Р.Л.Стратонович

ТЕОРИЯ ИНФОРМАЦИИ

М., «Сов. радио», 1975, 424 с.

Книга посвящена одному из главных направлений теоретической кибернетики. Дается систематическое изложение важнейших, ставших уже традиционными, результатов шенноновской теории информации, а также ряда новых вопросов, разработанных автором. К числу последних относятся теория ценности хартлиевского, больцмановского и шенноновского количеств информации, аппарат потенциальных функций, использующий параметры типа «температуры». Подчеркивается общность математического аппарата теории информации и статистической термодинамики. Содержание книги сгруппировано в соответствии с тремя вариационными задачами, характерными для теории информации.

Автор является крупным специалистом по случайным процессам, математической статистике и теории информации. Он опубликовал более ста оригинальных статей и три монографии.

Книга рассчитана на научных работников — специалистов в области кибернетики и статистической теории связи, а также аспирантов и студентов высших учебных заведений.

Оглавление

Предисловие	3
Введение	5
Глава 1. ОПРЕДЕЛЕНИЕ ИНФОРМАЦИИ И ЭНТРОПИИ ПРИ ОТСУТСТВИИ ПОМЕХ	**9**
§ 1.1. Определение энтропии в случае равновероятных возможностей	11
§ 1.2. Энтропия в случае неравновероятных возможностей и ее свойства	13
§ 1.3. Условная энтропия. Свойство иерархической аддитивности	16
§ 1.4. Асимптотическая эквивалентность неравновероятных возможностей равновероятным	21
§ 1.5. Асимптотическая равновероятность и энтропийная устойчивость	25
§ 1.6. Определение энтропии непрерывной случайной величины	30
§ 1.7. Свойства энтропии в обобщенной версии. Условная энтропия	37
Глава 2. КОДИРОВАНИЕ ДИСКРЕТНОЙ ИНФОРМАЦИИ ПРИ ОТСУТСТВИИ ПОМЕХ И ШТРАФОВ	**43**
§ 2.1. Основные принципы кодирования дискретной информации	44
§ 2.2. Основные теоремы для кодирования без помех. Равнораспределенные независимые сообщения	48
§ 2.3. Оптимальное кодирование по Хуфману. Примеры	53
§ 2.4. Погрешности кодирования без помех при конечной длине записи	57
Глава 3. КОДИРОВАНИЕ ПРИ НАЛИЧИИ ШТРАФОВ. ПЕРВАЯ ВАРИАЦИОННАЯ ЗАДАЧА	**61**
§ 3.1. Прямой способ вычисления информационной емкости записи для одного примера	62
§ 3.2. Дискретный канал без помех и его пропускная способность	64
§ 3.3. Решение первой вариационной задачи. Термодинамические	67

параметры и потенциалы
§ 3.4. Примеры применения общих методов вычисления пропускной способности 73
§ 3.5. Метод потенциалов в случае большего числа параметров 79
§ 3.6. Пропускная способность канала без шумов со штрафами в обобщенной версии 82
Глава 4. ПЕРВАЯ АСИМПТОТИЧЕСКАЯ ТЕОРЕМА И СВЯЗАННЫЕ С НЕЙ РЕЗУЛЬТАТЫ 84
§ 4.1. Потенциал Г или производящая функция семиинвариантов 85
§ 4.2. Некоторые асимптотические результаты статистической термодинамики. Устойчивость канонического распределения 89
§ 4.3. Асимптотическая эквивалентность двух видов ограничений 96
§ 4.4. Некоторые теоремы, касающиеся характеристического потенциала 101
Глава 5. ВЫЧИСЛЕНИЕ ЭНТРОПИИ ДЛЯ ЧАСТНЫХ СЛУЧАЕВ. ЭНТРОПИИ СЛУЧАЙНЫХ ПРОЦЕССОВ 109
§ 5.1. Энтропия отрезка стационарного дискретного процесса и удельная энтропия 110
§ 5.2. Энтропия марковской цепи 114
§ 5.3. Удельная энтропия части компонент дискретного марковского процесса и условного марковского процесса 120
§ 5.4. Энтропия гауссовых случайных величин 129
§ 5.5. Энтропия стационарной последовательности. Гауссова последовательность 134
§ 5.6. Энтропия случайных процессов в непрерывном времени. Общие понятия и соотношения 141
§ 5.7. Энтропия гауссового процесса в непрерывном времени 144
§ 5.8. Энтропия точечного случайного процесса 152
§ 5.9. Энтропия дискретного марковского процесса в непрерывном времени 162
§ 5.10. Энтропия диффузионных марковских процессов 165
§ 5.11. Энтропия комбинированного марковского процесса, условного процесса и части компонент марковского процесса 170
Глава 6. ИНФОРМАЦИЯ ПРИ НАЛИЧИИ ПОМЕХ. ШЕННОНОВСКОЕ КОЛИЧЕСТВО ИНФОРМАЦИИ 181
§ 6.1. Потери информации при вырожденных преобразованиях и при простых помехах 181
§ 6.2. Информация связи дискретных случайных величин 186
§ 6.3. Условная информация. Иерархическая аддитивность информации 189
§ 6.4. Количество информации связи в общем случае 195
§ 6.5. Информация связи гауссовых величин 198
§ 6.6. Удельная информация стационарных и стационарно связанных процессов. Гауссовы процессы 205
§ 6.7. Информация связи компонент марковского процесса 212
Глава 7. ПЕРЕДАЧА СООБЩЕНИЙ ПРИ НАЛИЧИИ ПОМЕХ. ВТОРАЯ АСИМПТОТИЧЕСКАЯ ТЕОРЕМА В РАЗЛИЧНЫХ 226

ФОРМУЛИРОВКАХ
§ 7.1. Принципы передачи и приема информации при наличии помех 227
§ 7.2. Случайный код и средняя вероятность ошибки 230
§ 7.3. Асимптотическая безошибочность декодирования. Теорема Шеннона (вторая асимптотическая теорема) 234
§ 7.4. Асимптотическая формула для вероятности ошибки 237
§ 7.5. Усиленные оценки для оптимального декодирования 241
§ 7.6. Некоторые общие соотношения между энтропиями и взаимными информациями при кодировании и декодировании 252
Глава 8. ПРОПУСКНАЯ СПОСОБНОСТЬ КАНАЛОВ. ВАЖНЫЕ ЧАСТНЫЕ СЛУЧАИ КАНАЛОВ 257
§ 8.1. Определение пропускной способности каналов 257
§ 8.2. Решение второй экстремальной задачи. Соотношения для пропускной способности и потенциала 260
§ 8.3. Вид оптимального распределения и статистическая сумма 267
§ 8.4. Симметричные каналы 270
§ 8.5. Двоичные каналы 272
§ 8.6. Гауссовы каналы 275
§ 8.7. Стационарные гауссовы каналы 285
§ 8.8. Аддитивные каналы 291
Глава 9. ОПРЕДЕЛЕНИЕ ЦЕННОСТИ ИНФОРМАЦИИ 296
§ 9.1. Уменьшение средних штрафов при уменьшении неопределенности 297
§ 9.2. Ценность хартлиевского количества информации. Пример 301
§ 9.3. Определение ценности шенноновского количества информации и а-информации 307
§ 9.4. Решение третьей вариационной задачи. Соответствующие ей потенциалы 311
§ 9.5. Решение вариационной задачи при некоторых дополнительных предположениях 320
§ 9.6. Ценность больцмановского количества информации 325
§ 9.7. Другой подход к определению ценности шенноновской информации 329
Глава 10. ЦЕННОСТЬ ШЕННОНОВСКОЙ ИНФОРМАЦИИ ДЛЯ ВАЖНЕЙШИХ БЕЙ-ЕСОВСКИХ СИСТЕМ 334
§ 10.1. Система с двумя состояниями 334
§ 10.2. Системы с однородной функцией штрафов 338
§ 10.3. Гауссовы бейесовские системы 344
§ 10.4. Стационарные гауссовы системы 353
Глава 11. АСИМПТОТИЧЕСКИЕ РЕЗУЛЬТАТЫ, КАСАЮЩИЕСЯ ЦЕННОСТИ ИНФОРМАЦИИ. ТРЕТЬЯ АСИМПТОТИЧЕСКАЯ ТЕОРЕМА 360
§ 11.1. О различии между ценностями различных родов информации. Предварительные формы 361
§ 11.2. Теорема об асимптотической равноценности различных количеств информации 365

§ 11.3. Быстрота исчезновения различия в ценности шенноновской и хартлиевской информации 376
§ 11.4. Другие способы записи основного результата. Обобщения и частные случаи 386
§ 11.5. Обобщенная теорема Шеннона 392
Глава 12. ТЕОРИЯ ИНФОРМАЦИИ И ВТОРОЙ ЗАКОН ТЕРМОДИНАМИКИ 398
§ 12.1. Информация о физической системе, находящейся в состоянии термодинамического равновесия. Обобщенный второй закон термодинамики 399
§ 12.2. Приток шенноновской информации и превращение теплоты в работу 402
§ 12.3. Энергетические затраты на создание и запись информации. Пример 407
§ 12.4. Энергетические затраты на создание и запись информации. Общая формулировка 410
§ 12.5. Энергетические затраты в физических каналах 413
Приложение. НЕКОТОРЫЕ МАТРИЧНЫЕ (ОПЕРАТОРНЫЕ) ТОЖДЕСТВА 416
§ П.1. Правило переноса оператора слева направо 416
§ П. 2. Детерминант составной матрицы 417
Список литературы 418
Предметный указатель 420

Предметный указатель

а-информация 308
Бит 13
Больцмана формула 10
Вероятность апостериорная финальная 123
— ошибки 229
— — средняя 230
Ветвь аномальная 265, 308
— нормальная 265, 308
Гиббса распределение каноническое 70, 85, 89, 398
— теорема 89
Длина записи 45, 51
W-процесс вторичный апостериорный 124
Емкость информационная 65
Задача вариационная первая 65, 67
— — вторая 257, 258, 259
— — третья 311
Закон сохранения количества информации 43
— термодинамики второй 400, 405
— — — обобщенный 400, 405
Иенсона неравенство 15
Информация связи парная (взаимная) 184
— — случайная 187
— — тройная 193
— — удельная 206
— — условная 189
Канал абстрактный 258, 260
— аддитивный 291
— гауссов 275
— — стационарный 285
— двоичный 272, 274
— дискретный без помех 64
— симметричный 270
— физический 413
— — пропускная способность 413
Код 45, 228
— дешифруемый 49

— — Крафта 49
— оптимальный 46
— Шеннона случайный 231
Кодирование информации 43, 44
— — «блочное» 44
— — оптимальное 43
— — скользящее (текущее) 44
Количество информации больцмановское 14, 326
— — хартлиевское 12
— — шенноновское 181, 186
Котельникова теорема 290
Леви формула 104
Лежандра преобразование 80, 88, 239
Маркова условие 122
Метод неопределенных множителей Лагранжа 67
«Микросостояние» 10
«Модель Изинга» 77
Нат 13
Негинформация 297
Область «активная» 67, 260, 314
Ошибка декодирования 58, 230
Параметр канонический 85, 86
— термодинамический внешний 79
— — внутренний 79, 85
— — сопряженный 80, 86
Плотность энтропии 165
Помеха простая 183
Последовательность бейесовских систем
информационно-устойчивая 374
Потенциал термодинамический 73
— характеристический 29, 86, 101, 133, 197, 203, 211
— — условный 249
Принцип аддитивности 12
Пропускная способность 61, 65, 258
Процесс вторичный апостериорный 119, 124
— дискретный 110
— — марковский 114
— — — стационарный 114
— — стационарный 110
— марковский диффузный 165
— — условный 120, 172, 215
— — —, энтропия 120
— стационарно-связанный 205
— стационарный периодический 145
— точечный случайный 152
Равноценность информации асимптотическая 367
Радона-Никодимова производная 34
Распределение каноническое 86
— экстремальное 308
Риск 307
Свободная энергия 70
Свойство иерархической аддитивности 19, 39, 190, 191
Система бейесовская 307
— — гауссовая 344
— — — стационарная 353
Случайный поток 152
Сообщение элементарное 47
Соотношение термодинамическое 70, 71, 263, 316
Стирлинга формула 158
Сумма статистическая 70, 82
Теорема асимптотическая вторая 234, 236
— — первая 99
— — третья 363, 367, 374
Условие мультипликативности 40, 42
— нормировки 67, 312
Условный марковский процесс 172, 215
Устойчивость информационная 236
— каноническая 95
— энтропийная 25, 28

— — достаточное условие 27
Фокера — Планка уравнение 165, 175
— — — стационарное 168
Функция правдоподобия 229
— семиинвариантов производящая 87
— ценности информации 303, 329
— штрафов 64, 303
Хартли формула 10. 13
Хинчина теорема 235
Ценность информации 296, 298, 303
— — больцмановской 326, 327
— — дифференциальная 298, 299
— — случайная 320
— — хартлиевской 304
— — шенноновской 308, 329
Цепь марковская 114
Чебышева неравенство 22
Чернова неравенство 102
Шеннона теорема 234, 259
— — обобщенная 360, 392, 394
Штраф средний 307
Энтропия 10, 11
— больцмановская 14
— конца отрезка 109, 113
— максимальное значение 15, 38
— непрерывной случайной величины 33, 34
— свойства 15, 16
— случайная 14
— удельная 23. 109, 111, 165
— условная 17, 39
Энергия свободная 70

Предисловие

Толчком к написанию настоящей книги послужило чтение автором курса лекций по теории информации на физическом факультете МГУ в 1963—1965 гг. Сначала книга писалась в соответствии с содержанием этих лекций. Планировалось построить ее так, чтобы отразить все важнейшие достижения шенноновской теории информации. Но в процессе работы над книгой автор «сбился» на более привычный для него стиль работы, в котором доминирующее место занимает самостоятельная разработка, вместо досконального штудирования имеющихся результатов. Это привело к включению в книгу оригинального материала и к оригинальной трактовке многих положений теории. Оригинальный материал вытеснил часть установившихся результатов, которые планировалось включить в книгу. Так, например, была опущена глава, посвященная известным конкретным способам кодирования и декодирования в каналах с помехами.

Включенный в книгу материал располагается по трем ступеням: первая, вторая и третья вариационные задачи и соответственно первая, вторая и третья асимптотические теоремы. Тем самым создается четкая панорама основного, наиболее принципиального содержания шенноновской теории информации.

Всякий стиль работы имеет свои достоинства и свои недостатки. Недостатком принятого стиля является то, что остается неотраженной (или недостаточно отраженной) работа многих ученых в данной области. Этот факт не следует расценивать как свидетельство недостаточного уважения к ним. Как правило, не дается оценка оригинальности материала, классификация авторской принадлежности результатов. Исключение делается лишь для немногих явных фактов.

В книге принят «принцип нарастающей сложности материала», заключающийся в том, что в начале книги (а также главы) помещается более простой и доступный материал. При этом от читателя не требуется знакомства с более сложным и специальным материалом, расположенным ближе к концу книги (главы). Такой принцип позволяет включать в книгу сложный материал, не делая ее недоступной для довольно широкого круга читателей. Хочется надеяться, что многие читатели почерпнут из книги кое-что полезное для себя.

При рассмотрении общих вопросов автор старался вести изложение с наибольшей возможной общностью. Для этого он иногда пользуется языком теории меры (например, распределение вероятностей записывает в виде $P(dx)$). Это не должно отпугивать тех, кто не владеет указанным языком. Дело в том, что, опуская детали, можно пользоваться простым словарем, переводящим эти «страшные» термины в привычные. Так, например, под «вероятностной мерой $P(dx)$» можно понимать вероятность $P(x)$ в случае дискретной случайной величины или произведение $p(x)dx$ для непрерывной случайной величины, где $p(x)$ — плотность распределения вероятностей, а $dx = dx_1 \ldots dx_r$ — дифференциал, соответствующий пространству размерности r.

Ориентируясь на различных читателей, автор при изложении материала не придавал значения выдержанности терминологии, считая, что безразлично сказать, допустим, «математическое ожидание» или «среднее значение», «вероятностная мера» или «распределение». Если пользоваться обобщенными функциями, то всегда существует плотность распределения вероятностей и ею можно пользоваться. Часто не возникает надобности различать знаки минимизации min и inf, имея в виду, что, если инфимум на рассматриваемом множестве не достигается, всегда можно «включить в действие» совершенно стандартную ε-процедуру и в сущности в результатах ничего не изменится. В тексте книги довольно свободно делается переход от дискретного вероятностного пространства к непрерывному. На несущественных деталях автор старался не задерживать внимание читателя, чтобы не отвлечь от главного.

Общие положения теории проиллюстрированы в книге многочисленными частными случаями и примерами. Вследствие общего значения теории и примеров при их изложении специальная радиофизическая терминология не употребляется. От читателя, интересующегося применением изложенного материала к проблеме передачи сообщений по радиоканалам, требуется умение наполнить абстрактные понятия радиофизическим содержанием. Так, например, при рассмотрении каналов с помехами (гл. 7) входной случайный процесс $x = \{x_t\}$ нужно трактовать как переменный информационный параметр сигнала $s(t, x_t)$, излучаемого радиопередатчиком, а выходной процесс $y = \{y_t\}$ — как сигнал на входе приемника. Правильная конкретизация понятий нужна для применения любой математической теории.

Автор выражает благодарность первому читателю этой книги Б. А. Гришанину, который оказал большую помощь при подготовке ее к печати, а также проф. В. И. Тихонову за обсуждение затронутых вопросов и ряд сделанных замечаний.

Введение

Термин «информация», указанный в заглавии книги, понимается здесь не в том широком смысле, в каком его понимают работники печати, радиовещания и народного хозяйства, а в том узком научном смысле, какой ему придал К. Шеннон. Другими словами, предметом настоящей книги является специальная математическая дисциплина — шенноновская теория информации, которая в состоянии решать не всеобъемлющие, а определенные, сравнительно специальные задачи, относящиеся к ее компетенции.

Содержанием этой дисциплины являются абстрактно формулируемые теоремы и результаты, которые по-разному могут конкретизироваться в различных отраслях знания. Теория информации имеет многочисленные приложения в теории передачи сообщений при наличии помех, в теории записывающих и регистрирующих устройств, в математической лингвистике и других науках, вплоть до генетики.

Теория информации вместе с другими математическими дисциплинами, такими, как теория оптимальных статистических решений, теория оптимального управления, теория алгоритмов и автоматов, теория игр и другие, входит в состав теоретической кибернетики — науки об управлении. По своему основному содержанию указанные дисциплины являются самостоятельными и не связанными между собой. Но это не значит, что они совершенно оторваны одна от другой, что мосты между ними не возможны. Без сомнения, возможно и вероятно появление комбинированных теорий, в которых используются понятия и результаты различных дисциплин и которые соединяют между собой различные дисциплины. Картина подобна деревьям в лесу: стволы их стоят отдельно, а кроны переплетаются. Первое время они растут независимо, а затем, переплетаясь ветвями, проникают друг в друга.

Конечно, в целом высказанное утверждение об объединении дисциплин является предположительным, однако срастание друг с другом некоторых первоначально оторванных дисциплин является уже реальностью, свершившимся фактом. Как видно из ряда работ и из предлагаемой книги, срастаются три дисциплины:

1) статистическая термодинамика как математическая теория,

2) шенноновская теория информации,

3) теория оптимальных статистических решений (вместе с ее многошаговыми (последовательностными) разновидностями — оптимальной фильтрацией и динамическим программированием).

Из содержания настоящей книги будет видно, что цементом, объединяющим указанные три дисциплины, являются «термодинамические» методы с такими типичными их атрибутами, как «термодинамические» параметры и потенциалы, преобразования Лежандра, экстремальные распределения, асимптотический характер важнейших теорем.

Статистическую термодинамику лишь условно можно относить к кибернетическим дисциплинам. Однако в некоторых вопросах статистической термодинамики ее кибернетическая подоплека выступает довольно отчетливо. Достаточно вспомнить второй закон термодинамики и «демона Максвелла», который является типичным автоматом, перерабатывающим информацию в физическую энтропию. Информация является «горючим» для вечного двигателя второго рода. Эти вопросы рассматриваются в гл. 12.

Если статистическую термодинамику считать кибернетической дисциплиной, то Л. Больцмана и Дж. К. Максвелла следует назвать первыми выдающимися кибернетиками. Важно учесть, что формулу, по которой энтропия выражается через вероятности, ввел Л. Больцман. Он также дал распределение вероятностей, которое является решением первой вариационной задачи (как называются входящие в формулу функции — энергией или штрафами, разумеется, совершенно безразлично).

При возникновении шенноновской теории информации появление в ней такого хорошо известного в термодинамике понятия как энтропия воспринималось некоторыми как курьез, и этому не придавалось серьезного значения. Считалось, что эта энтропия не имеет ничего общего с физической энтропией (вопреки деятельности вышеупомянутого «демона»). В связи с этим можно вспомнить бесчисленное множество кавычек, в которые заключено слово «энтропия» в первом издании русского перевода статей К. Шеннона (сборник под ред. А. Н. Железнова, ИЛ, 1953). Я думаю, что теперь даже такой термин как «температура» в теории информации можно писать без кавычек, понимая под этим просто параметр, определенным образом входящий в выражение для экстремального распределения. Однотипные закономерности имеют место и в теории информации и в статистической физике, и их условно можно называть «термодинамическими».

Сначала (с 1948 г. по 1959 г.) в шенноновской теории информации фигурировало лишь одно «термодинамическое» понятие — энтропия. В ней, казалось, нет места для энергии и других аналогичных термодинамических потенциалов. В этом отношении теория выглядела однобокой по сравнению со статистической термодинамикой. Но это было временным явлением. После осознания того, что в прикладной теории информации, понимаемой как теория передачи сигналов, аналогом энергии является функция штрафов, аналогом средней энергии является риск, положение изменилось. Стало очевидным сходство в числе основных понятий и в соотношениях между ними. Если, в частности, иметь в виду первую вариационную задачу, то можно говорить о подобии, «изоморфизме», данных двух теорий. Математические соотношения между соответствующими понятиями в обеих дисциплинах одни и те же, а они-то и составляют содержание математической теории — той теории, которая рассматривается в этой книге.

Указанными соотношениями не исчерпывается содержание теории информации. Помимо энтропии, в ней имеются и другие понятия, такие, как шенноновское количество информации. Кроме первой вариационной задачи, связанной с экстремумом энтропии при фиксированном риске — энергии, в ней возможны также вариационные задачи, в которых энтропия заменяется на шенноновское количество информации. Поэтому содержание теории информации шире математического содержания статистической термодинамики.

Новые вариационные задачи обнаруживают замечательную аналогию с первой вариационной задачей. В них имеется тот же набор сопряженных параметров и потенциалов, те же формулы связи потенциалов посредством преобразования Лежандра. И это не удивительно. Можно показать, что все это появляется при рассмотрении любой невырожденной вариационной задачи.

Остановимся на этом подробнее. Пусть заданы хотя бы два функционала $\Phi_1[P]$, $\Phi_2[P]$ от аргумента («распределения») P, и требуется обратить в экстремум один функционал при фиксированном значении второго, скажем,

$$\Phi_1[P] = \underset{P}{\mathrm{extr}} \quad \text{при} \quad \Phi_2[P] = A.$$

Вводя множитель Лагранжа α, который служит «термодинамическим» параметром, канонически сопряженным с параметром A, будем исследовать экстремум выражения

$$K = \Phi_1[P] + \alpha\Phi_2[P] = \underset{P}{\mathrm{extr}} \tag{1}$$

при фиксированном значении α. Экстремальное значение

$$K(\alpha) = \Phi_1[P^{\mathrm{extr}}] + \alpha\Phi_2[P^{\mathrm{extr}}] = \Phi_1 + \alpha A$$

служит «термодинамическим» потенциалом, так как $dK = d\Phi_1 + \alpha d\Phi_2 + \Phi_2 d\alpha = \Phi_2 d\alpha$, т. е.

$$dK/d\alpha = A. \tag{2}$$

Использованное здесь соотношение

$$d\Phi_1 + \alpha d\Phi_2 \equiv [\Phi_1(P^{\mathrm{extr}} + \delta P) - \Phi_1(P^{\mathrm{extr}})] + \alpha[\Phi_2(P^{\mathrm{extr}} + \delta P) - \Phi_2(P^{\mathrm{extr}})] = 0$$

следует из (1). Здесь взята частная вариация $\delta P = P^{\mathrm{extr}}(\alpha + d\alpha) - P^{\mathrm{extr}}(\alpha)$, обусловленная приращением параметра A, т. е. параметра α. Функция $L(A) = K - \alpha A = K - \alpha(dK/d\alpha)$ является потенциалом, сопряженным по Лежандру с $K(\alpha)$. При этом из (2) имеем

$$dL(A)/dA = -\alpha.$$

Эти обычные в термодинамике соотношения справедливы, очевидно, независимо от природы функционалов Φ_1, Φ_2 и аргумента P.

Излагаемые в данной книге результаты теории информации связаны с тремя вариационными задачами, решение которых приводит к ряду соотношений, параметров и потенциалов. Вариационные задачи играют в теоретической кибернетике большую роль потому, что в ней рассматриваются в первую очередь *оптимальные* конструкции и процедуры.

Как видно из содержания книги, с рассмотренными вариационными задачами связаны также важнейшие закономерности (теоремы), носящие асимптотический характер, т. е. справедливые для больших составных систем. Первой вариационной задаче соответствует факт устойчивости канонического распределения, который существен для статистической физики, и факт асимптотической эквивалентности ограничений, наложенных на точные и средние значения функции.

Со второй вариационной задачей связан известный результат Шеннона об асимптотической безошибочности передачи информации через каналы с помехами, а с третьей — факт асимптотической равноценности шенноновской и хартлиевской информации. Последние результаты являются замечательным примером единства дискретного и непрерывного, примером того, как при повышении сложности дискретной системы ее удобно описывать непрерывными математическими объектами; того, как сложная непрерывная система ведет себя асимптотически подобно сложной дискретной системе. Заманчиво было бы видеть нечто подобное, скажем, в будущей асимптотической теории алгоритмов или автоматов.

Глава 1

ОПРЕДЕЛЕНИЕ ИНФОРМАЦИИ И ЭНТРОПИИ ПРИ ОТСУТСТВИИ ПОМЕХ

В современной науке, технике и общественной жизни большую роль играет информация и связанные с ней операции: получение информации, передача информации, переработка ее, хранение и т. п.). Значение информации, по-видимому, переросло значение другого важного фактора, который играл доминирующую роль в прошлом веке, а именно — энергии.

В будущем, в связи с усложнением науки, техники, экономики и других отраслей значение правильного управления ими будет все возрастать, и поэтому будет возрастать значение информации.

Что же такое информация, возможна ли теория информации, существует ли для информации какие-либо общие закономерности, не зависящие от конкретного содержания информации, которое может быть весьма различным? Ответы на эти вопросы далеко не очевидны. Информация является более трудным для исследования понятием, чем, скажем, энергия, занимающая определенное, давно выясненное место в физике.

Информация имеет две стороны: количественную и качественную. Иногда важным является общее количество информации, а иногда — качественный вид сообщения, его конкретное содержание. Кроме того переработка информации из одного вида в другой является технически более сложной задачей, чем, скажем, превращение энергии из одной формы в другую. Все это затрудняет разработку теории информации и ее использование. Не исключено, что во многих практических задачах ситуация такова, что обращение к общей теории информации не принесет пользы, и их следует решать независимыми инженерными методами.

И тем не менее общая теория информации существует, существуют такие образцовые ситуации и задачи, в которых основную роль играют закономерности общей теории. Поэтому теория информации важна и с практической точки зрения, не говоря уже о ее большой принципиальной роли, философской роли, роли в формировании кругозора исследователя.

Из сказанного видно, насколько нелегким делом было открытие закономерностей теории информации. Важнейшим этапом в этом отношении явились работы К. Шеннона [1, 2], опубликованные в 1948—1949 гг. И по постановке задачи, и по результатам они были многими восприняты как неожиданность. Более внимательное осмысливание, однако, приводит к заключению, что новая теория продолжает и развивает прежние идеи, а именно, идеи статистической термодинамики, связанные с именем Л. Больцмана. Не случайной является глубокая общность математического аппарата этих двух направлений, доходящая до прямого совпадения формул (например, для энтропии дискретных случайных величин). Кроме того, логарифмическая мера для количества информации, являющаяся исходной в теории Шеннона, была предложена применительно к задачам связи еще в 1928 г. в работе Р. Хартли [1].

В настоящей главе мы введем эту логарифмическую меру количества информации и изложим ряд вытекающих из нее важных свойств информации, таких, как свойство аддитивности.

Понятие количества информации тесно связано с понятием энтропии, являющейся мерой неопределенности. Приобретение информации сопровождается уменьшением неопределенности, поэтому количество информации можно измерять количеством исчезнувшей неопределенности, т. е. энтропии.

В случае дискретного сообщения, т. е. дискретной случайной величины, энтропия определяется формулой Больцмана

$$H_\xi = -\sum_{\xi} P(\xi) \ln P(\xi),$$

где ξ — случайная величина, а $P(\xi)$ — ее распределение вероятностей.

В данной главе большое значение придается тому, что эта формула является следствием (в асимптотическом смысле) более простой формулы Хартли $H = \ln M$.

Важным результатом теории информации является тот факт, что множество реализаций энтропийно устойчивой (это понятие определено в § 1.5) случайной величины можно разбить на два подмножества. Первое из них имеет исчезающе малую вероятность, и поэтому его можно отбросить. Второе подмножество содержит приблизительно e^{H_ξ} реализации (т. е. вариантов записи), оно часто значительно меньше полного числа реализаций. Реализация этого второго подмножества приближенно можно считать равновероятными. Следуя терминологии, введенной Больцманом, эти равновероятные реализации можно называть «микросостояниями».

Теория информации является математической теорией, использующей понятия и методы теории вероятностей. При ее изложении нет особой надобности придерживаться специальной «информационной» терминологии, применяемой обычно в приложениях. Понятие «сообщения» без ущерба для теории можно заменить на понятие «случайной величины», понятие «последовательности сообщений» — на «случайный процесс» и т. п. Тогда для применения общей теории, разумеется, требуется правильная формализация прикладных понятий и перевод их с прикладного языка на язык теории вероятностей. Такому переводу мы не будем здесь уделять особого внимания.

Основные результаты, изложенные в § 1.1—1.3 и относящиеся к дискретным случайным величинам (или процессам), которые имеют конечное или счетное число состояний, могут быть обобщены на случай непрерывных (и вообще произвольных) величин, принимающих значения из многомерного действительного пространства. При таком обобщении приходится преодолевать известные трудности, мириться с определенным усложнением важнейших понятий и формул. Основное усложнение заключается в следующем. Если применительно к дискретным случайным величинам для определения энтропии было достаточно одной меры, одного распределения вероятностей, то в общем случае для определения энтропии необходимо ввести две меры. Поэтому энтропия теперь характеризует не одну меру (степень «неопределенности» в ней), а две меры, и характеризует тем самым соотношения между этими двумя мерами. В нашем изложении формула для энтропии в обобщенной версии аргументируется (выводится из больцмановской формулы) на примере уплотнения точек, изображающих случайные величины.

Специально теории информации посвящены вышедшие на русском языке книги Голдмана [1], Файнстейна [1], Фано [1], по которым удобно знакомиться с простейшими понятиями (см. также книгу Кульбака [1]).

1.1. Определение энтропии в случае равновероятных возможностей

Пусть имеется M равноправных и, следовательно, равновероятных возможностей. Например, при бросании правильной игральной кости $M = 6$. Конечно, не во всех примерах формализация условий проводится так просто и определенно, как в случае игральной кости. Мы предполагаем, тем не менее, что она проведена, что действительно осуществляется одна из M возможностей и никакая иная, что эти возможности равноправны. Тогда имеется априорная неопределенность, прямо связанная с M (т. е. чем больше M, тем больше неопределенность). Измеряющая ее численная величина носит название *энтропии* и обозначается H:

$$H = f(M), \tag{1.1.1}$$

где $f(\cdot)$ — некоторая возрастающая неотрицательная функция, определенная по меньшей мере для чисел натурального ряда.

При бросании кости и выяснении выпавшего числа приходит информация, количество которой обозначим I. После этого (т. е. апостериори) никакой неопределенности не остается: апостериорное $M = 1$ и этому значению должно соответствовать $H_{ps} = f(1) = 0$. Количество пришедшей информации естественно измерять величиной исчезнувшей неопределенности:

$$I = H_{pr} - H_{ps}. \tag{1.1.2}$$

Здесь индекс pr означает «априори», а ps — «апостериори».

Мы видим, что пришедшее количество информации I совпадает с первоначальной энтропией. Также и в других случаях (в частности для приводимой в дальнейшем формулы (1.2.3)) сообщение, имеющее энтропию H, может передавать количество информации I, равное H.

Для определения вида функции $f(\cdot)$ в (1.1.1.) используем вполне естественный *принцип аддитивности*. В применении к игральной кости он гласит: энтропия двух бросаний кости в два раза больше, чем энтропия одного бросания, трех бросаний — в три раза больше и т. д. В применении к другим примерам данный принцип указывает, что энтропия нескольких несвязанных систем, рукояток управления и т. п. равна сумме энтропий отдельных систем, рукояток управления и т. п. Но число равноправных возможностей сложной системы M равно произведению числа возможностей m каждой из подсистем («простых» по отношению к сложной). При двух бросаниях игральной кости число различных пар (ξ_1, ξ_2) (где ξ_1 принимает одно из шести значений и ξ_2—также одно из шести значений) равно $36 = 6^2$. Вообще, в случае n бросаний число равноправных возможностей равно 6^n. Применяя формулу (1.1.1) для этого числа, получаем энтропию $f(6^n)$. Согласно принципу аддитивности находим

$$f(6^n) = nf(6).$$

При других $m > 1$ эта формула имела бы вид

$$f(m^n) = n\,f(m). \tag{1.1.3}$$

Обозначая $x = m^n$, имеем $n = \ln x/\ln m$ и из (1.1.3) получаем

$$f(x) = K \ln x, \tag{1.1.4}$$

где $K = f(m)/\ln m$ — положительная константа, не зависящая от x. Она связана с выбором единиц информации. Итак, вид функции $f(\cdot)$ определен с точностью до выбора единиц измерения. Легко проверить, что отмеченное ранее условие $f(1) = 0$ действительно выполняется.

Впервые логарифмическую меру информации ввел Хартли [1], поэтому величину $H = K\ln M$ называем *хартлиевским количеством информации*.

Укажем три основных выбора единиц измерения информации:

1) если в (1.1.4) положить $K = 1$, то энтропия будет измеряться в натуральных единицах (натах)[1]:

$$H_{\text{нат}} = \ln M; \tag{1.1.5}$$

2) если положить $K = 1/\ln 2$, то будем иметь энтропию, выраженную в двоичных единицах (битах)[2]:

$$H_{\text{бит}} = \frac{1}{\ln 2} \ln M = \log_2 M; \tag{1.1.6}$$

3) наконец, мы будем иметь физическую шкалу, если в качестве K возьмем постоянную Больцмана $k = 1{,}38 \cdot 10^{-23}$ Дж/град. Энтропия, измеренная в этих единицах, будет

$$S = H_{\text{физ}} = k \ln M. \tag{1.1.7}$$

Из сопоставления (1.1.5) и (1.1.6) легко видеть, что 1 нат крупнее 1 бита в $\log_2 e = 1/\ln 2 \approx 1{,}44$ раза.

В дальнейшем, если не будет специальных оговорок, будем пользоваться натуральными единицами (формула (1.1.5)), опуская индекс «нат».

Пусть указанные равноправные возможности заключаются в том, что случайная величина ξ принимает одно из M значений, скажем $1, \ldots, M$. Вероятность каждого отдельного ее значения тогда равна $P(\xi) = 1/M$, $\xi = 1, \ldots, M$. Следовательно, формулу (1.1.5) можно записать

$$H = -\ln P(\xi). \tag{1.1.8}$$

1.2. Энтропия в случае неравновероятных возможностей и ее свойства

1. Пусть теперь вероятности различных возможностей (реализаций) не равны друг другу. Если, как и раньше, число возможностей равно M, можно рассматривать случайную величину ξ, принимающую одно из M значений. Взяв в качестве ξ номер возможности, получим, что эти значения равны $1, \ldots, M$. Вероятности $P(\xi)$ этих значений неотрицательны и подчинены условию нормировки: $\sum_{\xi} P(\xi) = 1$.

Если применить формально равенство (1.1.8) к этому случаю, то получим, что каждому значению ξ соответствует, вообще говоря, своя энтропия

$$H(\xi) = -\ln P(\xi). \tag{1.2.1}$$

---

[1] «нат» — сокращение английских слов natural digit, что означает «натуральная единица».

[2] «бит» — сокращение слов binary digit, означающих «двоичная единица (знак)».

Тем самым мы приписываем определенное значение энтропии каждой реализации величины ξ. Поскольку ξ — случайная величина, то эту энтропию можно рассматривать как случайную величину.

Как и в § 1.1, апостериорная энтропия, имеющаяся после выяснения реализации ξ, равна нулю. Поэтому информация, получаемая при выяснении реализации, численно равна первоначальной энтропии

$$I(\xi) = H(\xi) = -\ln P(\xi). \tag{1.2.2}$$

Она, как и $H(\xi)$, оказывается зависящей от вида реализации сообщения (от значения ξ), т. е. является случайной. Из этой формулы видно, что информация и энтропия велики, когда априорная вероятность данной реализации мала, и наоборот. Это вполне соответствует интуитивным представлениям.

Пример. Допустим, что нам интересно знать, сдал или не сдал экзамен данный студент. Примем следующие вероятности этих двух событий:

$$P(\text{сдал}) = {}^7/_8, \quad P(\text{не сдал}) = {}^1/_8.$$

Отсюда видно, что этот студент является довольно сильным. Если нам сообщили, что он сдал экзамен, мы вправе сказать: «Ваше сообщение мне мало что дало, я и без этого предполагал, что он сдал». Количественно по формуле (1.2.2) информация этого сообщения равна

$$I(\text{сдал}) = \log_2 ({}^8/_7) = 0{,}193 \text{ бита}.$$

Если нам сообщили, что не сдал, мы скажем: — «Неужели?» и почувствуем, что в большей степени обогатились знаниями. Количество информации такого сообщения равно

$$I(\text{не сдал}) = \log_2 8 = 3 \text{ бита}.$$

В теории, однако, большую роль играет не случайная энтропия (соответственно информация) (1.2.1), (1.2.2), а усредненная энтропия, определяемая формулой

$$H_\xi = \mathrm{M}H(\xi) = -\sum_\xi P(\xi) \ln P(\xi). \tag{1.2.3}$$

Будем называть ее *больцмановской* энтропией или больцмановской информацией. В приведенном выше примере усреднение по обоим сообщениям дает

$$I_\xi = H_\xi = \frac{7}{8}\ 0{,}193 + \frac{3}{8} = 0{,}544 \text{ бита}.$$

Случайная величина ξ, стоящая в индексе символа H_ξ (в отличие от величины, стоящей в скобках в $H(\xi)$), является «слепой», т. е. энтропия H_ξ описывает случайную величину ξ (зависит от вероятностей P_ξ), но не зависит от значения ξ, от реализации ξ. Такая система обозначений в расширенном виде, включающем условные энтропии, будет нами применяться и далее.

Неопределенность типа $0 \ln 0$, встречающаяся в (1.2.3), когда отдельные вероятности равны нулю, всегда понимается в смысле $0 \ln 0 = 0$. Вследствие этого множество из M возможностей всегда можно дополнить любыми возможностями нулевой вероятности, а также в индексе энтропии дописать детерминированные (не случайные) величины. Например, справедливо равенство $H_\xi = H_{\xi,7}$ при любой случайной величине ξ.

2. Свойства энтропии.

Теорема 1.1. *Как случайная, так и средняя энтропия всегда неотрицательны.*

Это свойство связано с тем, что вероятность не может превзойти единицу и с тем, что постоянная K в (1.1.4) берется обязательно положительной. Поскольку $P(\xi) \leqslant 1$, то $-\ln P(\xi) = H(\xi) \geqslant 0$. Это неравенство сохраняется, конечно, и после усреднения.

Теорема 1.2. *Энтропия имеет максимальное значение, равное* $\ln M$, *когда возможности (реализации) равновероятны, т. е. когда* $P(\xi) = 1/M$.

Доказательство. Это свойство является следствием неравенства Иенсена (см., например, Рао [1])

$$\mathbf{M}f(\zeta) \leqslant f(\mathbf{M}\zeta), \tag{1.2.4}$$

справедливого для любой выпуклой (вверх) функции $f(x)$. (Функция $f(x) = \ln x$ является выпуклой при $x > 0$, поскольку $f''(x) = -x^{-2} < 0$). В самом деле, обозначая $\zeta = 1/P(\xi)$, имеем

$$\mathbf{M}\zeta = \mathbf{M}\frac{1}{P(\xi)} = \sum_{\xi=1}^{M} \frac{P(\xi)}{P(\xi)} = M, \tag{1.2.5}$$

$$\mathbf{M}f(\zeta) = \mathbf{M}\ln\frac{1}{P(\xi)} = \mathbf{M}H(\xi) = H_\xi. \tag{1.2.6}$$

Подставляя (1.2.5), (1.2.6) в (1.2.4), получаем

$$H_\xi \leqslant \ln M.$$

Для частного вида функции $f(\zeta) = \ln\zeta$ неравенство (1.2.4) легко проверить непосредственно. Усредняя очевидное неравенство

$$\ln\frac{\zeta}{\mathbf{M}\zeta} \leqslant \frac{\zeta}{\mathbf{M}\zeta} - 1, \tag{1.2.7}$$

получаем (1.2.4).

В общем случае для доказательства (1.2.4) удобно рассмотреть касательную $f(\mathbf{M}\zeta) + f'(\mathbf{M}\zeta)(\zeta - \mathbf{M}\zeta)$ к функции $f(\zeta)$ в точке $\zeta = \mathbf{M}\zeta$. Вследствие выпуклости имеем

$$f(\zeta) \leqslant f(\mathbf{M}\zeta) + f'(\mathbf{M}\zeta)(\zeta - \mathbf{M}\zeta).$$

Усредняя это неравенство, получаем (1.2.4).

Как видно из приведенного доказательства, теорема 1.2 осталась бы справедливой, если в определении энтропии логарифмическую функцию заменить на любую другую выпуклую функцию. Перейдем к свойствам энтропии, которые являются специфическими для логарифмической функции, а именно, к свойствам, связанным с аддитивностью энтропии.

Теорема 1.3. *Если случайные величины ξ_1, ξ_2 независимы, то полная (совместная) энтропия $H_{\xi_1\xi_2}$ распадается на сумму энтропий:*

$$H_{\xi_1\xi_2} = H_{\xi_1} + H_{\xi_2}. \tag{1.2.8}$$

Доказательство. Допустим, что имеются две случайные величины ξ_1, ξ_2, первая из которых принимает значения 1, ..., m_1, а вторая — значения 1, ..., m_2. Существует $m_1 m_2$ пар $\xi = (\xi_1, \xi_2)$, причем они имеют вероятности $P(\xi_1, \xi_2)$. Нумеруя пары в произвольном порядке номером $\xi = 1, \ldots, m_1 m_2$, имеем

$$H_\xi = -\mathbf{M} \ln P(\xi) = -\mathbf{M} \ln P(\xi_1, \xi_2) = H_{\xi_1\xi_2}.$$

Вследствие независимости имеем

$$P(\xi_1, \xi_2) = P(\xi_1) P(\xi_2).$$

Поэтому $\ln P(\xi_1, \xi_2) = \ln P(\xi_1) + \ln P(\xi_2)$; $H(\xi_1 \xi_2) = H(\xi_1) + H(\xi_2)$. Усреднение последнего равенства дает $H_\xi = H_{\xi_1\xi_2} = H_{\xi_1} + H_{\xi_1}$, т. е. (1.2.7). Доказательство закончено.

Если имеются не две независимые случайные величины, а две группы $(\eta_1, \ldots, \eta_r$ и $\zeta_1, \ldots, \zeta_s)$ независимых величин $(P(\eta_1, \ldots, \eta_r, \zeta_1, \ldots, \zeta_s) = P(\eta_1, \ldots, \eta_r) P(\zeta_1, \ldots, \zeta_s))$, то приведенное рассуждение остается применимым при обозначении совокупности $(\eta_1, \ldots, \eta_r)$ или ее номера через ξ_1, а $(\zeta_1, \ldots, \zeta_s)$ — через ξ_2.

Свойство, указанное в теореме 1.3, является проявлением принципа аддитивности, который был взят нами за основу в § 1.1 и привел к логарифмической функции (1.1.1). Оно обобщается на случай нескольких независимых случайных величин $\xi_1, \ldots, \xi_n$. При этом

$$H_{\xi_1 \ldots \xi_n} = \sum_{j=1}^{n} H_{\xi_j}, \tag{1.2.9}$$

что легко доказывается аналогичным способом.

1.3. Условная энтропия. Свойство иерархической аддитивности

Обобщим формулы (1.2.1), (1.2.3) на случай условных вероятностей. Пусть имеются случайные величины $\xi_1, \ldots, \xi_n$, описываемые совместным распределением $P(\xi_1, \ldots, \xi_n)$. Условной вероятности

$$P(\xi_k, \ldots, \xi_n \mid \xi_1, \ldots, \xi_{k-1}) = \frac{P(\xi_1, \ldots, \xi_n)}{P(\xi_1, \ldots, \xi_{k-1})} \quad (k \leqslant n)$$

сопоставляем случайную условную энтропию

$$H(\xi_k, \ldots, \xi_n \mid \xi_1, \ldots, \xi_{k-1}) = -\ln P(\xi_k, \ldots, \xi_n \mid \xi_1, \ldots, \xi_{k-1}). \quad (1.3.1)$$

Введем особое обозначение для результата усреднения ее по ξ_k, $\ldots$, ξ_n:

$$H_{\xi_k \ldots \xi_n}(\mid \xi_1, \ldots, \xi_{k-1}) = -\sum_{\xi_k \ldots \xi_n} P(\xi_k, \ldots, \xi_n \mid \xi_1, \ldots, \xi_{k-1}) \times$$
$$\times \ln P(\xi_k, \ldots, \xi_n \mid \xi_1, \ldots, \xi_{k-1}), \quad (1.3.2)$$

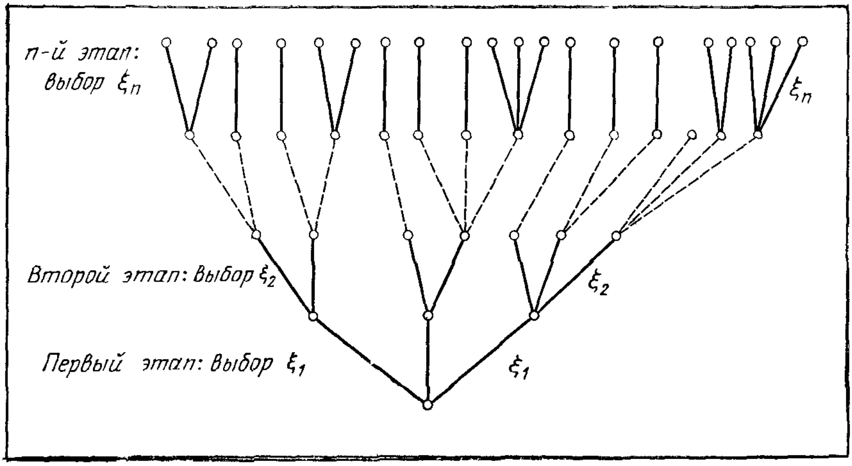


Рис. 1.1.
Этапы разбиения и «дерево выборов» в общем случае.

также для результата полного усреднения:

$$H_{\xi_k \ldots \xi_n \mid \xi_1 \ldots \xi_{k-1}} = \mathbf{M}H(\xi_k, \ldots, \xi_n \mid \xi_1, \ldots, \xi_{k-1}) =$$
$$= -\sum_{\xi_1 \ldots \xi_n} P(\xi_1, \ldots, \xi_n) \ln P(\xi_k, \ldots, \xi_n \mid \xi_1, \ldots, \xi_{k-1}). \quad (1.3.3)$$

Если еще менять k и n, то мы будем иметь большое число различных энтропий, условных и безусловных, случайных и неслучайных. Между ними существуют определенные тождественные соотношения, которые рассматриваются ниже.

Прежде чем перейти к формулировке основного иерархического тождества (1.3.4), покажем, как можно ввести иерархическое множество случайных величин ξ_1, $\ldots$, ξ_n, даже если первоначально была лишь одна случайная величина ξ.

Пусть ξ принимает одно из M значений с вероятностями $P(\xi)$. Выбор одной реализации будем производить в несколько этапов. На первом этапе пусть указывается принадлежность тому или иному подмножеству из полного набора неперекрывающихся подмножеств

$E_1, \ldots, E_{m_1}$. Пусть ξ_1 указывает номер этого подмножества. На втором этапе каждое подмножество делится на более мелкие подмножества $E_{\xi_1 \xi_2}$. Вторая случайная величина ξ_2 указывает, какому более мелкому подмножеству принадлежит реализация случайной величины. Эти более мелкие подмножества делятся, в свою очередь, до тех пор, пока мы не получим подмножества, состоящие из одного элемента. Число n нетривиальных этапов разбиения, очевидно, не может превосходить $M - 1$. Фиксированной схеме разбиений можно сопоставить «дерево», изображенное на рис. 1.1. Дальнейшие рассуждения будут относиться к какому-то одному выбранному «дереву». Чтобы указать реализацию ξ, необходимо и достаточно

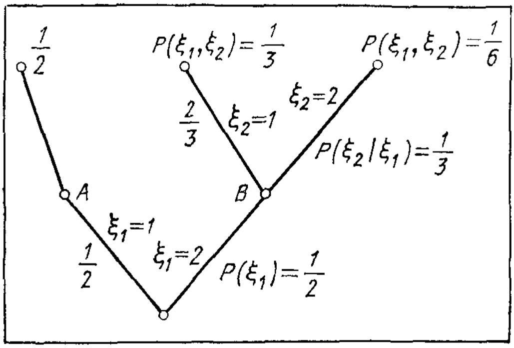


Рис. 1.2. «Дерево выборов» для одного конкретного примера.

фиксировать совокупность реализаций $(\xi_1, \xi_2, \ldots, \xi_n)$. Чтобы указать «узел» $(k+1)$-го этапа, нужно указать значения $\xi_1, \ldots, \xi_k$. Тогда значение ξ_{k+1} указывает ту ветвь, по которой мы будем выходить из этого узла.

В качестве примера «дерева» можно указать простое «дерево», изображенное на рис. 1.2, которое рассматривал Шеннон.

С выбором на каждом этапе связана некоторая неопределенность. Рассмотрим соответствующую ей энтропию. Первый этап имеет один узел, и энтропия выбора равна H_{ξ_1}. Фиксация ξ_1 определяет узел второго этапа. Вероятность пойти по ветви ξ_2 из узла ξ_1 равна условной вероятности

$$P(\xi_2 \mid \xi_1) = P(\xi_1, \xi_2)/P(\xi_1).$$

Энтропия, связанная с выбором одной ветви из этого узла, есть не что иное, как условная энтропия типа (1.3.2) при $n = 2$, $k = 2$:

$$H_{\xi_2}(\mid \xi_1) = -\sum_{\xi_2} P(\xi_2 \mid \xi_1) \ln P(\xi_2 \mid \xi_1).$$

Усреднение по всем узлам второго этапа, как и в (1.3.3), дает полную энтропию выбора на втором этапе:

$$H_{\xi_2 \mid \xi_1} = \mathrm{M} H_{\xi_2}(\mid \xi_1) = \sum_{\xi_1} P(\xi_1) H_{\xi_2}(\mid \xi_1).$$

Вообще энтропия выбора на k-м этапе в узле, который определяется значениями $\xi_1, \ldots, \xi_{k-1}$, равна

$$H_{\xi_k}(\,|\,\xi_1, \ldots, \xi_{k-1}),$$

а полная энтропия k-го этапа

$$H_{\xi_k|\xi_1 \ldots \xi_{k-1}} = \mathbf{M}H_{\xi_k}(\,|\,\xi_1, \ldots, \xi_{k-1}).$$

Для примера на рис. 1.2 энтропия первого этапа равна $H_{\xi_1} = 1$ бит. Узел A имеет энтропию $H_{\xi_2}(\,|\,\xi_1 = 1) = 0$, а узел B — энтропию $H_{\xi_2}(\,|\,\xi_1 = 2) = \frac{2}{3}\log\frac{3}{2} + \frac{1}{3}\log 3 = \log_2 3 - \frac{2}{3}$ бита.

Средняя энтропия второго этапа, очевидно, равна $H_{\xi_2|\xi_1} = \frac{1}{2}\, H_{\xi_2}(\,|2) = \frac{1}{2}\log_2 3 - \frac{1}{3}$ бита. Важная закономерность состоит в том, что сумма энтропий всех этапов равна полной энтропии H_ξ, которую можно вычислить без разбиений выбора на этапы. Для указанного примера

$$H_{\xi_1\xi_2} = \frac{1}{2}\log 2 + \frac{1}{3}\log 3 + \frac{1}{6}\log 6 = \frac{2}{3} + \frac{1}{2}\log_2 3 \text{ бита},$$

что, действительно, совпадает с суммой $H_{\xi_1} + H_{\xi_2|\xi_1}$. Это имеет место не только для данного примера.

Теорема 1.4. *Энтропия обладает свойством иерархической аддитивности*:

$$H_{\xi_1,\ldots,\xi_n} = H_{\xi_1} + H_{\xi_2|\xi_1} + H_{\xi_3|\xi_1\xi_2} + \ldots + H_{\xi_n|\xi_1,\ldots,\xi_{n-1}}. \quad (1.3.4)$$

Доказательство. По определению условных вероятностей они обладают следующим свойством иерархической мультипликативности:

$$P(\xi_1, \ldots, \xi_n) = P(\xi_1)\,P(\xi_2\,|\,\xi_1)\,P(\xi_3\,|\,\xi_1, \xi_2) \ldots P(\xi_n\,|\,\xi_1, \ldots, \xi_{n-1}). \quad (1.3.5)$$

Логарифмируя (1.3.5) и учитывая определение (1.3.1) случайной условной энтропии, имеем

$$H(\xi_1, \ldots, \xi_n) = H(\xi_1) + H(\xi_2\,|\,\xi_1) + H(\xi_3\,|\,\xi_1\xi_2) + \ldots + H(\xi_n\,|\,\xi_1, \ldots, \xi_{n-1}). \quad (1.3.6)$$

Усредняя это равенство в соответствии с (1.3.3), получаем равенство (1.3.4). Доказательство закончено.

Свойство, о котором идет речь в теореме 1.4, является проявлением того простого принципа аддитивности, который был принят в § 1.1. Оно является следствием выбора логарифмической функции в (1.1.1). Легко понять, что из этого свойства вытекает свойство аддитивности (1.2.8), (1.2.9). В самом деле, для независимых случайных величин условная вероятность совпадает с безусловной. Логарифмируя эти вероятности, имеем $H(\xi_2|\xi_1) = H(\xi_2)$ и после усреднения $H_{\xi_2|\xi_1} = H_{\xi_2}$. Следовательно, равенство $H_{\xi_1\xi_2} = H_{\xi_1} + H_{\xi_2|\xi_1}$ обращается в $H_{\xi_1\xi_2} = H_{\xi_1} + H_{\xi_2}$.

Частное проявление свойства (1.3.4) для двухэтапного разбиения было принято Шенноном [1], а также Файнстейном [1] в качестве одной из аксиом для вывода формулы (1.2.3), т. е. в сущности для специализации логарифмической меры информации. И при других аксиоматических способах определения количества информации свойство аддитивности в той или иной (может быть слабой и частной) форме приходится постулировать, чтобы выделить специальную логарифмическую меру.

В заключение этого параграфа докажем одну теорему, касающуюся условной энтропии. Сформулируем сначала следующее важное вспомогательное предложение.

Теорема 1.5. *Каковы бы ни были распределения вероятностей $P(\xi)$ и $Q(\xi)$, выполняется неравенство*

$$\sum_{\xi} P(\xi) \ln \frac{P(\xi)}{Q(\xi)} \geqslant 0. \tag{1.3.7}$$

Доказательство близко к доказательству теоремы 1.2. Оно основано на использовании неравенства (1.2.4) для функции $f(x) = \ln x$. Положим теперь $\zeta = Q(\xi)/P(\xi)$ и будем проводить усреднение с весом $P(\xi)$.

Тогда

$$\mathbf{M}\zeta = \sum_{\xi} P(\xi) \frac{Q(\xi)}{P(\xi)} = \sum_{\xi} Q(\xi) = 1$$

и

$$\mathbf{M}f(\zeta) = \sum_{\xi} P(\xi) \ln \frac{Q(\xi)}{P(\xi)}.$$

Подстановка этих значений в (1.2.4) дает

$$\sum_{\xi} P(\xi) \ln \frac{Q(\xi)}{P(\xi)} \leqslant \ln 1 = 0,$$

что завершает доказательство.

Теорема 1.6. *Условная энтропия не может превосходить безусловную*:

$$H_{\xi|\eta} \leqslant H_{\xi}. \tag{1.3.8}$$

Доказательство. Пользуясь теоремой 1.5, заменим в ней $P(\xi)$ на $P(\xi|\eta)$, $Q(\xi)$ на $P(\xi)$; тогда будем иметь

$$-\sum_{\xi} P(\xi|\eta) \ln P(\xi|\eta) \geqslant \sum_{\xi} P(\xi|\eta) \ln P(\xi).$$

Усреднение этого неравенства по η с весом $P(\eta)$ приводит к неравенству

$$-\sum_{\xi} P(\xi, \eta) \ln P(\xi|\eta) \leqslant -\sum_{\xi} P(\xi) \ln P(\xi),$$

что совпадает с (1.3.8). Доказательство закончено. Аналогично доказывается следующая теорема.

Теорема 1.6а. *При добавлении условий условная энтропия не увеличивается*:

$$H_{\xi|\eta\zeta} \leqslant H_{\xi|\zeta}. \tag{1.3.9}$$

1.4. Асимптотическая эквивалентность неравновероятных возможностей равновероятным

Идея о том, что общий случай неравновероятных возможностей (состояний) асимптотически сводится к случаю равновероятных, лежит в основе теории информации в отсутствие помех. Эта идея принадлежит Л. Больцману, который и вывел формулу (1.2.3) для энтропии. К. Шеннон возродил эту идею и широко использовал для получения новых результатов.

При рассмотрении данного вопроса в настоящем параграфе мы не будем стремиться к общности, поскольку эти результаты заведомо будут перекрыты более общими результатами в § 1.5. Возьмем набор независимых реализаций $\eta = (\xi_1, \ldots, \xi_n)$ случайной величины $\xi = \xi_j$, принимающей одно из двух значений 1 или 0 с вероятностями $\mathbf{P}[\xi = 1] = p < 1/2$; $\mathbf{P}[\xi = 0] = 1 - p = q$. Очевидно, что число различных таких наборов — реализаций равно 2^n. Пусть реализация η (обозначим ее η_{n_1}) имеет в своем составе n_1 единиц и $n - n_1 = n_0$ нулей. Тогда ее вероятность равна

$$P(\eta_{n_1}) = p^{n_1} q^{n-n_1}. \tag{1.4.1}$$

Конечно, эти вероятности при различных n_1 различны. Отношение $P(\eta_0)/P(\eta_n) = (q/p)^n$ самой большой вероятности к самой малой велико и сильно растет с ростом n. О какой же равновероятности можно говорить? Дело в том, что по закону больших чисел число единиц $n_1 = \xi_1 + \ldots + \xi_n$ имеет тенденцию принимать значения, близкие к своему среднему значению

$$\mathbf{M}n_1 = \sum_{j=1}^{n} \mathbf{M}\xi_j = n\mathbf{M}\xi_j = np.$$

Найдем дисперсию $\mathbf{D}n_1 = \mathbf{D}[\xi_1 + \ldots + \xi_n]$ числа единиц. В силу независимости слагаемых имеем

$$\mathbf{D}n_1 = n\mathbf{D}\xi = n[\mathbf{M}\xi^2 - (\mathbf{M}\xi)^2],$$

причем

$$\mathbf{M}\xi^2 = \mathbf{M}\xi = p; \qquad \mathbf{M}\xi^2 - (\mathbf{M}\xi)^2 = p - p^2 = pq.$$

Следовательно,

$$\mathbf{D}n_1 = npq; \qquad \mathbf{D}(n_1/n) = pq/n. \tag{1.4.2}$$

Мы получили, таким образом, что среднее отклонение

$$\Delta n_1 = n_1 - np \sim \sqrt{pqn}$$

растет с ростом n, но медленнее, чем растет среднее значение np и длина всего возможного диапазона $0 \leqslant n_1 \leqslant n$. Типичное относительное отклонение $\Delta n_1/n_1$ убывает по закону $\Delta n_1/n_1 \sim \sqrt{q/np}$. В пределах диапазона $|n_1 - pn| \sim \sqrt{pqn}$ неодинаковость вероятностей $P(\eta_{n_1})$ все-таки довольно велика:

$$\frac{P(\eta_{n_1})}{P(\eta_{n_1-\Delta n_1})} = \left(\frac{q}{p}\right)^{\Delta n_1} \approx \left(\frac{q}{p}\right)^{\sqrt{pqn}}$$

(как следует из (1.4.1)) и увеличивается с ростом n. Но это увеличение происходит относительно медленно, т. е. гораздо медленнее, чем увеличивается обратная величина самой вероятности. Соответствующее неравенство

$$\ln \frac{P(\eta_{n_1})}{P(\eta_{n_1-\Delta n_1})} \leqslant \ln \frac{1}{P(\eta_{n_1})}$$

при $\Delta n_1 \approx \sqrt{pqn}$ становится все более и более сильным с ростом n. Сформулируем сказанное точнее.

Теорема 1.7. *Все 2^n реализаций η можно разделить на два множества A_n и B_n так, что*

1) *суммарная вероятность реализаций одного множества (A_n) исчезает*:

$$\mathbf{P}(A_n) \to 0 \text{ при } n \to \infty; \tag{1.4.3}$$

2) *а реализации второго множества B_n становятся относительно равновероятными в следующем смысле*:

$$\left| \frac{\ln P(\eta) - \ln P(\eta')}{\ln P(\eta)} \right| \to 0, \quad \eta \in B_n;\ \eta' \in B_n. \tag{1.4.4}$$

Доказательство. Используя неравенство Чебышева (см., например, Гнеденко [1]), имеем

$$\mathbf{P}[|n_1 - pn| \geqslant \varepsilon] \leqslant \frac{\mathbf{D} n_1}{\varepsilon^2}.$$

Учитывая (1.4.2) и полагая $\varepsilon = n^{3/4}$, находим отсюда

$$\mathbf{P}[|n_1 - pn| \geqslant n^{3/4}] \leqslant pqn^{-1/2}. \tag{1.4.5}$$

Реализации η_{n_1}, для которых $|n_1 - pn| \geqslant n^{3/4}$, отнесем к множеству A_n, а остальные — к множеству B_n. Тогда первая часть (1.4.5) будет $P(A_n)$ и предельный переход $n \to 0$ в (1.4.5) докажет (1.4.3).

Для реализаций второго множества справедливо неравенство $pn - n^{3/4} < n_1 < pn + n^{3/4}$, из которого, учитывая (1.4.1), получаем

$$-n(p\ln p + q\ln q) - n^{3/4}\ln\frac{q}{p} < -\ln P(\eta_{n_1}) <$$
$$< -n(p\ln p + q\ln q) + n^{3/4}\ln\frac{q}{p}. \tag{1.4.6}$$

Отсюда имеем

$$|\ln P(\eta)-\ln P(\eta')|<2n^{3/4}\ln\frac{q}{p} \text{ при } \eta\in B_n,\ \eta'\in B_n$$

и

$$|\ln P(\eta)|>-n(p\ln p+q\ln q)-n^{3/4}\ln\frac{q}{p}.$$

Следовательно,

$$\left|\frac{\ln P(\eta)-\ln P(\eta')}{\ln P(\eta)}\right|<\frac{2n^{-1/4}\ln\frac{q}{p}}{-p\ln p-q\ln q-n^{-1/4}\ln\frac{q}{p}}.$$

Чтобы получить (1.4.4), здесь остается совершить предельный переход $n\to\infty$. Доказательство закончено.

Неравенство (1.4.6) позволяет, кроме того, оценить число элементов множества B_n.

Теорема 1.8. *Пусть B_n — множество, описанное в теореме 1.7. Число M его элементов таково, что*

$$\frac{\ln M}{n}\to -p\ln p-q\ln q\equiv H_{\xi_1} \text{ при } n\to\infty. \quad (1.4.7)$$

Эту теорему можно сформулировать также в следующей форме.

Теорема 1.8а. *Если считать реализации множества A_n имеющими нулевую вероятность, а реализации множества B_n равновероятными и вычислять энтропию H_η по простой формуле $H_\eta=\ln M$* (см. (1.1.5)), *то удельная энтропия H_η/n в пределе будет совпадать с энтропией, вычисленной по формуле* (1.2.3), *т. е.*

$$\lim_{n\to\infty}\frac{\widetilde{H}_\eta}{n}=-p\ln p-q\ln q. \quad (1.4.8)$$

В описанном смысле формула (1.2.3) получается как следствие более простого соотношения (1.1.5).

Доказательство. Согласно (1.4.5) сумма

$$\sum_{\eta\in B_n}P(\eta)=P(B_n)$$

оценивается неравенством

$$P(B_n)=1-P(A_n)\geqslant 1-\frac{pq}{\sqrt{n}}.$$

Указанная сумма распространяется на элементы множества B_n и имеет M членов. Вследствие (1.4.6) каждый член можно оценить сверху:

$$P(\eta)<\exp\left[n(p\ln p+q\ln q)+n^{3/4}\ln\frac{q}{p}\right].$$

Следовательно, число членов не может быть меньше определенного выражения

$$M \geqslant \frac{P(B_n)}{\max P(\eta)} > \left(1 - \frac{pq}{\sqrt{n}}\right) \exp\left[-n(p\ln p + q\ln q) - \right.$$

$$\left. - n^{3/4} \ln \frac{q}{p}\right]. \qquad (1.4.9)$$

С другой стороны, $\sum_{B_n} P(\eta) \leqslant 1$ и в силу (1.4.6)

$$P(\eta) > \exp\left[n(p\ln p + q\ln q) - n^{3/4} \ln \frac{q}{p}\right],$$

так что

$$M \leqslant \frac{P(B_n)}{\min P(\eta)} \leqslant \frac{1}{\min P(\eta)} < \exp\left[-n(p\ln p + q\ln q) + \right.$$

$$\left. + n^{3/4} \ln \frac{q}{p}\right]. \qquad (1.4.10)$$

Логарифмируя полученные неравенства (1.4.9), (1.4.10), имеем

$$H_{\xi_1} - n^{-1/4} \ln \frac{q}{p} + \ln\left(1 - \frac{pq}{\sqrt{n}}\right) < \frac{\ln M}{n} < H_{\xi_1} +$$

$$+ n^{-1/4} \ln \frac{q}{p}.$$

Предельным переходом $n \to \infty$ доказываем требуемые соотношения (1.4.7), (1.4.8).

В случае, когда ξ_1 принимает одно из m значений, имеется m^n различных реализаций процесса $\eta = (\xi_1, \ldots, \xi_n)$ с независимыми значениями. Согласно вышеизложенному из них заслуживают внимания лишь $M = e^{nH_{\xi_1}}$ реализаций, которые можно считать равновероятными. Когда $P(\xi_1) = 1/m$, $H_{\xi_1} = \ln m$, указанные два числа равны один другому, в противном случае, когда $H_{\xi_1} < \ln m$, доля $e^{nH_{\xi_1}}/m^n$ заслуживающих внимания реализаций неограниченно уменьшается с ростом n. Следовательно, подавляющее большинство реализаций при этом является несущественным и его можно отбросить. Этот факт лежит в основе теории кодирования (см. гл. 2).

Рассмотренный выше факт асимптотической равновероятности имеет место и при более общих предположениях в случае эргодических стационарных процессов. Указанные равновероятные реализации Больцман называл «микросостояниями», противопоставляя их «макросостояниям», образованным ансамблем «микросостояний».

1.5. Асимптотическая равновероятность и энтропийная устойчивость

1. Изложенные в предыдущем параграфе идеи, касающиеся асимптотической эквивалентности неравновероятных возможностей (сообщений) равновероятным, могут быть распространены на значительно более общий класс случайных последовательностей и процессов. Необязательно, чтобы случайные величины ξ_j, образующие отрезок последовательности $\eta^n = (\xi_1, \ldots, \xi_n)$, принимали лишь одно из двух значений. Необязательно, чтобы они имели одинаковый закон распределения $P(\xi_j)$. Необязательно, наконец, чтобы они были статистически независимыми и не обязательно даже, чтобы η^n была последовательностью $(\xi_1, \ldots, \xi_n)$. Что же является обязательным для описанной асимптотической эквивалентности?

Дать общую формулировку свойства асимптотической эквивалентности неравновероятных возможностей равновероятным помогает понятие энтропийной устойчивости семейства случайных величин.

Семейство случайных величин $\{\eta^n\}$ является *энтропийно устойчивым*, если отношение $H(\eta^n)/H_{\eta^n}$ при $n \to \infty$ сходится по вероятности к единице. Это значит, что каковы бы ни были $\varepsilon > 0$, $\eta > 0$, найдется такое $N(\varepsilon, \eta)$, что будет выполнено неравенство

$$\mathbf{P}\{|H(\eta^n)/H_{\eta^n} - 1| \geqslant \varepsilon\} < \eta \tag{1.5.1}$$

при любом $n \geqslant N(\varepsilon, \eta)$.

В приведенном выше определении подразумевается, что все $0 < H_{\eta^n} < \infty$ и H_{η^n} не убывают с ростом n. Обычно $H_{\eta^n} \to \infty$.

Факт асимптотической равновероятности можно сформулировать при помощи понятия энтропийной устойчивости в виде следующей общей теоремы.

Теорема 1.9. *Если семейство случайных величин $\{\eta^n\}$ является энтропийно устойчивым, то множество реализаций каждой случайной величины можно разбить на два подмножества A_n и B_n таким образом, что*

1) *суммарная вероятность реализаций подмножества A_n исчезает:*

$$P(A_n) \to 0 \text{ при } n \to \infty; \tag{1.5.2}$$

2) *реализации второго подмножества B_n становятся относительно равновероятными в смысле соотношения*

$$\left| \frac{\ln P(\eta) - \ln P(\eta')}{\ln P(\eta)} \right| \to 0 \text{ при } n \to \infty, \quad \eta \in B_n, \eta' \in B_n; \tag{1.5.3}$$

3) *число M_n реализаций множества B_n связано с энтропией H_{η^n} соотношением*

$$\ln M_n / H_{\eta^n} \to 1 \text{ при } n \to \infty. \tag{1.5.4}$$

Доказательство. Полагая, например, $\varepsilon = 1/m$, $\eta = 1/m$, из (1.5.1) имеем

$$\mathbf{P}\{|H(\eta^n)/H_{\eta^n}-1| \geqslant 1/m\} < 1/m \tag{1.5.5}$$

при $n \geqslant N(1/m,\ 1/m)$. Пусть m увеличивается, пробегая последовательные целые значения. Множества A_n определяем как множества реализаций, удовлетворяющих неравенству

$$|H(\eta^n)/H_{\eta^n}-1| \geqslant 1/m \tag{1.5.6}$$

при $N(1/m,\ 1/m) \leqslant n < N(1/(m+1),\ 1/(m+1))$. При таком определении, очевидно, свойство (1.5.2) выполняется в силу (1.5.5). Для дополнительного множества B_n имеем

$$(1-1/m)H_{\eta^n} < -\ln P(\eta^n) < (1+1/m)H_{\eta^n} \quad (\eta^n \in B_n), \tag{1.5.7}$$

откуда получаем

$$\left|\frac{\ln P(\eta^n)-\ln P(\eta'^n)}{-\ln P(\eta^n)}\right| \leqslant \frac{2H_{\eta^n}/m}{-\ln P(\eta^n)} \leqslant \frac{2/m}{1-1/m}$$

при $\eta^n \in B_n$, $\eta'^n \in B_n$, что доказывает свойство (1.5.3) (сходимость $n = N(1/m,\ 1/m)$ вызывает сходимость $m \to \infty$).

Вероятности $P(\eta^n)$ всех реализаций множества B_n согласно (1.5.7) лежат в диапазоне

$$e^{-(1+1/m)H_{\eta^n}} < P(\eta^n) < e^{-(1-1/m)H_{\eta^n}}.$$

В то же время суммарная вероятность $P(B_n) = \sum\limits_{B_n} P(\eta^n)$ заключена между $1-1/m$ и 1. Отсюда получаем следующий диапазон для числа членов:

$$(1-1/m)\,e^{(1-1/m)H_{\eta^n}} < M_n < e^{(1+1/m)H_{\eta^n}}$$

(слева наименьшее число делится на наибольшее, а справа — наибольшее на наименьшее). Поэтому

$$1-\frac{1}{m}+\frac{\ln(1-1/m)}{H_{\eta^n}} < \frac{\ln M_n}{H_{\eta^n}} < 1+\frac{1}{m},$$

откуда вытекает (1.5.4). (Член $\ln(1-1/m)/H_{\eta^n}$ стремится к нулю вследствие неубывания энтропии H_{η^n}.) Доказательство закончено.

Свойство энтропийной устойчивости, играющее согласно теореме 1.9 большую роль, удобно проверять для различных примеров, вычисляя дисперсию

$$\mathbf{D}H(\eta^n) = \mathbf{M}\,H^2(\eta^n) - H^2_{\eta^n}$$

случайной энтропии. Если эта дисперсия не слишком быстро растет с ростом n, то применением неравенства Чебышева можно доказать

(1.5.1), т. е. энтропийную устойчивость. Сформулируем три относящиеся к этому вопросу теоремы.

Теорема 1.10. *Если существует равный нулю предел*

$$\lim_{n\to\infty} \frac{\mathbf{D}H(\eta^n)}{H^2_{\eta^n}} = 0, \tag{1.5.8}$$

то семейство случайных величин $\{\eta^n\}$ *является энтропийно устойчивым.*

Доказательство. Согласно неравенству Чебышева для любой случайной величины с конечной дисперсией при любом $\varepsilon > 0$ имеем

$$\mathbf{P}\{|\xi - \mathbf{M}\xi| \geqslant \varepsilon\} \leqslant \mathbf{D}\xi/\varepsilon^2.$$

Полагая здесь $\xi = H(\eta_{\eta^n})/H_{\eta^n}$ и учитывая (1.5.8), получаем

$$\mathbf{P}\left\{\left|\frac{H(\eta^n)}{H_{\eta^n}} - 1\right| \geqslant \varepsilon\right\} \to 0 \quad \text{при} \quad n \to 0$$

при любом ε. Отсюда вытекает (1.5.1), т. е. энтропийная устойчивость.

Теорема 1.11. *Если энтропия* H_{η^n} *неограниченно возрастает и существует ограниченный верхний предел*

$$\lim_{n\to\infty} \sup \frac{\mathbf{D}H(\eta^n)}{H_{\eta^n}} < C, \tag{1.5.9}$$

то семейство случайных величин энтропийно устойчиво.

Для доказательства нужно лишь учесть, что величина $\mathbf{D}H(\eta^n)/H^2_{\eta^n}$, которую в силу (1.5.9) можно оценить как

$$\frac{\mathbf{D}H(\eta^n)}{H^2_{\eta^n}} \leqslant \frac{C}{H_{\eta^n}} + \varepsilon \quad (\text{где} \quad n > N(\varepsilon)),$$

стремится к нулю вследствие возрастания H_{η^n} (и произвольности ε) так что к данному случаю можно применить теорему 1.10.

Во многих важных для практики случаях существуют конечные пределы

$$H_1 = \lim_{n\to\infty} \frac{1}{n} H_{\eta^n}; \qquad D_1 = \lim_{n\to\infty} \frac{1}{n} \mathbf{D}H(\eta^n), \tag{1.5.10}$$

которые можно назвать удельной энтропией и удельной дисперсией. Для их вычисления разработан ряд более или менее общих методов. Согласно приводимой ниже теореме конечность этих пределов гарантирует энтропийную устойчивость.

Теорема 1.12. *Если пределы* (1.5.10) *существуют и конечны и первый из них отличен от нуля, то семейство случайных величин энтропийно устойчиво.*

Для доказательства учтем, что из (1.5.10) следуют соотношения

$$H_{\eta^n} = H_1 n + o(n); \tag{1.5.11}$$

$$\mathbf{D}H(\eta^n) = D_1 n + o(n). \tag{1.5.12}$$

Здесь, как обычно, $o(n)$ обозначает, что $o(n)/n \to 0$. Поскольку $H_1 > 0$, энтропия H_{η^n} неограниченно нарастает. Поделив выражение (1.5.11) на (1.5.12), получим конечный предел

$$\lim_{n\to\infty} \frac{\mathbf{D}H(\eta^n)}{H_{\eta^n}} = \lim_{n\to\infty} \frac{D_1 + o(1)}{H_1 + o(1)} = \frac{D_1}{H_1}.$$

Тем самым выполнены условия теоремы 1.11, что доказывает энтропийную устойчивость.

Пример. Пусть задана бесконечная последовательность $\{\xi_j\}$ дискретных случайных величин, статически независимых, но по разному распределенных. Предположим, что энтропии H_{ξ_j} всех этих величин конечны и равномерно ограничены снизу:

$$H_{\xi_j} > C_1, \tag{1.5.13}$$

а дисперсии равномерно ограничены сверху:

$$\mathbf{D}H(\xi_j) < C_2. \tag{1.5.14}$$

Случайные величины η^n определяем как набор n первых элементов последовательности $\eta^n = \{\xi_1, \ldots, \xi_n\}$. Тогда вследствие (1.15.10), (1.5.14) и статистической независимости, т. е. аддативности энтропий, будем иметь

$$H_{\eta^n} > C_1 n; \qquad \mathbf{D}H(\eta^n) < C_2 n.$$

В этом случае условия теоремы 1.11 являются выполненными и из нее вытекает энтропийная устойчивость семейства $\{\eta^n\}$.

Другие более сложные примеры энтропийно устойчивых случайных величин, не распадающихся на статистически независимые величины, будут рассмотрены в дальнейшем.

Понятие энтропийной устойчивости можно, правда менее строго, применять не к последовательности случайных величин $\{\eta^n\}$, а к одной случайной величине η. Тогда под энтропийно устойчивой случайной величиной η нужно понимать такую величину, для которой при некотором малом значении ε вероятность

$$\mathbf{P}\left\{\left|\frac{H(\eta)}{H_\eta} - 1\right| < \varepsilon\right\}$$

достаточно близка к единице, т. е. $H(\eta)/H_\eta$ достаточно близко к единице.

В дальнейшем будут введены другие понятия: понятие информационной устойчивости, канонической устойчивости и пр., которые во многих отношениях напоминают энтропийную устойчивость.

2. Для получения ряда результатов, связанных с энтропийно-устойчивыми случайными величинами, иногда удобно рассматривать характеристический потенциал $\mu_0(\alpha)$ энтропии $H(\eta)$, определяемый формулой

$$e^{\mu_0(\alpha)} = \sum_{\eta} e^{\alpha H(\eta)} P(\eta) = \sum_{\eta} P^{1-\alpha}(\eta), \tag{1.5.15}$$

подобный тем потенциалам, которые рассматриваются в дальнейшем (§ 4.1, 4.4). При помощи этого потенциала удобно исследовать быстроту сходимости (1.5.2)—(1.5.4) в теореме 1.9. Этот вопрос затрагивается в следующей теореме.

Теорема 1.13. *Пусть потенциал* (1.5.15) *определен и дифференцируем в отрезке* $s_1 < \alpha < s_2$ $(s_1 < 0;\ s_2 > 0)$ *и пусть уравнение*

$$\frac{d\mu_0}{d\alpha}(s) = (1+\varepsilon) H_\eta \qquad (\varepsilon > 0) \tag{1.5.16}$$

имеет корень $s \in [0,\ s_2]$. *Тогда подмножество* A *реализаций* η, *определенное условием*

$$\frac{H(\eta)}{H_\eta} - 1 > \varepsilon, \tag{1.5.17}$$

имеет вероятность

$$\mathbf{P}(A) \leqslant e^{-s\mu'_0(s) + \mu_0(s)}. \tag{1.5.18}$$

Остальные реализации, составляющие дополнительное подмножество B, *имеют вероятности, связанные соотношением*

$$\left| \frac{\ln P(\eta) - \ln P(\eta')}{\ln P(\eta)} \right| < \frac{2\varepsilon}{1-\varepsilon} \quad (\eta,\ \eta' \in B), \tag{1.5.19}$$

причем число M *этих реализаций удовлетворяет неравенству*

$$1 - \varepsilon + \frac{1}{H_\eta} \ln \left[1 - e^{-s\mu'_0(s) + \mu_0(s)}\right] < \frac{\ln M}{H_\eta} < 1 + \varepsilon. \tag{1.5.20}$$

Доказательство во многом аналогично доказательству теоремы 1.9. Соотношение (1.5.19)вытекает из (1.5.17). Неравенство (1.5.17) эквивалентно неравенству

$$e^{-(1+\varepsilon) H_\eta} < P(\eta) < e^{-(1-\varepsilon) H_\eta}.$$

Учитывая его и равенство

$$\sum_{B} P(\eta) = 1 - \mathbf{P}(A),$$

находим, что число членов в этой сумме, т. е. число реализаций в B, удовлетворяет неравенству

$$[1 - \mathbf{P}(A)]\, e^{(1-\varepsilon) H_\eta} < M < [1 - \mathbf{P}(A)]\, e^{(1+\varepsilon) H_\eta}.$$

Следовательно,

$$(1-\varepsilon) H_\eta + \ln[1-\mathbf{P}(A)] < \ln M < (1+\varepsilon) H_\eta + \ln[1-\mathbf{P}(A)].$$

Отсюда вытекает (1.5.20), если учесть отрицательность $\ln[1-\mathbf{P}(A)]$ и неравенство (1.5.18). Итак, для завершения доказательства теоремы остается обосновать (1.5.18). Это соотношение мы получим, если применим доказываемую в дальнейшем теорему 4.7 (§ 4.4) при $\xi = \eta$, $B(\xi) = H(\eta)$. Доказательство закончено.

Из формул, приведенных в предыдущей теореме, можно получить ряд простых приближенных соотношений, если использовать условие малости ε. При $\varepsilon = 0$ имеется нулевой корень $s = 0$ уравнения (1.5.16), так как $\frac{d\mu}{d\alpha}(0) = H_\eta$. При малых ε значение s мало, и правую часть уравнения (1.5.16) можно разложить в ряд Маклорена

$$\mu_0'(s) = \mu_0'(0) + \mu_0''(0)\, s + \ldots,$$

так что корень уравнения будет иметь вид

$$s = \frac{\varepsilon H_\eta}{\mu_0''(0)} + O(\varepsilon^2). \tag{1.5.21}$$

Разложим далее в ряд выражение, стоящее в экспоненте (1.5.18):

$$s\mu_0'(s) - \mu_0(s) = \frac{1}{2}\mu_0''(0)\, s^2 + O(s^3).$$

Подставляя сюда (1.5.21), получаем

$$\mathbf{P}(A) \leqslant \exp\left\{-\frac{\varepsilon^2 H_\eta^2}{2\mu_0''(0)}\right\}[1 + O(\varepsilon^3)]. \tag{1.5.22}$$

Здесь $\mu_0''(0) > 0$ — положительная величина, в силу общих свойств характеристического потенциала, равная дисперсии

$$\mu_0''(0) = \mathbf{D}H(\eta). \tag{1.5.23}$$

Мы видим, что характеристический потенциал энтропии помогает исследовать вопросы, связанные с энтропийной устойчивостью.

1.6. Определение энтропии непрерывной случайной величины

До сих пор предполагалось, что случайная величина ξ, энтропия H_ξ которой изучалась, может принимать значения из некоторого дискретного пространства, состоящего из конечного или счетного числа элементов, например, сообщений, символов и т. п. Между тем в технике большое распространение имеют также непрерывные величины, т. е. величины (скалярные или векторные), которые могут принимать значения из непрерывного пространства X, чаще всего пространства действительных чисел. При этом случайная величина ξ описывается плотностью распределения вероятностей $p(\xi)$, задающей вероятность

$$\Delta P = \int_{\xi \in \Delta X} p(\xi)\, d\xi \approx p(A)\, \Delta V \qquad (A \in \Delta X)$$

попадания ξ в область ΔX указанного пространства X, имеющую объем ΔV ($d\xi = dV$ — дифференциал объема).

Как определить энтропию H_ξ такой случайной величины? Один из возможных формальных путей таков. В формуле

$$H_\xi = -\sum_\xi P(\xi) \ln P(\xi) = -\mathbf{M} \ln P(\xi), \tag{1.6.1}$$

пригодной для дискретной величины, вероятности $P(\xi)$ под знаком логарифма формально заменяются на плотность вероятности и берется, следовательно, выражение

$$H_\xi = \mathbf{M} \ln p(\xi) = -\int_X p(\xi) \ln p(\xi)\, d\xi. \tag{1.6.2}$$

Такой способ определения энтропии не очень обоснован. Остается неясным, как определять энтропию в комбинированном случае, когда в непрерывном пространстве, кроме непрерывного распределения, имеется еще концентрация вероятности в отдельных точках, т. е. плотность вероятности содержит еще дельта-образные особенности. Энтропия (1.6.2) обладает также тем недостатком, что является неинвариантной, т. е. меняется при невырожденном преобразовании переменных $\eta = f(\xi)$ в отличие от энтропии (1.6.1), которая при этом остается инвариантной.

Поэтому целесообразно дать несколько другое определение энтропии непрерывной случайной величины. Подойдем к этому определению исходя из формулы (1.6.1). Будем предполагать, что ξ — дискретная случайная величина, вероятности которой сосредоточены в точках A_i непрерывного пространства X:

$$P(A_i) = w_i.$$

Это означает, что плотность распределения имеет вид

$$p(\xi) = \sum_i w_i \delta(\xi - A_i), \qquad \xi \in X. \tag{1.6.3}$$

Используя формулу (1.6.1), в этом случае имеем

$$H_\xi = -\sum_i P(A_i) \ln P(A_i). \tag{1.6.4}$$

Пусть далее точки A_i расположены в пространстве X довольно плотно, так что даже в сравнительно малой области ΔX, имеющей объем ΔV, располагается довольно большое число ΔN их. Область

ΔX предполагается малой в том смысле, что внутри ее вероятности $P(A_i)$ приблизительно равны

$$P(A_i)\approx\frac{\Delta P}{\Delta N}\quad\text{при}\quad A_i\in\Delta X. \tag{1.6.4a}$$

Тогда для суммы по точкам, лежащим внутри ΔX, будем иметь

$$-\sum_{A_i\in\Delta X}P(A_i)\ln P(A_i)\approx-\Delta P\ln\frac{\Delta P}{\Delta N}.$$

Суммируя по всем таким областям ΔX, увидим, что энтропия (1.6.4) запишется в виде

$$H_\xi\approx-\sum\Delta P\ln\frac{\Delta P}{\Delta N}. \tag{1.6.5}$$

Если ввести меру $\nu_0(\xi)$, указывающую плотность точек A_i, интегрированием которой вычисляется число точек

$$\Delta N=\int\limits_{\xi\in\Delta X}\nu_0(\xi)\,d\xi$$

внутри любой области ΔX, то энтропию (1.6.4) можно записать

$$H_\xi=-\int\limits_X p(\xi)\ln\frac{p(\xi)}{\nu_0(\xi)}\,d\xi. \tag{1.6.6}$$

Плотности распределения (1.6.3), очевидно, соответствует дельта-образная плотность $\nu_0(\xi)$ вида

$$\nu_0(\xi)=\sum_i\delta(\xi-A_i).$$

От этих дельта-образных плотностей можно перейти к сглаженным плотностям

$$\begin{aligned}\tilde p(\xi)&=\int K(\xi-\xi')\,p(\xi')\,d\xi',\\ \tilde\nu_0(\xi)&=\int K(\xi-\xi')\,\nu_0(\xi')\,d\xi'.\end{aligned} \tag{1.6.7}$$

При этом, если «радиус сглаживания» r_0, соответствующий ширине функции $K(x)$, относительно невелик (нужно, чтобы на таких расстояниях вероятности $P(A_i)$ не успевали существенно измениться), то сглаживание мало повлияет на отношение плотностей:

$$\frac{\tilde p(\xi)}{\tilde\nu_0(\xi)}\approx\frac{p(\xi)}{\nu_0(\xi)}.$$

Поэтому из (1.6.6) будем иметь

$$H_\xi\approx-\int\tilde p(\xi)\ln\frac{\tilde p(\xi)}{\tilde\nu_0(\xi)}\,d\xi. \tag{1.6.8}$$

Эту формулу можно получить также из (1.6.5), так как $\Delta P \approx \widetilde{p}\Delta V$, $\Delta N \approx \widetilde{v}_0 \Delta V$, когда «радиус» r_0 много меньше размеров областей ΔX. Но если «радиус сглаживания» r_0 значительно превосходит среднее расстояние между точками A_i, то сглаженные функции (1.6.7) будут иметь простой (гладкий) вид, который предполагался, скажем, в формуле (1.6.2).

Отбросив значки $\sim$ в (1.6.8) у обеих плотностей, мы, следовательно, получим в качестве формулы, определяющей энтропию H_ξ, вместо (1.6.2) формулу

$$H_\xi = -\int_X p(\xi) \ln \frac{p(\xi)}{\nu_0(\xi)} d\xi. \tag{1.6.9}$$

Здесь $\nu_0(\xi)$ — некоторая вспомогательная плотность, которая предполагается заданной. Она необязательно нормирована на единицу. Более того, в соответствии с приведенной выше интерпретацией энтропии (1.6.9) как частного случая энтропии (1.6.1) нормировочный интеграл

$$\int_X \nu_0(\xi) d\xi = N \tag{1.6.10}$$

предполагается достаточно большим числом ($N \gg 1$), интерпретируемым как общее число дискретных точек A_i, в которых сконцентрированы вероятности $P(A_i)$. Такая интерпретация, однако, необязательна, и можно, в частности, брать $N = 1$.

Если ввести нормированную плотность

$$q(\xi) = \nu_0(\xi)/N \tag{1.6.11}$$

(что можно сделать, когда N конечно), то, очевидно, из формулы (1.6.9) будем иметь

$$H_\xi = \ln N - \int p(\xi) \ln \frac{p(\xi)}{q(\xi)} d\xi. \tag{1.6.12}$$

Определение (1.6.9) энтропии H_ξ может быть обобщено и на комбинированный случай, и на общий случай абстрактной случайной величины. Последняя считается заданной, если фиксировано некоторое пространство X точек ξ и борелевское поле F (σ-алгебра) его подмножеств. При этом говорят, что задано измеримое пространство (X, F). На этом поле далее определена вероятностная мера $P(A)$ ($A \in F$), т. е. такая мера, что $P(X) = 1$. Для определения энтропии мы требуем, чтобы на измеримом пространстве (X, F) была задана также вспомогательная мера $\nu(A)$ такая, что мере P абсолютно непрерывна относительно ν.

Мера P называется абсолютно непрерывной относительно меры ν, если для каждого множества A из F, для которого $\nu(A) = 0$, имеет место равенство $P(A) = 0$. Согласно известной теореме Радона—Никодима из условия абсолютной непрерывности меры P относительно меры ν вытекает существование F-измеримой функции $f(x)$, обозначаемой $dP/d\nu$ и называемой производной Радона—Никодима, которая обобщает понятие плотности распределения. Она определена для всех точек пространства X, за исключением, может быть, некоторого подмножества Λ, для которого $\nu(\Lambda) = 0$, а значит и $P(\Lambda) = 0$.

Итак, если условие абсолютной непрерывности выполнено, то энтропия H_ξ определяется при помощи производной Радона—Никодима по формуле

$$H_\xi = -\int\limits_{X-\Lambda-\Lambda_0} \ln \frac{dP}{d\nu}(\xi)\, P(d\xi). \tag{1.6.13}$$

Подмножество Λ, для которого функция $dP/d\nu$ не определена, не влияет на результат интегрирования, так как ему соответствуют нулевые меры $P(\Lambda) = \nu(\Lambda) = 0$. Также несущественным является то подмножество Λ_0, в котором функция $dP/d\nu$ определена, но равна нулю, так как для него

$$P(\Lambda_0) = \int\limits_{\Lambda_0} \frac{dP}{d\nu}\nu(d\xi) = \int\limits_{\Lambda_0} 0 \cdot \nu(d\xi) = 0,$$

даже если $\nu(\Lambda_0) \neq 0$. Поэтому некоторая неопределенность функции $f(\xi) = dP/d\nu$ и бесконечные значения $\ln f(\xi)$ в точках, где $f(\xi) = 0$, не мешают определению (1.6.13) энтропии H_ξ в случае абсолютной непрерывности P относительно ν. Величина

$$H(\xi) = -\ln \frac{dP}{d\nu}(\xi) \tag{1.6.14}$$

при этом играет роль случайной энтропии, аналогичной случайной энтропии (1.2.2). Она определена почти всюду в X, т. е. во всем пространстве, за исключением, может быть, множества $\Lambda + \Lambda_0$ нулевой вероятности P.

По аналогии с (1.6.11), если $N = \nu(X) < \infty$, можно ввести вместо $\nu(A)$ нормированную, т. е. вероятностную меру

$$Q(A) = \nu(A)/N, \qquad A \in F, \tag{1.6.15}$$

и преобразовать (1.6.13) к виду

$$H_\xi = \ln N - \int \ln \frac{dP}{dQ}(\xi)\, P(d\xi), \tag{1.6.16}$$

аналогичному (1.6.12). Величина

$$H_\xi^{P/Q} = \int \ln \frac{dP}{dQ}(\xi)\, P(d\xi) \tag{1.6.17}$$

является неотрицательной. Это утверждение аналогично теореме 1.5 и может быть доказано тем же способом. Вследствие указанной неотрицательности величину (1.6.17) можно называть энтропией распределения вероятности P относительно распределения вероятности Q.

Данное в настоящем параграфе определение энтропии (1.6.9), (1.6.13) позволяет рассматривать энтропию (1.6.2) как частный случай этого общего определения. Именно формула (1.6.2) есть энтропия (1.6.13) для того случая, когда мере ν соответствует равномерная единичная плотность $\nu_0(\xi) = d\nu/d\xi = 1$.

Нужно отметить, что энтропия (1.6.17) вероятностной меры P относительно вероятностной меры Q может быть использована как показатель степени различия мер P и Q (по данному поводу см. книгу Кульбака [1]). Этому благоприятствует то обстоятельство, что она обращается в нуль для совпадающих мер $P(\cdot) = Q(\cdot)$ и положительна для несовпадающих.

Другим показателем различия мер, обладающим этими же свойствами, может служить «расстояние», определяемое формулой

$$s(P, Q) = \inf \int_P^Q ds, \tag{1.6.18}$$

причем

$$ds^2 = 2H^{P/P \pm \delta P} = 2 \int \ln \frac{P(d\zeta)}{P(d\zeta) \pm \delta P(d\zeta)} P(d\zeta). \tag{1.6.19}$$

Произведя разложение функции $-\ln(1 \pm \delta P/P)$ по $\delta P/P$, нетрудно убедиться, что эта метрика может быть задана также эквивалентной формулой

$$ds^2 = \int \frac{[\delta P(d\zeta)]^2}{P(d\zeta)} = \int [\delta \ln P(d\zeta)]^2 P(d\zeta). \tag{1.6.20}$$

Соединим точки P и Q «линией» — семейством точек P_λ, зависящих от параметра λ так, что $P_0 = P$, $P_1 = Q$. Тогда (1.6.18) можно записать

$$s(P, Q) = \inf \int_0^1 \frac{ds}{d\lambda} d\lambda, \tag{1.6.21}$$

где в силу (1.6.20)

$$\left(\frac{ds}{d\lambda}\right)^2 = \int \left[\frac{\partial \ln P_\lambda(d\zeta)}{\partial\lambda}\right]^2 P_\lambda(d\zeta). \tag{1.6.22}$$

Здесь и ниже предполагаются выполненными условия дифференцируемости. Из (1.6.22) вытекает, что

$$\left(\frac{ds}{d\lambda}\right)^2 = -\int \frac{\partial^2 \ln P_\lambda(d\zeta)}{\partial^2 \lambda} P_\lambda(d\zeta). \tag{1.6.23}$$

В самом деле, разность этих выражений

$$\int\left\{\left[\frac{\partial \ln P_\lambda(d\zeta)}{\partial\lambda}\right]^2+\frac{\partial^2 \ln P_\lambda(d\zeta)}{\partial^2\lambda}\right\}P_\lambda(d\zeta)=$$

$$=\frac{\partial}{\partial\lambda}\int\frac{\partial \ln P_\lambda(d\zeta)}{\partial\lambda}P_\lambda(d\zeta)$$

равна нулю вследствие того, что тождественно исчезает выражение

$$\int\frac{\partial \ln P_\lambda(d\zeta)}{\partial\lambda}P_\lambda(d\zeta)=\frac{\partial}{\partial\lambda}\int P_\lambda(d\zeta)=\frac{\partial}{\partial\lambda}1.$$

Из указанного определения видно, что при сближении точек P и Q энтропии $H^{P/Q}$, $H^{Q/P}$ и величина $\frac{1}{2}s^2(P, Q)$ перестают различаться. Можно сформулировать, однако, теорему, связывающую указанные величины не только в случае близких точек.

Теорема 1.14. *Квадрат «расстояния»* (1.6.18) *ограничен сверху суммой энтропии*:

$$s^2(P, Q)\leqslant H^{P/Q}+H^{Q/P}. \tag{1.6.24}$$

Доказательство. Соединим точки P и Q линией

$$P_\lambda(d\zeta)=e^{-\Gamma(\lambda)}[P(d\zeta)]^{1-\lambda}Q(d\zeta)^\lambda \tag{1.6.25}$$

$$(\Gamma(\lambda)=\ln\int[P(d\zeta)]^{1-\lambda}[Q(d\zeta)]^\lambda).$$

Для нее получаем

$$-\frac{\partial^2 \ln P_\lambda(d\zeta)}{\partial^2\lambda}=\frac{d^2\Gamma(\lambda)}{d\lambda^2}.$$

Следовательно, в силу (1.6.23)

$$\left(\frac{ds}{d\lambda}\right)^2=\frac{d^2\Gamma}{d\lambda^2},$$

также

$$-\frac{\partial^2 \ln P_\lambda(d\zeta)}{\partial^2\lambda}=\left(\frac{ds}{d\lambda}\right)^2. \tag{1.6.26}$$

Далее нетрудно убедиться, что выражение (1.6.17) можно записать в такой интегральной форме

$$H^{P/Q}=-\int_0^1 d\lambda\int\frac{\partial \ln P_\lambda(d\zeta)}{\partial\lambda}P_0(d\zeta)=$$

$$=-\int_0^1 d\lambda\int_0^\lambda d\lambda'\int\frac{\partial^2 \ln P_{\lambda'}(d\zeta)}{\partial\lambda'^2}P_0(d\zeta).$$

Здесь учтено, что $\int [\partial (\ln P_\lambda (d\zeta))/\partial\lambda]\, P_0 (d\zeta) = 0$ при $\lambda = 0$. Принимая во внимание (1.6.26), отсюда находим

$$H^{P/Q} = \int_0^1 d\lambda \int_0^\lambda \left(\frac{ds}{d\lambda}\right)^2 d\lambda' = \int_0^1 (1-\lambda') \left(\frac{ds}{d\lambda'}\right)^2 d\lambda'. \tag{1.6.27}$$

Аналогичным образом, меняя местами P и Q при неизменной соединительной линии (1.6.25), получаем

$$H^{Q/P} = \int_0^1 \lambda' \left(\frac{ds}{d\lambda'}\right)^2 d\lambda'. \tag{1.6.28}$$

Сложим (1.6.28), (1.6.27) и применим неравенство Коши—Шварца:

$$\left[\int_0^1 \frac{ds}{d\lambda'} d\lambda'\right]^2 \leqslant \int_0^1 \left(\frac{ds}{d\lambda'}\right)^2 d\lambda'.$$

Это даст

$$\left[\int_0^1 \frac{ds}{d\lambda'} d\lambda'\right]^2 \leqslant H^{P/Q} + H^{Q/P},$$

а, следовательно, и (1.6.24), если учесть (1.6.21). Доказательство закончено.

1.7. Свойства энтропии в обобщенной версии. Условная энтропия

Определенная в предыдущем параграфе энтропия (1.6.13), (1.6.16) обладает рядом свойств, аналогичных свойствам энтропии дискретной случайной величины, рассмотренным ранее. Такая аналогия является вполне естественной, если принять во внимание изложенную в § 1.6 интерпретацию энтропии (1.6.13) как асимптотического случая (при больших N) энтропии (1.6.1) дискретной случайной величины.

Свойство неотрицательности энтропии, о котором шла речь в теореме 1.1, для энтропии (1.6.13), (1.6.16) выполняется не всегда, а лишь при достаточно больших N. Условие

$$H_\xi^{P/Q} \leqslant \ln N$$

приводит к неотрицательности энтропии H_ξ.

Перейдем к теореме 1.2, в которой рассматривалось максимальное значение энтропии. В случае энтропии (1.6.13) при сравнении различных распределений P меру ν нужно оставлять фиксированной. Как указывалось, величина (1.6.17) является неотрицательной, поэтому из (1.6.16) имеем неравенство

$$H_\xi \leqslant \ln N.$$

В то же время, если положить $P = Q$, то, очевидно, будем иметь

$$H_\xi = \ln N.$$

Это доказывает следующее утверждение, являющееся аналогом теоремы 1.2.

Теорема 1.15. *Энтропия* (1.6.13) *принимает максимальное значение, равное* $\ln N$, *когда мера* P *пропорциональна мере* ν.

Этот результат является вполне естественным в свете данной в § 1.6 дискретной интерпретации формулы (1.6.13). В самом деле, пропорциональность мер P и ν означает именно равномерное распределение вероятности по дискретным точкам A_i, и теорема 1.15 становится перефразировкой теоремы 1.2. Аналогом теорем 1.2 и 1.15 для энтропии $H_\xi^{P/Q}$ является следующее утверждение: энтропия $H_\xi^{P/Q}$ принимает минимальное значение, равное нулю, когда распределение P совпадает с Q.

В соответствии с теоремой 1.15 энтропию $H_\xi^{P/Q}$, определенную формулой (1.6.17), целесообразно интерпретировать как дефект энтропии, т. е. как нехватку этой величины до ее максимального значения.

До сих пор предполагалось, что мера P является абсолютно непрерывной относительно меры Q или (что при конечных N то же самое) меры ν. Возникает вопрос, как определять энтропию H_ξ или $H_\xi^{P/Q}$, когда этой абсолютной непрерывности нет. Ответ на этот вопрос можно получить, рассматривая формулу (1.6.16) как асимптотический случай (при очень большом N) дискретной версии (1.6.1). Если, уплотняя точки A_i, о которых шла речь в § 1.6, вероятности $P(A_i)$ некоторых таких точек, вопреки формуле $P(A_i) \approx \Delta P/(N\Delta Q)$ (см. (1.6.4а)), оставлять конечными: $P(A_i) > c > 0$ (c не зависит от N), то в пределе $N \to \infty$ мера Q_N не будет абсолютно непрерывна относительно меры P. Дефект $H_\xi^{P/Q}$ при $N \to \infty$ в этом случае будет неограниченно возрастать. Это дает основание считать, что

$$H_\xi^{P/Q} = \infty,$$

если мера P не является абсолютно непрерывной относительно Q (т. е. является сингулярной относительно ее). Сказанное, правда, еще не определяет энтропии H_ξ при отсутствии абсолютной непрерывности, так как для нее согласно (1.6.16) получается неопределенность типа $\infty-\infty$. Для ее определения требуется более детальный анализ предельного перехода $N \to \infty$, связанного с уплотнением точек A_i.

Прочие свойства энтропии в дискретной версии, указанные в теоремах 1.3, 1.4, 1.6, касаются энтропии многих случайных величин и условной энтропии. При соответствующем определении последних понятий данные свойства будут иметь место и для обобщенной версии, основанной на определении (1.6.13).

Пусть имеются две случайные величины ξ, η. В соответствии с (1.6.13) они имеют энтропию

$$H_{\xi,\eta} = -\int \ln \frac{P(d\xi, d\eta)}{\nu(d\xi, d\eta)} P(d\xi, d\eta). \tag{1.7.1}$$

В то же время, применяя формулу (1.6.13) к одной случайной величине ξ или η, имеем

$$H_\xi = -\int \ln \frac{P(d\xi)}{\nu_1(d\xi)} P(d\xi),$$

$$H_\eta = -\int \ln \frac{P(d\eta)}{\nu_2(d\eta)} P(d\eta).$$

Здесь ν_1, ν_2 — некоторые меры; их связь с ν рассматривается ниже. Условную энтропию $H_{\xi|\eta}$ определим как разность

$$H_{\xi|\eta} = H_{\xi\eta} - H_\eta, \tag{1.7.2}$$

т. е. так, чтобы было выполнено свойство аддитивности

$$H_{\xi\eta} = H_\eta + H_{\xi|\eta}. \tag{1.7.3}$$

Учитывая (1.6.13), (1.7.2), легко видеть, что для $H_{\xi|\eta}$ будем иметь формулу

$$H_{\xi|\eta} = -\int \ln \frac{P(d\xi \mid \eta)}{\nu(d\xi \mid \eta)} P(d\xi\, d\eta), \tag{1.7.4}$$

где $P(d\xi|\lambda\eta)$, $\nu(d\xi|\eta)$ — условные меры, определяемые как производные Радона—Никодима при помощи обычных соотношений

$$P(\xi \in A, \eta \in B) = \int\limits_{\eta \in B} P(\xi \in A \mid \eta) P(d\eta),$$

$$\nu(\xi \in A, \eta \in B) = \int\limits_{\eta \in B} \nu(\xi \in A \mid \eta)\, \nu_2(d\eta)$$

(множества A, B произвольны). Определению (1.7.4) соответствует случайная энтропия

$$H(\xi \mid \eta) = -\ln \frac{P(d\xi \mid \eta)}{\nu(d\xi \mid \eta)}.$$

Данное выше определение условной энтропии $H_{\xi|\eta}$ можно многократно использовать при поэтапном рассмотрении цепочки случайных величин $\xi_1, \xi_2, \ldots, \xi_n$. При этом соотношение (1.7.3) приведет к выполнению свойства иерархической аддитивности:

$$H(\xi_1, \ldots, \xi_n) = H(\xi_1) + H(\xi_2 \mid \xi_1) + H(\xi_3 \mid \xi_1, \xi_2) + \ldots + \\ + H(\xi_n \mid \xi_1, \ldots, \xi_{n-1}), \tag{1.7.4a}$$

$$H_{\xi_1, \ldots, \xi_n} = H_{\xi_1} + H_{\xi_2 \mid \xi_1} + H_{\xi_3 \mid \xi_1, \xi_2} + \ldots + H_{\xi_n \mid \xi_1, \ldots, \xi_{n-1}},$$

аналогичного (1.3.4).

Чтобы энтропии H_ξ, $H_{\xi|\eta}$ и другие можно было интерпретировать как меры неопределенности, необходимо, чтобы в обобщенной версии выполнялось неравенство

$$H_\xi \geqslant H_{\xi\,|\,\eta}, \tag{1.7.4б}$$

обычное для дискретной версии. Между тем это неравенство удается доказать не в самом общем случае, как это можно видеть из последующей формулы (1.7.6).

Теорема 1.16. *Если выполнено условие*

$$\nu(d\xi,\, d\eta) \leqslant \nu_1(d\xi)\,\nu_2(d\eta), \tag{1.7.5}$$

то неравенство $H_\xi \geqslant H_{\xi|\eta}$ *выполняется.*

Доказательство. Для разности $H_\xi - H_{\xi|\eta}$ будем иметь

$$H_\xi - H_{\xi\,|\,\eta} = \int P(d\eta) \int \ln \frac{P(d\xi\,|\,\eta)}{P(d\xi)}\, P(d\xi\,|\,\eta) + \\ + \mathbf{M} \ln \frac{\nu_1(d\xi)}{\nu(d\xi/\eta)}. \tag{1.7.6}$$

Но для любых вероятностных мер P_1, P_2 справедливо неравенство

$$\int \ln \frac{P_1(d\xi)}{P_2(d\xi)}\, P_1(d\xi) \geqslant 0, \tag{1.7.7}$$

которое является обобщением неравенства (1.3.7) и может быть доказано тем же самым способом. Применяя (1.7.7) к (1.7.6) при $P_1(d\xi) = P(d\xi/\eta)$, $P_2(d\xi) = P(d\xi)$, получаем, что первый член в правой части (1.7.6) неотрицателен. Второй член можно записать так:

$$\mathbf{M} \ln \frac{\nu_1(d\xi)}{\nu(d\xi\,|\,\eta)} = \mathbf{M} \ln \frac{\nu_1(d\xi)\,\nu_2(d\eta)}{\nu(d\xi,\, d\eta)}.$$

Отсюда видна его неотрицательность в силу условия (1.7.5). Следовательно, разность (1.7.6) неотрицательна. Доказательство закончено.

Неравенство (1.7.5) естественным образом выполняется, когда $\nu(A)$ интерпретируется согласно сказанному в § 1.6 как число дискретных точек в A, имеющих ненулевую вероятность. Тогда $\nu_1(\Delta\xi)$ интерпретируется как число таких точек на интервале $[\xi,\, \xi + \Delta\xi]$, а $\nu_2(\Delta\eta)$ — как число точек на интервале $[\eta,\, \eta + \Delta\eta]$. В прямоугольнике $[\xi,\, \xi + \Delta\xi] \times [\eta,\, \eta + \Delta\eta]$, естественно, остается не более $\nu_1(\Delta\xi)\nu_2(\Delta\eta)$ разрешенных точек.

Если не придерживаться указанной интерпретации, то условие (1.7.5) придется постулировать независимо. Особенно удобно принимать условие мультипликативности

$$\nu(d\xi,\, d\eta) = \nu_1(d\xi)\,\nu_2(d\eta), \tag{1.7.8}$$

которое мы будем в дальнейшем постулировать.

Итак, энтропия в обобщенной версии имеет обычные свойства, т. е. не только обладает свойством иерархической аддитивности, но и удовлетворяет обычным неравенствам, если выполняется условие мультипликативности (1.7.8). Когда имеется несколько случайных величин $\xi_1, \ldots \ \xi_n$, удобно выбирать меру ν, удовлетворяющую условию полной мультипликативности

$$\nu(d\xi_1, \ldots, d\xi_n) = \prod_{k=1}^{n} \nu_k(d\xi_k), \tag{1.7.9}$$

которое мы понимаем как условие согласования вспомогательных мер. Согласно (1.7.9) мера ν комбинированной случайной величины $\xi = (\xi_1, \ldots, \xi_n)$ распадается на произведение «элементарных» мер ν_k. Тогда в соответствии с приведенными выше формулами будем иметь

$$H(\xi_k) = -\ln \frac{P(d\xi_k)}{\nu_k(d\xi_k)}, \tag{1.7.10}$$

$$H_{\xi_k} = \mathbf{M} H(\xi_k), \tag{1.7.11}$$

$$H(\xi_k \mid \xi_1, \ldots, \xi_{k-1}) = -\ln \frac{P(d\xi_k \mid \xi_1, \ldots, \xi_{k-1})}{\nu_k(d\xi_k)}, \tag{1.7.12}$$

$$H_{\xi_k \mid \xi_1, \ldots, \xi_{k-1}} = \mathbf{M} H(\xi_k \mid \xi_1, \ldots, \xi_{k-1}) = \\ = -\ln \frac{P(d\xi_k \mid \xi_1, \ldots, \xi_{k-1})}{\nu_k(d\xi_k)}. \tag{1.7.13}$$

Если случайные величины $\xi_1, \ldots, \xi_n$ являются статистически независимыми, то $P(d\xi_k \mid \xi_1, \ldots, \xi_{k-1}) = P(d\xi_k)$ (с вероятностью единица). Следовательно,

$$H(\xi_k \mid \xi_1, \ldots, \xi_{k-1}) = H(\xi_k),$$
$$H_{\xi_k \mid \xi_1, \ldots, \xi_{k-1}} = H_{\xi_k} \quad (k = 1, \ldots, n),$$

и из (1.7.4а) получаем свойство аддитивности

$$H_{\xi_1, \ldots, \xi_n} = \sum_{k=1}^{n} H_{\xi_k},$$

совпадающее с (1.2.9).

Перейдем к рассмотрению нормированных мер

$$Q(d\xi, d\eta) = \nu(d\xi, d\eta)/N, \ Q_1(d\xi) = \nu_1(d\xi)/N_1,$$
$$Q_2(d\eta) = \nu_2(d\eta)/N_2,$$

где

$$N = \int \nu(d\xi, d\eta), \quad N_1 = \int \nu_1(d\xi), \quad N_2 = \int \nu_2(d\eta),$$

и соответствующих им энтропий (1.6.17). Для них справедливы соотношения

$$H_{\xi\eta} = \ln N - H^{P/Q}; \tag{1.7.14}$$
$$H_\xi = \ln N_1 - H_\xi^{P/Q_1}; \quad H_\eta = \ln N_2 - H_\eta^{P/Q_2}$$

типа (1.6.16). Полагая

$$Q_1(d\xi) = Q(d\xi); \qquad Q_2(d\eta) = Q(d\eta) \tag{1.7.15}$$

и принимая условие мультипликативности

$$N = N_1 N_2; \ Q(d\xi\, d\eta) = Q(d\xi)\, Q(d\eta), \tag{1.7.16}$$

определяем условную энтропию естественной формулой

$$H_{\xi|\eta}^{P/Q} = \int \ln \frac{P(d\xi \mid \eta)}{Q(d\xi)} P(d\xi\, d\eta), \tag{1.7.17}$$

так что

$$H_{\xi \mid \eta} = \ln N_1 - H_{\xi \mid \eta}^{P/Q}. \tag{1.7.17a}$$

При этом выполняется свойство аддитивности

$$H_{\xi\eta}^{P/Q} = H_\eta^{P/Q} + H_{\xi \mid \eta}^{P/Q}. \tag{1.7.18}$$

Учитывая (1.7.2), приведем неравенство (1.7.4б) к виду

$$H_\xi \geqslant H_{\xi\eta} - H_\eta.$$

Используя (1.7.14), при (1.7.15), (1.7.16), отсюда получаем

$$H_{\xi\eta}^{P/Q} - H_\eta^{P/Q} \geqslant H_\xi^{P/Q}.$$

Вследствие (1.7.18) это неравенство можно записать

$$H_\xi^{P/Q} \leqslant H_{\xi \mid \eta}^{P/Q}. \tag{1.7.19}$$

Сравнивая его с (1.7.4б), видно, что знак $\geqslant$ для энтропии $H^{P/Q}$ заменился на обратный ($\leqslant$). Это является убедительным свидетельством тому, что энтропии $H_\xi^{P/Q}$, $H_{\xi/\eta}^{P/Q}$ нельзя трактовать как меру неопределенности, как это делается с энтропиями (1.6.1) или (1.6.13).

В случае многих случайных величин целесообразно принимать условия типа (1.7.15), (1.7.16) для многих величин и пользоваться свойством иерархической аддитивности

$$H_{\xi_1, \ldots, \xi_n}^{P/Q} = H_{\xi_1}^{P/Q} + H_{\xi_2 \mid \xi_1}^{P/Q} + H_{\xi_3 \mid \xi_1, \xi_2}^{P/Q} +$$
$$+ \ldots + H_{\xi_n \mid \xi_1, \ldots, \xi_{n-1}}^{P/Q}. \tag{1.7.20}$$

Это свойство аналогично (1.3.4) и (1.7.4а).

Глава 2

КОДИРОВАНИЕ ДИСКРЕТНОЙ ИНФОРМАЦИИ ПРИ ОТСУТСТВИИ ПОМЕХ И ШТРАФОВ

Определение количества информации, данное в гл. 1, получает свое оправдание при рассмотрении преобразования информации из одного вида в другой, т. е. при рассмотрении кодирования информации. Существенно, что при таком преобразовании выполняется закон *сохранения* количества информации. Полезно провести аналогию с законом сохранения энергии. Этот закон является главным основанием для введения понятия энергии. Правда, закон сохранения информации сложнее закона сохранения энергии в двух отношениях. Закон сохранения энергии устанавливает точное *равенство* энергий при переходе энергии из одного вида в другой. При превращении же информации имеет место более сложное соотношение, а именно «не больше» ($\leqslant$), т. е. информация не может вырасти. Знак равенства соответствует *оптимальному* кодированию. Поэтому при формулировке закона сохранения информации приходится говорить, что *возможно* (существует) такое кодирование, при котором имеет место равенство количеств информации.

Второе осложняющее обстоятельство заключается в том, что равенство не является точным. Оно является приближенным, асимптотическим, справедливым для сложных (больших) сообщений, для сложных случайных величин. Чем больше система сообщений, тем точнее выполняется указанное соотношение. Точный знак равенства имеет место лишь в предельном случае. В этом отношении имеется аналогия с законами статистической термодинамики, которые справедливы для больших термодинамических систем, состоящих из большого числа (практически порядка числа Авогадро) молекул.

При выполнении кодирования предполагается заданной (большая) последовательность, сообщений ξ_1, ξ_2, ... вместе со своими вероятностями, т. е. как последовательность случайных величин. Следовательно, может быть вычислено соответствующее ей количество информации (энтропия H). Эта информация может быть записана или передана различными реализациями записи или передачи. Если M — число таких реализаций, то указанный закон сохранения количества информации заключается в равенстве $H = \ln M$, осложненном двумя отмеченными обстоятельствами (т. е. в действительности $H \leq \ln M$).

Возможен двоякий подход к решению задачи кодирования. Может осуществляться кодирование бесконечной последовательности сообщений, т. е. *текущее* (или «скользящее»). Такой же характер будет носить и обратная процедура — процедура декодирования. Для текущей обработки характерно, что количественное равенство между кодируемой и закодированной информацией поддерживается лишь в среднем. При этом возникает и нарастает с течением времени случайная временная задержка или опережение. Для фиксированной длины последовательности сообщений длина ее записи будет иметь случайный разброс, растущий с течением времени, и наоборот: при фиксированной длине записи число элементарных переданных сообщений будет характеризоваться нарастающим разбросом.

Другой подход можно назвать *«блочным»*. При нем подлежит кодировке конечная совокупность (блок) элементарных сообщений. Различные блоки, если их несколько, кодируются и декодируются независимо. При таком подходе не происходит нарастания случайных временных сдвигов, но зато происходит потеря некоторых реализаций сообщения. Некоторая небольшая часть реализаций сообщений не может быть записана и теряется, так как для нее не хватает реализаций записи. Если блок является энтропийно устойчивой величиной, то вероятность такой потери достаточно мала. Следовательно, при блочном подходе необходимо исследование вопросов, связанных с энтропийной устойчивостью.

Следуя традиции, мы рассматриваем в этой главе в основном скользящее кодирование. В последнем параграфе изучаются погрешности рассмотренного ранее вида кодирования, возникающие при конечной длине кодирующей последовательности.

2.1. Основные принципы кодирования дискретной информации

Подтвердим правильность и эффективность принятого ранее определения энтропии и количества информации рассмотрением преобразования информации из одной конкретной формы в другую. Такое преобразование носит название *кодирования* и применяется при передаче и хранении информации.

Пусть *сообщение*, которое требуется передать или записать, представляет собой последовательность ξ_1, ξ_2, ξ_3, ... элементарных сообщений или случайных величин. Каждое элементарное сообщение пусть представляет собой дискретную случайную величину ξ, принимающую одно из m значений (скажем $1, \ldots, m$) с вероятностью $P(\xi)$. Требуется преобразовать сообщение в последовательность η_1, η_2, ... букв некоторого алфавита $\mathcal{A} = (А, Б, \ldots, Я)$. Число букв алфавита обозначим через D. Интервал между словами, знаки препинания и т. п. включаем в число букв алфавита, так что сообщение записывается в этих обозначениях слитно. Требуется так записать сообщение, чтобы по записи можно было восстановить сообщение без потерь и ошибок. Теория должна указывать, при каких условиях и как это можно сделать.

Ввиду того, что число элементарных сообщений никак не связано с числом букв алфавита, тривиальное «кодирование» одного сообщения одной особой буквой является в общем случае неприемлемым. Будем сопоставлять каждому отдельному сообщению, т.е. каждой реализации $\xi = i$ случайной величины свое «слово» $V(i) = (\eta_1^i, \ldots, \eta_{l_i}^i)$ в алфавите $\mathcal{A}$ (l_i — длина этого слова). Полная совокупность таких слов (число их равно m) образует *код.* Зная код, мы по реализации сообщения $\xi_1, \xi_2, \ldots$, записываем это сообщение в алфавите $\mathcal{A}$. Оно, очевидно, примет вид

$$V(\xi_1)\, V(\xi_2) \ldots = \eta_1\eta_2\ldots,$$

где под $V(\xi_i)$ следует понимать последовательность букв.

Ввиду того, что все слова $V(\xi_i)$ пишутся слитно друг с другом, возникает непростая задача разделения последовательности букв $\eta_1, \eta_2, \ldots$ на слова $V(\xi_j)$ (раздельная запись слов означала бы некоторый частный вид — и к тому же неэкономный — вышеописанного кодирования, поскольку интервал можно считать рядовой буквой). Об этой задаче будет речь ниже (см. § 2.2).

Желательно сделать запись сообщения как можно короче, если только это возможно без какой-либо потери содержания. Теория информации указывает границы, до которых можно сжимать длину записи (или время передачи сообщения, если сообщение передается в течение какого-то времени).

Поясним выводы теории рассуждением, используя результаты § 1.4, 1.5. Пусть полное сообщение $(\xi_1, \ldots, \xi_n)$ состоит из n одинаковых независимых элементарных сообщений $\xi_j = \xi$. Тогда $(\xi_1, \ldots, \xi_n) = \zeta$ будет энтропийно устойчивой случайной величиной. Если n велико, то согласно сказанному в § 1.4, 1.5, можно рассматривать следующую упрощенную модель полного сообщения: можно считать, что осуществляется только одно из e^{nH_ζ} равновероятных сообщений, где $H_\zeta = -\sum\limits_{\zeta} P(\zeta) \ln P(\zeta)$. В то же время в буквенной записи, содержащей последовательно L букв, реализуется одна из D^L возможностей. Приравнивая эти два числа, мы получаем, что длина L записи определяется уравнением

$$nH_\xi = L \ln D. \qquad (2.1.1)$$

Если использовать более короткую запись, то в некоторых случаях сообщение неизбежно будет искажаться, а если использовать запись длины больше $nH_\xi/\ln D$, то такая запись будет неэкономной, неоптимальной.

Выводы из § 1.4, 1.5 можно применить не только к последовательности $\xi_1, \xi_2, \ldots, \xi_n$, но и к буквенной последовательности $\eta_1, \eta_2, \ldots, \eta_L$. Максимальная энтропия (равная $L \ln D$) такой последовательности достигается при распределении для каждой буквы, независимом от соседних букв и при равновероятном распределении по всему алфавиту. Но вероятности букв однозначно определяются вероятностями сообщений и выбором кода. При фиксированных вероятностях сообщений остается подбирать кодовые слова и стараться достичь указанных выше оптимальных распределений букв. Если при некотором коде достигается максимальная энтропия $L \ln D$ буквенной записи, то тогда имеется $e^{L \ln D} = D^L$ существенных реализаций записи и устанавливается взаимно-однозначное соответствие между ними и существенными реализациями сообщения, число которых e^{nH_ξ}. При этом выполняется оптимальное соотношение (2.1.1). Такие коды назовем *оптимальными*. Поскольку буквы записи для них независимы, имеем $H(\eta_{j+1} | \eta_j, \eta_{j-1}, \ldots) = H(\eta_{j+1})$, а так как осуществляется равновероятное распределение $P(\eta_{j+1}) = 1/D$ по алфавиту, то каждая буква несет одинаковую информацию

$$H(\eta_{j+1}) = \ln D. \tag{2.1.2}$$

Это соотношение говорит о том, что каждая буква используется как бы на «полную мощность»; оно является признаком оптимального кода.

Возьмем какую-либо конкретную реализацию сообщения $\xi = j$. Она содержит согласно (1.2.1) случайную информацию $H(\xi) = -\ln P(j)$. Но каждая буква алфавита в силу (2.1.2) при оптимальном коде несет информацию $\ln D$. Отсюда следует, что кодовое слово $V(\xi)$ этого сообщения, в оптимальном коде также несущее информацию $H(\xi) = -\ln P(\xi)$, должно состоять из $l(\xi) = -\ln P(\xi)/\ln D$ букв. При неоптимальном кодировании каждая буква несет меньше информации, чем $\ln D$. Поэтому длина кодового слова $l(\xi)$, несущего информацию $-\ln P(\xi) = H(\xi)$, должна быть больше. Следовательно, для любого кода

$$l(\xi) \geqslant H(\xi)/\ln D.$$

Усреднение этого неравенства дает

$$l_{\text{ср}} \geqslant H_\xi/\ln D, \tag{2.1.3}$$

что согласуется с (2.1.1), поскольку $l_{\text{ср}} = L/n$.

Итак, теория указывает, чего нельзя достичь (нельзя сделать, чтобы длительность L записи была меньше, чем $nH_\xi/\ln D$), чего можно достичь (можно сделать, чтобы длительность L была близка к $nH_\xi/\ln D$) и как этого достичь (подбирая код, дающий по возможности независимые равновероятные распределения букв по алфавиту).

Пример. Пусть имеется случайная величина ξ со следующими значениями и вероятностями:

ξ	1	2	3	4	5	6	7	8
$P(\xi)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$

Сообщение записывается в двоичном алфавите (А, Б), так что $D = 2$. Возьмем кодовые слова

$$V(1) = (\text{ААА});\ V(2) = (\text{ААБ});\ V(3) = (\text{АБА});\ V(4) = (\text{АББ});$$
$$V(5) = (\text{БАА});\ V(6) = (\text{БАБ});\ V(7) = (\text{ББА});\ V(8) = (\text{БББ}). \tag{2.1.4}$$

Чтобы сказать, хорош этот код или нет, сравним его с оптимальным кодом, для которого выполняется соотношение (2.1.1). Вычисляя по формуле (1.2.3) энтропию элементарного сообщения, имеем

$$H_\xi = \ln 2 + \frac{3}{4} \ln 2 + \ln 2 = 2{,}75 \text{ бита}.$$

В коде (2.1.4) на элементарное сообщение тратится три буквы, тогда как в оптимальном коде на это сообщение может быть согласно (2.1.1) затрачено $L/n = H_\xi/\ln 2 = 2{,}75$ буквы. Следовательно, возможно сокращение записи на 8,4%. В качестве оптимального кода можно выбрать код

$$V(1) = (\text{АА});\ V(2) = (\text{АБ});\ V(3) = (\text{БАА});\ V(4) = (\text{БАБ});$$
$$V(5) = (\text{ББАА});\ V(6) = (\text{ББАБ});\ V(7) = (\text{БББА});\ V(8) = (\text{ББББ}). \tag{2.1.5}$$

Легко проверить, что вероятность появления буквы А на первом месте равна 1/2. Поэтому энтропия первого символа равна $H_{\eta_1} = 1$ бит. Далее, при фиксированном первом символе условная вероятность появления буквы А на втором месте по-прежнему равна 1/2. Это дает условные энтропии $H_{\eta_2}(\mid \eta_1) = 1$ бит; $H_{\eta_2/\eta_1} = 1$ бит. Аналогично

$$H_{\eta_3}(\mid \eta_1, \eta_2) = 1 \text{ бит};\ H_{\eta_3 \mid \eta_1 \eta_2} = 2 \text{ бит};\ H_{\eta_4 \mid \eta_1 \eta_2 \eta_3} = 1 \text{ бит}.$$

Итак, каждая буква независимо от ее места несет одну и ту же информацию 1 бит. Случайная информация элементарного сообщения ξ при этом условии равна длине соответствующего кодового слова:

$$H(\xi) = l(\xi) \text{ бит}. \tag{2.1.6}$$

Следовательно, средняя длина слова равна информации элементарного сообщения:

$$H_\xi = l_{\text{ср}} \text{ бит} = l_{\text{ср}} \ln 2 \text{ нат}.$$

Это согласуется с (2.1.1), так что код является оптимальным. Соотношение $P(\xi) = (1/2)^{l(\xi)}$, которое эквивалентно равенству (2.1.6), является следствием независимости распределения букв в буквенной записи сообщения.

Соотношение (2.1.3) со знаком равенства справедливо в предположении, что n велико, когда можно пренебречь вероятностью $P(A_n)$ несущественных реализаций последовательности $\xi_1, \ldots, \xi_n$ (см. теоремы 1.7, 1.9). В этом смысле оно является правильным асимптотически. Для конечного n подобного соотношения установить не удается. В самом деле, если мы потребуем пунктуального воспроизведения буквально *всех* сообщений конечной длины n (а их число m^n), то длина L записи будет определяться из соотношения $D^L = m^n$. Сократить ее (например до значения $nH_\xi/\ln D$) окажется невозможным без потери некоторых сообщений, вероятность которых при конечных n конечна. С другой стороны, допуская непостоянство длины $L = L(\xi_1, \ldots, \xi_n)$ при конечных n можно так закодировать сообщения, что (2.1.3) заменится на обратное неравенство $l_{ср} < H_\xi/\ln D$.

Покажем это для случая $n = 1$. Пусть имеются три возможности $\xi = 1, 2, 3$, с вероятностями 1/2, 1/4, 1/4. Выбрав код

ξ	1	2	3
$V(\xi)$	(А)	(Б)	(АБ)

, (2.1.7)

соответствующий $D = 2$, получаем $l_{ср} = \frac{1}{2} + \frac{1}{4} + \frac{2}{4} = 1{,}25$, в то время как $H_\xi = \frac{1}{2}\ln 2 + \frac{2}{4}\ln 2 + \frac{2}{4}\ln 2 = 1{,}5$ бита. Следовательно, для данного единичного сообщения неравенство (2.1.3) нарушается.

2.2. Основные теоремы для кодирования без помех. Равнораспределенные независимые сообщения

Рассмотрим тот случай, когда имеется неограниченная (в одну сторону) последовательность $\xi_1, \xi_2, \xi_3, \ldots$ одинаковых независимых сообщений. Требуется осуществить ее кодирование. Чтобы применить рассуждения предыдущего параграфа, требуется разбить эту последовательность на отрезки, называемые блоками, содержащие по n элементарных сообщений каждый. Тогда при больших n можно использовать описанные выше общие асимптотические («термодинамические») соотношения.

Для исследования кодирования в отсутствие помех, имеется, однако, и другая возможность — избегнуть разбиения последовательности на блоки и отбрасывания несущественных реализаций. Соответствующую теорию мы изложим в этом параграфе.

Код (2.1.7), пригодный для передачи (или записи) единичного сообщения, не приспособлен для передачи последовательности таких сообщений. Например, запись АБАБ можно истолковать как запись $V(1)\,V(2)\,V(3)$ сообщения $(\xi_1, \xi_2, \xi_3) = (1\ 2\ 3)$ и как запись $V(3)\,V(1)\,V(2)$ сообщения $(3\ 1\ 2)$ (при $n = 3$), не говоря уже о сообщениях $(\xi_1, \xi_2) = (3\ 3)$; $(\xi_1, \xi_2, \xi_3, \xi_4) = (1\ 2\ 1\ 2)$, которые соответствуют другим n. Он не дает возможности однозначно расшифровать сообщение, и мы должны его забраковать, если хотим передавать последовательность сообщений.

Обязательное условие, чтобы длинную последовательность кодовых символов можно было однозначно разбить на слова, существенно ограничивает класс допустимых кодов. Назовем *дешифруемыми* кодами такие коды, в которых любая полубесконечная последовательность кодовых слов разбивается на слова однозначным образом. Как доказывается, например в книге Файнстейна [1], для этих кодов выполняется неравенство

$$\sum_{\xi=1}^{m} D^{-l(\xi)} \leqslant 1. \tag{2.2.1}$$

Удобнее рассматривать несколько более узкий класс дешифруемых кодов — назовем их *дешифруемые коды Крафта* (см. работу

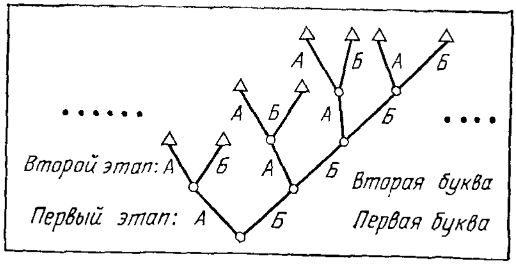


Рис. 2.1.
Пример кодового «дерева».

Крафта [1]). В этих кодах ни одно кодовое слово не является передней частью («префиксом») другого слова. В коде (2.1.7) это условие нарушено в слове (А), поскольку оно является передней частью слова (АВ). Коды можно рисовать в форме «дерева» (рис. 2.1), наподобие «деревьев», изображенных на рис. 1.1 и 1.2. Однако (если рассматривать код отдельно от сообщения ξ) в отличие от них «ветвям» «кодового дерева» не сопоставляются вероятности. Выбор ветви осуществляется по этапам записью очередной буквы η слова. Концу слова соответствует особый конечный узел, который мы на рисунке обозначаем треугольником. Каждому кодовому слову соответствует кодовая линия, идущая от начального узла до конечного. На рис. 2.1 изображено «дерево», соответствующее коду (2.1.5).

Для дешифруемых кодов Крафта ни один конечный узел не лежит на другой кодовой линии.

Теорема 2.1. *Неравенство* (2.2.1) *является необходимым и достаточным условием существования кода, дешифруемого по Крафту.*

Доказательство. Докажем сначала, что неравенство (2.2.1) выполняется для любого кода Крафта. Из каждого конечного узла кодового дерева выпустим максимальное число (равное D) вспомогательных линий, изображенных волнистыми линиями на рис. 2.2. Пусть они размножаются максимальным образом (каждая делится на D линий) на всех последующих этапах. Подсчитаем число вспомогательных линий на каком-то высоком этапе, соответствую щих номеру k. Предполагаем, что k больше длины самого длинного кодового слова. Конечный узел слова длины $l(\xi)$ даст $D^{k-l(\xi)}$ вспомогательных линий на k-м этапе. Полное число вспомогательных линий равно

$$\sum_{\xi} D^{k-l(\xi)}. \tag{2.2.2}$$

Будем теперь выпускать вспомогательные линии не только из конечных узлов, но и из каждого промежуточного узла, если из него выходит меньше, чем D кодовых линий. На других этапах пусть

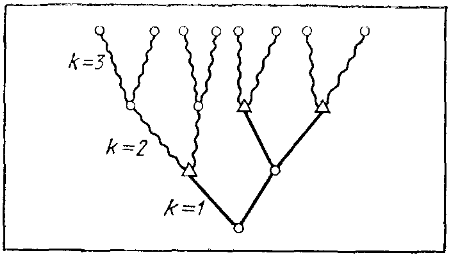


Рис. 2.2. Проведение дополнительных линий и их подсчет на уровне k (при $k=3$, $D=2$).

эти линии, в свою очередь, размножаются максимальным образом. На k-м этапе число вспомогательных линий возрастает по сравнению с (2.2.2) и станет равным

$$D^k \geqslant \sum_{\xi} D^{k-l(\xi)}.$$

Поделив обе части этого неравенства на D^k, получим (2.2.1). Очевидно, что знак равенства имеет место в том случае, если все внутренние узлы кодового дерева используются максимальным образом, т. е. из каждого узла выходит D кодовых линий.

Докажем теперь, что условие (2.2.1) достаточно для дешифруемости по Крафту, т. е. что всегда можно построить кодовое дерево с изолированными конечными узлами, если задан набор длин $l(\xi)$, удовлетворяющих условию (2.2.1). Проведение кодовой линии заданной длины на дереве означает заполнение слова заданной длины буквами. Какие препятствия при этом могут встретиться? Пусть имеется $m_1 = m\,[l(\xi) = 1)]$ однобуквенных слов. Заполнение этих слов буквами может встретить препятствие только тогда, когда не хватит алфавита, т. е. когда $m_1 > D$. Но это неравенство противоречит условию (2.2.1). В самом деле, оставив в левой части (2.2.1) только члены с $l(\xi) = 1$, получим лишь усиление неравенства:

$$\sum_{l(\xi)=1} D^{-l(\xi)} = \frac{m}{D} \leqslant 1.$$

Так что $m_1 \leqslant D$, и алфавита должно хватить для заполнения всех однобуквенных слов. Рассмотрим двухбуквенные слова. Кроме использованных букв, стоящих на первом месте, имеется еще $D - m_1$ букв, которые мы можем поставить на первое место. К первой букве можно прибавить любую из D букв. Всего имеется $(D - m_1) D$ возможностей. На «дереве» любой узел из $D - m_1$ неконечных узлов первого этапа дает D линий. Этих возможных двухбуквенных сочетаний (линий) должно хватить для заполнения всех двухбуквенных слов, число которых обозначим m_2. В самом деле, оставляя в левой части (2.2.1) лишь члены с $l(\xi) = 1$ и $l(\xi) = 2$, имеем

$$D^{-1} m_1 + D^{-2} m_2 \leqslant 1,$$

т. е. $m_2 \leqslant D^2 - Dm_1$. Это и означает, что двухбуквенных сочетаний хватит для заполнения двухбуквенных слов. После этого заполнения остаются свободными $D^2 - Dm_1 - m_2$ двухбуквенных сочетаний, которые можно использовать для прибавления новых букв. Числа трехбуквенных сочетаний, равного $(D^2 - Dm_1 - m_2) D$, достаточно для заполнения трехбуквенных слов и т. д. Каждый раз при доказательстве используется часть членов суммы $\sum\limits_{\xi} D^{-l(\xi)}$ в (2.2.1). Доказательство закончено.

Как видно из приведенного доказательства, не представляет труда фактическое построение кода (заполнение слов буквами), если задан приемлемый набор длин $l(1), \ldots, l(n)$.

Перейдем теперь к основным теоремам.

Теорема 2.2. *Средняя длина слова $l_{\text{ср}}$ меньшая, чем $H_\xi/\ln D$, является недостижимой ни при каком кодировании.*

Доказательство. Рассмотрим разность

$$l_{\text{ср}} \ln D - H_\xi = \mathbf{M}\left[l(\xi) \ln D + \ln P(\xi)\right] = -\mathbf{M} \ln \frac{D^{-l(\xi)}}{P(\xi)}.$$

Используя очевидное неравенство

$$-\ln x \geqslant 1 - x, \tag{2.2.3}$$

получаем

$$-\mathbf{M} \ln \frac{D^{-l(\xi)}}{P(\xi)} \geqslant 1 - \mathbf{M} \frac{D^{-l(\xi)}}{P(\xi)},$$

т. е.

$$-\mathbf{M} \ln \frac{D^{-l(\xi)}}{P(\xi)} \geqslant 1 - \sum_{\xi} D^{-l(\xi)},$$

так как

$$\mathbf{M} \frac{D^{-l(\xi)}}{P(\xi)} = \sum_{\xi} \left(\frac{D^{-l(\xi)}}{P(\xi)}\right) P(\xi).$$

Учитывая (2.2.1), отсюда находим

$$l_{\text{ср}} \ln D - H_\xi \geqslant 1 - \sum_{\xi} D^{-l(\xi)} \geqslant 0$$

Доказательство закончено.

Вместо неравенства (2.2.3), можно использовать неравенство (1.2.4).

Следующие теоремы утверждают, что к значению $H_\xi/\ln D$, ограничивающему среднюю длину $l_{\text{ср}}$ снизу, можно подойти довольно близко, выбирая подходящий способ кодирования.

Теорема 2.3. *Можно указать такой способ кодирования равнораспределенных независимых сообщений, что*

$$l_{\text{ср}} < \frac{H_\xi}{\ln D} + 1. \tag{2.2.4}$$

Доказательство. Выберем длины $l(\xi)$ кодовых слов так, чтобы они удовлетворяли неравенству

$$-\frac{\ln P(\xi)}{\ln D} \leqslant l(\xi) < -\frac{\ln P(\xi)}{\ln D} + 1. \tag{2.2.5}$$

Такой выбор возможен, ибо в этом диапазоне заведомо имеется одно целое число.

Из левого неравенства следует $P(\xi) \geqslant D^{-l(\xi)}$, так что

$$1 - \sum_\xi P(\xi) \geqslant \sum_\xi D^{-l(\xi)},$$

т. е. выполняется условие дешифруемости (2.2.1). Как видно из рассуждений перед теоремой 2.2, нетрудно фактически построить кодовые слова с выбранными длинами $l(\xi)$. Усредняя обе части правого неравенства (2.2.5), получаем (2.2.4). Доказательство закончено.

Приведем следующую теорему, которая носит асимптотический характер.

Теорема 2.4. *Существуют такие способы кодирования бесконечного сообщения $\xi_1, \xi_2, \ldots$, что средняя длина элементарного сообщения может быть сделана сколь угодно близкой к $H_\xi/\ln D$.*

Доказательство. Возьмем произвольное целое $n > 1$, разобьем последовательность $\xi_1, \xi_2, \ldots$ на группы по n случайных величин. Каждую такую группу будем рассматривать как одну случайную величину $\zeta = (\xi_1, \ldots, \xi_n)$ и применим к ней теорему 2.3. Неравенство (2.2.4) для нее примет вид

$$l_{\text{ср}\,\zeta} < \frac{H_\zeta}{\ln D} + 1, \tag{2.2.6}$$

где $H_\zeta = nH_\xi$, $l_{\text{ср}\zeta}$ — средняя длина слова, передающего сообщение $\zeta = (\xi_1, \ldots, \xi_n)$. Очевидно, что $\zeta_{\text{ср}\zeta} = nl_{\text{ср}}$, поэтому (2.2.6) дает

$$l_{\text{ср}} < \frac{H_\xi}{\ln D} + \frac{1}{n}.$$

Увеличивая n, величину $1/n$ можно сделать сколь угодно малой, что доказывает теорему.

Приведенные результаты получены для случая стационарной последовательности независимых случайных величин (сообщений). Эти результаты могут быть распространены на случай зависимых сообщений, на нестационарный случай и др. При этом существенным является то, чтобы последовательность удовлетворяла определенным свойствам энтропийной устойчивости (см. § 1.5).

2.3. Оптимальное кодирование по Хуфману. Примеры

Предыдущее рассмотрение не только решает принципиальные вопросы существования асимптотически оптимальных кодов, т. е. кодов, для которых средняя длина сообщения стремится к минимальному значению, но и дает практические способы их отыскания. Выбрав n и определив длины кодовых слов при помощи неравенства

$$-\frac{\ln P(\xi_1,\ldots,\xi_n)}{\ln D} \leqslant l(\xi_1,\ldots,\xi_n) < -\frac{\ln P(\xi_1,\ldots,\xi_n)}{\ln D} + 1,$$

получим код, который согласно теореме 2.4 является асимптотически оптимальным. Такой код, однако, может быть не вполне оптимальным для фиксированного n. Другими словами, может оказаться, что сообщение $\zeta = (\xi_1, \ldots, \xi_n)$ можно закодировать лучше, т. е. с меньшей длительностью $l_{\text{ср}}$.

Хуфман [1] (см. также Фано [1]) нашел несложный способ оптимального кодирования, которому соответствует минимальная средняя длительность из всех возможных для данного сообщения.

Рассмотрим сначала случай $D = 2$ и зададимся некоторым конкретным оптимальным кодом K для сообщения ζ. Исследуем, какими обязательными свойствами должен обладать оптимальный код или соответствующее ему «дерево».

На каждом этапе, кроме, может быть, последнего, из каждого неконечного узла должно выходить максимальное число D линий. В противном случае конечную линию с последнего этапа можно было бы перенести в этот «неполный» узел и этим сократить длину $l_{\text{ср}}$. На последнем этапе должно быть как минимум две линии. Иначе последний этап был бы излишним и его можно было бы исключить, укоротив соответствующую кодовую линию. Среди линий последнего этапа есть две линии, которым соответствуют две наименее вероятные реализации ζ_1, ζ_2 сообщения из всех реализаций $\zeta_1, \ldots$ $\ldots, \zeta_m$. В противном случае можно было бы поменять местами сообщения, приписав более длинному слову менее вероятную реализацию сообщения и тем самым укоротить среднюю длительность слова.

Но если два наименее вероятных сообщения имеют кодовые слова, кончающиеся на последнем этапе, то можно считать, что их линии выходят из одного узла предпоследнего этапа. Если в K это не так, то поменяв местами два сообщения и добившись этого, мы будем иметь другой конкретный код, имеющий ту же среднюю длительность и, следовательно, не хуже первоначального K.

Превратим узел предпоследнего этапа, из которого выходят две рассмотренные ранее линии в конечный узел. Соответственно этому две наименее вероятные реализации ζ_1, ζ_2 будем считать за одну. Число различных реализаций будет теперь равно $m - 1$ и им будут соответствовать вероятности $P(\zeta_1) + P(\zeta_2)$, $P(\zeta_3)$, ..., $P(\zeta_m)$. Можно считать, что имеется новая («упрощенная») случайная величина, новый код и новое («урезанное») «дерево». Средняя длительность слова для этой величины может быть выражена через прежнюю среднюю длительность по формуле

$$l_{\text{ср}} = l'_{\text{ср}} + P(\zeta_1) + P(\zeta_2).$$

Очевидно, если первоначальное кодовое «дерево» минимизирует $l_{\text{ср}}$, то «урезанное» кодовое «дерево» минимизирует $l'_{\text{ср}}$ и наоборот. Следовательно, задача отыскания оптимального кодового «дерева» свелась к задаче отыскания оптимального «урезанного» кодового «дерева» для сообщения с $m - 1$ возможностями. К этому «упрощенному» сообщению можно применить все те рассуждения, которые относились выше к первоначальному сообщению, и получить «дважды упрощенное» сообщение с $m - 2$ возможностями. К последнему также применяется вышеприведенное рассуждение и т. д., пока не получится тривиальное сообщение с двумя возможностями. В процессе описанного упрощения кодового «дерева» и сообщения две ветви кодового «дерева» на каждом шаге сливаются в одну и в конце полностью выясняется его структура.

Аналогичное рассмотрение можно провести и при $D > 2$. Предположим, что все узлы кодового «дерева» полностью заполнены. Это имеет место, когда $m - 1$ без остатка делится на $D - 1$. В самом деле, каждый узел (кроме конечных) добавляет $m - 1$ новых линий (если считать, что к первому узлу подходит снизу одна линия) и частное $(m - 1)/(D - 1)$ равно числу узлов. Каждое «упрощение» в этом случае уменьшает число реализаций случайной величины на $D - 1$. При этом сливаются воедино D наименее вероятных возможностей; D самых длинных ветвей кодового «дерева» из оставшихся укорачиваются на единицу и заменяются на один конечный узел.

Некоторое осложнение вносит тот случай, когда число $(m - 1)/(D - 1)$ не целое. Это значит, что из внутренних узлов кодового «дерева» хотя бы один должен быть неполным. Но в оптимальном дереве, как отмечалось, неполные узлы могут принадлежать лишь последнему этапу, относиться к последнему выбору. Без ухудшения «дерева» можно так переставить сообщения, относящиеся к словам максимальной длины, что: 1) неполным будет лишь один единственный узел и 2) сообщения, относящиеся к этому узлу, будут иметь самую меньшую вероятность. Обозначим через r $(0 < r < D - 2)$ остаток, получающийся от деления $m - 1$ на $D - 1$. Тогда единственный неполный узел будет иметь $r + 1$ ветвей. В соответствии с вышесказанным первое «обрезание» дерева и упрощение случайной величины будет заключаться в следующем: выбираются $r+1$ наименее вероятных реализаций из $\zeta_1, \ldots, \zeta_m$ и заменяются на одну реализацию с суммарной вероятностью. «Урезанное» дерево будет иметь уже заполненные внутренние узлы. При дальнейшем упрощении берутся D наименее вероятных реализаций из получившихся при предыдущем упрощении и заменяются на одну с суммарной вероятностью. То же самое производится и в дальнейшем. В этом заключается рецепт построения оптимального кода по Хуфману. Приведенные выше рассуждения говорят о том, что полученное в результате кодовое «дерево», возможно, не совпадает с оптимальным «деревом» K, но не хуже его, т. е. имеет такую же среднюю длину $l_{\text{ср}}$.

Пример. Рассмотрим кодирование в двоичном ($D=2$) алфавите (А, Б) бесконечной последовательности $\xi_1, \xi_2, \ldots$ одинаковых элементарных сообщений. Каждое элементарное сообщение ξ пусть будет такого же типа, как и в примере из § 1.2, т. е. $m=2$; $P(0)=7/8$, $P(1)=1/8$.

1) Если взять $n=1$, т. е. кодировать каждое элементарное сообщение в отдельности, то в качестве кода остается лишь выбрать тривиальный код:

$$V(0)=\text{А}; \qquad V(1)=\text{Б}.$$

Для него $l_{\text{ср}}=1$, в то время как

$$H_\xi/\ln D=0{,}544 \tag{2.3.1}$$

и, следовательно, возможно $l_{\text{ср}}=0{,}544$. Сравнение последней цифры с 1, показывает, что можно добиться значительного выигрыша в краткости, если перейти к усложненным вариантам кодирования — к $n>1$.

2) Положим $n=2$. Тогда будут такие возможности и вероятности

ζ	1	2	3	4
$\xi_1\xi_2$	11	10	01	00
$P(\zeta)$	0,015	0,110	0,110	0,765

Здесь ζ означает номер пары (ξ_1, ξ_2).

При первом «упрощении» можно объединить реализации $\zeta=1$ и $\zeta=2$ в одну реализацию, имеющую вероятность $0{,}015+0{,}110=0{,}125$. Среди оставшихся после «упрощения» реализаций, имеющих вероятность 0,110; 0,125; 0,765, снова объединяем две наименее вероятные. Схема этих «упрощений» и соответствующее им кодовое «дерево» показаны на рис. 2.3, 2.4.

Остается лишь расставить буквы А, Б вдоль ветвей, чтобы получить следующий оптимальный код:

ζ	1	2	3	4
$\xi_1\xi_2$	11	10	01	00
$V(\xi_1\xi_2)$	ААА	ААБ	АБ	Б

Для него средняя длительность слова равна

$$l_{\text{ср}\,\zeta} = 0{,}015 \cdot 3 + 0{,}110 \cdot 3 + 0{,}110 \cdot 2 + 0{,}765 = 1{,}36,$$
$$l_{\text{ср}} = \frac{1}{2}\, l_{\text{ср}\,\zeta} = 0{,}68, \tag{2.3.2}$$

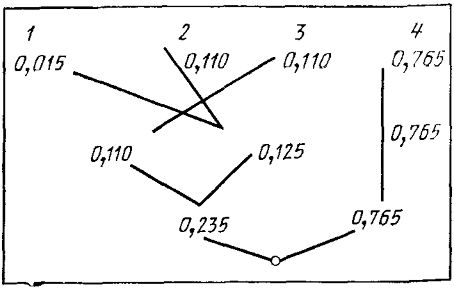


Рис. 2.3.
Схема упрощений для $n=2$.

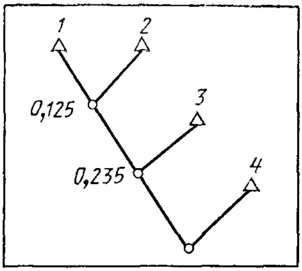


Рис. 2.4.
Кодовое «дерево» в рассмотренном примере при $n=2$.

3) Возьмем $n = 3$. Теперь имеем следующие возможности

ζ	1	2	3	4	5	6	7	8
$\xi_1\xi_2\xi_3$	111	110	101	011	100	010	001	000
$P(\zeta)$	0,002	0,014	0,014	0,014	0,096	0,096	0,096	0,668

На рис. 2.5 показано соответствующее кодовое «дерево». Расставляя буквы вдоль ветвей, находим оптимальный код

ζ	$V(\zeta)$	ζ	$V(\zeta)$
1	ААААА	5	ААБ
2	ААААБ	6	АБА
3	АААБА	7	АББ
4	АААББ	8	Б

Подсчитываем для него среднюю длительность кодовых слов:

$$l_{\text{ср}\,\zeta} = 2{,}308, \quad l_{\text{ср}} = \frac{1}{4}\, l_{\text{ср}\,\zeta} = 0{,}577.$$

Эта величина уже значительно ближе к предельному значению $l_{\text{ср}} = 0{,}544$ (2.3.1), чем $l_{\text{ср}} = 0{,}68$ (2.3.2). Увеличивая n далее

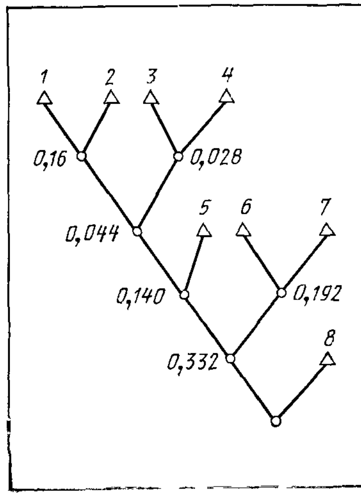


Рис. 2.5.
Кодовое «дерево» при $n=3$

и конструируя оптимальные коды описанным способом, можно сколь угодно близко приблизиться к значению 0,544. Это достигается, конечно, усложнением кодирующей системы.

2.4. Погрешности кодирования без помех при конечной длине записи

Методы кодирования, описанные в § 2.2, таковы, что длина записи фиксированного числа сообщений является случайной. Приведенная теория дает оценки для средней длины, но ничего не говорит о ее разбросе. Между тем практически длина записи или время передачи сообщения бывает ограниченно техническими условиями. Может оказаться, что запись сообщения не сможет уложиться в допустимые пределы и данная рализация сообщения поэтому не сможет быть записана (или передана). Это вносит определенные потери информации или искажения (зависящие от разброса длины записи), исследование которых является особой проблемой. Как видно из дальнейшего, определяющим фактором в этом отношении является энтропийная устойчивость случайных величин

$$\eta^n = (\xi_1, \ldots, \xi_n).$$

Пусть длина записи $l(\eta^n)$ сообщения η^n не может превосходить фиксированной величины L. Выберем код, определяемый согласно

(2.2.5) неравенствами

$$-\ln P(\eta^n)/\ln D \leqslant l(\eta^n) < -\ln P(\eta^n)/\ln D + 1.$$

Тогда запись тех реализаций сообщения, для которых выполнено условие

$$-\ln P(\eta^n)/\ln D + 1 \leqslant L, \tag{2.4.1}$$

заведомо сможет уложиться в заданных пределах. При декодировании эти реализации будут восстановлены безошибочно. Запись некоторых реализаций не сможет уложиться. При декодировании в этом случае можно указывать любую реализацию, что, как правило, будет сопряжено с появлением ошибки. Возникает вопрос, как оценить вероятность правильного и ошибочного декодирования.

Очевидно, что вероятность ошибки декодирования не будет превосходить вероятности неравенства

$$-\ln P(\eta^n)/\ln D + 1 > L, \tag{2.4.2}$$

которое обратно неравенству (2.4.1).

Неравенство (2.4.2) можно записать в виде

$$\frac{H(\eta^n)}{H_{\eta^n}} > \frac{L-1}{H_{\eta^n}} \ln D.$$

Но для энтропийно устойчивых случайных величин отношение $H(\eta^n)/H_{\eta^n}$ стремится к единице. Принимая во внимание определение энтропийно устойчивых величин (§ 1.5), получаем отсюда следующий результат.

Теорема 2.5. *Если случайные величины η^n энтропийно устойчивы и если длина записи $L = L_n$ так растет с ростом n, что выражение*

$$\frac{L_n - 1}{H_{\eta^n}} \ln D - 1$$

остается больше некоторой положительной постоянной ε, то вероятность $P_{\text{ош}}$ ошибки декодирования стремится к нулю при $n \to \infty$.

Очевидно, что условие

$$\frac{L_n - 1}{H_{\eta^n}} \ln D - 1 > \varepsilon$$

в приведенной теореме можно заменить на более простое условие

$$L_n \ln D > (1+\varepsilon) H_{\eta^n},$$

если $H_{\eta^n} \to \infty$ при $n \to \infty$.

Не поместившиеся в записи и поэтому неправильно декодированные реализации при этом относятся к множеству A_n несущественных реализаций, о котором идет речь в теореме 1.9.

Используя результаты § 1.5, можно получить более детальные оценки для вероятности ошибки $P_{\text{ош}}$, оценить быстроту ее исчезновения.

Теорема 2.6. *Вероятность ошибки декодирования при фиксированной длине записи* $L > (H_{\eta^n}/\ln D) + 1$ *удовлетворяет неравенству*

$$P_{\text{ош}} \leqslant \frac{\mathbf{D}H(\eta^n)}{[(L-1)\ln D - H_{\eta^n}]^2}. \tag{2.4.3}$$

Доказательство. Воспользуемся неравенством Чебышева

$$\mathbf{P}\left[|\xi - \mathbf{M}\xi| \geqslant \varepsilon\right] \leqslant \mathbf{D}\xi/\varepsilon^2$$

($\varepsilon > 0$ любое) для случайной величины $\xi = H(\eta^n)/H_{\eta^n}$. Полагая

$$\varepsilon = \frac{L-1}{H_{\eta^n}}\ln D - 1,$$

отсюда имеем

$$\mathbf{P}\left[\frac{H(\eta^n)}{H_{\eta^n}} > \frac{L-1}{H_{\eta^n}}\ln D\right] \leqslant \frac{\mathbf{D}H(\eta^n)}{H^2_{\eta^n}}\left[\frac{L-1}{H_{\eta^n}}\ln D - 1\right]^{-2},$$

что доказывает (2.4.3).

Для последовательности равнораспределенных независимых сообщений из (2.4.3) получаем убывание $P_{\text{ош}}$ с ростом n по закону $1/n$, если $L - 1$ растет линейно с ростом n и дисперсия $\mathbf{D}H(\xi)$ конечна. В самом деле, подставляя

$$H_{\eta^n} = nH_\xi;\ \ \mathbf{D}H(\eta^n) = n\mathbf{D}H(\xi);\ \ L - 1 = L_1 n \tag{2.4.4}$$

в (2.4.3), находим

$$P_{\text{ош}} \leqslant \frac{1}{n}\frac{\mathbf{D}H(\xi)}{[L_1 \ln D - H_\xi]^2}. \tag{2.4.5}$$

Используя теорему 1.13, нетрудно доказать более быстрый экспоненциальный закон убывания вероятности $P_{\text{ош}}$.

Теорема 2.7. *При фиксированной длине* $L > H_\eta/\ln D + 1$ *вероятность ошибки декодирования удовлетворяет неравенству*

$$P_{\text{ош}} \leqslant e^{\mu(s) - s\mu'(s)}, \tag{2.4.6}$$

где $\mu(\alpha)$ — *потенциал, определяемый формулой* (1.5.15), а s — *положительный корень уравнения*

$$\mu'(s) = (L-1)\ln D,$$

если этот корень лежит в области определения и дифференцируемости потенциала.

Для доказательства нужно лишь воспользоваться формулой (1.5.18) теоремы 1.13, полагая

$$\varepsilon = \frac{L-1}{H_\eta} \ln D - 1.$$

При малых ε в пределах применимости формулы (1.5.22) неравенство (2.4.6) можно заменить следующим:

$$P_{\text{ош}} \leqslant \exp\left\{-\frac{[(L-1)\ln D - H_\eta]^2}{2\mathbf{D}\, H(\eta)}\right\}. \tag{2.4.7}$$

В случае одинаково распределенных независимых сообщений, когда справедливы соотношения (2.4.4) и формула (2.4.5), в силу (2.4.7) имеем такое соотношение:

$$P_{\text{ош}} \leqslant \exp\left\{-\frac{n}{2}\,\frac{(L_1 \ln D - H_\xi)^2}{\mathbf{D}\, H(\xi)}\right\}. \tag{2.4.8}$$

Оно говорит об экспоненциальном законе убывания вероятности $P_{\text{ош}}$ с ростом n.

Полученные формулы (2.4.7), (2.4.8) соответствуют такому случаю, когда распределение вероятности для энтропии можно считать приближенно гауссовым в силу центральной предельной теоремы.

Глава 3

КОДИРОВАНИЕ ПРИ НАЛИЧИИ ШТРАФОВ. ПЕРВАЯ ВАРИАЦИОННАЯ ЗАДАЧА

Количество информации, которое можно записать или передать, определяется логарифмом числа различных реализаций записи или передачи. Подсчет этого числа, однако, не всегда является простым делом. Он может осложняться наличием каких-либо ограничений, наложенных на допустимые реализации. Вместо непосредственного вычисления числа реализаций во многих случаях целесообразно вычислять максимальное значение энтропии записи, беря максимум по распределениям, совместимым с условиями, наложенными на математическое ожидание определенной случайной величины штрафов. Это максимальное значение энтропии носит название *пропускной способности* канала без помех. Описанная вариационная задача является первой из ряда основных вариационных задач, играющих большую роль в теории информации.

Соотношения, полученные в результате решения указанной вариационной задачи, оказываются аналогичными соотношениям статистической термодинамики (см., например, учебники Леонтовича [1, 2]). Для нахождения пропускной способности и оптимальных распределений удобно применять разработанные там методы. Эти методы основаны на систематическом использовании термодинамических потенциалов и формул. Функция штрафов при этом служит аналогом энергии. В теории подобающее место находят температура, свободная энергия и другие термодинамические понятия. Тем самым математический аппарат этого раздела теории информации смыкается с аппаратом статистической термодинамики.

Подобная аналогия будет наблюдаться и в дальнейшем для второй и третьей вариационных задач (§ 8.2, 9.4 соответственно). Следует подчеркнуть, что указанная аналогия имеет не только принципиальное, но и практическое значение. Она позволяет использовать методы и результаты статистической термодинамики.

Рассмотрение и решение ряда частных задач (пример 3 из § 3.4) на оптимальное кодирование со штрафами математически эквивалентно расчету одномерной модели Изинга, хорошо известной в статистической физике.

В конце главы приведено распространение результатов на более общее определение энтропии, справедливое для непрерывных и произвольных случайных величин.

3.1. Прямой способ вычисления информационной емкости записи для одного примера

Рассмотрим один пример записи информации, для которого подсчет информационной емкости или пропускной способности $\ln M$, где M — число различных вариантов записи, не тривиален. Это случай неодинаковой длительности символов.

Пусть имеются m сивмволов $V_1, \ldots, V_m$, которые имеют длину $l(1), \ldots, l(m)$ соответственно. Возможно (но необязательно), что они являются комбинациями более элементарных равнодлинных букв $A_1, \ldots, A_D$, и $l(i)$ есть число таких букв. Тогда символы V_i будут, в сущности, кодовыми словами наподобие тех, которые были рассмотрены ранее. Теперь, однако, заполнение буквами является фиксированным и не меняется в процессе нашего рассмотрения.

Возьмем для примера четыре телеграфных символа:

Символы	V_i	$l(i)$	
Точка	$+-$	2	
Тире	$+++-$	4	(3.1.1)
Промежуток между буквами	$---$	3	
Промежуток между словами	$------$	6	

Во второй колонке приведена запись этих символов в двоичном ($D=2$) алфавите: А $=+$ и Б $=-$. В третьей колонке указана их длительность в этом алфавите. Конечно, символы V_i можно называть не «словами», а «буквами» некоторого нового, более сложного алфавита.

Пусть имеется запись (или передача)

$$V_{i_1} V_{i_2}, \ldots, V_{i_k} \tag{3.1.2}$$

из k символов. Ее длина, очевидно, равна $L = l(i_1) + l(i_2) + \ldots + l(i_k)$.

Зафиксируем эту суммарную длину записи и будем подсчитывать число $M(L)$ различных реализаций записи данной длины. Отбрасывая последний символ V_{i_k}, получаем последовательность меньшей длины $L - l(i_k)$. Каждая такая укороченная последовательность может быть осуществлена $M(L - l(i_k))$ различными способами. Суммируя различные возможности, получаем уравнение

$$M(L) = \sum_{i=1}^{n} M(L - l(i)), \tag{3.1.3}$$

которое позволяет найти $M(L)$ как функцию от L. После того, как число $M(L)$ найдено, легко определить информацию, которая может быть передана записью длины L. Как и в предыдущих случаях, максимальная информация достигается, когда все из $M(L)$ возможностей равновероятны. При этом

$$H^L/L = \ln M(L)/L$$

($H^L = H_{v_{i_1} \ldots v_{i_k}}$ — энтропия записи). Взяв предел этого отношения при $L \to \infty$, получим энтропию записи, рассчитанную на единицу длины:

$$H_1 = \lim_{L \to \infty} \ln M(L)/L. \qquad (3.1.4)$$

Как видно из этой формулы, нет надобности находить точное решение уравнения (3.1.3), а достаточно рассмотреть лишь асимптотическое поведение $\ln M(L)$ при больших L. Уравнение (3.1.3) является линейным и однородным. Как и для всякого такого уравнения, можно искать решение в виде

$$M(L) = C e^{\lambda L}. \qquad (3.1.5)$$

Подобное решение обычно оказывается возможным лишь при определенных («собственных») значениях $\lambda = \lambda_1, \lambda_2, \ldots$. С помощью спектра этих «собственных» значений общее решение записывается в виде

$$M(L) = \sum_{\alpha} C_{\alpha} e^{\lambda_{\alpha} L}, \qquad (3.1.6)$$

где постоянные $C_1, C_2, \ldots$ определяются из начальных условий. Подставляя (3.1.5) в (3.1.3), получаем «характеристическое» уравнение

$$1 = \sum_{i=1}^{m} e^{-\lambda l(i)}, \qquad (3.1.7)$$

позволяющее найти собственные значения $\lambda_1, \lambda_2, \ldots$. Функция

$$\Phi(X) \equiv \sum_{i} X^{-l(i)}, \qquad (3.1.8)$$

как легко видеть, является монотонной убывающей функцией от X на полупрямой $X > 0$, причем она убывает от ∞ до 0, поскольку все $l(i) > 0$. Следовательно, существует лишь один единственный положительный корень W уравнения

$$\sum_{i} W^{-l(i)} = 1. \qquad (3.1.9)$$

Уравнения (3.1.7), (3.1.9) переходят одно в другое, если положить

$$\lambda = \ln W. \qquad (3.1.10)$$

Поэтому, уравнение (3.1.7) также имеет лишь один действительный корень.

При достаточно больших L в сумме, стоящей в правой части (3.1.6), превалирующим будет член с максимальным значением действительной части собственного значения λ_α. Отметим это собственное значение λ_m индексом m:

$$\operatorname{Re} \lambda_m = \max_\alpha \operatorname{Re} \lambda_\alpha.$$

При этом из (3.1.6) будем иметь

$$M(L) \approx C_m e^{\lambda_m L}. \tag{3.1.11}$$

Поскольку функция $M(L)$ не может быть комплексной и знакопеременной, то собственное значение обязано быть действительным: $\operatorname{Im} \lambda_m = 0$. Но (3.1.10) есть единственное действительное собственное значение. Следовательно, значение $\ln W$ и есть искомое собственное значение λ_m с максимальной действительной частью. Формула (3.1.11) принимает вид

$$M(L) \approx C_m W^L,$$

а предел (3.1.4) оказывается равным

$$H = \lambda_m = \ln W, \tag{3.1.12}$$

что дает решение поставленной задачи. Оно впервые было получено Шенноном [1].

3.2. Дискретный канал без помех и его пропускная способность

Рассмотренная ранее запись или передача в заданном алфавите, имеющая фиксированную длину, является примером дискретного *канала без помех*. Дадим здесь более общее определение этого понятия.

Пусть имеется величина y, принимающая дискретное множество Y (не обязательно с конечным числом элементов) значений. Далее на Y задана некоторая (числовая) функция $c(y)$. По причинам, которые ясны из дальнейшего, ее будем называть *функцией штрафов*.

Пусть задано некоторое число a и фиксировано условие

$$c(y) \leqslant a. \tag{3.2.1}$$

Система $[Y, c(y), a]$ полностью характеризует *дискретный канал без помех*.

Условие (3.2.1) выделяет из множества Y некоторое подмножество Y_a, число M_a элементов которого предполагается конечным. Этими элементами можно записывать или передавать информацию, и число M_a интерпретируется как число различных реализаций записи, подобное числу $M(L)$ в § 3.1.

Для частного случая, рассмотренного в § 3.1, множество Y есть множество различных записей

$$y = V_{i_1}, \ldots, V_{i_k}. \tag{3.2.2}$$

В этом случае целесообразно выбрать функцию

$$c(y) = l(i_1) + \cdots + l(i_k). \tag{3.2.3}$$

Тем самым длинные записи сильнее штрафуются, в этом косвенно отражена желательность сокращения длины записи. При таком выборе условие (3.2.1) совпадает с условием

$$l(i_1) + \cdots + l(i_k) \leqslant L \quad (L \text{ играет роль } a), \tag{3.2.4}$$

т. е. с требованием, чтобы длина записи не превосходила заданной величины L. Рассмотрение этого примера будет продолжено в § 3.4 (пример 4).

Величина M_a (число реализаций) говорит о возможностях данной системы (канала) записывать или передавать информацию. Максимальное количество информации, которое может быть передано, очевидно, равно $\ln M_a$. Это количество можно назвать *информационной емкостью* или *пропускной способностью*. Непосредственное вычисление M_a, однако, иногда сопряжено с некоторыми трудностями, как это видно из примера, рассмотренного в § 3.1. Поэтому понятие пропускной способности канала удобнее определять не как $\ln M_a$, а несколько иначе.

Введем распределение вероятностей $P(y)$ на Y и заменим условие (3.2.1) аналогичным условием для среднего значения

$$\Sigma\, c(y)\, P(y) \leqslant a. \tag{3.2.5}$$

Его нужно понимать как ограничение, наложенное на распределение $P(y)$. *Пропускную способность* C или *информационную емкость* канала $[Y, c(y), a]$ определяем как максимальное значение энтропии

$$C = \sup_{P(y)} H_y. \tag{3.2.6}$$

Максимизация здесь производится по различным распределениям $P(y)$, которые совместимы с условием (3.2.5).

Таким образом, пропускная способность канала определена как решение некоторой вариационной задачи.

Как будет видно из дальнейшего (§ 4.3), между величиной (3.2.6) и M_a имеется прямая асимптотическая связь. Если говорить о приложениях этих идей к статистической физике, то соотношение (3.2.1) (взятое со знаком равенства) соответствует микроканоническому, а (3.2.5) — каноническому распределению. Асимптотическая эквивалентность этих распределений хорошо известна в статистической физике.

Приведенные выше определения канала и его пропускной способности можно несколько модифицировать, заменив, например, неравенства (3.2.1), (3.2.6) двухсторонними неравенствами

$$a_1 \leqslant c(y) \leqslant a_2; \qquad a_1 \leqslant \sum_y c(y)\, P(y) \leqslant a_2 \tag{3.2.7}$$

или неравенствами, направленными в другую сторону (если при этом число реализаций остается конечным). Кроме того, в более сложных случаях могут быть заданы несколько числовых функций или функция со значениями из какого-либо другого пространства. Все эти модификации обычно не связаны с принципиальными изменениями, и мы не будем на них особо останавливаться.

Функция штрафов $c(y)$ в различных задачах может иметь различный физический или технический смысл. Она может описывать «стоимость» отдельных символов, указывать неодинаковые затраты, идущие на запись или передачу того или иного символа, например, различное количество краски или электроэнергии. Она может соответствовать штрафам, налагаемым на различные неблагоприятные факторы, например, штрафовать чрезмерную высоту букв и т. п. В частности, если подвергается штрафам длина символов, то выполняется (3.2.3) или $c(y) = l(y)$ при $y = i$.

Введение штрафов, платежей и т. п. типично для математической статистики, теории оптимального управления, теории игр, и мы неоднократно будем их рассматривать в дальнейшем.

Когда функция $c(y)$ действительно имеет физический смысл штрафов или потерь, затрачиваемых на передачу информации, и затраты приносят пользу, то максимальное значение энтропии в (3.2.6), как правило, имеет место при наибольших затратах из допустимых, так что в (3.2.5) или (3.2.7) можно оставить лишь знак равенства, полагая

$$\sum_y c(y) P(y) = a. \tag{3.2.8}$$

К вариационной задаче (3.2.6), (3.2.8) близко стоит другая, обратная вариационная задача: задача отыскания распределения $P(y)$, минимизирующего средние потери (риск)

$$R \equiv \sum_y c(y) P(y) = \min \tag{3.2.9}$$

при фиксированном значении энтропии

$$-\sum_y P(y) \ln P(y) = I, \tag{3.2.10}$$

где I — заданное число.

Наконец, возможен третий вариант постановки задачи: требуется максимизировать некоторую линейную комбинацию указанных выражений

$$H_y - \beta R \equiv -\sum_y P(y) \ln P(y) - \beta \sum_y c(y) P(y) = \min. \tag{3.2.11}$$

Существенно, что все три указанные постановки задачи приводят к одному и тому же решению, если параметры a, I, β должным образом согласованы между собой. Любую из указанных разновидностей мы называем первой вариационной задачей.

Эти вопросы удобно исследовать при помощи вводимых ниже термодинамических потенциалов.

3.3. Решение первой вариационной задачи. Термодинамические параметры и потенциалы

1. Сформулированные в предыдущем параграфе вариационные задачи (3.2.6), (3.2.8) и задачи (3.2.9), (3.2.10) являются типичными задачами на условный экстремум, и их можно решать методом неопределенных множителей Лагранжа. К указанным условиям необходимо добавить условие нормировки

$$\sum_y P(y) = 1. \tag{3.3.1}$$

Далее должно выполняться требование неотрицательности вероятностей

$$P(y) \geqslant 0, \quad y \in Y. \tag{3.3.2}$$

Однако его можно не вводить в число условий, так как решение, получаемое без его учета, оказывается (как показывает последующая проверка) автоматически ему удовлетворяющим. Вводя множители Лагранжа β, γ, будем искать экстремум выражения

$$K = -\sum_y P(y)\ln P(y) - \beta\sum_y c(y)\,P(y) - \gamma\sum_y P(y). \tag{3.3.3}$$

Вследствие неопределенности множителей β, γ отыскание этого экстремума эквивалентно отысканию экстремума аналогичного выражения, построенного для задачи (3.2.9), (3.2.10) или для задачи (3.2.11) при помощи других неопределенных множителей. Условия экстремума имеют вид

$$-\frac{\partial K}{\partial P(y)} = \ln P(y) + 1 + \beta c(y) + \gamma = 0. \tag{3.3.4}$$

Дифференцирование здесь проводится по тем и только по тем $P(y)$, которые отличны от нуля в экстремальном распределении $P(\cdot)$. Предполагается, что это распределение существует и единственно. То подмножество из Y, элементы которого имеют ненулевые вероятности в экстремальном распределении, обозначим $\widetilde{Y}$ и будем называть «активным» множеством. Следовательно, уравнение (3.3.4) справедливо лишь для элементов y из множества $\widetilde{Y}$. Из (3.3.4) имеем

$$P(y) = e^{-1-\gamma-\beta c(y)}, \qquad y \in \widetilde{Y}. \tag{3.3.5}$$

Для прочих y, не входящих в $\widetilde{Y}$, вероятности равны нулю.

Мы видим, что экстремальные вероятности (3.3.5) оказываются всегда неотрицательными, так что условие (3.3.2) как дополнительное условие можно не принимать во внимание. Множитель $e^{-1-\gamma}$ немедленно определяется из условия нормировки (3.3.1), что дает

$$P(y) = e^{-\beta c(y)} \Big/ \sum_y e^{-\beta c(y)}, \tag{3.3.6}$$

если сумма в знаменателе сходится.

В задаче (3.2.11) параметр β является заданным и специально его определять не требуется. В задачах (3.2.6), (3.2.8), (3.2.9), (3.2.10) его следует определить из дополнительных условий (3.2.8) или (3.2.10) соответственно.

2. Полученное решение экстремальной задачи упрощается благодаря справедливости следующей теоремы.

Теорема 3.1. *При решении задачи на максимум энтропии (3.2.6) при условии (3.2.8) в экстремальном распределении отличны от нуля вероятности всех элементов множества Y, в которых функция штрафов $c\,(y)$ принимает конечные значения.* Следовательно, если функция $c\,(y)$ для всех элементов конечна, то множество $\widetilde{Y}$ совпадает со всем множеством Y.

Доказательство. Предположим противное, т. е. что в экстремальном распределении $P_0\,(y)$ некоторые элементы $y \in Y$ имеют нулевую вероятность $P_0\,(y) = 0$. Поскольку распределение P_0 экстремальное, то для другой меры P_1 (даже ненормированной) с тем же нулевым множеством $Y - \widetilde{Y}$ справедливо соотношение

$$\begin{aligned} &-\sum_y P_1\,(y) \ln P_1\,(y) - \beta \sum_y c\,(y)\, P_1\,(y) - \gamma \sum_y P_1\,(y) + \\ &+ \sum_y P_0\,(y) \ln P_0\,(y) + \beta \sum_y c\,(y)\, P_0\,(y) + \gamma \sum_y P_0\,(y) = \\ &= -\sum_{y \in Y} \frac{1}{P_0\,(y)} [P_1\,(y) - P_0\,(y)]^2 + \cdots = O\,[(P_1 - P_0)^2]. \end{aligned} \tag{3.3.7}$$

Возьмем теперь $P(y_1) \neq 0$, $y_1 \in Y - \widetilde{Y}$ и положим $P(y) = 0$ в других точках подмножества $Y - \widetilde{Y}$. Прочие вероятности $P\,(y), y \in \widetilde{Y}$ подберем так, чтобы были выполнены условия

$$\sum_y P\,(y) = 1; \qquad \sum_y c\,(y)\, P\,(y) = \sum c\,(y)\, P_0\,(y) \tag{3.3.8}$$

и разности $P\,(y) - P_0\,(y)$ $(y \in \widetilde{Y})$ линейно зависели от $P\,(y_1)$. Ненулевая вероятность $P\,(y_1) \neq 0$ приведет к появлению в выражении для энтропии H_y дополнительного члена $-P\,(y_1) \ln P\,(y_1)$. Полагая $P_1\,(y) = P\,(y)$ при $y \in \widetilde{Y}$ и учитывая (3.3.7), получаем

$$\begin{aligned} &-\sum P\,(y) \ln P\,(y) - \beta \sum c\,(y)\, P\,(y) - \gamma \sum P\,(y) + \\ &+ \sum P_0\,(y) \ln P_0\,(y) + \beta \sum c\,(y)\, P_0\,(y) + \gamma \sum P_0\,(y) = \\ &= -P\,(y_1) \ln P\,(y_1) - [\beta c\,(y_1) + \gamma]\, P\,(y_1) + O\,[(P - P_1)^2]. \end{aligned}$$

Используя (3.3.8), будем иметь

$$\begin{aligned} &-\sum P\,(y) \ln P\,(y) + \sum P_0\,(y) \ln P_0\,(y) = P\,(y_1) \ln \frac{1}{P\,(y_1)} - \\ &- [\beta c\,(y_1) + \gamma]\, P\,(y_1) + O\,(P^2\,(y_1)). \end{aligned} \tag{3.3.9}$$

При достаточно малых $P(y_1)$ (когда $-\ln P(y_1)-\beta c(y_1)-\gamma+$ $+O(P(y_1))>0$) выражение в правой части (3.3.9) будет заведомо положительным. Следовательно, энтропия распределения P, удовлетворяющего тем же условиям (3.3.8), будет превосходить энтропию экстремального распределения P_0, что невозможно. Значит, элемента y_1, имеющего в экстремальном распределении нулевую вероятность, не должно быть. Доказательство закончено.

Итак, вследствие установленного теоремой 3.1 факта совпадения $\tilde{Y}$ и Y при конечных значениях $c(y)$, мы будем в дальнейшем пользоваться формулой (3.3.5) или (3.3.6) для всего множества Y.

Выражение (3.3.6) аналогично соответствующему выражению для распределения Больцмана, хорошо известному в статистической физике. При этом $c(y)$ является аналогом энергии, а $\beta=1/T$ является параметром, обратным абсолютной температуре (если пользоваться шкалой, в которой постоянная Больцмана равна единице). Имея в виду эту аналогию, будем называть T температурой, а $c(y)$ — энергией. Тот факт, что термодинамически равновесное распределение типа (3.3.6) является решением задачи на экстремум энтропии при фиксированной энергии, т. е. экстремальной задачи (3.2.6), (3.2.8) также является известным фактом статистической физики.

3. Проверим, что найденное распределение (3.3.5), (3.3.6) действительно соответствует именно *максимуму* (а не минимуму, скажем) энтропии H_y. Вычисляя вторые производные от энтропии, имеем

$$\frac{\partial^2 H_y}{\partial P(y)\,\partial P(y')}=-\frac{1}{P(y)}\delta_{yy'}\ (y, y'\in\tilde{Y}),$$

где $\delta_{yy'}$ — символ Кронекера. Следовательно, учитывая исчезновение первых производных и пренебрегая членами третьего порядка относительно отклонения $\Delta P(y)$, $y\in\tilde{Y}$ от экстремального распределения P, будем иметь

$$\Delta H_y=H_y-C=-\sum_{y\in\tilde{Y}}\frac{1}{P(y)}[\Delta P(y)]^2. \qquad (3.3.10)$$

Поскольку входящие сюда $P(y)$ все положительны, то разность H_y-C является отрицательной, что и доказывает максимальность C.

Можно доказать, что если $\beta>0$, то найденное выше экстремальное распределение соответствует минимуму средних штрафов (3.2.9) при фиксированной энтропии (3.2.10). Для этого нужно учесть, что для функции (3.3.3), рассматриваемой как функция независимых переменных $P(y)$, $y\in\tilde{Y}$ (пусть их число равно m), по аналогии с (3.3.10) справедливо соотношение

$$\Delta K=K-K_{\text{экстр}}=-\sum_{y\in\tilde{Y}}\frac{1}{P(y)}[\Delta P(y)]^2<0.$$

Следовательно, в экстремальной точке имеет место максимум выражения K.

Теперь нужно учесть, что нас интересует условный экстремум, соответствующий условиям (3.2.10) и (3.3.1). Последние выделяют в m-мерном пространстве R_m гиперповерхность размерности $m - 2$, проходящую через экстремальную точку. Следует ограничиться лишь смещениями от экстремальной точки внутри этой гиперповерхности. Но если для любых смещений в экстремальной точке имеет место максимум, то и частному виду смещений вдоль гиперповерхности также соответствует максимум. Поскольку при смещении внутри гиперповерхности первый и третий члены (3.3.3) остаются неизменными, значит экстремальная точка является точкой условного максимума выражения $-\beta\Sigma c(y) P(y)$, т. е. точкой минимума средних штрафов (если $\beta > 0$). Доказательство закончено.

4. Найденное распределение (3.3.6) позволяет найти пропускную способность C и средние штрафы R как функцию от β или «температуры» $T = 1/\beta$. Исключая из этих зависимостей параметр β или T, нетрудно найти зависимость C от $R = a$ (для задачи (3.2.6), (3.2.8)) или зависимость R от $C = I$ (для задачи (3.2.11), (3.2.10)). При практическом проведении расчетов удобно использовать технику, которая разработана в статистической физике. Сначала вычисляется (предполагаемая конечной) сумма

$$Z = \sum_{y} e^{-c(y)/T}, \tag{3.3.11}$$

называемая *статистической суммой*. Она является нормировочным множителем в распределении (3.3.6). Через нее выражается *свободная энергия*

$$F(T) = -T \ln Z, \tag{3.3.12}$$

при помощи которой уже вычисляется энтропия C и энергия R. Докажем ряд обычных в термодинамике соотношений в форме отдельных теорем.

Теорема 3.2. *Свободная энергия связана с энтропией и средней энергией простым соотношением*

$$F = R - TC, \tag{3.3.13}$$

хорошо известным в термодинамике.

Доказательство. Используя формулы (3.3.12), (3.3.11), записываем распределение вероятностей (3.3.6) в виде

$$P(y) = e^{\frac{F - c(y)}{T}} \tag{3.3.14}$$

(распределение Гиббса).

Согласно (1.2.1) находим случайную энтропию

$$H(y) = \frac{c(y) - F}{T}.$$

Усреднение этого равенства по y приводит к соотношению (3.3.13). Доказательство закончено.

Теорема 3.3. *Энтропию можно вычислять, дифференцируя свободную энергию по температуре*

$$C = -dF/dT. \tag{3.3.15}$$

Доказательство. Дифференцируя выражение (3.3.12) по температуре и учитывая (3.3.11), получаем

$$-\frac{dF}{dT} = \ln Z + TZ^{-1}\frac{dZ}{dT} = \ln Z + TZ^{-1}\sum_y e^{-\frac{c(y)}{T}}\frac{c(y)}{T^2},$$

т. е. в силу (3.3.6), (3.3.11)

$$-\frac{dF}{dT} = \ln Z + T^{-1}\mathbf{M}c(y).$$

Принимая во внимание теперь равенства (3.3.12), (3.3.13), находим

$$-\frac{dF}{dT} = T^{-1}(-F+R) = C,$$

что и требовалось доказать.

Как видно из формул (3.3.13), (3.3.15), энергия (средние штрафы) выражаются через свободную энергию следующим образом:

$$R = F - T\frac{dF}{dT}. \tag{3.3.16}$$

Легко проверить, что эту формулу можно записать в следующем более компактном виде:

$$R = \frac{d}{d\beta}(\beta F) = -\frac{d\ln Z}{d\beta} \qquad \left(\beta = \frac{1}{T}\right). \tag{3.3.17}$$

После вычисления функций $C(T)$, $R(T)$ легко найти пропускную способность $C = C(a)$ (3.2.6). Уравнение (3.2.8), т. е. уравнение

$$R(T) = a \tag{3.3.18}$$

определит $T(a)$, и пропускная способность будет равна $C(a) = C(T(a))$.

Аналогично для задачи (3.2.9), (3.2.10) находятся минимальные средние штрафы (3.2.9) для данного количества информации I. Уравнение (3.2.10), принимающее вид

$$C(T) = I,$$

определяет температуру $T(I)$, соответствующую данному количеству информации I. При этом минимальные средние штрафы (3.2.9) будут равны $R(T(I))$.

В заключение этого параграфа сформулируем две теоремы, касающиеся фактов, хорошо известных в термодинамике.

Теорема 3.4. *Если распределение* (3.3.6) *существует, т. е. сумма* (3.3.11) *сходится, то справедлива формула*

$$\frac{dR}{dC} = T, \tag{3.3.19}$$

так что при $T > 0$ средние штрафы являются возрастающей функцией энтропии, а при $T < 0$ — убывающей.

Доказательство. Взяв дифференциал от выражения (3.3.13), имеем

$$dR = dF + CdT + TdC.$$

Но в силу (3.3.15) $dF + CdT = 0$, так что

$$dR = TdC. \tag{3.3.20}$$

Отсюда следует искомая формула (3.3.19).

Соотношение (3.3.20) является не чем иным, как известным термодинамическим равенством

$$dH = dQ/T,$$

где dQ — количество теплоты, пришедшей в систему и давшей увеличение dR ее внутренней энергии; $H = C$ — энтропия.

Теорема 3.5. *Если распределение* (3.3.6) *существует, а* $c(y)$ *непостоянна в* Y, *то* R *является возрастающей функцией температуры. Пропускная способность (энтропия)* C *при* $T > 0$ *является возрастающей функцией от* T.

Доказательство. Дифференцируя (3.3.17), получаем

$$\frac{dR}{dT} = -\frac{1}{T^2}\frac{d^2(\beta F)}{d\beta^2}. \tag{3.3.21}$$

Если подставить сюда (3.3.20), то будем иметь

$$\frac{dC}{dT} = -\frac{1}{T^3}\frac{d^2(\beta F)}{d\beta^2}. \tag{3.3.22}$$

Чтобы доказать теорему, остается показать, что

$$-\frac{d^2(\beta F)}{d\beta^2} > 0, \tag{3.3.23}$$

т. е. что βF является выпуклой функцией от β. Согласно (3.3.12) βF есть не что иное, как $-\ln Z$. Поэтому

$$-\frac{d^2(\beta F)}{d\beta^2} = \frac{1}{Z}\frac{d^2 Z}{d\beta^2} - \left(\frac{1}{Z}\frac{dZ}{d\beta}\right)^2. \tag{3.3.24}$$

Дифференцируя (3.3.11), получаем

$$\begin{gathered}\frac{1}{Z}\frac{dZ}{d\beta} = -\sum c(y)\,e^{-\beta c(y)}/\sum e^{-\beta c(y)} = -\mathbf{M}c(y),\\ \frac{1}{Z}\frac{d^2 Z}{d\beta^2} = \sum c^2(y)\,e^{-\beta c(y)}/\sum e^{-\beta c(y)} = \mathbf{M}c^2(y)\end{gathered} \tag{3.3.25}$$

в силу (3.3.6). Следовательно, выражение (3.3.24) есть не что иное, как дисперсия штрафов: $\mathbf{M}\,[c\,(y) - \mathbf{M}c\,(y)]^2$, которая является неотрицательной величиной (положительной, если $c\,(y)$ непостоянна в Y). Это доказывает неравенство (3.3.23) и в силу (3.3.21),(3.3.22) всю теорему.

Вследствие (3.3.15) формула (3.3.22) дает

$$\left(-\frac{dC}{dT}=\right)\frac{d^2\,F}{dT^2}=\frac{1}{T^3}\,\frac{d^2\,(\beta F)}{d\beta^2}\,,$$

откуда в силу (3.3.23) можно заключить о выпуклости функции $F\,(T)$ при $T > 0$ и вогнутости при $T < 0$.

Приведенные выше факты являются частным проявлением свойств термодинамических потенциалов, примерами которых являются функции $F\,(T)$, $\ln Z\,(\beta)$. Эти потенциалы, следовательно, играют роль не только в термодинамике, но и в теории информации. В следующей главе будут затронуты вопросы, связанные с асимптотической эквивалентностью условия типа (3.2.1) и условия типа (3.2.5).

3.4. Примеры применения общих методов вычисления пропускной способности

Пример 1. Пусть для простоты сначала имеются лишь два символа: $m = 2$; $y = 1, 2$, которым соответствуют различные штрафы

$$c\,(1) = b - a, \qquad c\,(2) = b + a. \tag{3.4.1}$$

Статистическая сумма (3.3.11) в этом случае равна

$$Z = e^{-\beta b + \beta a} + e^{-\beta b - \beta a} = 2e^{-\beta b}\,\mathrm{ch}\,\beta a.$$

По формуле (3.3.12) ей соответствует свободная энергия

$$F = b - T\ln 2 - T\ln\mathrm{ch}\,\frac{a}{T}\,.$$

Применяя формулы (3.3.12), (3.3.13), находим энтропию и среднюю энергию

$$\begin{aligned} C = H_y\,(T) &= \ln 2 + \ln\mathrm{ch}\,\frac{a}{T} - \frac{a}{T}\,\mathrm{th}\,\frac{a}{T}\,, \\ R\,(T) &= b - a\,\mathrm{th}\,\frac{a}{T}\,. \end{aligned} \tag{3.4.2}$$

Графики этих функций приведены на рис. 3.1. Там же показано, как графически легко определить пропускную способность при заданном уровне затрат $R = R_0$.

При изменении температуры от 0 до ∞ энтропия меняется от 0 до $\ln 2$, а энергия от $b - a$ до b ($a > 0$). Так же и в более общем случае при $T \to \infty$ энтропия H_y имеет своим пределом $\ln m$.

Это значение, являющееся пропускной способностью, соответствующей отсутствию условий, нельзя превзойти ни при каких средних штрафах.

Аддитивный параметр b в штрафах (3.4.1) является малосущественным. От него не зависит ни энтропия, ни распределение вероятностей. Это соответствует тому факту, что аддитивная постоянная в выражении для энергии (а R является аналогом энергии) может быть выбрана произвольно. Оптимальное распределение вероятностей в данном примере согласно (3.3.14) имеет вид

$$P(1, 2) = e^{\pm a/T}/2 \operatorname{ch}(a/T). \tag{3.4.2a}$$

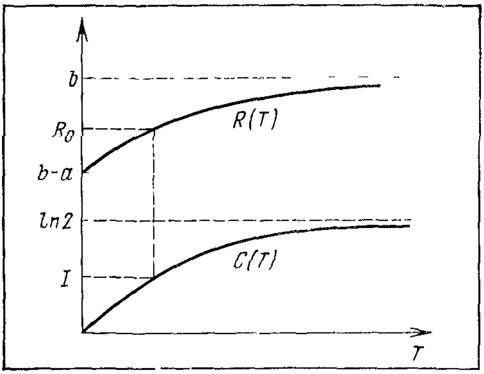


Рис. 3.1.
Термодинамические функции для примера 1.

Найденные функции (3.4.2) можно применять для расчета тех случаев, когда имеется последовательность $y^L = (y_1, \ldots, y_L)$ длины L из символов описанного вида. Если число различных элементарных символов равно 2, то число различных последовательностей, очевидно, есть $m = 2^L$. Будем предполагать, что штрафы, приходящиеся на всю последовательность, аддитивно складываются из штрафов, приходящихся на символы, из которых она состоит, так что

$$c(y^L) = \sum_{j=1}^{L} c(y_j). \tag{3.4.3}$$

Применение полученной ранее формулы (3.3.6) к последовательности показывает, что оптимальное распределение вероятностей для последовательности в этом случае распадается на произведение вероятностей для отдельных символов. При этом термодинамические функции F, R, H для последовательности равны сумме соответствующих термодинамических функций составляющих символов. В стационарном случае, когда символам, стоящим на различных местах, соответствует одинаковое распределение штрафов, имеем $F = LF_1$; $R = LR_1$; $H = LH_1$, где F_1, R_1, H_1 — функции для одиночного символа, которые были найдены ранее [см., например, (3.4.2)].

Пример 2. Рассмотрим более сложный пример, когда принцип аддитивности штрафов (3.4.3) не имеет места. Пусть выбор того или иного символа $y=1$ или $y=2$ не связан с какими-либо затратами сам по себе, но требует затрат смены символа, т. е. переключение с одного символа на другой. Если в последовательности символ 1 следует за 1 или 2 за 2, то штрафов нет; если же за 1 следует 2 или за 2 следует 1, то берется некоторый (один и тот же) штраф $2d$. Легко понять, что в этом случае суммарный штраф всей последовательности y^L записывается в виде

$$c(y^L)=2d\sum_{j=1}^{L-1}(1-\delta_{y_j y_{j+1}}). \tag{3.4.4}$$

Требуется найти условную пропускную способность такого канала и оптимальное распределение вероятностей для него. Начнем с вычисления статистической суммы (3.3.11):

$$Z=\mathrm{e}^{-2\beta d\,(L-1)}\sum_{y_1,\,\ldots,\,y_L}\prod_{j=1}^{L-1}\mathrm{e}^{2\beta d\delta_{y_j\,y_{j+1}}}. \tag{3.4.5}$$

Ее удобно записать при помощи матрицы

$$V=\|V_{yy'}\|=\begin{pmatrix}V_{11} & V_{12}\\ V_{21} & V_{22}\end{pmatrix}=\begin{pmatrix}1 & \mathrm{e}^{-2\beta d}\\ \mathrm{e}^{-2\beta d} & 1\end{pmatrix}$$

в виде

$$Z=\sum_{y_1,\,y_L}(V^{L-1})_{y_1\,y_L}. \tag{3.4.6}$$

Рассматривая ортогональное преобразование

$$U=\begin{pmatrix}1/\sqrt{2} & -1/\sqrt{2}\\ 1/\sqrt{2} & 1/\sqrt{2}\end{pmatrix},$$

приводящее указанную матрицу к диагональному виду, можно найти

$$V^{L-1}=\frac{1}{2}\begin{pmatrix}(1+k)^{L-1}+(1-k)^{L-1} & (1+k)^{L-1}-(1-k)^{L-1}\\ (1+k)^{L-1}-(1-k)^{L-1} & (1+k)^{L-1}+(1-k)^{L-1}\end{pmatrix}$$
$$(k=\mathrm{e}^{-2\beta d})$$

и

$$Z=2(1+k)^{L-1}=2^L\,\mathrm{e}^{-\beta d\,(L-1)}\,\mathrm{ch}^{L-1}\,\beta d.$$

Следовательно, по формуле (3.3.12)

$$F=-LT\ln 2+(L-1)\,d-(L-1)\,T\ln\mathrm{ch}\left(\frac{d}{T}\right).$$

В соответствии с (3.3.15), (3.3.16) отсюда получаем

$$C=L\ln 2+(L-1)\ln\mathrm{ch}\left(\frac{d}{T}\right)-(L-1)\left(\frac{d}{T}\right)\mathrm{th}\left(\frac{d}{T}\right),$$

$$R=(L-1)\,d-(L-1)\,d\,\mathrm{th}\left(\frac{d}{T}\right).$$

При помощи этих функций легко найти пропускную способность и средние штрафы, рассчитанные на один символ в асимптотическом пределе $L \to \infty$:

$$C_1 = \lim_{L\to\infty} \frac{C}{L} = \ln 2 + \ln \operatorname{ch} \frac{d}{T} - \frac{d}{T} \operatorname{th} \frac{d}{T};$$

$$R_1 = d - d \operatorname{th}(d/T).$$

Оптимальное распределение вероятностей для символов записи теперь уже не будет распадаться на произведение распределения вероятностей для отдельных символов. Другими словами, последовательность $y_1, y_2, \ldots, y_L$ не будет процессом с независимыми значениями. Она будет, однако, простой марковской цепью.

Пример 3. Перейдем к комбинированному случаю, когда имеются штрафы обоих видов: и аддитивные штрафы (3.4.3) того же вида, как и в примере 1, и парные штрафы (3.4.4). Полная функция штрафов имеет вид

$$c(y^L) = bL + a \sum_{j=1}^{L} (-1)^{y_j} + 2d \sum_{j=1}^{L-1} (1 - \delta_{y_j y_{j-1}}).$$

Вычисление статистической суммы, которая соответствует этой функции, удобно, как и в примере 2, проводить матричным методом, пользуясь несколько более сложной, чем (3.4.6), формулой

$$Z = \sum_{y_1,\, y_L} \exp\{-\beta a (-1)^{y_1}\} (V^{L-1})_{y_1 y_L}.$$

Теперь, однако, матрица имеет более сложный вид:

$$V = e^{-\beta b - \beta d} \begin{pmatrix} e^{\beta a + \beta d} & e^{-\beta a - \beta d} \\ e^{\beta a - \beta d} & e^{-\beta a + \beta d} \end{pmatrix}. \tag{3.4.7}$$

Практический интерес представляет асимптотический случай больших L. Если заранее ограничиться этим случаем, то вычисления сильно упростятся. Наиболее существенную роль в сумме Z будут играть члены, соответствующие наибольшему собственному значению λ_m матрицы (3.4.7). Как и в других подобных случаях (см., например, § 3.1), при взятии логарифма $\ln Z$ и делении его на L в пределе окажут влияние лишь эти члены, и мы будем иметь

$$\lim_{L\to\infty} \frac{\ln Z}{L} = \ln \lambda_m. \tag{3.4.8}$$

Матрице (3.4.7) соответствует характеристическое уравнение

$$\begin{vmatrix} e^{\beta a + \beta d} - \lambda' & e^{-\beta a - \beta d} \\ e^{\beta a - \beta d} & e^{-\beta a + \beta d} - \lambda' \end{vmatrix} = 0 \qquad (\lambda' = e^{-\beta b - \beta d}\lambda),$$

т. е. уравнение

$$\lambda'^2 - 2\lambda' e^{\beta d} \operatorname{ch} \beta a + e^{2\beta d} - e^{-2\beta d} = 0.$$

Взяв наибольший корень этого уравнения и учитывая (3.4.8), находим предельную свободную энергию, рассчитанную на один символ

$$F_1 = -T \ln \lambda_m = b - T \ln \left[\operatorname{ch} \frac{a}{T} + \sqrt{\operatorname{sh}^2 \frac{a}{T} + e^{-4d/T}} \right].$$

Обычным способом отсюда можно получить среднюю энергию, рассчитанную на один символ, и соответствующую ей пропускную способность.

Как и в примере 2, оптимальное распределение вероятностей соответствует марковскому процессу. Соответствующая ему вероятность перехода $P(y_{j+1}|y_j)$ связана с матрицей (3.4.7) и отличается от нее нормировочными множителями. Приведем результирующие выражения

$$P(1\,|\,1) = A_1 e^{\beta a + \beta d}; \qquad P(2\,|\,1) = A_1 e^{-\beta a - \beta d};$$
$$P(1\,|\,2) = A_2 e^{\beta a - \beta d}; \qquad P(2\,|\,2) = A_2 e^{-\beta a + \beta d}$$
$$(A_1^{-1} = e^{\beta a + \beta d} + e^{-\beta a - \beta d}; \qquad A_2^{-1} = e^{\beta a - \beta d} + e^{-\beta a + \beta d}).$$

Статистические системы, рассмотренные в двух последних примерах, исследовались в статистической физике под названием «модели Изинга» (см., например, Хилл [1], Стратонович [2]).

Пример 4. Рассмотрим теперь методами изложенной общей теории тот случай различных длительностей символов, который был исследован в § 3.1. Под величиной y будем понимать совокупность величин k, V_{i_1}, ..., V_{i_k}, где k — число последовательных символов в записи, а V_{i_j} — вид символа, стоящего на j-м месте. Если $l(i)$ — длительность символа вида i, то в качестве функции штрафа следует взять функцию (3.2.4).

Согласно общему методу вычислим для данного примера статистическую сумму

$$Z = \sum_k \sum_{i_1, \ldots, i_k} \exp\left[-\beta l(i_1) - \ldots - \beta l(i_k)\right] = \sum_{k=0}^{\infty} Z_1^k(\beta) =$$
$$= \frac{1}{1 - Z_1(\beta)}$$

(если выполнено условие сходимости $Z_1(\beta) < 1$). Здесь обозначено

$$Z_1(\beta) = \sum_i e^{-\beta l(i)}. \tag{3.4.9}$$

Применяя формулы (3.3.17), (3.3.13), получаем

$$R = \frac{d}{d\beta} \ln(1 - Z_1) = -\frac{dZ_1}{d\beta}(\beta)(1 - Z_1)^{-1},$$

(3.4.10)

$$C = \beta \ln(1 - Z_1) - \beta \frac{dZ_1}{d\beta}(1 - Z_1)^{-1} \quad (F = T \ln(1 - Z_1)).$$

Пусть L — фиксированная длина записи. Тогда условие (3.2.8), (3.3.18) будет иметь вид уравнения

$$-[1 - Z_1(\beta)]^{-1} \frac{dZ_1(\beta)}{d\beta} = L, \tag{3.4.11}$$

из которого определяется β.

Формулы (3.4.10), (3.3.11), (3.4.9) дают решение задачи вычисления пропускной способности $C(L)$. Представляет интерес также найти удельную пропускную способность

$$C_1 = C/L = C/R = \beta - \beta(F/R) \tag{3.4.12}$$

[использовано (3.3.13)] и особенно ее предельное значение при $L = R \to \infty$.

Дифференцируя (3.4.9), легко получить, что $-Z_1^{-1}\, dZ_1/d\beta$ имеет смысл средней длины символа:

$$-\frac{1}{Z_1}\frac{dZ_1}{d\beta} = \sum_i l(i)\, P(i) = l_{\text{ср}}$$

по аналогии с (3.3.25). Поэтому уравнению (3.4.11) можно придать вид

$$Z_1(\beta)/(1 - Z_1(\beta)) = L/l_{\text{ср}},$$

из которого видно, что $Z_1(\beta)$ стремится к единице:

$$Z_1(\beta) \to 1 \text{ при } L/l_{\text{ср}} \to \infty. \tag{3.4.13}$$

Но, очевидно,

$$(1 - Z_1)\ln(1 - Z_1) \to 0 \quad \text{при} \quad Z_1 \to 1.$$

Следовательно, в силу (3.4.10),

$$-\frac{\beta F}{R} = \frac{1}{dZ_1/d\beta}(1 - Z_1)\ln(1 - Z_1) \to 0 \quad \text{при } L \to \infty, \tag{3.4.14}$$

если $dZ_1/d\beta$ (т. е. $l_{\text{ср}}$) при этом стремится к конечному значению $dZ_1(\beta_0)/d\beta$. Согласно (3.4.12) и (3.4.14) имеем в этом случае

$$C_1 = \lim \beta = \beta_0. \tag{3.4.15}$$

Предельное значение β_0 в силу (3.4.13) определяется из уравнения

$$Z_1(\beta_0) = 1. \tag{3.4.16}$$

Это уравнение вследствие (3.4.9) есть не что иное, как полученные ранее уравнения (3.1.7), (3.1.9). Формула (3.4.15) при этом совпадает с соотношением (3.1.12). Итак, мы видим, что общий стандартизованный метод приводит к тем же результатам, что и примененный ранее специальный метод.

3.5. Метод потенциалов в случае бо́льшего числа параметров

Изложенный в § 3.3, 3.4 метод потенциалов может быть обобщен на более сложные случаи, когда имеется большее число внешних параметров. При этом данный метод еще в большей степени перерастает в методы, обычно применяемые в статистической термодинамике. Наметим здесь пути этого обобщения, отложив более подробное рассмотрение до следующей главы.

Пусть функция штрафов $c(y) = c(y, a)$ зависит теперь от числового параметра a и является дифференцируемой по этому параметру. Тогда свободная энергия

$$F(T, a) = -T \ln \sum_y e^{-c(y,a)/T} \tag{3.5.1}$$

и прочие функции, введенные в § 3.4, будут зависеть не только от температуры T (или параметра $\beta = 1/T$), но и от значения a. То же самое относится и к оптимальному распределению (3.3.14), имеющему теперь вид

$$P(y \mid a) = \exp\left[\frac{F(T, a) - c(y, a)}{T}\right]. \tag{3.5.2}$$

Формула (3.3.15) и прочие результаты из § 3.3, разумеется, сохранят свое значение, если при вариации параметра T параметр a считать постоянным, т. е. обыкновенные производные заменить частными.

Энтропия распределения (3.5.2), следовательно, равна

$$H_y = -\frac{\partial F(T, a)}{\partial T}. \tag{3.5.3}$$

В дополнение к этим результатам теперь можно получить формулу для частных производных по новому параметру a.

Дифференцируя (3.5.1) по a, находим

$$\frac{\partial F(T, a)}{\partial a} = \left[\sum_y \exp\left[-\frac{c(y, a)}{T}\right]\right]^{-1} \sum \frac{\partial c(y, a)}{\partial a} \exp\left[-\frac{c(y, a)}{T}\right]$$

или

$$\frac{\partial F(T, a)}{\partial a} = \sum_y \frac{\partial c(y, a)}{\partial a} P(y \mid a) \equiv -\mathbf{M}[B(y) \mid a], \tag{3.5.4}$$

если учесть (3.5.2), (3.5.1).

Функция

$$B(y) = -\frac{\partial c(y, a)}{\partial a}$$

называется случайным *внутренним термодинамическим параметром, сопряженным с внешним параметром a.*

Вследствие формул (3.5.3), (3.5.4) полный дифференциал свободной энергии можно записать:

$$dF(T, a) = -H_y\,dT - A\,da. \tag{3.5.5}$$

Отсюда, в частности, имеем

$$A = -\frac{\partial F}{\partial a}. \tag{3.5.5a}$$

Внутренний параметр A, определяемый подобной формулой, называется *сопряженным* с параметром a относительно потенциала F.

В формуле (3.5.5) H_y и $A = \mathbf{M}B$ мыслятся как функции от T и a. Их, однако, можно интерпретировать как независимые переменные, а вместо $F(T, a)$ рассматривать другой потенциал $\Phi(H_y, A)$, выражающийся через $F(T, a)$ преобразованием Лежандра:

$$\Phi(H_y, A) = F + H_y T + Aa = F - H_y\frac{\partial F}{\partial U_y} - a\frac{\partial F}{\partial a}. \tag{3.5.6}$$

Тогда параметры T, a будут получаться как функции от H_y, A дифференцированием:

$$T = \frac{\partial \Phi}{\partial H_y}, \quad a = \frac{\partial \Phi}{\partial A}, \tag{3.5.7}$$

поскольку из (3.5.5), (3.5.6) будем иметь $d\Phi = T\,dH_y + a\,dA$.

Преобразование Лежандра можно, конечно, совершить и по одной из двух переменных. Указанными потенциалами и их производными удобно пользоваться при решении различных вариационных задач, связанных с условной пропускной способностью. В качестве независимых переменных удобно брать те переменные, которые заданы в условии задачи.

Из многих возможных задач, касающихся выбора оптимального кодирования при наличии штрафов с учетом ряда условий, наложенных на количество информации, средние штрафы и прочее, мы рассмотрим один иллюстрационный пример.

Пример. Пусть сообщение следует закодировать символами ρ. Каждому символу ρ приписывается некоторый штраф $c_0(\rho)$. Кроме этого штрафа пусть требуется учесть еще один дополнительный расход $\eta(\rho)$, скажем, количество краски, которое идет на данный символ. Если ввести стоимость краски a, то полные штрафы будут иметь вид

$$c(\rho) = c_0(\rho) + a\eta(\rho). \tag{3.5.8}$$

Пусть требуется найти такое кодирование, при котором записывалось бы (передавалось) заданное количество информации I (в расчете на один символ), при этом тратилось бы фиксированное количество K краски (в среднем на один символ) и, кроме того, минимизировались бы средние затраты $\mathbf{M}c_0(\rho)$.

Для решения этой задачи вводятся параметры T и a как вспомогательные, которые сначала являются неопределенными, а затем находятся из дополнительных условий. Расход краски η (ρ) на символ ρ мы берем в качестве случайной переменной η. В качестве второй переменной ζ выбираем переменную, дополняющую η до ρ, так что $\rho = (\eta, \zeta)$. Штрафы (3.5.8) при этом можно записать

$$c(\eta, \zeta) = c_0(\eta, \zeta) + a\eta.$$

Теперь мы можем применить к данной задаче формулы (3.5.1)—(3.5.7) и другие при $B = -\eta$. Свободная энергия $F(T, a)$ согласно (3.5.1) находится по формуле

$$F(T, a) = -T \ln \sum_{\eta, \zeta} \exp[-\beta c_0(\eta, \zeta) - \beta a\eta] \quad \left(\beta = \frac{1}{T}\right), \tag{3.5.9}$$

а оптимальное распределение (3.5.2) имеет вид

$$P(\eta, \zeta) = \exp[\beta F - \beta c_0(\eta, \zeta) - \beta a\eta]. \tag{3.5.10}$$

Для окончательного определения вероятностей, энтропии и прочих величин остается конкретизировать параметры T и a, пользуясь условиями, сформулированными ранее. Именно, средняя энтропия $H_{\eta\zeta}$ и средний расход краски $\mathbf{M}\,\eta$ предполагаются фиксированными:

$$H_{\eta\zeta} = I; \qquad A \equiv \mathbf{M}\eta = K. \tag{3.5.11}$$

Используя формулы (3.5.3) и (3.5.4), получаем систему двух уравнений

$$-\frac{\partial F(T, a)}{\partial T} = I; \qquad \frac{\partial F(T, a)}{\partial a} = K \tag{3.5.12}$$

для определения параметров $T = T(I, K)$ и $a = a(I, K)$.

Оптимальное распределение (3. 5.10) минимизирует как полные средние штрафы $R = \mathbf{M}c(\eta, \zeta)$, так и частичные средние штрафы

$$\Phi = \mathbf{M}c_0(\eta, \zeta) = R - a\mathbf{M}\eta \equiv R + aA,$$

поскольку при вариациях разность $R - \Phi = aK$ остается постоянной. Ввиду того, что $R = F + TH_\eta$, легко видеть, что частичные штрафы $\Phi = F + TH_\eta + aA$, рассматриваемые как функция от H_η, A, получаются из $F(T, a)$ преобразованием Лежандра (3.5.6), т. е. являются примером потенциала $\Phi(H_y, A)$. Учитывая формулы (3.5.7), оптимальное распределение (3.5.10) можно записать в следующей форме:

$$P(\rho) = \text{const} \exp\left[-\frac{c_0(\rho) - \eta \dfrac{\partial \Phi}{\partial A}(I, -K)}{\dfrac{\partial \Phi}{\partial H_\eta}(I, -K)}\right].$$

После полного определения оптимальных вероятностей $P(\rho)$, можно осуществить фактическое кодирование по рецептам § 3.6.

3.6. Пропускная способность канала без шумов со штрафами в обобщенной версии

1. В § 3.2, 3.3 рассматривалась пропускная способность дискретного канала без шумов, но со штрафами. Как указывалось, ее вычисление сводится к решению первой вариационной задачи теории информации. Приведенные там результаты могут быть обобщены на случай произвольных случайных величин, могущих принимать, в частности, непрерывные значения. Некоторые относящиеся к обобщенной версии формулы, приведенные в § 1.6, подсказывают, как это сделать.

Будем считать, что задан канал без шумов (необязательно дискретный), если задано измеримое пространство (X, F), значения из которого может принимать величина ξ, и F-измеримая функция $c(\xi)$, называемая функцией штрафов, а также мера ν на (X, F) (нормировка которой не требуется). Для уровня потерь a определяем пропускную способность $C(a)$ как максимальное значение энтропии (1.6.9):

$$H_\xi = -\int P(d\xi) \ln \frac{P(d\xi)}{\nu(d\xi)}, \tag{3.6.1}$$

совместимое с условием

$$\int c(\xi) P(d\xi) = a. \tag{3.6.2}$$

Данная вариационная задача решается в принципе так же, как это делалось в § 3.3, частные производные при этом заменяются вариационными производными. После вариационного дифференцирования по $P(dx)$ вместо (3.3.4) будем иметь условие экстремума

$$\ln \frac{P(d\xi)}{\nu(d\xi)} = \beta F - \beta c(\xi), \tag{3.6.3}$$

где $\beta F = -1 - \gamma$.

Отсюда получаем экстремальное распределение

$$P(d\xi) = e^{\beta F - \beta c(\xi)} \nu(d\xi). \tag{3.6.3а}$$

Усредняя (3.6.3) и учитывая (3.6.1), (3.6.2), находим

$$H_\xi = \beta \mathsf{M} c(\xi) - \beta F, \qquad C = \beta R - \beta F. \tag{3.6.3б}$$

Последняя формула совпадает с равенством (3.3.13) дискретной версии. Как и в § 3.3, можно ввести статистическую сумму (интеграл)

$$Z = \int e^{-c(\xi)/T} \nu(d\xi), \tag{3.6.4}$$

служащую обобщением (3.3.11), и свободную энергию $F(T) = -T \ln Z$.

Таким же способом, как и в § 3.3, доказывается соотношение (3.3.15) и прочие результаты. Аналогичным образом обобщаются на данный случай формулы из § 3.5. При этом формулы (3.5.1), (3.5.2) заменяются на

$$F(T, a) = -T \ln \int e^{-c(\xi, a)/T} \nu(d\xi), \tag{3.6.5}$$

$$P(d\xi \mid a) = \exp\left[\frac{F(T,a) - c(\xi, a)}{T}\right] \nu(d\xi). \tag{3.6.6}$$

Результирующие же равенства (3.5.3), (3.5.5) и другие остаются неизменными.

2. Рассмотрим в качестве примера тот случай, когда X — r-мерное действительное пространство $\xi = (x_1, \ldots, x_r)$, а функция $c(\xi)$ имеет вид линейной квадратичной формы

$$c(\xi) = c_0 + \frac{1}{2} \sum_{i,\, j=1}^{r} (x_i - b_i)\, c_{ij}\, (x_j - b_j),$$

где $\|c_{ij}\|$ — невырожденная положительно определенная матрица.

Будем полагать, что $\nu(d\xi) = d\xi = dx_1, \ldots, dx_r$. Тогда формула (3.6.2) примет вид

$$Z \equiv e^{-F/T} \int \exp\left[-c_0\beta - \frac{\beta}{2} \sum y_i c_{ij} y_j\right] dy_1, \ldots, dy_r$$
$$(y_i = x_i - b_i).$$

Вычисляя интеграл и логарифмируя, находим

$$F(T) = c_0 + \frac{T}{2} \ln \det \|\beta c_{ij}\| = c_0 - \frac{rT}{2} \ln T + \frac{T}{2} \ln \det \|c_{ij}\|.$$

Следовательно, уравнение (3.3.18) при учете (3.3.16) принимает вид

$$c_0 + rT/2 = a, \tag{3.6.7}$$

а формула (3.3.15) дает

$$C(a) = \frac{r}{2}(1 + \ln T) - \frac{1}{2} \ln \det \|c_{ij}\|,$$

т. е. в силу (3.6.7)

$$C(a) = \frac{r}{2}\left[1 - \ln \frac{r}{2(a - c_0)}\right] - \frac{1}{2} \ln \det \|c_{ij}\|.$$

Экстремальное распределение в этом случае является гауссовым, и его энтропия $C(a)$ может быть также найдена при помощи формул § 5.4.

Глава 4

ПЕРВАЯ АСИМПТОТИЧЕСКАЯ ТЕОРЕМА И СВЯЗАННЫЕ С НЕЙ РЕЗУЛЬТАТЫ

В предыдущей главе для одного примера (см. § 3.1, 3.4) было показано, что при подсчете максимальной энтропии (т. е. пропускной способности каналов без помех) ограничение $c(y) \leqslant a$, наложенное на допустимые реализации, при достаточно большой записи эквивалентно ограничению $\mathsf{M}c(y) \leqslant a$ для среднего значения $\mathsf{M}c(y)$. В настоящей главе будет доказано (§ 4.3), что такая эквивалентность имеет место при определенных предположениях в общем случае. Это составляет содержание первой асимптотической теоремы. Кроме нее в дальнейшем будут рассмотрены еще две асимптотические теоремы (гл. 7 и 11). Указанные асимптотические теоремы представляют собой наиболее глубокие результаты теории информации. Всем им свойственна следующая общая черта: в конечном счете они утверждают, что для предельно больших систем исчезает разница между дискретным и непрерывным, что характеристики массовых дискретных объектов могут быть подсчитаны при помощи непрерывных функциональных зависимостей, затрагивающих усредненные величины. Применительно к первой вариационной задаче это выражается в том, что дискретная функция $H = \ln M$ от a, имеющая место при ограничении $c(y) \leqslant a$, асимптотически заменяется на непрерывную функцию $H(a)$, подсчитанную при решении первой вариационной задачи.

По методу доказательства первая асимптотическая теорема оказывается родственной теореме об устойчивости канонического распределения (§ 4.2), важной в статистической термодинамике. Там ее, по существу, доказывают при выводе канонического распределения из микроканонического. Здесь мы рассматриваем ее в более общей и абстрактной форме. Родственность первой асимптотической теоремы и теоремы о каноническом распределении еще раз говорит о единстве математического аппарата теории информации и статистической термодинамики.

В процессе доказательства указанных теорем используется потенциал $\Gamma(\alpha)$ и его свойства. Нужный для этого материал излагается в § 4.1. Он связан с содержанием § 3.3. Однако вместо обычной физической свободной энергии F рассматривается безразмерная свободная энергия — потенциал $\Gamma = -F/T$. Вместо обычных в термодинамике параметров T, a_2, a_3, ... вводятся симметрично определенные параметры $\alpha_1 = -1/T$, $\alpha_2 = a_2/T$, $\alpha_3 = a_3/T$, ... При таком выборе температура выступает как рядовой параметр наряду с прочими.

При систематическом использовании «термодинамических» потенциалов оказывается удобным трактовать логарифм $\mu(s) = \ln \Theta(s)$ характеристической функции $\Theta(s)$ как характеристический потенциал. В самом деле, он является производящей функцией семиинвариантов, как и потенциал $\Gamma(\alpha)$.

В § 4.4 излагаются вспомогательные теоремы, касающиеся характеристического потенциала, которые используются в дальнейшем при доказательстве второй и третьей асимптотических теорем.

4.1. Потенциал Г или производящая функция семиинвариантов

Для термодинамической системы или информационной системы, к которой относится формула (3.5.1), введем симметрично определенные параметры $\alpha_1, \ldots, \alpha_r$ и соответствующий им потенциал $\Gamma(\alpha)$, который математически эквивалентен свободной энергии F, но имеет перед ней то преимущество, что является производящей функцией семиинвариантов.

Для физической термодинамической системы равновесное распределение обычно задается формулой Гиббса

$$P(d\zeta) = \exp\left[\frac{F - \mathscr{H}(\zeta, a)}{T}\right] d\zeta, \tag{4.1.1}$$

где $\zeta = (p, q)$ — динамические переменные — координаты и импульсы; $\mathscr{H}(\zeta, a)$ — функция Гамильтона, зависящая от параметров $a_2, \ldots, a_r$. Формула (4.1.1) является аналогом формулы (3.5.2). Предположим, что функция Гамильтона линейна по указанным параметрам

$$\mathscr{H}(\zeta, a) = \mathscr{H}_0(\zeta) - a_2 F^2(\zeta) - \ldots - a_r F^r(\zeta). \tag{4.1.2}$$

Тогда (4.1.1) примет вид

$$P(d\zeta) = \exp[F - \mathscr{H}_0(\zeta) + a_2 F^2(\zeta) + \ldots + a_r F^r(\zeta)]\, d\zeta. \tag{4.1.3}$$

Введем новые параметры

$$\alpha_1 = -1/T \equiv -\beta; \quad \alpha_2 = a_2/T; \ \ldots; \ \alpha_r = a_r/T, \tag{4.1.4}$$

которые будем называть *каноническими внешними параметрами*. Параметры

$$B^1 = \mathscr{H}_0; \qquad B^2 = F^2; \ \ldots; \qquad B^r = F^r \tag{4.1.5}$$

называем *случайными внутренними параметрами*, а

$$A^1 = \mathbf{M}\mathscr{H}_0; \qquad A^2 = \mathbf{M}F^2; \ \ldots; \ A^r = \mathbf{M}F^r$$

—*каноническими внутренними* параметрами, *сопряженными с внешними параметрами* $\alpha_1, \ldots, \alpha_r$. Вводя также потенциал Γ, запишем распределение (4.1.3) в каноническом виде

$$P(d\zeta|d) = \exp\left[-\Gamma(\alpha) + B^1(\zeta)\,\alpha_1 + \ldots + B^r(\zeta)\,\alpha_r\right] d\zeta. \tag{4.1.6}$$

Здесь $-\Gamma(\alpha) = F/T$, или, как следует из условия нормировки,

$$\Gamma(\alpha) = \ln Z, \tag{4.1.7}$$

где

$$Z = \int \exp\left[B^1\alpha_1 + \ldots + B^r\alpha_r\right] d\zeta \tag{4.1.8}$$

— соответствующая статистическая сумма.

В симметричной форме (4.1.6) температурный параметр α_1 выступает на равных правах с другими параметрами $\alpha_2, \ldots, \alpha_r$, которые связаны с $a_2, \ldots, a_r$. В этой форме представляется также (при условии линейности $c(y, a)$ по a) выражение (3.5.2).

Проинтегрируем (4.1.6) по переменным η, таким, что (B, η) в совокупности образуют ζ. Тогда получим распределение по случайным внутренним параметрам

$$P(dB\,|\,\alpha) = \exp\left[-\Gamma(\alpha) + B\alpha - \Phi(B)\right] dB, \tag{4.1.9}$$

где

$$B\alpha = B^1\alpha_1 + \ldots + B^r\alpha_r; \qquad \int\limits_{\Delta B} e^{-\Phi(B)}\,dB = \int\limits_{\Delta B} d\zeta.$$

Его, в свою очередь, назовем каноническим.

В случае канонического распределения (4.1.6) легко выразить характеристическую функцию

$$\Theta(iu) = \int e^{iuB}\,P(dB\,|\,\alpha) \tag{4.1.10}$$

случайных величин $B_1, \ldots, B_r$ через Γ. В самом деле, подставляя (4.1.9) в (4.1.10), имеем

$$\Theta(iu) = e^{-\Gamma(\alpha)} \int \exp\left[(iu + \alpha)\,B - \Phi(B)\right] dB.$$

Отсюда, учитывая, что

$$e^{\Gamma(\alpha)} = \int \exp\left[B\alpha - \Phi(B)\right] dB \tag{4.1.10a}$$

в силу нормированности распределения (4.1.9), получаем

$$\Theta(iu) = \exp\left[\Gamma(\alpha + iu) - \Gamma(\alpha)\right]. \tag{4.1.11}$$

Логарифм

$$\mu(s) = \ln\Theta(s) = \ln\int e^{sB(\zeta)}\,P(d\zeta\,|\,\alpha) \tag{4.1.11a}$$

характеристической функции (4.1.10) представляет собой характеристический потенциал или производящую функцию семиинвариантов (кумулянтов), поскольку они, как известно, вычисляются его дифференцированием:

$$k_{j_1 \ldots j_m} = \frac{\partial^m \mu}{\partial s_{j_1} \ldots \partial s_{j_m}}(0). \qquad (4.1.12)$$

Подставляя сюда вытекающее из (4.1.11) равенство

$$\boldsymbol{\mu}(s) = \Gamma(\alpha + s) - \Gamma(\alpha), \qquad (4.1.12а)$$

находим окончательно

$$k_{j_1 \ldots j_m} = \frac{\partial^m \Gamma(\alpha)}{\partial \alpha_{j_1} \ldots \partial \alpha_{j_m}}. \qquad (4.1.13)$$

Отсюда видно, что потенциал $\Gamma(\alpha)$ является производящей функцией семиинвариантов для целого семейства распределений $P(dB|\alpha)$.

При $m = 1$ из (4.1.13) имеем

$$k_j \equiv A_j \equiv \mathbf{M}B_j - \frac{\partial \Gamma(\alpha)}{\partial \alpha_j}. \qquad (4.1.14)$$

Первая из этих формул

$$\mathbf{M}B_1 \equiv A_1 = \frac{\partial \Gamma(\alpha)}{\partial \alpha_1}$$

эквивалентна формулам (3.3.16), (3.3.17). Остальные формулы

$$A_l = \frac{\partial \Gamma(\alpha)}{\partial \alpha_l}, \qquad l = 2, \ldots, r$$

эквивалентны равенствам

$$A_l = -\frac{\partial F}{\partial a_l}, \qquad l = 2, \ldots, r,$$

типа (3.5.5а).

При данном определении параметров, когда энергия (средний штраф) и температура имеют вид рядовых параметров, своеобразную форму имеет соотношение, определяющее энтропию. Подставляя (4.1.1) в формулу для физической энтропии

$$H = -\mathbf{M}\ln\frac{P(d\zeta)}{d\zeta}$$

или, взяв аналогичную формулу для дискретной версии, имеем

$$H = -F/T + (1/T)\,\mathbf{M}\mathscr{H}(\zeta, a). \qquad (4.1.14а)$$

Подставляя сюда (4.1.2) и учитывая обозначения (4.1.4), (4.1.5), получаем

$$H = \Gamma - \alpha\mathbf{M}B$$

или, если учесть (4.1.14),

$$H = \Gamma(\alpha) - \sum_{j=1}^{r} \alpha_j \frac{\partial \Gamma(\alpha)}{\partial \alpha_j}. \qquad (4.1.14б)$$

Приведем еще одно следствие из формулы (4.1.13).

Теорема 4.1. *Канонический потенциал* $\Gamma(\alpha)$ *является вогнутой функцией от параметров* α *в области* Q_α *определения параметров.*

Доказательство проведем при дополнительном условии двукратной дифференцируемой потенциала. Воспользуемся формулой (4.1.13) при $m = 2$, принимающей вид

$$k_{ij} = \frac{\partial^2 \Gamma(\alpha)}{\partial \alpha_i \, \partial \alpha_j}, \tag{4.1.15}$$

и учтем, что корреляционная матрица k_{ij} является неотрицательно определенной. Поэтому и матрица вторых производных $\partial^2\Gamma/\partial\alpha_i\partial\alpha_j$ является неотрицательно определенной, что доказывает вогнутость потенциала. Доказательство закончено.

Следствие. *При наличии лишь одного параметра* α_1, $r = 1$, *функция* $H(A_1)$, *определяемая преобразованием Лежандра*

$$H(A_1) = \Gamma - \alpha_1(A_1)\,A_1 \qquad \left(A_1 = \frac{d\Gamma}{d\alpha_1}(\alpha_1)\right), \tag{4.1.16}$$

является выпуклой.

В самом деле, как следует из формулы (4.1.16) и формулы для обратного преобразования Лежандра

$$\Gamma(\alpha_1) = H + \alpha_1 \, A_1(\alpha_1) \qquad \left(\alpha_1 = -\frac{dH}{dA_1}(A_1)\right),$$

справедливы соотношения

$$\frac{d^2\,\Gamma(\alpha_1)}{d\alpha_1^2} = \frac{dA_1}{d\alpha_1}; \qquad \frac{d^2\,H(A_1)}{dA_1^2} = -\frac{d\alpha_1}{dA_1}.$$

Так как в силу теоремы 4.1 $d^2\Gamma/d\alpha_1^2 \geqslant 0$ (при условии дифференцируемости), то $dA_1/d\alpha_1 \geqslant 0$, и, следовательно,

$$\frac{d^2 H}{dA_1^2} = -\frac{d\alpha_1}{dA_1} \leqslant 0. \tag{4.1.16a}$$

Это утверждение можно доказать также, не используя условие дифференцируемости.

В заключение параграфа приведем формулы, касающиеся характеристического потенциала энтропии, определенного ранее формулой типа (1.5.15), но относящиеся теперь к обобщенной версии.

Теорема 4.2. *Для канонического семейства распределений*

$$P(d\zeta \mid \alpha) = e^{-\Gamma(\alpha) + \alpha B(\zeta)}\, \nu(d\zeta) \tag{4.1.17}$$

характеристический потенциал

$$\mu_0(s_0) = \ln \int e^{s_0 H(\zeta \mid \alpha)}\, P(d\zeta \mid \alpha) \tag{4.1.18}$$

случайной энтропии

$$H(\zeta \mid \alpha) = -\ln\left[\frac{P(d\zeta \mid \alpha)}{\nu(d\zeta)}\right]$$

имеет вид

$$\mu_0(s_0) = \Gamma((1-s_0)\alpha_1, \ldots, (1-s_0)\alpha_r) - \\ -(1-s_0)\Gamma(\alpha_1, \ldots, \alpha_r). \qquad (4.1.19)$$

Для доказательства достаточно подставить (4.1.17) в (4.1.18) и учесть формулу $\Gamma(\alpha) = \ln \int e^{\alpha B(\zeta)} \nu(d\zeta)$ типа (4.1.10а), определяющую $\Gamma(\alpha)$.

Дифференцируя (4.1.19) по s_0 и приравнивая s_0 нулю (по аналогии с (4.1.12) при $m = 1$), можно найти среднее значение энтропии, совпадающее с (4.1.14б). Повторным дифференцированием можно получить выражение для дисперсии.

4.2. Некоторые асимптотические результаты статистической термодинамики. Устойчивость канонического распределения

Наиболее глубокие результаты теории информации и статистической термодинамики носят асимптотический характер, т. е. имеют вид предельных теорем при увеличении совокупной системы. Прежде чем рассматривать первую асимптотическую теорему теории информации, приведем родственный (как это видно из доказательства) результат статистической термодинамики, а именно: важную теорему об устойчивости канонического распределения. В случае одного параметра последнее имеет вид

$$P(\xi \mid \alpha) = \exp[-\Gamma(\alpha) + \alpha B(\xi) - \varphi(\xi)]. \qquad (4.2.1)$$

Если под $B(\xi) = \mathscr{H}(p, q)$ понимать энергию системы — функцию Гамильтона, а $\varphi(\xi)$ полагать равной нулю, то указанное распределение принимает вид канонического распределения Гиббса:

$$\exp\left[\frac{F(T) - \mathscr{H}(p, q)}{T}\right] \qquad \left(F(T) = -T\Gamma\left(-\frac{1}{T}\right)\right),$$

где $T = -1/\alpha$ — температура. Теорема об устойчивости этого распределения, т. е. о том, что оно получается из «микроканонического» распределения для совокупной системы, включающей термостат, известна под названием теоремы Гиббса.

Следуя принятому в этой главе общему и формальному стилю изложения, сформулируем указанную теорему в абстрактном виде.

Предварительно введем несколько дополнительных понятий. Назовем условное распределение

$$P_n(\xi_1, \ldots, \xi_n|\alpha) \qquad (4.2.2)$$

n-й степенью распределения

$$P_1(\xi_1|\alpha), \qquad (4.2.3)$$

если

$$P_n(\xi_1, \ldots, \xi_n \mid \alpha) = P_1(\xi_1 \mid \alpha), \ldots, P_1(\xi_n \mid \alpha). \qquad (4.2.4)$$

Пусть распределение (4.2.3) является каноническим:

$$\ln P_1(\xi_1 \mid \alpha) = -\Gamma_1(\alpha) + \alpha B_n(\xi_1) - \varphi_1(\xi_1). \tag{4.2.5}$$

Тогда вследствие (4.2.4) для совместного распределения (4.2.2) имеем

$$\ln P_n(\xi_1, \ldots, \xi_n \mid \alpha) = -n\Gamma_1(\alpha) + \alpha B_n(\xi_1, \ldots, \xi_n) - \\ -\varphi_n(\xi_1, \ldots, \xi_n), \tag{4.2.6}$$

причем

$$B_n(\xi_1, \ldots, \xi_n) = \sum_{k=1}^{n} B_1(\xi_k), \quad \varphi_n(\xi_1, \ldots, \xi_n) = \sum_{k=1}^{n} \varphi_1(\xi_k).$$

Оно, очевидно, тоже является каноническим.

Параметров $\alpha = (\alpha_1, \ldots, \alpha_r)$ может быть несколько, тогда под αB нужно понимать скалярное произведение $\Sigma \alpha_i B^i$.

Канонические внутренние параметры $B_1(\xi_k)$, для которых справедливо соотношение $B_n = \sum\limits_k B_1(\xi_k)$, называются *экстенсивными*, а внешние соответствующие им параметры α — *интенсивными*.

Нетрудно доказать, что если n-я степень распределения является канонической, то и само это распределение каноническое. В самом деле, используя (4.2.4) и определение каноничности для (4.2.2), получаем

$$\sum_{k=1}^{n} \ln P_1(\xi_k \mid \alpha) = -\Gamma_n(\alpha) + \alpha B_n(\xi_1, \ldots, \xi_n) - \varphi_n(\xi_1, \ldots, \xi_n).$$

Положим здесь $\xi_1 = \xi_2 = \ldots = \xi_n$. Это даст

$$n \ln P_1(\xi_1 \mid \alpha) = -\Gamma_n(\alpha) + \alpha B_n(\xi_1, \ldots, \xi_1) - \varphi_n(\xi_1, \ldots, \xi_1),$$

т. е. для «малой» системы (4.2.3) действительно выполнено условие каноничности (4.2.1), причем

$$B_1(\xi_1) = \frac{1}{n} B_n(\xi_1, \ldots, \xi_1); \quad \varphi_1(\xi_1) = \frac{1}{n} \varphi_n(\xi_1, \ldots, \xi_1).$$

Получение канонического «малого» распределения из канонического «большого» распределения, конечно, является естественным. Более глубоким является доказываемый ниже факт, что приближенно каноническое «малое» распределение получается также из неканонического «большого» распределения. Каноническая форма распределения является, следовательно, устойчивой в том смысле, что она получается асимптотически из различных «больших» распределений. Этим и объясняется большая роль канонического распределения в теории, в частности, в статистической физике.

Теорема 4.3. *Пусть заданы функции* $B_1(\xi)$, $\varphi_1(\xi)$, *которым соответствует каноническое распределение*

$$\widetilde{P}_1(\xi \mid \alpha) = \exp[-\Gamma(\alpha) + \alpha B_1(\xi) - \varphi_1(\xi)]. \tag{4.2.7}$$

Кроме того, задано распределение «большой» системы

$$P_n(\xi_1, \ldots, \xi_n \mid A) = e^{-\Psi_n(A) - \sum_{k=1}^{n} \varphi_1(\xi_k)} \times$$
$$\times \delta\left(\sum_{k=1}^{n} B_1(\xi_k) - nA\right), \tag{4.2.8}$$

где функции $\Psi_n(A)$ *определяется из условия нормировки,* A *играет роль параметра. Тогда получаемое из него суммированием*

$$P_1(\xi_1 \mid A) = \sum_{\xi_2, \ldots, \xi_n} P_n(\xi_1, \ldots, \xi_n \mid A) \tag{4.2.9}$$

распределение «малой» системы асимптотически переходит в распределение (4.2.7) *при замене параметра* $\alpha = \alpha(A)$. *Именно*

$$\ln P_1(\xi_1 \mid A) = -\psi(A) + \alpha(A) B_1(\xi_1) -$$
$$- \varphi_1(\xi_1) + \frac{1}{n} B_1^2 \chi(A) + O(n^{-2}). \tag{4.2.10}$$

Вид функций $\psi(A)$, $\alpha(A)$, $\chi(A)$ *определяется нижеследующими формулами* (4.2.17). *Предполагается, что функция* (4.2.13) *дифференцируема достаточное число раз и что уравнение* (4.2.15) *имеет корень.*

Распределение типа (4.2.8) в статистической физике называется «*микроканоническим*».

Доказательство. Пользуясь интегральным представлением дельта-функции

$$\delta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\varkappa x} d\varkappa, \tag{4.2.11}$$

запишем «микроканоническое» распределение (4.2.8) в виде

$$P_n(\xi_1, \ldots, \xi_n \mid A) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left\{-\Psi_n(A) - i\varkappa nA + \right.$$
$$\left. + \sum_{k=1}^{n} [i\varkappa B_1(\xi_k) - \varphi_1(\xi_k)]\right\} d\varkappa$$

и подставим это равенство в (4.2.9). В получающемся выражении

$$P_1(\xi_1 \mid A) = \sum_{\zeta_2, \ldots, \zeta_n} \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left\{-\Psi_n(A) - i\varkappa nA + \right.$$
$$\left. + \sum_{k=1}^{n} [i\varkappa B_1(\xi_k) - \varphi_1(\xi_k)]\right\} d\varkappa$$

произведем суммирование по $\zeta_2, \ldots, \zeta_n$, пользуясь формулой (4.1.10a). Это приводит к результату

$$P_1(\xi_1 | A) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\{-\Psi(A) - i\varkappa nA + (n-1)\Gamma(i\varkappa) + $$
$$+ i\varkappa B_1(\xi_1) - \varphi_1(\xi_1)\} d\varkappa \equiv \exp[-\Psi(A) - \varphi_1(\xi_1)] I, \quad (4.2.12)$$

где

$$\Gamma(\alpha) = \ln \sum_{\xi_k} \exp[\alpha B_1(\xi_k) - \varphi_1(\xi_k)]. \quad (4.2.13)$$

Интеграл в (4.2.12) возьмем методом скорейшего спуска, используя то, что n принимает большие значения.

Определим седловую точку $i\varkappa = \alpha_0$ из условия экстремума выражения, стоящего в экспоненте (4.2.12), т. е. из уравнения

$$(n-1)\frac{d\Gamma}{d\alpha}(\alpha_0) = nA - B_1(\xi_1). \quad (4.2.14)$$

Ввиду того, что точка оказывается зависящей от ξ, удобно рассматривать также точку α_1, не зависящую от ξ_1, определяемую уравнением

$$\frac{d\Gamma}{d\alpha}(\alpha_1) = A, \quad (4.2.15)$$

которая при больших n близка к α_0.

Как следует из (4.2.14), (4.2.15)

$$\Gamma'(\alpha_0) - \Gamma'(\alpha_1) = \Gamma''(\alpha_1)\varepsilon + \frac{1}{2}\Gamma'''(\alpha_1)\varepsilon^2 + \ldots =$$
$$= \frac{A - B_1}{n-1} \quad (\varepsilon \equiv \alpha_0 - \alpha_1).$$

Отсюда

$$\varepsilon = \frac{A - B_1}{(n-1)\Gamma''} - \frac{1}{2}\frac{\Gamma'''}{\Gamma''}\varepsilon^2 - \ldots = \frac{A - B_1}{(n-1)\Gamma''} -$$
$$- \frac{1}{2}\frac{\Gamma'''(A - B_1)^2}{(n-1)^2(\Gamma'')^3} + O(n^{-3}). \quad (4.2.16)$$

Поскольку $\Gamma''(\alpha)$ согласно (4.1.15) можно интерпретировать как дисперсию $\sigma^2 = k_{11}$ некоторой случайной величины, то

$$\Gamma''(\alpha_0) > 0 \quad (4.2.16a)$$

и, следовательно, в точке α_0 (а также в α_1) направление скорейшего спуска функции $(n-1)\Gamma(\alpha) - \alpha nA + \alpha B_1$ перпендикулярно действительной оси. В самом деле, если разность $\alpha - \alpha_0 = iy$ мнимая, то

$$(n-1)\Gamma(\alpha) - n\alpha A + \alpha B_1 = (n-1)\Gamma(\alpha_0) - n\alpha_0 A +$$
$$+ \alpha_0 B_1 - \frac{1}{2}(n-1)\Gamma''(\alpha_0) y^2 + O(y^3).$$

Проводя контур интегрирования через точку α_0 в направлении скорейшего спуска, воспользуемся вытекающим из формулы

$$\int \exp\left[-\frac{a}{2}x^2+\frac{b}{6}x^3+\frac{c}{24}x^4+\ \dots\right]dx \approx \tag{4.2.16б}$$

$$\approx \sqrt{2\pi/a}\exp\,[c/8a^2+\dots] \quad \text{при}\ a>0;\ b/a_2^{3/2}\ll 1;\ \ c/a^2\ll 1;\ \dots$$

равенством

$$I \equiv \frac{1}{2\pi}\int \exp\{(n-1)\,\Gamma(i\varkappa)+i\varkappa\,[B_1(\xi_1)-nA]\}\,d\varkappa=$$

$$=[2\pi(n-1)\,\Gamma''(\alpha_0)]^{-1/2}\exp\Big\{(n-1)\,\Gamma(\alpha_0)+$$

$$+\alpha_0\,[B_1(\xi_1)-nA]+\frac{\Gamma^{\mathrm{IV}}(\alpha_0)}{8(n-1)\,\Gamma''(\alpha_0)^2}+O(n^{-2})\Big\}. \tag{4.2.16в}$$

Здесь, используя (4.2.16), значение α_0 нужно выразить через α_1, произведя разложение в ряд по $\varepsilon=\alpha_0-\alpha_1$ и учитывая требуемое число членов. Это приводит к результату

$$I=[2\pi(n-1)\,\Gamma''(\alpha_1)]^{-1/2}\exp\Big\{(n-1)\,\Gamma(\alpha_1)+\alpha_1\,[B_1(\xi_1)-nA]-$$

$$-\frac{\Gamma'''(\alpha_1)\,[A-B_1(\zeta)]}{2n\Gamma''(\alpha_1)^2}-\frac{[B_1(\xi_1)-A]^2}{2n\Gamma''(\alpha_1)}+\frac{\Gamma^{\mathrm{IV}}(\alpha_1)}{8n\Gamma''(\alpha_1)^2}+O(n^{-2})\Big\}.$$

Подставляя это выражение в (4.2.12) и обозначая

$$\psi(A)=\Psi_n(A)-(n-1)\,\Gamma(\alpha_1)+n\alpha_1A+\frac{1}{2}\ln[2\pi(n-1)\,\Gamma''(\alpha_1)]-$$

$$-\frac{\Gamma^{\mathrm{IV}}(\alpha_1)}{8n\Gamma''(\alpha_1)^2}+\frac{A^2}{2n\Gamma''(\alpha_1)}+\frac{A\Gamma'''(\alpha_1)}{2n\Gamma''(\alpha_1)^2}; \tag{4.2.17}$$

$$\alpha(A)=\alpha_1+\frac{A}{n\Gamma''(\alpha_1)}+\frac{\Gamma'''(\alpha_1)}{2n\Gamma''(\alpha_1)^2};$$

$$\chi(A)=-\frac{1}{2\Gamma''(\alpha_1)},$$

получаем (4.2.10). Доказательство закончено.

Первое равенство из (4.2.17) не обязательно принимать во внимание, так как функция $\psi(A)$ однозначно определяется функциями $\alpha(A)$, $\chi(A)$ в силу условия нормировки.

Поскольку ряд членов в (4.2.17) в пределе $n\to\infty$ исчезает, то предельное выражение (4.2.10) имеет вид

$$\ln P_1(\xi_1|A)=\mathrm{const}+\alpha B_1(\xi_1)-\varphi_1(\xi_1)\quad(\Gamma'(\alpha)=A),$$

или $\ln P_1(\xi|A)=-\Gamma(\alpha)+\alpha B_1(\xi_1)-\varphi_1(\xi_1)$, как это следует из соображений нормировки.

Теорема 4.3 может быть обобщена в различных направлениях. Тривиально обобщение на тот случай, когда имеется не один, а несколько (r) параметров $B^1, \dots, B^r$. При этом дельта-функцию в (4.2.8) нужно заменить на произведение дельта-функций, а выражение αB

понимать в смысле скалярного произведения. При другом обобщении дельта-функция в формуле (4.2.8) может быть заменена другими функциями. Для применения теории к рассмотренной в § 3.2 задаче вычисления пропускной способности каналов без помех большую роль играют функции

$$\vartheta_+(x)=\begin{cases}1 & \text{при } x>0\\ 0 & \text{при } x\leqslant 0,\end{cases} \qquad \vartheta_-(x)=1-\vartheta_+(x), \tag{4.2.18}$$

а также функции типа

$$\vartheta_+(x)-\vartheta_+(x-c)=\vartheta_-(x-c)-\vartheta_-(x).$$

В самом деле, введение функции $\vartheta_-(\cdot)$ в распределение

$$P_n(\xi_1, \ldots, \xi_n \mid A)=\vartheta_-\left(\sum_{k=1}^n B_1(\xi_k)-nA\right)e^{-\Psi_n(A)-\sum\limits_{k=1}^n \varphi_1(\xi_k)}$$

эквивалентно введению условия $\sum\limits_{k=1}^n B_1(\xi_k)\leqslant nA$, т. е. условия

(3.2.1) при $c(y)=\sum\limits_{k=1}^n B_1(\xi_k), \quad a=nA.$

Аналогично первое условие (3.2.7) соответствует введению функции

$$\vartheta_-\left(\sum_{k=1}^n B_1(\xi_k)-a_1\right)-\vartheta_-\left(\sum_{k=1}^n B_1(\xi_k)-a_2\right).$$

Желая охватить эти случаи, «микроканоническое» распределение (4.2.8) нужно заменить более общим распределением

$$P_n(\xi_1, \ldots, \xi_n \mid A)=\vartheta\left(\sum_{k=1}^n B_1(\xi_k)-nA\right)e^{-\Psi n(A)-\sum\limits_{k=1}^n \varphi_1(\xi_k)} \tag{4.2.19}$$

Здесь $\vartheta(\cdot)$ — заданная функция, имеющая спектральное представление

$$\vartheta(z)=\int e^{i\varkappa z}\,\theta(i\varkappa)\,d\varkappa. \tag{4.2.20}$$

Такое обобщение потребует весьма незначительных изменений в доказательстве теоремы. Разложение (4.2.11) нужно заменить разложением (4.2.20), после чего в экспоненте в формуле (4.2.12) и других появится дополнительный член $\ln\theta\ (i\varkappa)$. Это приведет к несущественному усложнению окончательных формул.

Результаты теоремы 4.3 допускают обобщение также в другом направлении. Можно не требовать того, чтобы в формулах (4.2.8), (4.2.19) не зависящий от A сомножитель $e^{-\sum\limits_{k=1}^n \varphi_1(\xi_k)}$, пропорциональный условной вероятности $P(\xi_1, \ldots, \xi_k|A, B_n) = P(\xi_1, \ldots, \xi_k \mid B_n)$, распадался на произведение сомножителей, тем более одинаковых. Однако для справедливости асимптотического результата тогда нужно ввести какое-то более слабое требование. Чтобы его сформулировать, введем понятие канонической устойчивости случайной величины. Назовем последовательность случайных величин $\{\zeta_n, n = 1, 2, \ldots\}$, описываемых распределениями вероятностей $P_n(\zeta_n)$ в пространстве Z, не зависящем от n, *канонически устойчивой* относительно $P_n(\zeta_n)$, если значения их характеристических потенциалов

$$\mu_n(\beta) = \ln \sum e^{\beta\zeta_n} P_n(\zeta_n) \tag{4.2.21}$$

при разных β неограниченно возрастают пропорционально друг другу (дружно стремятся к ∞). Последнее означает следующее:

$$\mu_n(\beta) \to \infty, \qquad \frac{\mu_n(\beta)}{\mu_n(\beta')} \to \frac{\mu_0(\beta)}{\mu_0(\beta')} \qquad (\beta, \beta' \in Q)$$

при $n \to \infty$, где $\boldsymbol{\mu}_0(\beta)$ — некоторая функция, не зависящая от n [Q — область значений β, для которых (4.2.21) имеет смысл].

Легко видеть, что случайные величины, равные расширяющимся суммам

$$\zeta_n = \sum_{k=1}^{n} B_1(\xi_k)$$

независимых и одинаково распределенных случайных величин, являются канонически устойчивыми, так как для них $\boldsymbol{\mu}_n(\beta) = n\boldsymbol{\mu}_1(\beta)$. Однако возможны и другие случаи канонически устойчивых семейств случайных величин. Формулируемая ниже теорема является поэтому обобщением теоремы 4.3.

Теорема 4.4. *Пусть имеется последовательность вероятностных распределений $P_n(\zeta_n, \eta_n)$, $n = 1, 2, \ldots$ такая, что последовательность случайных величин ζ_n является канонически устойчивой. При помощи них строится совместное распределение*

$$P(\xi_1, \zeta_n, \eta_n \mid A_n) = $$
$$= \theta\left(B_1(\xi_1) + \zeta_n - A_n\right) e^{-\Psi_n(A) - \varphi_1(\xi_1)} P_n(\zeta_n, \eta_n), \tag{4.2.22}$$

($\theta(\cdot)$ — *не зависящая от n функция со спектром $\theta(i\varkappa)$). Тогда для получаемого из (4.2.22) суммированием по η_n и ζ_n распределения $P(\xi_1|A_n)$ имеем выражение типа (4.2.10). В нем функции $\psi(A_n)$, $\alpha(A_n)$, $\chi(A_n)$ определяются соответствующими формулами наподобие (4.2.17); при $n \to \infty$ функция $\alpha(A_n)$ переходит в функцию, обратную функции*

$$A_n = \mu_n'(\alpha), \tag{4.2.23}$$

а $\psi(A_n)$ *переходит в* $\Gamma(\alpha(A_n))$, *где*

$$\mu_n(\beta)=\ln\sum_{\zeta_n} e^{\beta\zeta_n} P_n(\zeta_n),$$

$$\Gamma(\alpha)=\ln\sum_{\xi_1} e^{\alpha B_1(\xi_1)-\varphi_1(\xi_1)}.$$

Предполагается, что уравнение (4.2.23) *имеет корень, стремящийся при* $n\to\infty$ *к некоторому пределу.*

Доказательство аналогично доказательству теоремы 4.3. Разница лишь в том, что теперь дополнительно присутствует член $\ln\theta(i\varkappa)$ и что выражение $(n-1)\,\Gamma_1(i\varkappa)$ должно быть заменено на $\mu_n(i\varkappa)$. Вместо формулы (4.2.12), после суммирования по η_n и ζ_n, теперь будем иметь

$$P(\xi_1\,|A_n)=e^{-\Psi_n(A)-\varphi_1(\xi_1)}\int\exp\{i\varkappa B_1(\xi_1)-i\varkappa A_n+\mu_n(i\varkappa)+\ln\theta(i\varkappa)\}\,d\varkappa.$$

Седловая точка $i\varkappa=\alpha_0$ определяется из уравнения

$$\mu_n'(\alpha_0)-A_n+B_1(\xi)+\frac{\theta'(\alpha_0)}{\theta(\alpha_0)}=0,$$

корень которого асимптотически близок к корню α уравнения (4.2.23). Именно,

$$\alpha_0-\alpha=(\mu_n'')^{-1}\left[B_1+\frac{\theta'}{\theta}\right]+\ldots=O(\mu^{-1}).$$

Прочие изменения не требуют пояснений.

Приведенные в этом параграфе теоремы характеризуют большую роль канонических распределений подобно тому, как центральная предельная теорема характеризует роль гауссовых распределений. Этим, по существу, объясняется фундаментальное значение канонических распределений в статистической термодинамике.

4.3. Асимптотическая эквивалентность двух видов ограничений

Рассмотрим асимптотические результаты, связанные с содержанием гл. 3. Покажем, что при вычислении максимальной энтропии (пропускной способности каналов без помех) ограничения, наложенные на средние значения, и ограничения, наложенные на точные значения, асимптотически эквивалентны друг другу. Эти результаты тесно связаны с теоремой об устойчивости канонического распределения, доказанной в § 4.2.

Будем вести изложение в более общей форме, чем в § 3.2 и 3.3, используя вспомогательную меру $\nu(dx)$, подобно тому, как это делается в § 3.6.

Пусть задано пространство X с мерой $\nu(dx)$ (не нормированной на единицу). Энтропию будем определять формулой

$$H = -\int \ln \frac{P(dx)}{\nu(dx)} P(dx) \tag{4.3.1}$$

[см. (1.6.13)]. Имеется ограничение

$$B(x) \leqslant A \tag{4.3.2}$$

или, более общe,

$$B(x) \in E, \tag{4.3.3}$$

где E — некоторое (измеримое) множество, $B(x)$ — заданная функция. Энтропию уровня A (или множества E) определим максимизацией

$$\widetilde{H} = \sup_{P \in \widetilde{G}} \left[-\int P(dx) \ln \frac{P(dx)}{\nu(dx)} \right]. \tag{4.3.4}$$

Здесь множество $\widetilde{G}$ перебираемых распределений $P(\cdot)$ характеризуется тем, что вероятность сосредоточена в подпространстве $\widetilde{X} \subset X$, определяемом условиями (4.3.2) или (4.3.3), т. е.

$$P(\widetilde{X}) = 1; \quad P[B(x) \in E] = 1. \tag{4.3.5}$$

Условия (4.3.2), (4.3.3) будут ослаблены, если их заменить аналогичными условиями для математических ожиданий:

$$\mathbf{M}B(x) \leqslant A, \text{ или } \mathbf{M}B(x) \in E, \tag{4.3.6}$$

где символ $\mathbf{M}$ соответствует усреднению с мерой P. Усредняя (4.3.2), (4.3.3) с мерой P из $\widetilde{G}$, мы получим (4.3.6), следовательно, множество G распределений P, определяемое условием (4.3.6), включает в себя множество $\widetilde{G}$, определяемое условием (4.3.5). Поэтому, если ввести энтропию

$$\overline{H} = \sup_{P \in G} \left[-\int P(dx) \ln \frac{P(dx)}{\nu(dx)} \right], \tag{4.3.7}$$

то будем иметь

$$\overline{H} \geqslant \widetilde{H}. \tag{4.3.8}$$

Отыскание энтропии (4.3.7) при условии (4.3.6) есть не что иное. как первая вариационная задача (см. § 3.3, 3.6). Энтропия $\overline{H}$ совпадает с энтропией некоторого канонического распределения

$$P(dx) = \exp[-\Gamma(\alpha) + \alpha B(x)] \, \nu(dx) \tag{4.3.9}$$

(эта формула совпадает с (3.6. 3а) при $B(x) = c(x)$; $\xi = x$; $\alpha = -\beta$).

При этом $\mathbf{M}B(x) = \frac{d\Gamma}{d\alpha} \in E$. Значение $\mathbf{M}B(x)$ совпадает обычно с максимальной или с минимальной точкой интервала E.

Нетрудно убедиться, что энтропия (4.3.4) определяется формулой

$$\tilde{H} = \ln \nu (\tilde{X}), \tag{4.3.10}$$

в то время как для энтропии (4.3.7) из (4.3.9) получаем

$$\bar{H} = \Gamma(\alpha) - \alpha A \quad \left(A = \frac{d\Gamma}{d\alpha}(\alpha) \right) \tag{4.3.11}$$

[см. (3.6.3б)].

Глубокие асимптотические результаты, связанные с первой вариационной задачей, состоят в том, что энтропии (4.3.10) и (4.3.11) в некоторых случаях близки друг другу, и для вычисления одной величины можно находить другую. Обычно вместо (4.3.10) удобно вычислять энтропию (4.3.11) канонического распределения (4.3.9), применяя обычные в статистической термодинамике методы. Указанные результаты, подтверждающие большую роль канонического распределения (4.3.9), родственны результатам, изложенным в предыдущем параграфе как по содержанию, так и по методам доказательства.

Описанную выше «систему» (или канал без помех), к которой относятся $\nu\,(dx)$, $B\,(x)$, A, будем реализовать двояким образом. С одной стороны, будем полагать заданной «малую систему» (канал), которая имеет $\nu_1\,(dx_1)$, $B_1\,(x_1)$, A_1. С другой стороны, пусть имеется «большая система» (канал), для которой

$$\nu_n(dx_1 \ldots dx_n) = \nu_1(dx_1) \ldots \nu_1(dx_n);$$

$$B_n(x_1, \ldots, x_n) = \sum_{k=1}^{n} B_1(x_k); \quad A_n = n A_1.$$

Она является n-й степенью «малой системы». Приведенные ранее формулы (4.3.1)—(4.3.11) можно применять к обеим указанным системам. Определение энтропий $\tilde{H}_n$, $\bar{H}_n$ можно провести как для «малой», так и для «большой системы». Значения $\tilde{H}_1$, $\bar{H}_1$ для «малой системы» существенно различны, но для «большой системы» эти значения $\tilde{H}_n$, $\bar{H}_n$ согласно нижеизложенному относительно близки в асимптотическом смысле:

$$\tilde{H}_n / \bar{H}_n \to 1 \text{ при } n \to \infty. \tag{4.3.12}$$

Поскольку, как легко убедиться, $\bar{H}_n = n\bar{H}_1$, то соотношение (4.3.12) можно записать

$$\frac{1}{n} \tilde{H}_n \to \bar{H}_1. \tag{4.3.13}$$

Эта формула, а также различные ее обобщения и составляют основной результат первой асимптотической теоремы.

Для «большой системы» условия (4.3.3), (4.3.6) берем в виде

$$\sum_{k=1}^{n} B_1(x_k) - A_n \in E_0; \quad \mathbf{M} \sum_{k=1}^{n} B_1(x_k) - A_n \in E_0,$$

где E_0 не зависит от n. Если ввести функцию

$$\vartheta(z) = \begin{cases} 1 \text{ при } z \in E_0; \\ 0 \text{ при } z \text{ вне } E_0, \end{cases}$$

то условие $\Sigma B_1(x_k) - A_n \in E_0$ можно заменить формулой

$$P(dx_1 \ldots dx_n \mid A_n) = \\ = \text{const}\, \vartheta\left(\sum_k B_1(x_k) - A_n\right) P(dx_1 \ldots dx_n).$$

Экстремальное (дающее $\widetilde{H}_n$) распределение имеет вид

$$P(dx_1 \ldots dx_n \mid A_n) = N^{-1} \vartheta\left(\sum_k B_1(x_k) - A_n\right) \nu_1(dx_1) \ldots \nu_1(dx_n), \tag{4.3.14}$$

где N — нормировочная постоянная, весьма просто связанная с энтропией $\widetilde{H}$:

$$\widetilde{H} = \ln N = \ln \int \vartheta\left(\sum_k B_1(x_k) - A_n\right) \nu_1(dx_1) \ldots \nu_1(dx_n). \tag{4.3.15}$$

Очевидна аналогия формулы (4.3.14) с (4.2.19) и (4.2.22). Это говорит о том, что задача вычисления энтропии (4.3.15) родственн- задаче вычисления парциального распределения (4.2.9), рассмотренной в предыдущем параграфе. Как и там, условия точной мультипликатнвиости

$$\nu_n(dx_1 \ldots dx_n) = \nu_1(dx_1) \ldots \nu_1(dx_n)$$

и точной аддитивности

$$B_n(x_1, \ldots, x_n) = \sum_{k=1}^{n} B_1(x_k)$$

не являются обязательными для доказательства основного результата — сходимости (4.3.12). Сформулируем результат сразу в общем виде, используя введенное в § 4.2 понятие канонически устойчивой последовательности случайных величин (при этом под ξ_n будем подразумевать набор $x_1, \ldots, x_n$).

Теорема 4.5. (первая асимптотическая теорема). *Пусть задана последовательность мер $\nu_n(d\xi_n)$, $n = 1, 2, \ldots$ и последовательность случайных величин $B_n(\xi_n)$, канонически устойчивая относительно распределения*

$$P(d\xi_n) = \nu_n(d\xi_n)/\nu_n(\Xi_n) \tag{4.3.16}$$

(мера $\nu_n(\Xi_n)$ всего пространства значений ξ_n предполагается конечной). Тогда энтропию $\widetilde{H}_n$ можно подсчитывать *по асимптотической формуле*

$$\widetilde{H} = -\gamma_n(A_n) - \frac{1}{2}\ln\left[\Gamma''(\alpha_1)\right] + O(1), \qquad (4.3.17)$$

где

$$\gamma_n(A_n) = \alpha_1 A_n - \Gamma_n(\alpha_1) = -\bar{H}_n \quad \left(\Gamma_n'(\alpha_1) = A_n\right),$$
$$\Gamma_n(\alpha) = \ln\int \exp\left[\alpha B_n(\xi_n)\right] \nu_n(d\xi_n) \qquad (4.3.18)$$

— потенциал, входящий в (4.3.11). *Следовательно, при $n\to\infty$ имеет место сходимость* (4.3.12).

Условие $\nu_n(\Xi_n) < \infty$ нужно только для существования вероятностной меры (4.3.16). Если оно не выполняется, то нужно лишь модифицировать определение канонической устойчивости для $B_n(\xi_n)$, а именно, потребовать, чтобы при $n\to\infty$ дружно стремился к бесконечности не характеристический потенциал, а потенциал (4.3.18).

Для доказательства, как и в § 4.3, воспользуемся интегральным представлением (4.2.20). Произведем в соотношении

$$\widetilde{H} = \ln\int \theta\left(B_n(\xi_n) - A_n\right) \nu_n(d\xi_n)$$

интегрирование по ξ_n после подстановки (4.2.20). В силу (4.3.18) будем иметь

$$\widetilde{H} = \ln\int \exp\left\{-i\varkappa A_n + \Gamma_n(i\varkappa) + \ln\theta(i\varkappa)\right\} d\varkappa.$$

Вычисление интеграла можно проводить методом перевала (скорейшего спуска), т. е. при помощи формулы (4.2.16б) с различной степенью точности. Точка перевала $i\varkappa = \alpha_0$ определяется уравнением

$$\Gamma_n'(\alpha_0) = A_n - \theta'(\alpha_0)/\theta(\alpha_0).$$

Применимость метода перевала обеспечивается условием канонической устойчивости.

Для доказательства теоремы 4.4 достаточно такой точности:

$$\widetilde{H} = -\alpha_0 A_n + \Gamma_n(\alpha_0) - \frac{1}{2}\ln\left\{\left[\frac{1}{2\pi}\Gamma_n''(\alpha_0)\right]\left[1 + O(\Gamma^{-1})\right]\right\} =$$
$$= -\alpha_1 A_n + \Gamma_n(\alpha_1) - \frac{1}{2}\ln\Gamma_n''(\alpha_1) + O(1).$$

Здесь α_1 — величина, определяемая из уравнения $\Gamma_n'(\alpha_1) = A_n$, мало отличающаяся от α_0:

$$\alpha_1 - \alpha_0 = (\Gamma_n'')^{-1}\frac{\theta'}{\theta} + \ldots = O(\Gamma^{-1}).$$

Доказательство закончено.

Разумеется, аналогичным образом можно получать и более точные асимптотические результаты.

Как видно из доказательства, условия канонической устойчивости случайных величин B_n достаточно для справедливости теоремы. Однако ниоткуда не следует, что это условие является необходимым, что нельзя указать какого-либо более слабого условия, при котором теорема справедлива. Без сомнения, теорему 4.5 (а также, возможно, и теорему 4.4) можно распространить на более общий случай. Об этом косвенно говорит тот факт, что для примера, изложенного в § 3.1 и 3.4, условие канонической устойчивости не выполняется, а предельный переход (4.3.12) при $L \to \infty$, как мы видели, имеет место.

Наряду с другими асимптотическими теоремами, которые будут рассмотрены в дальнейшем (гл. 7 и 11), теоремы 4.3—4.5 составляют основное содержание теории информации, понимаемой в широком смысле как «термодинамической», т. е. асимптотической теории. Многие важные понятия и соотношения этой теории приобретают свое значение в процессе предельного перехода, связанного с увеличением рассматриваемых систем, и являются в этом смысле асимптотическими.

4.4. Некоторые теоремы, касающиеся характеристического потенциала

1. Характеристический потенциал $\mu(s) = \ln \Theta(s)$ случайных величин $B_i(\xi)$ (или производящая функция семиинвариантов) был ранее определен формулой (4.1.11а). Если задано семейство распределений $p(\xi \mid \alpha) = \exp[-\Gamma(\alpha) + \alpha B(\xi) - \varphi_0(\xi)]$, то $\mu(s)$ выражается через потенциал $\Gamma(\alpha)$ по формуле (4.1.12а). Если же задано не семейство распределений, а просто случайная величина ξ с распределением вероятностей $P(d\xi)$, то можно построить семейство распределений

$$P(d\xi \mid \alpha) = \text{const}\, e^{\alpha B(\xi)} P(d\xi).$$

Вследствие (4.1.11а) постоянная нормировки выражается через $\mu(\alpha)$, так что

$$P(d\xi \mid \alpha) = \exp\left[-\mu(\alpha) + \alpha B(\xi)\right] P(d\xi) \tag{4.4.1}$$

$$(\ln p(\xi) = -\varphi_0(\xi)).$$

Тем самым на основе $P(\xi)$ построено семейство $\{P(d\xi|\alpha), \alpha \in Q\}$. Помимо характеристического потенциала $\mu(s)$ исходного распределения по формуле (4.1.12а), т. е. по формуле

$$\mu(s|\alpha) = \mu(\alpha + s) - \mu(\alpha), \tag{4.4.2}$$

можно находить характеристический потенциал любого распределения (4.4.1) указанного семейства.

2. Рассмотрим сначала простой случай одной величины $B(\xi)$, $r=1$ и докажем несложную, но полезную теорему.

Теорема 4.6. *Пусть область Q значений параметра, входящего в (4.4.1), включает отрезок $s_1 < \alpha < s_2$ ($s_1 < 0$, $s_2 > 0$), причем потенциал $\mu(\alpha)$ дифференцируем на этом отрезке. Тогда функция распределения*

$$F(x) = \mathbf{P}[B(\xi) < x] \tag{4.4.3}$$

удовлетворяет неравенству Чернова

$$F(x) \leqslant \exp\left[-s\mu'(s) + \mu(s)\right], \tag{4.4.4}$$

где

$$\mu'(s) = x, \tag{4.4.5}$$

если последнее уравнение имеет корень $s \in (s_1, 0]$.

Доказательство. Учитывая (4.4.1), запишем (4.4.3) в виде

$$F(x) = \sum_{B(\xi) < x} \exp\left[\mu(\alpha) - \alpha B(\xi)\right] P(\xi \mid \alpha) \tag{4.4.6}$$

вариант дискретной случайной величины ξ). Очевидно, что

$$\sum_{B(\xi) < x} e^{-\alpha B(\xi)} P(\xi \mid \alpha) \leqslant e^{-\alpha x} \sum_{B(\xi) < x} P(\xi \mid \alpha) \text{ при } \alpha \leqslant 0 \tag{4.4.7}$$

и, кроме того,

$$\sum_{B(\xi) < x} P(\xi \mid \alpha) \leqslant 1. \tag{4.4.8}$$

Подставляя (4.4.7), (4.4.8) в (4.4.6), получаем

$$F(x) \leqslant e^{-\alpha x + \mu(\alpha)} \tag{4.4.9}$$

при любом $\alpha \in (s_1, 0]$, в том числе и при $\alpha = s$, где s — корень уравнения (4.4.5), если он существует. Для этого корня, очевидно, (4.4.9) переходит в (4.4.4). Доказательство закончено.

Поскольку значение $\mu'(0)$ есть не что иное, как среднее значение

$$A = \sum_{\xi} B(\xi) P(\xi),$$

то для выполнения условия $s \leqslant 0$, вследствие монотонности (неубывания) функции $\mu'(\alpha)$ ($\mu''(\alpha) \geqslant 0$ по теореме 4.1) необходимо выполнение условия $x < A$. Когда имеет место противоположное неравенство $x > A$, то теорему нужно заменить нижеследующей.

Теорема 4.7. *Если выполнены условия предыдущей теоремы с той лишь разницей, что уравнение (4.4.7) имеет положительный корень $s \in [0, s_2]$, то вместо (4.4.4) выполняется неравенство*

$$1 - F(x) \leqslant e^{-s\mu'(s) + \mu(s)}. \tag{4.4.10}$$

Доказательство этой теоремы аналогично предыдущему, и мы не будем на нем останавливаться.

Сравнивая (4.1.5) с (4.1.14) [см. также (3.5.5.а)], нетрудно видеть, что x играет роль параметра, сопряженного с α относительно потенциала $\mu(\alpha)$. Выражение

$$\gamma(x) = s(x)\,x - \mu(s) = s\mu'(s) - \mu(s), \tag{4.4.11}$$

рассматриваемое как функция от x, есть преобразование Лежандра потенциала $\mu(\alpha)$. При помощи потенциала (4.4.11) формулы (4.4.4), (4.4.10) принимают более короткий вид

$$\mathbf{P}\{[B(\xi) - x]\operatorname{sign}\gamma'(x) \geqslant 0\} \leqslant e^{-\gamma(x)}. \tag{4.4.12}$$

Сформулированные выше теоремы допускают также многомерное обобщение, т. е. обобщение на случай многих (r) переменных $B_1(\xi), \ldots, B_r(\xi)$. В этом случае возможны не две различные формулы (4.4.4), (4.4.10), а 2^r формул в зависимости от знаков корней $s_1, \ldots, s_r$. В правой части всех этих формул, которые можно записать так:

$$\mathbf{P}\left\{[B_1 - x_1]\operatorname{sign}\frac{\partial\gamma}{\partial x_1} \geqslant 0, \ldots, \right.$$
$$\left. [B_r - x_r]\operatorname{sign}\frac{\partial\gamma}{\partial x_r} \geqslant 0\right\} \leqslant e^{-\gamma(x_1, \ldots, x_r)}, \tag{4.4.13}$$

неизменно стоит одно и то же выражение $e^{-\gamma}$, где

$$\gamma(x_1, \ldots, x_r) = \sum x_i s_i - \mu(s)$$
$$\left(x_i = \frac{\partial\mu}{\partial s_i},\ i = 1, \ldots, r\right) \tag{4.4.14}$$

— преобразование Лежандра многомерного характеристического потенциала $\mu(s) = \ln \mathbf{M}e^{sB}$.

Вероятности в левой части соотношений (4.4.12), (4.4.13) при другом возможном обобщении могут браться в соответствии с распределением

$$p(\xi \mid \alpha) = \exp\left[-\Gamma(\alpha) + \alpha B(\xi) - \varphi_0(\xi)\right]$$

(в частности (4.4.1)), а не с $p(\xi)$. Тогда вместо (4.4.13), (4.4.14) будем иметь

$$\mathbf{P}\{[B_1 - x_1]\operatorname{sign} s_1 \geqslant 0, \ldots, [B_r - x_r]\operatorname{sign} s_r \geqslant 0\} \leqslant \exp\left[-\Gamma(\alpha) + \right.$$
$$\left. + \alpha x - \gamma(x)\right], \tag{4.4.15}$$

где

$$\gamma(x) = \sum_i a_i x_i - \Gamma(a) \qquad \left(\frac{\partial\Gamma}{\partial a_j}(a) = x_j\right) \tag{4.4.16}$$

— преобразование Лежандра потенциала $\Gamma(a)$. Чтобы убедиться в справедливости этих формул, нужно принять во внимание (4.1.12а) и формулы (4.4.13), (4.4.14) предыдущего случая.

3. Выведем формулы, выполняющиеся, в отличие от (4.4.3), (4.4.10), (4.4.13), (4.4.15), со знаком равенства, но являющиеся асимптотическими, т. е. справедливыми при больших n или Γ.

Теорема 4.8. *Пусть случайная величина B есть сумма одинаково распределенных независимых случайных величин $B_1(\xi), \ldots, \ldots, B_n(\xi)$* и *соответствующий ей характеристический потенциал $\mu(\alpha) = n\mu_1(\alpha)$ определен и дифференцируем (достаточное число раз) на отрезке $s_1 \leqslant \alpha \leqslant s_2$ $(s_1 < 0,\ s_2 > 0)$. Тогда при значениях x_1, для которых уравнение*

$$\mu'(s) = nx_1 \equiv x \tag{4.4.17}$$

имеет корень $s \in (s_1, s_2)$ и для которых выполняется неравенство

$$\gamma(x) < \min\left[s_1\mu'(s_1) - \mu(s_1),\quad s_2\mu'(s_2) - \mu(s_2)\right], \tag{4.4.18}$$

функция распределения (4.4.3) *случайной величины B определяется асимптотической формулой:*

$$\left.\begin{array}{ll} \text{при } x < \mathbf{M}B & F(x) \\ \text{при } x > \mathbf{M}B & 1 - F(x) \end{array}\right\} = \left[2\pi\mu''(s)s^2\right]^{-1/2} e^{-\gamma(x)}\left[1 + O(n^{-1})\right],$$

$$\gamma(x) = sx - \mu(s), \quad \mu'(s) = x. \tag{4.4.19}$$

Доказательство. Возьмем значения $x < \mathbf{M}B$ и $x' > \mathbf{M}B$ такие, что соответствующие им корни s и s' уравнения (4.4.17) лежат на отрезке (s_1, s_2), и воспользуемся для них известной формулой обращения

$$F(x') - F(x) = \lim_{c\to\infty} \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-itx} - e^{-itx'}}{it} e^{\mu(it)}\, dt \tag{4.4.20}$$

(формула Леви). Здесь $e^{\mu(it)} = e^{n\mu_1(it)}$ — характеристическая функция. Предел в правой части (4.4.20) представим как предел разности двух интегралов

$$\frac{1}{2\pi i}\int_L e^{-zx + \mu(z) - \ln z}\, dz - \frac{1}{2\pi i}\int_L e^{-zx' + \mu(z) - \ln z}\, dz = I - I'. \tag{4.4.21}$$

В качестве контура интегрирования L возьмем контур, идущий из $-ic$ в ic через седловую точку $z = z_0$ первого интеграла, которая определяется уравнением

$$\mu'(z_0) - x - \frac{1}{z_0} = 0 \text{ или } \mu_1'(z_0) - x_1 - \frac{1}{nz_0} = 0$$

$$(x_1 = x/n). \tag{4.4.22}$$

Оно получено приравниванием нулю производной от выражения, стоящего в экспоненте. Обозначим через L' контур, идущий из $-ic$

в ic через седловую точку z_0' второго интеграла, определяемую из уравнения

$$\mu_1'(z_0') - x_1' - \frac{1}{nz_0'} = 0,$$

и постараемся во втором интеграле (4.4.21) заменить контур интегрирования L на L'.

Поскольку $nx_1 < \mathbf{M}B = n\boldsymbol{\mu}_1'(0)$, а $nx_1' > \mathbf{M}B = n\boldsymbol{\mu}_1'(0)$, то указанные две седловые точки при достаточно больших n лежат по разные стороны от начала координат $z = 0$. Поэтому, чтобы изменить указанным образом контур интегрирования L на L', во втором интеграле в (4.4.21) нужно учесть вычет в точке $z = 0$, что дает

$$I' = \frac{1}{2\pi i}\int_L e^{-zx'+\mu(z)}\,\frac{dz}{z} = \frac{1}{2\pi i}\int_{L'} e^{-zx'+\mu(z)}\,\frac{dz}{z} - e^{-\mu(0)}. \qquad (4.4.23)$$

Интегралы в (4.4.21), (4.4.23) будем брать методом скорейшего спуска. Учитывая, что $\boldsymbol{\mu}_1''(z_0) \geqslant 0$ (теорема 4.1), легко видеть, что направление скорейшего спуска перпендикулярно действительной оси:

$$-zx + \mu(z) - \ln z = -z_0 x + \mu(z_0) - \ln z_0 - \frac{1}{2}\mu''(z_0)\,y^2 - \frac{y^2}{2z_0^2} + \frac{1}{6}\mu'''(z_0)(iy)^3 - \frac{1}{3}\left(\frac{iy}{z_0}\right)^3 + y^4\ldots .$$

Используя разложение

$$e^{-nzx_1 + n\mu_1(z) - \ln z} = \frac{1}{z_0}\exp\left\{-z_0 x + \mu(z_0) - \frac{n}{2}\left[\mu_1''(z_0) + \frac{1}{nz_0^2}\right]y^2 + ny^3\ldots\right\} = \frac{1}{z_0}\exp\left\{-z_0 x + \mu(z_0) - \frac{n}{2}\left[\mu_1'' + \frac{1}{nz_0^2}\right]y^2\right\}[1 + ny^3\ldots]$$

согласно (4.2.16б) получаем

$$I \equiv \frac{1}{2\pi i}\int_L e^{-nzx_1 + n\mu_1(z)}\,\frac{dz}{z} = \frac{1}{z_0}\,e^{-z_0 x + \mu(z_0)}\left\{2\pi n\left[\mu_1''(z_0) + \frac{1}{nz_0^2}\right]\right\}^{-1/2}\left[1 + O(n^{-1})\right]. \qquad (4.4.24)$$

(Наибольший член остатка $O(n^{-1})$ дается четвертой производной Γ^{IV} и имеет вид $\frac{1}{8n}\,\boldsymbol{\mu}_1^{IV}(\boldsymbol{\mu}_1'')^{-2}$.) Сравнивая (4.4.22) с уравнением

$\mu_1'(s) - x_1 = 0$, эквивалентным уравнению (4.4.17), имеем

$$\mu_1''(s)(z_0 - s) + (z-s)^2 \ldots = \frac{1}{nz_0}; \quad z_0 - s = \frac{1}{nz_0\,\mu_1''(s)} + O(n^{-2}). \tag{4.4.25}$$

Для вывода требуемого соотношения (4.4.19) достаточно взять формулу, вытекающую из (4.4.24) и (4.4.25)

$$I = \frac{1}{s} e^{-sx+\mu(s)} \left\{2\pi n \mu_1''(s)\right\}^{-1/2} \left[1 + O(n^{-1})\right], \tag{4.4.26}$$

где $sx - \mu(s) = \gamma(x)$, $\mu'(s) = x$ $(s < 0)$.

Аналогично берется второй интеграл (4.4.23), что дает

$$I' = \frac{1}{s'} \left\{2\pi\mu''(s')\right\}^{-1/2} e^{-\gamma(x')} \left[1 + O(n^{-1})\right] - 1 \quad (s' > 0). \tag{4.4.27}$$

Подставляя (4.4.26) и (4.4.27) в (4.4.21), в силу (4.4.20) имеем

$$F(x') - F(x) = 1 - \left[2\pi s'^2 \mu''(s')\right]^{-1/2} e^{-\gamma(x')} \left[1 + O(n^{-1})\right] -$$
$$- \left[2\pi s^2 \mu''(s)\right]^{-1/2} e^{-\gamma(x)} \left[1 + O(n^{-1})\right] \tag{4.4.28}$$

(здесь учтено, что $s < 0$; корни $[\ldots]^{-1/2}$ берутся положительные).

Это равенство определяет $F(x)$ и $F(x')$ с точностью до некоторой аддитивной постоянной:

$$F(x) = \left[2\pi\,\mu''(s)\,s^2\right]^{-1/2} e^{-\gamma(x)} \left[1 + O(n^{-1})\right] + K(n);$$
$$F(x') = 1 - \left[2\pi\,\mu''(s')\,s'^2\right]^{-1/2} e^{-\gamma(x')} \left[1 + O(n^{-1})\right] + K(n).$$

Чтобы оценить постоянную $K(n)$, возьмем точку s_* из отрезка (s_1, s) и s^* из отрезка (s', s_2) и соответствующие им значения x_*, x^*. Учитывая неравенство $F(x_*) \leqslant 1 - F(x^*) + F(x_*)$, и подставляя (4.4.28) в его правую часть, будем иметь

$$F(x_*) \leqslant \left[2\pi\,\mu''(s_*)\,s_*^2\right]^{-1/2} e^{-\gamma(x_*)} \left[1 + O(n^{-1})\right] +$$
$$+ \left[2\pi\,\mu''(s^*)\,s^{*2}\right]^{-1/2} e^{-\gamma(x^*)} \left[1 + O(n^{-1})\right].$$

Следовательно,

$$F(x_*) = O\left(e^{-n\gamma_1(x_{*1})}\right) + O\left(e^{-n\gamma_1(x_1^*)}\right)$$
$$\left(\gamma_1(x_1) = \frac{1}{n}\gamma(x_1) = x_1\mu_1' - \mu_1; \qquad x_{*1} = x_*/n, \quad x_1^* = x^*/n\right).$$

Подставляя это выражение в равенство

$$1 - F(x') = -F(x_*) + \left[2\pi\mu''(s')\,s'^2\right]^{-1/2} e^{-\gamma(x')} \left[1 + O(n^{-1})\right] +$$
$$+ \left[2\pi\,\mu''(s_*)\,s_*^2\right]^{-1/2} e^{-\gamma(x_*)} \left[1 + O(n^{-1})\right],$$

вытекающее из (4.4.28) при $x = x_*$, находим

$$1 - F(x') = 2\pi \left[\mu''(s') s'^2\right]^{-1/2} e^{-\gamma(x')}\left[1 + O(n^{-1})\right] + \\ + O\left(e^{-n\gamma_1(x_{*1})}\right) + O\left(e^{-n\gamma_1\left(x_1^*\right)}\right). \quad (4.4.29)$$

Из (4.4.28), (4.4.29) нетрудно получить

$$F(x) = 2\pi \left[\mu''(s) s^2\right]^{-1/2} e^{-\gamma(x)} \left[1 + O(n^{-1})\right] + \\ + O\left(e^{-n\gamma_1(x_{*1})}\right) + O\left(e^{-n\gamma_1\left(x_1^*\right)}\right). \quad (4.4.30)$$

Если x_{*1}, s_* и x_1^*, s^* не зависят от n и выбраны таким образом, что

$$\gamma_1(x_{*1}) > \gamma_1(x_1); \quad \gamma_1(x_1^*) > \gamma_1(x_1), \quad (4.4.31)$$

то члены $O\left(e^{-n\gamma_1(x_{*1})}\right)$, $O\left(e^{-n\gamma_1(x_1)}\right)$ в (4.4.30) стремятся к нулю при $n \to \infty$ быстрее, чем $e^{-n\gamma_1(x_1)}\, O(n^{-1})$, и мы из (4.4.30) получаем первое равенство (4.4.19). Второе равенство вытекает из (4.4.29) при выполнении неравенств

$$\gamma_1(x_{*1}) > \gamma_1(x_1'); \quad \gamma_1(x_1^*) > \gamma_1(x_1'). \quad (4.4.32)$$

Поскольку $\gamma(x)$ является монотонной на отрезках $\boldsymbol{\mu}'(s_1) \leqslant x \leqslant \boldsymbol{\mu}'(0)$ и $\boldsymbol{\mu}'(0) \leqslant x \leqslant \boldsymbol{\mu}'(s_2)$, точки x_{*1} и x_1^*, для которых выполнялись бы неравенства (4.4.31), (4.4.32), заведомо можно подобрать, если выполнено условие (4.4.18). Доказательство закончено.

Как видно из приведенного доказательства, отмеченное в условии теоремы 4.8 требование, чтобы случайная величина B равнялась сумме одинаково распределенных независимых величин, является необязательным. Достаточно дружного (пропорционального) возрастания результирующего потенциала $\boldsymbol{\mu}(s)$ для того, чтобы члены, подобные члену $\boldsymbol{\mu}^{IV}/(\boldsymbol{\mu}'')^2$ в правой части (4.4.24), были малы. Поэтому формула (4.4.19) справедлива и в более общем случае, если заменить $O(n^{-1})$ на $O(\boldsymbol{\mu}^{-1})$ и понимать эту оценку в только что отмеченном смысле.

Если формулу (4.4.19) применить не к распределению $p(B)$, а к распределению $p(B|\alpha)$, зависящему от параметра, то вследствие (4.1.12а) будем иметь $\boldsymbol{\mu}(s) = \Gamma(\alpha + s) - \Gamma(\alpha)$; $\boldsymbol{\mu}''(s) = \Gamma''(\alpha + s)$; $s\,\boldsymbol{\mu}'(s) - \boldsymbol{\mu}(s) = s\,\Gamma'(\alpha + s) - \Gamma(\alpha + s) + \Gamma(\alpha) = \gamma(x) - \alpha x + \Gamma(\alpha)$, и формула (4.4.19) примет вид

$$\left.\begin{matrix} \text{при } x < \Gamma'(\alpha) & F(x) \\ \text{при } x > \Gamma'(\alpha) & 1 - F(x) \end{matrix}\right\} = 2\pi \left[\Gamma''(\alpha + s) s^2\right]^{1/2} \times \\ \times e^{-\Gamma(\alpha) + \alpha x - \gamma(x)} \left[1 + O(\Gamma^{-1})\right], \quad (4.4.33)$$

где $\gamma(x) = ax - \Gamma(a)$ $(\Gamma'(a) = x)$ — преобразование Лежандра.

Не представляет труда также обобщить теорему 4.8 на случай многих случайных величин $B_1, \ldots, B_r$. Соответствующее этому обобщение формулы (4.4.19), если применить ту же форму записи, что и в (4.4.13), будет

$$\mathbf{P}\left\{[B_1 - x_1]\operatorname{sign} s_1 \geqslant 0, \ldots, [B_r - x_r]\operatorname{sign} s_r \geqslant 0\right\} =$$
$$= \det{}^{-1/2}\left\| 2\pi \frac{\partial^2 \mu(s)}{\partial s_i\, \partial s_j} s_i s_j \right\| e^{-\gamma(x)}\left[1 + O(\Gamma^{-1})\right], \qquad (4.4.34)$$

где $\gamma(x)$ определена посредством (4.4.14).

Наконец, обобщение последней формулы на случай параметрического распределения [а формулы (4.4.33) на многомерный случай] запишется следующим образом:

$$\mathbf{P}\left\{(B_1 - x_1)\operatorname{sign} s_1 \geqslant 0, \ldots, (B_r - x_r)\operatorname{sign} s_r \geqslant 0 \,|\, \alpha\right\} =$$
$$= (2\pi)^{-r/2}\,|s_1 \ldots s_r|^{-1} \det{}^{-1/2}\left\| \frac{\partial^2 \Gamma(\alpha + s)}{\partial s_i\, \partial s_j} \right\| \times$$
$$\times e^{-\Gamma(\alpha) + \alpha x - \gamma(x)}\left[1 + O(\Gamma^{-1})\right]. \qquad (4.4.35)$$

Приведенные выше результаты свидетельствуют о важной роли потенциалов и их отображений по Лежандру. Примененный метод доказательства объединяет теорему 4.8 с теоремами 4.3 — 4.5.

Глава 5

ВЫЧИСЛЕНИЕ ЭНТРОПИИ ДЛЯ ЧАСТНЫХ СЛУЧАЕВ. ЭНТРОПИЯ СЛУЧАЙНЫХ ПРОЦЕССОВ

В настоящей главе излагаются методы подсчета энтропии многих случайных величин или случайного процесса в дискретном и непрерывном времени.

Особый интерес с принципиальной, а также прикладной точки зрения представляют стационарные случайные процессы и их теоретико-информационные характеристики, в частности, энтропия. Такие процессы являются относительно простыми объектами, особенно дискретный процесс, т. е. стационарный процесс с дискретными состояниями, протекающий в дискретном времени. Поэтому на его примере удобно иллюстрировать основные положения, и мы с него начнем.

Основное внимание будет уделено определению такой важной характеристики стационарного процесса, как удельная энтропия, т. е. энтропия, отнесенная к единице времени, к одному такту. Кроме того вводится энтропия Г конца отрезка. Она вместе с удельной энтропией H_1 определяет энтропию длинного отрезка длины T по приближенной формуле

$$H_T \approx H_1 T + 2\,\Gamma,$$

которая тем точнее, чем больше T. Обе постоянные H_1, Г вычисляются для дискретного марковского процесса.

Обобщенное определение энтропии, данное в § 1.6, позволяет применять это понятие как к непрерывно-значным случайным величинам, так и к тому случаю, когда число этих случайных величин континуально, т. е. к случайному процессу с непрерывным параметром (временем).

Ниже показано, что на случай непрерывного пространства значений и на случай непрерывного времени могут быть распространены многие результаты, относящиеся к дискретному процессу. Так, для стационарных процессов в непрерывном времени может быть введена удельная энтропия, отнесенная уже не к одному такту, а к единице времени, и энтропия конца отрезка. Энтропия случайного процесса на отрезке приближенно представляется в виде двух членов по аналогии с приведенной выше формулой.

Для нестационарных процессов с непрерывным временем, вместо постоянной удельной энтропии, нужно рассматривать плотность энтропии, которая, вообще говоря, непостоянна во времени.

Энтропия и ее плотность подсчитываются для различных важных случаев процессов с непрерывным временем: гауссовых процессов и диффузионных марковских процессов.

Проводимый здесь подсчет энтропии случайных процессов дает возможность подсчитывать для случайных процессов шенноновское количество информации, о котором будет сказано в гл. 6.

5.1. Энтропия отрезка стационарного дискретного процесса и удельная энтропия

Будем предполагать, что имеется последовательность случайных величин $\ldots, \xi_{k-1}, \xi_k, \xi_{k+1}, \ldots$. Индекс (параметр) k можно трактовать как дискретное время t, принимающее целочисленные значения $\ldots, k-1, k, k+1, \ldots$. Число различных значений индекса может быть неограниченно в обе стороны: $-\infty < k < \infty$; неограниченно в одну сторону, например: $0 < k < \infty$, или конечно: $1 \leqslant k \leqslant N$. Указанные значения образуют область K определения параметра. Каждая случайная величина ξ_k пусть принимает одно из конечного или счетного числа значений, например, $\xi_k = (A_1, A_2, \ldots, A_m)$ (для дальнейшего существенна не столько конечность m, сколько конечность энтропий H_{ξ_k}). Описанный процесс — дискретную случайную величину как функцию дискретного параметра k, будем называть *дискретным процессом*.

Дискретный процесс является *стационарным*, если все законы распределения $P(\xi_{k_1}, \ldots, \xi_{k_r})$ (произвольной кратности r) не меняются при сдвиге:

$$\mathbf{P}[\xi_{k_1} = x_1, \ldots, \xi_{k_r} = x_r] = \mathbf{P}[\xi_{k_1+a} = x_1, \ldots, \xi_{k_r+a} = x_r] \qquad (5.1.1)$$

(a — любое целое).

Величина сдига a предполагается такой, что значения $k_1 + a, \ldots, k_r + a$ не выходят из области K определения параметра k. В дальнейшем мы это не будем оговаривать, предполагая, скажем, что область определения параметра не ограничена в обе стороны.

Рассмотрим различные условные энтропии одной из случайных величин ξ_k дискретного стационарного случайного процесса. Ее безусловная энтропия $H_{\xi_k} = -\sum\limits_{\xi_k} P(\xi_k) \ln P(\xi_k)$ вследствие свойства (5.1.1) не зависит от выбранного значения параметра k. Аналогично согласно (5.1.1) условная энтропия $H_{\xi_k \mid \xi_{k-1}}$ не зависит от k. Применяя теорему 1.6, имеем неравенство

$$H_{\xi_k \mid \xi_{k-1}} \leqslant H_{\xi_k}$$

или, учитывая стационарность,

$$H_{\xi_2|\xi_1}=H_{\xi_3|\xi_2}=\ldots=H_{\xi_k|\xi_{k-1}}\leqslant H_{\xi_1}=H_{\xi_2}=\ldots=H_{\xi_k}.$$

Если ввести условную энтропию $H_{\xi_k|\xi_{k-1}\xi_{k-2}}$, то применение теоремы 1.6а при $\xi=\xi_k$, $\eta=\xi_{k-1}$, $\zeta=\xi_{k-2}$ даст неравенство

$$H_{\xi_k|\xi_{k-1},\,\xi_{k-2}}\leqslant H_{\xi_k|\xi_{k-1}}.$$

Согласно условию стационарности (5.1.1) эта энтропия не зависит от k.

Аналогичным образом, увеличивая число случайных величин в условии, мы и дальше будем иметь монотонное изменение (невозрастание) условной энтропии:

$$\begin{aligned}H_{\xi_k}&\geqslant H_{\xi_k|\xi_{k-1}}\geqslant H_{\xi_k|\xi_{k-1},\,\xi_{k-2}}\geqslant\ldots\geqslant\\&\geqslant H_{\xi_k|\xi_{k-1},\ldots,\,\xi_{k-l}}\geqslant\ldots\geqslant 0.\end{aligned}\tag{5.1.2}$$

Кроме того, все условные энтропии неотрицательны, т. е. ограничены снизу. Отсюда вытекает существование неотрицательного предела

$$H_1=\lim_{l\to\infty}H_{\xi_k|\xi_{k-1},\ldots,\,\xi_{k-l}},\tag{5.1.3}$$

который мы будем обозначать также h, чтобы избежать увеличения числа индексов.

Этот предел мы определяем как *удельную энтропию*, рассчитанную на один элемент последовательности $\{\xi_k\}$. Основанием для такого наименования является следующая теорема.

Теорема 5.1. *Если $\{\xi_k\}$ — стационарный дискретный процесс, такой, что $H_{\xi_k}<\infty$, то предел*

$$\lim_{l\to\infty}H_{\xi_1\ldots\xi_l}\,/\,l$$

существует и равен пределу (5.1.3).

Доказательство. Рассмотрим энтропию $H_{\xi_1\ldots\xi_{m+n}}$ и представим ее в виде

$$\begin{aligned}H_{\xi_{m+n}\ldots\xi_1}&=H_{\xi_m\ldots\xi_1}+H_{\xi_{m+1}|\xi_m\ldots\xi_1}+H_{\xi_{m+2}|\xi_{m+1}\ldots\xi_1}+\\&+\ldots+H_{\xi_{m+n}|\xi_{m+n-1}\ldots\xi_1}.\end{aligned}\tag{5.1.4}$$

Поскольку

$$H_{\xi_{m+i}|\xi_{m+i-1}\ldots\xi_1}=H_1+o_{m+i}(1)=H_1+o_m(1)\qquad(i\geqslant 1)$$

согласно (5.1.3) (здесь $o_j(1)\to 0$ при $j\to\infty$), то из (5.1.4) имеем (после деления на $m+n$)

$$\frac{H_{\xi_{m+n}\ldots\xi_1}}{m+n}=\frac{H_{\xi_m\ldots\xi_1}}{m+n}+\frac{n}{m+n}H_1+\frac{n}{m+n}o_m(1).\tag{5.1.5}$$

Пусть m и n стремятся к бесконечности так, что $n/m \to \infty$; тогда $n/(m+n)$ будет стремиться к 1, а отношение $H_{\xi_m \ldots \xi_1}/(m+n)$, которое можно оценить

$$\frac{1}{m+n} H_{\xi_m \ldots \xi_1} \leqslant \frac{m}{m+n} H_{\xi_1},$$

будет, очевидно, стремиться к 0.

Поэтому из равенства (5.1.5) получаем утверждение теоремы. Доказательство закончено.

Нетрудно доказать также, что с ростом l отношение $H_{\xi_1 \ldots \xi_l}/l$ меняется монотонно, т. е. не возрастает. Для этого образуем разность

$$\delta = \frac{1}{l} H_{\xi_1 \ldots \xi_l} - \frac{1}{l+1} H_{\xi_1 \ldots \xi_{l+1}} = \frac{1}{l} H_{\xi_1 \ldots \xi_l} - \frac{1}{l+1}\left[H_{\xi_1 \ldots \xi_l} + H_{\xi_{l+1} \mid \xi_1 \ldots \xi_l}\right] = \frac{1}{l(l+1)} H_{\xi_1 \ldots \xi_l} - \frac{1}{l+1} H_{\xi_{l+1} \mid \xi_1 \ldots \xi_l}$$

и представим ее в виде

$$\delta = \frac{1}{l(l+1)}\left[H_{\xi_1 \ldots \xi_l} - l H_{\xi_{l+1} \mid \xi_1 \ldots \xi_l}\right] = \frac{1}{l(l+1)} \sum_{i=1}^{l} \left[H_{\xi_i \mid \xi_1, \ldots, \xi_{i-1}} - H_{\xi_i \mid \xi_{i-l} \ldots \xi_{i-1}}\right]. \qquad (5.1.6)$$

Вследствие неравенств (5.1.2) слагаемые в правой части (5.1.6) являются неотрицательными, откуда вытекает неотрицательность разности δ.

Согласно теореме 5.1 имеет место равенство

$$H_{\xi_1 \ldots \xi_n} = nH_1 + n o_n(1). \qquad (5.1.7)$$

Образуем комбинацию

$$H_{\xi_1 \ldots \xi_m} + H_{\xi_1 \ldots \xi_n} - H_{\xi_1 \ldots \xi_{m+n}} = H_{\xi_1 \ldots \xi_m} + H_{\xi_{m+1} \ldots \xi_{m+n}} - H_{\xi_1 \ldots \xi_{m+n}}. \qquad (5.1.8)$$

Легко доказать, что она является монотонной функцией от m при фиксированном n (или наоборот). В самом деле, ее можно записать в форме

$$H_{\xi_{m+1} \ldots \xi_{m+n}} - H_{\xi_{m+1} \ldots \xi_{m+n} \mid \xi_1 \ldots \xi_m}. \qquad (5.1.9)$$

Поскольку условные энтропии не возрастают с ростом m согласно (5.1.2), то выражение (5.1.9) не убывает при увеличении m. То же самое можно сказать и о зависимости от n, так как в (5.1.8) m и n

входят симметрично. Из (5.1.9) очевидна также неотрицательность комбинации (5.1.8) (в силу теоремы 1.6 при $\xi=(\xi_{m+1}, \dots, \xi_{m+n})$).

Рассмотрим предел

$$2\Gamma=\lim_{n\to\infty}\lim_{m\to\infty}\left[H_{\xi_1 \dots \xi_m}+H_{\xi_1 \dots \xi_n}-H_{\xi_1 \dots \xi_{m+n}}\right]= \\ =\lim_{m\to\infty}\lim_{n\to\infty}\left[H_{\xi_1 \dots \xi_m}+H_{\xi_1 \dots \xi_n}-H_{\xi_1 \dots \xi_{m+n}}\right]. \qquad (5.1.10)$$

Менять порядок пределов здесь можно вследствие уже отмеченной симметрии относительно перестановки m и n. В силу указанной монотонности этот предел (конечный или бесконечный) всегда существует. Переходя от формы записи (5.1.8) к форме (5.1.9), с использованием иерархического соотношения

$$H_{\xi_{m+1}, \dots \xi_{m+n} \mid \xi_1, \dots, \xi_m}=\sum_{i=1}^{n} H_{\xi_{m+i} \mid \xi_1, \dots, \xi_{m+i-1}}$$

(типа (1.3.4)) равенство (5.1.10) можно записать, выполнив предельный переход $m \to \infty$:

$$2\Gamma=\lim_{n\to\infty}\left[H_{\xi_1 \dots \xi_n}-nH_1\right], \qquad (5.1.11)$$

так как $H_{\xi_{m+i} \mid \xi_1 \dots \xi_{m+i-1}} \to H_1$ при $m \to \infty$.

Если энтропию $H_{\xi_1 \dots \xi_n}$ здесь представить в виде иерархической суммы (1.3.4), то формула (5.1.11) примет вид

$$2\Gamma=\sum_{j=0}^{\infty}\left[H_{\xi_{j+1} \mid \xi_1 \dots \xi_j}-H_1\right]. \qquad (5.1.12)$$

Согласно (5.1.11) имеем

$$H_{\xi_1 \dots \xi_n}=nH_1+2\Gamma+o_n(1). \qquad (5.1.13)$$

Эта формула уточняет (5.1.7).

Поскольку на один элемент последовательности $\{\xi_k\}$ приходится в среднем энтропия H_1, то на n таких элементов приходится nH_1. Согласно (5.1.13) энтропия 2Г отличается от nH_1 при больших n на величину 2Г, которую можно трактовать как энтропию, приходящуюся на концы отрезка. Таким образом, *энтропия каждого конца отрезка стационарного процесса в пределе равна* Г.

Если один длинный отрезок стационарного процесса разбить на два отрезка и ликвидировать статистические связи (корреляции) между процессами на этих отрезках, то энтропия возрастет приблизительно на 2Г, так как появятся два новых конца.

Применяя формулу (5.1.13) для трех отрезков длиной m, n и $m+n$ и образуя комбинацию (5.1.8), теперь имеем

$$H_{\xi_1 \dots \xi_m}+H_{\xi_1 \dots \xi_n}-H_{\xi_1 \dots \xi_{m+n}}=2\Gamma+o_m(1)+o_n(1)+o_{m+n}(1).$$

Это согласуется с (5.1.10) и подтверждает сказанное об увеличении энтропии на 2Г.

5.2. Энтропия марковской цепи

1. Пусть дискретный (необязательно стационарный) процесс $\{\xi_k\}$ является марковским. Это значит, что совместные законы распределения последовательно идущих случайных величин распадаются на произведение

$$P(\xi_k, \xi_{k+1}, \ldots, \xi_{k+l}) = \\ = P(\xi_k)\, \pi_k(\xi_k, \xi_{k+1}) \ldots \pi_{k+l-1}(\xi_{k+l-1}, \xi_{k+l}) \qquad (5.2.1)$$

функций $\pi_j(\xi, \xi') = P(\xi' \mid \xi)$, которые называются вероятностями перехода. В (5.2.1) вероятности $P(\xi_k)$ соответствуют однократному распределению случайной величины ξ_k.

Вероятность перехода $\pi_j(\xi_j, \xi_{j+1})$, равняясь условным вероятностям $P(\xi_{j+1} \mid \xi_j)$, обладает свойством неотрицательности и нормировки

$$\sum_{\xi'} \pi_j(\xi, \xi') = 1. \qquad (5.2.2)$$

Дискретный марковский процесс называют также марковской цепью.

Переходя от вероятностей (5.2.1) к условным вероятностям, легко получить, что в случае марковского процесса

$$P(\xi_{k+m+1} \mid \xi_k, \ldots, \xi_{k+m}) = \pi_{k+m}(\xi_{k+m}, \xi_{k+m+1}) \quad (\text{любое } m \geqslant 1)$$

и, следовательно,

$$P(\xi_{k+m+1} \mid \xi_k, \ldots, \xi_{k+m}) = P(\xi_{k+m+1} \mid \xi_{k+m}).$$

Поэтому

$$H(\xi_{k+m+1} \mid \xi_k, \ldots, \xi_{k+m}) = -\ln \pi_{k+m}(\xi_{k+m}, \xi_{k+m+1}) = \\ = H(\xi_{k+m+1} \mid \xi_{k+m}). \qquad (5.2.3)$$

Благодаря (5.2.3) иерархическая формула (1.3.6) принимает вид

$$H(\xi_1, \ldots, \xi_n) = H(\xi_1) + H(\xi_2 \mid \xi_1) + H(\xi_3 \mid \xi_2) + \ldots + H(\xi_n \mid \xi_{n-1}). \qquad (5.2.4)$$

Аналогичным образом для средних энтропий имеем

$$H_{\xi_{k+m+1} \mid \xi_k \ldots \xi_{k+m}} = H_{\xi_{k+m+1} \mid \xi_{k+m}}, \qquad (5.2.5)$$

$$H_{\xi_1 \ldots \xi_n} = H_{\xi_1} + H_{\xi_2 \mid \xi_1} + H_{\xi_3 \mid \xi_2} + \ldots + H_{\xi_n \mid \xi_{n-1}}. \qquad (5.2.6)$$

Дискретный марковский процесс является стационарным, если вероятности перехода $\pi_k(\xi, \xi')$ не зависят от значения параметра k и если все однократные распределения $P(\xi_k)$ одинаковы (равны $P_{\text{ст}}(\xi)$). Образуя многократные распределения по формуле (5.2.1), легко доказать, что они также удовлетворяют условию стационарности (5.1.1).

Стационарность однократного распределения $P_{\text{ст}}(\xi)$ означает, что выполняется уравнение

$$\sum_{\xi} P_{\text{ст}}(\xi)\,\pi(\xi, \xi') = P_{\text{ст}}(\xi'). \tag{5.2.7}$$

Последнее легко вывести, записывая согласно (5.2.1.) двукратное распределение $P(\xi_k, \xi_{k+1}) = P_{\text{ст}}(\xi_k)\,\pi(\xi_k, \xi_{k+1})$ и суммируя его по $\xi_k = \xi$, что дает $P(\xi_{k+1})$.

Учитывая (5.2.3), легко видеть, что для стационарного марковского процесса удельная энтропия (5.1.3) совпадает с энтропией, соответствующей вероятностям перехода $\pi(\xi, \xi')$ при стационарном распределении вероятностей:

$$H_1 = -\sum_{\xi} P_{\text{ст}}(\xi) \sum_{\xi'} \pi(\xi, \xi') \ln \pi(\xi, \xi'). \tag{5.2.8}$$

В самом деле, усредняя (5.2.3) со стационарными вероятностями $P_{\text{ст}}(\xi)\,\pi(\xi, \xi')$, убеждаемся, что всем значениям $m \geqslant 1$ соответствует одна и та же условная энтропия $H_{\xi_{k+m+1} \mid \xi_{k+1} \ldots \xi_{k+m}}$, равная (5.2.8).

Усредним далее формулу (5.2.4) со стационарными вероятностями. Это даст равенство

$$H_{\xi_1 \ldots \xi_n} = H_{\xi_1} + (n-1) H_1. \tag{5.2.9}$$

Сравнивая его с (5.1.13), видим, что в стационарном марковском случае

$$2\Gamma = H_{\xi_1} - H_1 = \sum_{\xi, \xi'} P_{\text{ст}}(\xi)\,\pi(\xi, \xi') \ln \frac{\pi(\xi, \xi')}{P_{\text{ст}}(\xi')} \tag{5.2.10}$$

и $o_n(1) = 0$, т. е. формула

$$H_{\xi_1 \ldots \xi_n} = nH_1 + 2\Gamma \tag{5.2.11}$$

выполняется точно. К результату (5.2.10) можно прийти также при помощи формулы (5.1.12). В самом деле, поскольку, как отмечалось, $H_{\xi_{j+1} \mid \xi_1 \ldots \xi_j} = H_1$, то в сумме в правой части (5.1.14) остается лишь единственный неиулевой член: $2\Gamma = H_{\xi_1} - H_1$, что совпадает с (5.2.10).

2. Итак, для вычисления энтропии, если задана матрица вероятностей перехода

$$\boldsymbol{\pi} = \| \pi(\xi, \xi') \|, \tag{5.2.12}$$

следует найти стационарное распределение и воспользоваться формулами (5.2.8), (5.2.9). Уравнение (5.2.7) вполне однозначно определяет стационарное распределение $P_{\text{ст}}(\xi)$, если марковский процесс является эргодическим, т. е. если собственное значение $\lambda = 1$ является невырожденным собственным значением матрицы

(5.2.12). Согласно теореме о разложении определителей, из уравнения $\det(\boldsymbol{\pi}-1)=0$ будет вытекать (5.2 7), если положить

$$P_{\text{ст}}(\xi)=A_{\xi\xi'}\Big/\sum_{\xi''}A_{\xi''\xi'}\,, \tag{5.2.13}$$

где $A_{\xi\xi'}$ — алгебраические дополнения элемента $a_{\xi\xi'}$ матрицы $\|a_{\xi\xi'}\|=\|\pi(\xi,\xi')-\delta_{\xi\xi'}\|\equiv\boldsymbol{\pi}-1$. Значение ξ' здесь любое, так что достаточно вычислить алгебраические дополнения одного столбца указанной матрицы.

Если марковский процесс неэргодический, то стационарное распределение не определяется полностью матрицей перехода, а зависит еще от начального (или какого-либо другого однократного) распределения. В этом случае в формуле (5.2.13) алгебраические дополнения равны нулю и, следовательно, имеет место неопределенность типа 0/0. Стационарное распределение при этом может быть выражено через миноры меньшего порядка и начальное распределение.

Как известно [Дуб, 1], состояния дискретного марковского процесса можно расположить в таком порядке, чтобы переходная матрица имела следующий «ящичный» вид:

$$\boldsymbol{\pi}=\begin{pmatrix}\pi^{00} & \pi^{01} & \pi^{02} & \pi^{03}\ldots\\ 0 & \pi^{11} & 0 & 0 & \ldots\\ 0 & 0 & \pi^{22} & 0 & \ldots\\ & & & & \end{pmatrix}\equiv\Pi^{00}+\sum_{l}(\Pi^{0l}+\Pi^{ll}), \tag{5.2.14}$$

где

$$\Pi^{00}=\begin{pmatrix}\pi^{00} & 0 & \ldots\\ 0 & 0 & \ldots\\ & \vdots & \end{pmatrix},\qquad \Pi^{01}=\begin{pmatrix}0 & \pi^{01} & 0 & \ldots\\ 0 & 0 & 0 & \ldots\\ \vdots & \vdots & \vdots & \end{pmatrix},$$

и т. д.

Здесь π^{ij} означает матрицу, имеющую размерность $r_i\times r_j$ и описывающую переход из подмножества E_i, содержащего r_i состояний, в подмножество E_j с r_j состояниями. На прочих местах стоят нули. Множества E_1, E_2, ... образуют эргодические классы. Из множества E_0 происходят переходы в эргодические классы E_1, E_2, ... Из каждого такого класса нет выхода. Поэтому внутри каждого класса устанавливается свое стационарное распределение. Оно может быть найдено по формуле типа (5.2.13)

$$P^{l}_{\text{ст}}(\xi)=A^{ll}_{\xi\xi'}\Big/\sum_{\xi''\in E_l}A^{ll}_{\xi''\xi'}\quad\text{при }\xi\in E_l \tag{5.2.15}$$

с той лишь разницей, что теперь берутся алгебраические дополнения подматрицы $\|\pi^{ll}_{\xi\xi'}-\delta_{\xi\xi'}\|$, ξ, $\xi'\in E_l$, которые уже не равны нулю. Вероятности $P^{l}_{\text{ст}}(\xi)$ сосредоточены на E_l. Полное стационарное распределение есть линейная комбинация

$$P_{\text{ст}}(\xi)=\sum q_i P^i_{\text{ст}}(\xi) \tag{5.2.16}$$

частных распределений (5.2.15), которые, как легко видеть, являются ортогональными

$$\sum_{\xi} P^i_{\text{ст}}(\xi) P^j_{\text{ст}}(\xi)=0 \qquad \text{при } i \neq j.$$

Коэффициенты q_i этой линейной комбинации определяются начальным распределением $P(\xi_1) = P_1(\xi_1)$. Они совпадают с результирующей вероятностью попадания в тот или иной эргодический класс и удовлетворяют условию нормировки $\sum\limits_i q_i = 1$. Принимая во внимание вид (5.2.14) переходной матрицы и суммируя переходы из E_0 в E_i на различных этапах, можно найти, как они явно выражаются через начальное распределение:

$$q_i=\sum_{\xi\in E_i} P_1(\xi)+\sum_{\xi\xi'} P_1(\xi)\left[\Pi^{0i}_{\xi\xi'}+\right.$$
$$\left.+(\Pi^{00}\Pi^{0i})_{\xi\xi'}+(\Pi^{00}\Pi^{00}\Pi^{0i})_{\xi\xi'}+\ldots\right],\ i\neq 0.$$

В круглых скобках здесь стоит матричное произведение. Суммируя степени матрицы Π^{00}, имеем

$$q_i=\sum_{\xi\in E_i} P_1(\xi)+\sum_{\xi,\,\xi'} P_1(\xi)\left([1-\Pi^{00}]^{-1}\Pi^{0i}\right)_{\xi\xi'},\ q_0=0.$$

Учитывая (5.2.16), (5.2.14), получаем, что удельную энтропию (5.2.8) удобно представить суммой

$$H_1=\sum q_i H^i_1, \tag{5.2.17}$$

где частные энтропии

$$H^i_1=-\sum_{\xi,\,\xi'\in E_i} P^i_{\text{ст}}(\xi)\,\pi^{ii}(\xi,\xi')\ln\pi^{ii}(\xi,\xi') \tag{5.2.18}$$

вычисляются совершенно аналогично эргодическому случаю. Причина этого в том, что неэргодический процесс является статистической смесью (с вероятностями q_i) эргодических процессов, имеющих меньшее число состояний r_i.

Суммирование в (5.2.17), (5.2.18) проводится только по эргодическим классам E_1, E_2, ... Подпространство же E_0 имеет нулевую стационарную вероятность. Соединение всех эргодических классов $E_1+E_2+\ldots=E_a$, на котором концентрируется стационарная вероятность, можно назвать «активным» подпространством. На удельную энтропию H_1 распределения и переходы в «пассивном» пространстве E_0 оказывают влияние лишь постольку, поскольку они влияют на вероятности q_i. Если процесс эргодический, т. е. имеется, кроме E_0, лишь один эргодический класс E_1, тогда $q_1 = 1$ и пассивное пространство вообще не оказывает никакого влияния на удельную энтропию. В этом случае удельная энтропия марковского процесса в пространстве $E_0 + E_a$ совпадает с удельной энтропией марковского процесса, протекающего в подпространстве $E_a = E_1$ и имеющего матрицу перехода π^{11}.

3. Чтобы проиллюстрировать применение формул, полученных в этом параграфе, рассмотрим несколько несложных примеров.

Пример 1. Рассмотрим сначала простейший дискретный марковский процесс — процесс с двумя состояниями. Матрица (5.2.12) является при этом 2×2-матрицей. Вследствие условия нормировки (5.2.2) ее элементы не являются независимыми. Имеются лишь два независимых параметра μ и ν, определяющих матрицу π:

$$\boldsymbol{\pi} = \begin{pmatrix} 1-\mu & \mu \\ \nu & 1-\nu \end{pmatrix}.$$

Поскольку в данном случае

$$\boldsymbol{a} = \boldsymbol{\pi} - \mathbf{1} = \begin{pmatrix} -\mu & \mu \\ \nu & -\nu \end{pmatrix}; \quad A_{11} = -\nu; \quad A_{21} = -\mu,$$

то согласно (5.2.13) имеем стационарное распределение

$$P_{\text{ст}}(1) = \frac{\nu}{\mu+\nu}; \qquad P_{\text{ст}}(2) = \frac{\mu}{\mu+\nu} \tag{5.2.19}$$

и кроме того

$$\| P_{\text{ст}}(\xi_k, \xi_{k+1}) \| = \frac{1}{\mu+\nu} \begin{pmatrix} \nu-\mu\nu & \mu\nu \\ \mu\nu & \mu-\mu\nu \end{pmatrix}.$$

Применяя формулу (5.2.8), находим удельную энтропию

$$H_1 = \frac{\nu}{\mu+\nu} h_2(\mu) + \frac{\mu}{\mu+\nu} h_2(\nu), \tag{5.2.20}$$

где

$$h_2(x) = -x \ln x - (1-x) \ln(1-x). \tag{5.2.20a}$$

Далее по формуле (5.2.10) нетрудно получить граничную энтропию

$$2\Gamma = h_2\left(\frac{\mu}{\mu+\nu}\right) - \frac{\nu h_2(\mu) + \mu h_2(\nu)}{\mu+\nu}.$$

Пример 2. Пусть теперь имеется процесс с тремя состояниями, имеющий переходную матрицу

$$\boldsymbol{\pi} = \begin{pmatrix} 1-\mu & \mu' & \mu'' \\ \nu'' & 1-\nu & \nu' \\ \lambda' & \lambda'' & 1-\lambda \end{pmatrix} \qquad (\mu = \mu' + \mu'' \text{ и т. д.})$$

При этом, очевидно,

$$\boldsymbol{a} = \begin{pmatrix} -\mu & \mu' & \mu'' \\ \nu'' & -\nu & \nu' \\ \lambda' & \lambda'' & -\lambda \end{pmatrix}.$$

По формуле (5.2.13) находим стационарное распределение

$$P_{\text{ст}}(\xi)=A_{\xi 1}/(A_{11}+A_{21}+A_{31}), \tag{5.2.21}$$

где

$$A_{11}=\begin{vmatrix}-\nu & \nu' \\ \lambda'' & -\lambda\end{vmatrix}=\lambda\nu-\lambda''\nu'=\lambda'\nu'+\lambda'\nu''+\lambda''\nu'';$$

$$A_{21}=\mu'\lambda'+\mu'\lambda''+\lambda''\mu'';$$

$$A_{31}=\mu'\nu'+\nu'\mu''+\nu''\mu''.$$

Удельная энтропия (5.2.8) оказывается равной

$$H_1 \quad P_{\text{ст}}(1)h_3(\mu', \mu'')+P_{\text{ст}}(2)h_3(\nu', \nu'')+P_{\text{ст}}(3)h_3(\lambda', \lambda''),$$

где обозначено

$$h_3(\mu', \mu'')=-\mu'\ln\mu'-\mu''\ln\mu''-(1-\mu'-\mu'')\ln(1-\mu'-\mu''). \tag{5.2.22}$$

Данный процесс с тремя состояниями будет неэргодическим, если, например, $\lambda' = \lambda'' = 0$; $\mu'' = \nu' = 0$, так что матрица перехода имеет «ящичный» вид

$$\pi=\begin{pmatrix}1-\mu & \mu & 0 \\ \nu & 1-\nu & 0 \\ 0 & 0 & 1\end{pmatrix}.$$

При такой матрице третье состояние остается неизменным, а переходы совершаются только между первым и вторым состояниями. Алгебраические дополнения (5.2.21) обращаются в нуль. Стационарными, как легко понять, являются следующие распределения:

$$P^1_{\text{ст}}(\xi)=\begin{cases}\nu/(\mu+\nu) & \text{при } \xi=1; \\ \mu/(\mu+\nu) & \text{при } \xi=2; \\ 0 & \text{при } \xi=3;\end{cases} \qquad P^2_{\text{ст}}(\xi)=\delta_{\xi 3}.$$

Первое из них совпадает с (5.2.19). Указанные функции $P^1_{\text{ст}}(\xi)$, $P^2_{\text{ст}}(\xi)$ являются ортогональными. Задаваясь начальным распределением $P_1(\xi)$, по формуле (5.2.16) находим результирующее стационарное распределение

$$P_{\text{ст}}(\xi)=[P_1(1)+P_1(2)]\,P^1_{\text{ст}}(\xi)+P_1(3)\,P^2_{\text{ст}}(\xi).$$

Теперь согласно (5.2.17) нетрудно записать удельную энтропию

$$H_1=[P_1(1)+P_1(2)]\,(\nu h(\mu)+\mu h(\nu))/(\mu+\nu).$$

5.3. Удельная энтропия части компонент дискретного марковского процесса и условного марковского процесса

1. Вычислим удельную энтропию стационарного немарковского дискретного процесса $\{y_k\}$, который, однако, может быть дополнен до марковского процесса присоединением добавочных компонент x_k. Совокупность $\xi = (x, y)$ будет образовывать фазовое пространство марковского процесса, и процесс $\{\xi_k\} = \{x_k, y_k\}$ уже будет марковским. К нему можно применять результаты предыдущего параграфа и, в частности, находить удельную энтропию, которую будем обозначать $H^1_{xy} = h_{xy}$. Энтропия $H_{y_1 \ldots y_n}$ исходного процесса отличается от энтропии

$$H_{\xi_1 \ldots \xi_n} = H_{y_1 \ldots y_n} + H_{x_1 \ldots x_n \mid y_1 \ldots y_n} \tag{5.3.1}$$

марковского процесса на условную энтропию $H_{x_1 \ldots x_n \mid y_1 \ldots y_n}$ называемую *энтропией условного марковского процесса* $\{x_k\}$ (при фиксированном $\{y_k\}$).

Наряду с удельной энтропией $h_{xy} = h_\xi$ стационарного марковского процесса введем удельную энтропию исходного y-процесса

$$h_y = \lim_{n \to \infty} \frac{1}{n} H_{y_1 \ldots y_n} \tag{5.3.2}$$

и условного x-процесса

$$h_{x \mid y} = \lim_{n \to \infty} \frac{1}{n} H_{x_1 \ldots x_n \mid y_1 \ldots y_n}. \tag{5.3.3}$$

Они вследствие (5.3.1) связаны с h_{xy} соотношением

$$h_y + h_{x \mid y} = h_{xy}. \tag{5.3.4}$$

Нетрудно понять, что можно также рассматривать x-процесс как априорный немарковский (в общем случае) процесс и условный y-процесс при фиксированном x. Им будут соответствовать удельные энтропии

$$h_x = \lim_{n \to \infty} \frac{1}{n} H_{x_1 \ldots x_n}; \qquad h_{y \mid x} = \lim_{n \to \infty} H_{y_1 \ldots y_n \mid x_1 \ldots x_n}. \tag{5.3.5}$$

Соотношения (5.3.1), (5.3.4) при этом заменяются на

$$H_{x_1 \ldots x_n} + H_{y_1 \ldots y_n \mid x_1 \ldots x_n} = H_{\xi_1 \ldots \xi_n}; \tag{5.3.6}$$

$$h_x + h_{y \mid x} = h_{xy}. \tag{5.3.7}$$

Поскольку энтропию марковского процесса мы уже умеем находить, достаточно научиться вычислять лишь одну из величин h_y или $h_{x|y}$. Вторая находится при помощи (5.3.4). Вследствие симметрии соответствующая величина из h_x, $h_{y|x}$ будет вычисляться тем же способом, а вторая — из соотношения (5.3.7).

Излагаемый ниже (п.2) метод можно применять для вычисления энтропии условного марковского процесса как в стационарном, так и в нестационарном случае. Стационарные же распределения вероятностей и пределы (5.3.2), (5.3.3), (5.3.5), как правило, существуют только в стационарном случае.

Условную энтропию $H_{x_1 \ldots x_n | y_1 \ldots y_n}$ можно представить в виде суммы

$$H_{x_1 \ldots x_n | y_1 \ldots y_n} = H_{x_1 | y_1 \ldots y_n} + H_{x_2 | x_1 y_1 \ldots y_n} + \\ + H_{x_3 | x_1 x_2 y_1 \ldots y_n} + \ldots + H_{x_n | x_1 \ldots x_{n-1} y_1 \ldots y_n}$$

и предел (5.3.3) заменить на предел

$$h_{x|y} = \lim_{k \to \infty,\ n-k \to \infty} H_{x_k | x_1 \ldots x_{k-1}, y_1 \ldots y_n} = \\ = H_{x_k | \ldots \xi_{k-2} \xi_{k-1} y_k y_{k+1} \ldots}. \tag{5.3.8}$$

Об эквивалентности двух таких представлений удельной энтропии была речь в теореме 5.1. Конечно, теперь процесс $\{x_k\}$ является условным и поэтому нестационарным. Прямо применять к нему теорему 5.1. нельзя, поэтому необходимо произвести ее обобщение.

Теорема 5.2. *Для стационарного процесса* $\{\xi_k\} = \{x_k, y_k\}$ *пределы (5.3.3) и (5.3.8) совпадают.*

Доказательство. Поскольку

$$\lim_{n-k \to \infty} H_{x_k | x_1 \ldots x_{k-1} y_1 \ldots y_n} = H_{x_k | \xi_1 \ldots \xi_{k-1} y_k y_{k+1} \ldots},$$
$$\lim_{k \to \infty} H_{x_k | \xi_1 \ldots \xi_{k-1} y_k y_{k+1} \ldots} = H_{x_k | \ldots \xi_{k-2} \xi_{k-1} y_k y_{k+1} \ldots},$$

то имеем

$$H_{x_k | x_1 \ldots x_{k-1} y_1 \ldots y_n} = H_{x_k | \ldots \xi_{k-2} \xi_{k-1} y_k y_{k+1} \ldots} + \\ + o_k(1) + o_{n-k}(1) \tag{5.3.9}$$

(предполагается, что оценка $o_{n-k}(1)$ равномерна по k).

Представим n в виде суммы трех чисел: $n = m + r + s$. Подставляя (5.3.9) в равенство

$$H_{x_1 \ldots x_n | y_1 \ldots y_n} = H_{x_1 \ldots x_m | y_1 \ldots y_n} + H_{x_{m+1} | \xi_1 \ldots \xi_m y_{m+1} \ldots y_n} + \\ + \ldots + H_{x_{m+r} | \xi_1 \ldots \xi_{m+r-1} y_{m+r} \ldots y_n} + \\ + H_{x_{m+r+1} \ldots x_n | \xi_1 \ldots \xi_{m+r} y_{m+r+1} \ldots y_n},$$

получаем

$$\frac{1}{n} H_{x_1 \ldots x_n | y_1 \ldots y_n} = \frac{1}{m+r+s} H_{x_1 \ldots x_m | y_1 \ldots y_n} + \\ + \frac{r}{m+r+s} H_{x_k | \ldots \xi_{k-1} y_k \ldots} + \frac{r}{m+r+s} [o_m(1) + o_s(1)] +$$

$$+\frac{1}{m+r+s} H_{x_{m+r+1} \ldots x_n \mid \xi_1 \ldots \xi_{m+r} y_{m+r+1} \ldots y_n}. \qquad (5.3.10)$$

Здесь

$$H_{x_1 \quad x_m \mid y_1 \quad y_n} = \sum_{k=1}^{m} H_{x_k \mid x_1 \quad x_{k-1} y_1 \quad y_n} \leqslant mH_{x_k}$$

и

$$H_{x_{m+r+1} \cdot \quad x_n \mid \xi_1 \quad \xi_{m+r} y_{m+r+1} \quad y_n} \leqslant sH_{x_k},$$

так как условная энтропия не больше безусловной. Поэтому, если в (5.3.10) совершить предельный переход $m \to \infty$, $r \to \infty$, $s \to \infty$, причем $r/m \to \infty$; $r/s \to \infty$, то в этом выражении останется лишь один член $H_{x_k \mid \quad, \xi_{k-1}, y_k}$. Это доказывает теорему

Как видно из приведенного доказательства, теорема 5.2 справедлива не только в случае марковского совокупного процесса $\{x_k, y_k\}$. Для марковского же процесса вследствие условия Маркова, имеем

$$P(x_k \mid x_1, \ldots, x_{k-1} y_1, \ldots, y_n) =$$
$$= P(x_k \mid x_{k-1}, y_{k-1}, y_k, \ldots, y_n) \quad (n \geqslant k)$$

и

$$H_{x_k \mid x_1 \quad x_{k-1}, y_1 \quad y_n} = H_{x_k \mid \xi_{k-1} y_k \quad y_n}.$$

Следовательно, формула (5.3.8) принимает вид

$$h_{x \mid y} = H_{x_k \mid x_{k-1} y_{k-1} y_k y_{k+1}} \quad .$$

2. Перейдем к вычислению энтропии y-процесса и условного процесса для марковского совокупного процесса. Для этого рассмотрим условную энтропию

$$H_{y_k \mid y_1 \ldots y_{k-1}} = -\mathrm{M} \ln P(y_k \mid y_1, \ldots, y_{k-1}), \qquad (5.3.11)$$

определяющую согласно (5.1 3) в пределе

$$h_y = \lim_{k \to \infty} H_{y_k \mid y_1 \ldots y_{k-1}} \qquad (5.3.12)$$

удельную энтропию h_y.

Пользуясь условием марковости (5 2 1), запишем многомерное распределение вероятностей

$$P(\xi_1, \ldots, \xi_n) = P(\xi_1)\, \pi(\xi_1, \xi_2) \ldots \pi(\xi_{k-1}, \xi_k),$$

где ξ_j обозначает пару x_j, y_j

Применяя формулу обратной вероятности (формулу Бейеса), отсюда получаем

$$P(y_k \mid y_1, \ldots, y_{k-1}) = \frac{\sum\limits_{x_1 \ldots x_k} P(\xi_1)\, \pi(\xi_1, \xi_2) \ldots \pi(\xi_{k-1}, \xi_k)}{\sum\limits_{x_1 \ldots x_{k-1}} P(\xi_1)\, \pi(\xi_1, \xi_2) \ldots \pi(\xi_{k-2}, \xi_{k-1})}. \qquad (5.3\ 13)$$

Следуя теории условных марковских процессов (Стратонович [9]), введем *финальные апостериорные вероятности*

$$P(x_{k-1} \mid y_1, \ldots, y_{k-1}) \equiv W_{k-1}(x_{k-1}) = \frac{\sum\limits_{x_1 \ldots x_{k-2}} P(\xi_1)\,\pi(\xi_1, \xi_2) \ldots \pi(\xi_{k-2}, \xi_{k-1})}{\sum\limits_{x_1 \ldots x_{k-1}} P(\xi_1)\,\pi(\xi_1, \xi_2) \ldots \pi(\xi_{k-2}, \xi_{k-1})}. \tag{5.3.14}$$

При помощи них выражение (5.3.13) запишется так:

$$P(y_k \mid y_1, \ldots, y_{k-1}) = \sum_{x_{k-1}, x_k} W_{k-1}(x_{k-1})\,\pi(x_{k-1}, y_{k-1}; x_k, y_k)$$

или

$$P(y_k \mid y_1, \ldots, y_{k-1}) = \sum_x W_{k-1}(x)\,\pi(x, y_{k-1}; y_k), \tag{5.3.15}$$

если обозначить $\sum\limits_{x'} \pi(x, y; x', y') = \pi(x, y; y')$, и формула (5.3.11) принимает вид

$$H_{y_k \mid y_1 \ldots y_{k-1}} = -\mathbf{M} \ln \left[\sum_x W_{k-1}(x)\,\pi(x, y_{k-1}; y_k)\right]$$

Здесь символ $\mathbf{M}$ указывает усреднение по $y_1, \ldots, y_{k-1}, y_k$. Если же подставить (5.3.15) в формулу

$$H_{y_k \mid y_1 \ldots y_{k-1}} = -\mathbf{M} \sum_{y_k} P(y_k \mid y_1, \ldots, y_{k-1}) \ln P(y_k \mid y_1, \ldots, y_{k-1}),$$

то будем иметь

$$H_{y_k \mid y_1 \ldots y_{k-1}} = -\mathbf{M} \sum_{y_k} \left[\sum_x W_{k-1}(x)\,\pi(x, y_{k-1}; y_k)\right] \times \\ \times \ln \left[\sum_x W_{k-1}(x)\,\pi(x, y_{k-1}; y_k)\right], \tag{5.3.16}$$

де символ $\mathbf{M}$ обозначает усреднение по $y_1, \ldots, y_{k-1}$.

Удобно рассматривать апостериорные вероятности $W_{k-1}(\cdot)$ как распределение по обеим переменным $(x, y) = \xi$, доопределяя их формулой

$$W_{k-1}(x, y) = \begin{cases} W_k(x) & \text{при } y = y_{k-1}, \\ 0 & \text{при } y \neq y_{k-1}. \end{cases} \tag{5.3.17}$$

Тогда формула (5.3.15) примет вид

$$P(y_k \mid y_1 \ldots y_{k-1}) = \sum_\xi W_{k-1}(\xi)\,\pi(\xi, y_k)$$

и равенство (5.3.16) можно будет записать

$$H_{y_k \mid y_1 \ldots y_{k-1}} = \\ = -\mathbf{M} \sum_{y_k} \sum_\xi W_{k-1}(\xi)\,\pi(\xi, y_k) \ln \sum_{\xi'} W_{k-1}(\xi')\,\pi(\xi', y_k). \tag{5.3.18}$$

Здесь, как и в (5.3.16), **M** указывает усреднение по $y_1, \ldots, y_{k-1}$, но подлежащее усреднению выражение, очевидно, зависит от $y_1, \ldots, y_{k-1}$ лишь через посредство апостериорных вероятностей $W_{k-1}(\cdot)$. Поэтому в (5.3.18) можно считать, что символ **M** соответствует усреднению по случайным величинам $W_{k-1}(A_1), \ldots, W_{k-1}(A_L)$, ($A_1, \ldots, A_L$ — пространство состояний процесса $\xi = (x, y)$, L — число его состояний). Вводя закон распределения $P(dW_{k-1})$ для этих случайных величин, придаем формуле (5.3.18) вид

$$H_{y_k \mid y_1 \ldots y_{k-1}} = \\ = -\int P(dW_{k-1}) \sum_{y'} \sum_{\xi} W_{k-1}(\xi)\, \pi(\xi, y') \ln \sum_{\xi'} W_{k-1}(\xi')\, \pi(\xi', y'). \tag{5.3.19}$$

Это и есть основной результат. Мы видим, что для вычисления условной энтропии части компонент марковского процесса нужно исследовать апостериорные вероятности $\{W_k(\cdot)\}$ как самостоятельный случайный процесс. Этот процесс изучается в теории условных марковских процессов. Известно, что с ростом k указанные вероятности преобразуются по определенным рекуррентным соотношениям. Чтобы вывести последние, запишем равенство, аналогичное (5.3.14), но с заменой $k-1$ на k:

$$W_k(x_k) = \frac{\sum\limits_{x_{k-1}} \left[\sum\limits_{x_1 \ldots x_{k-2}} P(\xi_1)\, \pi(\xi_1, \xi_2) \ldots \pi(\xi_{k-2}, \xi_{k-1}) \right] \pi(\xi_{k-1}, \xi_k)}{\sum\limits_{x_{k-1}\, x_k} \left[\sum\limits_{x_1 \ldots x_{k-2}} P(\xi_1)\, \pi(\xi_1, \xi_2) \ldots \pi(\xi_{k-2}, \xi_{k-1}) \right] \pi(\xi_{k-1}, \xi_k)}.$$

Подставляя сюда (5.3.14), получаем рекуррентные соотношения

$$W_k(x_k) = \frac{\sum\limits_{x_{k-1}} W_{k-1}(x_{k-1})\, \pi(x_{k-1}, y_{k-1}; x_k, y_k)}{\sum\limits_{x_{k-1},\, x_k} W_{k-1}(x_{k-1})\, \pi(x_{k-1}, y_{k-1}; x_k, y_k)}. \tag{5.3.20}$$

Процесс $\{W_k(\cdot)\}$, рассматриваемый как самостоятельный случайный процесс, носит название вторичного апостериорного W-процесса. Как известно из теории условных марковских процессов, он является марковским. Рассмотрим его вероятность перехода.

Преобразование (5.3.20), которое можно также записать

$$W_k(\xi) = \frac{\sum\limits_{\xi'} W_{k-1}(\xi')\, \pi(\xi', \xi)\, \delta_{y y_k}}{\sum\limits_{\xi''} W_{k-1}(\xi'')\, \pi(\xi'', y_k)}, \tag{5.3.21}$$

явно зависит от случайной величины y_k. Ее апостериорные вероятности $P(y_k \mid y_1, \ldots, y_{k-1})$ совпадают с (5.3.15) и, следовательно, полностью определяются «состоянием» $W_{k-1}(\cdot)$ вторичного апостериорного процесса в момент $k-1$. Итак, каждое преобразование (5.3.21) осуществляется с вероятностью (5.3.15). Это значит, что «плотность вероятности» перехода вторичного процесса $\{W_k\}$ можно записать

$$\pi^W(W_{k-1}, W_k) = \sum_{y_k} \delta\left(W_k(\xi) - \frac{\delta_{yy_k} \sum_{\xi'} W_{k-1}(\xi')\pi(\xi', \xi)}{\sum_{\xi''} W_{k-1}(\xi'')\pi(\xi'', y_k)} \right) \times$$
$$\times \sum_{\xi'} W_{k-1}(\xi')\pi(\xi', y_k). \tag{5.3.22}$$

Здесь $\delta(W)$ — L-мерная дельта-функция, определяемая формулой

$$\int f(W)\delta(W)\prod_{\xi} dW(\xi) = f(0).$$

Она соответствует мере, сосредоточенной в начале координат L-мерного пространства.

Итак, мы видим, что *энтропия процесса* $\{y_k\}$, *немарковского, но являющегося частью марковского процесса, может быть определена методами теории марковских процессов* после перехода к вторичному апостериорному уже марковскому процессу. В стационарном случае обычным способом может быть найдено стационарное распределение $P_{\text{ст}}(dW)$, к которому сходится при $k \to \infty$ распределение $P_k(dW)$, входящее в (5.3.19). Подставляя (5.3.19) в (5.3.12) и переходя к пределу $k \to \infty$, получаем удельную энтропию

$$h_y = -\int P_{\text{ст}}(dW) \sum_{y\xi} W(\xi)\pi(\xi, y) \ln \sum_{\xi'} W(\xi')\pi(\xi', y). \tag{5.3.23}$$

Затем уже по формуле (5.3.4), если нужно, можно найти $h_{x|y}$.

Прежде чем перейти к рассмотрению примера, приведем следующую теорему, имеющую прямое отношение к вышеизложенному.

Теорема 5.3. *Энтропия $H_{y_2 \ldots y_l \mid y_1}$ немарковского y-процесса совпадает с аналогичной энтропией вторичного апостериорного (марковского) процесса:*

$$H_{y_2 \ldots y_l \mid y_1} = H_{W_2 \ldots W_l \mid W_1}. \tag{5.3.24}$$

Следовательно, совпадают и удельные энтропии: $h_y = h_W$.

Доказательство. Поскольку справедливы равенства

$$H_{y_2 \ldots y_l \mid y_1} = \sum_{k=2}^{l} H_{y_k \mid y_1 \ldots y_{k-1}},$$
$$H_{W_2 \ldots W_l \mid W_1} = \sum_{k=2}^{l} H_{W_k \mid W_1 \ldots W_{k-1}},$$

а также формула (5.3.12) и аналогичная формула для $\{W_k\}$, то достаточно доказать равенство

$$H_{y_k \mid y_1 \ldots y_{k-1}} = H_{W_k \mid W_1 \ldots W_{k-1}} = H_{W_k \mid W_{k-1}}. \tag{5.3.25}$$

Здесь $H_{w_k | w_1 \ldots w_{k-1}} = H_{w_k | w_{k-1}}$ в силу марковского свойства.

Пусть S_0 — некоторая исходная точка пространства значений $W(\cdot)$. В соответствии с (5.3.21) из нее возможны переходы в различные точки $S(y_k)$ в зависимости от значения y_k ($S(y_k)$ — точка с координатами (5.3.21)). Эти точки не совпадают при несовпадающих y_k. В самом деле, если бы две точки из $S(y_k)$, скажем $S' = S(y')$ и $S'' = S(y'')$, совпадали, т. е. имело бы место равенство

$$W'_k(x, y) = W''_k(x, y),$$

из него можно было бы получить

$$\sum_x W'_k(x, y) = \sum_x W''_k(x, y). \tag{5.3.26}$$

Но согласно (5.3.17)

$$\sum_x W'_k(x, y) = \delta_{yy_k} = \delta_{yy'},$$

и равенство (5.3.26) означало бы совпадение $y' = y''$.

Итак, между точками $S(y_k)$ и значениями y_k можно установить взаимно однозначное соответствие. Поэтому

$$P(S(y_k) | S_0) = P(y_k | y_1, \ldots, y_{k-1}),$$
$$H_{S(y_k)}(\,| S_0) = H_{y_k}(\,| y_1, \ldots, y_{k-1}).$$

Но при должном выборе точки $S_0 = W_{k-1}$ события, стоящие в условии, совпадают, и после усреднения мы получаем (5.3.25). Это завершает доказательство.

3. **Пример.** Пусть немарковский процесс $\{y_k\}$ есть процесс с двумя состояниями, т. е. y_k может принимать одно из двух значений, скажем 1 или 2. Пусть далее его можно превратить в марковский, если добавлением переменной x разбить состояние $y = 2$ на два состояния: $\xi = 2$ и состояние $\xi = 3$. Именно, $\xi = 2$ соответствует $x = 1$, $y = 2$, а $\xi = 3$ соответствует $x = 2$, $y = 2$. С состоянием $y = 1$ можно сопоставить, скажем, $\xi = 1$ и $x = 1$.

Совокупный процесс $\{\xi_k\} = \{x_k, y_k\}$ является стационарным марковским и описывается матрицей перехода

$$\pi_{\xi\xi'} = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix}.$$

Апостериорные распределения $W_k(\xi) = W_k(x, y)$ согласно (5.3.17) имеют вид

$$(W_k(\xi = 1),\ W_k(\xi = 2),\ W_k(\xi = 3)) = (1,\ 0,\ 0) \quad \text{при } y_k = 1,$$
$$\tag{5.3.27}$$
$$W_k(\xi = 1),\ W_k(\xi - 2),\ W_k(\xi = 3)) = (0,\ W_k(2),\ W_k(3)) \quad \text{при } y_k = 2.$$

Значению $y_k = 2$ согласно (5.3.21) соответствует преобразование

$$W_k(1)=0;$$

$$W_k(2)=\frac{W_{k-1}(1)\,\pi_{12}+W_{k-1}(2)\,\pi_{22}+W_{k-1}(3)\,\pi_{32}}{W_{k-1}(1)\,(\pi_{12}+\pi_{13})+W_{k_1}(2)\,(\pi_{22}+\pi_{23})+W_{k-1}(3)\,(\pi_{32}+\pi_{33})};\tag{5.3.28}$$

$$W_k(3)=1-W_k(2).$$

Точку (1, 0, 0) трехмерного пространства значений (W (1), W (2), W (3)), соответствующую распределению (5.3.27), обозначим S_0. Исследуем возможные переходы из этой точки. Подставляя значение $W_{k-1}=(1,\ 0,\ 0)$ в (5.3.28), получаем переход в точку

$$W(0)=0;$$

$$W(2)=\frac{\pi_{12}}{\pi_{12}+\pi_{13}};\tag{5.3.29}$$

$$W(3)=\frac{\pi_{13}}{\pi_{12}+\pi_{13}},$$

которую обозначим S_1. Вследствие (5.3.18) такой переход осуществляется с вероятностью

$$p_1=\pi_{12}+\pi_{13}.\tag{5.3.30}$$

С вероятностью $1-p_1=\pi_{11}$ сохраняется пребывание в точке S_0.

Рассмотрим теперь переходы из точки S_1. Подставляя значения (5.3.29) в качестве W_{k-1} в формулу (5.3.28) при $y_k=2$, получаем координаты

$$W(1)=0;$$

$$W(2)=\frac{\pi_{12}\,\pi_{22}+\pi_{13}\,\pi_{32}}{\pi_{12}\,(\pi_{22}+\pi_{23})+\pi_{13}\,(\pi_{32}+\pi_{33})};\tag{5.3.31}$$

$$W(3)=1-W(2)$$

новой точки S_2. Переход из S_1 в S_2 происходит с вероятностью

$$p_2=\frac{\pi_{12}\,(\pi_{22}+\pi_{23})+\pi_{13}\,(\pi_{32}+\pi_{33})}{\pi_{12}+\pi_{13}}.\tag{5.3.32}$$

Это выражение получено подстановкой (5.3.29) в формулу

$$\mathbf{P}\,(y_k=2\mid y_1,\ \dots,\ y_{k-1})=P\,(S_k\mid S_{k-1})=$$
$$=W_{k-1}(2)\,(\pi_{22}+\pi_{23})+W_{k-1}(3)\,(\pi_{32}+\pi_{33}),\tag{5.3.33}$$

получаемую из (5.3.17а). С вероятностью $1-p_2$ происходит возврат в точку S_0. Аналогично, подставляя значения (5.3.31) в (5.3.33), находим вероятность

$$p_3=\frac{(\pi_{12}\,\pi_{22}+\pi_{13}\,\pi_{32})\,(\pi_{22}+\pi_{23})+(\pi_{12}\,\pi_{23}+\pi_{13}\,\pi_{33})\,(\pi_{32}+\pi_{33})}{\pi_{12}\,(\pi_{22}+\pi_{23})+\pi_{13}\,(\pi_{32}+\pi_{33})}\tag{5.3.34}$$

перехода из S_2 в следующую точку S_3 и т. д. Описанным способом последовательно вычисляются вероятности перехода p_k в следующие друг за другом точки S_k. Каждый раз с вероятностью $1 - p_k$ происходит возврат в точку S_0. Вероятность того, что к моменту k еще не произошел возврат в точку S_0, очевидно, равна $p_1 p_2 \ldots p_k$. Если последовательные значения p_k не стремятся к единице, то эта вероятность стремится к нулю при $k \to \infty$. Поэтому обычно с вероятностью, равной единице, происходит в итоге возврат в точку S_0. Если бы в качестве начальной точки была взята какая-либо другая точка S_0', то происходило бы движение по какой-то другой последовательности точек с последующим переходом (сколь угодно близким к достоверности) в точку S_0. После такого перехода будет происходить уже описанное выше движение по точкам S_0, S_1, S_2, ... с вычисленными вероятностями переходов.

Описанная выше картина переходов вторичного марковского процесса позволяет легко найти стационарное распределение вероятностей. Оно будет сосредоточено в точках S_0, S_1, S_2, ...

Для этих точек матрица вероятностей перехода имеет вид

$$\pi^W = \begin{pmatrix} 1-p_1 & p_1 & 0 & 0 & \cdot \\ 1-p_2 & 0 & p_2 & 0 & \cdot \\ 1-p_3 & 0 & 0 & p_3 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}. \tag{5.3.35}$$

Учитывая равенства $P_{\text{ст}}(S_k) = p_k P_{\text{ст}}(S_{k-1})$, $k = 1, 2, \ldots$, вытекающие из (5.2.7), (5.3.35), и условие нормировки, находим стационарные вероятности указанных точек

$$P_{\text{ст}}(S_0) = \frac{1}{1+p_1+p_1 p_2+\ldots}; \quad P_{\text{ст}}(S_k) = \frac{p_1 \ldots p_k}{1+p_1+p_1 p_2+\ldots}, \quad k = 1, 2, \ldots \tag{5.3.36}$$

Для вычисления удельной энтропии (5.3.12) остается лишь подставить выражения (5.3.36) в формулу

$$h_y = -\sum_{k=0}^{\infty} P_{\text{ст}}(S_k)\,[p_{k+1} \ln p_{k+1} + (1-p_{k+1}) \ln (1-p_{k+1})], \tag{5.3.37}$$

которая вытекает из (5.3.23) или из теоремы 5.3 и формулы (5.2.8), примененной к вероятностям перехода (5.3.35).

В том частном случае, когда значение x не влияет на переходы от одного значения y к другому, имеем

$$\pi_{22} + \pi_{23} = \pi_{32} + \pi_{33}. \tag{5.3.38}$$

Здесь выписана вероятность $\mathbf{P}(y_k = 2 \mid y_{k-1} = 2, x) = 1 - \nu$ (ν — вероятность перехода $\mathbf{P}(y_k = 1 \mid y_{k-1} = 2, x)$), которая теперь не зависит от x. В этом случае (5.3.33) дает

$$\begin{aligned} p_2 = p_3 = \ldots &= [W_{k-1}(2) + W_{k-1}(3)](\pi_{22}+\pi_{23}) = \\ &= \pi_{22}+\pi_{23} = 1-\nu \end{aligned} \tag{5.3.39}$$

и из (5.3.37) получаем

$$h_y = P_{\text{ст}}(S_0)\, h_2(\mu) + [1 - P_{\text{ст}}(S_0)]\, h_2(\nu) \qquad (\mu = p_1 = \pi_{12} + \pi_{13}), \tag{5.3.40}$$

причем в силу (5.3.36), (5.3.39)

$$P_{\text{ст}}(S_0) = \left[1 + \frac{p_1}{1-p_2}\right]^{-1} = \frac{\nu}{\nu+\mu}.$$

Формула (5.3.40) поэтому совпадает с (5.2.20), и это естественно, поскольку при выполнении условия (5.3.38) процесс $\{y_k\}$ становится марковским сам по себе.

В рассмотренном примере стационарные вероятности были сосредоточены на счетном множестве точек. Применяя терминологию из § 5.2, можно сказать, что активное пространство E_a W-процесса состояло из точек S_0, S_1, S_2, ... и, следовательно, было счетным, хотя полное пространство $E_a + E_0$ представляло собой континуум. Можно ожидать, что и в других случаях дискретных процессов $\{x_k, y_k\}$ положение такое же, то есть в довольно общем случае конечного или счетного числа состояний совокупного процесса $\{x_k, y_k\}$ стационарное распределение $P_{\text{ст}}(dW)$ сосредоточено на счетном множестве E_a точек.

5.4. Энтропия гауссовых случайных величин

1. Рассмотрим l гауссовых случайных величин $\xi_1, \ldots, \xi_l$, которые описываются вектором средних значений $\mathbf{M}\,\xi_k = m_k$ и невырожденной корреляционной матрицей

$$R_{ij} = \mathbf{M}(\xi_i - m_i)(\xi_j - m_j). \tag{5.4.1}$$

Как известно, такие случайные величины имеют плотность распределения вероятностей

$$p(\xi_1, \ldots, \xi_l) = (2\pi)^{-l/2} \det{}^{-1/2} \|R_{ij}\| \times \\ \times \exp\left\{-\frac{1}{2} \sum_{i,\, j=1}^{l} (\xi_i - m_i)\, a_{ij}\, (\xi_j - m_j)\right\}, \tag{5.4.2}$$

где $\|a_{ij}\| = \|R_{ij}\|^{-1}$ — матрица, обратная корреляционной.

Чтобы определить энтропию указанных величин, введем вспомогательную меру $\nu\,(d\xi_1, \ldots, d\xi_l)$, удовлетворяющую условию мультипликативности (1.7. 9). Для простоты выберем меры $\nu_j\,(d\xi_j)$, которым соответствует одинаковая равномерная плотность распределения

$$\frac{\nu_j(d\xi_j)}{d\xi_j} = \nu_1 = (\nu_0)^{1/l}. \tag{5.4.3}$$

Как будет видно из дальнейшего, удобно полагать

$$\nu_1 = (2\pi e)^{-1/2}. \tag{5.4.4}$$

При этом

$$\nu(d\xi_1, \ldots, d\xi_l) = \nu_1^l\, d\xi_1 \ldots d\xi_l = (2\pi e)^{-l/2}\, d\xi_1 \ldots d\xi_l.$$

Учитывая (5.4.2), можно видеть, что при постоянной плотности

$$\nu(d\xi_1, \ldots, d\xi_l)/d\xi_1 \ldots d\xi_l$$

выполнено условие абсолютной непрерывности меры P относительно меры ν. В самом деле, равенство $\nu(A) = 0$ для некоторого множества A, состоящего из точек l-мерного действительного пространства R_l, означает, что l-мерный объем множества A равен нулю. Но для таких множеств вероятность $P(A)$ также равна нулю. Указанное условие абсолютной непрерывности не было бы выполнено, если бы корреляционная матрица (5.4.1) была вырожденной. Поэтому условие невырожденности является существенным.

При описанном выборе мер случайная энтропия (1.6.14) имеет вид

$$H(\xi_1, \ldots, \xi_l) = \frac{l}{2}\ln 2\pi + l\ln\nu_1 + \frac{1}{2}\ln\det\|R_{ij}\| + \\ + \frac{1}{2}\sum_{i,j}^{l}(\xi_i - m_i)\,a_{ij}\,(\xi_j - m_j). \tag{5.4.5}$$

Чтобы вычислить энтропию (1.6.13) остается произвести усреднение этого выражения. Учитывая (5.4.1), получаем

$$H_{\xi_1, \ldots, \xi_l} = \frac{l}{2}\ln 2\pi + l\ln\nu_1 + \frac{1}{2}\ln\det\|R_{ij}\| + \\ + \frac{1}{2}\sum_{i,j=1}^{l} a_{ij}R_{ij} = \frac{l}{2}\ln 2\pi + l\ln\nu_1 + \\ + \frac{1}{2}\ln\det\|R_{ij}\| + \frac{l}{2}. \tag{5.4.6}$$

При выборе (5.4.4) имеет место простой результат

$$H_{\xi_1 \ldots \xi_l} = \frac{1}{2}\ln\det\|R_{ij}\|. \tag{5.4.6a}$$

Матрица $R = \|R_{ij}\|$ является симметричной. Поэтому, как известно, существует унитарное преобразование U, приводящее эту матрицу к диагональному виду:

$$\sum_{i,j} U_{ir}^* R_{ij} U_{js} = \lambda_r \delta_{rs}.$$

Здесь λ_r — собственные значения корреляционной матрицы, удовлетворяющие уравнению

$$\sum_k R_{jk} U_{kr} = \lambda_r U_{jr}. \tag{5.4.7}$$

При помощи этих собственных значений энтропию $H_{\xi_1,\ldots,\xi_l}$ можно записать в виде

$$H_{\xi_1,\ldots,\xi_l}=\frac{1}{2}\sum_{r=1}^{l}\ln\lambda_r. \tag{5.4.8}$$

Этой формуле можно придать также вид

$$H_{\xi_1\ldots\xi_l}=\frac{1}{2}\operatorname{Sp}\ln\mathrm{R}. \tag{5.4.8a}$$

Полученный результат позволяет, в частности, легко найти условную энтропию $H_{\xi_k|\xi_1\ldots\xi_{k-1}}$. Используя иерархическое свойство энтропии, можно записать

$$H_{\xi_1,\ldots,\xi_k}=H_{\xi_1}+H_{\xi_2|\xi_1}+\ldots+H_{\xi_k|\xi_1,\ldots,\xi_{k-1}},$$
$$H_{\xi_1,\ldots,\xi_{k-1}}=H_{\xi_1}+H_{\xi_2|\xi_1}+\ldots+H_{\xi_{k-1}|\xi_1,\ldots,\xi_{k-1}}$$

и, вычтя одно выражение из другого, получить

$$H_{\xi_k|\xi_1,\ldots,\xi_{k-1}}=H_{\xi_1,\ldots,\xi_k}-H_{\xi_1,\ldots,\xi_{k-1}}.$$

Каждую из энтропий, входящих в правую часть этого равенства, можно вычислить по формуле (5.4.8). Это приводит к соотношению

$$H_{\xi_k|\xi_1,\ldots,\xi_{k-1}}=\frac{1}{2}\operatorname{Sp}\ln\mathrm{r}^{(k)}-\frac{1}{2}\operatorname{Sp}\ln\mathrm{r}^{(k-1)}, \tag{5.4.9}$$

где

$$\mathrm{r}^{(k)}=\begin{pmatrix} R_{11} & . R_{1k}\\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot\\ R_{k1}\ldots & R_{kk}\end{pmatrix},\qquad \mathrm{r}^{(k-1)}=\begin{pmatrix} R_{11} & \ldots & R_{1,k-1}\\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot\\ R_{k-1,1} & \ldots & R_{k-1,k-1}\end{pmatrix}.$$

Аналогичным способом можно вычислить энтропию

$$H_{\xi_k\xi_{k-1}|\xi_1,\ldots,\xi_{k-2}}=H_{\xi_1,\ldots,\xi_k}-H_{\xi_1,\ldots,\xi_{k-2}}\ \text{и т. п.}$$

2. Выберем теперь более сложные меры $\nu_k(d\xi_k)$. Именно, положим, что они имеют плотность распределения $N_k q_k(\xi_k)$, где $q_k(\xi_k)$ — гауссова плотность распределения

$$q_k(\xi_k)=\left(2\pi\widetilde{\lambda}_k\right)^{-1/2}\exp\left\{-\frac{1}{2}\frac{(\xi_k-\widetilde{m}_k)^2}{\widetilde{\lambda}_k}\right\}.$$

Условие мультипликативности (1.7.9) предполагается выполненным. При этом случайная энтропия (1.6.14) оказывается равной

$$H(\xi_1,\ldots,\xi_l)=\ln N+\frac{1}{2}\ln\det\|R_{ij}\|-\frac{1}{2}\sum_k\ln\widetilde{\lambda}_k-$$
$$-\frac{1}{2}\sum_k\frac{(\xi_k-\widetilde{m}_k)^2}{\widetilde{\lambda}_k}+\frac{1}{2}\sum_{i,j}(\xi_i-m_i)\,a_{ij}\,(\xi_j-m_j), \tag{5.4.10}$$

где $N=\Pi N_k$.

Ее усреднение приводит теперь к соотношению

$$H_{\xi_1, \ldots, \xi_l} = \ln N + \frac{1}{2} \ln \det \| R_{ij} \| - \frac{l}{2} \sum_k \ln \widetilde{\lambda}_k -$$

$$- \frac{1}{2} \sum_k \frac{1}{\widetilde{\lambda}_k} \left[R_{kk} + (m_k - \widetilde{m}_k)^2 \right] + \frac{l}{2}. \tag{5.4.11}$$

Вводя матрицу $\widetilde{\mathrm{R}} = \|\widetilde{\lambda}_k \delta_{kr}\|$ и используя матричную форму записи, равенству (5.4.11) можно придать вид

$$H_{\xi_1, \ldots, \xi_l} = \ln N + \frac{1}{2} \operatorname{Sp} \ln (\widetilde{\mathrm{R}}^{-1} \mathrm{R}) - \frac{1}{2} \operatorname{Sp} \widetilde{\mathrm{R}}^{-1} (\mathrm{R} - \widetilde{\mathrm{R}}) -$$

$$- \frac{1}{2} (\mathrm{m} - \widetilde{\mathrm{m}})^T \widetilde{\mathrm{R}}^{-1} (\mathrm{m} - \widetilde{\mathrm{m}}). \tag{5.4.12}$$

Здесь мы учли, что $\det \mathrm{R} - \det \widetilde{\mathrm{R}} = \det \widetilde{\mathrm{R}}^{-1} \mathrm{R}$; $\operatorname{Sp} \widetilde{\mathrm{R}}^{-1} \widetilde{\mathrm{R}} = \operatorname{Sp} 1 = l$; $\mathrm{m} - \widetilde{\mathrm{m}}$ — матрица-столбец; T означает транспонирование. Сравнивая (5.4.12) с (1.6.16), нетрудно видеть, что мы нашли тем самым энтропию $H^{P/Q}_{\xi_1, \ldots, \xi_l}$ распределения P по распределению Q, которая оказалась равной

$$H^{P/Q}_{\xi_1, \ldots, \xi_l} = \operatorname{Sp} G (\widetilde{\mathrm{R}}^{-1} \mathrm{R}) + \frac{1}{2} (\mathrm{m} - \widetilde{\mathrm{m}})^T \widetilde{\mathrm{R}}^{-1} (\mathrm{m} - \widetilde{\mathrm{m}}), \tag{5.4.13}$$

где $G(x) = (x - 1 - \ln x)/2$.

Последняя формула получена в предположении мультипликативности (1.7.9) меры $\nu(d\xi_1, \ldots, d\xi_l)$, а, следовательно, и $Q(d\xi_1, \ldots, d\xi_l)$. Однако ее нетрудно распространить и на более общий случай. Пусть мера Q является гауссовой и определяется вектором $\widetilde{m}_k = \mathbf{M}_Q \xi_k$ и корреляционной матрицей $\widetilde{R}_{kr}$, необязательно диагональной. Энтропия $H^{P/Q}$ является инвариантной относительно ортогональных преобразований (и вообще не вырожденных линейных преобразований) l-мерного действительного пространства R_l. Совершив преобразование поворота, можно добиться того, чтобы матрица $\widetilde{\mathrm{R}}$ стала диагональной, и после этого применить формулу (5.4.13). Но формула (5.4.13) уже является инвариантной относительно линейных невырожденных преобразований и поэтому справедлива не только в случае диагональной, но и недиагональной матрицы $\widetilde{\mathrm{R}}$. В самом деле, при линейном преобразовании $\xi'_k = \sum_r C_{kr} \xi_r$, т. е. $\xi' = \mathrm{C}\xi$, имеют место преобразования $\mathrm{m}' - \widetilde{\mathrm{m}}' = \mathrm{C} (\mathrm{m} - \widetilde{\mathrm{m}})$, $\mathrm{R}' = \mathrm{CRC}^T$, $\widetilde{\mathrm{R}}' = \mathrm{C}\widetilde{\mathrm{R}}\mathrm{C}^T$ и, следовательно, $(\widetilde{\mathrm{R}}^{-1})' = (\mathrm{C}^T)^{-1} \widetilde{\mathrm{R}}^{-1} \mathrm{C}^{-1}$, $(\widetilde{\mathrm{R}}^{-1}\mathrm{R})' = \mathrm{C}^{T^{-1}} \widetilde{\mathrm{R}}^{-1} \mathrm{R} \mathrm{C}^T$. Поэтому комбинации $\operatorname{Sp} f (\widetilde{\mathrm{R}}^{-1}\mathrm{R})$ и $(\mathrm{m} - \widetilde{\mathrm{m}})^T \widetilde{\mathrm{R}}^{-1} (\mathrm{m} - \widetilde{\mathrm{m}})$ остаются инвариантными. Это доказывает, что формула (5.4.13) сохраняет свое значение не только в мультипликативном случае (1.7.9), но и в более общем случае, когда формулы (5.4.8), (5.4.9), (5.4.10) могут быть несправедливы.

3. В заключение этого параграфа вычислим дисперсию энтропии, которую полезно знать при исследовании вопроса об энтропийной устойчивости (см. § 1.5) семейства гауссовых случайных величин.

Начнем со случайной энтропии (5.4.5). Вычитая (5.4.6), находим случайное отклонение

$$H(\xi_1, \ldots, \xi_l) - H_{\xi_1, \ldots, \xi_l} = \frac{1}{2} \sum a_{jk} (\eta_j \eta_k - R_{jk}) =$$
$$= \frac{1}{2} \eta^T \mathrm{a} \eta - \frac{1}{2} \operatorname{Sp} \mathbf{1} = \frac{1}{2} \eta^T \mathrm{a} \eta - \frac{l}{2}, \tag{5.4.14}$$

средний квадрат которого совпадает с искомой дисперсией. Мы обозначили $\eta_j = \xi_j - m_j$. При усреднении квадрата этого выражения нужно принять во внимание, что

$$\mathbf{M} \eta_j \eta_k \eta_r \eta_s = R_{jk} R_{rs} + R_{jr} R_{ks} + R_{js} R_{kr} \tag{5.4.15}$$

согласно известным свойствам гауссовых величин. Поэтому будем иметь

$$\mathbf{D} H(\xi_1, \ldots, \xi_l) = \mathbf{M} \left[H(\xi_1, \ldots, \xi_l) - H_{\xi_1, \ldots, \xi_l} \right]^2 =$$
$$= \frac{1}{4} \sum a_{jk} a_{rs} (R_{jr} R_{ks} + R_{js} R_{kr}) = \frac{1}{2} \operatorname{Sp} \mathrm{a R a R}, \tag{5.4.16}$$

т. е.

$$\mathbf{D} H(\xi_1, \ldots, \xi_l) = \frac{1}{2} \operatorname{Sp} \mathbf{1} = \frac{l}{2}.$$

Переходя к случайной энтропии (5.4.10), имеем

$$H(\xi_1, \ldots, \xi_l) - H_{\xi_1, \ldots, \xi_l} = \frac{1}{2} \eta^T (\mathrm{a} - \widetilde{\mathrm{R}}^{-1}) \eta -$$
$$- \frac{1}{2} \operatorname{Sp} (\mathrm{a} - \widetilde{R}^{-1}) \mathrm{R} - \frac{1}{2} (\mathrm{m} - \widetilde{\mathrm{m}})^T \widetilde{\mathrm{R}}^{-1} \eta.$$

Используя, в свою очередь, (5.4.15), получаем

$$\mathbf{D} H(\xi_1, \ldots, \xi_l) = \frac{1}{2} \operatorname{Sp} \left[(\mathrm{a} - \widetilde{\mathrm{R}}^{-1}) \mathrm{R} \right]^2 + \frac{1}{4} (\mathrm{m} - \widetilde{\mathrm{m}})^T \times$$
$$\times \widetilde{\mathrm{R}}^{-1} \mathrm{R} \widetilde{\mathrm{R}}^{-1} (\mathrm{m} - \widetilde{\mathrm{m}}) = \frac{1}{2} \operatorname{Sp} (1 - \widetilde{\mathrm{R}}^{-1} \mathrm{R})^2 +$$
$$+ \frac{1}{4} (\mathrm{m} - \widetilde{\mathrm{m}})^T \widetilde{\mathrm{R}}^{-1} \mathrm{R} \widetilde{\mathrm{R}}^{-1} (\mathrm{m} - \widetilde{\mathrm{m}}). \tag{5.4.17}$$

Не представляет особого труда вычислить и другие статистические характеристики случайной энтропии гауссовых переменных, в частности, ее характеристический потенциал

$$\mu_0(s) = \ln \mathbf{M} e^{sH(\xi_1, \ldots, \xi_l)}$$

(см. (1.5.15), (4.1.18)). Так, подставляя сюда (5.4.5) и принимая во внимание вид (5.4.2) плотности распределения вероятностей, получаем

$$\mu_0(s) = \ln\left\{(2\pi)^{-l/2} (\det{}^{-1/2} R) \exp\left[sH_{\xi_1,\ldots,\xi_l} - \frac{sl}{2}\right] \times \right.$$
$$\left. \times \int \exp\left\{-\frac{1-s}{2} \sum_{i,j} \eta_i a_{ij} \eta_j\right\} d\eta_1, \ldots, d\eta_l\right\} =$$
$$= sH_{\xi_1,\ldots\xi_l} - \frac{sl}{2} + \ln \det{}^{-1/2}\left[(1-s)\,a\right] +$$
$$+ \ln \det{}^{-1/2} R \quad (\eta_i = \xi_i - m_i). \tag{5.4.18}$$

Здесь использована формула

$$\int \exp\left[-\frac{1}{2} \sum_{i,j} \eta_i d_{ij} \eta_j\right] d\eta_1, \ldots, d\eta_l = \det{}^{-1/2} \left\| \frac{d_{ij}}{2\pi} \right\|, \tag{5.4.19}$$

справедливая при любой невырожденной положительно определенной матрице $\|d_{ij}\|$.

Поскольку $a = R^{-1}$, то в (5.3.18) члены с $\ln \det^{-1/2} R$ сокращаются, и мы получаем

$$\mu_0(s) = -\frac{l}{2} \ln(1-s) - \frac{sl}{2} + sH_{\xi_1,\ldots,\xi_l},$$

что справедливо при $s < 1$. Отсюда можно получить, в частности, формулу (5.4.16).

5.5. Энтропия стационарной последовательности. Гауссова последовательность

1. В § 5.1 была рассмотрена энтропия отрезка стационарного процесса $\{\xi_k\}$ в дискретном времени, т. е. стационарной последовательности. При этом предполагалось, что каждый элемент последовательности представляет собой дискретную случайную величину. Данное в § 1.6 обобщение понятия энтропии позволяет рассмотреть энтропию стационарной последовательности, составленную из произвольных, в том числе непрерывных случайных величин, обобщая тем самым результаты § 5.1.

Если вспомогательная мера ν удовлетворяет условию мультипликативности (1.7.9), то, как показано выше, условные энтропии в обобщенной версии обладают всеми свойствами, которыми обладали условные энтропии в дискретной версии. Указанные свойства (и по существу только они) были существенно использованы при изложении материала в § 5.1. Поэтому все сказанное в § 5.1 может быть отнесено к произвольным случайным величинам, к энтропии в обобщенной версии.

Мера ν предполагается мультипликативной

$$\nu(d\xi_{k_1}\dots d\xi_{k_r})=\prod_{i=1}^{r}\nu_{k_i}(d\xi_{k_i}), \tag{5.5.1}$$

причем из соображений стационарности «элементарные» меры ν_k предполагаются одинаковыми. Предполагается также, что выполнено условие абсолютной непрерывности вероятности $P(d\xi_k)$ относительно $\nu_k(d\xi_k)$. Процесс $\{\xi_k\}$ является стационарным по отношению к распределению P, т. е. выполнено условие типа (5.1.1) при любых $k_1, \dots, k_r, a$.

Формулой (5.1.3) вводится удельная энтропия H_1. Теперь под $H_{\xi_k \mid \xi_{k-l}, \dots, \xi_{k-1}}$, однако, следует понимать энтропию (1.7.13), так что указанное определение соответствует формуле

$$H_1=-\int \ln\frac{P(d\xi_k\mid\xi_{k-1},\xi_{k-2},\dots)}{\nu_k(d\xi_k)}P(d\xi_k\mid\xi_{k-1},\xi_{k-2},\dots). \tag{5.5.2}$$

В обобщенной версии также справедлива теорема 5.1, которая доказывается точно так же, как и раньше. Теперь она означает равенство

$$H_1=-\lim_{l\to\infty}\frac{1}{l}\int\dots\int\ln\frac{P(d\xi_1\dots d\xi_l)}{\nu_1(d\xi_1)\dots\nu_l(d\xi_l)}P(d\xi_1\dots d\xi_l). \tag{5.5.3}$$

Далее, по аналогии с (5.1.11), (5.1.12) может быть введена величина

$$2\Gamma=\lim_{n\to\infty}\left[H_{\xi_1\dots\xi_n}-nH_1\right]=\sum_{j=0}^{\infty}\left[H_{\xi_{j+1}\mid\xi_j\dots\xi_1}-H_1\right], \tag{5.5.4}$$

которая неотрицательна, поскольку

$$H_{\xi_{j+1}\mid\xi_j\dots\xi_1}\geqslant H_{\xi_{j+1}\mid\xi_j\xi_{j-1}\dots}=H_1. \tag{5.5.5}$$

Она может быть интерпретирована как энтропия концов рассматриваемого отрезка последовательности.

Кроме приведенных выше величин и соотношений, основанных на определении энтропии (1.6.13), могут быть рассмотрены аналогичные величины и соотношения, основанные на определении 1.6.17). Именно, по аналогии с (5.5.2), (5.5.4) можно ввести

$$H_1^{P/Q}=H^{P/Q}_{\xi_k\mid\xi_{k-1}\xi_{k-2}\dots}=\int\ln\frac{P(d\xi_k\mid\xi_{k-1},\xi_{k-2},\dots)}{Q(d\xi_k)}\times$$

$$\times P(d\xi_k\mid\xi_{k-1},\xi_{k-2},\dots), \tag{5.5.6}$$

$$2\Gamma^{P/Q}=\sum_{j=0}^{\infty}\left[H^{P/Q}_{\xi_{j+1}\mid\xi_j\dots\xi_1}-H_1^{P/Q}\right]. \tag{5.5.7}$$

Для энтропии $H^{P/Q}$, однако, согласно (1.7.19) имеет место неравенство

$$H^{P/Q}_{\xi_{j+1}\mid\xi_j\dots\xi_i}\leqslant H^{P/Q}_{\xi_{j+1}\mid\xi_j\dots\xi_1},\dots$$

обратное неравенству (5.5.5). Поэтому «этропия конца» $\Gamma^{P/Q}$ обязана быть неположительной.

Подставляя выражения типа (1.7.13), (5.5.2) для условных энтропий, нетрудно убедиться, что разность $H_{\xi_{j+1}\mid\xi_1,\dots,\xi_j}-H_1$, а следовательно, и граничная энтропия Γ оказывается не зависящей от ν_k. Аналогично $\Gamma^{P/Q}$ оказывается не зависящей от Q (при выполнении условия мультипликативности), причем справедливо тождество $\Gamma^{P/Q}=-\Gamma$. Это полезно иметь в виду при записи формулы (5.1.13) для обеих энтропий, которая принимает вид

$$H_{\xi_1\dots\xi_l}=lH_1+2\Gamma+o_l(1),\qquad H^{P/Q}_{\xi_1\dots\xi_l}=lH_1^{P/Q}-2\Gamma+o_l(1). \tag{5.5.7a}$$

Последние соотношения позволяют найти этропию отрезка стационарного процесса точнее, чем простым умножением удельной энтропии H_1 на его длину l.

2. Найдем удельную энтропию H_1 для случая стационарной гауссовой последовательности.

а) Предположим сначала, что задана стационарная последовательность $\xi_1,\dots,\xi_l$ на круге, распределение которой инвариантно относительно вращений круга. При этом элементы корреляционной матрицы R_{ij} будут зависеть лишь от разности $j-k$ и удовлетворять условию периодичности: $R_{jk}=R_{j-k}$, $R_{j-k+l}=R_{j-k}$. Уравнение (5.4.7) при этом будет иметь решения:

$$U_{jr}=\frac{1}{\sqrt{l}}\,e^{2\pi ijr/l}, \tag{5.5.8}$$

$$\lambda_r=\sum_{s=0}^{l-1}R_s\,e^{-2\pi isr/l},\qquad r=0,1,\dots,l-1. \tag{5.5.9}$$

В самом деле, подставляя (5.5.8) в (5.4.7), получаем равенство

$$\sum_k R_{j-k}\,e^{2\pi ikr/l}=\lambda_r\,e^{2\pi ijr/l},$$

которое удовлетворяется в силу (5.5.9). Итак (5.5.8) определяет преобразование, которое диагонализует корреляционную матрицу R_{j-k}. Легко проверить его унитарность. В самом деле, эрмитово-сопряженный оператор

$$U^+=\|U^*_{jr}\|=\left\|\frac{1}{\sqrt{l}}\,e^{-2\pi ijr/l}\right\|$$

вследствие равенств

$$\sum_r U_{jr}U^*_{kr}=\frac{1}{l}\sum_{r=1}^{l}e^{2\pi i\frac{(j-k)r}{l}}=\frac{\varepsilon}{l}\sum_{r=0}^{l-1}\varepsilon^r=\frac{\varepsilon}{l}\,\frac{1-\varepsilon^l}{1-\varepsilon}=\delta_{jk}$$

$$(\varepsilon=e^{2\pi i(j-k)/l})$$

совпадает с обратным оператором U^{-1}.

После вычисления собственных значений (5.5.9) для получения энтропии $H_{\xi_1,\dots,\xi_l}$ остается воспользоваться формулой (5.4.8).

В рассматриваемом случае инвариантности относительно поворотов легко вычислить также энтропию (5.4.13). Предполагается, конечно, что описанной симметрией («круговой стационарностью») обладает не только мера P, но и мера Q. Корреляционная матрица $\widetilde{R}_{jk}$ последней обладает, следовательно, теми же свойствами, что и R_{jk}. Кроме того, для обеих мер средние значения m_k, $\widetilde{m}_k$ постоянны (соответственно равны m и $\widetilde{m}$).

Унитарное преобразование $\mathrm{U} = \| U_{jr} \|$ диагонализирует не только матрицу R, но и матрицу $\widetilde{R}$ (даже, если условие мультипликативности не выполнено), причем по аналогии с (5.5.9) ее средние значения имеют вид

$$\widetilde{\lambda}_r = \sum_{s=0}^{l-1} \widetilde{R}_s e^{-2\pi i s r/l}, \quad r = 1,\dots, l-1. \tag{5.5.10}$$

Вектор $\mathrm{m} - \widetilde{\mathrm{m}}$ при этом преобразовании переходит в вектор

$$\mathrm{U}^+(\mathrm{m} - \widetilde{\mathrm{m}}) = \frac{1}{\sqrt{l}} \begin{pmatrix} \sum_{j=1}^{l} e^{-2\pi i 0 j/l}(m_j - \widetilde{m}_j) \\ \dots\dots\dots\dots \\ \sum_{j=1}^{l} e^{-2\pi i (l-1) j/l}(m_j - \widetilde{m}_j) \end{pmatrix} =$$

$$= \frac{m - \widetilde{m}}{\sqrt{l}} \begin{pmatrix} \sum_{j=1}^{l} e^{-2\pi i 0 j/l} \\ \vdots \\ \sum_{j=1}^{l} e^{-2\pi i (l-1) j/l} \end{pmatrix} = \begin{pmatrix} \sqrt{l}\,(m - \widetilde{m}) \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

В результате в соответствии с формулой (5.4.13) будем иметь

$$H^{P/Q}_{\xi_1\dots\xi_l} = \sum_{r=0}^{l-1} G\left(\frac{\lambda_r}{\widetilde{\lambda}_r}\right) + \frac{l}{2} \frac{(m - \widetilde{m})^2}{\widetilde{\lambda}_0}, \tag{5.5.11}$$

где $G(x) = \frac{1}{2}(x - 1 - \ln x)$.

Для энтропии, отнесенной к одному элементу последовательности, согласно (5.4.8), (5.5.11) получаем

$$H_1 = \frac{1}{2l} \sum_{r=0}^{l-1} \ln \lambda_r, \tag{5.5.12}$$

$$H_1^{P/Q} = \frac{1}{l} \sum_{r=0}^{l-1} \ln\left(\frac{\lambda_r}{\widetilde{\lambda}_r}\right) + \frac{1}{2} \frac{(m - \widetilde{m})^2}{\widetilde{\lambda}_0}. \tag{5.5.13}$$

б) Пусть теперь $\xi_1, \ldots, \xi_l$ составляют отрезок стационарной последовательности, так что $R_{jk} = R_{j-k}$, $j, k = 1, \ldots, l$, но условие периодичности $R_{j-k+l} = R_{j-k}$ не выполняется. Тогда каждый из концов отрезка играет некоторую роль и дает свой вклад Γ в суммарную энтропию (5.5.7а). Если этим вкладом пренебречь, то можно соединить концы отрезка и перейти к разобранному ранее случаю круговой симметрии. При этом нужно образовать новую корреляционную функцию

$$\overline{R}_{j,\, j+s} = \overline{R}_s = \sum_{n=-\infty}^{\infty} R_{s+nl}, \tag{5.5.14}$$

которая уже обладает свойством периодичности. Если R_s заметно отличны от нуля лишь при $\| s \| \ll l$, то добавочные члены в (5.5.14) заметно повлияют лишь на небольшое число элементов корреляционной матрицы $\overline{R}_{jk}$, стоящих в углах, где $j \ll l$, $l-k \ll l$ или $l-j \ll l$, $k \ll l$.

После перехода к корреляционной матрице (5.5.14) (и, если нужно, после аналогичного перехода для второй матрицы $\widetilde{R}$) можно воспользоваться полученными ранее формулами (5.5.12), (5.5.13), (5.5.9), (5.5.10).

Учитывая (5.5.9), для собственных значений теперь будем иметь

$$\lambda_r = \sum_{n=-\infty}^{\infty} \sum_{s=0}^{l-1} R_{s+nl}\, e^{-2\pi i s r/l} = \sum_{\sigma=-\infty}^{\infty} R_\sigma\, e^{-2\pi i \sigma r/l},$$

или $\lambda_r = \varphi\ (r/l)$, если ввести обозначение

$$\varphi(\mu) = \sum_{\sigma=-\infty}^{\infty} R_\sigma\, e^{-2\pi i \mu\sigma}. \tag{5.5.15}$$

Аналогично

$$\widetilde{\lambda}_r = \widetilde{\varphi}\left(\frac{r}{l}\right) = \sum_{\sigma=-\infty}^{\infty} e^{-2\pi i \sigma r/l}\, \widetilde{R}_\sigma.$$

Формулы (5.5.12), (5.5.13), очевидно, примут вид

$$\overline{H}_1 = \frac{1}{2l} \sum_{r=0}^{l-1} \ln \varphi\left(\frac{r}{l}\right), \tag{5.5.16}$$

$$\overline{H}_1^{P/Q} = \frac{1}{l} \sum_{r=0}^{l-1} G\left(\frac{\varphi(r/l)}{\widetilde{\varphi}(r/l)}\right) + \frac{1}{2}\,\frac{(m-\widetilde{m})^2}{\widetilde{\varphi}(0)}.$$

Будем теперь увеличивать длину l выбранного отрезка последовательности. Совершая в последних формулах предельный переход $l \to \infty$, получаем вместо сумм интегралы

$$H_1 = \frac{1}{2} \int_0^1 \ln \varphi(\mu)\, d\mu = \frac{1}{2} \int_{-1/2}^{1/2} \ln \varphi(\mu)\, d\mu, \tag{5.5.17}$$

$$H_1^{P/Q}=\int_0^1 G\left(\frac{\varphi(\mu)}{\widetilde{\varphi}(\mu)}\right)d\mu+\frac{1}{2}\frac{(m-\widetilde{m})^2}{\widetilde{\varphi}(0)}=$$
$$=\int_{-1/2}^{1/2} G\left(\frac{\varphi(\mu)}{\widetilde{\varphi}(\mu)}\right)d\mu+\frac{(m-\widetilde{m})^2}{2\widetilde{\varphi}(0)}. \tag{5.5.18}$$

Здесь при изменении пределов интегрирования учтено свойство $\varphi(\mu+1)=\varphi(\mu)$, вытекающее из (5.5.15).

При больших l концы отрезка дают относительно малый вклад по сравнению с большой полной энтропией, имеющей порядок lH_1. Переход (5.5.14) к корреляционной функции $\overline{R}_s$ изменяет энтропию $H_{\xi_1,\ldots,\xi_l}$ на некоторое число, не возрастающее с ростом l. Поэтому имеет место совпадение пределов

$$\lim_{l\to\infty}\frac{1}{l}\overline{H}_{\xi_1,\ldots,\xi_l}=\lim_{l\to\infty}\frac{1}{l}H_{\xi_1,\ldots\xi_l}.$$

Здесь $\overline{H}$ соответствует корреляционной функции $\overline{R}_s$, а H — функции R_s. Следовательно, выражения (5.5.17), (5.5.18) совпадают с определенными ранее удельными энтропиями (5.5.3), (5.5.6). Тем самым мы вычислили удельные энтропии для случая стационарных гауссовых последовательностей. Они оказались выраженными через спектральные плотности $\varphi(\mu)$, $\widetilde{\varphi}(\mu)$.

Условие абсолютной непрерывности меры P относительно меры Q принимает вид условия интегрируемости функции $G(\varphi(\mu)/\widetilde{\varphi}(\mu))$, входящей в (5.5.18).

Формула (5.5.18) справедлива не только при выполнении условия мультипликативности (5.5.1). Для стационарного гауссового случая это условие означает, что матрица $\widetilde{R}_{jk}$ кратна единичной: $\widetilde{R}_{jk}=\widetilde{\varphi}\delta_{jk}$, $\widetilde{\varphi}(\mu)=\widetilde{\varphi}=\text{const}$. Тогда формула (5.5.18) дает

$$H_1^{P/Q}=\int_0^1 G\left(\frac{\varphi(\mu)}{\widetilde{\varphi}}\right)d\mu+\frac{1}{2}\frac{(m-\widetilde{m})^2}{\widetilde{\varphi}}.$$

Приведенные результаты обобщаются и на тот случай, когда имеется не одна случайная последовательность $\{\ldots,\xi_1,\xi_2,\ldots\}$, а несколько ($r$) стационарных и стационарно связанных последовательностей $\{\ldots,\xi_1^\alpha,\xi_2^\alpha,\ldots\}$, $\alpha=1,\ldots,r$, описываемых корреляционной матрицей

$$\| R_{j-k}^{\alpha,\beta}\|,\qquad \alpha,\beta=1,\ldots,r$$

или матрицей спектральных функций $\varphi(\mu)=\|\varphi^{\alpha\beta}(\mu)\|$, $\varphi^{\alpha\beta}(\mu)=\sum_{\sigma=-\infty}^{\infty}R_\sigma^{\alpha\beta}e^{-2\pi i\mu\sigma}$ и вектором-столбцом (по индексу α) средних значений $\mathrm{m}=\|m^\alpha\|$.

При этом формула (5.5.17) заменяется на матричное обобщение

$$H_1=\frac{1}{2}\int_{-1/2}^{1/2}\mathrm{Sp}\left[\ln\varphi(\mu)\right]d\mu=\frac{1}{2}\int_{-1/2}^{1/2}\ln\left[\det\varphi(\mu)\right]d\mu. \quad (5.5.19)$$

Мера Q пусть описывается матрицей спектральных функций $\widetilde{\varphi}\ (\mu)=\|\varphi^{\alpha\beta}(\mu)\|$ и средних значений $\widetilde{\mathrm{m}}=\|\widetilde{m}^{\alpha}\|$. Тогда, вместо (5.5.18), будем иметь аналогичную матричную формулу

$$H_1^{P/Q}=\int_{-1/2}^{1/2}\mathrm{Sp}\,G\left(\widetilde{\varphi}^{-1}(\mu)\varphi(\mu)\right)d\mu+\frac{1}{2}(\mathrm{m}-\widetilde{\mathrm{m}})^T\widetilde{\varphi}^{-1}(0)(\mathrm{m}-\widetilde{\mathrm{m}}). \quad (5.5.20)$$

Приведенные результаты, разумеется, вытекают из формул (5.4.6а), (5.4.13). По виду же они представляют собой синтез формул (5.4.6а), (5.4.13) и (5.5.17), (5.5.18).

3. Полученные результаты позволяют сделать заключение об энтропийной устойчивости (см. § 1.5) семейства случайных величин $\{\xi^l\}$, где $\xi^l=\{\xi_1,\ldots,\xi_l\}$ — отрезок стационарной гауссовой последовательности. Энтропия $H_{\xi_1,\ldots,\xi_l}$ растет с ростом l приблизительно линейно. Дисперсия энтропии согласно (5.4.16) также растет линейно. Поэтому отношение $\mathbf{D}\,H\,(\xi_1,\ldots,\xi_l)/H^2_{\xi_1,\ldots,\xi_l}$ стремится к нулю, так что условие (1.5.8) энтропийной устойчивости для энтропии (5.4.5) оказывается выполненным.

Перейдем к энтропии (5.4.10). Условия

$$\frac{\mathbf{D}H(\xi_1,\ldots,\xi_l)}{H^2_{\xi_1,\ldots\,\xi_l}}\to 0,\quad \frac{\mathbf{D}H^{P/Q}(\xi_1,\ldots,\xi_l)}{\left(H^{P/Q}_{\xi_1,\ldots,\xi_l}\right)^2}\to 0$$

для нее будут выполнены, если дисперсия $\mathbf{D}H\ (\xi_1,\ \ldots,\ \xi_l)=\mathbf{D}H^{P/Q}\ (\xi_1,\ldots,\xi_l)$, определяемая формулой (5.4.17), возрастает с ростом l приблизительно линейно, т. е. если существует конечный предел

$$D_1=\lim_{l\to\infty}\frac{1}{l}\mathbf{D}H^{P/Q}(\xi_1,\ldots,\ \xi_l). \quad (5.5.21)$$

Для вычисления предела (5.5.21) выражения (5.4.17) можно применить те же методы, какие были использованы для вычисления предела $\lim\limits_{l\to\infty}\frac{1}{l}\,H^{P/Q}_{\xi_1,\ldots,\xi_l}$, соответствующего выражению (5.4.13). Подобно тому, как из (5.4.13) мы получили (5.5.18), из формулы (5.4.17) находим предел (5.5.21)

$$D_1=\frac{1}{2}\int_{-1/2}^{1/2}\left[1-\frac{\varphi(\mu)}{\widetilde{\varphi}(\mu)}\right]^2d\mu+\frac{(m-\widetilde{m})^2\varphi(0)}{4\widetilde{\varphi}^2(0)}. \quad (5.5.22)$$

5.6. Энтропия случайных процессов в непрерывном времени. Общие понятия и соотношения

1. Обобщенное определение энтропии, данное в гл. 6, позволяет вычислять энтропию случайных процессов $\{\xi_t\}$, зависящих от непрерывного параметра t (времени). Будем предполагать, что процесс ξ_t задан на некотором интервале $a \leqslant t \leqslant b$. Для произвольного подынтервала $\alpha \leqslant t \leqslant \beta$, принадлежащего интервалу определения процесса, будем пользоваться обозначением $\xi_\alpha^\beta = \{\xi_t, \alpha \leqslant t \leqslant \beta\}$. Следовательно, ξ_α^β обозначает совокупность значений процесса $\{\xi_t\}$ на подынтервале $[\alpha, \beta]$.

Исходный процесс $\{\xi_t\}$ описывается вероятностной мерой P. В соответствии с определением энтропии, данным в § 1.6, чтобы определить энтропию $H_{\xi_\alpha^\beta}$ для любых различных интервалов $[\alpha, \beta]$, нужно ввести вспомогательную ненормированную меру ν или соответствующую ей вероятностную меру Q. Мера ν (или Q) должна быть определена на том же измеримом пространстве, т. е. на том же поле событий, касающихся поведения процесса $\xi(t)$ на всем интервале $[a, b]$. При этом процесс $\{\xi(t)\}$, имеющий вероятности Q, можно интерпретировать как новый вспомогательный случайный процесс $\{\eta(t)\}$, отличающийся от исходного процесса $\{\xi(t)\}$.

Мера P должна быть абсолютно непрерывна относительно меры Q (или ν) для всего поля событий, относящихся к поведению процесса $\xi(t)$ на всем интервале определения $[a, b]$. Следовательно, условие абсолютной непрерывности будет выполнено и для любого его подынтервала $[\alpha, \beta]$.

Применяя формулу (1.6.17) для значений случайного процесса на некотором выбранном подынтервале $[\alpha, \beta]$, получаем следующее определение энтропии для этого подынтервала:

$$H^{P/Q}_{\xi_\alpha^\beta} = \int \ln \frac{P(d\xi_\alpha^\beta)}{Q(d\xi_\alpha^\beta)} P(d\xi_\alpha^\beta). \tag{5.6.1}$$

Далее в соответствии с содержанием § 1.7 (см. (1.7.17)) можно ввести условную энтропию

$$H^{P/Q}_{\xi_\alpha^\beta \mid \xi_\gamma^\delta} = \int \ln \frac{P(d\xi_\alpha^\beta \mid \xi_\gamma^\delta)}{Q(d\xi_\alpha^\beta)} P(d\xi_\alpha^\beta\, d\xi_\gamma^\delta), \tag{5.6.2}$$

где $[\gamma, \delta]$ — другой подынтервал, не перекрывающийся с $[\alpha, \beta]$.

Введенные энтропии подчиняются обычным соотношениям, встречающимся в дискретной версии, например, соотношению аддитивности

$$H^{P/Q}_{\xi_\alpha^\delta} = H^{P/Q}_{\xi_\alpha^\beta} + H^{P/Q}_{\xi_\beta^\delta \mid \xi_\alpha^\beta}, \quad \alpha < \beta < \delta. \tag{5.6.3}$$

При записи формул (5.6.2), (5.6.3) предполагается, что меры Q, ν удовлетворяют условию мультипликативности

$$Q\left(d\xi_\alpha^\beta\, d\xi_\gamma^\delta\right)=Q\left(d\xi_\alpha^\beta\right)Q\left(d\xi_\gamma^\delta\right),\ \nu\left(d\xi_\alpha^\beta\, d\xi_\gamma^\delta\right)=\nu_1\left(d\xi_\alpha^\beta\right)\nu_2\left(d\xi_\gamma^\delta\right) \quad (5.6.4)$$

($[\alpha,\ \beta]$ не перекрывается с $[\gamma,\ \delta]$), которое аналогично (1.7.8). Приведенное условие мультипликативности для меры Q означает, что вспомогательный процесс $\{\eta_t\}$ таков, что его значения ξ_α^β и ξ_γ^δ для неперекрывающихся интервалов $[\alpha,\ \beta]$, $[\gamma,\ \delta]$ должны быть независимыми. Условие мультипликативности для меры ν означает кроме того, что постоянные

$$N_\alpha^\beta=\int \nu\left(d\xi_\alpha^\beta\right)$$

определяются некоторой возрастающей функцией $F\ (t)$ по формуле

$$N_\alpha^\beta=e^{F\,(\beta)-F\,(\alpha)}. \quad (5.6.5)$$

В случае стационарного процесса функция $F\ (t)$ линейна, так что

$$F(\beta)-F(\alpha)=(\beta-\alpha)\,h_\nu, \quad N_\alpha^\beta=e^{(\beta-\alpha)\,h_\nu}.$$

Учитывая (5.6.5), безусловную энтропию типа (1.6.16) и условную энтропию типа (1.7.4) можно определить по формулам

$$H_{\xi_\alpha^\beta}=F(\beta)-F(\alpha)-H_{\xi_\alpha^\beta}^{P/Q}, \quad (5.6.6)$$

$$H_{\xi_\alpha^\beta \mid \xi_\gamma^\delta}=F(\beta)-F(\alpha)-H_{\xi_\alpha^\beta \mid \xi_\gamma^\delta}^{P/Q}, \quad (5.6.7)$$

где стоящие справа величины определены соотношениями (5.6.1), (5.6.2).

2. В дальнейшем будем рассматривать стационарный случайный процесс $\{\xi_t\}$, определенный при всех t. В этом случае вспомогательный процесс $\{\eta_t\}$ естественно выбирать также стационарным.

Ввиду того, что энтропия в обобщенной версии при выполнении условия мультипликативности обладает теми же свойствами, что и энтропия в дискретной версии, на рассматриваемый случай непрерывного времени могут быть перенесены рассуждения и результаты, изложенные в § 5.1, 5.5 применительно к стационарному процессу в дискретном времени.

Вследствие обычных общих свойств энтропии условная энтропия $H_{\xi_0^\tau \mid \xi_{-\sigma}^0}$ $(\sigma > 0)$ монотонно не возрастает с ростом σ. Отсюда вытекает существование предела

$$\lim_{\sigma\to\infty} H_{\xi_0^\tau \mid \xi_{-\sigma}^0}=H_{\xi_0^\tau \mid \xi_{-\infty}^0}, \quad (5.6.8)$$

который мы определяем как $H_{\xi_0^\tau \mid \xi_{-\infty}^0}$.

Поскольку в силу общего свойства (1.7.3)

$$H_{\xi_0^{\tau_1+\tau_2} \mid \xi_{-\sigma}^0} = H_{\xi_0^{\tau_1} \mid \xi_{-\sigma}^0} + H_{\xi_{\tau_1}^{\tau_1+\tau_2} \mid \xi_{-\sigma}^{\tau_1}}, \tag{5.6.9}$$

а в силу условия стационарности

$$H_{\xi_{\tau_1}^{\tau_1+\tau_2} \mid \xi_{-\sigma}^{\tau_1}} = H_{\xi_0^{\tau_2} \mid \xi_{-\tau_1-\sigma}^0},$$

то, переходя в (5.6.9) к пределу $\sigma \to \infty$, получаем

$$H_{\xi_0^{\tau_1+\tau_2} \mid \xi_{-\infty}^0} = H_{\xi_0^{\tau_1} \mid \xi_{-\infty}^0} + H_{\xi_0^{\tau_2} \mid \xi_{-\infty}^0}.$$

Следовательно, условная энтропия $H_{\xi_0^\tau \mid \xi_{-\infty}^0}$ линейно зависит от τ. Соответствующий коэффициент пропорциональности определяем как удельную энтропию:

$$h = \frac{1}{\tau} H_{\xi_0^\tau \mid \xi_{-\infty}^0}. \tag{5.6.10}$$

Аналогом теоремы 5.1 в непрерывной версии будет следующая теорема.

Теорема 5.4. *Если энтропия* $H_{\xi_0^t}$ *конечная, то удельную энтропию* (5.6.10) *можно вычислять предельным переходом*

$$h = \lim_{t \to \infty} \frac{1}{t} H_{\xi_0^t}. \tag{5.6.11}$$

Доказательство проводится тем же методом, что и доказательство теоремы 5.1. Пользуясь свойством аддитивности (1.7.4а), представим энтропию $H_{\xi_0^t}$ в виде

$$H_{\xi_0^t} = H_{\xi_0^\sigma} + H_{\xi_\sigma^{\sigma+1} \mid \xi_0^\sigma} + H_{\xi_{\sigma+1}^{\sigma+2} \mid \xi_0^{\sigma+1}} + \ldots + \\ + H_{\xi_{t-1}^t \mid \xi_0^{t-1}} \quad (\sigma = t - n). \tag{5.6.12}$$

Вследствие (5.6.8), (5.6.10) имеем

$$H_{\xi_{\sigma+k}^{\sigma+k+1} \mid \xi_0^{\sigma+k}} = h + o_{\sigma+k}(1) \quad (0 \leqslant k \leqslant t - \sigma - 1), \tag{5.6.13}$$

где $o_{\sigma+k}(1)$ стремится к 0 при $\sigma \to \infty$. Подставляя (5.6.13) в (5.6.12), получаем

$$\frac{1}{t} H_{\xi_0^t} = \frac{1}{\sigma+n} H_{\xi_0^\sigma} + \frac{n}{\sigma+n} h + \frac{n}{\sigma+n} o_\sigma(1). \tag{5.6.14}$$

Устремим здесь σ и n к бесконечности, но так, чтобы $n/\sigma \to \infty$. Поскольку

$$H_{\xi_0^\sigma} = H_{\xi_0^1} + H_{\xi_1^2 \mid \xi_0^1} + \ldots + H_{\xi_{m-1}^\sigma \mid \xi_0^{m-1}} \leqslant m H_{\xi_0^1} \quad (m = [\sigma] + 1),$$

то при этом член $\frac{1}{\sigma+n} H_{\xi_0^\sigma} \leqslant \frac{m}{\sigma+n} H_{\xi_0^1}$ будет стремиться к 0. Член $\frac{n}{\sigma+n} o_\sigma(1)$ также будет стремиться к 0, а $\frac{n}{\sigma+n} h$ — к h, поскольку $n/(\sigma+n) \to 1$. Следовательно, из (5.6.14) получим требуемое соотношение (5.6.11). Доказательство закончено.

На случай непрерывного времени обобщаются и те высказывания из § 5.1, которые касаются граничной энтропии Г. По аналогии с (5.1.10) ее можно определить по формуле

$$2\Gamma = \lim_{\sigma\to\infty,\ \tau\to\infty} \left(H_{\xi_0^\sigma} + H_{\xi_0^\tau} - H_{\xi_0^{\sigma+\tau}} \right) \tag{5.6.15}$$

и представить в виде

$$\Gamma = \frac{1}{2} \int_0^\infty \left(H_{\xi_0^{d\tau} \mid \xi_{-\tau}^0} - h d\tau \right), \tag{5.6.16}$$

аналогичном (5.1.12).

При помощи величин h, Г энтропия $H_{\xi_0^t}$ конечного отрезка стационарного процесса выражается по формуле

$$H_{\xi_0^t} = th + 2\Gamma + o_t(1). \tag{5.6.17}$$

Из определения (5.6.15) граничной энтропии Г, если учесть (5.6.1), нетрудно усмотреть, что она не зависит от выбора меры Q или ν подобно тому, как это было в § 5.4.

Для энтропии $H^{P/Q}$ при выполнении условия мультипликативности по аналогии с (5.6.17) будет справедлива формула

$$H_{\xi_0^t}^{P/Q} = th^{P/Q} - 2\Gamma + o_t(1), \tag{5.6.18}$$

где Г — та же величина, что и в (5.6.17).

5.7. Энтропия гауссового процесса в непрерывном времени

1. Гауссов случайный процесс $\xi(t)$ при непрерывном времени t $(a \leqslant t \leqslant b)$, подобно случайным величинам, рассмотренным в § 5.4, характеризуется вектором средних значений $\mathbf{M}\,\xi(t) = m(t)$ и корреляционной матрицей $R(t, t') = \mathbf{M}\,[\xi(t) - m(t)]\,[\xi(t') - m(t')]$. Разница лишь в том, что теперь вектор представляет собой заданную на интервале $[a, b]$ функцию от t, тогда как в § 5.4 вектор состоял из l компонент. Матрица (скажем, R)

теперь является функцией двух переменных t, t', определяющей линейное преобразование

$$y(t)=\int_a^b R(t, t')\,x(t')\,dt' \qquad \left(t \in [a, b]\right)$$

вектора $x(t)$. На этот случай, как известно, могут быть перенесены все основные результаты теории конечномерных векторов. При этом требуется произвести такое тривиальное видоизменение формул, как замена суммы на интеграл и т. п. С описанными видоизменениями на случай непрерывного времени могут быть распространены изложенные в § 5.4 методы вычисления энтропии. Результирующие матричные формулы (5.4.8а), (5.4.13) сохраняют свое значение при новом понимании матриц и векторов.

Разумеется, указанные выражения теперь не обязаны быть конечными. Условие их конечности связано с условием абсолютной непрерывности меры P относительно меры ν или Q.

При обобщенном понимании векторов и матриц формулы (5.4.8а), (5.4.13) справедливы как для конечных, так и для бесконечных интервалов $[a, b]$ определения процесса для стационарного и для нестационарного случаев. В дальнейшем мы ограничимся рассмотрением стационарных процессов и вычислим для них удельные энтропии h, $h^{P/Q}$.

Для этой цели может быть применен прием, использованный в п. 2 § 5.5, заключающийся в переходе к периодическому стационарному процессу. Рассматривая на интервале $[0, T]$ процесс, имеющий корреляционную функцию $R(t, t') = R(t - t')$, можно сконструировать новую корреляционную функцию

$$\overline{R}(\tau)=\sum_{n=-\infty}^{\infty} R(\tau+nT), \tag{5.7.1}$$

которая помимо стационарности будет обладать еще свойством периодичности. Формула (5.7.1) аналогична формуле (5.5.14). Такую корреляционную функцию имеет процесс

$$\overline{\xi}(t)-\frac{1}{\sqrt{N}}\sum_{j=0}^{N-1}\xi(t+jT)$$

в пределе при $N \to \infty$.

Стационарная периодическая матрица $\overline{R}(t-t')$ типа (5.7.1) диагонализуется унитарным преобразованием U, имеющим матрицу

$$U_{tr}=\frac{1}{\sqrt{T}}\,e^{2\pi i t r/T} \quad t \in [0, T], \ \ r=\ldots, -1, 0, 1, 2, \ldots$$

При этом собственные значения λ матрицы $\overline{R}\ (t\ -\ t')$ оказываются такими:

$$\lambda_r = \int_0^T \overline{R}_\tau \, e^{-2\pi i \tau r/T} \, d\tau.$$

Если учесть (5.7.1), то отсюда будем иметь

$$\lambda_r = \int_{-\infty}^{\infty} e^{-2\pi i \tau r/T} R(\tau) \, d\tau = S\left(2\pi \frac{r}{T}\right), \tag{5.7.2}$$

где через $S\ (\omega)$ обозначена спектральная плотность

$$S(\omega) = \int_{-\infty}^{\infty} e^{-i\omega\tau} R(\tau)\, d\tau = S(-\omega) \tag{5.7.3}$$

процесса $\xi\ (t)$.

Подставляя (5.7.2) в (5.4.8а), получаем

$$\overline{H}_{\xi_0^T} = \frac{1}{2} \operatorname{Sp} \ln \mathrm{R} = \frac{1}{2} \sum_r \ln \lambda_r = \frac{1}{2} \sum_r \ln S\left(2\pi \frac{r}{T}\right). \tag{5.7.4}$$

Перейдем к энтропии $H^{P/Q}_{\xi_0^T}$, которая находится по формуле (5.4.13). Эту формулу удобно применить после диагонализации матриц $R\ (t\ -\ t')$, $\widetilde{R}\ (t\ -\ t')$. Тогда

$$\operatorname{Sp} G\left(\widetilde{\mathrm{R}}^{-1}\mathrm{R}\right) = \sum_r G\left(\frac{\lambda_r}{\widetilde{\lambda}_r}\right), \tag{5.7.5}$$

где

$$\widetilde{\lambda}_r = \widetilde{S}\left(2\pi \frac{r}{T}\right), \qquad \widetilde{S} = \int_{-\infty}^{\infty} e^{-i\omega\tau} \widetilde{R}(\tau)\, d\tau \tag{5.7.6}$$

по аналогии с (5.7.2), (5.7.3). Второй член в правой части (5.4.13) после диагонализации примет вид

$$\frac{1}{2}\left(\mathrm{m}^T - \widetilde{\mathrm{m}}^T\right) \widetilde{\mathrm{R}}^{-1} \left(\mathrm{m} - \widetilde{\mathrm{m}}\right) = \mathrm{c}^+ \mathrm{U}^{-1} \widetilde{\mathrm{R}}^{-1} \mathrm{U}\mathrm{c} =$$

$$= \sum \widetilde{\lambda}_r^{-1} c_r \quad \left(\mathrm{c} = \mathrm{U}^+ (\mathrm{m} - \widetilde{\mathrm{m}})\right). \tag{5.7.7}$$

Поскольку

$$c_r = \int U^+_{tr} \left[m(t) - \widetilde{m}(t)\right] dt = \begin{cases} \sqrt{T}\,(m - \widetilde{m}) & \text{при } r = 0, \\ 0 & \text{при } r \neq 0 \end{cases}$$

то из (5.4.13) вследствие (5.7.5), (5.7.7) будем иметь

$$\overline{H}^{P/Q}_{\xi_0^T} = \sum_{r=-\infty}^{\infty} G\left(\frac{S(2\pi r/T)}{\widetilde{S}(2\pi r/T)}\right) + T\left(\frac{(m - \widetilde{m})^2}{\widetilde{S}(0)}\right). \tag{5.7.8}$$

Сумма по r в правой части (5.7.8) содержит бесконечное число членов. Чтобы ряд сходился к конечному пределу, необходимо чтобы отношение $S(2\pi r/T)/\widetilde{S}(2\pi r/T)$ стремилось к 1 при $|r| \to \infty$, так как функция $G(x) = (x - 1 - \ln x)/2$ обращается в нуль (в области положительных значений x) лишь при $x = 1$. Вблизи точки $x = 1$ функция ведет себя так:

$$G(x) = \frac{1}{4}(x-1)^2 + O\left((x-1)^3\right).$$

Поэтому при выполнении условий

$$\sum_{r=0}^{\infty}\left[\frac{S(2\pi r/T)}{\widetilde{S}(2\pi r/T)} - 1\right]^2 < \infty, \qquad 0 < \frac{S(2\pi r/T)}{\widetilde{S}(2\pi r/T)} < \infty \tag{5.7.9}$$

энтропия (5.7.8) оказывается конечной, что свидетельствует об абсолютной непрерывности меры P относительно Q.

2. Перейдем к вычислению удельной энтропии для стационарного процесса, заданного на бесконечной временной оси. Поскольку замена (5.7.1) корреляционной функции оказывает существенное влияние лишь на граничные эффекты, то энтропии (5.7.4), (5.7.8) отличаются от энтропии, соответствующих корреляционной функции $R(\tau)$ величиной порядка 1, а не порядка $T \gg 1$. Поэтому найденные выражения (5.7.4), (5.7.8) можно использовать для вычисления удельной энтропии

$$h = \lim_{T\to\infty}\frac{1}{T}H_{\xi_0^T} = \lim_{T\to\infty}\frac{1}{T}\overline{H}_{\xi_0^T},$$

$$h^{P/Q} = \lim_{T\to\infty}\frac{1}{T}H^{P/Q}_{\xi_0^T} = \lim_{T\to\infty}\frac{1}{T}\overline{H}^{P/Q}_{\xi_0^T}.$$

Итак, в полученных формулах следует совершить предельный переход $T \to \infty$. При этом суммы по r обратятся в интеграл. Из (5.7.4) будем иметь

$$h = \frac{1}{4\pi}\lim_{T\to\infty}\sum_{r=-\infty}^{\infty}\ln S\left(\frac{2\pi r}{T}\right)\frac{2\pi}{T} = \frac{1}{4\pi}\int_{-\infty}^{\infty}\ln S(\omega)\,d\omega, \tag{5.7.10}$$

а (5.7.8) даст

$$h^{P/Q} = \frac{1}{2\pi}\int_{-\infty}^{\infty}G\left(\frac{S(\omega)}{\widetilde{S}(\omega)}\right)d\omega + \frac{(m-\widetilde{m})^2}{\widetilde{S}(0)} =$$

$$= \frac{1}{\pi}\int_{0}^{\infty}G\left(\frac{S(\omega)}{\widetilde{S}(\omega)}\right)d\omega + \frac{(m-\widetilde{m})^2}{\widetilde{S}(0)}. \tag{5.7.11}$$

Эти результаты можно получить также из формул (5.5.17), (5.5.18), определяющих удельную энтропию стационарных последовательностей как предельный результат при неограниченном уплотнении точек на временной оси. Выбрав точки $t_k = k\Delta$ из стационарного гауссова процесса в непрерывном времени t, можно получить стационарную гауссову последовательность $\{\xi\ (t_k)\}$, которая будет иметь корреляционную матрицу

$$R_{jk} = R((j-k)\Delta)$$

и те же средние значения m. Сравнивая формулу (5.5.15) с (5.7.3), нетрудно видеть, что

$$\varphi\left(\frac{\omega\Delta}{2\pi}\right)\Delta = \sum_{\sigma=-\infty}^{\infty} R(\sigma\Delta)\,\mathrm{e}^{-i\omega\sigma\Delta}\,\Delta = \int_{-\infty}^{\infty} R(\tau)\,\mathrm{e}^{-i\omega\tau}d\tau + o_\Delta(1),$$

т. е. функции $\varphi\ (\mu)$, $S\ (\omega)$ связаны между собой соотношением

$$\varphi(\mu)\,\Delta = S(\omega) + o_\Delta(1) \quad \text{при} \quad \mu = \omega\Delta/2\pi, \ |\omega| \sim 1. \tag{5.7.12}$$

Относя энтропию не к одному элементу последовательности, а к единице времени, имеем

$$h = \lim_{\Delta\to\infty} H_1/\Delta, \qquad h^{P/Q} = \lim_{\Delta\to 0} H_1^{P/Q}/\Delta.$$

Подставляя сюда (5.5.17), (5.5.18) при учете равенства (5.7.12) и аналогичного равенства для $\widetilde{\varphi}$, $\widetilde{S}$, получаем

$$h = \frac{1}{4\pi}\lim_{\Delta\to 0}\int_{-\pi/\Delta}^{\pi/\Delta}[-\ln\Delta + \ln S(\omega)]\,d\omega, \tag{5.7.13}$$

$$h^{P/Q} = \frac{1}{2\pi}\int_{-\infty}^{\infty} G\left(\frac{S(\omega)}{\widetilde{S}(\omega)}\right)d\omega + \frac{(m-\widetilde{m})^2}{2\,\widetilde{S}(0)}. \tag{5.7.14}$$

Формула (5.7.14) совпадает с (5.7.11), а (5.7.13) отличается от (5.7.10) членом $-\ln\Delta$, который может быть отнесен к соответствующим образом подобранной мере ν. Пользуясь свободой выбора меры ν, можно представить формулы (5.7.10), (5.7. 13) в другой, более удобной форме. Предположим, что спектральная плотность $S\ (\omega)$ стремится при $\omega\to\infty$ к конечному отличному от нуля пределу

$$S(\infty) = \lim_{|\omega|\to\infty} S(\omega).$$

Учитывая, что согласно (5.7.12) $\varphi(\mu)\,\Delta = S\ (2\pi\mu/\Delta) + o_\Delta(1)$, будем иметь, следовательно, что при фиксированном μ в процессе предельного перехода $\Delta\to 0$ значение $\varphi(\mu)\,\Delta$ будет стремиться к $S\ (\infty)$. Это предельное значение целесообразно выделить, записав формулу (5.5.16) в виде

$$\overline{H}_1 = \frac{1}{2}\ln\frac{S(\infty)}{\Delta} + \frac{1}{2l}\sum_{r=0}^{l-1}\ln\frac{\varphi(r/l)\,\Delta}{S(\infty)}. \tag{5.7 15}$$

Определим теперь меру ν не формулой (5.4.4), а формулой

$$\frac{\nu(d\xi(t_j))}{d\xi(t_j)} = (2\pi e)^{-1/2}(S(\infty)/\Delta)^{-1/2}. \tag{5.7.16}$$

Тогда первый член в (5.7.15) будет отнесен к мере ν и вместо (5.7.15) будем иметь

$$\overline{H}_1 = \frac{1}{2l}\sum_{r=0}^{l-1} \ln \frac{\varphi(r/l)\,\Delta}{S(\infty)}. \tag{5.7.17}$$

Соответственно этому формула (5.5.17) заменится формулой

$$H_1 = \frac{1}{2}\int_{-1/2}^{1/2} \ln \frac{\varphi(\mu)\,\Delta}{S(\infty)}\, d\mu,$$

а из последней по аналогии с (5.7.13) получим

$$h = \frac{1}{4\pi}\int_{-\infty}^{\infty} \ln \frac{S(\omega)}{S(\infty)}\, d\omega. \tag{5.7.18}$$

Здесь подынтегральное выражение стремится к нулю при $|\omega| \to \infty$. Поскольку $\ln(1+y) = y - \frac{y^2}{2} + \ldots$, то интеграл в (5.7.18) сходится, если

$$\int_c^{\infty} \frac{S(\omega)-S(\infty)}{S(\infty)}\, d\omega < \infty, \qquad \int_c^{\infty}\left[\frac{S(\omega)}{S(\infty)} - 1\right]^2 d\omega < \infty \tag{5.7.19}$$

(c — некоторое число). Таким образом, мы получаем для удельной энтропии h конечное значение, если выполняется условие (5.7.19) и все особенности спектральной плотности (где она обращается в бесконечность или нуль) являются логарифмически интегрируемыми, как, например, нули и полюса вида

$$S(\omega) - |\omega - \omega_1|^r, \quad |r| < \infty.$$

При выполнении этих условий мера P является абсолютно непрерывной относительно специальным образом сконструированной меры ν, определенной соотношением (5.7.16) и условием мультипликативности.

Условие сходимости другого интеграла (5.7.11) в верхнем пределе имеет вид

$$\int_c^{\infty}\left[\frac{S(\omega)}{\widetilde{S}(\omega)} - 1\right]^2 d\omega < \infty, \tag{5.7.19a}$$

аналогичный (5.7.9). Оно является необходимым условием абсолютной непрерывности меры P относительно Q.

Если для меры Q выполнено условие мультипликативности то (в стационарном случае) ее корреляционная матрица должна быть кратна единичной матрице, а спектральная плотность постоянна: $\widetilde{S}(\omega) = \widetilde{S}$. Для абсолютной непрерывности необходимо равенство $S(\infty) = \widetilde{S}(\infty)$, следовательно, $\widetilde{S}(\omega) = S(\infty)$. Подставляя это значение в (5.7.11), получаем при совпадающих средних формулу

$$h^{P/Q} = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} G\left(\frac{S(\omega)}{S(\infty)}\right) d\omega \quad (G(x) = (x - 1 - \ln x)/2). \quad (5.7.20)$$

Сходство последней с (5.7.18) является очевидным, различие заключается лишь в выборе подынтегральной функции.

Пример 1. Пусть имеется стационарный гауссов процесс со спектральной плотностью

$$S(\omega) = S_0 \frac{\omega^2 + \gamma^2}{\omega^2 + \beta^2}.$$

Вычислим для него удельные энтропии. Поскольку $S(\infty) = S_0$, то в соответствии с (5.7.18) имеем

$$h = \frac{1}{2\pi} \int\limits_{0}^{\infty} \ln \frac{\omega^2 + \gamma^2}{\omega^2 + \beta^2} d\omega = \frac{1}{2}(\gamma - \beta).$$

Применяя далее формулу (5.7.20), находим

$$h^{P/Q} = \frac{1}{2\pi} \int\limits_{0}^{\infty} \left(\frac{\omega^2 + \gamma^2}{\omega^2 + \beta^2} - 1 \right) d\omega - \frac{1}{2\pi} \int\limits_{0}^{\infty} \ln \frac{\omega^2 + \gamma^2}{\omega^2 + \beta^2} =$$

$$= \frac{\gamma^2 - \beta^2}{4\beta} - h = \frac{1}{4\beta}(\gamma - \beta)^2. \quad (5.7.21)$$

Граничная энтропия Γ, входящая в соотношения (5.6.17), (5.6.18), может быть вычислена методом, который предложен Стратоновичем [6]. Для данного примера она оказывается такой:

$$\Gamma = \frac{1}{2} \ln \frac{(\gamma/\beta)^{1/2} + (\beta/\gamma)^{1/2}}{2}.$$

Пример 2. Пусть теперь рассматриваемый процесс имеет спектральную плотность $S(\omega) = S_0/(\omega^2 + \beta^2)$. Тогда применение формулы (5.7.18) не дает конечного результата. Чтобы получить конечную удельную энтропию по формуле (5.7.14), нужно подобрать подходящую спектральную плотность $\widetilde{S}(\omega)$. Положим

$$\widetilde{S}(\omega) = S_0/\omega^2. \quad (5.7.22)$$

Тогда интеграл в (5.7.14) сведется к интегралам, стоящим в (5.7.21) (при $\gamma = 0$), и мы будем иметь

$$h^{P/Q} = \beta/4. \tag{5.7.23}$$

При выборе спектральной плотности (5.7.22) мера Q, описывающая случайный процесс $\eta\,(t)$, не удовлетворяет условию мультипликативности. Однако это условие будет выполнено для производной $\dot{\eta}\,(t) = d\eta\,(t)/dt$, так как последняя будет иметь равномерную спектральную плотность S_0. При переходе к производной $\dot{\eta}\,(t)$ процесс $\xi\,(t)$ также нужно заменить на его производную $\dot{\xi}\,(t)$. Можно считать поэтому, что результат (5.7.23) согласуется с условием мультипликативности.

3. Для стационарного гауссового процесса, как и для гауссовой последовательности (см. § 5.4., п. 3), из линейного роста дисперсии $\mathbf{D}H\,(\xi_0^T)$ с ростом T при условии линейного нарастания средней энтропии $H_{\xi_0^T}$ вытекает выполнение условия энтропийной устойчивости (1.5.8). Поэтому в случае ненулевой конечной удельной энтропии h для доказательства энтропийной устойчивости следует проверить конечность предела

$$D_0 = \lim_{T\to\infty} \frac{1}{T}\,\mathbf{D}\,H\,(\xi_0^T).$$

Чтобы его вычислить, нужно применить формулу (5.4.17). Диагонализуя входящие в нее матрицы, нетрудно получить (подобно тому, как из (5.4.13) было получено выражение (5.7.14)) следующее равенство:

$$D_0 = \frac{1}{2\pi}\int_0^\infty \left[\frac{S\,(\omega)}{\widetilde{S}\,(\omega)} - 1\right]^2 d\omega + \frac{(m-\widetilde{m})^2\,S\,(0)}{4\widetilde{S}^2\,(0)}\,. \tag{5.7.24}$$

Оно является обобщением на случай непрерывного времени равенства (5.5.22). Условие конечности интеграла в (5.7.24), очевидно, связано с указанным ранее условием (5.7.19а). Таким образом, условие энтропийной устойчивости для гауссовых мер оказывается тесно связанным с условием абсолютной непрерывности меры P относительно Q.

Вышеизложенное может быть обобщено и на случай нескольких стационарных и стационарно связанных процессов $\{\xi^\alpha\,(t)\}$, которые характеризуются столбцом средних значений $\mathrm{m} = \|m^\alpha\|$ и матрицей спектральных плотностей

$$\mathrm{S}\,(\omega) = \|S^{\alpha\beta}\,(\omega)\|,\quad S^{\alpha\beta}\,(\omega) = \int_{-\infty}^{\infty} \mathrm{e}^{-i\omega\tau}\,\times$$
$$\times\,\mathbf{M}[\xi^\alpha\,(t) - m^\alpha]\,[\xi^\beta\,(t+\tau) - m^\beta]\,d\tau.$$

Так, если мера Q описывается матрицами $\widetilde{\mathrm{m}}$, $\widetilde{\mathrm{S}}(\omega)$, то формула, обобщающая формулу (5.7.14), имеет вид

$$h^{P/Q} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \operatorname{Sp} G\left(\widetilde{\mathrm{S}}^{-1}(\omega)\, S(\omega)\right) d\omega + \\ + \frac{1}{2}\left(\mathrm{m}^T - \widetilde{\mathrm{m}}^T\right) \widetilde{\mathrm{S}}(0)^{-1}\left(\mathrm{m} - \widetilde{\mathrm{m}}\right). \tag{5.7.25}$$

Очевидна аналогия этой формулы с (5.5.20).

5.8. Энтропия точечного случайного процесса

Рассмотрим точечный случайный процесс на интервале $a \leqslant t \leqslant b$, представляющий собой совокупность случайных точек, положение и число которых случайно. Иногда подобный процесс называют «случайным потоком», однако, этот термин нам представляется неудачным, поскольку слово «поток» естественнее употреблять в другом смысле, связывая его с движением в пространстве, подобно тому, как это делается в физике.

Обозначим через n число выпавших точек, а через $\tau_1, \ldots, \tau_n$ места их выпадения (так что все $\tau_i \in [a, b]$). Будем предполагать, что величины $\tau_1, \ldots, \tau_n$ пронумерованы в порядке неубывания: $\tau_1 \leqslant \tau_2 \leqslant \ldots \leqslant \tau_n$. Совпадение $\tau_i = \tau_j$, $i \neq j$ предполагаем имеющим нулевую вероятность.

1. Для описанного процесса мера $\mathbf{P}$ характеризуется заданием вероятностей

$$\mathbf{P}(n) = P_n \tag{5.8.1}$$

и системы плотностей распределения

$$p(\tau_1 \mid 1), \qquad p(\tau_1, \tau_2 \mid 2), \ldots, \tag{5.8.2}$$

удовлетворяющих условию нормировки

$$\int_a^b d\tau_1 \int_{\tau_1}^b d\tau_2 \int_{\tau_2}^b d\tau_3 \ldots \int_{\tau_{n-1}}^b d\tau_n\, p(\tau_1, \ldots, \tau_n \mid n) = 1. \tag{5.8.3}$$

Требуется определить энтропию данного точечного процесса. Вероятности (5.8.1) и плотности (5.8.3) определяют вероятность

$$\Delta P = P_n\, p_n(t_1, \ldots, t_n)\, \Delta_1 \ldots \Delta_n \left(1 + O(\Delta_{\max})\right) \qquad (t_1 < \ldots < t_n), \\ \Delta_{\max} = \max(\Delta_1, \ldots, \Delta_n) \tag{5.8.4}$$

того, что первая точка τ_1 из n последовательных случайных точек $\tau_1 < \ldots < \tau_n$ попадет в интервал $[t_1, t_1 + \Delta_1]$, вторая — в интервал $[t_2, t_2 + \Delta_2]$ и т. д., последняя — в интервал $[t_n, t_n + \Delta_2]$

при условии, что в других местах интервала $[a, b]$ не выпадет ни одной случайной точки. Простейшей системой случайных точек является пуассоновская система, для которой

$$P_n = \frac{1}{n!}\left(\int_a^b \beta(t)\,dt\right)^n \exp\left[-\int_a^b \beta(t)\,dt\right],$$

$$p_n(\tau_1, \ldots, \tau_n) = n!\,\beta(\tau_1)\ldots\beta(\tau_n) \Big/ \left[\int_a^b \beta(t)\,dt\right]^n. \tag{5.8.5}$$

Здесь $\beta(\tau)$ — средняя плотность точек, которая в стационарном случае постоянна.

Рассмотрение случайных точек $\tau_1, \ldots, \tau_n$ эквивалентно рассмотрению случайного процесса

$$\xi(t) = \sum_{j=1}^{n} \delta(t-\tau_j).$$

Выпадение случайных точек на неперекрывающихся интервалах $[\alpha, \beta]$, $[\gamma, \delta]$ для пуассоновского процесса является взаимно независимым (или, иначе говоря, ξ_α^β независимы от ξ_γ^δ). Поэтому для выполнения условия мультипликативности (5.6.4) удобно брать энтропию точечного процесса, определенную в соответствии с (1.6.16), (1.6.17), выбирая меру Q, соответствующую пуассоновскому процессу. Будем полагать для простоты, что мере Q соответствует постоянная пуассоновская плотность β.

В соответствии с (5.8.5) для отношения элементарных вероятностей типа (5.8.4) в стационарном случае будем иметь

$$\frac{dP}{dQ} = \frac{P_n\, p_n(t_1, \ldots, t_n)}{\beta^n e^{-\beta(b-a)}}.$$

Поэтому энтропия (1.6.17) будет определяться формулой

$$H_{\xi_a^b}^{P/Q} = \mathbf{M} \ln \frac{dP}{dQ} = \sum_{n=0}^{\infty} \int_{a<t_1<\ldots<t_n<b} \cdots \int P_n\, p_n(t_1, \ldots, t_n) \times$$
$$\times \ln [P_n e^{\beta(b-a)} p_n(t_1, \ldots, t_n)\, \beta^{-n}]\, dt_1 \ldots dt_n. \tag{5.8.6}$$

Выделяя β из-под знака логарифма, это выражение, очевидно, можно записать в виде

$$H_{\xi_a^b}^{P/Q} = (\beta - 1)\,T - \ln\beta\, \mathbf{M}n + H_{\xi_a^b}^{P/Q_1} \quad (T = b - a), \tag{5.8.7}$$

$$H_{\xi_a^b}^{P/Q_1} = \sum_{n=0}^{\infty} \int_{a<t_1<\ldots<t_n<b} \cdots \int P_n\, p_n(t_1, \ldots, t_n) \times$$
$$\times \ln [e^T P_n\, p_n(t_1, \ldots, t_n)]\, dt_1 \ldots dt_n. \tag{5.8.8}$$

— энтропия, соответствующая пуассоновской мере Q_1 с единичной плотностью.

Из энтропии (5.8.8) можно выделить энтропию H_n случайного числа точек

$$H_n = -\sum_{n=0}^{\infty} P_n \ln P_n, \tag{5.8.9}$$

представив (5.8.8) в форме

$$H_{\xi_a^b}^{P/Q_1} = H_{\xi_a^b|n}^{P/Q_1} - H_n,$$

где

$$H_{\xi_a^b|n}^{P/Q_1} = \sum_{n=0}^{\infty} P_n \int\limits_{a<t_1<\ldots<t_n<b}\ldots\int p_n(t_1, \ldots, t_n) \times$$
$$\times \ln [e^T p_n(t_1, \ldots, t_n)]\, dt_1 \ldots dt_n. \tag{5.8.10}$$

Вследствие общих свойств энтропии выражения (5.8.6), (5.8.8) — (5.8.10) всегда неотрицательны.

Возьмем для примера пуассоновскую систему точек с постоянной плотностью γ. Тогда в силу (5.8.5), (5.8.8) будем иметь

$$H_{\xi_a^b}^{P/Q_1} = \mathbf{M} \ln (e^{T-\gamma T} \gamma^n) = \gamma T \ln \gamma + (1-\gamma) T, \tag{5.8.11}$$

поскольку $\mathbf{M} n = \gamma T$.

Далее из (5.8.7) получаем

$$H_{\xi_a^b}^{P/Q} = (\beta - \gamma)(b-a) + \gamma (b-a) \ln \frac{\gamma}{\beta}. \tag{5.8.12}$$

Поделив этот результат на $b - a$, найдем удельную энтропию одной пуассоновской меры по другой:

$$h^{P/Q} = \beta - \gamma + \gamma \ln \frac{\gamma}{\beta} = \gamma \left(\frac{\beta}{\gamma} - 1 - \ln \frac{\beta}{\gamma} \right) \geqslant 0.$$

Энтропии (5.8.11), (5.8.12) пропорциональны $b - a$ вследствие того, что пуассоновские процессы являются процессами с независимыми значениями. Более сложно зависят от длины интервала энтропии (5.8.9), (5.8.10). Вводя функцию

$$S(x) = e^{-x} \sum_{n=0}^{\infty} \frac{x^n}{n!} \ln n!,$$

эти выражения для пуассоновской меры P можно записать

$$H_n = \gamma T [1 - \ln (\gamma T)] + S(\gamma T), \qquad H_{\xi_0^T|n}^{P/Q_1} = T + S(\gamma T) - \gamma T \ln T.$$

2. Приведенные выше формулы основывались на определении (1.6.17) энтропии одной вероятностной меры по другой. Можно перейти к энтропии (1.6.13), (1.6.16), соответствующей ненормированной мере ν. Эта мера будет удовлетворять условию мульти-

пликативности, если будет пропорциональна пуассоновской вероятностной мере Q.

Согласно сказанному в § 1.6 обобщенную энтропию (1.6.16) можно интерпретировать как предельный случай энтропии дискретной версии, в которой число возможных исходов является конечным или счетным. Чтобы перейти от случая непрерывного времени t к дискретному случаю, разобьем рассматриваемый интервал на конечное число элементарных интервалов $[t'_m, t'_{m+1})$ и вместо интервала $[a, b]$ будем рассматривать множество Z точек $t'_0 = a$; $t'_1, \ldots, t'_{N-1}$; $t'_N = b$. Точечный процесс на Z определим так: пусть $\xi'(t'_m) = 1$, если на интервал $[t'_m, t'_{m+1})$ попадает хотя бы одна точка τ_k и $\xi'(t'_m) = 0$ в противном случае. Имея в виду рассмотреть в дальнейшем стационарный процесс, будем полагать разбиение равномерным:

$$t'_{m+1} - t'_m = \Delta, \qquad m = 0, 1, \ldots, N-1 \qquad (N = (b-a)/\Delta).$$

Вероятность выпадения на элементарном интервале больше одной точки предполагается величиной порядка Δ^2, т. е. высшего порядка малости. Тогда при достаточно большом N число n случайных точек, выпавших в $[a, b]$ с вероятностью, близкой к единице, не будет отличаться от числа точек $n' = \sum\limits_{k=0}^{N-1} \xi'(t'_k)$, выпавших в Z.

Точечный процесс на Z является дискретным процессом, имеющим в общей сложности 2^N различных реализаций. Событие $n' = 0$ реализуется одним способом. Событие $n' = 1$ реализуется N различными способами, т. е. включает N различных реализаций. Событие $n' = 2$ состоит из $N(N-1)/2$ различных реализаций и т. д. Вводя как в § 1.6 меру, указывающую число различных реализаций, имеем

$$\nu(n' = 0) = 1, \qquad \nu(n' = 1) = N, \quad \nu(n' = 2) = N(N-1)/2!, \ldots$$
$$\ldots, \nu(n') = N!/n'!(N-n')!, \ldots$$

Причем

$$\sum_{n'=0}^{\infty} \nu(n') = 2^N.$$

Нетрудно подсчитать также число реализаций того события, что τ_1 попадет в интервал $[t_1, t_1 + \Delta_1)$, в то же время τ_2 попадает в интервал $[t_2, t_2 + \Delta_2)$ и т. д., τ_n попадет в интервал $[t_n, t_n + \Delta_n)$. В предположении, что $\Delta_k \gg \Delta$, $n \ll N$, число таких реализаций равно

$$\Delta\nu \approx \frac{\Delta_1}{\Delta} \ldots \frac{\Delta_n}{\Delta}.$$

Напомним, что вероятность такого множества задается формулой (5.8.4). Поскольку

$$\frac{\Delta P}{\Delta \nu} \approx P_n p_n (t_1, \ldots, t_n) \Delta^n,$$

то применяя формулы (1.6.5), (1.6.6), получаем

$$H_{\xi'} \approx -\sum P_n p_n (t_1, \ldots, t_n) \ln [P_n p_n (t_1, \ldots, t_n) \Delta^n].$$

Вынося Δ из-под знака логарифма, полученный результат можно выразить через энтропию (5.8.8):

$$H_{\xi'} \approx \ln \frac{1}{\Delta} \mathbf{M} n + b - a - H_{\xi_a^b}^{P/Q_1}. \tag{5.8.13}$$

Это обосновывает введение энтропии H_ξ^{P/Q_1} меры P относительно пуассоновской меры Q_1. Если разбиение интервала $[a, b]$ производить не равномерным образом, а согласно соотношению $t_{m+1} - t_m = \beta(t_m)\Delta$, то энтропия $H_{\xi'}$ по аналогии с (5.8.13) будет выражаться через энтропию $H_\xi^{P/Q}$ меры P по пуассоновской мере Q, имеющей неравномерную плотность $\beta(t)$.

3. В случае стационарного точечного процесса можно рассматривать удельную энтропию, приходящуюся в среднем на единицу времени. Укажем два способа ее вычисления.

1) В соответствии с теоремой 5.2 ее можно вычислять по формуле (5.6.11). Имея в виду применить эту формулу, исследуем поведение энтропий (5.8.8), (5.8.10) при больших n.

В случае стационарного процесса среднее число точек $\mathbf{M}n \equiv \bar{n}$ пропорционально длине интервала $b - a = T$. Для эргодического процесса, кроме того, зависимость дисперсии $\mathbf{D}n$ от T при больших n, как правило, приближается к линейной:

$$\mathbf{D}n = D_0 T + O(1),$$

и случайная величина n подчиняется центральной предельной теореме. Вероятность неравенства $n_0 \leqslant n \leqslant n_0 + \Delta N$ при больших n может быть вычислена при помощи гауссового распределения:

$$\Delta P \approx \frac{1}{\sqrt{2\pi \mathbf{D}n}} e^{-(n_0 - \bar{n})^2/2\mathbf{D}n} \Delta N \qquad (\Delta N^2 \ll \mathbf{D}n), \tag{5.8.14}$$

где ΔN имеет тот же смысл, что и в (1.6.4а). Это позволяет вычислить H_n приближенно как энтропию гауссовой переменной. Из (5.8.14) имеем

$$H(n) = -\ln P_n \approx -\ln \frac{\Delta P}{\Delta N} \approx \frac{1}{2} \ln (2\pi \mathbf{D}n) + \frac{(n - \bar{n})^2}{2\mathbf{D}n}, \tag{5.8.15}$$

что является частным случаем формулы (5.4.5). Усредняя (5.8.15), находим

$$H_n \approx \frac{1}{2}\ln 2\pi e\,\mathbf{D}n \approx \frac{1}{2}\ln(2\pi e D_0 T).$$

Из найденной зависимости следует, в частности, исчезновение предела

$$\lim_{T\to\infty}\frac{1}{T}H_n = 0.$$

Это означает, что вклад энтропии H_n в удельную энтропию (5.6.11) сводится к нулю. На удельную энтропию, следовательно, оказывает влияние лишь энтропия (5.8.10):

$$h^{P/Q_1} = \lim_{T\to\infty}\frac{1}{T}H^{P/Q_1}_{\xi_0^T|n}. \qquad (5.8.16)$$

Вводя обозначение

$$\Phi(n) = H^{P/Q_1}_{\xi_0^T}(|\,n) = \int\limits_{0\leqslant t_1<}\cdots\int\limits_{<t_n\quad T} p_n(t_1,\ldots,t_n)\times$$

$$\times\ln[e^T p_n(t_1,\ldots,t_n)]\,dt_1\ldots dt_n, \qquad (5.8.17)$$

ее можно записать так

$$H^{P/Q_1}_{\xi_0^T|n} = \mathbf{M}\Phi(n). \qquad (5.8.18)$$

Величина n при больших T является квазинепрерывной. Для нее $\bar{n}$ мало отличается от $[\bar{n}]$ (скобки $[\ldots]$ означают целую часть). Разности

$$\Phi'(n) = \Phi(n+1) - \Phi(n),$$

$$\Phi''(n) = \Phi(n+1) - 2\Phi(n) + \Phi(n-1)$$

можно интерпретировать как производные. Пренебрегая несущественными усложнениями, связанными с дискретным характером величины n, из (5.8.17) будем иметь

$$\mathbf{M}\Phi(n) = \mathbf{M}\left[\Phi(\bar{n}) + \Phi'(\bar{n})(n-\bar{n}) + \frac{1}{2}\Phi''(\bar{n})(n-\bar{n})^2 + \ldots\right] =$$

$$= \Phi(\bar{n}) + \frac{1}{2}\Phi''(\bar{n})\,\mathbf{D}n + \ldots.$$

Подставляя это выражение в (5.8.18) и (5.8.16), получаем, что

$$h^{P/Q_1} = \lim_{T\to\infty}\frac{1}{T}\Phi(\bar{n}), \qquad (5.8.19)$$

если

$$\lim_{T\to\infty}\Phi''(\bar{n})\frac{\mathbf{D}n}{T} = 0. \qquad (5.8.20)$$

Уже указывалось, что обычно отношение $\mathbf{D}n/T$ стремится при $T \to \infty$ к конечному пределу D_0. Следовательно, условие (5.8.20) выполняется, если

$$\Phi''(\bar{n}) \to 0 \text{ при } T \to \infty. \tag{5.8.21}$$

Как показывает исследование, условие (5.8.21), говорящее о приближении зависимости $\Phi(n)$ при $n - cT$ к прямой пропорциональности, выполняется для большого числа практических случаев.

Итак, в соответствии с формулой (5.8.19) удельную энтропию можно вычислить, считая число случайных точек, выпавших на большом интервале $[0, T]$, неслучайным, заранее известным и равным $\bar{n} = \mathbf{M}n$.

Пример. Вычислим этим способом энтропию пуассоновского процесса. Предполагается, что число точек на всем интервале является неслучайным и равным $\bar{n} = \gamma T$. При этом $p_n(t_1 \ldots, t_n) = n! T^{-n}$. Подставляя эту функцию в (5.8.17), находим

$$\Phi(n) = T + \ln n! - n \ln T.$$

Пользуясь формулой Стирлинга

$$n! = n^n e^{-n} \sqrt{2\pi n}\, [1 + O_n(1)],$$

получаем

$$\Phi(n) = T + n \ln n - n + \frac{1}{2} \ln(2\pi n) - n \ln T.$$

Условие (5.8.21) действительно имеет место, поскольку

$$\Phi''(\bar{n}) \approx \frac{1}{\bar{n}} = \frac{1}{\gamma T}.$$

Подставляя найденное выражение для $\Phi(n)$ в (5.8.19) при $\bar{n} = \gamma T$ и переходя к пределу $T \to \infty$, имеем

$$h^{P/Q_1} = 1 + \gamma \ln\gamma - \gamma, \tag{5.8.22}$$

что, конечно, совпадает с удельной энтропией, находимой из (5.8.11).

2) Другой способ вычисления удельной энтропии основан на ее определении (5.6.10) как условной энтропии. Фиксация процесса $\xi_{-\infty}^0$ означает фиксацию всех случайных точек $\ldots, \tau_{-2}, \tau_{-1}, \tau_0 < 0$, выпавших до момента $t = 0$. Поэтому (5.6.10) можно записать:

$$h^{P/Q_1} = \frac{1}{\tau} H^{P/Q_1}_{\xi_0^\tau \mid \ldots, \tau_{-1}, \tau_0} = \frac{1}{\tau} \mathbf{M} H^{P/Q_1}_{\xi_0^\tau}(\mid \ldots, \tau_{-1}, \tau_0). \tag{5.8.23}$$

Энтропию $H^{P/Q}_{\xi_0^\tau}(\mid \ldots, \tau_{-1}, \tau_0)$, здесь следует определять формулой (5.8.8), но при замене вероятностей P_n и плотностей

$p_n\ (\tau_1, \ldots, \tau_n)$ условными вероятностями и плотностями

$$P(n \mid \ldots, \tau_{-1}, \tau_0), \ p_n(\tau_1, \ldots, \tau_n \mid \ldots, \tau_{-1}, \tau_0), \tag{5.8.24}$$

т. е. формулой

$$H_{\xi_0^\tau}^{P/Q_1}(\mid \ldots, \tau_{-1}, \tau_0) = \sum_{n=0}^{\infty} P(n \mid \ldots, \tau_{-1}, \tau_0) \times$$
$$\times \int_{t_1 <} \cdots \int_{t_n} p_n(t_1, \ldots, t_n \mid \ldots, \tau_1, \tau_0) \ln [e^\tau P_n(n \mid \ldots, \tau_{-1}, \tau_0) \times$$
$$\times p_n(t_1, \ldots, t_n \mid \ldots, \tau_{-1}, \tau_0)]\, dt_1 \ldots dt_n. \tag{5.8.24a}$$

Длину τ интервала $[0, \tau]$ желательно брать небольшой, чтобы вероятности выпадения двух и более точек на этом интервале были пренебрежимо малы и их можно было бы не учитывать. Тогда в выражении (5.8.24) можно будет оставить лишь два члена, полагая при этом $p_1\ (t_1 \mid \ldots, \tau_{-1}, \tau_0) = 1/\tau$. Обозначая кроме того

$$\lim_{\tau \to 0} \mathbf{P}(n=1 \mid \ldots, \tau_{-1}, \tau_0)/\tau = p(0 \mid \ldots, \tau_{-1}, \tau_0),$$

из (5.8.24) будем иметь

$$H_{\xi_0^\tau}^{P/Q_1}(\mid \ldots, \tau_{-1}, \tau_0) - [1 - p(0 \mid \ldots, \tau_{-1}, \tau_0)\tau] \times$$
$$\times \ln \{e^\tau [1 - p(0 \mid \ldots, \tau_{-1}, \tau_0)\tau]\} +$$
$$+ p(0 \mid \ldots, \tau_{-1}, \tau_0)\tau \ln [e^\tau p(0 \mid \ldots, \tau_{-1}, \tau_0)] + O(\tau^2),$$

или

$$H_{\xi_0^\tau}^{P/Q_1}(\mid \ldots, \tau_{-1}, \tau_0) = \tau - p(0 \mid \ldots, \tau_{-1}, \tau_0)\tau +$$
$$+ p(0 \mid \ldots, \tau_{-1}, \tau_0)\tau \ln p(0 \mid \ldots, \tau_{-1}, \tau_0) + O(\tau^2). \tag{5.8.25}$$

Подставляя это выражение в (5.8.23) и переходя к пределу $\tau \to 0$, получаем искомую удельную энтропию

$$h^{P/Q_1} = \mathbf{M}[1 - p(0 \mid \ldots, \tau_{-1}, \tau_0) + p(0 \mid \ldots, \tau_{-1}, \tau_0) \times$$
$$\times \ln p(0 \mid \ldots, \tau_{-1}, \tau_0)], \tag{5.8.26}$$

где $\mathbf{M}$ обозначает усреднение по случайным точкам $\ldots, \tau_{-1}, \tau_0$.

Пример 1. Применим найденную формулу к стационарному точечному процессу с ограниченным последействием. Будем полагать, что интервалы $\sigma = t_{m+1} - t_m$ между соседними случайными точками являются независимыми случайными величинами, имеющими одинаковую плотность распределения вероятностей $w\ (\sigma)$. Тогда, очевидно,

$$\mathbf{P}(n=1 \mid \ldots, \tau_{-1}, \tau_0) = w(-\tau_0)\tau \Big/ \int_{-\tau_0}^{\infty} w(\sigma)\, d\sigma + O(\tau^2),$$

так что

$$p(0 \mid \ldots, \tau_{-1}, \tau_0) = w(-\tau_0) \Big/ \int_{-\tau_0}^{\infty} w(\sigma)\, d\sigma.$$

Подстановка в (5.8.26) дает

$$h^{P/Q_1} = 1 + \int_0^\infty \left[\ln \frac{w(\sigma)}{\int_\sigma^\infty w(\rho)\,d\rho} - 1 \right] \frac{w(\sigma)}{\int_\sigma^\infty w(\rho)\,d\rho} P(d\sigma), \qquad (5.8.27)$$

где $P(d\sigma) = \mathbf{P}[\sigma \leqslant -\tau_0 < \sigma + d\sigma] \equiv p(\sigma)\,d\sigma$ — закон распределения для случайного расстояния от фиксированной точки $t = 0$ до ближайшей слева случайной точки. Его можно выразить через плотность $w(\sigma)$ несложной формулой

$$p(\sigma)\,d\sigma = \frac{d\sigma}{\bar{\sigma}} \int_\sigma^\infty w(\rho)\,d\rho, \qquad \bar{\sigma} = \int \rho w(\rho)\,d\rho. \qquad (5.8.28)$$

Вследствие равенств (5.8.27), (5.8.28) получаем

$$h^{P/Q_1} = 1 + \frac{1}{\bar{\sigma}} \int_0^\infty \left[\ln w(\sigma) - \ln \int_\sigma^\infty w(\rho)\,d\rho - 1 \right] w(\sigma)\,d\sigma, \qquad (5.8.29)$$

что дает решение задачи вычисления удельной энтропии.

Пример 2. Рассмотрим несколько более сложный пример. Пусть дан стационарный точечный процесс. Интервалы $\tau_{m+1} - \tau_m$, по-прежнему, взаимно независимы, но имеют (через один) различные плотности распределения; $w_1(\sigma)$ или $w_2(\sigma)$. Если $\tau_{m+1} - \tau_m$ распределен по w_1, то $\tau_{m+2} - \tau_{m+1}$ распределен по w_2; интервал $\tau_{m+3} - \tau_{m+2}$ снова распределен по w_1, а $\tau_{m+4} - \tau_{m+3}$ — по w_2 и т. д. Такой точечный процесс эквивалентен стационарному процессу с двумя состояниями A_1 и A_2, когда времена пребывания в каждом состоянии взаимно независимы и имеют законы распределения w_1 (для времени пребывания в A_1) и w_2 (в A_2). Случайные точки можно при этом классифицировать дополнительным параметром ϑ, полагая $\vartheta_m = 1$, если в точке τ_m происходит скачок из A_1 в A_2, и $\vartheta_m = 2$, если происходит обратный скачок из A_2 в A_1.

В описанном случае плотность вероятности $p(0 \mid \ldots, \tau_{-1}, \tau_0)$ зависит не только от времени τ_0 выпадения последней случайной точки, но и от ее типа ϑ_0. Именно

$$p(0 \mid \ldots, \tau_{-1}, \tau_0) = p(0 \mid \tau_0, \vartheta_0) = w_{\vartheta_0}(-\tau_0) \Big/ \int_{-\tau_0}^\infty w_{\vartheta_0}(\rho)\,d\rho. \qquad (5.8.29\text{a})$$

Усреднение в (5.8.26) будет проводиться как по τ_0, так и по ϑ_0.

Обозначим через $P(\vartheta, d\sigma)$ совместное распределение для случайных величин $\vartheta = \vartheta_0$, $\sigma = -\tau_0$. Тогда из (5.8.26), (5.8.29a) будем иметь

$$h^{P/Q_1} = 1 + \sum_{\vartheta=1}^{2} \int \left[\ln \frac{w_\vartheta(\sigma)}{\int_\sigma^\infty w_\vartheta(\rho)\,d\rho} - 1 \right] \times$$

$$\times \frac{w_\vartheta(\sigma)}{\int\limits_\sigma^\infty w_\vartheta(\rho)\,d\rho} P(\vartheta, d\sigma). \tag{5.8.30}$$

Остается вычислить $P(\vartheta, d\sigma)$. Априори вероятность попадания на элементарный интервал $[-\sigma, -\sigma + d\sigma]$ случайной точки любого из двух типов одна и та же и равна

$$dP = \frac{d\sigma}{\bar\sigma_1 + \bar\sigma_2}, \qquad \bar\sigma_\vartheta = \int\limits_0^\infty \rho w_\vartheta(\rho)\,d\rho,$$

поскольку для каждого типа средняя плотность точек равна $(\bar\sigma_1 + \bar\sigma_2)^{-1}$. Если выпала точка $\vartheta_0 = 1$, то с вероятностью $\int\limits_\sigma^\infty w_1(\rho)\,d\rho$ она будет последней, т. е. в $[-\sigma + d\sigma, 0]$ не выпадет ни одной другой точки. Если $\vartheta_0 = 2$, то точка τ_0 будет последней с вероятностью $\int\limits_\sigma^\infty w_2(\rho)\,d\rho$. Поэтому формулу (5.8.30) можно записать

$$h^{P/Q_1} = 1 + \frac{1}{\bar\sigma_1 + \bar\sigma_2} \sum_{\vartheta=1}^{2} \int\limits_0^\infty \left[\ln w_\vartheta(\sigma) - \ln \int\limits_\sigma^\infty w_\vartheta(\rho)\,d\rho - 1\right] \times$$
$$\times w_\vartheta(\sigma)\,d\sigma. \tag{5.8.31}$$

Пусть, например w_1, w_2 имеют экспоненциальный вид

$$w_\vartheta(\sigma) = \mu_\vartheta\, e^{-\mu_\vartheta \sigma}.$$

Тогда

$$\bar\sigma_\vartheta = \frac{1}{\mu_\vartheta}, \qquad \int\limits_\sigma^\infty w_\vartheta(\rho)\,d\rho = e^{-\mu_\vartheta \sigma},$$

и из (5.8.31) получаем

$$h^{P/Q_1} = 1 + \frac{\mu_1\mu_2}{\mu_1 + \mu_2} \sum_{\vartheta=1}^{2} [\ln \mu_\vartheta - 1] =$$
$$= 1 - \frac{2\mu_1\mu_2}{\mu_1 + \mu_2} + \frac{\mu_1\mu_2}{\mu_1 + \mu_2} (\ln \mu_1 + \ln \mu_2). \tag{5.8.32}$$

Если, в частности, $\mu_1 = \mu_2 = \gamma$, то формула (5.8.32) совпадает с (5.8.22), так как при этом рассматриваемый точечный процесс переходит в пуассоновский.

5.9. Энтропия дискретного марковского процесса в непрерывном времени

Рассмотрим марковский процесс $\xi(t)$ с дискретным пространством состояний, т. е. имеющий конечное или счетное число возможных состояний. В отличие от процесса, рассмотренного в § 5.2, он протекает теперь в непрерывном времени. Пусть его вероятности переходов определяются дифференциальной вероятностью перехода $\pi_t(x, x')$ в соответствии с формулой

$$P[\xi(t+\Delta) \mid \xi(t)] = \delta_{\xi(t+\Delta),\, \xi(t)} + \pi_t\left(\xi(t), \xi(t+\Delta)\right) \Delta + O(\Delta^2).$$

В силу условия нормировки имеем $\sum\limits_{x'} \pi_t(x, x') = 0$, что соответствует (5.2.2).

Если процесс $\xi(t)$ стационарный, то матрица $\pi_t(x, x') = \pi(x, x')$ не зависит от времени t.

На данный марковский процесс могут быть распространены методы и результаты, изложенные в § 5.2.

Любая конечная совокупность случайных значений $\xi(t_1), \ldots, \xi(t_k)$ является для данного процесса совокупностью дискретных случайных величин. Поэтому их энтропию можно вычислить по формулам дискретной версии. В частности, если процесс задан на интервале $[0, T]$, энтропия начального значения $\xi(0)$ равна

$$H_{\xi(0)} = -\sum_{\xi(0)} P\left(\xi(0)\right) \ln P\left(\xi(0)\right). \tag{5.9.1}$$

Если, однако, мы хотим вычислить энтропию непрерывного множества значений ξ_0^T, то мы должны применить формулы обобщенной версии, т. е. теорию § 1.6 и 5.6.

Обозначим через $\{\tau_j\}$ те точки временной оси, т. е. те моменты времени, в которых происходят скачкообразные изменения состояния процесса. Вспомогательную меру $Q_1(\xi_0^T)$ определим как пуассоновскую меру (с единичной плотностью) для системы случайных точек $\{\tau_j\}$. Выделяя энтропию начального значения (5.9.1) по формуле

$$H_{\xi_0^T} = H_{\xi(0)} + H_{\xi_{0+0}^T \mid \xi(0)},$$

условную энтропию $H_{\xi_{0+0}^T \mid \xi(0)}$ или короче $H_{\xi_0^T \mid \xi(0)}$ определим формулами (1.7.17а), (1.7.17), (5.6.1) обобщенной версии. Тогда будем иметь

$$H^{P/Q_1}_{\xi_0^T} = -H_{\xi(0)} + H^{P/Q_1}_{\xi_0^T \mid \xi(0)}. \tag{5.9.1a}$$

Вследствие мультипликативного свойства меры Q_1 энтропия H^{P/Q_1} обладает свойством иерархической аддитивности

$$H^{P/Q_1}_{\xi_\alpha^\gamma | \xi(\alpha)} = H^{P/Q_1}_{\xi_\alpha^\beta | \xi(\alpha)} + H^{P/Q_1}_{\xi_\beta^\gamma | \xi_\alpha^\beta}.$$

Согласно условию марковости процесса последнюю формулу можно записать

$$H^{P/Q_1}_{\xi_\alpha^\gamma | \xi(\alpha)} = H^{P/Q_1}_{\xi_\alpha^\beta | \xi(\alpha)} + H^{P/Q_1}_{\xi_\beta^\gamma | \xi(\beta)}.$$

Так что

$$H^{P/Q_1}_{\xi_0^T} = -H_{\xi(0)} + H^{P/Q_1}_{\xi_0^{t_1} | \xi(0)} + H^{P/Q_1}_{\xi_{t_1}^{t_2} | \xi(t_1)} + \ldots +$$
$$+ H^{P/Q_1}_{\xi_{t_N}^T | \xi(t_N)} \qquad (0 < t_1 < \ldots < t_N < T). \tag{5.9.1б}$$

Это равенство является аналогом равенства (5.2.6), относящегося к случаю дискретного времени.

Как отмечалось в § 5.6, для стационарного процесса $\xi(t)$ энтропия $H^{P/Q_1}_{\xi_0^\tau | \xi_{-\infty}^0}$ пропорциональна τ. Вследствие условия Маркова это относится и к энтропии $H^{P/Q_1}_{\xi_0^\tau | \xi(0)}$. Формула (5.6.10) принимает вид

$$H^{P/Q_1}_{\xi_0^\tau | \xi(0)} = \tau h^{P/Q_1}. \tag{5.9.2}$$

Учитывая это, из (5.9.1а) или (5.9.2) имеем

$$H^{P/Q_1}_{\xi_0^T} = -H_{\xi(0)} + h^{P/Q_1} T.$$

Сравнивая эту формулу с (5.6.18), видим, что в данном случае $o_t(1) = 0$ и

$$2\Gamma = H_{\xi(0)} = -\sum_{\xi(0)} P(\xi(0)) \ln P(\xi(0)). \tag{5.9.3}$$

Соотношением (5.9.2) удобно воспользоваться при малых τ, чтобы вычислить удельную энтропию h^{P/Q_1} по аналогии с тем, как это было сделано в § 5.8 (см. (5.8.23), (5.8.24а)).

При малых τ можно пренебречь вероятностью выпадения на $(0, \tau]$ больше чем одной точки перехода. Тогда останутся лишь такие возможности: не произойдет перехода с вероятностью

$$1 + \pi(\xi(0), \xi(0))\tau + O(\tau^2)$$

или произойдет переход в состояние $x' \neq \xi(0)$ с вероятностью $\pi(\xi(0), x')\tau + O(\tau^2)$. Аналогично (5.8.25) записываем энтропию этих событий:

$$H^{P/Q_1}_{\xi_0^\tau}(\,|\xi(0)) =$$

$$= [1+\pi(\xi(0),\ \xi(0))\,\tau]\ln\{e^\tau[1+\pi(\xi(0),\ \xi(0))\,\tau]\} +$$
$$+ \sum_{\xi'\neq\xi(0)} \pi(\xi(0),\ \xi')\,\tau\ln[e^\tau\pi(\xi(0),\ \xi')] + O(\tau^2).$$

Усредняя по ξ (0) и совершая предельный переход

$$h^{P/Q_1} = \lim_{\tau\to 0}\frac{1}{\tau}H^{P/Q_1}_{\xi_0^\tau\,|\,\xi(0)},$$

отсюда получаем

$$h^{P/Q_1} = 1 + \sum_\xi [\pi(\xi,\ \xi) + \sum_{\xi'\neq\xi}\pi(\xi,\xi')\ln\pi(\xi,\xi')]\,P(\xi). \qquad (5.9.4)$$

Учитывая, что $\pi(\xi,\xi) = -\sum_{\xi'\neq\xi}\pi(\xi,\xi')$, в силу (5.9.4) имеем

$$h^{P/Q_1} = \sum_\xi P(\xi)\,(1 + \sum_{\xi'\neq\xi}[\pi(\xi,\xi')\ln\pi(\xi,\xi') - \pi(\xi,\xi')]).$$

Здесь, как и в (5.9.4), P (ξ) — стационарные вероятности, определяемые из уравнения

$$\frac{dP(\xi')}{dt} = \sum_\xi P(\xi)\,\pi(\xi,\xi') = 0. \qquad (5.9.5)$$

Последнее уравнение аналогично (5.2.7), а формула (5.9.4) является обобщением на случай непрерывного времени приведенной ранее формулы (5.2.8).

Пример. Пусть имеется процесс с двумя состояниями, характеризуемый дифференциальными вероятностями перехода

$$\begin{pmatrix}\pi(1,1) & \pi(1,2)\\ \pi(2,1) & \pi(2,2)\end{pmatrix} = \begin{pmatrix}-\mu & \mu\\ \nu & -\nu\end{pmatrix}. \qquad (5.9.6)$$

Уравнение (5.9.5) принимает вид

$$-\mu P(1) + \nu P(2) = 0,$$

так что стационарные вероятности оказываются такими:

$$P(1) = \frac{\nu}{\mu+\nu}, \qquad P(2) = \frac{\mu}{\mu+\nu} \qquad (5.9.7)$$

(ср. с (5.2.19)). По формуле (5.9.4) записываем удельную энтропию

$$h^{P/Q_1} = 1 + \frac{\mu\nu}{\mu+\nu}[\ln\mu + \ln\nu - 2].$$

Этот результат совпадает с (5.8.32) и это естественно, поскольку рассмотренный в примере 2 § 5.8 процесс становится марковским процессом с двумя состояниями, если распределения

$w_1(\sigma)$, $w_2(\sigma)$ экспоненциальные. При других видах распределений процесс с двумя состояниями будет немарковским, но станет марковским, если к $\xi(t)$ добавить еще одну переменную, а именно время $\tau = t - \tau_j$, прошедшее после момента τ_j последнего перескока. Комбинированный процесс $\{\xi(t), \sigma(t)\}$ будет марковским. Пространством состояний для него будут две полупрямых. Несмотря на такое усложнение, к нему применима почти без усложнений описанная выше теория.

Сформулируем в заключение этого параграфа указанное обобщение на произвольное пространство в более полном виде.

Пусть пространство состояний X марковского процесса произвольно. Дифференциальная вероятность перехода $\pi(x, x')$ такова, что из каждой точки $x \in X$ возможны переходы лишь в конечное или счетное множество точек. Тогда удельная энтропия определяется формулой

$$h^{P/Q} = \int P_{\text{ст}}(d\xi)\left[1 + \sum_{\xi' \neq \xi} \pi(\xi, \xi')(\ln \pi(\xi, \xi') - 1)\right], \tag{5.9.8}$$

где $P_{\text{ст}}(d\xi)$ — стационарное распределение, определяемое из уравнения

$$\int P(d\xi) \sum_{\xi' \in A} \pi(\xi, \xi') = 0$$

(A произвольно), обобщающего (5.9.5). В данном случае формула (5.9.1) может оказаться несправедливой, и для определения $H_{\xi(0)}$ может потребоваться мера Q.

Приведенные формулы (5.9.4), (5.9.8) можно применять не только в стационарном случае для вычисления средней по времени удельной энтропии. Они пригодны также вследствие марковских свойств процесса и в нестационарном случае, а именно, определяют плотность энтропии $h(t)$, рассчитанную на единицу времени, которая может зависеть от времени t. При этом усреднение в них следует производить с нестационарным распределением $P(d\xi)$.

5.10. Энтропия диффузионных марковских процессов

Пусть $\{x, (t), t \in [0, T]\}$ — соответствующий мере P диффузионный марковский процесс, характеризуемый сносом $a(x, t)$ и локальной дисперсией $b(x, t) > 0$. Он описывается уравнением Фоккера — Планка

$$\dot{p}_t(x) = -\frac{\partial}{\partial x}[a p_t(x) + \frac{1}{2}\frac{\partial^2}{\partial x^2}[b p_t(x)], \tag{5.10.1}$$

где

$$p_t(x)\,dx = \mathbf{P}\{x_t \in [x, x + dx]\}.$$

Чтобы определить для него энтропию $H_{x_0^T}$, нужно подобрать меру Q, относительно которой мера P является абсолютно непрерывной. При этом желательно, чтобы она была максимально простой.

Известно (например, Стратонович [4]), что такой мерой является мера, соответствующая диффузионному процессу с той же локальной дисперсией $b(x, t)$ и нулевым сносом, т. е. мера, для которой (5.10.1) заменяется уравнением

$$\dot{q}_t(x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}[b_t q(x)] \quad (q_t(x)\,dx = \mathbf{Q}\{x_t \in [x, x+dx]\}).$$

При этом производная Радона — Никодима имеет вид

$$\frac{P(dx_0^T)}{Q(dx_0^T)} = \frac{p(x(0))}{q(x(0))}\exp\left\{\int_0^T \frac{a(x(t), t)}{b(x(t), t)}\left[d^*x(t) - \right.\right.$$

$$\left.\left. - \frac{1}{2}a\left(x(t), t\right)dt\right]\right\}, \tag{5.10.2}$$

где стохастический интеграл $\int \ldots d^*x(t)$ понимается в смысле Ито. Постараемся теперь удовлетворить условию мультипликативности (5.6.4), чтобы была применима теория § 5.6. Для этого представим процесс $\{x(t), t \in [0, T]\}$ как процесс $\{\xi(0), \xi(t), t \in (0, T]\}$, где $\xi(0) = x(0)$, $\xi(t) = \dot{x}(t)$, $t > 0$, и потребуем, чтобы $b(x, t) = b(t)$ не зависела от x. Тогда мера Q будет соответствовать гауссовому дельта-коррелированному процессу:

$$\int \xi(t)\,dQ = 0, \qquad \int \xi(t)\,\xi(t')\,dQ = b(t)\,\delta(t-t'),$$

так что условие мультипликативности (5.6.4) будет выполнено. Используя (5.10.2), нетрудно вычислить энтропию (5.6.1), а также в предположении стационарности удельную энтропию

$$h^{P/Q} = \frac{1}{\tau}H^{P/Q}_{\xi_t^{t+\tau}\,|\,\xi_0^t} = \frac{1}{\tau}H^{P/Q}_{x_t^{t+\tau}\,|\,x(t)}.$$

При этом энтропию $H^{P/Q}_{x_t^{t+\tau}|x(t)}$ удобно вычислять приближенно, пользуясь малостью τ, а затем совершить предельный переход

$$h^{P/Q} = \lim_{\tau\to 0}\frac{1}{\tau}H^{P/Q}_{x_t^{t+\tau}\,|\,x(t)}. \tag{5.10.3}$$

При малых τ из (5.10.2) имеем

$$\mathbf{M}\ln\frac{P\left[dx_t^{t+\tau}\,|\,x(t)\right]}{Q\{d[x(\sigma)-x(t)],\ \sigma\in(t, t+\tau]\}} = \mathbf{M}\int_t^{t+\tau}\frac{a(x(\sigma))}{b}\left[d^*x(\sigma) - \right.$$

$$\left. - \frac{1}{2}a\left(x(\sigma)\right)d\sigma\right] \approx \mathbf{M}\frac{a(x(t))}{b}\left[x(t+\tau) - x(t) - \right.$$

$$-\frac{1}{2} a\left(x(t)\right)\tau\Big] + o(\tau).$$

Для вычисления $H^{P/Q_1}_{x^{t+\tau}_t \mid x(t)}$ проводим усреднение

$$H^{P/Q}_{x_t^{t+\tau} \mid x(t)} = \frac{a(x(t))}{b}\Big\{\mathbf{M}\left[x(t+\tau) - x(t)\right] -$$

$$-\frac{1}{2} a\left(x(t)\right)\tau\Big\} + o(\tau).$$

Учитывая, что в соответствии с определением сноса a справедливо соотношение

$$\mathbf{M}\ [x(t+\tau) - x(t) \mid x(t)] = a(x(t))\tau + o\ (\tau),$$

будем иметь

$$\frac{1}{\tau} H^{P/Q}_{x_t^{t+\tau} \mid x(t)} = \frac{1}{2}\mathbf{M}\ \frac{a(x(t))^2}{b} + o(1).$$

Подставляя это выражение в (5.10.3), находим плотность энтропии

$$h^{P/Q} = \frac{1}{2b}\mathbf{M}\left[a\left(x(t)\right)\right]^2, \qquad (5.10.4)$$

которая не зависит от t в стационарном случае.

Выше предполагалось, что a и b не зависят от времени t в соответствии с условием стационарности. Если бы это условие не было выполнено, мы получили бы описанным способом зависящую от времени энтропийную плотность

$$h^{P/Q}(t) = \frac{1}{2}\mathbf{M}\frac{a(x(t),\ t)^2}{b\ (x(t),\ t)} \qquad (5.10.5)$$

(условие независимости b от x здесь также может быть не выполнено). Через нее получаемая из (5.10.2) полная энтропия

$$H^{P/Q}_{x_0^T} = \mathbf{M}\ln\frac{P\ (dx_0^T)}{Q\ (dx_0^T)}$$

выражается интегрированием

$$H^{P/Q}_{x_0^T} = H^{P/Q}_{x(0)} + \int_0^T H_0^{P/Q}(t)\,dt. \qquad (5.10.6)$$

В стационарном же случае при $b = \text{const}$ плотность энтропии может быть вычислена по формуле (5.10.4) при помощи стационарной плотности распределения $p_{\text{ст}}(x)$, удовлетворяющей стационарному уравнению Фоккера — Планка

$$-\frac{d}{dx}[a(x)p_{\text{ст}}(x)]+\frac{b}{2}\frac{d^2}{dx^2}p_{\text{ст}}(x)=0.$$

Если ввести потенциальную функцию

$$U(x)=-\int\limits_c^x a(x)\,dx, \tag{5.10.7}$$

то стационарное распределение можно записать

$$p_{\text{ст}}(x)=\frac{1}{N}e^{-2U(x)/b}, \qquad N=\int e^{-2U(x)/b}\,dx, \tag{5.10.8}$$

и из (5.10.4) будем иметь

$$h^{P/Q}=\frac{1}{2bN}\int e^{-2U(x)/b}\left[\frac{dU(x)}{dx}\right]^2 dx.$$

Интегрированием по частям (при учете исчезновения плотности $p_{\text{ст}}(x)$ на границах, например, при $t\to\pm\infty$) эту формулу можно привести к виду

$$h^{P/Q}=\frac{1}{4N}\int e^{-2U(x)/b}\frac{d^2U(x)}{dx^2}dx=-\frac{1}{4}\mathbf{M}\frac{da(x)}{dx}. \tag{5.10.9}$$

Пример. Пусть функция $a(x)$ линейна: $a(x)=-\beta x+\gamma$. Чтобы процесс был стационарным, необходима положительность $\beta>0$. Функция (5.10.7) имеет вид

$$U(x)=\frac{1}{2}\beta x^2-\gamma x,$$

и распределение (5.10.8) является гауссовым:

$$p_{\text{ст}}(x)=\sqrt{\frac{\beta}{\pi b}}\exp\left[-\frac{\beta}{b}\left(x-\frac{\gamma}{\beta}\right)^2\right], \qquad N=\sqrt{\frac{\pi b}{\beta}}.$$

Применение формулы (5.10.9) дает

$$h^{P/Q}=\beta/4. \tag{5.10.10}$$

Данный процесс является, как известно, гауссовым процессом, имеющим спектральную плотность $S(\omega)=2b/(\omega^2+\beta^2)$. Поэтому он совпадает с процессом, рассмотренным в § 5.7 (пример 2). Результат (5.10.10), естественно, совпадает с соответствующим результатом (5.7.23), полученным на основе теории гауссовых процессов.

Приведенные в этом параграфе результаты допускают обобщение на многомерный случай, когда имеется многокомпонентный марковский процесс $\{x(t)\}=\{x_1(t), \ldots, x_l(t)\}$. Пусть он характеризуется сносами $a_\rho(x, t)$, $\rho = 1, \ldots, l$, и матрицей локальных дисперсий $b_{\rho\sigma}(x, t)$, $\rho, \sigma = 1, \ldots, l$. Тогда выбирая меру Q, описанную в формулировке теоремы 4.1 монографии Стратоновича [4], имеем производную Радона — Никодима

$$\frac{P(dx_0^T \mid x(0))}{Q(dx_0^T \mid x(0))} = \exp\left\{\int_0^T \sum_{\rho', \sigma'} a_{\rho'} b_{\rho'\sigma'}^{-1}\left(d^* x_{\sigma'} - \frac{1}{2} a_{\sigma'}\, dt\right)\right\}, \quad (5.10.11)$$

что служит обобщением формулы (5.10.2). Здесь $b_{\rho'\sigma'}^{-1}$ — матрица, обратная невырожденной подматрице $b_{\rho'\sigma'}$, $\rho', \sigma' = 1, \ldots, l' \leqslant l$, матрице локальных дисперсий $b_{\rho\sigma}$, $\rho, \sigma = 1, \ldots, l$.

Тем же приемом, что и в одномерном случае, получаем из (5.10.11) плотность энтропии

$$h^{P/Q}(t) = \frac{1}{2} \mathbf{M} \sum_{\rho', \sigma'} a_{\rho'}(x(t), t)\, b_{\rho'\sigma'}^{-1}(x(t), t)\, a_{\sigma'}(x(t), t). \quad (5.10.12)$$

В частности, если $l \times l$-матрица $b_{\rho\sigma}$ невырожденна, то суммирование в (5.10.12) по ρ' и σ' нужно проводить от 1 до l, а мера Q соответствует той же матрице локальных дисперсий $b_{\rho\sigma}$, но нулевому вектору сносов.

В стационарном случае усреднение в (5.10.12) означает интегрирование со стационарной плотностью распределения $p_{\text{ст}}(x)$, удовлетворяющей стационарному уравнению Фоккера — Планка

$$-\sum_\rho \frac{\partial}{\partial x_\rho}(a_\rho p_{\text{ст}}) + \frac{1}{2} \sum_{\rho, \sigma} \frac{\partial^2}{\partial x_\rho \partial x_\sigma}(b_{\rho\sigma} p_{\text{ст}}) = 0.$$

В важном частном случае стационарное распределение таково, что потоки вероятности обращаются в нуль:

$$G_\rho = a_\rho p_{\text{ст}} - \frac{1}{2} \sum_\sigma \frac{\partial}{\partial x_\sigma}(b_{\rho\sigma} p_{\text{ст}}) = 0.$$

В этом случае из (5.10.12) имеем

$$h^{P/Q}(t) = \frac{1}{4} \sum_{\rho', \sigma', \pi} \int a_{\rho'} b_{\rho'\sigma'}^{-1} \frac{\partial}{\partial x_\pi}[b_{\sigma'\pi} p_{\text{ст}}]\, dx_1 \ldots dx_l.$$

Применяя формулу Грина и учитывая исчезновение интеграла по границе, получаем

$$h^{P/Q}(t) = -\frac{1}{4} \sum_{\rho', \sigma', \pi} \int b_{\sigma'\pi} p_{\text{ст}} \frac{\partial}{\partial x_\pi}[a_{\rho'} b_{\rho'\sigma'}^{-1}]\, dx_1 \ldots dx_l =$$

$$= -\frac{1}{4} \mathbf{M}\, b_{\sigma'\pi} \frac{\partial}{\partial x_\pi}[a_{\rho'} b_{\rho'\sigma'}^{-1}].$$

Если матрица $b_{\rho\sigma}$ невырожденна и не зависит от x, то имеет место следующая простая формула:

$$h^{P/Q} = -\frac{1}{4}\sum_{\rho,\sigma,\pi} \mathbf{M}\, b_{\sigma\pi} b_{\sigma\rho}^{-1} \frac{\partial a_\rho}{\partial x_\pi} =$$

$$= -\frac{1}{4}\mathbf{M}\sum_\rho \frac{\partial a_\rho}{\partial x_\rho} \equiv -\frac{1}{4}\mathbf{M}\operatorname{div} a,$$

которая является многомерным обобщением формулы (5.10.9).

5.11. Энтропия комбинированного марковского процесса, условного процесса и части компонент марковского процесса

1. На случай непрерывного времени могут быть обобщены результаты и методы, изложенные в § 5.3. Предполагается, что совокупный процесс $\{\xi(t)\} = \{x(t),\ y(t)\}$ является марковским. Теория марковских процессов позволяет вычислять энтропию $H^{P/Q}_{\xi_a^T} = H^{P/Q}_{x_a^T y_a^T}$, где Q — вероятностная мера, относительно которой мера P является абсолютно непрерывной. Меру Q удобно подбирать таким образом, чтобы процессы $\{x\ (t)\}$ и $\{y\ (t)\}$ были для этой меры независимыми:

$$Q\left(d\xi_a^T\right) = Q_1\left(dx_a^T\right) Q_2\left(dy_a^T\right)$$

и марковскими, т. е. чтобы

$$Q\left(dx_t^{t+\tau}\,|\,x_a^t\, y_a^t\right) = Q\left(dx_t^{t+\tau}\,|\,x(t)\right),\quad Q\left(dy_t^{t+\tau}\,|\,x_a^t\, y_a^t\right) =$$
$$= Q\left(dy_t^{t+\tau}\,|\,y(t)\right).$$

Тогда будут справедливы соотношения

$$H^{P/Q}_{x_a^T y_a^T} = H^{P/Q_2}_{y_a^T} + H^{P/Q_1}_{x_a^T\,|\,y_a^T}, \tag{5.11.1}$$

$$H^{P/Q}_{x_a^T y_a^T} = H^{P/Q_1}_{x_a^T} + H^{P/Q_2}_{y_a^T\,|\,x_a^T}, \tag{5.11.2}$$

$$H^{P/Q}_{x_t^{t+\tau}\, y_t^{t+\tau}\,|\,x_a^t\, y_a^t} = H^{P/Q}_{x_t^{t+\tau}\, y_t^{t+\tau}\,|\,x(t)\, y(t)}, \tag{5.11.3}$$

аналогичные соотношениям (5.3.1), (5.3.6), (5.2.5) дискретной версии.

В случае непрерывного времени удобно ввести энтропийные плотности

$$h^{P/Q}_{xy}(t) = \lim_{\tau\to 0}\frac{1}{\tau} H^{P/Q}_{x_t^{t+\tau}\, y_t^{t+\tau}\,|\,x(t)\, y(t)}, \tag{5.11.4}$$

$$h_y^{P/Q_2}(t)=\lim_{\tau\to 0}\frac{1}{\tau}H^{P/Q_2}_{y_t^{t+\tau}\,|\,y_a^t},\tag{5.11.5}$$

$$h_{x|y}^{P/Q_1}(t)=\lim_{\tau\to 0}\frac{1}{\tau}\left[H^{P/Q_1}_{x_a^{t+\tau}\,|\,y_a^{t+\tau}}-H^{P/Q_1}_{x_a^t\,|\,y_a^t}\right].\tag{5.11.6}$$

Полагая в (5.11.1) сначала $T=t+\tau$, а затем $T=t$, взяв разность этих выражений и используя свойство аддитивности типа (1.7.18), получаем соотношение

$$H^{P/Q}_{x_t^{t+\tau}\,y_t^{t+\tau}\,|\,x_a^t\,y_a^t}=H^{P/Q_2}_{y_t^{t+\tau}\,|\,y_a^t}+H^{P/Q_1}_{x_a^{t+\tau}\,|\,y_a^{t+\tau}}-H^{P/Q_1}_{x_a^t\,|\,y_a^t}.\tag{5.11.7}$$

Поделив (5.11.7) на τ и перейдя к пределу $\tau\to 0$, получим соотношение

$$h_{xy}^{P/Q}(t)=h_y^{P/Q_2}(t)+h_{x|y}^{P/Q_1}(t).\tag{5.11.8}$$

Аналогичное соотношение

$$h_{xy}^{P/Q}(t)=h_x^{P/Q_1}(t)+h_{y|x}^{P/Q_2}(t)\tag{5.11.8a}$$

справедливо, очевидно, и для другой пары энтропийных плотностей

$$\begin{aligned}h_x^{P/Q}(t)&=\lim_{\tau\to 0}\frac{1}{\tau}H^{P/Q_1}_{x_t^{t+\tau}\,|\,x_a^t},\\ h_{y|x}^{P/Q_2}(t)&=\lim_{\tau\to 0}\frac{1}{\tau}\left[H^{P/Q_2}_{y_a^{t+\tau}\,|\,x_a^{t+\tau}}-H^{P/Q_2}_{y_a^t\,|\,x_a^t}\right].\end{aligned}\tag{5.11.9}$$

В стационарном случае плотность $h_{xy}^{P/Q}$, как видно из (5.11.4), не зависит от t. Другие энтропийные плотности при этом целесообразно определить с помощью дополнительного предельного перехода $a\to-\infty$. В этом случае энтропии будут строго пропорциональны τ и предельный переход $\tau\to 0$ станет излишним. Формулы (5.11.5), (5.11.6) заменяются следующими:

$$h_y^{P/Q_2}(t)=\frac{1}{\tau}H^{P/Q_2}_{y_t^{t+\tau}\,|\,y_{-\infty}^t},$$

$$h_{x|y}^{P/Q_1}(t)=\frac{1}{\tau}\lim_{a\to-\infty}\left[H^{P/Q_1}_{x_a^{t+\tau}\,|\,y_a^{t+\tau}}-H^{P/Q_1}_{x_a^t\,|\,y_a^t}\right]$$

(и аналогично для другой пары h^{P/Q_1}_x, $h^{P/Q_2}_{y/x}$). При этом соотношения (5.11.8), (5.11.9) сохраняют свое значение. Все эти энтропийные плотности в стационарном случае будут постоянными. Можно доказать, что они совпадают с удельными энтропиями, т. е. доказать равенства:

$$h_y^{P/Q_2}=\lim_{t\to\infty}\frac{1}{t}H^{P/Q_2}_{y_0^t},\qquad h_{x|y}^{P/Q_1}=\lim_{t\to\infty}\frac{1}{t}H^{P/Q_1}_{x_0^t\,|\,y_0^t}.$$

Можно доказать кроме того, что для плотности $h_{x|y}^{P/Q_1}$ (и аналогично для $h_{y|x}^{P/Q_2}$) в стационарном случае справедлива формула

$$h_{x|y}^{P/Q_1} = \frac{1}{\tau} H^{P/Q_1}_{x_0^\tau | x(0)\, y_0^\infty} = \frac{1}{\tau} H^{P/Q_1}_{x_0^\tau | x(\tau)\, y_{-\infty}^\tau} \quad (\tau > 0 \text{ любое}).$$

Все эти утверждения распространяют на случай непрерывного времени соответствующие утверждения, доказанные в § 5.3.

Обобщая методы § 5.3, можно вычислить энтропию $H^{P/Q_2}_{y_0^T}$ или ее плотность h_y^{P/Q_2} для части компонент $\{y\}$ марковского процесса ξ и энтропию $H^{P/Q_1}_{x_0^T | y_0^T}$ или $h_{x|y}^{P/Q_1}$ условного марковского процесса $\{x(t)\}$, а также аналогичные энтропии $H^{P/Q_1}_{x_0^T}$, h_x^{P/Q_1} и $H^{P/Q_2}_{y_0^T | x_0^T}$, $h_{y|x}^{P/Q_2}$.

Рассмотрим энтропию $H^{P/Q_2}_{y_t^{t+\tau} | y_0^t}$. Ее можно представить в таком виде:

$$H^{P/Q_2}_{y_t^{t+\tau} | y_0^t} = \mathbf{M} \ln \frac{P\left[dy_t^{t+\tau} | y_0^t\right]}{Q_2\left[dy_t^{t+\tau} | y(t)\right]}. \tag{5.11.10}$$

Тогда имеем

$$\frac{P\left[dy_t^{t+\tau} | y_0^t\right]}{Q_2\left[dy_t^{t+\tau} | y(t)\right]} = \mathbf{M}\left\{\frac{P\left[dy_t^{t+\tau} | x(t),\ y_0^t\right]}{Q_2\left[dy_t^{t+\tau} | y(t)\right]} \,\middle|\, y_0^t\right\}.$$

В силу условия Маркова

$$P\left[dy_t^{t+\tau} | x(t),\ y_0^t\right] = P\left[dy_t^{t+\tau} | x(t),\ y(t)\right]$$

указанное равенство можно записать так:

$$\frac{P\left[dy_t^{t+\tau} | y_0^t\right]}{Q_2\left[dy_t^{t+\tau} | y(t)\right]} = \int \frac{P\left[dy_t^{t+\tau} | x(t),\, y(t)\right]}{Q_2\left[dy_t^{t+\tau} | y(t)\right]} W\left(dx(t)\right), \tag{5.11.11}$$

где обозначено

$$W\left[dx(t)\right] = P\left[dx(t) | y_0^t\right].$$

Подставляя (5.11.11) в (5.11.10), получаем формулу

$$H^{P/Q_2}_{y_t^{t+\tau} | y_0^t} = \mathbf{M} \ln \left[\int \frac{P\left[dy_t^{t+\tau} | x(t),\, y(t)\right]}{Q_2\left[dy_t^{t+\tau} | y(t)\right]} W(dx)\right]. \tag{5.11.12}$$

Если апостериорную меру W трактовать как распределение в комбинированном пространстве $\xi = (x,\ y)$ (см. (5.3.17)), то формула

(5.11.12) запишется в виде

$$H^{P/Q_2}_{y_t^{t+\tau} \mid y_0^t} = \mathbf{M} \ln \left[\int \frac{P\left(dy_t^{t+\tau} \mid \xi\right)}{Q_2\left(dy_t^{t+\tau} \mid y\right)} W(d\xi) \right],$$

или

$$H^{P/Q_2}_{y_t^{t+\tau} \mid y_0^t} = \mathbf{M} \int P\left(dy_t^{t+\tau} \mid \xi\right) W(d\xi) \ln \left[\frac{P\left(dy_t^{t+\tau} \mid \xi\right)}{Q_2\left(dy_t^{t+\tau} \mid y\right)} W(d\xi) \right], \tag{5.11.12a}$$

где усреднение $\mathbf{M}$ проводится только по случайным переменным $W(\cdot)$, образующим вторичный апостериорный W-процесс, являющийся марковским процессом с известными вероятностями перехода.

Формулы (5.11.12), (5.11.12а) являются обобщением формул (5.3.15), (5.3.16) дискретной версии. Они справедливы при любых τ, но для вычисления h_y^{P/Q_2} ими удобнее пользоваться при малых τ. Это будет видно из рассматриваемых ниже частных случаев.

2. Пусть теперь $\{x(t)\}$ — марковский процесс с дискретным числом состояний, подобный процессу, рассмотренному в § 5.9. Он характеризуется введенной в § 5.9 дифференциальной матрицей перехода $\pi_t(x, x')$, определяющей вероятности перехода $P(\xi(t+\Delta) \mid \xi(t))$. Выбирая в качестве Q_1 пуассоновскую меру для точек перехода, получаем плотность $h_x^{P/Q_1}(t)$ энтропии $H^{P/Q_1}_{x_0^T}$

$$h_x^{P/Q_1}(t) = \sum_{x_t} P\left(x(t)\right) \Big\{1 + \sum_{x' \neq x_t} \left[\ln \pi_t\left(x(t), x'\right) - 1\right] \pi_t\left(x(t), x'\right)\Big\}. \tag{5.11.13}$$

Процесс $\{y(t)\} = \{y_1(t), \ldots, y_l(t)\}$ пусть строится следующим образом. Задан зависящий не только от $y(t)$, но и от $x(t)$ вектор сносов $a_\rho(x(t), y(t), t)$, а также матрица локальных дисперсий $b_{\rho\sigma}(y(t), t)$. Последнюю предполагаем невырожденной и не зависящей от $x(t)$.

Тогда при фиксированной реализации $\{x(t)\}$ процесс $y(t)$ будет диффузионным процессом, рассмотренным в § 5.10. Применяя полученные там результаты, мы можем найти плотность $h^{P/Q_2}_{y \mid x}(t)$ энтропии $H^{P/Q_2}_{y_0^T \mid x_0^T}$. В качестве меры Q_2 выбираем меру диффузионного процесса с нулевыми сносами и с той же матрицей локальных дисперсий $b_{\rho\sigma}(y, t)$. В соответствии с (5.10.12) имеем

$$h_y^{P/Q_2}\left(t \mid x_0^T\right) = \frac{1}{2} \sum_{\rho, \sigma} \int a_\rho\left(x(t), y(t), t\right) b_{\rho\sigma}^{-1}\left(y(t), t\right) \times a_\sigma\left(x(t), y(t), t\right) P\left(dy(t) \mid x_0^T\right). \tag{5.11.14}$$

Для получения плотности $h_{y|x}^{P/Q_2}$ энтропии $H_{y_0^T|x_0^T}^{Q/P_2}$ остается в (5.11.14) произвести дополнительное усреднение по x_0^T:

$$h_{y|x}^{P/Q_2}(t) = \frac{1}{2}\sum_{\rho,\sigma} a_\rho\left(x(t),\ y(t),\ t\right) b_{\rho\sigma}^{-1}\left(y(t),\ t\right) \times$$
$$\times a_\sigma\left(x(t),\ y(t),\ t\right) P\left(dy(t)\ dx(t)\right). \tag{5.11.15}$$

В силу аддитивности (5.11.2), (5.11.9) тем самым найдена энтропийная плотность $h_{xy}^{P/Q}$ для комбинированного марковского процесса $\{\xi(t)\} = \{x(t),\ y(t)\}$. Именно, в (5.11.9) следует подставить выражения (5.11.13), (5.11.15).

Другие энтропийные плотности h_y^{P/Q_2} и $h_{x|y}^{P/Q_1}$ согласно (5.11.1) связаны аналогичным соотношением (5.11.8). Поэтому для завершения вычисления всех плотностей остается вычислить одну из них.

Описанный выше процесс $\{y(t)\}$ (при нефиксированной реализации $\{x(t)\}$) является немарковским процессом. Соответствующую ему энтропию вычислим, используя формулу (5.11.12).

При малых τ из (5.10.11) имеем

$$P\left[dy_t^{t+\tau}\middle|\, x(t),\ y(t)\right]/Q_2\left[dy_t^{t+\tau}\middle|\, y(t)\right] = \exp\left\{\sum_{\rho,\sigma} a_\rho b_{\rho\sigma}^{-1} \times\right.$$
$$\left.\times\left[y_\sigma(t+\tau) - y_\sigma(t) - \frac{\tau}{2}a_\sigma\right] + o(\tau)\right\} = 1 + \sum_{\rho,\sigma} a_\rho b_{\rho\sigma}^{-1} \times$$
$$\times\left[y_\sigma(t+\tau) - y_\sigma(t) - \frac{\tau}{2}a_\sigma\right] + \frac{1}{2}\sum a_\rho b_{\rho\sigma}^{-1}\left[y_\sigma(t+\tau) -\right.$$
$$\left.- y_\sigma(t)\right]\left[y_\pi(t+\tau) - y_\pi(t)\right] b_{\pi\nu}^{-1} a_\nu + o(\tau).$$

Подставляя это выражение в (5.11.11) и для краткости обозначая $\sum\limits_{x(t)} \ldots\ W_t(x_t) = \mathbf{M}_{\mathrm{ps}} \ldots$, будем иметь

$$P\left[dy_t^{t+\tau}\middle|\, y_0^t\right]/Q_2\left[y_t^{t+\tau}\middle|\, y(t)\right] = 1 + \sum_{\rho,\sigma}(\mathbf{M}_{\mathrm{ps}}\, a_\rho)\, b_{\rho\sigma}^{-1} \times$$
$$\times\left[y_\sigma(t+\tau) - y_\sigma(t)\right] + \frac{1}{2}\sum\left\{b_{\rho\sigma}^{-1}\left[y_\sigma(t+\tau) - y_\sigma(t)\right] \times\right.$$
$$\left.\times\left[y_\pi(t+\tau) - y_\pi(t)\right] b_{\pi\nu}^{-1} - \tau b_{\rho\nu}^{-1}\right\}\mathbf{M}_{\mathrm{ps}}(a_\nu a_\rho) + o(\tau).$$

Согласно (5.11.10), (5.11.12) следует взять логарифм от последнего выражения. При его логарифмировании нужно оставить лишь следующие члены:

$$\ln\frac{P\left[dy_t^{t+\tau}\middle|\, y_0^t\right]}{Q_2\left[dy_t^{t+\tau}\middle|\, y(0)\right]} = (\mathbf{M}_{\mathrm{ps}}\, a_\rho)\, b_{\rho\sigma}^{-1}\left[y_\sigma(t+\tau) - y_\sigma(t)\right] +$$

$$+\frac{1}{2}\left\{b_{\rho\sigma}^{-1}\left[y_\sigma(t+\tau)-y_\sigma(t)\right]\left[y_\pi(t+\tau)-y_\pi(\tau)\right]b_{\pi\nu}^{-1}-b_{\rho\nu}^{-1}\right\}\times$$
$$\times \mathbf{M}_{ps}(a_\nu a_\rho)-\frac{1}{2}\left\{(\mathbf{M}_{ps}\,a_\rho)\,b_{\rho\sigma}^{-1}\left[y_\sigma(t+\tau)-y_\sigma(t)\right]\right\}^2+o(\tau). \tag{5.11.16}$$

Усреднение его будем проводить в несколько этапов: сначала усредним по $y_t^{t+\tau}$ при фиксированных $x(t)$, y_0^t, потом по $x(t)$ (с весом $W_t\ (x(t))$) при фиксированных y_0^t и в последнюю очередь по y_0^t или, что то же самое, по W. На первом этапе усреднения нужно учесть формулы

$$\mathbf{M}\left\{y_\sigma(t+\tau)-y_\sigma(t)\,|\,x(t),\ y(t)\right\}=a_\sigma\left(x(t),\ y(t),\ t\right)\tau+o(\tau),$$
$$\mathbf{M}\left\{\left[y_\sigma(t+\tau)-y_\sigma(t)\right]\left[y_\pi(t+\tau)-y_\pi(t)\right]\,|\,x(t),\ y(t)\right\}=$$
$$=b_{\sigma\pi}\left(x(t),\ y(t),\ t\right)\tau+o(\tau),$$

так что

$$\frac{1}{\tau}\mathbf{M}\left[\ln\frac{P\left[dy_t^{t+\tau}\,|\,y_0^t\right]}{Q_2\left[dy_t^{t+\tau}\,|\,y(0)\right]}\,\middle|\,x(0),\ y_0^t\right]=(\mathbf{M}_{ps}\,a_\rho)\,b_{\rho\sigma}^{-1}\,a_\sigma-$$
$$-\frac{1}{2}(\mathbf{M}_{ps}\,a_\rho)\,b_{\rho\sigma}^{-1}(\mathbf{M}_{ps}\,a_\sigma)+o(1).$$

Дальнейшее усреднение приведет к результату

$$\frac{1}{\tau}H^{P/Q_2}_{y_t^{t+\tau}|\,y_0^t}=\frac{1}{2}\mathbf{M}(\mathbf{M}_{ps}\,a_\rho)\,b_{\rho\sigma}^{-1}(\mathbf{M}_{ps}\,a_\sigma)+o(1).$$

Совершая предельный переход $\tau\to 0$, найдем

$$h_y^{P/Q_2}(t)=\frac{1}{2}\mathbf{M}\sum_{\rho,\ \sigma}(\mathbf{M}_{ps}\,a_\rho)\,b_{\rho\sigma}^{-1}(\mathbf{M}_{ps}a_\sigma)=$$
$$=\frac{1}{2}\int P\left(dW_t,\ dy(t)\right)\sum_{x,\ x'}\sum_{\rho,\ \sigma}W_t(x)\,a_\rho\left(x,\ y(t),\ t\right)\times$$
$$\times\ b_{\rho\sigma}^{-1}\left(y(t),\ t\right)a_\sigma\left(x',\ y(t),\ t\right)W_t(x'). \tag{5.11.17}$$

Найти входящее сюда распределение $P(dW_t,\ dy(t))$ помогает то обстоятельство, что переменные $\{W_t,\ y_t\}$ образуют марковский процесс, названный в монографии Стратоновича [4] вторичным апостериорным W-процессом. Там найдено уравнение Фоккера — Планка, которому удовлетворяет плотность распределения вероятностей $p_t(W,\ y)$ процесса $\{W_t,\ y(t)\}$. Оно имеет вид

$$\frac{dp_t}{dt}=\frac{\partial}{\partial W(x')}\left[\pi_{xx'}\,W(x)\,p_t\right]-\frac{\partial}{\partial y_\rho}\left[(\mathbf{M}_{ps}\,a_\rho)\,p_t\right]+$$
$$+\frac{1}{2}\sum\frac{\partial^2}{\partial W(x)\,\partial W(x')}\left\{W(x)\left[a_\rho(x)-\mathbf{M}_{ps}\,a_\rho\right]b_{\rho\sigma}^{-1}\times\right.$$

$$\times\left[a_\sigma(x')-\mathbf{M}_{ps}\,a_\sigma\right]W(x')\,p_t\Big\}+\sum\frac{\partial^2}{\partial W(x)\,\partial y_\rho}\Big\{W(x)\times$$

$$\times\left[a_\rho(x)-\mathbf{M}_{ps}\,a_\rho\right]p_t\Big\}+\frac{1}{2}\sum\frac{\partial^2}{\partial y_\rho\partial y_\sigma}[a_{\rho\sigma}\,p_t]. \qquad (5.11.18)$$

Указанные результаты справедливы как в стационарном, так и в нестационарном случае. В стационарном случае функции $\pi_t(x, x')$, $a_\rho(x, y, t)$, $b_{\rho\sigma}(y, t)$ не зависят от времени. Плотность энтропии h_y^{P/Q_2} при этом постоянна и вычисляется по формуле (5.11.17) при помощи стационарного распределения $P_{\text{ст}}(dW_t, dy(t))$, являющегося решением стационарного уравнения (5.11.18), в котором следует положить $dp_t/dt = 0$.

По аналогии с (5.3.17) вместо $W(x)$ можно рассматривать $W(\cdot)$ как распределение $W_t(dxdy)$ во всем «фазовом пространстве» $\xi = (x, y)$ марковского процесса (такой процесс $\{W_t(.)\}$ будет марковским). Тогда формулу (5.11.17) можно будет записать следующим образом:

$$h_y^{P/Q_2}(t)=\frac{1}{2}\int P(dW_t)\int\int\sum_{\rho,\sigma}W_t(d\xi)\,a_\rho(\xi, t)\times$$

$$\times\, b_{\rho\sigma}^{-1}(y, t)\,a_\sigma(\xi', t)\,W_t(d\xi').$$

Итак, удельная энтропия h_y^{P/Q_2}, а значит и энтропия $h_{x|y}^{P/Q_2} = h_{xy}^{P/Q_2} - h_y^{P/Q_2}$ вычисляется методами, использующими результаты теории условных марковских процессов.

Пример. Пусть $\{x(t)\}$ — процесс с двумя состояниями и матрицей перехода (5.9.6). В § 5.9 была найдена соответствующая ему удельная энтропия

$$h_x^{P/Q_1}=1+\frac{\mu\nu}{\mu+\nu}\ln\frac{\mu\nu}{e^2}. \qquad (5.11.19)$$

Процесс $\{y(t)\}$ определим как одномерный диффузионный процесс с постоянной локальной дисперсией b и с не зависящими от y и t, но зависящими от x сносами $a(x)$.

Тогда формула (5.11.15) дает

$$h_{y|x}^{P/Q_2}=\frac{1}{2b}\left[\mathbf{P}(x=1)\,a^2(1)+\mathbf{P}(x=2)\,a^2(2)\right],$$

или

$$h_{y|x}^{P/Q_2}=\frac{\nu a^2(1)+\mu a^2(2)}{2b(\mu+\nu)}, \qquad (5.11.20)$$

если учесть (5.9.7).

Сумма выражений (5.11.19), (5.11.20) дает удельную энтропию $h_{xy}^{P/Q}$ комбинированного процесса.

Перейдем к вычислению удельной энтропии h_y^{P/Q_2} для данного примера. Вводя переменную $z_t = W_t(1) - W_t(2)$, средний снос $\mathbf{M}_{\text{ps}}\,a$ можно представить в виде

$$\mathbf{M}_{\text{ps}}\,a = W_t(1)\,a(1) + W_t(2)\,a(2) = \frac{a(1)+a(2)}{2} + \frac{a(1)-a(2)}{2}\,z.$$

Подставляя это выражение в (5.11.17), имеем

$$h_y^{P/Q_2} = \frac{1}{8b}\left[a(1)+a(2)\right]^2 + \frac{1}{4b}\,\mathbf{M}\left\{\left[a^2(1)-a^2(2)\right]z + \right.$$
$$\left. + \frac{1}{2}\left[a(1)-a(2)\right]^2 z^2\right\}. \tag{5.11.21}$$

Нетрудно понять, что усреднение апостериорных вероятностей $W_t(x) = \mathbf{P}\,[x_t = x\,|y_0^t]$ приводит к априорным вероятностям $\mathbf{P}\,[x_t = x]$, которые в стационарном случае имеют вид (5.9.7). Поэтому

$$\mathbf{M}z = \mathbf{M}\left[W_t(1) - W_t(2)\right] = \mathbf{P}\,[x=1] - \mathbf{P}\,[x=2) = \frac{\nu-\mu}{\mu+\nu}$$

и формулу (5.11.21) можно преобразовать к виду

$$h_y^{P/Q_2} = \frac{1}{8b}\left[a(1)+a(2)\right]^2 + \frac{(\nu-\mu)\left[a^2(1)-a^2(2)\right]}{4b\,(\mu+\nu)} +$$
$$+ \frac{1}{8b}\left[a(1)-a(2)\right]^2 \int_{-1}^{1} z^2\,p_{\text{ст}}(z)\,dz$$

или

$$h_y^{P/Q_2} = \frac{\nu a^2(1)+\mu a^2(2)}{2b\,(\mu+\nu)} - \frac{\left[a(1)-a(2)\right]^2}{8b}\int_{-1}^{1}(1-z^2)\,p_{\text{ст}}(z)\,dz. \tag{5.11.22}$$

Процесс $\{z_t\}$ в данном случае является марковским сам по себе. Ему соответствует уравнение Фоккера — Планка

$$\dot{p}(z) = -\frac{\partial}{\partial z}\left\{\left[\nu-\mu+(\mu+\nu)\,z\right]p(z)\right\} +$$
$$+ \frac{1}{2b}\,\frac{\partial^2}{\partial z^2}\left[(1-z^2)^2\,p(z)\right],$$

вытекающее из (5.11.18).

Приравнивая производную $\dot{p}_{\text{ст}}(z)$ нулю и интегрируя получающееся уравнение, получаем стационарную плотность распределения вероятностей

$$p_{\text{ст}}(z) = \frac{\text{const}}{(1-z^2)^2}\exp\left\{2b\int_0^z\left[\nu-\mu-(\mu+\nu)\,x\right]\frac{dx}{(1-x^2)^2}\right\}. \tag{5.11.23}$$

Входящий в выражение (5.11.22) интеграл с этой плотностью распределения был вычислен в [4] и оказался равным

$$1-\int z^2 p(z)\,dz = 4K_q\left(b\sqrt{\mu\nu}\right)\left\{2K_q\left(b\sqrt{\mu\nu}\right)+\right.$$

$$\left.+\sqrt{\frac{\nu}{\mu}}K_{q+1}\left(b\sqrt{\mu\nu}\right)+\sqrt{\frac{\mu}{\nu}}K_{q-1}\left(b\sqrt{\mu\nu}\right)\right\}^{-1}, \tag{5.11.24}$$

где $q=\frac{b}{2}(\nu-\mu)$, а $K_q(z)$ — функция Макдонольда (Рыжик И. М., Градштейн И. С. [1]).

Подстановка (5.11.24) в (5.11.22) решает задачу вычисления удельной энтропии h_y^{P/Q_2} стационарного немарковского процесса y. Комбинируя (5.11.19), (5.11.20), (5.11.22), легко найти энтропию $h_{x|y}^{P/Q_1}=h_x^{P/Q_1}+h_{y|x}^{P/Q_2}-h_y^{P/Q_2}$ условного марковского процесса $x(t)$.

3. Изложенный выше вывод формулы (5.11.17) применим и в других случаях, например, когда процесс $\{x(t)\}$ является диффузионным марковским процессом или образует часть компонент комбинированного диффузионного марковского процесса $\{\xi(t)\}=\{x(t),\,y(t)\}$. Рассмотрим последний случай подробнее. Пусть комбинированный марковский процесс, имеющий компоненты

$$\xi_\alpha(t)=x_\alpha(t),\quad \alpha=1,\ldots,m;\qquad \xi_\rho(t)=y_\rho(t),\ \rho=m+1,\ldots,m+l,$$

описывается вектором сносов $a_j(\xi,\ t)$ и $(m+l)\times(m+l)$-матрицей локальных дисперсий $b_{jk}(\xi,\ t)$, $j,\ k=1,\ \ldots,\ m+l$. Предполагается, что $l\times l$ — подматрица последней $b_{\rho\sigma}(\xi,\ t)$, $\rho,\ \sigma=m+1,\ \ldots,\ m+l$ является невырожденной и не зависящей от x.

Тогда по аналогии с (5.11.17) справедлива формула

$$h_y^{P/Q_2}(t)=\frac{1}{2}\mathbf{M}\sum_{\rho,\ \sigma=m+1}^{m+l}(\mathbf{M}_{\mathrm{ps}}a_\rho)\,b_{\rho\sigma}^{-1}\,(\mathbf{M}_{\mathrm{ps}}a_\sigma), \tag{5.11.25}$$

где усреднение $\mathbf{M}_{\mathrm{ps}}$ соответствует усреднению с апостериорной плотностью вероятности

$$\mathbf{M}_{\mathrm{ps}}a_\sigma=\int a_\sigma(x,\ y,\ t)\,w_t(x)\,dx_1\ldots dx_m,$$

которая удовлетворяет полученному в указанной монографии [4] уравнению

$$\frac{\partial}{\partial t}w_t(x)=\frac{1}{2}\sum_{\alpha,\ \beta}\frac{\partial^2}{\partial x_\alpha\,\partial x_\beta}[b_{\alpha\beta}w_t]-$$

$$-\sum_{\alpha=1}^{m}\frac{\partial}{\partial x_\alpha}\left[a_\alpha w_t+\sum_{\sigma,\rho=m+1}^{m+l}b_{\alpha\rho}\,b_{\rho\sigma}^{-1}\,w_t\left(\frac{d^*y_\sigma}{dt}-\mathbf{M}_{\mathrm{ps}}a_\sigma\right)\right]+$$

$$+\sum_{\rho,\,\sigma=m+1}^{m+l}(a_\rho-\mathbf{M}_{\mathrm{ps}}\,a_\rho)\,b_{\rho\sigma}^{-1}\,w_t\left[\frac{d^*y_\sigma}{dt}-\mathbf{M}_{\mathrm{ps}}\,a_\sigma\right].$$

Второе усреднение в (5.11.25) соответствует усреднению по марковскому процессу $\{w_t(x),\ y(t)\}$ как стационарному случайному процессу.

4. До сих пор процесс $\{y(t)\}$ предполагался диффузионным Аналогичные результаты можно получить, скажем для того случая, когда процесс $\{y\ (t)\}$ есть процесс с конечным или счетным числом состояний.

Пусть $\{x(t)\}$ — марковский процесс с матрицей перехода $\pi_t\ (x,\ x')$, подобный процессу в п. 2. При фиксированной реализации $\{x\ (t)\}$ процесс $\{y(t)\}$ пусть является марковским процессом с дискретными состояниями, описываемыми дифференциальными вероятностями перехода $\pi_t\ (x,\ y,\ y')$, зависящими от $x=x\ (t)$. Тогда комбинированный процесс $\{\xi\,(t)\}=\{x(t),\ y(t)\}$ также будет марковским процессом с дискретными состояниями. Ему будет соответствовать дифференциальная матрица перехода

$$\pi_t\,(\xi,\,\xi')=\pi_t\,(x,\,x')\,\delta_{yy'}+\pi_t\,(x,\,y,\,y')\,\delta_{xx'}. \tag{5.11.26}$$

В данном случае энтропия процесса $\{x\,(t)\}$ и энтропия комбинированного процесса $\{x\,(t),\ y\,(t)\}$ могут быть подсчитаны путем применения формулы (5.9.4). Подсчет энтропии $H^{P/Q_2}_{y_0^T}$ или ее плотности $h_y^{P/Q_2}\ (t)$ требует особого рассмотрения, поскольку процесс $\{y\ (t)\}$, взятый в отдельности, является немарковским.

В принципиальном отношении соответствующий расчет может быть проведен так же, как и в п. 2. Воспользуемся формулой (5.11.12), выбирая теперь в качестве меры Q_2 пуассоновскую меру (для моментов перескока процесса $y(t)$) с единичной плотностью. При малом τ будем иметь

$$\frac{P\left[dy_t^{t+\tau}\,\middle|\,x\,(t),\ y\,(t)\right]}{Q_2\left[dy_t^{t+\tau}\,\middle|\,y\,(t)\right]}=$$
$$=\begin{cases} e^{\tau}\left[1+\pi_t\left(x\,(t),\ y\,(t),\ y\,(t)\right)\tau\right]+O\,(\tau^2) & \text{при } y\,(t+\tau)=y\,(t),\\ e^{\tau}\,\pi_t\left(x\,(t),\ y\,(t),\ y\,(t+\tau)\right)\tau+O\,(\tau^2) & \text{при } y\,(t+\tau)\neq y\,(t). \end{cases}$$

Подставляя это выражение в (5.11.12), находим

$$H^{P/Q_2}_{y_t^{t+\tau}\,|\,y_0^t}=\mathbf{M}\left[1+\mathbf{M}_{\mathrm{ps}}\,\pi_t\left(x\,(t),\ y\,(t),\ y\,(t)\right)\tau\right]\ln\left\{e^{\tau}\left[1+\right.\right.$$
$$\left.\left.+\mathbf{M}_{\mathrm{ps}}\,\pi_t\left(x\,(t),\ y\,(t),\ y\,(t)\right)\tau\right]\right\}+\sum_{y\,(t+\tau)\,\neq\,y\,(t)}\mathbf{M}_{\mathrm{ps}}\,\pi_t\left(x\,(t),\ y\,(t),\right.$$
$$\left.y\,(t+\tau)\right)\tau\ln\left[e^{\tau}\,\mathbf{M}_{\mathrm{ps}}\,\pi_t\left(x\,(t),\ y\,(t),\ y\,(t+\tau)\right)\right]+O\,(\tau^2). \tag{5.11.27}$$

Здесь, в первую очередь, произведено усреднение $\mathbf{M}\,[\ldots\,|x(t), y_0^t(t)]$ по $y_t^{t+\tau}$, затем усреднение $\mathbf{M}\,[\ldots|y_0^t]$ по $x(t)$ с весом W_t, а потом усреднение по прочим переменным.

Из (5.11.27) вытекает следующая формула для энтропийной плотности:

$$h_y^{P/Q_2}(t)=\mathbf{M}\Big\{1+\sum_{y'\neq y(t)}\Big[\mathbf{M}_{\mathrm{ps}}\,\pi\big(x(t),\,y(t),\,y'\big)\Big]\times$$
$$\times\Big[\ln\Big[\mathbf{M}_{\mathrm{ps}}\,\pi\big(x(t),\,y(t),\,y'\big)\Big]-1\Big]\Big\}. \qquad (5.11.28)$$

Отличие этой формулы от формулы

$$h_y^{P/Q_2}=\mathbf{M}\Big\{1+\sum_{y'\neq y(t)}\pi\big(y(t),\,y'\big)\Big[\ln\pi\big(y(t),\,y'\big)-1\Big]\Big\}, \qquad (5.11.29)$$

(т. е. от формулы (5.9.4), взятой при $\xi=y$) в том, что $\pi(y,\,y')$ заменены на апостериорные средние

$$\mathbf{M}_{\mathrm{ps}}\,\pi\big(x(t),\,y(t),\,y'\big)=\sum_x W_t(x)\,\pi\big(x,\,y(t),\,y'\big)$$

и имеется усреднение не по $y(t)$ с весом $P(dy)$, а усреднение по $\{W_t,\,y(t)\}$ с весом $P\,[dW_t dy\,(t)]$. Процесс $\{W_t,\,y(t)\}$ является вторичным апостериорным марковским процессом и для него нетрудно найти вероятности перехода, которые определяют, в частности, стационарное распределение $P_{\text{ст}}\,[dW_t\,dy(t)\}$.

Формулу (5.11.28) интересно сравнить с условной энтропией

$$h_{y|x}^{P/Q_2}(t)=\mathbf{M}\Big\{1+\sum_{y'\neq y(t)}\pi\big(x(t),\,y(t),\,y'\big)\times$$
$$\times\Big[\ln\pi\big(x(t),\,y(t),\,y'\big)-1\Big]\Big\}. \qquad (5.11.30)$$

Это выражение мы записали по аналогии с (5.11.29), считая значение $x(t)$ фиксированным, а затем усредняя по нему.

Изложенный метод вычисления плотности энтропии h_y может быть распространен и на случай, когда процесс $\{x(t)\}$ в отдельности не является марковским, а марковским процессом (с дискретными состояниями) является комбинированный процесс $\{x(t),\ y(t)\}$. Пусть он описывается дифференциальной матрицей перехода $\pi_t\,(x,\,y,\,x',\,y')$. В этом случае вид результирующей формулы (5.11.28) остается без изменения. Аналогично вычисляется и $h_x^{P/Q}$.

Из вышеизложенного видно, что описанный метод вычисления энтропии для части компонент марковского процесса имеет широкую область применения. В нем наиболее трудным этапом является отыскание распределения $P\,(dW_t,\,dy\,(t))$ для вторичного апостериорного W-процесса.

Глава 6

ИНФОРМАЦИЯ ПРИ НАЛИЧИИ ПОМЕХ. ШЕННОНОВСКОЕ КОЛИЧЕСТВО ИНФОРМАЦИИ

В настоящей главе рассматривается введенное Шенноном количество информации связи двух случайных величин или двух групп случайных величин. Это понятие является центральным в теории информации, самостоятельное развитие которой начато в работах Шеннона (1948 г.). Оно определяется как разность априорной и апостериорной (условной) энтропий.

Явное вычисление количества информации в сложных случаях, когда в качестве рассматриваемых случайных величин берутся случайные процессы, является, вообще говоря, непростой задачей. В двух последних параграфах главы эта задача решается для частных случаев: вычисляется удельное количество информации в случае стационарных гауссовых и марковских процессов. При этом используются методы и результаты гл. 5.

6.1. Потери информации при вырожденных преобразованиях и при простых помехах

1. Пусть имеется случайная величина x с конечным числом m состояний (числом реализаций) и соответствующие ей вероятности $P(x)$. Рассмотрим некоторое преобразование $\eta = f(x)$ этой случайной величины. Функция f, естественно, определена для всех возможных значений случайной величины.

Если преобразование $\eta = f(x)$ является невырожденным, т. е. существует обратная функция $f^{-1}(\eta) = x$ и может быть установлено взаимно однозначное соответствие между x и η, то энтропия при преобразовании не меняется:

$$H_x = H_\eta.$$

В самом деле, вследствие взаимно однозначного соответствия между значениями x и значениями η выражения типа (1.2.3) для энтропии H_x и H_η будут отличаться лишь порядком следования членов, нумерацией этих членов. Указанные значения можно нумеровать также числами $1, \ldots, m$. Вследствие этого мы можем, как это делали ранее, заменить множество значений случайной величины x на множество $1, \ldots, m$.

Сложнее обстоит дело, если преобразование $\eta = f(x)$ является вырожденным. Тогда фиксированному значению η соответствует подмножество $E(\eta)$ элементов x, число $\Omega(\eta)$ элементов которого больше единицы. Зная η, мы не в состоянии сказать, какое значение x из $E(\eta)$ породило данное значение η. Чтобы ответить на этот вопрос, требуется помимо η знать номер ζ исходного элемента x в множестве $E(\eta)$. Пара (η, ζ) уже в состоянии заменить x. Между возможными значениями этой пары и множеством возможных значений x можно осуществить взаимно однозначное соответствие и, следовательно,

$$H_x = H_{\eta\zeta}.$$

Случайная величина ζ является дополнительной по отношению к η; она дополняет η до x. Пусть, например, на натуральных числах задана функция $\eta = [x/10]$, где скобки обозначают целую часть. Тогда число η представляет собой номер десятка, в котором находится натуральное число x. Дополнительной величиной ζ в данном случае является номер числа в десятке. Пара (η, ζ) есть не что иное, как представление числа x в десятичной системе и, разумеется, заменяет x.

Применяя формулы из § 1.3 [типа (1.3.4)], имеем

$$H_x = H_\eta + H_{\zeta \mid \eta}.$$

Отсюда получаем

$$H_x \geqslant H_\eta, \tag{6.1.1}$$

так как $H_{\zeta|\eta} \geqslant 0$.

Рассмотрим, когда имеет место знак равенства в (6.1.1). Поскольку

$$H_{\zeta \mid \eta} = \sum P(\eta)\, H_\zeta(\mid \eta),$$

то $H_{\zeta|\eta} = 0$ тогда и только тогда, когда условная энтропия $H_\zeta(|\eta)$ исчезает для всех зачений η, имеющих ненулевую вероятность. Но энтропия

$$H_\zeta(\mid \eta) = -\sum_\zeta P(\zeta \mid \eta) \ln P(\zeta \mid \eta)$$

обращается в нуль, лишь если вероятности $P(\zeta|\eta)$ сосредоточены на одном единственном элементе из $E(\eta)$. Выбирая этот элемент, мы можем осуществить взаимно однозначное соответствие между значениями η, имеющими ненулевую вероятность, и значениями $(\eta, \zeta) = x$, имеющими также ненулевую вероятность. Следовательно, равенство $H_x - H_\eta = 0$, т. е. $H_{\zeta/\eta} = 0$, означает, что данное преобразование может быть сделано невырожденным в результате удаления элементов, имеющих нулевую вероятность, и что оно эквивалентно невырожденному преобразованию. Если же преобразование существенно вырожденное (т. е. объединяет хотя бы два каких-то элемента с ненулевыми вероятностями в один), тогда имеет место строгое неравенство

$$H_x > H_\eta. \tag{6.1.2}$$

Итак, мы доказали следующее:

Теорема 6.1. *Существенно вырожденное преобразование случайных величин $x \to \eta$ уменьшает (разрушает) информацию H, которая может содержаться в случайной величине.*

2. Перейдем теперь к рассмотрению помех. Передача сигналов в канале связи обычно сопровождается помехами или искажениями. Если x — сигнал на одном конце канала, то сигнал y на другом (приемном) конце, вообще говоря, отличается от x вследствие случайных искажений. Если x обозначает запись, сделанную в запоминающем устройстве в некоторый момент времени, то через определенное время запись может исказиться и вместо записи x в запоминающем устройстве будет содержаться другая запись y.

Такие случайные искажения описываются вероятностями перехода $P(y \mid x)$. Если через $P(x)$ обозначить вероятности первоначального сообщения или записи, то искаженное сообщение или запись будет, очевидно, иметь вероятности

$$P(y) = \sum_x P(x) P(y \mid x).$$

Совместное распределение вероятностей будет

$$P(x, y) = P(x) P(y \mid x). \tag{6.1.3}$$

В отличие от рассмотренного раньше преобразования $\eta = f(x)$ теперь преобразование x в y, связанное с помехами, является рандомизированным, т. е. носит случайный характер (y случайно, даже если x фиксировано). Покажем, что иногда, несмотря на это кардинальное отличие, между вырожденным преобразованием и между случайными искажениями имеется много общего.

Рассмотрим специальный случай простых помех. Будем называть помехи простыми, если они перемешивают значения x лишь внутри некоторых классов. Точнее говоря, пусть область значений x можно разбить на неперекрывающиеся подмножества $E_1, E_2, \ldots$, а область значений y — на подмножества $G_1, G_2, \ldots$, обладающие следующими свойствами:

а) из области E_k можно попасть лишь в область G_k:

$$\mathbf{P}(y \in G_l \mid x \in E_k) = 0, \text{ если } l \neq k; \tag{6.1.4}$$

б) Из всех точек области E_k происходит переход в $y \in G_k$ с одинаковыми вероятностями

$$P(y \mid x \in E_k) = P(y \mid k). \tag{6.1.5}$$

Легко видеть, что из а) следует

$$H_{l \mid k} = 0 \tag{6.1.6}$$

или, что эквивалентно,

$$H_k = H_{kl} \tag{6.1.7}$$

(k — номер области E_k, l — номер области G_l). Из б) же вытекает

$$H_{y|x} = H_{y|k}. \tag{6.1.8}$$

Теорема 6.2. *Простые помехи с информационной точки зрения эквивалентны нерандомизированному вырожденному преобразованию $k = k(x)$ (где $k(x)$ — номер той области E_k, которой принадлежит x)*, т. е. $H_{x|y} = H_{x|k}$.

Для доказательства следует рассмотреть апостериорное распределение вероятностей $P(x|y)$. Подставляя (6.1.5) в формулу обратной вероятности (формулу Байеса), приводя ее к виду

$$P(x|y) = \frac{P(x)\,P(y|x)}{\sum\limits_{x'} P(x')\,P(y|x')} = \frac{P(x)\,P(y|l)}{\sum\limits_{x' \in E_l} P(x')P(y|l)}, \qquad \text{где } y \in G_l,$$

($l = k$) и сокращая на общий множитель $P(y|l)$, получаем

$$P(x|y) = \frac{P(x)}{P(E_l)}, \quad \text{если } x \in E_l \quad P(x|y) = 0, \ \text{если } X \notin E_l.$$

Это выражение зависит от номера l множества G_l, которому принадлежит y, но не зависит от конкретного значения y внутри G_l, т. е.

$$P(x|y) = p(x|l) \quad (y \in G_l). \tag{6.1.9}$$

Если выполняется подобное равенство, то говорят, что переменная l является достаточной переменной или достаточной статистикой, заменяющей y. Мы получили, следовательно, что номер l множества служит в данном случае достаточной статистикой. Из равенства (6.1.9) вытекает, что $H_x(|y) = H_x(|l)$ и (после усреднения этого равенства) что

$$H_{x|y} = H_{x|l}. \tag{6.1.10}$$

Равенства (6.1.9), (6.1.10) означают, что наблюдение переменной y эквивалентно наблюдению переменной $l=k$. Доказательство закончено.

Из определения простых помех и из теоремы 6.2 видно, что понятие простых помех является обратимым: помехи, соответствующие обратному переходу с вероятностями $P(x|y)$, являются простыми, если простыми являются помехи прямого перехода с вероятностями $P(y|x)$.

В самом деле, подставляя (6.1.4) в (6.1.3), легко убедиться, что отличны от нуля лишь те вероятности $P(x, y)$, для которых x и y попадают в области E_k, G_k с одинаковым номером k. Вероятности $P(x|y)$ равны нулю, если номера k и l областей $E_k \ni x$, $G_l \ni y$

не совпадают. Следовательно, свойство (6.1.4) обратимо. В дополнение к (6.1.6), (6.1.7) выполняются соотношения

$$H_{k \mid l}=0, \quad H_l=H_{kl}=H_k. \tag{6.1.11}$$

Далее равенство (6.1.9), очевидно, является обращением равенства (6.1.5). Тем самым указанная обратимость доказана: кроме вырожденного преобразования $k=k(x)$ можно рассматривать также нерандомизированное вырожденное преобразование $l=l(y)$, где l — номер области G_l, содержащей точку y.

Из теоремы 6.2. следует, что помехи разрушают информацию, поскольку такое разрушение имеет место при вырожденном преобразовании $l=l(x)$ по теореме 6.1.

3. Чтобы безошибочно передавать информацию при вырожденном преобразовании или при простых помехах, нужно связывать информацию не с переменной x, которая искажается при преобразовании, а с переменной $\eta=k$, которая останется неизменной, так как $l=k$. Количество передаваемой информации, следовательно, будет равно

$$I=H_k. \tag{6.1.12}$$

Преобразуем это соотношение к другому виду, используя тождество

$$H_k=H_{xk}-H_{x \mid k}=H_x+H_{k \mid x}-H_{x \mid k}, \tag{6.1.13}$$

вытекающее из определения условных энтропий (§ 1.3). При фиксированном x номер $k(x)$ области $E_k \ni x$ является полностью определенным, поэтому

$$H_{k \mid x}=0, \tag{6.1.14}$$

и из (6.1.12), (6.1.13) получаем

$$I=H_x-H_{x \mid k}.$$

Согласно (6.1.10) это соотношение можно записать

$$I=H_x-H_{x \mid y}. \tag{6.1.15}$$

Далее, по аналогии с (6.1.14) имеем $H_{l \mid y}=0$ или $H_{k \mid y}=0$ (что то же самое, поскольку $l=k$ с вероятностью 1). Следовательно,

$$H_k=H_k-H_{k \mid y}. \tag{6.1.16}$$

Учитывая (6.1.12), (6.1.11), (6.1.16), (6.1.15), будем иметь

$$I=H_k-H_{k \mid l}=H_k-H_{k \mid y}=H_x-H_{x \mid y}. \tag{6.1.17}$$

Приведенные результаты относятся к случаю простых помех однако в менее явной форме (в асимптотическом смысле) их можно перенести и на случай произвольных помех, как это видно из последующего (§ 7.6). Там вместо точных соотношений (6.1.17) выведены приближенные соотношения (7.6.19). Для этого нужно рассматривать не отдельные случайные величины x и y, а последовательности $x_1, \ldots, x_n$ и $y_1, \ldots, y_n$ этих величин при $n \rightarrow \infty$. Подобно тому как случай произвольных вероятностей асимптотически сводится к случаю равновероятных возможностей (см. § 1.4, 1.5), так и случай произвольных помех асимптотически (при $n \to \infty$) сводится к случаю простых помех. Поэтому, если информацию связывать с соответствующими (приближенными) достаточными статистиками $k(x) = 1, \ldots, M$ (для этого ее должно быть не больше, чем $H_k - H_{k \mid l}$: $\ln M \leqslant H_k - H_{k \mid l}$), то в пределе $n \to \infty$ можно избежать ошибок, вызванных искажениями. Этот факт связан с известной теоремой Шеннона и является одной из возможных ее интерпретаций (см. гл. 7).

6.2. Информация связи дискретных случайных величин

Переходя к случаю произвольных помех, рассмотрим величину

$$I_{xy} = H_x - H_{x \mid y}, \tag{6.2.1}$$

называемую взаимной (парной) информацией связи случайных величин x и y. Ее можно интерпретировать в некотором смысле как количество информации об x, содержащееся в y.

Количество информации (6.2.1) было введено Шенноном [1], который показал также значение этой величины в теории информации. Будем называть его *шенноновским количеством информации*.

Имеется немало косвенных причин для того, чтобы трактовать величину (6.2.1) как меру количества информации (например неотрицательность, аддитивность и пр.), однако главной причиной в конечном счете является тот факт, что количество (6.2.1) асимптотически сводится к хартлиевскому количеству информации $\ln M$. Об этом говорит рассматриваемая в дальнейшем теорема Шеннона (см. гл. 7). Без этого использование (6.2.1) в качестве меры информации носило бы более умозрительный, нежели практический характер.

Поскольку $H_x = H_{\text{pr}}$ — количество априорной неопределенности, а $H_{x \mid y} = H_{\text{ps}}$ — среднее количество апостериорной неопределенности при наблюдении величины y, то количество информации (6.2.1), которое можно записать

$$I_{xy} = H_{\text{pr}} - H_{\text{ps}},$$

указывает среднюю величину неопределенности, исчезнувшей при приеме информации. Подобная трактовка количества информации давалась раньше в § 1.1 (1.1.2).

Используя обычное соотношение $H_{x \mid y} = H_{xy} - H_y$ (см. (1.3.4)), мы можем записать формулу (6.2.1) в виде

$$I_{xy} = H_x + H_y - H_{xy}, \tag{6.2.2}$$

откуда видна симметрия этой величины — она остается неизменной, если x и y меняются ролями.

Следовательно, такое же количество неопределенности в среднем исчезнет, если наблюдаемой переменной является x, а неизвестной переменной — y:

$$I_{xy} = H_y - H_{y \mid x}. \tag{6.2.3}$$

На рис. 6.1. наглядно изображены соотношения, имеющиеся между величинами H_x, H_y, H_{xy}, $H_{x \mid y}$, $H_{y \mid x}$, I_{xy}. Поскольку условная энтропия не превосходит безусловную (теорема 1.6), информация связи I_{xy} является неотрицательной.

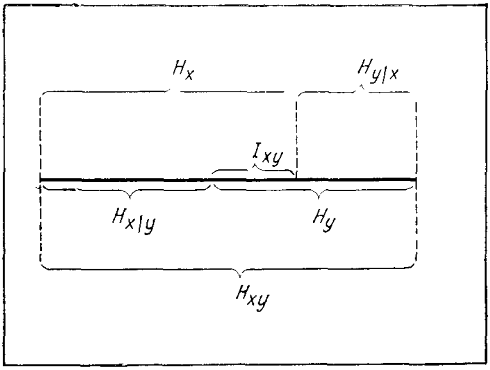


Рис. 6.1.
Соотношения между информационными характеристиками двух случайных величин.

Раскрывая в (6.2.2) энтропии H_x, H_y, H_{xy} по обычной формуле типа (1.2.3), записываем количество информации связи двух случайных величин в виде

$$I_{xy} = \mathbf{M}\left[H(x) + H(y) - H(x, y)\right] = \mathbf{M} \ln \frac{P(x, y)}{P(x)\, P(y)}. \tag{6.2.4}$$

Очевидно, что возможны также и такие эквивалентные формы записи:

$$I_{xy} = \mathbf{M} \ln \frac{P(x \mid y)}{P(x)} = \mathbf{M} \ln \frac{P(y \mid x)}{P(y)}.$$

Подобно тому, как в § 1.2 помимо средней энтропии H_ξ мы рассматривали случайную энтропию $H(\xi) = -\ln P(\xi)$, можно ввести случайную информацию связи

$$I(x, y) = H(x) + H(y) - H(x, y) = \ln \frac{P(x, y)}{P(x)\, P(y)} = \\ = \ln \frac{P(x \mid y)}{P(x)}. \tag{6.2.5}$$

Тогда, как легко видеть, (6.2.4) примет вид

$$I_{xy} = \mathbf{M} I(x, y). \tag{6.2.6}$$

Из (6.2.5) вытекает следующая формула для совместного распределения:

$$P(x, y) = e^{-H(x) - H(y) + I(x, y)}. \tag{6.2.7}$$

Следовательно, парная информация $I(x, y)$ доопределяет однократные распределения вероятностей

$$P(x) = e^{-H(x)}, \qquad P(y) = e^{-H(y)} \tag{6.2.8}$$

до двукратного распределения.

Поскольку из двукратного распределения суммированием можно получить однократные

$$\sum_y P(x, y) = P(x), \qquad \sum_x P(x, y) = P(y),$$

то вследствие (6.2.7), (6.2.8) случайная информация обязана удовлетворять уравнениям

$$\sum_y e^{I(x, y) - H(y)} = 1, \qquad \sum_x e^{I(x, y) - H(x)} = 1.$$

Средняя информация связи (6.2.4), как отмечалось, является заведомо неотрицательной величиной. Этого нельзя сказать о случайной информации (6.2.5). Возможны и отрицательные ее значения, хотя положительные превалируют, так что усреднение приводит к неотрицательной величине. Докажем, что уже *условное усреднение* $I(x, y)$ *только по* x *или по* y *дает неотрицательную величину*.

Доказательство. Формулу (6.2.5) запишем в виде

$$I(x, y) = -\ln \frac{P(x)}{P(x \mid y)}. \tag{6.2.9}$$

Используя неравенство

$$\ln \frac{P(x)}{P(x \mid y)} \leqslant -1 + \frac{P(x)}{P(x \mid y)},$$

будем иметь

$$I(x, y) \geqslant 1 - \frac{P(x)}{P(x \mid y)}. \tag{6.2.10}$$

Если это неравенство условно усреднить по x, т. е. взять условное математическое ожидание $\mathbf{M}[\dots \mid y]$, то поскольку

$$\mathbf{M}\left[\frac{P(x)}{P(x \mid y)} \,\middle|\, y\right] = \sum_x \frac{P(x)}{P(x \mid y)} P(x \mid y) = 1,$$

из (6.2.10) получим

$$\mathbf{M}\left[I(x, y) \mid y\right] \geqslant 0.$$

Заменив в этом рассуждении x на y, а y на x, получим второе требуемое неравенство : $\mathbf{M}\,[I\,(x,\ y)\,|\,x] \geqslant 0$.

Любопытно отметить, что будет иметь место обратное неравенство

$$\sum_x I\,(x,\ y)\,P\,(x) \leqslant 0 \quad \left(\text{или } \sum_y I\,(x,\ y)\,P\,(y) \leqslant 0\right),$$

если усреднение проводить не с весом $P\,(x|y)$, а с весом $P\,(x)$ (или $P\,(y)$).

В самом деле, записав правую часть (6.2.9) или (6.2.5) в виде $\ln \frac{P(y|x)}{P(y)}$ и использовав неравенство

$$\ln \frac{P\,(y\,|\,x)}{P\,(y)} \leqslant \frac{P\,(y\,|\,x)}{P\,(y)} - 1,$$

получим вместо (6.2.10)

$$I\,(x,\ y) \leqslant \frac{P\,(y\,|\,x)}{P\,(y)} - 1.$$

Усредняя это неравенство с весом $P\,(x)$, будем иметь

$$\sum_x I\,(x,\ y)\,P\,(x) \leqslant \frac{1}{P\,(y)} \sum_x P\,(y\,|\,x)\,P\,(x) - 1 = 0,$$

поскольку $\sum_x P\,(y\,|\,x)\,P\,(x) = P\,(y)$.

Таким образом, отрицательных значений случайной информации $I\ (x,\ y)$ не так уже мало. Это одна из причин для того, чтобы называть количеством информации не $I\ (x,\ y)$, а соответствующее среднее значение I_{xy}.

6.3. Условная информация. Иерархическая аддитивность информации

1. Если имеется несколько случайных величин $x,\ y,\ z,\ \ldots$, то по аналогии с формулой (6.2.5) можно определить условную информацию связи, взяв вместо безусловных условные вероятности:

$$I\,(x,\ y\,|\,z) = \ln \frac{P\,(x\,|\,yz)}{P\,(x\,|\,z)} \tag{6.3.1}$$

или

$$I\,(x,\ y\,|\,z) = H\,(x\,|\,z) + H\,(y\,|\,z) - H\,(x,\ y\,|\,z). \tag{6.3.2}$$

Результат частичного или полного усреднения обозначим так:

$$\begin{aligned} I_{xy}\,(|\,z) &= \mathbf{M}\left[I\,(x,\ y\,|\,z)\,|\,z\right] = \sum_{x,\ y} I\,(x,\ y\,|\,z)\,P\,(x,\ y\,|\,z), \\ I_{xy\,|\,z} &= \mathbf{M}I\,(x,\ y\,|\,z) = \sum_{x,\ y,\ z} I\,(x,\ y\,|\,z)\,P\,(x,\ y,\ z). \end{aligned} \tag{6.3.3}$$

Подставляя сюда (6.3.1) или (6.3.2), будем иметь

$$I_{xy}(|z) = H_x(|z) + H_y(|z) - H_{xy}(|z) = H_x(|z) - H_{x|y}(|z),$$
$$I_{xy|z} = H_{x|z} + H_{y|z} - H_{xy|z} = H_{x|z} - H_{x|yz}. \tag{6.3.4}$$

Мы видим, что информации, условные или безусловные, могут быть выражены через соответствующие энтропии, условные или безусловные.

Поскольку энтропия обладает свойством иерархической аддитивности (§ 1.3), то аналогичным же свойством обладает и информация. Пусть x в формуле (6.2.1) содержит несколько составляющих $x = (x_1, \ldots, x_n)$. Тогда к энтропиям H_x, $H_{x|y}$ можно применить формулу иерархической аддитивности (1.3.4), что даст

$$H_x = H_{x_1} + H_{x_2|x_1} + \ldots + H_{x_n|x_1 \ldots x_{n-1}},$$
$$H_{x|y} = H_{x_1|y} + H_{x_2|x_1 y} + \ldots + H_{x_n|x_1 \ldots x_{n-1} y}.$$

Взяв в соответствии с (6.2.1) разность $H_x - H_{x|y}$ и группируя члены попарно, имеем

$$I_{(x_1 \ldots x_n) y} = [H_{x_1} - H_{x_1|y}] + [H_{x_2|x_1} - H_{x_2|x_1 y}] + \ldots +$$
$$+ [H_{x_n|x_1 \ldots x_{n-1}} - H_{x_n|x_1 \ldots x_{n-1} y}].$$

Но каждая разность $H_{x_k|x_1 \ldots x_{k-1}} - H_{x_k|x_1 \ldots x_{k-1} y}$ в силу (6.3.4) есть не что иное, как условная информация $I_{x_k y|x_1 \ldots x_{k-1}}$. Следовательно,

$$I_{(x_1 \ldots x_n) y} = I_{x_1 y} + I_{x_2 y|x_1} + I_{x_3 y|x_1 x_2} + \ldots + I_{x_n y|x_1 \ldots x_{n-1}}. \tag{6.3.5}$$

Нетрудно понять, что такая же формула справедлива и для условной информации

$$I_{(x_1 \ldots x_n) y|z} = \sum_{k=1}^{n} I_{x_k y|x_1 \ldots x_{k-1} z}. \tag{6.3.6}$$

Предположим теперь, что вторая случайная величина также является составной: $y = (y_1 \ldots y_r)$. Тогда, применяя формулу (6.3.6) к каждой информации $I_{x_k y|x_1 \ldots x_{k-1}}$, будем иметь

$$I_{x_k (y_1 \ldots y_r)|x_1 \ldots x_{k-1}} = \sum_{l=1}^{r} I_{x_k y_l|x_1 \ldots x_{k-1} y_1 \ldots y_{l-1}}.$$

Следовательно, из (6.3.5) получим формулу, содержащую двойное суммирование:

$$I_{(x_1 \ldots x_n)(y_1 \ldots y_r)} =$$
$$= \sum_{k=1}^{n} \sum_{l=1}^{r} I_{x_k y_l|x_1 \ldots x_{k-1} y_1 \ldots y_{l-1}}. \tag{6.3.7}$$

Согласно этой формуле вклад в информацию $I_{(x_1 \dots x_n)(y_1 \dots y_r)}$ дают всевозможные пары x_k, y_l. Конечно, фактический, ненулевой вклад дают лишь те пары x_k, y_l, которые не являются статистически независимыми случайными величинами.

Рассмотрим, например, информацию связи $I_{(x_1 x_2)(y_1 y_2)}$, где величина x статистически связана лишь с y_1, а x_2 — лишь с y_2. Тогда в сумме (6.3.7) вместо четырех будет только два члена:

$$I_{(x_1 x_2)(y_1 y_2)} = I_{x_1 y_1} + I_{x_2 y_2 \mid x_1 y_1}. \tag{6.3.8}$$

Здесь $I_{x_2 y_2 | x_1 y_1} = I_{x_2 y_2}$, если $P(x_2, y_2 | x_1, y_1) = P(x_2, y_2)$, т. е. если и x_2 и y_2 не зависят от x_1, y_1.

Подобно тому как в случае энтропии свойство иерархической аддитивности справедливо не только для средних энтропий (1.3.4), но и для случайных энтропий (1.3.6), так и в случае информаций связи соотношения, аналогичные (6.3.5)—(6.3.8), могут быть записаны для случайных информаций. Так, например,

$$I(x_1, \dots, x_n; y_1, \dots, y_r) = \sum_{k=1}^{n} \sum_{l=1}^{r} I(x_k, y_l \mid x_1, \dots, x_{k-1}, y_1, \dots, y_{l-1}).$$

Это обоснование совершенно аналогично предыдущему при использовании (1.3.6) вместо (1.3.4).

2. Условные информации связи (6.3.3) являются неотрицательными, что может быть выведено, например, из формул (6.3.4) при учете теорем 1.6, 1.6а. Это обстоятельство позволяет получить из формул иерархической аддитивности (6.3.5), (6.3.7) различные неравенства. Именно, информация $I_{(x_1 \dots x_n) y}$ или $I_{(x_1 \dots x_n)(y_1 \dots y_r)}$, стоящая в левой части, не меньше суммы любой части членов, входящих в правую часть равенства. Проиллюстрируем это на простом примере, когда рассматривается информация $I_{(x_1 x_2) y}$ связи пары x_1, x_2 случайных величин с величиной y. Формула (6.3.5) дает $I_{(x_1 x_2) y} = I_{x_1 y} + I_{x_2 y \mid x_1}$. Поскольку $I_{x_2 y \mid x_1} \geqslant 0$, отсюда имеем неравенство

$$I_{(x_1 x_2) y} \geqslant I_{x_1 y}. \tag{6.3.9}$$

Знак равенства

$$I_{(x_1 x_2) y} = I_{x_1 y} \tag{6.3.10}$$

имеет место в том и только в том случае, когда

$$I_{x_2, y \mid x_1} = H_{x_2 \mid x_1} - H_{x_2 \mid x_1 y} = H_{y \mid x_1} - H_{y \mid x_1 x_2} = 0.$$

Это условие выполняется, если

$$P(x_2 \mid x_1, y) = P(x_2 \mid x_1) \text{ или } P(y \mid x_1 x_2) = P(y \mid x_1). \tag{6.3.11}$$

Последнее равенство есть условие марковской связи тройки x_2, x_1, y.

Из (6.3.7) можно также получить соотношение

$$I_{(x_1, x_2)\,(y_1, y_2)} \geqslant I_{x_1 y_1}. \tag{6.3.12}$$

Итак, мы видим, что информация связи заданных случайных величин не меньше информации связи части указанных величин. Это аналогично неравенству $H_{x_1} \leqslant H_{x_1 x_2}$, для энтропии (так как $H_{x_2 \mid x_1} \geqslant 0$). Между тем соотношение $H_{x \mid z} \leqslant H_x$ не имеет своего аналога для информации. Неравенство $I_{xy \mid z} \leqslant I_{xy}$ в общем случае не имеет места.

Пример 1. Пусть x, y, z — случайные величины с двумя значениями, описываемые вероятностями

$$P(z_1) = P(z_2) = 1/2, \qquad \| P(y_i, x_j \mid z_1) \| = \begin{pmatrix} 3/8 & 1/8 \\ 1/8 & 3/8 \end{pmatrix},$$

$$\| P(y_i, x_j \mid z_2) \| = \begin{pmatrix} 3/4 & 1/4 \\ 0 & 0 \end{pmatrix} \qquad (i, j = 1,2).$$

Тогда

$$H_x(\mid z_1) = h(1/2), \qquad H_{x \mid y}(\mid z_1) = h(1/4), \qquad H_x(\mid z_2) = h(1/4),$$
$$H_{x \mid y}(\mid z_2) = h(1/4),$$

где $h(p) = -p \ln p - (1-p) \ln (1-p)$.

По формуле (6.3.4) имеем

$$2I_{xy \mid z} = I_{xy}(\mid z_1) + I_{xy}(\mid z_2) = h(1/2) - h(1/4). \tag{6.3.13}$$

В то же время

$$\| P(y_i, x_j) \| = \begin{pmatrix} 9/16 & 3/16 \\ 1/16 & 3/16 \end{pmatrix}, \qquad P(x_i) = (5/8, 3/8).$$

Следовательно,

$$H_x = h(3/8), \qquad H_{x \mid y} = h(1/4),$$

так что

$$I_{xy} = h(3/8) - h(1/4). \tag{6.3.14}$$

Из (6.3.13), (6.3.14) находим разность

$$I_{xy} - I_{xy \mid z} = h(3/8) - (1/2) h(1/4) - (1/2) h(1/2) =$$
$$= 0{,}183 \text{ бита} > 0. \tag{6.3.15}$$

Пример 2. Предположим теперь, что случайные величины x, y, z с двумя значениями описываются вероятностями

$$P(z_1) = P(z_2) = \frac{1}{2}, \qquad \| P(y_i, x_j \mid z_1) \| = \begin{pmatrix} 3/8 & 1/8 \\ 1/8 & 3/8 \end{pmatrix},$$

$$\| P(y_i, x_j \mid z_2) \| = \begin{pmatrix} 1/3 & 1/6 \\ 1/6 & 1/3 \end{pmatrix}.$$

В этом случае

$$H_x(|z_1) = H_x(|z_2) = H_x = h(1/2),$$
$$I_{xy}(|z_1) = h(1/2) - h(1/4), \qquad I_{xy}(|z_2) = h(1/2) - h(1/3),$$
$$I_{xy|z} = h(1/2) - (1/2)h(1/4) - (1/2)h(1/3).$$

Далее, поскольку

$$P(y_i, x_j) = \begin{pmatrix} 17/48 & 7/48 \\ 7/48 & 17/48 \end{pmatrix}, \qquad \| P(x_i \mid y_j) \| = \begin{pmatrix} 17/24 & 7/24 \\ 7/24 & 7/24 \end{pmatrix},$$

то

$$I_{xy} = h(1/2) - h(7/24).$$

Следовательно,

$$I_{xy} - I_{xy\,|\,z} = (1/2)h(1/4) + (1/2)h(1/3) - h(7/24) = \\ = -0{,}04 \text{ бита} < 0. \tag{6.3.16}$$

На знак разности $I_{xy} - I_{xy\,|\,z}$ в этих примерах оказала влияние выпуклость функции $h(p) = -p\ln p - (1-p)\ln(1-p)$. В самом деле, для выпуклой функции имеем

$$h(\mathbf{M}\xi) - \mathbf{M}h(\xi) \geqslant 0; \tag{6.3.17}$$

значения же 3/8, 7/24 есть не что иное, как среднее:

$$3/8 = (1/4 + 1/2)/2 = \mathbf{M}\xi, \qquad 7/24 = (1/4 + 1/3)/2 = \mathbf{M}\xi.$$

Вследствие (6.3.17) разность (6.3.15) положительна, а (6.3.16) отрицательна.

Итак, мы убедились, что знак указанной разности может быть любым.

3. Способ определения (6.2.1) парной информации связи двух случайных величин можно обобщить так, чтобы определить информацию связи трех или большего числа случайных величин. Определим тройную информацию связи формулой

$$I_{xyz} = I_{xy} - I_{xy\,|\,z}. \tag{6.3.18}$$

Подставляя сюда (6.3.4) и выражая условные энтропии через безусловные, будем иметь

$$I_{xyz} = H_x + H_y + H_z - H_{xy} - H_{xz} - H_{yz} + H_{xyz} \tag{6.3.19}$$

или

$$I_{xyz} = H_{xyz} + I_{xy} + I_{yx} + I_{zx} - H_x - H_y - H_z. \tag{6.3.20}$$

Очевидна симметрия этой формулы относительно x, y, z, т. е. ее инвариантность при перестановках этих случайных величин.

Формулой

$$I(x, y, z) = H(x, y, z) + I(x, y) + I(y, z) + I(z, x) - \\ - H(x) - H(y) - H(z), \tag{6.3.21}$$

которая аналогична (6.3.20), определяем случайную тройную информацию. Учитывая, что $H(x, y, z) = -\ln P(x, y, z)$, получаем из (6.3.21) следующую формулу для трехкратного закона распределения:

$$P(x, y, z) = \exp\{-H(x) - H(y) - H(z) + I(x, y) + \\ + I(y, z) + I(z, x) - I(x, y, z)\}.$$

Она является обобщением формулы (6.2.7).

В рассмотренных выше примерах случайная информация (6.3.21) такова:

$$I(y_i, x_j, z_1)\| = \begin{pmatrix} \ln\frac{4}{5} & \ln\frac{4}{3} \\ \ln\frac{4}{5} & \ln\frac{4}{3} \end{pmatrix},$$

$$\| I(y_1, x_j, z_2)\| = \left(\ln\frac{6}{5}, \ln\frac{2}{3}\right)$$

(при $y = y_2$, $z = z_2$ информация не определена) для примера 1 и

$$\| I(y_i, x_j, z_1)\| = \begin{pmatrix} \ln\frac{17}{18} & \ln\frac{7}{6} \\ \ln\frac{7}{16} & \ln\frac{17}{18} \end{pmatrix},$$

$$\| I(y_i, x_j, z)\| = \begin{pmatrix} \ln\frac{51}{48} & \ln\frac{21}{24} \\ \ln\frac{21}{24} & \ln\frac{51}{48} \end{pmatrix}$$

для примера 2. Среднее же I_{xyz}, соответственно равно 0,183 бита и —0,04 бита в силу (6.3.15), (6.3.16).

Таким образом, неотрицательность тройной корреляции не является обязательной.

По аналогии с формулой (6.3.19) строится информация связи и для большего числа случайных величин. В общем случае n-кратная информация связи определяется формулой

$$I_{x_1 \ldots x_n} = \sum_{i=1}^{n} H_{x_i} - \sum H_{x_i x_j} + \sum H_{x_i x_j x_k} - \ldots - \\ -(-1)^n H_{x_1 \ldots x_n}, \tag{6.3.22}$$

где суммирование производится по всевозможным несовпадающим парам ($n(n-1)/1 \cdot 2$ членов), тройкам ($n(n-1)(n-2)/3!$ членов) и другим сочетаниям индексов $1, \ldots, n$.

Можно доказать, что информация связи (6.3.22) записывается аналогично (6.3.18) в виде разности безусловной и условной информации меньшей кратности

$$I_{x_1 \dots x_{n+1}} = I_{x_1 \dots x_n} - I_{x_1 \dots x_n \mid x_{n+1}}. \qquad (6.3.23)$$

Доказательство эквивалентности формул (6.3.22) и (6.3.23) целесообразно проводить по индукции. Из выражения, стоящего в правой части (6.3.22), следует вычесть выражение для условной информации

$$I_{x_1 \dots x_n \mid x_{n+1}} = \sum H_{x_i \mid x_{n+1}} - \sum H_{x_i x_k \mid x_{n+1}} - \dots$$
$$\dots - (-1)^n H_{x_1 \dots x_n \mid x_{n+1}} = \sum H_{x_i x_{n+1}} - nH_{x_n} -$$
$$- \sum H_{x_i x_k x_{n+1}} + \frac{n(n-1)}{2} H_{x_n} - \dots$$
$$- (-1)^n H_{x_1 \dots x_n x_{n+1}} + (-1)^n H_{x_{n+1}}.$$

Анализируя члены, входящие в получившееся выражение, и учитывая равенство

$$-n + \frac{n(n-1)}{2} - \dots + (-1)^n = (1-1)^n - 1 = -1,$$

убеждаемся, что это выражение совпадает с суммой в правой части формулы (6.3.22), если в последней заменить n на $n+1$. Равенство (6.3.22) справедливо для $n = 2$ и $n = 3$, следовательно, оно будет справедливо и для бо́льших значений n.

В случае высоких кратностей, начиная с рассмотренной выше тройной связи, уже нельзя высказать определенных суждений о неотрицательности информации, о которой шла речь в случае парной информации.

6.4. Количество информации связи в общем случае

1. В предыдущих параграфах предполага́лось, что рассматриваемые случайные величины являются дискретными, т. е. принимают значения из конечного или счетного множества состояний. При этом по существу использовались лишь общие свойства энтропии. Совершенно ясно, что приведенные выше формулы и утверждения для информации связи могут быть распространены на случай произвольных непрерывных или комбинированных случайных величин. В самом деле, в § 1.6, 1.7 было показано как для произвольных случайных величин ввести понятие энтропии, обладающей всеми свойствами энтропии дискретной версии. Нам остается лишь воспользоваться приведенными там соотношениями. Наличие обычных свойств энтропии обобщенной версии обеспечит выполнение для информации связи в обобщенной версии тех же соотношений, что и в дискретной.

Рассмотрим две произвольные случайные величины x, y. На абстрактном языке теории меры это означает, что задано вероятностное пространство (Ω, F, P) и два борелевских подполя $F_1 \subset F$, $F_2 \subset F$. Первое F_1 определяется условиями, наложенными на $x(\omega)$, $\omega \in \Omega$ (т. е. условиями типа $x(\omega) < c$), а второе F_2 — условиями, наложенными на $y(\omega)$, $\omega \in \Omega$. Предполагается, что кроме вероятностной меры P заданы:

1) мера ν на объединенном борелевском поле $F_{12} = \sigma(F_1 \cup F_2)$; 2) мера ν_1 на поле F_1; 3) мера ν_2 на F_2.

Эти меры таковы, что выполнено условие

$$\nu(AB) = \nu_1(A)\nu_2(B), \qquad A \in F_1, \qquad B \in F_2, \tag{6.4.1}$$

соответствующее условию (1.7.8), мера P абсолютна непрерывна на F_{12} относительно меры ν (а значит, и меры $\nu_1(A)\ \nu_2(B)$). При описанных условиях можно определить энтропии

$$H_x = -\int \ln \frac{P(dx)}{\nu_1(dx)} P(dx), \qquad H_{x\mid y} = -\int \ln \frac{P(dx\mid y)}{\nu_1(dx)} P(dx\,dy)$$
и другие. (6.4.2)

В соответствии с формулами (6.2.2), (6.2.5), (6.2.6) информация связи определяется как разность

$$I_{xy} = H_x - H_{x\mid y} \tag{6.4.3}$$

или, если подставить (6.4.2),

$$I_{xy} = \int \ln \frac{P(dx\mid y)}{P(dx)} P(dx\,dy). \tag{6.4.4}$$

Последнее равенство можно принять за исходное определение информации связи независимо от использования понятия энтропии, что в принципе весьма удобно. Такое определение позволяет не вводить в рассмотрение вспомогательные меры ν_1, ν_2, ν.

Формуле (6.4.4) соответствует следующее выражение для случайной энтропии:

$$I(x, y) = \ln \frac{P(dx\mid y)}{P(dx)}. \tag{6.4.5}$$

Функция, стоящая под знаком логарифма, определяется как производная Радона — Никодима меры $P(A|y)$ по мере $P(A)$, $A \in F_1$.

При таком ее определении, как известно, остается некоторый произвол. Различные ее определения могут не совпадать на множестве нулевой вероятности. Этот произвол, конечно, не сказывается на величине средней информации (6.4.4). Соотношениям (6.4.5), (6.4.4) можно придать также вид

$$I(x, y) = \ln \frac{P(dx\,dy)}{P(dx)\ P(dy)} = \ln \frac{P(dy\mid x)}{P(dy)}, \tag{6.4.6}$$

$$I_{xy} = \mathbf{M} I(x, y). \tag{6.4.7}$$

Эти соотношения аналогичны формулам (6.2.4), (6.2.5) дискретной версии.

Формулы (6.2.8), (6.2.7) в обобщенном случае принимают вид

$$P(dx) = \nu_1(dx)\,e^{-H(x)}, \qquad P(dy) = \nu_2(dy)\,e^{-H(y)},$$

$$P(dx\,dy) = \nu(dx\,dy)\,e^{-H(x,y)} = \nu(dx\,dy)\,e^{-H(x)-H(y)+I(x,y)}.$$

Без труда могут быть продублированы и другие формулы из § 6.2, 6.3. В дальнейшем будем писать соответствующие формулы в том виде, в каком потребуется.

Выполнение условия мультипликативности (6.4.1) для нормированной меры

$$Q(dx\,dy) = \nu(dx\,dy)/N = \nu(dx\,dy) \Big/ \int \nu(d\,x\,dy)$$

означает равенство

$$Q(dx\,dy) = Q(dx)\,Q(dy), \tag{6.4.8}$$

(см. (1.7.16)). Следовательно, можно будет воспользоваться формулой (1.7.17). Учитывая (1.6.17), (1.7.17), нетрудно убедиться, что формулу (6.4.3) можно записать

$$I_{xy} = H^{P/Q}_{x\,|\,y} - H^{P/Q}_{x} = H^{P/Q}_{xy} - H^{P/Q}_{x} - H^{P/Q}_{y}. \tag{6.4.9}$$

Таким образом, энтропии $H^{P/Q}$ также позволяют вычислять информацию связи как разность энтропий наподобие (6.4.3), но с другим знаком.

2. Информация связи (6.4.5), (6.4.6) является случайной величиной. Для некоторых целей, например для исследования вероятности ошибки при передаче сообщений по каналу с помехами (§ 7.2, 7.3), важно знать не только ее среднее значение $I_{xy} = \mathbf{M}\,I(x, y)$, но и другие статистические характеристики. Важной характеристикой этой случайной величины является ее характеристический потенциал $\mu(s)$. Применяя формулу типа (4.1.11а) к случайной информации (6.4.5) (взятой с обратным знаком), получаем

$$\mu(s) = \ln \int \exp\left[-s \ln \frac{P(dx\,dy)}{P(dx)\;P(dy)}\right] P(dx\,dy) =$$

$$= \ln \int P^{1-s}(dx\,dy)\,P^{s}(dx)\,P^{s}(dy). \tag{6.4.10}$$

Этот потенциал, как и характеристический потенциал (4.1.11а) любой случайной величины $B(\xi)$, обладает свойством

$$\mu(0) = 0. \tag{6.4.11}$$

Кроме того, он обладает дополнительным свойством

$$\mu(1) = 0, \tag{6.4.12}$$

поскольку $\int P(dx)\,P(dy) = \int P(dx) \int P(dy) = 1.$

Через характеристический потенциал (6.4.10) выражаются важные результаты теории оптимального кодирования при наличии помех (теоремы 7.2, 7.3).

6.5. Информация связи гауссовых величин

1. Рассмотрим две группы гауссовых случайных величин: $\mathrm{x} = (x_1, \ldots, x_r)$, $\mathrm{y} = (y_1, \ldots, y_s) \equiv (x_{r+1}, \ldots, x_{r+s})$ и найдем информацию связи (6.4.4) между ними. Для этой цели удобно воспользоваться полученными в § 5.3 выражениями для энтропии гауссовых переменных, вычисляя информацию как разность $H_x - H_{\mathrm{x}|\mathrm{y}} = H_x + H_y - H_{xy}$. Эту информацию можно вычислить в данном случае также по формуле (6.4.9), используя найденные в § 5.4 выражения для энтропии $H^{P/Q}$.

Указанные гауссовские переменные характеризуются векторами средних значений

$$m_\alpha = \mathbf{M} x_\alpha, \qquad m_\rho = n_{\rho - r} = \mathbf{M} y_{\rho - r} = \mathbf{M} x_\rho$$

и корреляционными матрицами

$$\begin{gathered}
\mathrm{R} = \| R_{\alpha\beta} \| = \| \mathbf{M} (x_\alpha - m_\alpha)(x_\beta - m_\beta) \|, \quad \mathrm{S} = \| S_{\rho\sigma} \| = \\
= \| \mathbf{M} (y_{\rho - r} - n_{\rho - r})(y_{\sigma - r} - n_{\sigma - r}) \|, \\
\mathrm{U} = \| U_{\alpha\sigma} \| = \| \mathbf{M} (x_\alpha - m_\alpha)(y_{\sigma - r} - n_{\sigma - r}) \| \\
(\alpha, \beta = 1, \ldots, r; \ \rho, \sigma = r+1, \ldots, r+s).
\end{gathered}$$

Матрицы R, S предполагаем невырожденными (в противном случае следует отбросить часть переменных, оставляя столько переменных, каков ранг матрицы). Матрица взаимных корреляций U в общем случае может быть не квадратной (когда $r \neq s$). Совместная корреляционная матрица имеет вид

$$\mathrm{K} = \begin{pmatrix} \mathrm{R} & \mathrm{U} \\ \mathrm{U}^T & \mathrm{S} \end{pmatrix}, \tag{6.5.1}$$

где T означает транспонирование.

Используя (5.4.5), по формуле (6.2.5) находим случайную информацию

$$\begin{gathered}
I(x_1, \ldots, x_r; \ y_1, \ldots, y_s) = -\frac{1}{2} \ln \det \mathrm{K} + \frac{1}{2} \ln \det \mathrm{R} + \\
+ \frac{1}{2} \ln \det \mathrm{S} - \frac{1}{2} \sum_{i, j = 1}^{r+s} (x_i - m_i) K_{ij}^{-1} (x_j - m_j) + \frac{1}{2} \times \\
\times \sum_{\alpha, \beta = 1}^{r} (x_\alpha - m_\alpha) R_{\alpha\beta}^{-1} (x_\beta - m_\beta) + \frac{1}{2} \sum_{\rho, \sigma = r+1}^{r+s} (x_\rho - m_\rho) \times \\
\times S_{\rho\sigma}^{-1} (x_\sigma - m_\sigma).
\end{gathered} \tag{6.5.2}$$

Для отыскания средней информации I_{xy} целесообразно использовать простую формулу (5.4.6а), которая дает

$$I_{xy} = -\frac{1}{2}\ln\det \mathrm{K} + \frac{1}{2}\ln\det \mathrm{R} + \frac{1}{2}\ln\det \mathrm{S}. \tag{6.5.3}$$

Этому результату можно придать другую форму. Поскольку

$$\ln\det \mathrm{A} = \mathrm{Sp}\ln \mathrm{A} \tag{6.5.4}$$

для любой матрицы A, то формулу (6.5.3) можно преобразовать к виду

$$I_{xy} = -\frac{1}{2}\mathrm{Sp}\,[\ln \mathrm{K} + \ln \mathrm{R}^{-1} + \ln \mathrm{S}^{-1}] =$$

$$= -\frac{1}{2}\mathrm{Sp}\ln\left[\mathrm{K}\begin{pmatrix}\mathrm{R}^{-1} & 0\\ 0 & \mathrm{S}^{-1}\end{pmatrix}\right]$$

или, если учесть (6.5.1) и перемножить матрицы,

$$I_{xy} = -\frac{1}{2}\mathrm{Sp}\ln\begin{pmatrix}1 & \mathrm{US}^{-1}\\ \mathrm{U}^T\mathrm{R}^{-1} & 1\end{pmatrix} =$$

$$= -\frac{1}{2}\ln\det\begin{pmatrix}1 & \mathrm{US}^{-1}\\ \mathrm{U}^T\mathrm{R}^{-1} & 1\end{pmatrix},$$

где 1 — единичные матрицы.

Применим далее формулу

$$\det\begin{pmatrix}\mathrm{A} & \mathrm{B}\\ \mathrm{C} & \mathrm{D}\end{pmatrix} = \det\,[(\mathrm{A} - \mathrm{BD}^{-1}\mathrm{C})\,\mathrm{D}] \tag{6.5.5}$$

см. приложение (П.2.4)] и получим

$$I_{xy} = -\frac{1}{2}\ln\det\,(1 - \mathrm{US}^{-1}\mathrm{U}^T\mathrm{R}^{-1}) =$$

$$= -\frac{1}{2}\mathrm{Sp}\ln\,(1 - \mathrm{US}^{-1}\mathrm{U}^T R^{-1}) \tag{6.5.6}$$

или

$$I_{xy} = -\frac{1}{2}\mathrm{Sp}\ln\,(1 - \mathrm{U}^T R^{-1}\mathrm{US}^{-1}). \tag{6.5.7}$$

Если воспользоваться разложением логарифмической функции в ряд

$$\ln(1 - z) = -\sum_{k=1}^{\infty}\frac{z^k}{k}, \tag{6.5.8}$$

то из (6.5.7) будем иметь

$$I_{xy} = \frac{1}{2}\sum_{k=1}^{\infty}\frac{\mathrm{Sp}\,(\mathrm{U}^T\mathrm{R}^{-1}\mathrm{US}^{-1})^k}{k} = \frac{1}{2}\sum_{k=1}^{\infty}\frac{1}{k}\times$$

$$\times\,\mathrm{Sp}\,(\mathrm{US}^{-1}\mathrm{U}^T\mathrm{R}^{-1})^k. \tag{6.5.9}$$

Мы получили несколько различных эквивалентных формул для вычисления информации связи гауссовых переменных. Некоторыми из них, например (6.5.7), можно пользоваться, приводя матрицы к диагональному виду, а некоторые формулы, например (6.5.3), (6.5.9), позволяют избежать этого.

2. Частный случай 1. Пусть в каждой группе имеется по одной переменной. Совместная корреляционная матрица (6.5.1) имеет вид

$$\mathrm{K}=\begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2 R_1 \\ \sigma_1\sigma_2 R_1 & \sigma_2^2 \end{pmatrix}$$

(R_1 — коэффициент корреляции, а матрицы R, S, U состоят из одного элемента: $\mathrm{R}=(\sigma_1^2)$, $\mathrm{S}=(\sigma_2^2)$, $\mathrm{U}=(\sigma_1\sigma_2 R_1)$). Тогда

$$\mathrm{US}^{-1}\mathrm{U}^T\mathrm{R}^{-1}=(\sigma_1\sigma_2 R_1)(\sigma_2^2)^{-1}(\sigma_1\sigma_2 R_1)(\sigma_1^2)^{-1}=R_1^2 \qquad (6.5.10)$$

и из (6.5.6) имеем

$$I_{xy}=-\frac{1}{2}\ln(1-R_1^2). \qquad (6.5.11)$$

Частный случай 2. Возьмем три гауссовы величины с корреляционной матрицей

$$\mathrm{K}=\begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2 R_3 & \sigma_1\sigma_3 R_2 \\ \sigma_1\sigma_2 R_3 & \sigma_2^2 & \sigma_2\sigma_3 R_1 \\ \sigma_1\sigma_3 R_2 & \sigma_2\sigma_3 R_1 & \sigma_3^2 \end{pmatrix} \qquad (6.5.12)$$

и найдем информацию связи первой случайной величины с двумя остальными. В этом случае матрица K распадается на подматрицы следующим образом:

$$R=(\sigma_1^2), \qquad \mathrm{U}=(\sigma_1\sigma_2 R_3, \quad \sigma_1\sigma_3 R_2),$$

$$\mathrm{U}^T=\begin{pmatrix} \sigma_1\sigma_2 R_3 \\ \sigma_1\sigma_3' R_2 \end{pmatrix}, \qquad \mathrm{S}=\begin{pmatrix} \sigma_2^2 & \sigma_2\sigma_3 R_1 \\ \sigma_2\sigma_3 R_1 & \sigma_3^2 \end{pmatrix}.$$

Поскольку

$$\mathrm{S}^{-1}=\frac{1}{\sigma_2^2\sigma_3^2(1-R_1^2)}\begin{pmatrix} \sigma_3^2 & -\sigma_2\sigma_3 R_1 \\ -\sigma_2\sigma_3 R_1 & \sigma_2^2 \end{pmatrix},$$

$$\mathrm{US}^{-1}\mathrm{U}^T=\sigma_1^2\,\frac{R_3^2-2R_1R_2R_3+R_2^2}{1-R_1^2},$$

имеем

$$\mathrm{US}^{-1}\mathrm{U}^T\mathrm{R}^{-1}=(R_3^2+R_2^2-2R_1R_2R_3)/(1-R_1^2).$$

Эта матрица состоит лишь из одного элемента. По формуле (6.5.6) получаем

$$I_{x_1,\,(x_2\,x_3)}=-\frac{1}{2}\ln\left(1-\frac{R_3^2+R_2^2-2R_1\,R_2\,R_3}{1-R_1^2}\right). \qquad (6.5.13)$$

Этот результат можно получить и из формулы (6.5.3), вычисляя детерминанты

$$\begin{aligned}&\det \mathrm{K}=\sigma_1^2\,\sigma_2^2\,\sigma_3^2\,(1-R_1^2-R_2^2-R_3^2+2R_1\,R_2\,R_3),\\ &\det \mathrm{S}=\sigma_2^2\,\sigma_3^2\,(1-R_1^2), \qquad \det R=\sigma_1^2.\end{aligned} \qquad (6.5.14)$$

При этом (6.5.3) дает

$$I_{xy}=\frac{1}{2}\ln(1-R_1^2)-\frac{1}{2}\ln(1-R_1^2-R_2^2-R_3^2+2R_1\,R_2\,R_3),$$

что совпадает с (6.5.13).

В дополнение к предыдущему вычислим тройную информацию связи (6.3.19) трех гауссовых случайных величин. Без ограничения общности их корреляционную матрицу можно брать в виде (6.5.12). Используя (5.4.6а), (6.5.14), имеем

$$H_{x_1\,x_2\,x_3}-H_{x_1}-H_{x_2}-H_{x_3}=\frac{1}{2}\ln\det\mathrm{K}-\frac{1}{2}\ln\sigma_1^2-$$
$$-\frac{1}{2}\ln\sigma_2^2-\frac{1}{2}\ln\sigma_3^2=\frac{1}{2}\ln(1-R_1^2-R_2^2-R_3^2+2R_1\,R_2\,R_3).$$

Прибавляя сюда согласно (6.3.20) парные информации $I_{x_1x_2}$, $I_{x_2x_3}$, $I_{x_3x_1}$, которые имеют вид (6.5.11), получаем

$$I_{x_1\,x_2\,x_3}=\frac{1}{2}\ln(1-R_1^2-R_2^2-R_3^2+2R_1\,R_2\,R_3)-$$
$$-\frac{1}{2}\ln(1-R_1^2)-\frac{1}{2}\ln(1-R_2^2)-\frac{1}{2}\ln(1-R_3^2).$$

Если произвести разложение найденного выражения по R_1, R_2, R_3, то, как легко видеть, получим для тройной информации формулу

$$I_{x_1\,x_2\,x_3}=R_1\,R_2\,R_3+O(R^4).$$

Ее полезно сопоставить с вытекающей из (6.5.11) аналогичной формулой

$$I_{x_1\,x_2}=\frac{1}{2}R_1^2+O(R^4)$$

для парной информации.

3. Частный случай 3. Рассмотрим случай аддитивных независимых помех, когда переменные второй группы y_1, y_2, ..., y_r ($s=r$) есть результат прибавления к переменным пер-

вой группы $x_1, \ldots, x_r$ независимых гауссовых случайных величин $\xi_1, \ldots, \xi_r$:

$$y_\alpha = x_\alpha + \xi_\alpha, \quad \alpha = 1, \ldots, r. \tag{6.5.15}$$

В этом случае число переменных первой и второй групп одно и то же ($s = r$). Пусть R есть корреляционная матрица переменных $x_1, \ldots, x_r$, а N — корреляционная матрица аддитивных помех $\xi_1, \ldots, \xi_r$. Из условия независимости помех ξ от x вытекает, что корреляционные матрицы для суммы (6.5.15) имеют вид

$$\mathrm{S} = \mathrm{R} + \mathrm{N}, \qquad \mathrm{U} = \mathrm{R} \quad (\mathrm{U}^T = \mathrm{R}). \tag{6.5.16}$$

Чтобы применить формулу (6.5.7), вычислим $\mathrm{U}^T\mathrm{R}^{-1}\mathrm{U}\mathrm{S}^{-1}$. В данном случае вследствие (6.5.16) имеем

$$\mathrm{U}^T \mathrm{R}^{-1} \mathrm{U} \mathrm{S}^{-1} = \mathrm{R}(\mathrm{R} + \mathrm{N})^{-1},$$

но

$$1 - \mathrm{R}(\mathrm{R} + \mathrm{N})^{-1} = [\mathrm{R} + \mathrm{N} - \mathrm{R}](\mathrm{R} + \mathrm{N})^{-1} = \mathrm{N}(\mathrm{R} + \mathrm{N})^{-1} =$$
$$= [(\mathrm{R} + \mathrm{N})\mathrm{N}^{-1})^{-1} = [\mathrm{R}\mathrm{N}^{-1} + 1]^{-1},$$

поэтому

$$1 - \mathrm{U}^T \mathrm{R}^{-1} \mathrm{U} \mathrm{S}^{-1} = (\mathrm{R}\mathrm{N}^{-1} + 1)^{-1}.$$

Следовательно, формула (6.5.7) дает

$$I_{xy} = \frac{1}{2} \operatorname{Sp} \ln (1 + \mathrm{R}\mathrm{N}^{-1}) = \frac{1}{2} \ln \det (1 + \mathrm{R}\mathrm{N}^{-1}). \tag{6.5.17}$$

Пусть λ_k — собственные значения матрицы $\mathrm{R}\mathrm{N}^{-1}$, тогда, очевидно

$$I_{xy} = \frac{1}{2} \sum_k \ln (1 + \lambda_k).$$

Если переменные $x_1, \ldots, x_r$ и $\xi_1, \ldots, \xi_r$ взаимно независимы, то обе матрицы имеют диагональный вид

$$R = \| R_\alpha \delta_{\alpha\beta} \|, \qquad \mathrm{N} = \| \mathrm{N}_\alpha \delta_{\alpha\beta} \| \tag{6.5.18}$$

и (6.5.17) обращается в соотношение

$$I_{xy} = \frac{1}{2} \sum_\alpha \ln \left(1 + \frac{R_\alpha}{N_\alpha} \right). \tag{6.5.19}$$

В противном случае можно добиться диагональности (6.5.18) невырожденным линейным преобразованием C:

$$\mathrm{C}\mathrm{R}\mathrm{C}^T = \| R_\alpha \delta_{\alpha\beta} \|, \qquad \mathrm{C}\mathrm{N}\mathrm{C}^T = \| N_\alpha \delta_{\alpha\beta} \|.$$

Это преобразование оставляет информацию (6.5.17) инвариантной, так как

$$\mathrm{C}\mathrm{R}\mathrm{C}^T (\mathrm{C}\mathrm{N}\mathrm{C}^T)^{-1} = \mathrm{C}\mathrm{R}\mathrm{C}^T \mathrm{C}^{T\,-1} \mathrm{N}^{-1} = \mathrm{C}(\mathrm{R}\mathrm{N}^{-1})\mathrm{C}^{-1}. \tag{6.5.20}$$

Поэтому и в данном случае можно пользоваться формулой (6.5.19) после указанного преобразования.

4. Вычислим для гауссовых переменных характеристический потенциал (6.4.10) случайной информации связи. Для этого запишем случайную информацию (6.4.6), (6.5.2) в следующем виде:

$$I(x_1, \ldots, x_r; y_1, \ldots, y_s) = I_{\mathrm{xy}} - \frac{1}{2} \sum_{i, j=1}^{r+s} (x_i - m_i) \times$$

$$\times A_{ij} (x_j - m_j), \tag{6.5.20a}$$

где обозначено

$$\| A_{ij} \| = \mathrm{A} = \mathrm{K}^{-1} - \begin{pmatrix} \mathrm{R}^{-1} & 0 \\ 0 & \mathrm{S}^{-1} \end{pmatrix} = \begin{pmatrix} \mathrm{R} & \mathrm{U} \\ \mathrm{U}^T & \mathrm{S} \end{pmatrix}^{-1} - \begin{pmatrix} \mathrm{R}^{-1} & 0 \\ 0 & \mathrm{S}^{-1} \end{pmatrix}.$$

Подставляя (6.5.20а) в формулу

$$\mu(s) = \ln \mathbf{M} \exp[-sI(x_1, \ldots, x_r; y_1, \ldots, y_s)] =$$

$$= \ln \Big\{ \det{}^{-1/2}(2\pi \mathrm{K}) \int \exp[-sI(x_1, \ldots, x_r; y_1, \ldots, y_s) -$$

$$- \frac{1}{2}(x_i - m_i) K_{ij}^{-1} (x_j - m_j)]\, dx_1, \ldots, dx_{r+s} \Big\}$$

и учитывая, что

$$\int \exp\left[-\frac{1}{2} \sum \xi_i D_{ij} \xi_j \right] d\xi_1, \ldots, d\xi_l = \det{}^{-1/2}(\mathrm{D}/2\pi)$$

при произвольной положительно определенной матрице $\mathrm{D} = \|D_{ij}\|$, получаем

$$\mu(s) = -sI_{\mathrm{xy}} + \ln \det{}^{-1/2} \mathrm{K} + \ln \det{}^{-1/2} \Big[(1-s) \mathrm{K}^{-1} +$$

$$+ s \begin{pmatrix} \mathrm{R}^{-1} & 0 \\ 0 & \mathrm{S}^{-1} \end{pmatrix} \Big].$$

Объединим последние два члена:

$$\mu(s) = -sI_{\mathrm{xy}} - \frac{1}{2} \ln \det \left\{ 1 - s + s\mathrm{K} \begin{pmatrix} \mathrm{R}^{-1} & 0 \\ 0 & \mathrm{S}^{-1} \end{pmatrix} \right\}. \tag{6.5.21}$$

Матрица $(1-s)\,\mathrm{K}^{-1} + s \begin{pmatrix} \mathrm{R}^{-1} & 0 \\ 0 & S^{-1} \end{pmatrix}$ заведомо является невырожденной и положительно определенной при $0 \leqslant s \leqslant 1$ как сумма двух таких положительно определенных матриц. Поэтому формула (6.5.21) и последующие формулы справедливы, по крайней мере, на интервале $[0,1] \ni s$.

Полученному результату можно придать различную форму. Поскольку

$$1-s+s\mathrm{K}\begin{pmatrix}\mathrm{R}^{-1} & 0\\ 0 & \mathrm{S}^{-1}\end{pmatrix}=\left[\begin{pmatrix}(1-s)\,\mathrm{R} & 0\\ 0 & (1-s)\,\mathrm{S}\end{pmatrix}+s\mathrm{K}\right]\times$$

$$\times\begin{pmatrix}\mathrm{R}^{-1} & 0\\ 0 & \mathrm{S}^{-1}\end{pmatrix}=\begin{pmatrix}\mathrm{R} & s\mathrm{U}\\ s\mathrm{U}^{T} & \mathrm{S}\end{pmatrix}\begin{pmatrix}\mathrm{R}^{-1} & 0\\ 0 & \mathrm{S}^{-1}\end{pmatrix}, \qquad (6.5.22)$$

то, учитывая (6.5.3), характеристический потенциал (6.5.21) можно записать

$$\mu(s)=\frac{s}{2}\ln\det\mathrm{K}+\frac{1-s}{2}\ln\det\mathrm{R}+\frac{1-s}{2}\ln\det\mathrm{S}-$$

$$-\frac{1}{2}\ln\det\begin{pmatrix}\mathrm{R} & s\mathrm{U}\\ s\mathrm{U}^{T} & S\end{pmatrix}. \qquad (6.5.23)$$

Полученное выражение, как легко проверить, дает $\mu(s)=0$ при $s=0$ и $s=1$ в соответствии с отмеченными ранее свойствами (6.4.11), (6.4.12).

После перемножения матриц в правой части равенства (6.5.22) получим матрицу

$$\begin{pmatrix}1 & s\mathrm{U}\mathrm{S}^{-1}\\ s\mathrm{U}^{T}\mathrm{R}^{-1} & 1\end{pmatrix}.$$

Поэтому формуле (6.5.23) можно придать вид

$$\mu(s)=-\frac{1}{2}\ln\det\begin{pmatrix}1 & s\mathrm{U}\mathrm{S}^{-1}\\ s\mathrm{U}^{T}\mathrm{R}^{-1} & 1\end{pmatrix}-sI_{xy}. \qquad (6.5.24)$$

Применяя формулу (6.5.5), придем к результату

$$\mu(s)=-\frac{1}{2}\ln\det(1-s^{2}\mathrm{U}\mathrm{S}^{-1}\mathrm{U}^{T}\mathrm{R}^{-1})-sI_{xy}=$$

$$=-\frac{1}{2}\ln\det(1-s^{2}\mathrm{B})+\frac{s}{2}\ln\det(1-\mathrm{B}) \qquad (6.5.25)$$

$$(\mathrm{B}=\mathrm{U}\mathrm{S}^{-1}\mathrm{U}^{T}\mathrm{R}^{-1})$$

или

$$\mu(s)=-\frac{1}{2}\ln\det(1-s^{2}\mathrm{U}^{T}\mathrm{R}^{-1}\mathrm{U}\mathrm{S}^{-1})-sI_{xy}=$$

$$=-\frac{1}{2}\ln\det(1-s^{2}\widetilde{\mathrm{B}})+\frac{s}{2}\ln\det(1-\widetilde{\mathrm{B}})$$

$$(\widetilde{\mathrm{B}}=\mathrm{U}^{T}\mathrm{R}^{-1}\mathrm{U}\mathrm{S}^{-1}),$$

что соответствует формулам (6.5.6), (6.5.7). Если использовать разложение (6.5.8), то будем иметь

$$\mu(s)=\frac{s}{2}\sum_{k=1}^{\infty}\frac{1}{k}(s^{2k-1}-1)\operatorname{Sp}\mathrm{B}^k \tag{6.5.26}$$

по аналогии с (6.5.9). В частности, легко найти дисперсию случайной информации связи гауссовых переменных. Для этого нужно взять коэффициент при $\frac{1}{2}s^2$ в указанном разложении (6.5.26), что даст

$$\mathbf{D}I(x_1,\ldots,x_r;\,y_1,\ldots,y_s)=\operatorname{Sp}\mathrm{B}=\operatorname{Sp}\widetilde{\mathrm{B}}.$$

Итак, мы видим, что все статистические свойства случайной информации связи гауссовых переменных определяются лишь одной единственной матрицей $\mathrm{B}=\mathrm{US}^{-1}\mathrm{U}^T\mathrm{R}^{-1}$ или $\widetilde{\mathrm{B}}$.

Для частного случая 1, рассмотренного ранее, в соответствии с формулами (6.5.25), (6.5.9) имеем

$$\mu(s)=-\frac{1}{2}\ln(1-s^2R_1^2)+\frac{s}{2}\ln(1-R_1^2).$$

Не представляет труда применить приведенные формулы и к другим частным случаям.

6.6. Удельная информация стационарных и стационарно связанных процессов. Гауссовы процессы

1. Будем предполагать теперь, что как первая, так и вторая группа случайных величин представляют собой стационарные процессы

$$x_{-\infty}^{\infty}=\{x_t,-\infty<t<\infty\},\qquad y_{-\infty}^{\infty}=\{y_t,-\infty<t<\infty\} \tag{6.6.1}$$

в дискретном или непрерывном времени t. Эти процессы предполагаются не только стационарными, но и стационарно связанными, так что комбинированный процесс $z_{-\infty}^{\infty}=\{x_t,\,y_t,-\infty<t<\infty\}$ является стационарным. Для отрезка $[0,\,T]$ можно рассматривать энтропии $H_{x_0^T}$, $H_{y_0^T}$, $H_{z_0^T}$, определенные в соответствии с формулами и результатами гл. 5. Здесь

$$x_0^T=\{x_t,0\leqslant t<T\},\qquad y_0^T=\{y_t,0\leqslant t<T\},$$
$$z_0^T=\{x_t,y_t,0\leqslant t<T\}.$$

Согласно общей формуле (6.2.2) эти энтропии позволяют вычислить информацию связи

$$I_{x_0^T,\,y_0^T}=H_{x_0^T}+H_{y_0^T}-H_{x_0^T y_0^T}. \tag{6.6.2}$$

Определим удельную информацию связи процессов $\{x_t\}$ и $\{y_t\}$ как предел

$$i_{xy} = \lim_{T\to\infty} \frac{1}{T} I_{x_0^T, y_0^T}. \tag{6.6.3}$$

Если сюда подставить (6.6.2), то, очевидно,

$$i_{xy} = \lim_{T\to\infty} \frac{1}{T} H_{x_0^T} + \lim_{T\to\infty} \frac{1}{T} H_{y_0^T} - \lim_{T\to\infty} \frac{1}{T} H_{x_0^T y_0^T}. \tag{6.6.4}$$

Пределы, стоящие в правой части этого равенства, существуют в соответствии с теоремами 5.1 и 5.4 и соответственно равны

$$h_x = H_{x_0^1 \mid x_{-\infty}^0}, \qquad h_y = H_{y_0^1 \mid y_{-\infty}^0}, \qquad h_{xy} = H_{x_0^1 y_0^1 \mid x_{-\infty}^0 y_{-\infty}^0} \tag{6.6.5}$$

(для дискретного времени, когда t — целое, $H_{\xi_0^1 \mid \xi_{-\infty}^0}$ совпадает с $H_{\xi_0 \mid \ldots \xi_{-2}\xi_{-1}} = H^1$, а для непрерывного времени мы пользуемся формулой (5.6.10), полагая в ней $\tau = 1$). Поэтому (6.6.4) принимает вид

$$i_{xy} = h_x + h_y - h_{xy}, \qquad I_{xy}^1 = H_x^1 + H_y^1 - H_{xy}^1. \tag{6.6.6}$$

Разумеется, информация (6.6.2) и удельная информация (6.6.6) могут быть конечными не только в том случае, когда $H_{x_0^T}$, $H_{y_0^T}$, $H_{x_0^T y_0^T}$ или h_x, h_y, h_{xy} конечны по отдельности. Поэтому, в принципе, можно вычислять информацию, минуя вычисление энтропий. Однако в практических случаях, когда информация конечна, всегда так можно подобрать вспомогательные меры ν, Q, входящие в определение энтропии (§ 1.6), чтобы были конечны все члены в (6.6.2), (6.6.6). Тогда задача вычисления информации связи сведется к рассмотренной в гл. 5 более простой задаче вычисления удельных энтропий, хотя бы при одном (самом удобном) выборе меры ν или Q. Согласно сказанному в § 6.4 нужно следить, чтобы выполнялось условие мультипликативности (6.4.1) или (6.4.8), принимающее для процессов (6.6.1) вид

$$\nu(dx_{-\infty}^{\infty} dy_{-\infty}^{\infty}) = \nu_1(dx_{-\infty}^{\infty})\, \nu_2(dy_{-\infty}^{\infty})$$

или

$$Q(dx_{-\infty}^{\infty} dy_{-\infty}^{\infty}) = Q_1(dx_{-\infty}^{\infty})\, Q_2(dy_{-\infty}^{\infty}). \tag{6.6.7}$$

При этом согласно (6.4.9) вместо (6.6.2), (6.6.6) можно брать формулы

$$I_{x_0^T, y_0^T} = H_{x_0^T y_0^T}^{P/Q} - H_{x_0^T}^{P/Q_1} - H_{y_0^T}^{P/Q_2}, \quad i_{xy} = h_{xy}^{P/Q} - h_x^{P/Q_1} - h_y^{P/Q_2}, \tag{6.6.8}$$

где $h_{xy}^{P/Q}$, h_x^{P/Q_1}, h_y^{P/Q_2}, — энтропийные плотности типа

$$\lim_{T\to\infty}\frac{1}{T}H_{\xi_0^T}^{P/Q}=H_{\xi_0^1\mid\xi_{-\infty}^0}^{P/Q}.$$

Кроме удельной информации рассмотрим информацию конца интервала, аналогичную энтропии Г конца интервала, входящей в формулу

$$H_{\xi_0^T}=hT+2\Gamma+o_T(1) \tag{6.6.9}$$

[см. (5.6.17)].

Сравнивая соотношение (5.6.15) с формулой (6.2.2), легко видеть, что постоянную 2Г можно интерпретировать как информацию связи случайного процесса на одной полупрямой ($-\infty$, 0) с процессом на другой полупрямой (0, ∞):

$$2\Gamma=I_{\xi_{-\infty}^0\,\xi_0^\infty}. \tag{6.6.10}$$

Формулы (6.6.9), (6.6.10) справедливы для каждого из процессов $\{x_t\}$, $\{y_t\}$, $\{x_t, y_t\}$. Подставляя подобные выражения для каждой энтропии в (6.6.2) и учитывая (6.6.6), получаем

$$I_{x_0^T y_0^T}=i_{xy}T+I_{x_{-\infty}^0,\,x_0^\infty}+I_{y_{-\infty}^0,\,y_0^\infty}-I_{(x_{-\infty}^0\,y_{\infty}^0),\,(x_0^\infty\,y_0^\infty)}+$$
$$+o_T(1). \tag{6.6.11}$$

Это равенство позволяет вычислить информацию для конечного отрезка [0, T) более точно, чем по формуле $I_{x_0^T y_0^T}\approx i_{xy}T$, вытекающей из (6.6.3).

2. Применим приведенные выше формулы для вычисления удельной информации связи двух гауссовых стационарных случайных последовательностей $\{x_t,\ t=\dots, 1, 2, \dots\}$, $\{y_t,\ t=\dots, 1,2, \dots\}$. Поскольку средние значения гауссовых переменных не влияют на величину информации связи (см. например, (6.5.3)) без ограничения общности можно полагать, что средние значения равны нулю:

$$\mathbf{M}x_t=0,\qquad \mathbf{M}y_t=0.$$

Корреляционные матрицы

$$R_{t-q}^{11}=\mathbf{M}x_t\,x_q,\qquad R_{t-q}^{22}=\mathbf{M}y_t\,y_q,\qquad R_{t-q}^{12}=\mathbf{M}x_t\,y_q$$

или соответствующие спектральные плотности

$$\varphi^{\alpha\beta}(\mu)=\sum_{\sigma=-\infty}^{\infty}e^{-2\pi i\mu\sigma}R_\sigma^{\alpha\beta}$$

предполагаются заданными.

Для вычисления удельной информации применим формулу (6.6.6). Входящие в нее удельные энтропии для стационарных гауссовых последовательностей были вычислены в § 5.5. Энтропии

H_x^1, H_y^1 определяются равенством (5.5.17), а для отыскания энтропии H_{xy}^1 можно использовать формулу (5.5.19). Итак, имеем

$$H_{xy}^1 = \frac{1}{2}\int_{-1/2}^{1/2} \ln \det \| \varphi^{\alpha\beta}(\mu) \| \, d\mu = \frac{1}{2}\int_{-1/2}^{1/2} \ln [\varphi^{11}(\mu)\varphi^{22}(\mu) - $$

$$- \varphi^{12}(\mu)\varphi^{21}(\mu)] \, d\mu, \quad H_x = \frac{1}{2}\int_{-1/2}^{1/2} \ln \varphi^{11}(\mu) \, d\mu,$$

$$H_y = \frac{1}{2}\int_{-1/2}^{1/2} \ln \varphi^{22}(\mu) \, d\mu.$$

Подстановка этих выражений в (6.6.6) приводит к результату

$$I_{xy}^1 = \frac{1}{2}\int_{-1/2}^{1/2} \ln\left[1 - \frac{|\varphi^{12}(\mu)|^2}{\varphi^{11}(\mu)\varphi^{22}(\mu)}\right] d\mu. \qquad (6.6.12)$$

Поскольку $\frac{|\varphi^{12}(\mu)|^2}{\varphi^{11}(\mu)\varphi^{22}(\mu)} = R_\mu^2$ является как бы квадратом коэффициента корреляции для спектральных составляющих, которые соответствуют значению μ, выражение в правой части равенства (6.6.12) можно интерпретировать как сумму информации различных спектральных составляющих. Каждое же слагаемое определяется простой формулой (6.5.11).

Перейдем к многомерному случаю. Найдем удельную информацию связи группы стационарных гауссовых последовательностей $\{x_t^\alpha\}$, $\alpha = 1, \ldots, r$ с другой группой последовательностей $\{y_t^{\rho-r}\} = \{x_t^\rho\}$, $\rho = r+1, \ldots, r+s$. В совокупности эти последовательности описываются корреляционной матрицей R_{t-q}^{ij}, $i,j = 1, \ldots, r+s$ или матрицей спектральных плотностей

$$\varphi(\mu) = \| \varphi^{ij}(\mu) \| = \begin{pmatrix} \varphi_x(\mu) & \varphi_{xy}(\mu) \\ \varphi_{yx}(\mu) & \varphi_y(\mu) \end{pmatrix} \qquad (6.6.13)$$

($\varphi_{yx}(\mu) = \varphi_{xy}^+(\mu)$ — эрмитово-сопряженная матрица). Здесь $\varphi_x(\mu) = \|\varphi^{\alpha\beta}(\mu)\|$ — матрица плотностей для группы процессов $\{x_t^\alpha\}$, $\varphi_y(\mu) = \|\varphi^{\rho\sigma}(\mu)\|$ — матрица для процессов $\{y_t^{\rho-r}\}$, а $\varphi_{xy}(\mu) = \|\varphi^{\alpha\sigma}(\mu)\|$ — матрица взаимных спектральных функций.

Применяя (5.5.19), по формуле (6.6.6) находим удельную энтропию

$$I_{xy}^1 = \frac{1}{2}\int_{-1/2}^{1/2} [\ln \det \varphi_x(\mu) + \ln \det \varphi_y(\mu) -$$

$$- \ln \det \varphi_{xy}(\mu)] \, d\mu. \qquad (6.6.14)$$

К подынтегральному выражению можно применить все те преобразования, которые от формулы (6.5.3) привели к (6.5.6). После этого равенство (6.6.14) примет вид

$$I_{xy}^1 = -\frac{1}{2}\int_{-1/2}^{1/2} \ln\det\left[1 - \varphi_{xy}(\mu)\,\varphi_y^{-1}(\mu)\,\varphi_{yx}(\mu)\,\varphi_x^{-1}(\mu)\right] d\mu, \tag{6.6.15}$$

что является матричным обобщением формулы (6.6.12).

Приведенные здесь результаты можно получить также при помощи формул (6.6.8), (5.5.20), как будет видно из дальнейшего.

3. Вычислим теперь удельную информацию связи двух групп стационарных гауссовых процессов: $\{x_t^\alpha\}$, $\alpha = 1, \ldots, r$ и $\{y_t^{\rho-r}\} = \{x_t^\rho\}$, $\rho = r+1, \ldots, r+s$, протекающих в непрерывном времени. Они описываются совокупной корреляционной матрицей $R_{t-t'}^{jk}$, $j, k = 1, \ldots, r+s$, или совокупной матрицей спектральных плотностей

$$\mathrm{S}(\omega) = \| S^{jk}(\omega) \| = \begin{pmatrix} \mathrm{S}_x(\omega) & \mathrm{S}_{xy}(\omega) \\ \mathrm{S}_{yx}(\omega) & \mathrm{S}_y(\omega) \end{pmatrix},$$

$$S^{jk}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega\tau} R_\tau^{jk}\, d\tau = S^{*kj}(\omega)$$

($*$ обозначает комплексное сопряжение), аналогичной матрице (6.6.13). Здесь $\mathrm{S}_x(\omega) = \|S^{\alpha\beta}(\omega)\|$ — матрица спектральных плотностей для первой группы $\{x_t^\alpha\}$, $\mathrm{S}_y(\omega) = \|S^{\rho\sigma}(\omega)\|$ — матрица для второй группы $\{y_t^{\rho-r}\}$, $\mathrm{S}_{xy}(\omega) = \|S^{\rho\sigma}(\omega)\|$ — матрица взаимных спектральных функций, связывающих процессы первой и второй групп.

Для вычисления удельной информации связи применим теперь формулу (6.6.8), где удельные энтропии $h_x^{P/Q}$, $h_y^{P/Q}$, $h_{xy}^{P/Q}$ (при $Q_1 = Q_2 = Q$) вычисляются по формуле (5.7.25). В качестве меры Q возьмем гауссову меру, задаваемую совокупной матрицей спектральных плотностей

$$\widetilde{\mathrm{S}}(\omega) = \begin{pmatrix} \widetilde{\mathrm{S}}_x(\omega) & 0 \\ 0 & \widetilde{\mathrm{S}}_y(\omega) \end{pmatrix}. \tag{6.6.16}$$

Взаимные спектральные плотности $\widetilde{S}^{\alpha\sigma}(\omega)$ здесь положены равными нулю для того, чтобы было выполнено условие мультипликативности (6.6.7). Если средние значения $\widetilde{m}$ выбрать совпадающими с m_j ($j = 1, \ldots, r+s$) (что, конечно, совершенно необязательно для получения окончательного результата), из (5.7.25) будем иметь

$$h_x^{P/Q} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \operatorname{Sp} G\left(\widetilde{\mathrm{S}}_x^{-1}(\omega)\,\mathrm{S}_x(\omega)\right) d\omega,$$

$$h_y^{P/Q} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \operatorname{Sp} G(\widetilde{S}_y^{-1}(\omega) S_y(\omega))\, d\omega, \tag{6.6.17}$$

$$h_{xy}^{P/Q} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \operatorname{Sp} G(\widetilde{S}^{-1}(\omega) S(\omega))\, d\omega,$$

где

$$G(z) = (z - 1 - \ln z)/2.$$

Матрицы $\widetilde{S}_x(\omega)$, $\widetilde{S}_y(\omega)$ подбираются таким образом, чтобы указанные *интегралы по ω сходились*. Остается подставить выражения (6.6.17) в (6.6.8). Нетрудно убедиться, что при этом вклады первых двух членов $z/2 - 1/2$ функции $G(z)$, взаимно уничтожатся. В самом деле, вследствие специального вида матрицы (6.6.16) имеем

$$\operatorname{Sp}[\widetilde{S}^{-1} S - 1] = \operatorname{Sp}\begin{pmatrix} \widetilde{S}_x^{-1} S_x - 1 & \widetilde{S}_x^{-1} S_{xy} \\ \widetilde{S}_y^{-1} S_{yx} & \widetilde{S}_y^{-1} S_y - 1 \end{pmatrix} =$$
$$= \operatorname{Sp}[\widetilde{S}_x^{-1} S_x - 1] + \operatorname{Sp}[\widetilde{S}_y^{-1} S_y - 1].$$

Но это есть в точности сумма соответствующих выражений от $G(\widetilde{S}_x^{-1} S_x)$ и $G(\widetilde{S}_y^{-1} S_y)$.

Поэтому в подынтегральном выражении останутся лишь логарифмические члены

$$i_{xy} = h_{xy}^{P/Q} - h_x^{P/Q} - h_y^{P/Q} = \frac{1}{4\pi} \int_{-\infty}^{\infty} [\operatorname{Sp}\ln S_x + \operatorname{Sp}\ln S_y -$$
$$- \operatorname{Sp}\ln S - \operatorname{Sp}\ln\widetilde{S}_x - \operatorname{Sp}\ln\widetilde{S}_y + \operatorname{Sp}\ln\widetilde{S}]\, d\omega.$$

Члены с вспомогательными спектральными плотностями $\widetilde{S}_x$, $\widetilde{S}_y$ полностью выпадают опять-таки из-за отсутствия взаимных корреляций $\widetilde{S}^{\alpha\sigma}(\omega)$, так как

$$\operatorname{Sp}\ln\widetilde{S}_x + \operatorname{Sp}\ln\widetilde{S}_y = \operatorname{Sp}\left[\ln\begin{pmatrix} \widetilde{S}_x & 0 \\ 0 & 1 \end{pmatrix} + \ln\begin{pmatrix} 1 & 0 \\ 0 & \widetilde{S}_y \end{pmatrix}\right] = \operatorname{Sp}\ln\widetilde{S},$$

и мы получаем

$$i_{xy} = \frac{1}{4\pi} \int_{-\infty}^{\infty} [\ln\det S_x(\omega) + \ln\det S_y(\omega) -$$
$$- \ln\det S(\omega)]\, d\omega. \tag{6.6.18}$$

Здесь использована также формула (6.5.4). Подынтегральное выражение аналогично выражению, стоящему в правой части (6.5.3).

Точно так же, как и в § 6.5, его можно преобразовать к виду (6.5.6) или (6.5.7). При этом найденная формула (6.6.18) примет вид

$$i_{xy} = -\frac{1}{4\pi} \int_{-\infty}^{\infty} \ln\det\left[1 - \mathrm{S}_{xy}(\omega)\,\mathrm{S}_y^{-1}(\omega)\,\mathrm{S}_{yx}(\omega)\,\mathrm{S}_x^{-1}(\omega)\right] d\omega. \tag{6.6.19}$$

Очевидна аналогия этого результата с соответствующей формулой (6.6.15) для стационарных последовательностей. В частном случае, когда рассматривается информация связи одного процесса $\{x_t\}$ с одним процессом $\{y_t\}$ из (6.6.19), имеем

$$i_{xy} = -\frac{1}{4\pi} \int_{-\infty}^{\infty} \ln\left[1 - \frac{|S_{xy}(\omega)|^2}{S_x(\omega)\,S_y(\omega)}\right] d\omega.$$

Нетрудно применить полученные формулы к случаю аддитивных независимых помех (частный случай 3 из § 6.5). В этом случае имеют место соотношения

$$y_t^\alpha = x_t^\alpha + \xi_t^\alpha, \qquad \alpha = 1, \ldots, r, \tag{6.6.20}$$

$$\mathrm{S}_{xy}(\omega) = \mathrm{S}_x(\omega), \qquad \mathrm{S}_y(\omega) = \mathrm{S}_x(\omega) + \mathrm{S}_\xi(\omega), \tag{6.6.21}$$

аналогичные (6.5.15), (6.5.16). Подробно тому, как формула (6.5.6) приняла вид (6.5.17), вследствие соотношений (6.6.21) выражение (6.6.19) перейдет в выражение

$$i_{xy} = \frac{1}{4\pi} \int_{-\infty}^{\infty} \ln\det\left[1 + \mathrm{S}_x(\omega)\,\mathrm{S}_\xi^{-1}(\omega)\right] d\omega. \tag{6.6.22}$$

(О затронутых здесь вопросах см. работу Пинскера [1]).

В случае стационарных гауссовых процессов кроме средней удельной информации связи можно вычислить также удельный характеристический потенциал случайной информации. Удельный потенциал выражается через полный характеристический потенциал (6.4.10) обычным предельным переходом

$$\mu^1(s) = \lim_{T \to \infty} \frac{1}{T}\,\mu(s),$$

так что $\mu(s) \approx \mu^1(s)T$. Его нетрудно вычислить при помощи формулы (6.5.25), подобно тому, как удельная информация (6.6.19) может быть вычислена из (6.5.6). Принимая во внимание выше изложенное, нетрудно сообразить, какой вид будет иметь выражение для удельного потенциала в различных случаях. Так, в том случае, когда справедлива формула (6.6.19), удельный потенциал имеет вид

$$\mu^1(s) = -\frac{1}{4\pi} \int_{-\infty}^{\infty} \{\ln\det\left[1 - s^2\,\mathrm{S}_{xy}(\omega)\,\mathrm{S}_y^{-1}(\omega)\,\mathrm{S}_{yx}(\omega)\,\mathrm{S}_x(\omega)\right] +$$
$$+ s\ln\det\left[1 - \mathrm{S}_{xy}(\omega)\,\mathrm{S}_y^{-1}(\omega)\,\mathrm{S}_{yx}(\omega)\,\mathrm{S}_x^{-1}(\omega)\right]\}\, d\omega.$$

Из этого результата можно получить как удельную информацию (6.6.19), так и удельную дисперсию

$$(\mathbf{D}I_{xy})^1 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{Sp}\,[\mathrm{S}_{xy}(\omega)\,\mathrm{S}_y^{-1}(\omega)\,\mathrm{S}_{yx}(\omega)\,\mathrm{S}_x^{-1}(\omega)]\,d\omega,$$

а также прочие удельные статистические характеристики случайной информации.

6.7. Информация связи компонент марковского процесса

1. Пусть заданы произвольные (необязательно стационарные) случайные процессы $\{x_t\}$, $\{y_t\}$ в непрерывном или дискретном времени t. Информация связи этих процессов на фиксированном отрезке $a \leqslant t \leqslant b$ согласно (6.4.9) равна

$$I_{x_a^b,\, y_a^b} = H^{P/Q}_{x_a^b y_a^b} - H^{P/Q}_{x_a^b} - H^{P/Q}_{y_a^b}. \tag{6.7.1}$$

Здесь

$$x_a^b = \{x_t,\, a \leqslant t \leqslant b\}, \qquad y_a^b = \{y_t,\, a \leqslant t \leqslant b\};$$

мера Q предполагается мультипликативной, т. е. удовлетворяющей условию (6.4.8)

$$Q\,(dx_a^b\, dy_a^b) = Q_1\,(dx_a^b)\, Q_2\,(dy_a^b) \tag{6.7.2}$$

при любых a, b. Можно обозначать $Q_1\,(dx_a^b) = Q\,(dx_a^b)$, $Q_2\,(dy_a^b) = Q\,(dy_a^b)$, имея в виду, что Q_1 и Q_2 порождаются мерой $Q\,(dx_a^b dy_a^b)$.

Полагая в (6.7.1) сначала $b = t + \tau$, а затем $b = t$ и взяв разность этих выражений, найдем приращение информации

$$I^\tau_{x,\,y} = H^{P/Q}_{x_a^{t+\tau} y_a^{t+\tau}} - H^{P/Q}_{x_a^t y_a^t} + \left[H^{P/Q}_{x_a^t} - H^{P/Q}_{x_a^{t+\tau}}\right] + \left[H^{P/Q}_{y_a^t} - H^{P/Q}_{y_a^{t+\tau}}\right].$$

Это выражение можно записать при помощи условных энтропий в виде

$$I^\tau_{xy} = H^{P/Q}_{y_t^{t+\tau} \mid x_a^t y_a^t} + H^{P/Q}_{x_t^{t+\tau} \mid x_a^t y_a^{t+\tau}} - H^{P/Q}_{x_t^{t+\tau} \mid x_a^t} - H^{P/Q}_{y_t^{t+\tau} \mid y_a^t}. \tag{6.7.3}$$

В самом деле,

$$H^{P/Q}_{x_a^{t+\tau} y_a^{t+\tau}} - H^{P/Q}_{x_a^t y_a^t} = H^{P/Q}_{x_t^{t+\tau} y_t^{t+\tau} \mid x_a^t y_a^t} =$$

$$= H^{P/Q}_{y_t^{t+\tau} \mid x_a^t y_a^t} + H^{P/Q}_{x_t^{t+\tau} \mid x_a^t y_a^{t+\tau}}.$$

Поскольку каждая разность

$$H^{P/Q}_{y_t^{t+\tau}\mid x_a^t y_a^t}-H^{P/Q}_{y_t^{t+\tau}\mid y_a^t}=I_{y_t^{t+\tau},\, x_a^t\mid y_a^t},$$

$$H^{P/Q}_{x_t^{t+\tau}\mid x_a^t y_a^{t+\tau}}-H^{P/Q}_{x_t^{t+\tau}\mid x_a^t}=I_{x_t^{t+\tau},\, y_a^{t+\tau}\mid x_a^t}$$

неотрицательна [см. (1.7.19), (6.4.11)], то очевидна и неотрицательность приращения (6.7.3).

Если процессы протекают в непрерывном времени, то (6.7.3) можно поделить на τ и перейти к пределу $\tau\to 0$, получив аналогичную формулу

$$i_{xy}(t)=h^{P/Q}_{y\mid x_a^t y_a^t}(t)+h^{P/Q}_{x\mid x_a^t y_a^t}(t)-h_x^{P/Q}(t)-h_y^{P/Q}(t), \qquad (6.7.4)$$

затрагивающую энтропийные плотности

$$h^{P/Q}_{y\mid x_a^t y_a^t}=\lim_{\tau\to 0}\frac{1}{\tau}H^{P/Q}_{y_t^{t+\tau}\mid x_a^t y_a^t},$$

$$h^{P/Q}_{x\mid x_a^t y_a^t}=\lim_{\tau\to 0}\frac{1}{\tau}H^{P/Q}_{x_t^{t+\tau}\mid x_a^t y_a^{t+\tau}},\qquad h^{P/Q}_{y\mid x_a^t y_a^t}+h^{P/Q}_{x\mid x_a^t y_a^t}=h^{P/Q}_{xy},$$

$$h_x^{P/Q}=\lim_{\tau\to 0}\frac{1}{\tau}H^{P/Q}_{x_t^{t+\tau}\mid x_a^t},\qquad h_y^{P/Q}=\lim_{\tau\to 0}\frac{1}{\tau}H^{P/Q}_{y_t^{t+\tau}\mid y_a^t} \qquad (6.7.4a)$$

[см. (5.11.4), (5.11.5), (5.11.9)].

Если процессы $\{x_t\}$, $\{y_t\}$ стационарные и стационарно связанные, то удобно полагать в (6.7.3), (6.7.4) $a=-\infty$. Тогда выражение (6.7.3) будет пропорционально τ (и можно положить $\tau=1$), а плотности в (6.7.4a) не будут зависеть от t. При этом выражение (6.7.4) совпадает с удельной информацией (6.6.3).

Возвращаясь к произвольному случаю, сравним энтропии $H^{P/Q}_{y_t^{t+\tau}\mid x_a^t y_a^t}$ и $H^{P/Q}_{y_t^{t+\tau}\mid y_a^t}$, входящие в правую часть (6.7.3). Очевидно, что вследствие (6.7.2),

$$H^{P/Q}_{y_t^{t+\tau}\mid x_a^t y_a^t}=\mathbf{M}\int\ln\left[\frac{P\left(dy_t^{t+\tau}\mid x_a^t,\, y_a^t\right)}{Q\left(dy_t^{t+\tau}\mid y_a^t\right)}\right]P\left(dy_t^{t+\tau}\mid x_a^t,\, y_a^t\right), \qquad (6.7.5)$$

$$H^{P/Q}_{y_t^{t+\tau}\mid y_a^t}=\mathbf{M}\int\ln\left[\frac{P\left(dy_t^{t+\tau}\mid y_a^t\right)}{Q\left(dy_t^{t+\tau}\mid y_a^t\right)}\right]P\left(dy_t^{t+\tau}\mid y_a^t\right). \qquad (6.7.6)$$

Но вероятность $P\left(dy_t^{t+\tau}\mid y_0^t\right)$ можно записать

$$P\left(dy_t^{t+\tau}\mid y_a^t\right)=\mathbf{M}\left[P\left(dy_t^{t+\tau}\mid x_a^t,\, y_a^t\right)\mid y_a^t\right]. \qquad (6.7.7)$$

Обозначим через $\mathbf{M}_1$ условное усреднение по x_a^t, y_t с весом $P\left(dx_a^t dy_t\mid y_a^t\right)$ и через $\mathbf{M}_2$ усреднение по y_a^t с весом $P(dy_a^t)$. Тогда (6.7.7), (6.7.6) будут в виде $P\left(dy_t^{t+\tau}\mid y_a^t\right)=\mathbf{M}_1 P\left(dy_t^{t+\tau}\mid x_a^t,\, y_a^t\right)$,

$$H^{P/Q}_{y_t^{t+\tau}|y_a^t} = \mathbf{M}_2 \int \ln \frac{P\left(dy_t^{t+\tau} \mid y_a^t\right)}{Q\left(dy_t^{t+\tau} \mid y_a^t\right)} \, P\left(dy_t^{t+\tau} \mid y_a^t\right).$$

Усреднения в (6.7.5) можно представить как последовательные усреднения $\mathbf{M}_2\mathbf{M}_1$. Тогда разность энтропий (6.7.5), (6.7.6). запишется в форме

$$H^{P/Q}_{y_t^{t+\tau}|x_a^t y_a^t} - H^{P/Q}_{y_t^{t+\tau}|y_a^t} = \mathbf{M}_2 \left\{ \mathbf{M}_1 \int \ln \frac{P\left(dy_t^{t+\tau} \mid x_a^t,\, y_a^t\right)}{Q\left(dy_t^{t+\tau} \mid y_a^t\right)} \times \right.$$
$$\times P\left(dy_t^{t+\tau} \mid x_a^t,\, y_a^t\right) - \int \ln \left[\mathbf{M}_1 \frac{P\left(dy_t^{t+\tau} \mid x_a^t,\, y_a^t\right)}{Q\left(dy_t^{t+\tau} \mid y_a^t\right)} \right] \times$$
$$\left. \times \mathbf{M}_1 P\left(dy_t^{t+\tau} \mid x_a^t,\, y_a^t\right) \right\}. \qquad (6.7.8)$$

Аналогично можно представить и вторую разность $H^{P/Q}_{x_t^{t+\tau}|x_a^t y_a^{t+\tau}} - H^{P/Q}_{x_t^{t+\tau}|x_a^t}$, входящую в (6.7.3). Обозначим через $\mathbf{M}_3$ усреднение по $y_a^{t+\tau}$, x_t с весом $P(dy_a^{t+\tau} dx_t \mid x_a^t)$ и через $\mathbf{M}_4$ усреднение по x_a^t с весом $P(dx_a^t)$. Тогда будем иметь

$$H^{P/Q}_{x_t^{t+\tau}|x_a^t y_a^{t+\tau}} - H^{P/Q}_{x_t^{t+\tau}|x_a^t} = \mathbf{M}_4 \left\{ \mathbf{M}_3 \int \ln \frac{P\left(dx_t^{t+\tau} \mid x_a^t,\, y_a^{t+\tau}\right)}{Q\left(dx_t^{t+\tau} \mid x_a^t\right)} \times \right.$$
$$\times P\left(dx_t^{t+\tau} \mid x_a^t,\, y_a^{t+\tau}\right) - \int \ln \left[\mathbf{M}_3 \frac{P\left(dx_t^{t+\tau} \mid x_a^t,\, y_a^{t+\tau}\right)}{Q\left(dx_t^{t+\tau} \mid x_a^t\right)} \right] \times$$
$$\left. \times \mathbf{M}_3 P\left(dx_t^{t+\tau} \mid x_a^t,\, y_a^{t+\tau}\right) \right\}. \qquad (6.7.9)$$

Информация (6.7.3) равна сумме указанных выражений (6.7.8), (6.7.9). Вычитаемые члены в фигурных скобках отличаются друг от друга разным порядком проведения операций усреднения и нелинейного преобразования. Нетрудно убедиться, что мера Q из них, по существу, полностью выпадает.

Этими выражениями, выведенными без использования марковских свойств, удобно пользоваться для вычисления информации связи одной части компонент марковского процесса с другой частью его компонент. Совокупный процесс $\{x_t,\ y_t\} = \{\xi_t\}$ предполагается марковским относительно меры P, а процесс $\{y_t\}$ марковским относительно меры Q. Тогда

$$P\left(dy_t^{t+\tau} \mid x_a^t,\, y_a^t\right) = P\left(dy_t^{t+\tau} \mid x_t,\, y_t\right) \qquad (\tau > 0),$$
$$Q\left(dy_t^{t+\tau} \mid x_a^t,\, y_a^t\right) = Q\left(dy_t^{t+\tau} \mid y_t\right)$$

[см. также (6.7.2)].

Поэтому усреднение $\mathbf{M}_1$ в формуле (6.7.8) сведется к усреднению по $dx_t dy_t$, проводимому с весом

$$P(dx_t\, dy_t \mid y_a^t) \equiv W_t(dx_t\, dy_t).$$

Указанная формула примет вид

$$H^{P/Q}_{y_t^{t+\tau} \mid x_a^t y_a^t} - H^{P/Q}_{y_t^{t+\tau} \mid y_a^t} = \mathbf{M}_2 \left\{ \int\limits_{\xi_t} W_t(d\xi_t) \left[\int\limits_{y_t^{t+\tau}} \ln \frac{P(dy_t^{t+\tau} \mid \xi_t)}{Q(dy_t^{t+\tau} \mid y_t)} \times \right.\right.$$

$$\left. \times P(dy_t^{t+\tau} \mid \xi_t) \right] - \int\limits_{y_t^{t+\tau}} \ln \left[\int\limits_{\xi_t} \frac{P(dy_t^{t+\tau} \mid \xi_t)}{Q(y_t^{t+\tau} \mid y_t)} W_t(d\xi_t) \right] \times$$

$$\left. \times \int\limits_{\xi_t} P(dy_t^{t+\tau} \mid \xi_t) W_t(d\xi_t) \right\}. \tag{6.7.10}$$

Второе усреднение $\mathbf{M}_2$ здесь, очевидно, относится к W_t, и его можно проводить с весом $P(dW_t)$. Если в последнем члене (6.7.10) изменить порядок интегрирования по ξ_t и $y_t^{t+\tau}$, то (6.7.10) можно записать

$$H^{P/Q}_{y_t^{t+\tau} \mid x_a^t y_a^t} - H^{P/Q}_{y_t^{t+\tau} \mid y_a^t} = \mathbf{M}_2 \int\limits_{\xi_t} W_t(d\xi_t) \int\limits_{y_t^{t+\tau}} P(dy_t^{t+\tau} \mid \xi_t) \times$$

$$\times \left\{ \ln \left[\frac{P(dy_t^{t+\tau} \mid \xi_t)}{Q(dy_t^{t+\tau} \mid y_t)} \right] - \ln \left[\frac{\int P(dy_t^{t+\tau} \mid \xi_t') W_t(d\xi_t')}{Q(dy_t^{t+\tau} \mid y_t)} \right] \right\}.$$

Отсюда видно, что мера Q выпадает, и мы имеем

$$H^{P/Q}_{y_t^{t+\tau} \mid x_a^t y_a^t} - H^{P/Q}_{y_t^{t+\tau} \mid y_a^t} = \mathbf{M}_2 \int\limits_{\xi_t} W_t(d\xi_t) \int\limits_{y_t^{t+\tau}} P(dy_t^{t+\tau} \mid \xi_t) \times$$

$$\times \ln \frac{P(dy_t^{t+\tau} \mid \xi_t)}{\int\limits_{\xi_t'} P(dy_t^{t+\tau} \mid \xi_t') W_t(d\xi_t')}. \tag{6.7.11}$$

Рассмотрим второе выражение (6.7.8). Для марковского совокупного процесса $\{x_t, y_t\}$ относительно P и марковского $\{x_t\}$ относительно Q справедливы соотношения

$$P(dx_t^{t+\tau} \mid x_a^t y_a^{t+\tau}) = P(dx_t^{t+\tau} \mid x_t, y_t^{t+\tau}),$$

$$Q(dx_t^{t+\tau} \mid x_a^t) = Q(dx_t^{t+\tau} \mid x_t) \quad (\tau > 0).$$

Кроме того вес, соответствующий усреднению $\mathbf{M}_3$, равен

$$P(dy_t^{t+\tau} dx_t \mid x_a^t) = \int P(dy_t^{t+\tau} \mid x_a^t, y_t) P(dx_t\, dy_t \mid x_a^t) =$$

$$= \int P\left(dy_t^{t+\tau} \mid x_t y_t\right) \widetilde{W}_t\left(dx_t dy_t\right),$$

где $\widetilde{W}_t\left(dx_t dy_t\right) = P\left(dx_t dy_t \mid x_a^t\right)$.

Поэтому формула (6.7.9) переписывается в виде

$$H^{P/Q}_{x_t^{t+\tau} \mid x_a^t y_a^{t+\tau}} - H^{P/Q}_{x_t^{t+\tau} \mid x_a^t} = \mathbf{M}_4 \left\{ \int\limits_{\xi_t y_a^{t+\tau}} \widetilde{W}_t(d\xi_t) P\left(dy_t^{t+\tau} \mid \xi_t\right) \times \right.$$

$$\times \int\limits_{x_t^{t+\tau}} \ln\left[\frac{P\left(dx_t^{t+\tau} \mid x_t y_t^{t+\tau}\right)}{Q\left(dx_t^{t+\tau} \mid x_t\right)}\right] P\left(dx_t^{t+\tau} \mid x_t, y_t^{t+\tau}\right) -$$

$$\left. - \int\limits_{x_t^{t+\tau}} \ln\left[\int\limits_{\xi_t} \frac{P\left(dx_t^{t+\tau} \mid \xi_t\right)}{Q\left(dx_t^{t+\tau} \mid x_t\right)} \widetilde{W}_t(d\xi_t)\right] \int\limits_{\xi_t} P\left(dx_t^{t+\tau} \mid \xi_t\right) \widetilde{W}_t(d\xi_t) \right\}. \tag{6.7.12}$$

Усреднение $\mathbf{M}_4$ здесь относится лишь к $\widetilde{W}_t$ и может проводиться с весом $P\,(d\widetilde{W}_t)$. Формулу (6.7.12) по аналогии с (6.7.11) можно записать также в форме, не содержащей Q:

$$H^{P/Q}_{x_t^{t+\tau} \mid x_a^t y_a^{t+\tau}} - H^{P/Q}_{x_t^{t+\tau} \mid x_a^t} = \mathbf{M}_4 \int\limits_{\xi_t} \widetilde{W}_t(d\xi_t) \times$$

$$\times \int\limits_{x_t^{t+\tau} y_t^{t+\tau}} P\left(dx_t^{t+\tau} dy_t^{t+\tau} \mid \xi_t\right) \ln \frac{P\left(dx_t^{t+\tau} \mid x_t, y_t^{t+\tau}\right)}{\int\limits_{\xi_t'} P\left(dx_t^{t+\tau} \mid \xi_t'\right) \widetilde{W}_t\left(d\xi_t'\right)}. \tag{6.7.13}$$

Сумма выражений (6.7.11), (6.7.13) дает искомую информацию (6.7.3):

$$I_{xy}^{\tau} = \mathbf{M}_2 \int\limits_{\xi_t} W_t(d\xi_t) \int\limits_{y_t^{t+\tau}} P\left(dy_t^{t+\tau} \mid \xi_t\right) \ln \frac{P\left(dy_t^{t+\tau} \mid \xi_t\right)}{\int\limits_{\xi_t'} P\left(dy_t^{t+\tau} \mid \xi_t'\right) W_t(d\xi_t')} +$$

$$+ \mathbf{M}_4 \int\limits_{\xi_t} \widetilde{W}_t(d\xi_t) \int\limits_{x_t^{t+\tau} y_t^{t+\tau}} P\left(dx_t^{t+\tau} dy_t^{t+\tau} \mid \xi_t\right) \times$$

$$\times \ln \frac{P\left(dx_t^{t+\tau} \mid x_t, y_t^{t+\tau}\right)}{\int\limits_{\xi_t'} P\left(dx_t^{t+\tau} \mid \xi_t'\right) \widetilde{W}_t\left(d\xi_t'\right)}.$$

Коротко эту формулу можно записать, объединив оба члена:

$$I_{xy}^{\tau} = \mathbf{M} \ln \frac{P\left(dx_t^{t+\tau} dy_t^{t+\tau} \mid x_t, y_t\right)}{\int\limits_{\xi_t} P\left(dx_t^{t+\tau} \mid \xi_t\right) \widetilde{W}_t(d\xi_t) \int\limits_{\xi_t} P\left(dy_t^{t+\tau} \mid \xi_t\right) W_t(d\xi_t)}. \tag{6.7.14}$$

Для марковского совокупного процесса сумму двух первых членов в (6.7.10), (6.7.12) можно представить в более простом виде

$$H^{P/Q}_{y_t^{t+\tau} \mid x_a^t y_a^t} + H^{P/Q}_{x_t^{t+\tau} \mid x_a^t y_a^{t+\tau}} = H^{P/Q}_{x_t^{t+\tau} y_t^{t+\tau} \mid x_a^t y_a^t} = H^{P/Q}_{\xi_t^{t+\tau} \mid \xi_t} =$$

$$= \int \left[\int_{\xi_t^{t+\tau}} \ln \left[\frac{P\left(d\xi_t^{t+\tau} \mid \xi_t\right)}{Q\left(dx_t^{t+\tau} \mid x_t\right) Q\left(dy_t^{t+\tau} \mid y_t\right)} \right] P\left(d\xi_t^{t+\tau} \mid \xi_t\right) P\left(d\xi_t\right).\right.$$

Учитывая это при суммировании выражений (6.7.10), (6.7.12), получаем результат также в другой форме

$$I_{xy}^{\tau} = \mathbf{M} \int_{\xi_t^{t+\tau}} \ln \left[\frac{P\left(d\xi_t^{t+\tau} \mid \xi_t\right)}{Q\left(d\xi_t^{t+\tau} \mid \xi_t\right)} \right] P\left(d\xi_t^{t+\tau} \mid \xi_t\right) -$$

$$- \mathbf{M}_2 \int_{y_t^{t+\tau}} \ln \left[\int_{\xi_t} \frac{P\left(dy_t^{t+\tau} \mid \xi_t\right)}{Q\left(dy_t^{t+\tau} \mid y_t\right)} W_t\left(d\xi_t\right) \right] \int_{\xi_t} P\left(dy_t^{t+\tau} \mid \xi_t\right) W_t\left(d\xi_t\right) -$$

$$- \mathbf{M}_4 \int_{x_t^{t+\tau}} \ln \left[\int_{\xi_t} \frac{P\left(dx_t^{t+\tau} \mid \xi_t\right)}{Q\left(dx_t^{t+\tau} \mid y_t\right)} \widetilde{W}_t\left(d\xi_t\right) \right] \int_{\xi_t} P\left(dx_t^{t+\tau} \mid \xi_t\right) \widetilde{W}_t\left(d\xi_t\right), \tag{6.7.15}$$

где усреднения $\mathbf{M}_1$, $\mathbf{M}_2$, $\mathbf{M}_3$ проводятся по ξ_t, W_t, $\widetilde{W}_t$ соответственно. В стационарном случае для этой цели нужно использовать стационарные распределения $P_{\text{ст}}(d\xi_t)$, $P_{\text{ст}}(dW_t)$, $P_{\text{ст}}(d\widetilde{W}_t)$. При этом формула (6.7.15) дает (при $\tau = 1$) удельную информацию i_{xy}. В случае непрерывного времени при вычислении по формуле (6.7.15) удобно полагать τ малым и (поделив предварительно выражение на τ) перейти к пределу $\tau \to 0$.

Изложенный метод вычисления информации связи между частями компонент марковского процесса тесно соприкасается с методом вычисления энтропии части компонент марковского процесса, приведенным в § 5.11. Формулу (6.7.15) можно получить короче при помощи формул (6.7.3), (5.11.12а).

Особый интерес представляет тот случай, когда один из процессов, скажем $\{x_t\}$, является марковским сам по себе (относительно P). Тогда

$$H^{P/Q}_{x_t^{t+\tau} \mid x_a^t} = H^{P/Q}_{x_t^{t+\tau} \mid x_t}, \quad H^{P/Q}_{x_t^{t+\tau} \mid x_a^t y_a^t} = H^{P/Q}_{x_t^{t+\tau} \mid x_t},$$

и в формуле

$$I_{xy}^{\tau} = H^{P/Q}_{x_t^{t+\tau} \mid x_a^t y_a^t} + H^{P/Q}_{y_t^{t+\tau} \mid x_a^{t+\tau} y_a^t} - H^{P/Q}_{x_t^{t+\tau} \mid x_a^t} - H^{P/Q}_{y_t^{t+\tau} \mid y_a^t},$$

аналогичной (6.7.3), остаются лишь два члена

$$I_{xy}^{\tau}=H^{P/Q}_{y^{t+\tau}\mid x_a^{t+\tau}y_a^t}-H^{P/Q}_{y_t^{t+\tau}\mid y_a^t}=H^{P/Q}_{y_t^{t+\tau}\mid x_t^{t+\tau}y_t}-H^{P/Q}_{y_t^{t+\tau}\mid y_a^t}.$$

Поэтому для вычисления информации I_{xy}^{τ} оказывается достаточно формулы (6.7.12), где нужно поменять местами x и y. Именно

$$I_{xy}^{\tau}=\mathbf{M}\int\limits_{y_t^{t+\tau}}\ln\frac{P\left(dy_t^{t+\tau}\mid x_t^{t+\tau},\ y_t\right)}{Q\left(dy_t^{t+\tau}\mid y_t\right)}P\left(dy_t^{t+\tau}\mid x_t^{t+\tau},\ y_t\right)-$$

$$-\mathbf{M}_2\int\limits_{y_t^{t+\tau}}\ln\left[\int\limits_{\xi_t}\frac{P\left(dy_t^{t+\tau}\mid\xi_t\right)}{Q\left(dy_t^{t+\tau}\mid y_t\right)}W_t(d\xi_t)\right]\int\limits_{\xi_t}P\left(dy_t^{t+\tau}\mid\xi_t\right)W_t(d\xi_t). \tag{6.7.16}$$

2. Переходя к рассмотрению различных частных случаев начнем с того случая, когда $\{x_t,\ y_t\}$ есть дискретный стационарный марковский процесс в дискретном времени, т. е. марковская цепь. При этом в формулах предыдущего пункта нет надобности вводить меру Q, ее можно опустить, заменив $P(dx_\beta^\alpha/Q\ (dx_\alpha^\beta)$ на $P(x_\alpha^\beta)$ и $H^{P/Q}$ на $-H$. Можно также непосредственно использовать результаты § 5.2 и 5.3.

Пусть, как и в § 5.2, 5.3, марковская цепь описывается вероятностями перехода $\pi\ (x,\ y;\ x',\ y')$. Применяя формулу (5.2.8), находим удельную энтропию комбинированного процесса

$$h_{xy}=-\sum_{x,\ y}P_{\text{ст}}(x,\ y)\sum_{x',\ y'}\pi(x,\ y;\ x',\ y')\ln\pi(x,\ y;\ x',\ y'). \tag{6.7.17}$$

Для компонент x и y по отдельности удельная энтропия выражена формулой (5.3.23):

$$h_x=-\int P_{\text{ст}}(d\widetilde{W})\sum_{x,\ y,\ x',\ y'}\widetilde{W}(x,\ y)\pi(x,\ y,\ x',\ y')\times$$

$$\times\ln\sum_{x'',\ y'',\ y'''}\widetilde{W}(x'',\ y'')\pi(x'',\ y'';\ x',\ y'''),$$

$$h_y=-\int P_{\text{ст}}(dW)\sum_{x,y,\ x',\ y'}W(x,\ y)\pi(x,\ y;\ x',\ y')\times$$

$$\times\ln\sum_{x'',\ y'',\ x'''}W(x'',\ y'')\pi(x'',\ y'',\ x''',\ y').$$

В приведенных формулах стационарное распределение $P_{\text{ст}}(x,\ y)$ является решением уравнения

$$\sum_{x,\ y}P_{\text{ст}}(x,\ y)\pi(x,\ y;\ x',\ y')=P_{\text{ст}}(x',\ y') \tag{6.7.18}$$

[см. (5.2.7)], а распределение $P_{\text{ст}}(dW)=p_{\text{ст}}\ (W)\prod_{\xi}dW(\xi)$ — решением аналогичного уравнения, но соответствующего вторичному апостериорному марковскому процессу, имеющему вероятности перехода (5.3.22). Это уравнение имеет вид

$$\sum_{y_k}\int p_{\text{ст}}(W(x,\ y))\times$$

$$\times\prod_{x',\ y'}\delta\left(W'(x',\ y')-\frac{\delta_{y'y_k}\sum\limits_{x,\ y}W(x,\ y)\,\pi(x,\ y;\ x',\ y')}{\sum\limits_{x,\ y,\ x'''}W(x,\ y)\,\pi(x,\ y;\ x''',\ y_k)}\right)\times$$

$$\times\sum_{x,\ y,\ x'''}W(x,\ y)\,\pi(x,\ y;\ x''',\ y_k)\prod_{x,\ y}dW(x,\ y)=$$

$$=p_{\text{ст}}(W'(x',\ y')).$$

Аналогично записывается и уравнение

$$\sum_{x_k}\int p_{\text{ст}}(\widetilde{W}(x,\ y))\times$$

$$\times\prod_{x',\ y'}\delta\left(\widetilde{W}'(x',\ y')-\frac{\delta_{x'x_k}\sum\limits_{x,\ y}\widetilde{W}(x,\ y)\,\pi(x,\ y;\ x',\ y')}{\sum\limits_{x,\ y,\ y'''}\widetilde{W}(x,\ y)\,\pi(x,\ y;\ x_k,\ y''')}\right)\times$$

$$\times\sum_{x,\ y,\ y'''}\widetilde{W}(x,\ y)\,\pi(x,\ y,\ x_k,\ y''')\prod_{x,\ y}d\widetilde{W}(x,\ y)=$$

$$=p_{\text{ст}}(\widetilde{W}'(x',\ y')),$$

служащее для определения стационарного распределения $p_{\text{ст}}(W)$, которое соответствует вторичному апостериорному марковскому процессу $\widetilde{W}$ (когда наблюдается процесс $\{x_t\}$). Приведенные формулы, в принципе, решают задачу вычисления удельной информации

$$i_{xy}=h_x+h_y-h_{xy}. \tag{6.7.19}$$

Пример. Возьмем для примера марковский процесс $\{\xi_t\}$ с тремя состояниями $\xi=a,\ b,\ c$, рассмотренный в § 5.3. Его представляем как комбинацию $\{x_t,\ y_t\}$ процессов $\{x_t\}$ и $\{y_t\}$, каждый из которых имеет два состояния $x,\ y=1,\ 2$. Состояния $\xi=a$, $\xi=b$, $\xi=c$ интерпретируются как комбинированные состояния

$$(\xi=a)=(x=1,\ y=1),\qquad(\xi=b)=(x=1,\ y=2),$$
$$(\xi=c)=(x=2,\ y=2).$$

Берем матрицу перехода

$$\|\pi(x,\ y;\ x',\ y')\|=\begin{pmatrix}\pi(1,\ 1;\ 1,\ 1) & \pi(1,\ 1;\ 1,\ 2) & \pi(1,\ 1;\ 2,\ 2)\\ \pi(1,\ 2;\ 1,\ 1) & \pi(1,\ 2;\ 1,\ 2) & \pi(1,\ 2;\ 2,\ 2)\\ \pi(2,\ 2;\ 1,\ 1) & \pi(2,\ 2;\ 1,\ 2) & \pi(2,\ 2;\ 2,\ 2)\end{pmatrix}=$$

$$=\begin{pmatrix}1-\lambda-\mu & \lambda & \mu\\ \nu & 1-2\nu & \nu\\ \mu & \lambda & 1-\lambda-\mu\end{pmatrix}, \tag{6.7.20}$$

которая обладает определенной симметрией. Она не меняется, если процессы $\{x\}$, $\{y\}$ поменять ролями (т. е. при замене $x \rightleftarrows y$, $1 \rightleftarrows 2$), поэтому при такой матрице перехода имеем

$$h_x = h_y. \tag{6.7.20a}$$

Найдем энтропию h_{xy} по формуле (6.7.17). Используя обозначение (5.2.22), имеем

$$\begin{gathered}
-\sum_{x',\,y'} \pi(1,\ 1;\ x',\ y') \ln \pi(1,\ 1;\ x',\ y') = \\
= -\sum_{x',\,y'} \pi(2,\ 2;\ x',\ y') \ln \pi(2,\ 2;\ x',\ y') = h_3(\lambda,\ \mu), \\
-\sum_{x',\,y'} \pi(1,\ 2;\ x',\ y') \ln \pi(1,\ 2;\ x',\ y') = \\
= h_3(\nu,\ \nu) = -2\nu \ln \nu - (1-2\nu) \ln (1-2\nu).
\end{gathered} \tag{6.7.21}$$

Стационарное распределение $P_{\text{ст}}(x,\ y)$ обладает свойством симметрии $P_{\text{ст}}(1,1) = P_{\text{ст}}(2,2)$. При его использовании уравнение (6.7.18) дает

$$P_{\text{ст}}(1,\ 1)(1-\lambda) + P_{\text{ст}}(1,\ 2)\nu = P_{\text{ст}}(1,\ 1);$$

отсюда

$$P_{\text{ст}}(1,\ 1)\lambda = P_{\text{ст}}(1,\ 2)\nu, \qquad P_{\text{ст}}(1,\ 1) = \frac{\nu}{\lambda+2\nu},$$

$$P_{\text{ст}}(1,\ 2) = \frac{\lambda}{\lambda+2\nu}, \qquad P_{\text{ст}}(2,\ 2) = \frac{\nu}{\lambda+2\nu},$$

и из (6.7.17) в силу (6.7.21) получаем

$$h_{xy} = \frac{2\nu}{\lambda+2\nu} h_3(\lambda,\ \mu) + \frac{\lambda}{\lambda+2\nu} h_3(\nu,\ \nu). \tag{6.7.22}$$

Для отыскания информации (6.7.19) вследствие (6.7.20а) остается найти энтропию h_x или h_y. Для данного примера она была найдена в п. 3 § 5.3. Согласно (5.3.36), (5.3.37) она вычисляется по формуле

$$h_y = \frac{h_2(p_1) + \sum\limits_{k=1}^{\infty} p_1 \ldots p_k h_2(p_{k+1})}{1 + \sum\limits_{k=1}^{\infty} p_1 \ldots p_k}, \tag{6.7.23}$$

где p_1, p_2, ... — величины, определяемые соотношениями (5.3.30), (5.3.32), (5.3.34) и т. д. Для рассматриваемого случая (6.7.20) эти соотношения принимают вид

$$p_1 = \lambda + \mu, \qquad p_2 = \frac{\lambda(1-\nu) + \mu(1-\mu)}{\lambda+\mu},$$

$$p_3 = \frac{[\lambda(1-2\nu) + \mu\nu](1-\nu) + [\lambda\nu + \mu(1-\lambda-\mu)](1-\mu)}{\lambda(1-\nu) + \mu(1-\mu)},$$

$$\ldots \qquad \ldots\ldots\ldots\ldots\ldots \qquad \ldots$$

Любые p_k находятся последовательно по методу, описанному в в § 5.3. Взяв разность $2h_y - h_{xy}$ выражений (6.7.22), (6.7.23), получим искомую удельную информацию связи i_{xy}.

3. Предположим теперь, что $\{x_t, y_t\}$ есть дискретный марковский процесс в непрерывном времени. Он описывается дифференциальной матрицей перехода $\pi(x, y; x', y')$, т. е. уравнением

$$\frac{dP_t(x_t', y_t')}{dt} = \sum_{x_t, y_t} P_t(x_t, y_t)\, \pi(x_t, y_t; x_t', y_t'). \tag{6.7.24}$$

При помощи результатов § 5.9, 5.11 или формул п. 1 настоящего параграфа может быть вычислена плотность информации связи i_{xy}, совпадающая в стационарном случае с удельной информацией связи.

Чтобы плотность информации i_{xy} была конечна (см. ниже) будем предполагать, что матрица $\pi(x, y; x', y')$ имеет следующий специальный вид [ср. с (5.11.26)]:

$$\pi(x, y; x', y') = \pi_1(x, y; x')\,\delta_{yy'} + \pi_2(x, y; y')\,\delta_{xx'} \tag{6.7.25}$$

$$\left(\pi_1(x, y; x) = -\sum_{x' \neq x} \pi_1(x, y; x'); \qquad \pi_2(x, y; y) = \right.$$
$$\left. = -\sum_{y' \neq y} \pi_2(x, y; y')\right)$$

Вместо пуассоновской меры в (5.11.27) выбираем теперь марковские меры $Q_1(x_0^T)$ и $Q_2(y_0^T)$ с дифференциальными матрицами перехода

$$\pi^{Q_1}(x, x') = \begin{cases} -1 & \text{при } x' = x, \\ 1/(m-1) & \text{при } x' \neq x; \end{cases}$$
$$\pi^{Q_2}(y, y') = \begin{cases} -1 & \text{при } y' = y, \\ 1/(l-1) & \text{при } y' \neq y \end{cases}$$

(m — число состояний процесса $x(t)$, а l — процесса $y(t)$). Это означает, что комбинированная мера $Q(x_0^T y_0^T) = Q_1(x_0^T)\, Q_2(y_0^T)$ описывается матрицей

$$\pi^Q(x, y; x', y') = -2\delta_{xx'}\delta_{yy'} + \frac{1}{m-1}(1 - \delta_{xx'})\,\delta_{yy'} +$$
$$+ \frac{1}{l-1}\,\delta_{xx'}(1 - \delta_{yy'}), \tag{6.7.26}$$

вид которой напоминает (6.7.25).

Вычисления будем проводить по формуле

$$i_{xy} = h_{xy}^{P/Q} - h_x^{PQ_1} - h_y^{P/Q_2} \tag{6.7.27}$$

[см. (6.7.4)]. Вследствие (5.11.12) по аналогии с (5.11.28) имеем

$$h_y^{P/Q_2} = 1 + \mathbf{M} \sum_{y' \neq y} \pi_W(y, y') \ln\left[\frac{l-1}{e}\, \pi_W(y, y')\right], \tag{6.7.28}$$

где

$$\pi_W(y, y') = \sum_x W(x) \sum_{x'} \pi(x, y; x', y'). \qquad (6.7.29)$$

Аналогично

$$h_x^{P/Q_1} = 1 + \mathbf{M} \sum_{x' \neq x} \pi_{\widetilde{W}}(x, x') \ln\left[\frac{m-1}{e} \pi_{\widetilde{W}}(x, x')\right], \qquad (6.7.30)$$

где

$$\pi_{\widetilde{W}}(x, x') = \sum_y \widetilde{W}(y) \sum_{y'} \pi(x, y; x', y'). \qquad (6.7.31)$$

Плотность $h_{xy}^{P/Q}$ подсчитывается методами, изложенными в § 5.9. Для того чтобы эта величина была конечной, и нужен специальный вид (6.7.25) комбинированной матрицы перехода. При помощи формулы (5.9.8) для матриц (6.7.25), (6.7.26) получаем

$$h_{xy}^{P/Q} = 2 + \mathbf{M} \sum_{x' \neq x} \pi_1(x, y; x') \{\ln[(m-1)\pi_1(x, y; x')] - 1\} +$$
$$+ \mathbf{M} \sum_{y' \neq y} \pi_2(x, y; y') \{\ln[(l-1)\pi_2(x, y; y')] - 1\}. \qquad (6.7.32)$$

Вследствие (6.7.25) равенства (6.7.29), (6.7.31) принимают вид

$$\pi_W(y, y') = \sum_x W(x) \pi_2(x, y; y'),$$
$$(6.7.33)$$
$$\pi_{\widetilde{W}}(x, x') = \sum_y \widetilde{W}(y) \pi_1(x, y; x').$$

При подстановке (6.7.28), (6.7.30), (6.7.32) в (6.7.27) учтем, что

$$\mathbf{M}\pi_2(x, y; y') = \mathbf{MM}[\pi_2(x, y; y') \mid y_0^t] =$$
$$= \mathbf{M} \sum_x \pi_2(x, y; y') P(x \mid y_0^t) = \mathbf{M}\pi_W(y, y')$$

и аналогично

$$\mathbf{M}\,\pi_1(x, y; x') = \mathbf{M}\pi_{\widetilde{W}}(x, x').$$

Последние равенства позволяют сократить линейные по π_1 и π_2 члены и записать результат в следующей форме:

$$i_{xy} = \sum_{x, y} P(x, y) \left\{ \sum_{x' \neq x} \pi_1(x, y; x') \ln \pi_1(x, y; x') + \right.$$
$$\left. + \sum_{y' \neq y} \pi_2(x, y; y') \ln \pi_2(x, y; y') \right\} - \int P(d\widetilde{W} dx) \times$$
$$\times \sum_{x' \neq x} \pi_{\widetilde{W}}(x, x') \ln \pi_{\widetilde{W}}(x, x') -$$
$$- \int P(dW\,dy) \sum_{y' \neq y} \pi_W(y, y') \ln \pi_W(y, y'). \qquad (6.7.34)$$

Параметры мер Q_1, Q_2 из окончательного результата, разумеется, выпали.

Распределения $P(x, y)$, $P(d\widetilde{W}dx)$, $P(d\widetilde{W}dy)$ отыскиваются как распределения вероятностей марковских процессов с известными вероятностями перехода. В стационарном случае для вычисления удельной информации нужно брать соответствующие стационарные распределения.

В п. 4 § 5.11 был рассмотрен частный случай, когда процесс $\{x_t\}$, взятый в отдельности является марковским. В этом случае формула (6.7.34) упрощается, из нее выпадают члены с суммированием по x'. Информация связи при этом оказывается равной разности энтропий (5.11.28) и (5.11.30).

4. Перейдем к рассмотрению диффузионных процессов. Будем предполагать, что комбинированный процесс $\{x_t, y_t\} = \{\xi_t\}$ является диффузионным процессом, описываемым вектором сносов $a_j(\xi, t)$, $j = 1, \ldots, r+s$, и совокупной матрицей локальных дисперсий

$$b(\xi, t) = \begin{pmatrix} b_x(x, t) & 0 \\ 0 & b_y(y, t) \end{pmatrix}.$$

Подматрица $b_x = \| b_{\alpha\beta} \|$, $\alpha, \beta = 1, \ldots, r$ относится к компонентам $x_\alpha = \xi_\alpha$ x-процесса; подматрица $b_y = \| b_{\sigma\rho} \|$, $\sigma, \rho = r + 1, \ldots, r+s$ соответствует y-процессу, т. е. другой части компонент совокупного процесса $y_{\rho-r}(t) = \xi_\rho$. Перекрестные локальные коэффициенты предполагаются равными нулю: $b_{\alpha\sigma} = 0$, $\alpha = 1, \ldots, r$; $\sigma = r+1, \ldots, r+s$, чтобы не было бесконечных значений плотности информации i_{xy}. Предполагается также, что подматрица b_x не зависит от $y_1, \ldots, y_s$, а b_y не зависит от $x_1, \ldots, x_r$, и эти подматрицы являются невырожденными.

При этих условиях в качестве меры Q удобно взять меру, при которой процессы $\{x_t\}$, $\{y_t\}$ являются независимыми диффузионными марковскими процессами каждый по отдельности. Первый имеет локальные дисерсии b_x и нулевые сносы, второй — локальные дисперсии b_y и нулевые сносы. При таком выборе можно найти плотности энтропии $h_{xy}^{P/Q}$, $h_x^{P/Q}$, $h_y^{P/Q}$ по формулам (5.10.12), (5.11.17), а значит, и плотность информации $i_{xy} = h_{xy}^{P/Q} - h_x^{P/Q} - h_y^{P/Q}$. Иначе, не вводя меру Q, информацию можно находить по формуле (6.7.14). Любым способом получаем следующий результат:

$$i_{xy} = \frac{1}{2} \mathbf{M} \sum_{j,k=1}^{r+s} a_j b_{jk}^{-1} a_k -$$

$$- \frac{1}{2} \mathbf{M} \sum_{\alpha,\beta=1}^{r} \left[\int \widetilde{W}_t(dz) a_\alpha \right] b_{\alpha\beta}^{-1} \left[\int \widetilde{W}_t(dz) a_\beta \right] -$$

$$-\frac{1}{2}\mathbf{M}\sum_{\rho,\,\sigma=r+1}^{r+s}\left[\int a_\rho W(dz)\right] b_{\rho\sigma}^{-1}\left[\int a_\sigma W(dz)\right].$$

Его можно записать в следующей форме:

$$i_{xy}=\frac{1}{2}\mathbf{M}\sum_{\alpha,\,\beta=1}^{r}\left[a_\alpha-\int a_\alpha \widetilde{W}(dz)\right] b_{\alpha\beta}^{-1}\left[a_\beta-\int a_{\alpha\beta}\widetilde{W}(dz)\right]+$$
$$+\frac{1}{2}\mathbf{M}\sum_{\rho,\,\sigma=r+1}^{r+s}\left[a_\rho-\int a_\rho W(dz)\right] b_{\rho\sigma}^{-1}\left[a_\sigma-\int a_\sigma W(dz)\right], \tag{6.7.35}$$

из которой видна неотрицательность найденного выражения (в силу положительной определенности матрицы $\| b_{\rho\sigma}^{-1}\|$).

5. В заключение этого параграфа положим, что процесс $\{x_t\}$ является марковским в отдельности, процесс же $\{y_t\}$ при фиксированной реализации $\{x_t\}$ является многомерным диффузионным процессом с параметрами $a_\rho(x, y, t)$, $b_{\rho\sigma}(y, t)$, $\rho, \sigma = 1, \ldots, s$. Матрицу локальных дисперсий $b_{\rho\sigma}$ предполагаем не зависящей от x и невырожденной.

Для вычисления информации I_{xy}^{τ} или i_{xy} в данном случае можно воспользоваться формулой (6.7.16), определяя каждый из двух входящих в правую часть членов при помощи формул (5.10.12), (5.11.17) или (5.11.25). Получаемой в результате этого разности

$$i_{xy}=\frac{1}{2}\mathbf{M}\sum_{\rho,\sigma=1}^{s} a_\rho(x, y)\, b_{\rho\sigma}^{-1}\, a_\sigma(x, y)-$$
$$-\frac{1}{2}\mathbf{M}\sum_{\rho,\,\sigma=1}^{s}\left[\int a_\rho(x, y)\, W_t(dx)\right] b_{\rho\sigma}^{-1}\left[\int a_\sigma(x, y))\, W_t(dx)\right]$$

можно придать вид

$$i_{xy}=\frac{1}{2}\mathbf{M}\sum_{\rho,\,\sigma=1}^{s}\left[a_\rho-\int a_\rho W_t(dx)\right] b_{\rho\sigma}^{-1}\left[a_\sigma-\int a_\sigma W_t(dx)\right]\equiv$$
$$\equiv\frac{1}{2}\mathbf{M}_2\mathbf{M}_{ps}\sum_{\rho,\,\sigma=1}^{s}(a_\rho-\mathbf{M}_{ps}\, a_\rho)\, b_{\rho\sigma}^{-1}(a_\sigma-\mathbf{M}_{ps}\, a_\sigma). \tag{6.7.36}$$

Здесь, как и в (6.7.35), $W_t(dx)$ — апостериорное распределение $W_t(dx_t) = P(dx_t \mid y_a^t)$, представляющее собой (вместе с y_t) вторичный марковский W-процесс с вероятностями перехода, определяемыми теорией условных марковских процессов.

Пример. В п. 2 § 5.11 был рассмотрен в качестве примера марковский процесс x_t с двумя состояниями и матрицей перехода (5.9.6). Процесс y_t представлял собой одномерный диффузионный процесс, для которого снос $a(x)$ (но не локальная дисперсия b) зависел от x. Зависимость параметров a и b от y и t предполагалась отсутствующей.

Апостериорное распределение $W(dx)$ в этом примере определяется лишь одной переменной $z_t = W_t(1) - W_t(2)$, причем

$$W_t(1) = (1 + z_t)/2, \qquad W_t(2) = (1 - z_t)/2.$$

Поэтому

$$a(1) - \sum_z a(z) W(z) = [a(1) - a(2)] W(2) = [a(1) - a(2)] \frac{1-z}{2},$$

$$a(2) - \sum_z a(z) W(z) = [a(2) - a(1)] W(1) = -[a(1) - a(2)] \frac{1+z}{2},$$

и формула (6.7.36) принимает вид

$$i_{xy} = \frac{1}{8b} \left[a(1) - a(2) \right]^2 \int_{-1}^{1} (1 - z^2) p_{\text{ст}}(z) \, dz. \tag{6.7.37}$$

Плотность распределения $p_{\text{ст}}(z)$ указана в п. 2 § 5.11. Интеграл в (6.7.37) вычисляется, и согласно (5.11.24) имеем

$$i_{xy} = \frac{1}{2b} [a(1) - a(2)]^2 K_q(b\sqrt{\mu\nu}) \Big\{ 2K_q(b\sqrt{\mu\nu}) +$$
$$+ \sqrt{\frac{\nu}{\mu}} K_{q+1}(b\sqrt{\mu\nu}) + \sqrt{\frac{\mu}{\nu}} K_{q-1}(b\sqrt{\mu\nu}) \Big\}^{-1}$$
$$\left(q = \frac{1}{2} b(\nu - \mu) \right).$$

Найденная информация есть не что иное, как разность вычисленных ранее энтропий (5.11.20) и (5.11.22).

В заключение отметим, что результат (6.7.36) справедлив не только в том случае, когда процесс $\{x_t\}$ является марковским. Для его справедливости нужно лишь, чтобы условный процесс y_t, описываемый мерой $P(dy_a^b \mid x_a^b)$, был диффузионным и каузально-рандомизированно зависящим от x_a^t, так что $H^{P/Q}_{x_t^{t+\tau} \mid x_a^t y_a^t} = H^{P/Q}_{x_t^{t+\tau} \mid x_a^t}$. Данное обобщение нетрудно получить из формулы (6.7.12) (при замене в ней x на y), не ограниченной условием Маркова для x_t. Изложенная в п. 1 теория, не использующая марковские свойства, является полезной также для получения и других результатов.

Глава 7

ПЕРЕДАЧА СООБЩЕНИЙ ПРИ НАЛИЧИИ ПОМЕХ. ВТОРАЯ АСИМПТОТИЧЕСКАЯ ТЕОРЕМА В РАЗЛИЧНЫХ ФОРМУЛИРОВКАХ

В этой главе приводятся важнейшие асимптотические результаты, касающиеся существования оптимальных кодов для каналов с помехами. Доказывается, что шенноновское количество информации является границей для асимптотически безошибочно передаваемого хартлиевского количества информации. В этом состоит вторая асимптотическая теорема. Приводятся формулы, показывающие быстроту убывания вероятности ошибки декодирования при увеличении длины блока. Изложение этих результатов в отличие от общепринятого ведется в терминах шенноновского количества информации, а не пропускной способности (т. е. не производится максимизация предельного количества информации по распределению вероятностей входной переменной).

Теоремы 7.1, 7.3, 7.5 последовательно усиливают одна другую. Такое расположение целесообразно потому, что в таком виде легче знакомиться с этим материалом. С точки зрения результатов каждая из них, конечно, делает излишними предыдущие. Усиление результатов, однако, достигается усложнением доказательства. Имеет некоторое основание та точка зрения, что это усложнение не окупается значимостью усиления (переход от теоремы 7.3 к значительно более сложной теореме 7.5 мало сказывается на поведении коэффициента α в области, где R близко к I_{xy}); мы излагаем все эти теоремы, чтобы читатель по желанию мог остановиться на любой из них.

Все указанные результаты могут быть распространены со случая блоков из независимых одинаково распределенных случайных величин на случай произвольного семейства информационно-устойчивых случайных величин. Это обобщение производится стандартным способом, и мы касаемся его лишь один раз (в теореме 7.2).

Изложение в настоящей главе ведется в простейшем дискретном варианте, однако результаты имеют общее значение. Распространение их на общий случай связано, по существу, лишь с изменением способа записи формул.

7.1. Принципы передачи и приема информации при наличии помех

Рассмотрим некоторый канал связи. Его входную переменную (в выбранный момент времени), которую условно будем называть передаваемым символом или буквой, обозначим через x. Она может принимать дискретные значения из некоторого множества X. Удобно предполагать заданными также вероятности $P(x)$, тогда x будет заданной случайной величиной.

Будем рассматривать канал с помехами. Это значит, что при фиксированном значении x переменная на выходе канала (в фиксированный момент) является случайной, т. е. описывается условными вероятностями $P(y|x)$. Случайную величину y можно назвать принимаемым символом или буквой.

Предполагается, что процесс передачи буквы x и приема буквы y может совершаться многократно с теми же самыми вероятностями $P(x)$, $P(y|x)$ (хотя возможно обобщение на случай переменных вероятностей, см. теорему 7.2). Пусть n букв образуют блок или слово, например

$$\xi = (x_1, \dots, x_n), \qquad \eta = (y_1, \dots, y_n) \quad (\xi \in X^n, \ \eta \in Y^n). \qquad (7.1.1)$$

Желая передать по каналу какие-либо сообщения, мы естественно должны связать эти сообщения с входными словами при помощи некоторого кода. Тогда получатель на приемном конце, прочтя принимаемое слово и зная код, будет пытаться восстановить переданное сообщение.

Поскольку в канале имеются помехи, есть возможность, что получатель ошибется и примет не то сообщение, которое было передано. Код нужно подбирать из тех соображений, чтобы вероятность подобной ошибки была как можно меньше. Представляет принципиальный и практический интерес вопрос о том, чего можно добиться, выбирая хорошие коды, а также, какие коды являются хорошими и как их выбирать.

Особенно ясным является случай простых помех (см. § 6.1). В этом случае передаваемое сообщение, очевидно, нужно связывать с той или иной областью — подмножеством E_k значений x, а значит и областью G_k значений y. Для простых помех при передаче буквы x из области E_k принимаемая буква y обязательно принадлежит области G_k. Поэтому, если сообщение сопоставлять с областями E_k и G_k (или, что то же, с их номером k), то прием сообщений будет безошибочным несмотря на наличие помех в канале связи. При этом, конечно, число передаваемых сообщений не должно превышать число L областей E_k ($k = 1, \dots, L$) равное числу областей G_k). Каждая буква, очевидно, способна безошибочно передавать $\ln L$ единиц информации (нат), а слово из n букв — $n \ln L$ единиц. Попытка передать через канал больше информации неизбежно приведет к появлению ошибок.

Большой заслугой Шеннона является открытие того факта, что нечто аналогичное имеет место для произвольных (не простых) помех в асимптотическом случае $n \to \infty$. Если количество информации, приходящееся на букву, меньше некоторого предела, то вероятность ошибочного приема сообщения можно сделать в среднем сколь угодно малой, увеличивая длину n слова, при использовании хорошего кода. Оказалось далее, что такие хорошие коды найти нетрудно: для этого достаточно выбирать кодовые слова наудачу случайным образом. Эти вопросы, а также приемы декодирования будут рассмотрены нами в настоящей главе.

Предполагая независимость процессов передачи последовательных букв, легко записать вероятности слов (7.1.1) через исходные вероятности

$$P(\xi) = \prod_{i=1}^{n} P(x_i), \quad P(\eta \mid \xi) = \prod_{i=1}^{n} P(y_i \mid x_i),$$
$$P(\xi, \eta) = P(\xi) P(\eta \mid \xi) \tag{7.1.2}$$

(в подобном случае будем говорить, что канал $[P(\xi), P(\eta|\xi)]$ является n-й *степенью канала* $[P(x), P(y|x)]$).

Зададимся целью передать словами из n букв M сообщений, т. е. передать количество информации $\ln M$. Можно попытаться сделать это следующим образом. Из всех возможных слов типа $\xi = (x_1, \ldots, x_n)$ выберем M различных слов

$$\xi_1 = (x_{11}, \ldots, x_{n1}), \ \ldots, \ \xi_M = (x_{1M}, \ldots, x_{nM}). \tag{7.1.3}$$

Их совокупность составит код, который должен быть известен как на передающем, так и на приемном конце канала. Каждое из M сообщений сопоставляется с одним из кодовых слов (7.1.3), скажем k-е сообщение передается словом ξ_k. При этом слово на приемном конце может быть случайным, иметь разброс из-за действия помех. Ему соответствуют вероятности $P(\eta|\xi)$. Приняв слово η, получатель информации еще не может сказать точно, что было передано слово ξ_k, а не ξ_l ($l \neq k$), если для обоих слов вероятности $P(\eta|\xi_k)$, $P(\eta|\xi_l)$ не равны нулю. Он может судить лишь об апостериорных вероятностях того или иного кодового слова. Если априорные вероятности $P(k)$ всех M сообщений, а значит и всех кодовых слов ξ_k считать одинаковыми (равными $1/M$), то по формуле обратной вероятности получим для кодового слова ξ_k апостериорную вероятность

$$P[\xi_k \mid \eta] = \frac{\frac{1}{M} P(\eta \mid \xi_k)}{\sum_l \frac{1}{M} P(\eta \mid \xi_l)} = \frac{P(\eta \mid \xi_k)}{\sum_l P(\eta \mid \xi_l)}. \tag{7.1.4}$$

Если при фиксированном η получатель информации выберет кодовое слово ξ_k, то с вероятностью $P[\xi_k|\eta]$ он получит правильное сообщение, а с вероятностью $1 - P[\xi_k|\eta]$ ошибется. Чтобы минимизировать вероятность ошибки, он, очевидно, должен выбрать то слово ξ_k, которому соответствует максимальная (из M возможных) вероятность (7.1.4) или, что то же самое, максимальная *функция правдоподобия* $P(\eta|\xi_k)$. Средняя вероятность $P_{\text{ош}}$ ошибки будет при этом получаться усреднением вероятности ошибки $1 - P[\xi_k|\eta]$ по η:

$$P_{\text{ош}} = 1 - \mathbf{M} \max_k P[\xi_k|\eta] = 1 - \sum_\eta \max_k \left[\frac{P(\eta|\xi_k)}{\sum_l P(\eta|\xi_l)}\right] P(\eta)$$

(вследствие (7.1.4)) или

$$P_{\text{ош}} = 1 - \frac{1}{M} \sum_\eta \max_k P(\eta|\xi_k), \tag{7.1.5}$$

поскольку

$$P(\eta) = \sum_l P(\eta|\xi_l) \frac{1}{M}.$$

Выбор максимальной функции правдоподобия из функций

$$P(\eta|\xi_1), \ldots, P(\eta|\xi_M), \tag{7.1.6}$$

очевидно, эквивалентен выбору максимальной случайной информации $I(\xi_k, \eta) = \ln[P(\eta|\xi_k)/P(\eta)]$ из множества

$$I(\xi_1, \eta), \ldots, I(\xi_M, \eta)$$

или выбору минимального «расстояния» из множества

$$D_0(\xi_1, \eta), \ldots, D_0(\xi_M, \eta), \tag{7.1.7}$$

если «расстояние» $D_0(\xi, \eta)$ между точками ξ и η определено формулой

$$D_0(\xi, \eta) = -I(\xi, \eta) = \ln \frac{P(\eta)}{P(\eta|\xi)}. \tag{7.1.8}$$

Согласно описанному правилу декодирования получатель информации решает, что было передано то сообщение, которое связано с кодовым словом, «ближайшим» к полученному слову η.

Совокупность выходных слов η, которые «ближе» к кодовому слову ξ_k, чем к какому-либо другому кодовому слову, обозначим G_k. Тогда при фиксированном коде множество значений разобьется на области декодирования $G_1, \ldots, G_M$, и решение о том, какое сообщение было передано, будет производиться по принадлежности принятого слова η той или иной области декодирования.

Схема кодирования и декодирования при некотором выбранном коде представлена на рис. 7.1. Как видно, используются не все входные слова ξ, а лишь кодовые слова $\xi_1, \ldots, \xi_M$.

Передаваемое сообщение, по существу, совпадает с номером кодового слова, а принятое сообщение — с номером той области декодирования, которой принадлежит принятое слово.

Предположив, что код фиксирован, найдем условную и среднюю вероятность ошибки. При фиксированном сообщении k вероят-

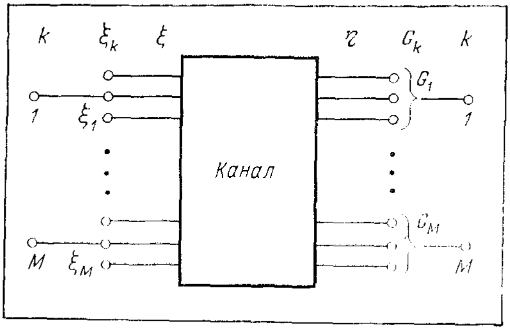


Рис. 7.1.
Схема кодирования и декодирования для канала с помехами.

ность ошибки совпадает с вероятностью того, что точка η, имеющая вероятности $P(\eta|\xi_k)$, выпадает вне области G_k, т. е.

$$P_{\text{ош}}(k) = 1 - \sum_{\eta \in G_k} P(\eta \mid \xi_k). \tag{7.1.9}$$

Поскольку $P(\eta|\xi_k) = \max P(\eta|\xi_l)$ внутри области $G_k \ni \eta$ (по определению этой области), то

$$P_{\text{ош}}(k) = 1 - \sum_{\eta \in G_k} \max_l P(\eta \mid \xi_l). \tag{7.1.10}$$

Если теперь произвести усреднение

$$P_{\text{ош}} = \frac{1}{M} \sum_{k=1}^{M} P_{\text{ош}}(k) \tag{7.1.11}$$

по всем сообщениям, каждое из которых имеет вероятность $1/M$, то сумма в (7.1.10) распространится на все пространство значений η, и мы получим формулу (7.1.5).

7.2. Случайный код и средняя вероятность ошибки

При фиксированном правиле декодирования, например, при описанном в предыдущем параграфе оптимальном правиле, вероятность ошибки (т. е. вероятность выбора получателем не того кодового слова ξ_k, которое было передано) зависит от выбранного кода. Чтобы реже были ошибки декодирования, обусловленные помехами, кодовые слова желательно выбирать по возможности более «непохожими», лежащими в некотором смысле как можно «дальше» друг от друга. Ввиду того, что одновременно увеличивать «расстояние» между кодовыми точками $\xi_1, \ldots, \xi_M$ нельзя без уменьшения их числа M, то желательно располагать кодовые точки «как можно равномерней» в пространстве X^n значений ξ. Желаемая «равномерность» достигается по законам массовых явлений при больших M (и n), если выбирать кодовые точки независимо друг от друга случайным образом.

Случайный код Шеннона конструируется следующим образом. Кодовая точка ξ_1 получается в результате выбрасывания случайной величины ξ с вероятностями $P(\xi)$. Вторая точка (а также третья и прочие) выбрасывается независимо от других подобным же образом и является, следовательно, независимой случайной величиной с вероятностями $P(\xi_2)$. Все кодовые точки $\xi_1, \ldots, \xi_M$ в совокупности описываются распределением вероятностей $P(\xi_1) \cdot \ldots \cdot P(\xi_M)$.

Для каждого фиксированного кода $(\xi_1, \ldots, \xi_M)$, полученного описанным способом, и фиксированного сообщения k имеется некоторая вероятность ошибки декодирования. Обозначим эту вероятность $P_{\text{ош}}(|k, \xi_1, \ldots, \xi_M)$. Согласно (7.1.9) она равна

$$P_{\text{ош}}(|k, \xi_1, \ldots, \xi_M) = 1 - \sum_{\eta \in G_k} P(\eta \mid \xi_k). \tag{7.2.1}$$

Согласно определению области G_k, данному в § 7.1, суммирование в (7.2.1) должно проводиться по той области, где все

$$P(\eta \mid \xi_l) < P(\eta \mid \xi_k), \quad l = 1, \ldots, k-1, \ k+1, \ldots, M.$$

поэтому

$$P_{\text{ош}}(|k, \xi_1, \ldots, \xi_k) = 1 - \sum_{\text{все } P(\eta|\xi_l) < P(\eta \mid \xi_k)} P(\eta \mid \xi_k). \tag{7.2.2}$$

Вместо того чтобы вычислить вероятности ошибок

$$P_{\text{ош}}(|k, \xi_1, \ldots, \xi_M) \quad \text{и}$$

$$P_{\text{ош}}(|\xi_1, \ldots, \xi_M) = \frac{1}{M} \sum_{k=1}^{M} (|k, \xi_1, \ldots, \xi_M),$$

значительно удобнее вычислить вероятность

$$P_{\text{ош}}(|k) = \sum_{\xi_1, \ldots, \xi_M} P_{\text{ош}}(|k, \xi_1, \ldots, \xi_M) P(\xi_1) \ldots P(\xi_M), \tag{7.2.3}$$

усредненную по различным случайным кодам. В этом удобстве и состоит преимущество случайных кодов. Вероятность (7.2.3), как это очевидно из соображений симметрии, не зависит от номера передаваемого сообщения, и дополнительного усреднения по k, стоящего в (7.1.11), в данном случае не требуется:

$$P_{\text{ош}} = P_{\text{ош}}(|k). \tag{7.2.4}$$

Вычисления показывают, что средняя вероятность (7.2.3) имеет удовлетворительные асимптотические свойства (при $n \to \infty$).

Отсюда вытекает, что среди реализаций случайных кодов имеются коды со свойствами, по меньшей мере, такими же удовлетворительными.

Перейдем к вычислению вероятности (7.2.3), (7.2.4). Подставляя (7.2.2) в (7.2.3) и учитывая (7.2.4), находим

$$P_{\text{ош}} = 1 - \sum_{\xi_k,\ \eta} P(\xi_k, \eta) \sum_{\text{все}\ P(\eta \mid \xi_l) < P(\eta \mid \xi_k)} \prod_{l \neq k} P(\xi_l) =$$

$$= \sum_{\xi_k,\ \eta} P(\xi_k, \eta) \left\{ 1 - \left[\sum_{P(\eta \mid \xi) \quad P(\eta \mid \xi_k)} P(\xi) \right]^{M-1} \right\}. \qquad (7.2.5)$$

Иначе

$$P_{\text{ош}} = \sum_{\xi_k,\ \eta} P(\xi_k, \eta) f \left(\sum_{P(\eta \mid \xi_k) < P(\eta \mid \xi)} P(\xi) \right), \qquad (7.2.6)$$

если ввести возрастающую функцию $f(x) = 1 - (1 - x)^{M-1}$ и учесть, что

$$\sum_{P(\eta \mid \xi) < P(\eta \mid \xi_k)} P(\xi) = 1 - \sum_{P(\eta \mid \xi) > P(\eta \mid \xi_k)} P(\xi). \qquad (7.2.7)$$

В формулах (7.2.2.), (7.2.5), (7.2.6) знак $<$ не исключает знака $\leqslant$. Дело в том, что могут иметься «спорные» точки η, равноудаленные от нескольких конкурирующих кодовых точек, скажем $P(\eta|\xi_l) = P(\eta|\xi_k)$ $(l \neq k)$. Такие точки с равным основанием можно отнести и к области E_k и к области E_l. Если учесть неоднозначность, связанную с такими точками, то, как легко понять, вместо (7.2.6) будем иметь неравенство

$$\sum_{\xi_k,\ \eta} P(\xi_k, \eta) f \left(\sum_{P(\eta \mid \xi) \quad P(\eta \mid \xi_k)} P(\xi) \right) \leqslant P_{\text{ош}} \leqslant$$

$$\leqslant \sum_{\xi_k,\ \eta} P(\xi_k, \eta) f \left(\sum_{P(\eta \mid \xi) \geqslant P(\eta \mid \xi_k)} P(\xi) \right), \qquad (7.2.8)$$

где под знаком суммы слева стоит строгое неравенство. В выражении слева исключены вклады всех «спорных» точек, в выражении справа они учтены несколько раз.

Рассмотрим выражение $\sum_{P(\eta \mid \xi) \geqslant P(\eta \mid \xi_k)} P(\xi)$, являющееся аргументом функции f. Неравенство $P(\eta|\xi) \geqslant P(\eta|\xi_k)$ или $P(\eta|\xi)/P(\eta) \geqslant P(\eta|\xi_k)/P(\eta)$ эквивалентно неравенствам

$$e^{I(\xi,\ \eta)} \geqslant e^{I(\xi_k,\ \eta)}; \qquad \frac{P(\xi \mid \eta)}{P(\xi)} \geqslant e^{I(\xi_k,\ \eta)}, \qquad (7.2.9)$$

поскольку в силу (6.2.5)

$$\frac{P(\eta \mid \xi)}{P(\eta)} = \frac{P(\xi \mid \eta)}{P(\xi)} = e^{I(\xi,\ \eta)}.$$

Из неравенства $P(\xi) \leqslant P(\xi|\eta)\ e^{-I(\xi_k, \eta)}$ (т. е. второго неравенства (7.2.9)) суммированием получаем

$$\sum_{P(\eta \mid \xi) \geqslant P(\eta \mid \xi_k)} P(\xi) \leqslant e^{-I(\xi_k, \eta)} \sum_{I(\xi, \eta) \geqslant I(\xi_k, \eta)} P(\xi \mid \eta).$$

Данное неравенство может только усилиться, если сумму в правой части заменить на единицу:

$$\sum_{P(\eta \mid \xi) \geqslant P(\eta \mid \xi_k)} P(\xi) \leqslant e^{-I(\xi_k, \eta)}.$$

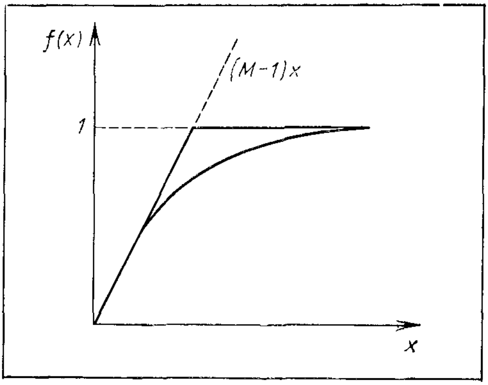


Рис. 7.2.
Функция $f(x) = 1 - (1-x)^{M-1}$ и мажорирующая ломаная.

Используя это неравенство при рассмотрении правого неравенства (7.2.8) и учитывая возрастающий характер функции f, находим

$$P_{\text{ош}} \leqslant \sum_{\xi_k, \eta} P(\xi_k, \eta)\, f\left(e^{-I(\xi_k, \eta)}\right) = \mathbf{M} f\left(e^{-I(\xi_k, \eta)}\right) =$$
$$= \int f\left(e^{-I}\right) dF(I). \tag{7.2.10}$$

Здесь через $F(I) = \mathbf{P}\{I(\xi, \eta) < I\}$ обозначена функция распределения случайной информации связи $I(\xi, \eta)$ величин ξ и η, имеющих распределение $P(\xi, \eta)$. Полученная оценка выражается, таким образом, только через эту функцию распределения.

Поведение функции $f(x)$ показано на рис. 7.2. Эта функция на интервале $0 \leqslant x \leqslant 1$ возрастает от 0 до 1. Она имеет производные

$$f'(x) = (M-1)(1-x)^{M-2},$$
$$f''(x) = -(M-1)(M-2)(1-x)^{M-3} \leqslant 0,$$

и следовательно, является выпуклой. Очевидно, ее можно мажорировать ломаной

$$f(x) \leqslant \min[(M-1)x, 1] \leqslant \min[Mx, 1]. \tag{7.2.11}$$

После этого формула (7.2.10) примет вид

$$P_{\text{ош}} \leqslant \int \min [M e^{-I}, 1]\, dF(I) \leqslant M \int\limits_{I \geqslant \ln M} e^{-I}\, dF(I) + \\ + \int\limits_{I < \ln M} dF(I), \tag{7.2.12}$$

т. е.

$$P_{\text{ош}} \leqslant M \int\limits_{I=\ln M}^{\infty} e^{-I}\, dF(I) + F(\ln M). \tag{7.2.13}$$

Точку излома $I = \ln M$ в (7.2.12) можно заменить на любую другую точку $I = \lambda$; от этого неравенство только усилится. Полученные неравенства будут использованы в дальнейшем.

7.3. Асимптотическая безошибочность декодирования. Теорема Шеннона (вторая асимптотическая теорема)

Важным и далеко не тривиальным фактом является то, что средняя ошибка декодирования может быть сделана подходящим выбором кода сколь угодно малой при увеличении числа символов n без уменьшения уровня помех в канале и без уменьшения количества передаваемой информации, рассчитанного на один символ. Этот результат получен Шенноном в 1948 г. [1] и (будучи формулируем в терминах пропускной способности) обычно носит название теоремы Шеннона.

Теорема 7. 1. *Пусть* (как в § 7.1) *имеется канал* $[P(y \mid x), P(x)]$ *и канал* $[P(\eta \mid \xi), P(\xi)]$ [(см. (7.1.1), (7.1.2)], *являющийся* n-*й степенью первого. Пусть далее при* $n \to \infty$ *количество* $\ln M$ *передаваемой информации растет по закону*

$$\ln M = \ln [e^{nR}] \leqslant nR \tag{7.3.1}$$

(*скобки обозначают целую часть*), *где* R — *не зависящая от* n *величина, удовлетворяющая неравенству*

$$R < I_{xy} < \infty. \tag{7.3.2}$$

Тогда существует последовательность кодов $K^{(n)}$, *таких, что*

$$P_{\text{ош}}(K^{(n)}) \to 0 \quad \text{при} \quad n \to \infty. \tag{7.3.3}$$

Доказательство. Вследствие независимости и одинаковой распределенности различных символов x_i и y_i, составляющих слова ξ и η (7.1.1), имеем, что случайная информация $I(\xi, \eta)$ равна сумме

$$I(\xi, \eta) = \sum_{i=1}^{n} I(x_i, y_i) \tag{7.3.4}$$

одинаково распределенных независимых случайных величин $I(x_i, y_i)$. Каждая из них имеет конечное математическое ожидание

$$I_{xy} = \sum_{x,\,y} I(x, y) P(x, y) = I_{\xi\eta}/n.$$

Применяя к (7.3.4) закон больших чисел (теорему Хинчина, см., например Гнеденко [1]), получаем, что

$$\mathbf{P}\left\{\left|\frac{1}{n} I(\xi, \eta) - I_{xy}\right| < \varepsilon\right\} \to 1$$

и, следовательно,

$$\mathbf{P}\{|I(\xi, \eta) - I_{\xi\eta}| \geqslant n\varepsilon\} \to 0 \qquad \text{при } n \to \infty, \tag{7.3.5}$$

каково бы ни было $\varepsilon > 0$.

Рассмотрим среднюю вероятность ошибки по ансамблю случайных кодов, описанных в § 7.2. Положим $\varepsilon = (I_{xy} - R)/2 = (I_{\xi\eta} - nR)/2n$, так что $n(R + \varepsilon) = I_{\xi\eta} - n\varepsilon$. Очевидно, что $\varepsilon > 0$ в силу (7.3.2). Поскольку

$$\min[M\,\mathrm{e}^{-I}, 1] \leqslant \begin{cases} M\,\mathrm{e}^{-I} & \text{при} \quad I > I_{\xi\eta} - n\varepsilon, \\ 1 & \text{при} \quad I \leqslant I_{\xi\eta} - n\varepsilon, \end{cases}$$

то из формулы (7.2.12) имеем

$$P_{\text{ош}} \leqslant M \int\limits_{I > nR + \varepsilon} \mathrm{e}^{-I}\, dF(I) + \mathbf{P}\{I(\xi, \eta) \leqslant I_{\xi\eta} - n\varepsilon\}. \tag{7.3.6}$$

Рассмотрим первый член, стоящий в правой части. Из неравенства $I > nR + n\varepsilon$ следует, что

$$\mathrm{e}^{-I} < \mathrm{e}^{-nR - n\varepsilon}, \quad \int\limits_{I > nR + n\varepsilon} \mathrm{e}^{-I}\, dF(I) < \mathrm{e}^{-nR - n\varepsilon} \int\limits_{I > nR + n\varepsilon} dF(I) \leqslant$$
$$\leqslant \mathrm{e}^{-nR - n\varepsilon}.$$

Поэтому

$$M \int\limits_{I > n(R + \varepsilon)} \mathrm{e}^{-I}\, dF(I) \leqslant M\mathrm{e}^{-nR - n\varepsilon}$$

и в силу (7.3.1)

$$M \int\limits_{I > nR + n\varepsilon} \mathrm{e}^{-I}\, dF(I) \leqslant \mathrm{e}^{nR}\,\mathrm{e}^{-nR - n\varepsilon} = \mathrm{e}^{-n\varepsilon}.$$

Это выражение стремится к нулю при $n \to \infty$. Второй член в правой части (7.3.6) стремится к нулю в силу (7.3.5). Следовательно,

$$P_{\text{ош}} \to 0 \qquad \text{при} \quad n \to \infty. \tag{7.3.7}$$

Среди совокупности случайных кодов заведомо есть код $(\xi_1^n, \ldots$ $\ldots, \xi_M^n)$, для которого вероятность ошибки не превосходит среднюю вероятность

$$P_{\text{ош}}(\xi_1^n, \ldots, \xi_M^n) \leqslant P_{\text{ош}}. \tag{7.3.8}$$

Из (7.3.7), (7.3.8) вытекает, что для последовательности таких кодов

$$P_{\text{ош}}(\xi_1^n, \ldots, \xi_M^n) \to 0 \quad \text{при} \quad n \to \infty.$$

Этим заканчивается доказательство теоремы.

Доказанная асимптотическая безошибочность передачи информации имеет место не только для каналов, являющихся степенями какого-то канала. Приведенное доказательство легко обобщается и на другие случаи. Существенным здесь является вводимое ниже условие информационной устойчивости.

Определение. Последовательность случайных величин ξ^n, η^n, $n = 1, 2, \ldots$ (n — индекс) называется *информационно устойчивой*, если

А. $\infty > I_{\xi^n \eta^n} \to \infty$ при $n \to \infty$;

Б. Отношение $I(\xi^n, \eta^n)/I_{\xi^n \eta^n}$ сходится при $n \to \infty$ по вероятности к 1.

Соответствующее обобщение теоремы 7.1 можно сформулировать так.

Теорема 7.2. (Общий вид второй асимптотической теоремы). *Пусть имеется последовательность информационно устойчивых случайных величин ξ^n, η^n, $n = 1, 2, \ldots$. Пусть далее количество передаваемой информации растет при $n \to \infty$ по закону*

$$\ln M = I_{\xi^n \eta^n}(1 - \mu) \left(\textit{точнее } M = [e^{I_{\xi^n \eta^n}(1-\mu)}]\right), \tag{7.3.9}$$

где $\mu > 0$ не зависит от n. Тогда существует последовательность кодов таких, что $P_{\text{ош}} \to 0$ при $n \to \infty$.

Доказательство. Непосредственно из определения информационной устойчивости (свойство Б) имеем

$$\begin{gathered} \mathbf{P}\{|I(\xi^n, \eta^n)/I_{\xi^n \eta^n} - 1| < \varepsilon\} \to 1, \\ \mathbf{P}\{|I(\xi^n, \eta^n) - I_{\xi^n \eta^n}| \geqslant \varepsilon I_{\xi^n \eta^n}\} \to 0, \end{gathered} \tag{7.3.10}$$

при любом $\varepsilon > 0$. Положим

$$\varepsilon = \mu/2. \tag{7.3.11}$$

Вследствие неравенства

$$\min[M e^{-I}, 1] \leqslant \begin{cases} M e^{-I} & \text{при } I > (1 - \varepsilon) I_{\xi^n \eta^n}, \\ 1 & \text{при } I \leqslant (1 - \varepsilon) I_{\xi^n \eta^n} \end{cases}$$

из формулы (7.2.12) находим

$$P_{\text{ош}} \leqslant M \int\limits_{I > (1-\varepsilon) I_{\xi^n \eta^n}} e^{-I} dF(I) + \mathbf{P}\{I(\xi, \eta) \leqslant I_{\xi^n \eta^n} - \varepsilon I_{\xi^n \eta^n}\}.$$

Второй член в правой части стремится к нулю при $n \to \infty$ в силу (7.3.10), а первый член, используя (7.3.9), можно оценить так:

$$M \int\limits_{I > (1-\varepsilon) I_{\xi^n \eta^n}} e^{-I}\, dF(I) \leqslant e^{I_{\xi^n \eta^n}(1-\mu)}\, e^{-(1-\varepsilon) I_{\xi^n \eta^n}} \times$$

$$\times \int\limits_{I > (1-\varepsilon) I_{\xi^n \eta^n}} dF(I) \leqslant e^{(\varepsilon-\mu) I_{\xi^n \eta^n}}.$$

Следовательно, он стремится к нулю в силу (7.3.11) и в силу свойства А определения информационной устойчивости. Доказательство теоремы завершают такие же рассуждения, что и в предыдущей теореме.

7.4. Асимптотическая формула для вероятности ошибки

В дополнение к результатам предыдущего параграфа можно получить более сильные результаты, касающиеся быстроты исчезновения вероятности ошибки. Оказывается, что вероятность ошибки для удовлетворительных кодов убывает с ростом n в основном экспоненциально:

$$P_{\text{ош}} \leqslant e^{a - \alpha n}, \tag{7.4.1}$$

где a — слабо зависящая от n величина, а α — постоянная, представляющая основной интерес. Для нее удается получить довольно общие формулы.

1. Теорема 7.3. *В условиях теоремы* 7.1 *имеет место неравенство*

$$P_{\text{ош}} \leqslant 2e^{-[s\mu'(s) - \mu(s)]n}, \tag{7.4.2}$$

где

$$\mu(t) = \ln \sum_{x, y} P^{1-t}(x, y)\, P^t(x)\, P^t(y) \tag{7.4.3}$$

(см. (6.4.10), причем аргумент s заменен на t) *а* s — *положительный корень уравнения*

$$\mu'(s) = -R. \tag{7.4.4}$$

Предполагается, что R достаточно близко к I_{xy} для того, чтобы последнее уравнение имело решение, и что значение s принадлежит отрезку дифференцируемости потенциала $\mu(t)$.

Доказательство. Введем скошенную функцию распределения

$$\widetilde{F}(\lambda) = \int_{-\infty}^{\lambda} e^{-I}\, dF(I) \Big/ \int_{-\infty}^{\infty} e^{-I}\, dF(I). \tag{7.4.5}$$

Формулу (7.4.5) можно записать в таком виде

$$\int_{\lambda}^{\infty} e^{-I}\, dF(I) = [1-\widetilde{F}(\lambda)] \int_{-\infty}^{\infty} e^{-I}\, dF(I).$$

Подставляя это равенство в неравенство типа (7.2.13) (но при выборе точки излома $\lambda = nR$) будем иметь

$$P_{\text{ош}} \leqslant M\,[1-\widetilde{F}(nR)] \int_{\infty}^{\infty} e^{-I}\, dF(I) + F(nR)$$

или

$$P_{\text{ош}} \leqslant e^{nR}\,[1-\widetilde{F}(nR)] \int_{-\infty}^{\infty} e^{-I}\, dF(I) + F(nR) \tag{7.4.6}$$

вследствие (7.3.1).

Чтобы оценить полученное выражение, надо уметь оценивать «хвост» функций распределения $F(nR)$, $\widetilde{F}(nR)$, соответствующих сумме независимых случайных величин. Это составляет содержание теорем 4.6, 4.17. Применяя теорему 4.7 к функции $F(nR)$ (при $B(\xi) = -I(\xi, \eta)$), имеем

$$F(nR) = \mathbf{P}[-I > nR] \leqslant e^{-[s\mu'(s)-\mu(s)]\,n}, \tag{7.4.7}$$

где

$$e^{n\mu(t)} = \int e^{-tI} F(dI), \tag{7.4.8}$$

т. е.

$$e^{n\mu(t)} = \mathbf{M}e^{-tI(\xi,\eta)} = \left[\sum_{x,y} e^{-tI(x,y)} P(x,y)\right]^n, \tag{7.4.9}$$

а s — корень уравнения

$$\mu'(s) = -R. \tag{7.4.10}$$

Корень s является положительным, поскольку $-R > -I_{xy} = \mu'(0)$, а $\mu'(t)$ — возрастающая функция: $\mu''(t) > 0$.

Аналогично для скошенного распределения $1-\widetilde{F}(-x) = \widetilde{\mathbf{P}}[-I < x]$ по теореме 4.6 имеем

$$1-\widetilde{F}(nR) \leqslant e^{-[\widetilde{s}\widetilde{\mu}'(\widetilde{s})-\widetilde{\mu}(\widetilde{s})]n} \tag{7.4.11}$$

где

$$e^{n\widetilde{\mu}(\widetilde{s})} = \int e^{-\widetilde{s}I} \widetilde{F}(dI), \tag{7.4.12}$$

причем

$$\widetilde{\mu}'(\widetilde{s}) = -R, \tag{7.4.13}$$

$\widetilde{s} < 0$, как как $-R < \mu'(0)$ $(\mu'(0) = \mu'(1) > 0)$.

Подставляя (7.4.5) в (7.4.12), находим

$$e^{n\widetilde{\mu}(\widetilde{s})} = \int_{-\infty}^{\infty} e^{-\widetilde{s}I - I}\, dF(I) \Big/ \int_{-\infty}^{\infty} e^{-I}\, dF(I) = e^{n\mu(\widetilde{s}+1) - n\mu(1)}. \tag{7.4.14}$$

Последнее справедливо в силу соотношения (7.4.8), из которого вытекает

$$\int_{-\infty}^{\infty} e^{-I}\, dF(I) = e^{n\mu(1)}. \tag{7.4.15}$$

Следовательно,

$$\widetilde{\mu}(\widetilde{s}) = \mu(\widetilde{s}+1) - \mu(1). \tag{7.4.16}$$

Поэтому (7.4.13) принимает вид

$$\mu'(\widetilde{s}+1) = -R. \tag{7.4.17}$$

Сравнение этого равенства с (7.4.10) дает:

$$\widetilde{s} + 1 = s,$$

и формула (7.4.11) записывается в виде

$$1 - \widetilde{F}(nR) \leqslant e^{-[(s-1)\mu'(s) - \mu(s) + \mu(1)]n}, \tag{7.4.18}$$

где учтено (7.4.16).

Подставляя (7.4.7), (7.4.18), (7.4.15) в (7.4.6) и учитывая (7.4.10), получаем требуемую формулу (7.4.2). Равенство (7.4.3) следует из (7.4.9). Теорема доказана.

2. Согласно приведенной теореме потенциал $\mu(s)$ определяет стоящий в экспоненте (7.4.1) коэффициент

$$\alpha = s\mu'(s) - \mu(s) \tag{7.4.19}$$

как преобразование Лежандра от характеристического потенциала:

$$\alpha(R) = -sR - \mu(s) \qquad (d\mu/ds = -R). \tag{7.4.20}$$

Согласно общим закономерностям преобразований Лежандра из вогнутости функции $\mu(s)$ вытекает вогнутость функции $\alpha(R)$. Это можно доказать аналогично тому, как было доказано неравенство (4.1.16а). Используя свойство вогнутости, можно экстраполировать функцию $\alpha(R)$ касательными. Если функция $\alpha(R)$ вычислена, скажем, на интервале $[R_1, I_{xy}]$, то можно произвести экстраполяцию

$$\alpha_{\text{экс}}(R)=\begin{cases}\alpha(R) & \text{при } R_1\leqslant R\leqslant I_{xy}\\ \alpha(R_1)+\dfrac{d\alpha(R_1)}{\alpha R}(R-R_1), & R\leqslant R_1\end{cases}$$

и заменить формулу

$$P_{\text{ош}}\leqslant 2e^{-n\alpha(R)} \qquad (7.4.21)$$

менее сильной формулой

$$P_{\text{ош}}\leqslant 2e^{-n\alpha_{\text{экс}}(R)}.$$

Типичный ход зависимости $\alpha(R)$ и указанная экстраполяция касательной показаны на рис. 7.3.

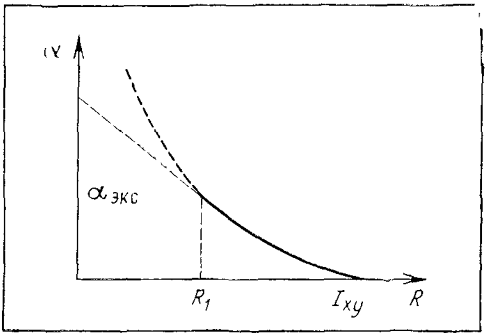


Рис. 7.3. Типичное поведение коэффициента $\alpha(R)=\lim\limits_{n\to\infty}\dfrac{-\ln P_{\text{ош}}}{n}$ в экспоненте формулы (7.4.1).

Точка $R=I_{xy}$ соответствует значению $s=0$, поскольку уравнение (7.4.4) при этом имеет вид $\mu'(s)=\mu'(0)$. Далее $\mu(0)=0$, как это следует из определения функции $\mu(s)$; поэтому значению $R=I_{xy}$ соответствует нулевое значение коэффициента $\alpha=0$.

Исследуем поведение зависимости $\alpha(R)$ вблизи этой точки. Дифференцируя (7.4.4) и (7.4.19), получаем

$$-\mu''(s)\,ds=dR, \qquad d\alpha=s\mu''(s)\,ds.$$

Следовательно, $d\alpha/dR=-s$.

Взяв дифференциал от этой производной и поделив на $dR=-\mu''(s)ds$, найдем

$$d^2\alpha/dR^2=1/\mu''(s).$$

В частности, в точке $R=I_{xy}$, $s=0$ имеем

$$\frac{d^2\alpha}{dR^2}(I_{xy})=\frac{1}{\mu''(0)}. \qquad (7.4.22)$$

Здесь $\mu''(0)$, как легко видеть из определения (7.4.9) функции $\mu(s)$, совпадает с дисперсией случайной информации $I(x, y)$

$$\mu''(0)=\mathbf{M}[I(x,y)-I_{xy}]^2=\sum_{x,y}[I(x,y)]^2P(x,y)-I_{xy}^2.$$

Вычисляя аналогичным образом третью производную, имеем

$$\frac{d^3\alpha}{dR^3} = \frac{\mu'''(s)}{[\mu''(s)]^3}. \tag{7.4.23}$$

Представляя функцию $\alpha(R)$ в виде ряда Тейлора и опуская члены с более высокими производными, согласно (7.4.22), (7.4.23) будем иметь

$$\alpha(R) = \frac{(I_{xy} - R)^2}{2\mu''(0)} - \frac{\mu'''(0)}{6[\mu''(0)]^3}(I_{xy} - R)^3 + \ldots, \qquad R < I_{xy}. \tag{7.4.24}$$

Приведенные выше результаты, связанные с формулой (7.4.1), могут быть дополнены и улучшены в различных отношениях. Некоторые более сильные результаты будут указаны в дальнейшем (см. § 7.5).

Эти результаты могут быть распространены на более общий [по сравнению с (7.1.1)] случай произвольных информационно-устойчивых случайных величин, т. е. может быть произведено усиление теоремы 7.2 в направлении учета быстроты исчезновения вероятности ошибки. Мы не будем на этом останавливаться, а ограничимся указанием, что это делается совершенно стандартным способом. Коэффициент α не следует рассматривать в отдельности от n. Вместо этого нужно оперировать с комбинацией $\alpha_n = n\alpha$. Аналогично следует рассматривать лишь комбинации

$$R_n = nR, \qquad \mu_n(t) = n\mu(t), \qquad I_{\xi^n\eta^n} = nI_{xy}.$$

Формулы (7.4.2)—(7.4.4) при этом заменятся формулами

$$P_{\text{ош}} \leqslant 2e^{-s\mu_n'(s) + \mu_n(s)}, \quad \mu_n' = -R_n, \tag{7.4.24a}$$

$$\mu_n(t) = \ln \int P^{1-t}(d\xi^n, d\eta^n)\, P^t(d\xi^n)\, P^t(d\eta^n).$$

В приведенном выше изложении изменится лишь способ записи формул.

Формулы, оценивающие поведение вероятности ошибки, выводятся в работах Шеннона [3] и Фано [1].

7.5. Усиленные оценки для оптимального декодирования

1. В предыдущем изложении декодирование производилось по принципу максимальной функции правдоподобия (7.1.6) или, что то же самое, минимального расстояния (7.1.8). Представляет интерес исследовать, чему будет равна оценка вероятности ошибки, если декодирование производить на основе «расстояния» $D(\xi, \eta)$, определенного как-нибудь по-другому. В настоящем параграфе мы будем считать «расстояние» $D(\xi, \eta)$ некоторой произвольно задан-

ной функцией. Переход к новому «расстоянию», конечно, не может уменьшить вероятности ошибки декодирования, но, в принципе, может уменьшить верхнюю границу оценивания указанной вероятности.

Теорема 7.4. *Пусть (как в теореме 7.1) имеется канал* $[P(\eta \mid \xi), P(\xi)]$, *являющийся n-й степенью канала* $[P \mid (y|x), P(x)]$. *Пусть декодирование проводится по принципу минимального расстояния*

$$D(\xi, \eta) = \sum_{j=1}^{n} d(x_j, y_j), \tag{7.5.1}$$

где $d(x, y)$ — *заданная функция.*

Количество передаваемой информации $\ln M$ *растет с увеличением n по закону*

$$\ln M = \ln [e^{nR}] \leqslant nR \tag{7.5.2}$$

($R < I_{xy}$ *не зависит от* n). *Тогда существует последовательность кодов, имеющих вероятность ошибки декодирования*

$$P_{\text{ош}} \leqslant 2e^{-n[s_0\gamma'(s_0) - \gamma(s_0)]}. \tag{7.5.3}$$

Здесь s_0 — *один из корней* s_0, t_0, r_0 *системы уравнений:*

$$\begin{gathered} \gamma'(s_0) = \varphi'_r(r_0, t_0), \qquad \varphi'_t(r_0, t_0) = 0, \\ (s_0 - r_0)\gamma'(s_0) - \gamma(s_0) + \varphi(t_0, t_0) + R = 0, \end{gathered} \tag{7.5.4}$$

а $\gamma(s)$, $\varphi(r, t)$ — *следующие функции:*

$$\gamma(s) = \ln \sum_{x, y} e^{sd(x, y)} P(x) P(y \mid x), \tag{7.5.5}$$

$$\varphi(r, t) = \ln \sum_{x, y, x'} e^{(r-t)d(x, y) + td(x', y)} P(x) P(y \mid x) P(x'), \tag{7.5.6}$$

Кроме того $\varphi'_r = \dfrac{\partial\varphi}{\partial r}$, $\varphi'_t = \dfrac{\partial\varphi}{\partial t}$.

Предполагается, что уравнения (7.5.4) *имеют корни, лежащие в области определения и дифференцируемости функций* (7.5.5), (7.5.6), *причем* $s_0 > 0$, $r_0 < 0$, $t_0 < 0$.

Доказательство. Как и раньше, будем рассматривать случайные коды и усреднять ошибку декодирования по ним.

Запишем для средней ошибки равенства, аналогичные (7.2.1)—(7.2.4), но с произвольно заданным расстоянием $D(\xi, \eta)$. Будем теперь производить усреднение по η в последнюю очередь:

$$P_{\text{ош}} = P_{\text{ош}}(\mid k) = \sum_{\eta} P_{\text{ош}}(\mid k, \eta) P(\eta), \tag{7.5.7}$$

где

$$\begin{aligned} P_{\text{ош}}(\mid k, \eta) = \sum_{\xi_1 \ldots \xi_m} \vartheta(k, \xi_1, \ldots, \xi_M, \eta) P(\xi_1) \ldots P(\xi_{k-1}) \times \\ \times P(\xi_k \mid \eta) P(\xi_{k+1}) \ldots P(\xi_M) \end{aligned} \tag{7.5.8}$$

и

$$\vartheta(k, \xi_1, \ldots, \xi_M, \eta) = \begin{cases} 0, & \text{если все } D(\xi_l, \eta) > D(\xi_k, \eta), \ l \neq k, \\ 1, & \text{если хотя бы одно } D(\xi_l, \eta) \leqslant D(\xi_k, \eta), \ l \neq k. \end{cases} \tag{7.5.9}$$

Подставляя (7.5.9) в (7.5.8) по аналогии с (7.2.5)—(7.2.8), будем иметь

$$P_{\text{ош}}(\,|\,k, \eta) = P_{\text{ош}}(\,|\,\eta) = 1 - \sum_k P(\xi_k\,|\,\eta) \prod_{l \neq k} \sum_{D(\xi_l, \eta) > D(\xi_k, \eta)} P(\xi_l) \times \\ \times \sum_{\xi_k} P(\xi_k\,|\,\eta) = f\left(\sum_{D(\xi_l, \eta) \leqslant D(\xi_k, \eta)} P(\xi_l)\right). \tag{7.5.10}$$

Здесь неравенство $D(\xi_l, \eta) < D(\xi_k, \eta)$ нестрогое.

Включив все те случаи, когда $D(\xi_l, \eta) = D(\xi_k, \eta)$, обозначим

$$F_\eta[\lambda] = \sum_{D(\xi_l, \eta) \leqslant \lambda} P(\xi_l). \tag{7.5.11}$$

Тогда из (7.5.10), используя (7.2.11), получаем

$$P_{\text{ош}}(\,|\,\eta) \leqslant \sum_\xi P(\xi\,|\,\eta)\, f(F_\eta[D(\xi, \eta)]) \leqslant \\ \leqslant \mathbf{M}[\min\{MF_\eta[D(\xi, \eta)], 1\}\,|\,\eta]. \tag{7.5.12}$$

Мы опустили здесь номер k, поскольку выражения в (7.5.10), (7.5.12) оказываются не зависящими от него. Выбирая некоторое граничное значение nd (не зависящее от η) и пользуясь неравенством

$$\min\{MF_\eta[D(\xi, \eta)], 1\} \leqslant \begin{cases} MF_\eta[D(\xi, \eta)] & \text{при } D(\xi, \eta) \leqslant nd, \\ 1 & \text{при } D(\xi, \eta) > nd, \end{cases} \tag{7.5.13}$$

из (7.5.12) имеем

$$P_{\text{ош}}(\,|\,\eta) \leqslant \sum_{D(\xi, \eta) \leqslant nd} MF_\eta[D(\xi, \eta)]\, P(\xi\,|\,\eta) + \sum_{D(\xi, \eta) > nd} P(\xi\,|\,\eta). \tag{7.5.14}$$

Усредняя это неравенство по η, получаем для средней вероятности ошибки оценку

$$P_{\text{ош}} \leqslant MP_1 + P_2, \tag{7.5.15}$$

где

$$P_2 = \sum_{D(\xi, \eta) > nd} P(\xi, \eta), \tag{7.5.16}$$

$$P_1 = \sum_{D(\xi, \eta) \leqslant nd} \sum_{D(\xi_l, \eta) \leqslant D(\xi, \eta)} P(\xi, \eta)\, P(\xi_l) \tag{7.5.17}$$

(в последнем выражении учтено (7.5.11)).

Функцию распределения (7.5.16) оценим при помощи теоремы 4.7. В предположении, что $d > \frac{1}{n} MD(\xi, \eta) = \gamma'(0)$, имеем

$$P_2 \leqslant e^{-n[s_0\gamma'(s_0) - \gamma(s_0)]}, \tag{7.5.18}$$

где

$$\gamma'(s_0) = d, \qquad s_0 > 0; \tag{7.5.19}$$
$$n\gamma(s) = \ln \mathbf{M}\, e^{sD(\xi, \eta)}, \; \gamma(s) = \ln \mathbf{M}\, e^{sd(x,y)}$$

— функция (7.5.5).

Вероятность (7.5.17) выражается через двумерную функцию распределения

$$F(\lambda_1, \lambda_2) = \sum_{D(\xi,\eta) \leqslant \lambda_1} \; \sum_{D(\xi_l, \eta) - D(\xi, \eta) \leqslant \lambda_2} P(\xi) P(\eta \mid \xi) P(\xi_l). \tag{7.5.20}$$

Чтобы получить для нее аналогичную оценку, нужно воспользоваться многомерным обобщением теоремы 4.6 или 4.7, т. е. формулой (4.4.13). Это дает

$$P_1 \leqslant e^{-n[r_0\varphi'_r + t\varphi'_t - \varphi(r_0, t_0)]}$$
$$\left(\varphi'_r = \frac{\partial\varphi(r_0, t_0)}{\partial r}, \quad \varphi'_t = \frac{\partial\varphi(r_0, t_0)}{\partial t}\right), \tag{7.5.21}$$

где r_0, t_0 — корни уравнений

$$\varphi'_r(r_0, t_0) = d, \qquad \varphi'_t(r_0, t_0) = 0, \tag{7.5.22}$$

а $n\varphi(r, t)$ — двумерный характеристический потенциал

$$n\varphi(r, t) = \ln \mathbf{M} e^{rD(\xi, \eta) + t[D(\xi_l, \eta) - D(\xi, \eta)]} =$$
$$= n \ln \mathbf{M} e^{rd(x, y) + t[d(x', y) - d(x, y)]} \tag{7.5.23}$$

[см. (7.5.6)]. Предполагается, что

$$r_0 < 0, \; t_0 < 0.$$

Подставим (7.5.18), (7.5.21) в (7.5.15) и заменим M на e^{nR}. Пользуясь свободой выбора постоянной d, распорядимся ею так, чтобы оценки для обоих членов в правой части формулы

$$P_{\text{ош}} \leqslant e^{nR} P_1 + P_2 \tag{7.5.24}$$

равнялись друг другу. Это даст уравнение

$$r_0\varphi'_r(r_0, t_0) - \varphi(r_0, t_0) - R = s_0\gamma'(s_0) - \gamma(s_0),$$

которое вместе с другими уравнениями, вытекающими из (7.5.19), (7.5.22), составляет систему уравнений (7.5.4). Неравенство (7.5.24) при этом перейдет в (7.5.3). Доказательство закончено.

2. Остановимся отдельно на том частном случае, когда $R\,(<I_{xy})$ настолько далеко от I_{xy}, что корень r_0 [см. уравнения (7.5.4)] становится положительным. Тогда в неравенстве (7.5.13) в качестве nd целесообразно выбрать ∞, после чего оно примет вид

$$\min\{MF_\eta[D(\xi,\eta)],\,1\}\leqslant MF_\eta[D(\xi,\eta)].$$

Вместо (7.5.15)—(7.5.17) будем иметь

$$P_{\text{ош}}\leqslant M\sum_{D(\xi_l,\eta)<D(\xi,\eta)}P(\xi,\eta)\,P(\xi_l)=MF(\infty,0) \qquad (7.5.25)$$

[см. (7.5.20); $F(\infty,\lambda_2)$ — одномерная функция распределения разности $D(\xi_l,\eta)-D(\xi,\eta)$].

Чтобы применить к неравенству (7.5.25) теорему 4.6, нужно взять характеристический потенциал

$$\ln\sum\exp\{t\,[D(\xi_l,\eta)-D(\xi,\eta)]\}\,P(\xi,\eta)\,P(\xi_l),$$

который равен $n\varphi(0,t)$ в силу (7.5.23). По теореме 4.6 (при усиливающей неравенство замене M на e^{nR}) имеем

$$P_{\text{ош}}\leqslant\exp\{nR-n\,[t^*\varphi_t'(0,t^*)-\varphi(0,t^*)]\}=$$
$$=e^{n[\varphi(0,t^*)+R]}. \qquad (7.5.26)$$

Здесь t^* — отрицательный корень уравнения

$$\varphi_t'(0,t^*)=0. \qquad (7.5.27)$$

Обратимся к уравнениям (7.5.4). Обозначим через R^* такое значение R, при котором корень r_0 равен 0.

Соответствующие этому значению другие корни s_0, t_0 обозначим через s_0^*, t_0^*. Второе уравнение из (7.5.4) примет вид

$$\varphi_t'(0,t_0^*)=0. \qquad (7.5.28)$$

Сравнивая (7.5.28) с (7.5.27), видим, что $t^*=t_0^*$. Третье уравнение из (7.5.4) при этом запишется так:

$$s_0^*\gamma'(s_0^*)-\gamma(s_0^*)+\varphi(0,t^*)+R^*=0. \qquad (7.5.29)$$

Вследствие этого неравенство (7.5.26) можно записать

$$P_{\text{ош}}\leqslant\exp\{-n\,[s_0^*\gamma'(s_0^*)-\gamma(s_0^*)+R^*-R]\}. \qquad (7.5.30)$$

Последний результат справедлив, когда $r_0>0$ и когда формулой (7.5.3) нельзя пользоваться. Учитывая характер изменения потенциалов $\varphi(r,t)$, $\gamma(s)$, можно убедиться, что условие $r_0>0$ эквивалентно условию $R<R^*$ или $s_0>s_0^*$. Введя равенствами (7.4.19), (7.4.20) образ Лежандра $\alpha(R)$ функции $\gamma(s)$, формулы (7.5.3), (7.5.30) запишем в следующем виде:

$$P_{\text{ош}}=\begin{cases}2e^{-n\alpha(R)} & \text{при } R^*<R<I_{xy},\\ e^{-n[\alpha(R^*)+R^*-R]} & \text{при } R<R^*.\end{cases} \qquad (7.5.31)$$

3. Рассмотрим важное следствие из полученных результатов. Выберем функцию $d(x, y)$ такого вида:

$$d(x, y) = -\ln P(y \mid x) + f(y), \tag{7.5.32}$$

где $f(y)$ — функция, которая будет специализирована ниже. Соответствующее этому расстояние $D(\xi, \eta) = \sum_i d(x_i, y_i)$ является несколько более общим, чем (7.1.8). Вводя обозначение

$$\gamma_y(\beta) = \ln \sum_x P^\beta(y \mid x) P(x), \tag{7.5.33}$$

функции (7.5.5), (7.5.6) в этом случае можно записать

$$\gamma(s) = \ln \sum_y e^{\gamma_y(1-s)+sf(y)},$$

$$\varphi(r, t) = \ln \sum_y \exp[\gamma_y(1-r+t) + \gamma_y(-t) + rf(y)]. \tag{7.5.34}$$

Далее входящие в (7.5.4) производные приобретают вид

$$\gamma'(s_0) = e^{-\gamma(s_0)} \sum_y e^{\gamma_y(1-s_0)+s_0 f(y)} [f(y) - \gamma'_y(1-s_0)], \tag{7.5.35}$$

$$\varphi'_r(r_0, t_0) = e^{-\varphi(r_0, t_0)} \sum_y \exp[\gamma_y(1-r_0+t_0) + \gamma_y(-t_0) + r_0 f(y)] [f(y) - \gamma'_y(1-r_0+t_0)], \tag{7.5.36}$$

$$\varphi'_t(r_0, t_0) = e^{-\varphi(r_0, t_0)} \sum_y \exp[\gamma_y(1-r_0+t_0) + \gamma_y(-t_0) + r_0 f(y)] [\gamma'_y(1-r_0+t_0) - \gamma'_y(-t_0)]. \tag{7.5.37}$$

Второе уравнение (7.5.4) будет удовлетворено, если положить

$$1 + t - r = -t, \qquad r = 1 + 2t, \tag{7.5.38}$$

в частности $r_0 = 1 + 2t_0$, так как при этом каждый член последней суммы обратится в нуль. Согласно (7.5.35), (7.5.36) первое уравнение (7.5.4) в этом случае примет вид

$$e^{-\gamma(s_0)} \sum_y e^{\gamma_y(1-s_0)+s_0 f(y)} [f(y) - \gamma'_y(1-s_0)] = e^{-\varphi(1+2t_0, t_0)} \sum_y e^{2\gamma_y(-t_0)+(1+2t_0)f(y)} [f(y) - \gamma'_y(-t_0)]. \tag{7.5.39}$$

Чтобы удовлетворить этому уравнению, положим

$$1 - s_0 = -t_0, \tag{7.5.40}$$

$$\gamma_y(1-s_0) + s_0 f(y) = 2\gamma_y(-t_0) + (1+2t_0) f(y), \tag{7.5.41}$$

т. е. выберем конкретный вид функции f:

$$f(y) = \frac{1}{1-s_0} \gamma_y(1-s_0). \tag{7.5.42}$$

Тогда суммы в левой и правой частях (7.5.39) отождествятся, и это уравнение сведется к уравнению

$$\gamma(s_0)=\varphi(1+2t_0,\ t_0)=\varphi(2s_0-1,\ s_0-1). \tag{7.5.43}$$

Но это уравнение удовлетворяется в силу тех же самых соотношений (7.5.38), (7.5.40), (7.5.42), в чем легко убедиться подстановкой их в (7.5.34).

Итак, удовлетворены два уравнения системы (7.5.4). Оставшееся уравнение вследствие (7.5.38), (7.5.40), (7.5.43) приводится к виду

$$(1-s_0)\gamma'(s_0)+R=0, \tag{7.5.44}$$

причем в силу (7.5.34), (7.5.42)

$$\gamma(s)=\ln\sum_y \exp\left[\gamma_y(1-s)+\frac{s}{1-s_0}\gamma_y(1-s_0)\right]. \tag{7.5.45}$$

Дифференцируя это выражение или учитывая (7.5.35), получим, что уравнение (7.5.44) можно записать

$$R=e^{-\gamma(s_0)}\sum_y e^{\frac{1}{1-s_0}\gamma_y(1-s_0)}[(1-s_0)\gamma_y'(1-s_0)-$$
$$-\gamma_y(1-s_0)]. \tag{7.5.46}$$

Остается проверить знаки корней s_0, r_0, t_0. Полагая $s_0=0$, из (7.5.46) находим граничное значение

$$R_{\max}=-[\gamma'(s_0)]_{s_0=0}=\sum_y e^{\gamma_y(1)}[\gamma_y'(1)-\gamma_y(1)], \tag{7.5.47}$$

которое совпадает с информацией I_{xy}. Действительно, вследствие (7.5.33), поскольку

$$\gamma_y'(1)=\sum_x\frac{P(x,\ y)}{P(y)}\ln P(y\mid x),$$
$$\gamma_y(1)=\ln P(y),$$
$$\gamma_y'(1)-\gamma_y(1)=H(y)-\frac{1}{P(y)}\sum_x P(x,\ y)\,H(y\mid x),$$

имеем

$$\sum_y e^{\gamma_y(1)}[\gamma_y'(1)-\gamma_y(1)]=H_y-H_{y\mid x}=I_{xy}. \tag{7.5.47a}$$

Анализируя выражение (7.5.46), можно получить, что $s_0>0$, если $R<I_{xy}$. Прочие корни вследствие (7.5.38), (7.5.40) равны

$$r_0=2s_0-1, \qquad t_0=s_0-1. \tag{7.5.48}$$

Они, очевидно, отрицательны, если $0<s_0<1/2$, т. е., если R достаточно близко к I_{xy}. Если же корень s_0 превосходит $1/2$, так что r_0 в силу первого равенства (7.5.48) становится положительным, то следует, вместо (7.5.3) использовать формулу (7.5.30), как об этом говорилось в предыдущем пункте. Значения s_0^*, t_0^*, получаемые из условия $r_0 = 0$, согласно (7.5.48) оказываются такими:

$$s_0^* = 1/2,\ t_0^* = -1/2.$$

«Критическое» же значение R^* получается из уравнения (7.5.46) подстановкой $s_0 = 1/2$, т. е. оказывается равным

$$R^* = -\frac{1}{2}\gamma'\left(\frac{1}{2}\right) = e^{-\gamma(1/2)} \sum_y e^{2\gamma_y(1/2)}\left[\frac{1}{2}\gamma_y'\left(\frac{1}{2}\right) - \gamma_y'\left(\frac{1}{2}\right)\right],$$

или

$$R^* = \frac{\sum\limits_y e^{2\gamma_y(1/2)}\left[\frac{1}{2}\gamma_y'\left(\frac{1}{2}\right) - \gamma\left(\frac{1}{2}\right)\right]}{\sum\limits_y e^{2\gamma_y(1/2)}} \tag{7.5.49}$$

здесь учтены (7.9.35), (7.5.42)).

Полученные выше результаты можно сформулировать в виде следующей теоремы.

Теорема 7.5. *В условиях теорем* 7.1, 7.3 *существует последовательность кодов такая, что вероятность ошибки декодирования удовлетворяет неравенству*

$$P_{\text{ош}} \leqslant \begin{cases} 2e^{n[s_0 R/(1-s_0)+\gamma(s_0)]} & \text{при } R^* \leqslant R < I_{xy}, \\ e^{n[\gamma(1/2)+R]} & \text{при } R < R^*, \end{cases} \tag{7.5.50}$$

где

$$\gamma(s_0) = \ln \sum_y e^{\gamma_y(1-s_0)/(1-s_0)} \equiv \ln \sum_y \left(\sum_x P(x) P^{1-s_0}(y \mid x)\right)^{1/(1-s_0)};$$

$s_0 \in (0, 1/2)$ — *корень уравнения* (7.5.46), *а* R^* — *значение* (7.5.49).

Коэффициент $\alpha_0 = -s_0 R/(1 - s_0) - \gamma(s_0)$ в экспоненте (7.5.50) получен подстановкой равенства (7.5.44) в выражение

$$\alpha_0 = s_0\gamma'(s_0) - \gamma(s_0), \tag{7.5.51}$$

стоящее в показателе формул (7.5.3), (7.5.30), (7.5.31).

4. Исследуем поведение выражений, приведенных в теореме 7.5, при значениях R, близких к предельному значению I_{xy}. Это исследование позволит сравнить указанные результаты с результатами теоремы 7.3.

При любых расстояниях $d(x, y)$, как видно из (7.5.5), функция $\gamma(s)$ обладает свойством $\gamma(0) = 0$. Это справедливо и для частного случая (7.5.32), (7.5.42). Поэтому разложение функции (7.5.45) в двойной ряд по s и s_0 будет иметь следующие члены:

$$\gamma(s) = \gamma_{10}s + \gamma_{20}s^2 + \gamma_{11}ss_0 + \gamma_{30}s^3 + \gamma_{21}s^2 s_0 + \\ + \gamma_{12}ss_0^2 + \ldots, \tag{7.5.52}$$

причем

$$\gamma_{10} = -I_{xy} \tag{7.5.53}$$

в силу (7.5.47), (7.5.47а).

Подставляя (7.5.52) в (7.5.51), получаем

$$\alpha_0 = \gamma_{20} s_0^2 + (2\gamma_{30} + \gamma_{21}) s_0^3 + s_0^4 \dots . \tag{7.5.54}$$

Чтобы выразить s_0 через $I_{xy} - R$, подставим (7.5.52) в уравнение (7.5.44), что дает

$$-R = (1 - s_0)[\gamma_{10} + (2\gamma_{20} + \gamma_{11}) s_0 + (3\gamma_{30} + 2\gamma_{21} + \gamma_{12}) s_0^2 + \dots].$$

Учитывая (7.5.53), отсюда имеем

$$\begin{aligned} I_{xy} - R = (2\gamma_{20} + \gamma_{11} - \gamma_{10}) s_0 + (3\gamma_{30} + 2\gamma_{21} + \\ + \gamma_{12} - 2\gamma_{20} - \gamma_{11}) s_0^2 + \dots s_0^3 + \dots \end{aligned} \tag{7.5.55}$$

и, разрешая относительно s_0,

$$\begin{aligned} s_0 &= \frac{I_{xy} - R}{2\gamma_{20} + \gamma_{11} - \gamma_{10}} - \frac{3\gamma_{30} + 2\gamma_{21} + \gamma_{12} - 2\gamma_{20} - \gamma_{11}}{2\gamma_{20} + \gamma_{11} - \gamma_{10}} s_0^2 - s_0^3 \dots = \\ &= \frac{I_{xy} - R}{2\gamma_{20} + \gamma_{11} - \gamma_{10}} - \frac{3\gamma_{30} + 2\gamma_{21} + \gamma_{12} - 2\gamma_{20} - \gamma_{11}}{2\gamma_{20} + \gamma_{21} - \gamma_{10}} \times \\ &\times \left(\frac{I_{xy} - R}{2\gamma_{20} + \gamma_{21} - \gamma_{10}} \right)^2 - \dots . \end{aligned} \tag{7.5.56}$$

Подстановка этого выражения в (7.5.54) позволяет найти величину α_0 с точностью порядка $(I_{xy} - R)^3$.

Коэффициенты γ_{ik} могут быть вычислены при помощи (7.5.45). Для удобства вычисления преобразуем это выражение к несколько другому виду, вводя условный характеристический потенциал случайной информации

$$\mu(t \mid y) = \ln \sum_x e^{-tI(x,\, y)} P(x \mid y) = \ln \sum_x P^{1-t}(y \mid x) P(x) P^{t-1}(y), \tag{7.5.57}$$

который связан с функцией (7.5.33) очевидным соотношением

$$\gamma_y(1 - t) = \mu(t \mid y) + (1 - t) \ln P(y). \tag{7.5.58}$$

Производные от (7.5.57) имеют смысл условных семиинвариантов от $-I(x, y)$

$$\begin{aligned} &\mu(0 \mid y) = 0, \\ &\mu'(0 \mid y) = -\mathbf{M}[I(x, y) \mid y] = -m, \\ &\mu''(0 \mid y) = \mathbf{M}[I^2(x, y) \mid y] - \{\mathbf{M}[I(x, y) \mid y]\}^2 = \\ &\qquad = \mathbf{D}[I(x, y) \mid y] = D, \\ &\mu'''(0 \mid y) = -k, \\ &\dots\dots\dots\dots\dots\dots \end{aligned} \tag{7.5.59}$$

[см. формулу (4.1.12)].

Подстановка (7.5.58) в (7.5.45) дает

$$\gamma(s) = \ln \sum_{y} \exp\left[\mu(s \mid y) + \frac{s}{1-s_0}\mu(s_0 \mid y)\right] P(y). \tag{7.5.60}$$

Стоящее в экспоненте выражение, учитывая (7.5.59), представляем в форме

$$\mu(s \mid y) + s(1 + s_0 + s_0^2 + \ldots)\mu(s_0 \mid y) =$$
$$= -ms + \frac{1}{2}Ds^2 - \frac{1}{6}ks^3 + \ldots + s\Big(-ms_0 +$$
$$+ \left(\frac{1}{2}D - m\right)s_0^2 + \ldots\Big].$$

Следовательно,

$$\exp\left[\mu(s \mid y) + \frac{s}{1-s_0}\mu(s_0 \mid y)\right] = 1 - ms + \frac{1}{2}m^2 s^2 - \frac{1}{6}m^3 s^3 +$$
$$+ \frac{1}{2}Ds^2 + \ldots - \frac{1}{6}ks^3 + \ldots - \frac{1}{2}mDs^3 + m^2s^2s_0 - mss_0 +$$
$$+ \ldots + \left(\frac{1}{2}D - m\right)ss_0^2 + \ldots = 1 - ms + \frac{1}{2}(D + m^2)s^2 - mss_0 -$$
$$- \frac{1}{6}(k + 3Dm + m^3)s^3 + m^2s^2s_0 + \left(\frac{1}{2}D - m\right)ss_0^2 + \ldots.$$

После усреднения этого выражения по y в соответствии с (7.5.60), обозначая среднее значение чертой сверху, будем иметь

$$\gamma(s) = \ln\Big[1 - \overline{m}s + \frac{1}{2}(\overline{D + m^2})s^2 - \overline{m}ss_0 -$$
$$- \frac{1}{6}(\overline{k + 3Dm + m^3})s^3 + \overline{m^2}s^2s_0 + \left(\frac{1}{2}\overline{D} - \overline{m}\right)ss_0^2 + \ldots\Big].$$

Отсюда

$$\gamma(s) = -\overline{m}s + \frac{1}{2}\left[\overline{D} + \overline{m^2} - (\overline{m})^2\right]s^2 - \overline{m}ss_0 -$$
$$- \frac{1}{6}\left[\overline{k} + 3\overline{Dm} + \overline{m^3} + 2(\overline{m})^3 - 3(\overline{D} + \overline{m^2})\overline{m}\right]s^3 +$$
$$+ \overline{m^2}s^2s_0 - (\overline{m})^2s^2s_0 + \left(\frac{1}{2}\overline{D} - \overline{m}\right)ss_0^2 + \ldots. \tag{7.5.61}$$

Если в выражении (7.5.60) положить $s_0 = 0$, то оно в силу (7.5.57) обратится в характеристический потенциал

$$\mu(s) = \ln \sum_{x,\, y} e^{-sI(x,\, y)} P(x,\, y),$$

совпадающий с (7.4.3). Поэтому члены с s, s^2, s^3, ... в (7.5.61) заведомо пропорциональны полным семиинвариантам

$$\gamma_{10} = \mu'(0) = -I_{xy},$$

$$\gamma_{20}=\frac{1}{2}\mu''(0)=\frac{1}{2}\mathbf{D}[I(x, y)], \tag{7.5.62}$$
$$\gamma_{30}=\frac{1}{6}\mu'''(0)=-\frac{1}{6}\mathbf{K}_3[I(x, y)],$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$

где $\mathbf{K}_3$ обозначает третий кумулянт.

В дополнение к данным соотношениям из (7.5.61) имеем

$$\gamma_{11}=\gamma_{10}=-I_{xy}, \qquad \gamma_{21}=\overline{m^2}I_{xy}^2, \qquad \gamma_{12}=\frac{1}{2}\overline{D}-I_{xy}.$$

Вследствие этого соотношение (7.5.56) принимает вид

$$s_0=\frac{I_{xy}-R}{\mu''(0)}-\frac{\left[\frac{1}{2}\mu'''(0)+2\overline{m^2}2I_{xy}^2+\frac{1}{2}\overline{D}-\mu''(0)\right]}{[\mu''(0)]^3}(I_{xy}-R)^2.$$

Подставим это равенство в формулу (7.5.54), принимающую вид

$$\alpha_0=\frac{1}{2}\mu''(0)\,s_0^2+\left[\frac{1}{3}\mu'''(0)\overline{m^2}-I_{xy}^2\right]s_0^3+\ldots .$$

В результате получим

$$\alpha_0=\frac{(I_{xy}-R)^2}{2\mu''(0)}+\left[2\mu''(0)-\frac{3}{2}\overline{D}-\right.$$
$$\left.-\frac{1}{6}\mu'''(0)\right]\frac{(I_{xy}-R)^3}{[\mu''(0)]^3} \quad (R<I_{xy}). \tag{7.5.63}$$

Здесь мы учли, что согласно третьему равенству (7.5.59)

$$\mu''(0)=\overline{D}-MI^2(x, y)-I_{xy}^2-\overline{D}=\overline{m^2}-I_{xy}^2.$$

Сравним этот результат с формулой (7.4.24), которая справедлива в том же приближении. При этом учтем, что

$$\mu''(0)\geqslant 0 \text{ и } \mu''(0)\geqslant\overline{D},$$

поскольку условная дисперсия $\overline{D}$ не превосходит безусловную $\mu''(0)$.

В (7.5.63) присутствуют добавочные положительные члены по сравнению с (7.4.24), благодаря которым $\alpha_0>\alpha$. Следовательно, неравенство (7.5.50) является более сильным, чем неравенство (7.4.2) (по крайней мере, при значениях R, достаточно близких к I_{xy}). Таким образом, теорема 7.5 более сильная, чем теорема 7.3.

Ряд других результатов, дающих усиленную оценку поведения вероятности ошибки декодирования, приведены в книге Фано [1].

7.6. Некоторые общие соотношения между энтропиями и взаимными информациями при кодировании и декодировании

1. Рассматриваемый нами канал с помехами характеризуется условными вероятностями $P(\eta \mid \xi) = \prod_i P(y_i \mid x_i)$. Вероятности $P(\xi) = \prod_i P(x_i)$ входной величины ξ определяют способ получения случайного кода $\xi_1, \ldots, \xi_M$. Передаваемое сообщение, имеющее номер $k = 1, \ldots, M$, сопоставляется с k-й кодовой точкой ξ_k. При декодировании наблюдаемая случайная величина η определяет номер $l(\eta)$ принимаемого сообщения. Как отмечалось ранее, выбирается та кодовая точка ξ_l, которая в смысле некоторого «расстояния» является «ближайшей» к наблюденной точке η.

Преобразование $l = l(\eta)$ является вырожденным преобразованием. Поэтому в силу неравенства (6.3.9) имеем

$$I_{k,l}(\mid \xi_1, \ldots, \xi_M) \leqslant I_{k,\eta}(\mid \xi_1, \ldots, \xi_M). \tag{7.6.1}$$

Применяя (6.3.9), нужно k отождествить с y, l отождествить с x_1, а под x_2 понимать случайную величину, дополняющую l до η (так что η будет совпадать с x_1, x_2).

Код $\xi_1, \ldots, \xi_M$ в (7.6.1) предполагается фиксированным. Усредняя (7.6.1) по различным кодам с весом $P(\xi_1) \ldots P(\xi_M)$ и обозначая результаты усреднения через $I_{kl \mid \xi_1 \ldots \xi_M}$, $I_{k\eta \mid \xi_1 \ldots \xi_M}$, получаем

$$I_{kl \mid \xi_1 \ldots \xi_M} \leqslant I_{k\eta \mid \xi_1, \ldots, \xi_M}. \tag{7.6.2}$$

2. Сравним теперь количество информации $I_{k\eta \mid \xi_1 \ldots \xi_M}$ с $I_{\xi\eta}$. Первое количество информации определяется формулой

$$I_{k\eta \mid \xi_1 \ldots \xi_M} = \sum I_{k\eta}(\mid \xi_1 \ldots \xi_M) P(\xi_1) \ldots P(\xi_M), \tag{7.6.3}$$

где

$$\begin{aligned} I_{k\eta}(\mid \xi_1, \ldots, \xi_M) &= H_\eta(\mid \xi_1, \ldots, \xi_M) - H_{\eta \mid k}(\mid \xi_1, \ldots, \xi_M) = \\ &= \sum_\eta f\Bigl(\sum_k P(k) P(\eta \mid \xi_k)\Bigr) - \sum_k P(k) H_\eta(\mid \xi_k); \end{aligned} \tag{7.6.4}$$

$$f(z) = -z \ln z. \tag{7.6.5}$$

Информацию же $I_{\xi\eta}$ можно записать

$$I_{\xi\eta} = H_\eta - H_{\eta\mid\xi} = \sum_\eta f\Bigl(\sum_\xi P(\xi) P(\eta \mid \xi)\Bigr) - \sum_\xi P(\xi) H_\eta(\mid \xi). \tag{7.6.6}$$

Нетрудно убедиться, что после усреднения (7.6.4) по $\xi_1, \ldots, \xi_k$ второй (вычитаемый) член совпадает со вторым членом формулы (7.6.6). Действительно, математическое ожидание

$$\sum_{\xi_k} P(\xi_k) H\ \ (\mid \xi_k)$$

энтропии $H_\eta(\mid\xi)$ не зависит от k в силу равноправности всех k (см. § 7.2), так что

$$\sum_k P(k)\sum_{\xi_k} P(\xi_k)\,H_\eta(\mid\xi_k)=\sum P(\xi)\,H_\eta(\mid\xi).$$

Разность информаций (7.6.6) и (7.6.3), следовательно, равна

$$I_{\xi\eta}-I_{k\eta\mid\xi_1\ldots\xi_M}=\sum_\eta f\Big(\sum_\xi P(\xi)\,P(\eta\mid\xi)\Big)-\\-\mathbf{M}\sum_\eta f\Big(\sum_k P(k)\,P(\eta\mid\xi_k)\Big), \qquad (7.6.7)$$

где $\mathbf{M}$ обозначает усреднение по $\xi_1, \ldots, \xi_M$. Вследствие выпуклости функции (7.6.5) можно воспользоваться формулой

$$\mathbf{M}f(\zeta)\leqslant f(\mathbf{M}\zeta)$$

[см. (1.2.4)] при $\zeta=\sum_k P(k)\,P(\eta\mid\xi_k)$,

т. е. неравенством

$$f\Big(\sum_k P(k)\,\mathbf{M}P(\eta\mid\xi_k)\Big)-\mathbf{M}f\Big(\sum_k P(k)\,P(\eta\mid\xi_k)\Big)\geqslant 0.$$

Но $\mathbf{M}P(\eta\mid\xi_k)=\sum_k P(\xi_k)P(\eta\mid\xi_k)$ не зависит от k и совпадает с аргументом функции f в первом члене соотношения (7.6.7). Поэтому при каждом η

$$f\Big(\sum_\xi P(\xi)\,P(\eta\mid\xi)\Big)-\mathbf{M}f\Big(\sum_k P(k)\,P(\eta\mid\xi_k)\Big)\geqslant 0,$$

и из (7.6.7) получаем

$$I_{\xi\eta}-I_{k\eta\mid\xi_1\ldots\xi_M}\geqslant 0. \qquad (7.6.8)$$

Объединяя неравенства (7.6.2), (7.6.8), имеем

$$I_{kl}\leqslant I_{k\eta}\leqslant I_{\xi\eta}, \qquad (7.6.9)$$

где для краткости опущены $\mid\xi_1\ldots\xi_M$.

3. Полезно связать информацию I_{kl} между входными и выходными сообщениями с вероятностью ошибки. Рассмотрим энтропию $H_l(\mid k, \xi_1, \ldots, \xi_k)$, соответствующую фиксированному переданному сообщению k. После передачи сообщения k остается некоторая неопределенность, касающаяся того, какое сообщение будет принято. Эту неопределенность, численно равную $H_l(\mid k; \xi_1, \ldots, \xi_M)$, пользуясь иерархическим свойством энтропии (§ 1.3), можно представить в виде суммы двух членов. Эти члены соответствуют устранению неопределенности в два этапа. На первом этапе указывается, является ли принятое сообщение правильным, т. е. совпадают ли l и k. Эта неопределенность равна

$$-P_{\text{ош}}(k;\ \xi_1, \ldots, \xi_M)\ln P_{\text{ош}}(k, \xi_1, \ldots, \xi_M)-$$

$$-[1-P_{\text{ош}}(k, \xi_1, \dots, \xi_M)]\ln[1-P_{\text{ош}}(k_1, \xi_1, \dots, \xi_M)] \equiv$$
$$\equiv h_2[P_{\text{ош}}(k, \xi_1, \dots, \xi_M)],$$

где $P_{\text{ош}}(k, \xi_1, \dots, \xi_M)$ — вероятность ошибки декодирования при условии, что передано сообщение k (код фиксирован). На втором этапе, если $l \neq k$, следует указать, какое же сообщение из остальных сообщений принято. Соответствующая неопределенность не может быть больше $P_{\text{ош}} \ln(1-M)$. Следовательно,

$$H_l(|k, \xi_1, \dots, \xi_M) < h_2(P_{\text{ош}}(k, \xi_1, \dots, \xi_M)) +$$
$$+ P_{\text{ош}}(k, \xi_1, \dots, \xi_M)\ln(M-1).$$

Усредним это неравенство по k, пользуясь формулой

$$\mathbf{M}h_2(\zeta) \leqslant h_2(\mathbf{M}\zeta), \tag{7.6.10}$$

справедливой в силу выпуклости функции $h_2(z) = -z\ln z - (1-z)\ln(1-z)$ и неравенства (1.2.4). Это дает

$$H_{l\,|\,k}(|\xi_1, \dots, \xi_M) \leqslant h_2(P_{\text{ош}}(\xi_1, \dots, \xi_M) -$$
$$-P_{\text{ош}}(\xi_1, \dots, \xi_M)\ln(M-1).$$

Можно далее произвести усреднение по ансамблю случайных кодов и аналогично, используя (7.6.10) еще раз, получить

$$H_{l|k,\,\xi_1 \dots \xi_M} \leqslant h_2(P_{\text{ош}}) + P_{\text{ош}}\ln(M-1). \tag{7.6.11}$$

Поскольку $I_{kl\,|\,\xi_1 \dots \xi_M} = H_{l\,|\,\xi_1 \dots \xi_M} - H_{l\,|\,k,\,\xi_1 \dots \xi_M}$, то из (7.6.11) следует, что

$$I_{kl\,|\,\xi_1 \dots \xi_M} \geqslant H_{l|\xi_1 \dots \xi_M} - P_{\text{ош}}\ln(M-1) - h_2(P_{\text{ош}}) \tag{7.6.12}$$

Точно те же самые рассуждения можно провести, поменяв местами k и l. Тогда по аналогии с (7.6.12) будем иметь

$$I_{kl|\xi_1 \dots \xi_M} \geqslant H_k - P_{\text{ош}}\ln(M-1) - h_2(P_{\text{ош}})$$
$$(P_{\text{ош}} = \mathbf{P}(k \neq l) = \mathbf{MP}(k \neq l\,|\,k) = \mathbf{MP}(k \neq l\,|\,l)). \tag{7.6.13}$$

В предположении равновероятности всех M возможных сообщений $k = 1, \dots, M$, имеем $H_k = \ln M$. Кроме того, очевидно, $h_2(P_{\text{ош}}) \leqslant \ln 2 = 1$ бит. Поэтому (7.6.13) можно записать

$$\ln M - P_{\text{ош}}\ln(M-1) - \ln 2 \leqslant I_{kl\,|\,\xi_1 \dots \xi_M}. \tag{7.6.14}$$

Ранее мы полагали $M = [e^{nR}]$; при этом

$$e^{nR} - 1 \leqslant M, \quad e^{nR} \geqslant M-1,$$

и (7.6.14) принимает вид

$$\ln(e^{nR}-1) - P_{\text{ош}}\,nR - \ln 2 \leqslant I_{kl\,|\,\xi_1 \dots \xi_M},$$

т. е.

$$1 + \ln(1-e^{-nR})/nR - P_{\text{ош}} - \ln 2/nR \leqslant I_{kl\,|\,\xi_1 \dots \xi_M}/nR.$$

Учитывая (7.6.9) и соотношение $I_{\xi\eta} = nI_{xy}$, отсюда имеем результирующее неравенство

$$P_{\text{ош}} \geqslant 1 - \frac{I_{xy}}{R} - \frac{\ln 2}{nR} + \frac{1}{nR} \ln(1 - e^{-nR}), \tag{7.6.15}$$

определяющее нижнюю границу для вероятности ошибки декодирования. При $nR \to \infty$ неравенство (7.6.15) переходит в асимптотическую формулу

$$P_{\text{ош}} \geqslant 1 - I_{xy}/R,$$

которой имеет смысл пользоваться при $R > I_{xy}$ (если $R < I_{xy}$, то неравенство становится тривиальным). Согласно этой формуле асимптотическая безошибочность декодирования заведомо не имеет места при $R > I_{xy}$, так что граница I_{xy} для R является существенной.

4. Объединяя формулы (7.6.9), (7.6.14) и заменяя множитель при $P_{\text{ош}}$ на $\ln M$, будем иметь результат

$$\ln M - P_{\text{ош}} \ln M - \ln 2 \leqslant I_{kl} \leqslant I_{k\eta} \leqslant I_{\xi\eta}. \tag{7.6.16}$$

Величина $\ln M = I$ есть количество информации в хартлиевском понимании. Из асимптотической безошибочности декодирования следует, что это количество информации близко к шенноновскому количеству информации $I_{\xi\eta}$. Поделив (7.6.16) на $\ln M = I$, имеем

$$1 - P_{\text{ош}} - \ln 2/I \leqslant I_{kl}/I \leqslant I_{k\eta}/I \leqslant I_{\xi\eta}/I. \tag{7.6.17}$$

Из теоремы 7.1 и других следует, что можно увеличивать I и производить кодирование и декодирование таким образом, что $I_{\xi\eta}/I \to 1$ при $n \to 0$, и в то же самое время $P_{\text{ош}} \to 0$. Тогда, очевидно, длина диапазона $\left[1 - P_{\text{ош}} - \frac{\ln 2}{I}, \frac{I_{\xi\eta}}{I}\right]$ будет стремиться к нулю:

$$I_{\xi\eta}/I - 1 + P_{\text{ош}} + \ln 2/I \to 0. \tag{7.6.18}$$

Это значит, что с ростом n все точнее выполняются приближенные равенства

$$I \approx I_{kl}/I \approx I_{k\eta}/I \approx I_{\xi\eta}/I$$

или

$$I/n \approx I_{kl}/n \approx I_{k\eta}/n \approx I_{\xi\eta}/n. \tag{7.6.19}$$

Эти асимптотические соотношения обобщают равенства (6.1.17), относящиеся к простым помехам. В соответствии с этим произвольные помехи можно считать асимптотически эквивалентными простым помехам. Номер l кодовой области G_l является асимптотически достаточной координатой (см. § 6.1).

Подобно тому, как в случае простых помех (§ 6.1) обоснованием шенноновского количества информации

$$I_{xy} = H_x - H_{x \mid y} \tag{7.6.20}$$

являлась его сводимость (в соответствии с (6.1.17)) к более простому «больцмановскому» количеству информации

$$H_k = -\sum_k P(k) \ln P(k),$$

так в случае произвольных помех наиболее убедительным обоснованием количества информации (7.6.20) является асимптотическое равенство $I_{\xi\eta}/\ln M \approx 1$.

Глава 8

ПРОПУСКНАЯ СПОСОБНОСТЬ КАНАЛОВ. ВАЖНЫЕ ЧАСТНЫЕ СЛУЧАИ КАНАЛОВ

Данная глава посвящена второй вариационной задаче, в которой отыскивается экстремум шенноновского количества информации по различным входным распределениям. Канал, т. е. условное распределение на его выходе при фиксированном входном сигнале, предполагается известным. Максимальное значение количества информации называется пропускной способностью. В отличие от традиционных способов изложения здесь с самого начала вводится дополнительное условие, касающееся среднего значения некоторой функции от входных переменных, т. е. рассматривается условная вариационная задача. Результаты, относящиеся к случаю отсутствия условия, получаются как частный случай приводимых общих результатов.

В соответствии с характером изложения, принятым в настоящей книге, вводятся потенциалы, через которые выражается условная пропускная способность. Более подробно разбирается ряд важных частных случаев каналов, для которых удается получить результаты в более явном виде. Так, для случая гауссовых каналов общие результирующие формулы, имеющие матричный вид, находятся путем применения матричной техники.

Изложение в настоящей главе ведется, в основном, применительно к случаю дискретных случайных величин x, y. Многие рассуждения и результаты, однако, непосредственно переносятся на более общий случай путем изменения обозначения (например, замены $P(y \mid x)$, $P(x)$ на $P(dy \mid x)$, $P(dx)$ и пр.).

8.1. Определение пропускной способности каналов

В предыдущей главе предполагалось, что статистически заданы не только помехи в канале, описываемые условными вероятностями $P(y|x)$, но и сигналы на входе канала, описываемые априорными вероятностями $P(x)$. Поэтому в качестве канала связи бралась система, характеризуемая совокупностью распределений $[P(y|x), P(x)]$ или, иначе, совместным распределением $P(x, y)$.

Между тем, распределение $P(x)$ обычно не является неотъемлемой частью реального канала связи, как условное распределение

$P(y|x)$. Иногда имеет смысл не конкретизировать заранее распределение $P(x)$, а фиксировать лишь некоторые технически важные требования, скажем вида

$$a_1 \leqslant \sum_x c(x) P(x) \leqslant a_2, \tag{8.1.1}$$

где $c(x)$ — известная функция. Обычно достаточно рассматривать лишь одностороннее условие типа

$$\mathbf{M} c(x) \leqslant a_0. \tag{8.1.2}$$

К требованиям такого типа относится, например, ограничение средней мощности передатчика. Тогда канал будет характеризоваться совокупностью распределения $P(y|x)$ и условия (8.1.1) или (8.1.2). В соответствии с этим, ссылаясь на такой канал, мы будем употреблять выражения «канал $[P(y|x),\ c(x),\ a_1,\ a_2]$», «канал $[P(y|x),\ c(x),\ a_0]$» или коротко «канал $[P(y|x),\ c(x)]$». В некоторых случаях условия (8.1.1), (8.1.2) могут вообще отсутствовать. Тогда канал характеризуется лишь условным распределением $P(y|x)$. Как видно из результатов предыдущей главы, важнейшей информационной характеристикой канала $[P(y|x),\ P(x)]$ является величина I_{xy}, которая определяет верхний предел для асимптотически безошибочно передаваемого количества информации. Аналогичной величиной для канала $[P(y|x),\ c(x)]$ является пропускная способность C. Пользуясь свободой выбора распределения $P(x)$, остающейся после фиксации условия (8.1.1) или (8.1.2), естественно подобрать наиболее выгодное распределение с точки зрения количества информации I_{xy}. Это приводит к следующему определению.

Пропускной способностью канала $[P(y|x),\ c(x)]$ называется максимальное количество информации связи между входом и выходом:

$$C = C\,[P(y,\ x),\ c(x)] = \sup_{P(x)} I_{xy}, \tag{8.1.3}$$

где максимизация идет по всем $P(x)$, совместимым с условием (8.1.1) или (8.1.2).

В результате указанной максимизации можно найти оптимальное распределение $P_0(x)$, для которого

$$C = I_{xy}, \tag{8.1.4}$$

или хотя бы ε-оптимальное $P_\varepsilon(x)$, для которого

$$0 \leqslant C - I_{xy} < \varepsilon,$$

где ε сколь угодно мало. После этого можно рассмотреть систему $[P(y|x),\ P_0(x)]$ или $[P(y|x),\ P_\varepsilon(x)]$ и использовать для нее результаты предыдущей главы. Так, теорема 7.1 даст следующее утверждение:

Теорема 8.1. *Пусть имеется стационарный канал, являющийся n-й степенью канала $[P(y \mid x), c(x)]$. Пусть при $n \to \infty$ количество передаваемой информации $\ln M$ растет по закону*

$$\ln M = \ln [e^{nR}],$$

где R — не зависящая от n величина, удовлетворяющая неравенству

$$R < C, \tag{8.1.5}$$

а $C < \infty$ — пропускная способность канала $[P(y \mid x), c(x)]$. Тогда существует последовательность кодов таких, что

$$P_{\text{ош}} \to 0, \text{ при } n \to \infty.$$

Для того чтобы вывести эту теорему из теоремы 7.1, очевидно достаточно подобрать такое распределение $P(x)$, совместимое с условием (8.1.1) или (8.1.2), чтобы выполнялось неравенство $R < I_{xy} \leqslant C$. Вследствие (8.1.3), (8.1.5) это можно сделать.

Аналогичным образом на каналы $[P(y \mid x), c(x)]$ распространяются и другие результаты, полученные в предыдущей главе для каналов $[P(y\ x), P(x)]$. Мы не будем на этом более останавливаться.

Согласно определению (8.1.3) пропускной способности канала с помехами ее вычисление сводится к решению определенной экстремальной задачи. Аналогичная ситуация встречалась в § 3.2, 3.3, 3.6, где рассматривалась пропускная способность канала без помех. Разница между этими двумя случаями в том, что раньше максимизировалась энтропия, а теперь шенноновское количество информации. Несмотря на это различие, между данными экстремальными задачами имеется много общего. В отличие от прежней экстремальной задачи настоящую задачу будем называть второй экстремальной задачей теории информации.

Экстремум (8.1.3) обычно достигается на границе допустимого диапазона (8.1.1) средних штрафов, поэтому условие (8.1.1) можно заменить односторонним неравенством типа (8.1.2) или даже равенством

$$\mathbf{M} c(x) = a \tag{8.1.6}$$

(где a совпадает с a_0 или a_2). В этом случае канал можно специализировать как систему $[P(y, x), c(x), a]$ и говорить, что пропускная способность канала соответствует *уровню потерь a*.

Рассматриваемый канал не обязан быть дискретным. Все сказанное выше относится и к тому случаю, когда случайные величины x, y являются произвольными: непрерывными, комбинированными и т. п. При абстрактной формулировке задачи следует рассматривать абстрактное пространство Ω, составленное из точек ω, и два борелевских поля $\mathcal{F}_1$, $\mathcal{F}_2$ его подмножеств. Поле $\mathcal{F}_1$ соответствует случайной величине x, а $\mathcal{F}_2$ — величине y. Функция штрафов $c(x) = c(\omega)$ есть $\mathcal{F}_1$ — измеримая функция от ω. Кроме того должно быть задано условное распределение $P(\Lambda \mid \mathcal{F}_1)$, $\Lambda \in \mathcal{F}_2$.

Тогда система $[P(\cdot \mid \mathscr{F}_1),\ c(\cdot),\ a]$ составляет абстрактный канал Его пропускная способность определяется как максимальное значение (8.1.3) шенновского количества информации

$$I_{xy} = \int P(d\omega) \int\limits_{d\omega' \in \mathscr{F}_2} \ln\left[\frac{P(d\omega' \mid \mathscr{F}_1)}{P(d\omega')}\right] P(d\omega' \mid \mathscr{F}_1) P(d\omega)$$

$$\left(P(\Lambda) = \int P(\Lambda \mid \mathscr{F}_1) P(d\omega), \quad \Lambda \in \mathscr{F}_2\right)$$

в обобщенной версии (см. § 6.4). При этом сравниваются различные распределения $P(\cdot)$ на $\mathscr{F}_1$, удовлетворяющие условию типа (8.1.1), (8.1.2) или (8.1.6).

8.2. Решение второй экстремальной задачи. Соотношения для пропускной способности и потенциала

1. Обозначим черех X пространство значений, которые может принимать входная величина x. В экстремальном распределении $P_0(dx)$, соответствующем пропускной способности (8.1.3), вероятность может быть сконцентрирована лишь на части указанного пространства. Обозначим через $\widetilde{X}$ минимальное подмножество $\widetilde{X} \in X$, для которого $P_0(\widetilde{X}) = 1$ (т. е. $P_0(X - \widetilde{X}) = 0$), и будем называть его «активной областью».

При решении экстремальной задачи для удобства записи будем полагать, что x — дискретная величина. Тогда можно рассматривать вероятности $P(x)$ отдельных точек x и брать по ним частные производные. В противном случае пришлось бы вводить вариационные производные, что связано с некоторым усложнением, не носящим, впрочем, принципиального характера.

Ищется условный экстремум по $P(x)$ выражения

$$I_{xy} = \sum_{x,\,y} P(x) P(y \mid x) \ln \frac{P(y \mid x)}{\sum\limits_x P(x) P(y \mid x)}$$

при дополнительных условиях

$$\sum_x c(x) P(x) = a, \tag{8.2.1}$$

$$\sum_x P(x) = 1. \tag{8.2.2}$$

Условие неотрицательности вероятностей $P(x)$ пока можно не фиксировать, а проверить его после решения задачи.

Вводя неопределенные множители Лагранжа $-\beta$, $1 + \beta\varphi$, которые будут затем определяться из условий (8.2.1), (8.2.2), составим выражение

$$K = \sum_{x,\,y} P(x) P(y \mid x) \ln \frac{P(y \mid x)}{\sum\limits_x P(x) P(y \mid x)} -$$

$$-\beta \sum_x c(x) P(x) + (1+\beta\varphi) \sum_x P(x). \tag{8.2.3}$$

Будем искать его экстремум, варьируя величины $P(x)$, $x \in \widetilde{X}$, соответствующие активной области $\widetilde{X}$. Приравнивая нулю частную производную от (8.2.3) по $P(x)$, $x \in \widetilde{X}$, получаем уравнение

$$\sum_y P(y \mid x) \ln \frac{P(y \mid x)}{P(y)} - \beta c(x) + \beta\varphi = 0 \quad \text{при } x \in \widetilde{X}, \tag{8.2.4}$$

являющееся необходимым условием экстремальности. Здесь

$$P(y) = \sum_x P(x) P(y \mid x). \tag{8.2.5}$$

Умножая (8.2.4) на $P(x)$ и суммируя по x с учетом (8.2.1), (8.2.2), имеем $I_{xy} = \beta(a - \varphi)$, т. е.

$$C = \beta(a - \varphi). \tag{8.2.6}$$

Это соотношение позволяет исключить φ из уравнения (8.2.4), записав его в виде

$$\sum_y P(y \mid x) \ln \frac{P(y \mid x)}{P(y)} = C - \beta a + \beta c(x) \quad \text{при } x \in \widetilde{X}. \tag{8.2.7}$$

Уравнения (8.2.7), (8.2.1), (8.2.2) образуют систему уравнений, служащую для совместного определения неизвестных C, β, $P(x)$ ($x \in \widetilde{X}$), если область $\widetilde{X}$ является уже выбранной. При правильном выборе этой области в результате решений указанных уравнений должны получиться положительные вероятности $P(x)$, $x \in \widetilde{X}$.

Основное уравнение (8.2.4) или (8.2.7), умножая на $P(x)$, можно записать также в форме равенства

$$\sum_y P(x) P(y|x) \ln \frac{P(y \mid x)}{P(y)} = [C - \beta a + \beta c(x)] P(x), \tag{8.2.8}$$

которое удобно тем, что справедливо при всех значениях $x \in X$, а не только при $x \in \widetilde{X}$. Вне области $\widetilde{X}$ уравнение (8.2.7) не обязано выполняться.

Не представляет труда записать обобщение приведенных уравнений на тот случай, когда случайные переменные не являются дискретными. Вместо (8.2.7), (8.2.5) будем иметь

$$\int P(dy \mid x) \ln \frac{P(dy \mid x)}{P(dy)} = C - \beta a + \beta c(x), \quad x \in \widetilde{X}$$

$$\left(P(dy) = \int P(dy \mid x) P(dx)\right). \tag{8.2.9}$$

Искомым распределением при этом будет $P(dx) = p(x)dx$, $x \in \widetilde{X}$.

2. Возвращаясь к дискретной версии, докажем следующее положение.

Теорема 8.2. *Решение уравнений* (8.2.7), (8.2.1), (8.2.2) *действительно соответствует максимуму информации* I_{xy} *относительно вариаций распределения* $P(x)$, *оставляющих неизменной активную область* X.

Доказательство. Вычислим матрицу вторых производных от выражения (8.2.3):

$$\frac{\partial^2 K}{\partial P(x)\,\partial P(x')} = \frac{\partial^2 I_{xy}}{\partial P(x)\,dP(x')} = -\sum_y \frac{P(y\mid x)\,P(y\mid x')}{P(y)} \quad (8.2.10)$$

(β, φ не варьируются). Легко видеть, что эта матрица является неположительно определенной (поскольку $P(y) \geqslant 0$), так что K является выпуклой функцией от $P(x)$, $x \in \widetilde{X}$. В самом деле

$$-\sum_{x,\,x'}\left(\sum_y \frac{P(y\mid x)\,P(y\mid x')}{P(y)}\right) f(x) f(x') =$$

$$= -\sum_y \frac{1}{P(y)}\left[\sum_x P(y\mid x) f(x)\right]^2 \leqslant 0$$

при любых функциях $f(x)$.

Неизменность области $\widetilde{X}$ при вариациях означает, что варьируются лишь вероятности $P(x)$, $x \in \widetilde{X}$. При таких вариациях функция K имеет нулевые производные (8.2.4). Из выпуклости функции K следует поэтому максимальность функции в экстремальной точке. Максимальность, следовательно, имеет место и для тех частных вариаций переменных $P(x)$, $x \in \widetilde{X}$, которые оставляют неизменными равенства (8.2.1), (8.2.2). Доказательство закончено.

К сожалению, при записи и решении уравнения (8.2.7), активное множество $\widetilde{X}$ заранее неизвестно, и это усложняет задачу. Для отыскания $\widetilde{X}$, вообще говоря, следует производить максимизацию $C(\widetilde{X})$ по $\widetilde{X}$. В некоторых важных случаях, например, когда для всего пространства $\widetilde{X} = X$ указанные уравнения дают распределение $P(x) > 0$, этого можно избежать. Тогда $C(X)$ есть искомая пропускная способность C.

В самом деле, вследствие доказанной в теореме 8.2 максимальности информации I_{xy} в этом случае нет надобности рассматривать меньшие активные области $\widetilde{X}_1 \subset \widetilde{X}$. Распределение $P_{01}(x)$, имеющее любую меньшую активную область $\widetilde{X}_1$, можно получить как предел распределений $P'(x)$, имеющих $\widetilde{X}$ в качестве активной области. Но для P' информация I'_{xy} не больше экстремальной I^0_{xy} по теореме 8.2. Следовательно, и предельная информация $\lim I'_{xy}$, совпадающая с информацией I^{01}_{xy} для P_{01}, не больше, чем найденная экстремальная информация I^0_{xy}. Таким образом, указанное решение является искомым решением.

Утверждение теоремы 8.2 с дискретного варианта можно распространить на произвольный случай, соответствующий уравнениям (8.2.9).

3. Для рассматриваемой вариационной задачи можно ввести термодинамические потенциалы, которые играют такую же роль, как и потенциалы в первой вариационной задаче, рассмотренной в § 3.2, 3.3, 3.6. Уже соотношение (8.2.6) аналогично известному соотношению $H = (E - F)/T$ [см., например, (4.1.14а)] термодинамики. В нем C является аналогом энтропии H, a — аналогом внутренней энергии E, $-\varphi$ — свободной энергии F; $\beta = 1/T$ — параметр, обратный температуре. В дальнейшем средние штрафы $\mathsf{M}c(x)$ как функцию термодинамических параметров мы будем обозначать буквой R. Следующая теорема подтверждает указанную аналогию.

Теорема 8.3. *Пропускную способность можно вычислять дифференцированием потенциала* φ *по температуре*:

$$C = -d\varphi(T)/dT. \tag{8.2.11}$$

Доказательство. Будем варьировать параметр $\beta = 1/T$ или, что то же самое, параметр a в условии (8.2.1). Эта вариация сопровождается вариациями параметра φ и распределения $P(x)$. Возьмем некоторую точку x активной области $\widetilde{X}$. Если вариация da не слишком велика, то после вариации сохранится условие $P(x) + dP(x) > 0$, т. е. x будет принадлежать проварьированной активной области. В точке x и до и после вариации выполняется равенство типа (8.2.4). Дифференцируя его, получаем уравнение для вариаций

$$-\sum_{y \in \widetilde{Y}} P(y \mid x) \frac{dP(y)}{P(y)} - [c(x) - \varphi]\, d\beta + \beta d\varphi = 0. \tag{8.2.12}$$

Суммирование здесь проводится по той области $\widetilde{Y}$, где $P(y) > 0$. Умножим это равенство на $P(x)$ и просуммируем по x. Учитывая (8.2.5), (8.2.1) и сохранение условия нормировки

$$\sum_y dP(y) = d\sum_y P(y) = d1 = 0,$$

получаем $(a - \varphi)d\beta = \beta\, d\varphi$. В силу (8.2.6) это дает

$$C = \beta^2 \frac{d\varphi}{d\beta}, \tag{8.2.13}$$

а, следовательно, и (8.2.11). Доказательство закончено.

Если известен потенциал $\varphi(T)$ как функция температуры T, то для определения пропускной способности по формуле (8.2.11) остается конкретизировать значение температуры T (или параметра β), соответствующее условию (8.2.1). Для этого удобно рассмотреть новую функцию

$$R = -T \frac{d\varphi}{dT} + \varphi. \tag{8.2.14}$$

Подставляя (8.2.11) в (8.2.6), имеем уравнение

$$-T\frac{d\varphi}{dT}(T)+\varphi(T)=a, \quad \text{т. е. } R=a, \tag{8.2.15}$$

служащее для определения величины T.

Формулу (8.2.14) удобно рассматривать как преобразование Лежандра функции $\varphi(T)$:

$$R(S)=TS+\varphi(T(S)) \quad \left(S=-\frac{\partial\varphi}{\partial T}\right).$$

Тогда согласно (8.2.15) пропускная способность C будет являться корнем уравнения

$$R(C)=a. \tag{8.2.16}$$

Кроме того, если рассматривать потенциал

$$\Gamma(\beta)=-\beta\varphi\left(\frac{1}{\beta}\right), \tag{8.2.17}$$

то уравнение (8.2.15) принимает вид

$$\frac{d\Gamma}{d\beta}(\beta)=-a. \tag{8.2.18}$$

Из этого уравнения находится β.

Формула (8.2.11), определяющая пропускную способность, при этом преобразуется к виду

$$C=\Gamma(\beta)-\beta\frac{d\Gamma(\beta)}{d\beta}. \tag{8.2.19}$$

Формулы (8.2.18), (8.2.19) аналогичны формулам (4.1.14б) и (3.2.9) при учете (4.1.14). Это является следствием того, что в случае второй вариационной задачи выполняются те же обычные соотношения (8.2.6), (8.2.11), что и в случае первой вариационной задачи.

Взяв дифференциал от (8.2.15), получим

$$da/T=-d\,(d\varphi/dT)=dC, \tag{8.2.20}$$

что соответствует известной термодинамической формуле $dH=dE/T$ для дифференциала энтропии.

4. Полезным является следующий результат.

Теорема 8.4. *Функция $C(a)$ является выпуклой*:

$$\frac{d^2C(a)}{da^2}\leqslant 0. \tag{8.2.21}$$

Доказательство. Пусть вариации da соответствует вариация $dP(x)$ и вариация $dP(y)=\sum_x P(y\mid x)dP(x)$. Умножим

(8.2.12) на $dP(x)$ и просуммируем по $x \in \widetilde{X}$, учитывая, что

$$\sum_x dP(x) = 0, \qquad \sum_x c(x)\, dP(x) = da$$

в силу (8.2.1), (8.2.2). Это дает

$$\sum_{y \in \widetilde{Y}} \frac{[dP(y)]^2}{P(y)} + da\, d\beta = 0. \tag{8.2.22}$$

Первый член здесь, очевидно, не может быть отрицательным. Поэтому $da\, d\beta \leqslant 0$. Поделив это неравенство на положительную величину $(da)^2$, имеем

$$d\beta/da \leqslant 0. \tag{8.2.23}$$

Искомое неравенство (8.2.21) вытекает отсюда, если учесть, что $\beta = dC/da$ в силу (8.2.20). Доказательство закончено

Знак равенства в (8.2.21), (8.2.23), как видно из (8.2.22), относится к тому случаю, когда все $dP(y)/da = 0$ в области $\widetilde{Y}$.

Типичный ход кривой $C(a)$ представлен на рис. 8.1. Точка максимума функции $C(a)$ вследствие соотношения (8.2.20), которое можно записать в форме $dC/da = \beta$, соответствует значению $\beta = 0$. При этом частном значении уравнение (8.2.7) принимает вид

$$\sum_{y \in \widetilde{Y}} P(y \mid x) \ln \frac{P(y \mid x)}{P(y)} = C. \tag{8.2.24}$$

Здесь совершенно отсутствуют $c(x)$, a. Уравнение (8.2.24) соответствует каналу $P(y|x)$, определенному без учета условий (8.1.1), (8.1.2), (8.1.6). Действительно, решение вариационной задачи без учета условия (8.1.6) приводит именно к уравнению (8.2.24), которое нужно дополнить еще условием (8.2.2). Это максимальное значение C обозначим $C_{\max}$.

Обсудим некоторые следствия теоремы 8.4. Функция $a = R(C)$ является обратной функции $C(a)$. Поэтому она определена лишь при $C \leqslant C_{\max}$ и является двузначной (хотя бы на некотором примыкающем к $C_{\max}$ интервале), если в (8.2.21) имеет место знак «$<$» при $\beta = 0$. Для одной ветви

$$\frac{dR(C)}{dC} > 0, \qquad \text{т. е. } T > 0 \quad \text{или} \quad \beta > 0;$$

эту ветвь мы назовем *нормальной*. Для другой, *аномальной* ветви имеем

$$\frac{dR(C)}{dC} < 0, \qquad \beta < 0.$$

Легко понять, что нормальная ветвь является вогнутой:

$$\frac{d^2 R(C)}{dC^2} \equiv \frac{dT}{dC} > 0 \qquad (C < C_{\max}),$$

а аномальная ветвь — выпуклой:

$$\frac{d^2 R(C)}{dC^2} \equiv \frac{dT}{dC} < 0 \qquad (C < C_{\max}),$$

как это следует из выпуклости функции $C(a)$.

Если теперь рассмотреть функцию $\varphi(T)$, являющуюся преобразованием Лежандра от $R(C)$:

$$\varphi(T) = -CT + R(C(T)) \qquad \left(T = \frac{dR}{dC}\right)$$

[(см. (8.2.15)], то, поскольку,

$$\frac{d\varphi}{dT} = -C, \qquad \frac{d^2\varphi}{dT^2} = -\frac{dC}{dT},$$

она будет выпуклой в нормальной ветви и вогнутой в аномальной.

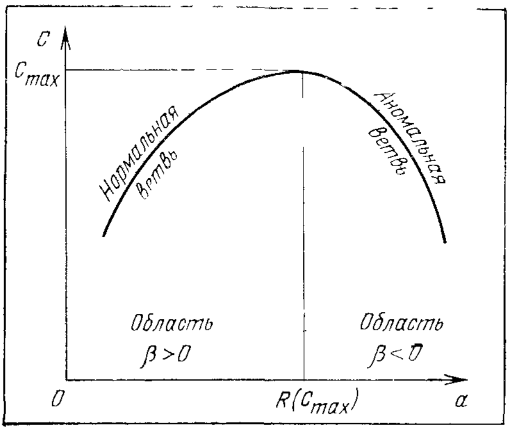


Рис. 8.1. Типичный вид зависимости между $R = a$ и C.

В заключение этого параграфа вернемся к тому случаю, когда задано условие (8.1.1) на интервале $[a_1, a_2]$. Предыдущее исследование поведения функций $C(a)$, $R(C)$ позволяет немедленно найти для него значение пропускной способности C. Могут встретиться три случая.

1. Если

$$a_1 \leqslant R(C_{\max}) \leqslant a_2,$$

то, очевидно, максимальное значение пропускной способности

$$C = C_{\max}$$

относится к числу допустимых. В этом случае фиксация условия (8.1.1) не приводит к уменьшению пропускной способности канала.

2. Если

$$a_2 < R(C_{\max}),$$

то интервал $[a_1, a_2]$ относится к нормальной ветви зависимости между C и a. Функция $C(a)$ здесь является неубывающей, и пропускная способность поэтому равна

$$C = C(a_2).$$

3. Если

$$R(C_{\max}) < a_1,$$

то нужно рассматривать аномальную ветвь. Поскольку функция $C(a)$ при $a < C_{\max}$ является невозрастающей, максимальное значение $C(a)$ достигается при $a = a_1$, т. е. $C = C(a_1)$.

8.3. Вид оптимального распределения и статистическая сумма

Рассмотрение, проводимое в этом параграфе, будет иметь меньшую степень общности, чем результаты предыдущего параграфа. Оно ограничено условием существования описываемого ниже обратного линейного преобразования L^{-1}.

Как видно из формы уравнений (6.2.8), (6.2.9), их решение легко записать, если найдено преобразование

$$f(y) = L^{-1} g(x), \tag{8.3.1}$$

обратное преобразованию

$$g(x) = Lf(y) = \int P(dy \mid x) f(y), \tag{8.3.2}$$

или в другой записи

$$Lf(y) = \sum_y P(y \mid x) f(y), \qquad Lf(y) = \int p(y \mid x) f(y)\, dy.$$

Обратное преобразование (8.3.1) записывается при помощи ядра $Q(y, dx) = q(y, x) dx$ следующим образом:

$$L^{-1} g(x) = \int Q(y, dx) g(x)$$

или

$$L^{-1} g(x) = \sum_x Q(y, x) g(x), \qquad L^{-1} g(x) = \int q(y, x) g(x)\, dx. \tag{8.3.3}$$

Для простоты остановимся на дискретном варианте. Тогда $\| Q(y, x) \| = \| P(y \mid x) \|^{-1}$ будет матрицей, обратной матрице $\| P(y \mid x),\ y \in Y,\ x \in \widetilde{X} \|$.

При помощи этой же (но транспонированной) матрицы можно записать также преобразование

$$G(x) = F(y) L^{-1} = \sum_y F(y) Q(y, x), \tag{8.3.4}$$

обратное преобразованию

$$F(y)=G(x)L=\sum_x G(x)P(y\mid x). \tag{8.3.4a}$$

Придавая уравнению (8.2.7) вид

$$\sum_y P(y\mid x)\ln P(y)=-C+\beta a-\beta c(x)-H_y(\mid x)$$

или

$$\sum_y P(y\mid x)[\ln P(y)+C-\beta a]=-\beta c(x)-H_y(\mid x),\ \ x\in\widetilde{X} \tag{8.3.5}$$

и используя обратное преобразование (8.3.3), будем иметь

$$\ln P(y)+C-\beta a=-L^{-1}[\beta c(x)+H_y(\mid x)],$$

т. е.

$$P(y)=\exp\left\{\beta a-C-\sum_{x\in\widetilde{X}}Q(y,\ x)[\beta c(x)+H_y(\mid x)]\right\}. \tag{8.3.6}$$

Левую часть можно записать в виде $\sum\limits_{x\in X}P(x)P(y\mid x)$ или $\sum\limits_{x\in X}P(x)P(y\mid x)$. Следовательно,

$$P(x)L=\exp\left\{\beta a-C-\sum_x Q(y,\ x)[\beta c(x)+H_y(\mid x)]\right\}.$$

Применяя обратное преобразование (8.3.4), находим

$$P(x)=\sum_y\exp\left\{\beta a-C-\sum_{x_1}Q(y,x_1)[\beta c(x_1)+H_y(\mid x_1)]\right\}Q(y,\ x). \tag{8.3.7}$$

Неизвестные параметры C, β определяем при помощи (8.2.1), (8.2.2). Суммируя (8.3.7) по x и учитывая, что левая часть обращается в единицу, получаем уравнение

$$\sum_y\exp\left\{-\sum_x Q(y,x)[\beta c(x)+H_y(\mid x)]\right\}=e^{C-\beta a}, \tag{8.3.8}$$

которое эквивалентно (8.2.2). (Использовано, что $\sum\limits_x Q(y,\ x)=1$, поскольку в силу (8.3.2), (8.3.3) функции $f(y)=1$ соответствует $g(x)=1$ и наоборот). Подставляя далее (8.3.7) в (8.2.1), имеем

$$\sum_y\sum_{x\in\widetilde{X}}\exp\left\{\beta a-C-\sum_x Q(y,\ x)[\beta c(x)+H_y(\mid x)]\right\}\times$$
$$\times Q(y,\ x)c(x)=a. \tag{8.3.9}$$

Равенства (8.3.8), (8.3.9) образуют систему двух уравнений относительно двух неизвестных C и β.

Решать указанные уравнения можно с использованием обычных термодинамических методов и понятий. Удобно обозначить

$$b(y) = \sum_{x \in X} Q(y,x)\, c(x),$$

$$\nu(y) = \exp\left[-\sum_x Q(y,\ x)\, H_y(|\,x)\right]. \tag{8.3.10}$$

Тогда выражение (8.3.8) примет вид обычной статистической суммы

$$Z = \mathrm{e}^{\Gamma} = \sum_y \mathrm{e}^{-\beta b(y)}\, \nu(y), \tag{8.3.11}$$

определенной ранее формулами (3.3.11), (3.6.4). Здесь $b(y)$ является аналогом энергии, а $\nu(y)$ — «кратности вырождения». Равенство (8.3.9) при этом можно записать так:

$$a = \sum_y Z^{-1} \mathrm{e}^{-\beta b(y)}\, b(y)\, \nu(y) = -\frac{d \ln Z}{d\beta}. \tag{8.3.12}$$

Если использовать (8.3.11) для определения потенциалов $\Gamma = \ln Z$, $\varphi = -\beta^{-1} \ln Z$, которые аналогичны ранее введенным потенциалам Γ, F [см. (4.1.7), (3.3.12)], то они будут не чем иным, как потенциалами, рассмотренными в § 8.2. Они удовлетворяют приведенным в § 8.2 обычным термодинамическим соотношениям, что совершенно естественно в свете формул (8.3.11), (8.3.12), имеющих вид, обычный в статистической термодинамике [см. (3.6.4), (3.3.17)]. Они соответствуют каноническому распределению

$$P(y) = \mathrm{e}^{-\Gamma - \beta b(y)}\, \nu(y) \tag{8.3.13}$$

по переменной y.

Искомое же экстремальное распределение согласно (8.3.7) имеет менее обычный вид

$$P(x) = \sum_y \mathrm{e}^{-\Gamma - \beta b(y)}\, Q(y,\ x).$$

В дополнение к теории, развитой в § 8.2, формула (8.3.11) указывает способ явного вычисления потенциалов Γ, φ.

Приведем в заключение следующий результат.

Т е о р е м а 8. 5. *Если преобразование* (8.3.2) *обратимо и если условная энтропия* $H_y(|\,x) = H_{y|x}$ *не зависит от* x, *то отношение* $C/(a - a_0)$ *заключено в интервале между* β *и* β_0. *Здесь* a_0 *и* β_0, β *определены равенствами*

$$\Gamma(\beta_0) - \beta_0 \frac{d\Gamma}{d\beta}(\beta_0) = 0, \qquad a_0 = -\frac{d\Gamma}{d\beta}(\beta_0), \quad a = -\frac{d\Gamma}{d\beta}(\beta), \tag{8.3.14}$$

т. е.

$$\beta_0 = \frac{dC}{da}(a_0), \qquad \beta = \frac{dC}{da}(a).$$

Доказательство. При $H_y(x) = H_{y|x}$ из (8.3.10), (8.3.11) вследствие равенства $\sum_x Q(y, x) = 1$ (уже упомянутого ранее) имеем

$$\Gamma(\beta) = -H_{y|x} + \Gamma_0(\beta), \tag{8.3.15}$$

где

$$\Gamma_0(\beta) = \sum_y e^{-\beta b(y)}.$$

Потенциал $\Gamma_0(\beta)$ очевидно обладает тем же свойством вогнутости, что и $\Gamma(\beta)$. Поэтому его образ по Лежандру

$$S_0(R) = \Gamma_0(\beta) + \beta R \quad \left(R = -\frac{d\Gamma_0(\beta)}{d\beta}\right) \tag{8.3.16}$$

является выпуклой функцией. Из (8.2.18), (8.2.19), если учесть (8.3.15), находим

$$C + H_{y|x} = \Gamma_0 - \beta \frac{d\Gamma_0}{d\beta}, \qquad \frac{d\Gamma_0}{d\beta} = -a \tag{8.3.17}$$

или согласно (8.3.16)

$$C + H_{y|x} = S_0(a). \tag{8.3.18}$$

Вследствие указанной выпуклости отношение $[S_0(a) - S_0(a_0)]/(a - a_0)$ лежит на интервале между $\frac{dS_0}{da}(a) = \beta$ и $\frac{dS_0}{da}(a_0) = \beta_0$. Выбирая a_0 из условия $S_0(a_0) = H_{y|x}$, получаем в силу (8.3.18), (8.2.19) утверждение теоремы. Доказательство закончено.

Укажем одно следствие из данной теоремы. Если $a \geqslant a_0$ (тогда $\beta \leqslant \beta_0$), то будем иметь

$$\beta(a - a_0) \leqslant C \leqslant \beta_0(a - a_0), \tag{8.3.19}$$

в частности

$$C \leqslant \beta_0 a, \tag{8.3.20}$$

если к тому же $\beta_0 a_0 > 0$.

8.4. Симметричные каналы

В некоторых частных случаях пропускная способность каналов может быть подсчитана более простыми методами, чем методы § 8.3. Рассмотрим сначала так называемые симметричные каналы. Изложение будем вести в дискретном варианте.

Канал $[P(y|x), c(x)]$ называется симметричным по входу, если каждая строка матрицы

$$\|P(y|x)\| = \begin{pmatrix} P(y_1|x_1) & \dots & P(y_r|x_1) \\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ P(y_1|x_{r'}) & \dots & P(y_r|x_{r'}) \end{pmatrix} \tag{8.4.1}$$

является перестановкой одних и тех же величин $p_1, \ldots, p_r$, имеющих смысл вероятностей. Очевидно, что для таких каналов выражение

$$H_y(\mid x) = -\sum_y P(y\mid x)\ln P(y\mid x) \tag{8.4.2}$$

не зависит от x и равно

$$H \equiv -\sum_j p_j \ln p_j.$$

Нетрудно понять, что (8.4.2) совпадает с $H_{y\mid x}$. Формуле (8.3.5) при этом можно придать вид

$$\sum_y P(y\mid x)\left[\ln\sum_x P(x)P(y\mid x) + C - \beta a + H_{y\mid x}\right] = \\ = -\beta c(x) \quad (x = x_1, \ldots, x_r), \tag{8.4.3}$$

или

$$\sum_y P(y\mid x)\left[\ln\sum_x P(x)P(y\mid x) + C + H_{x\mid y}\right] = 0 \quad (x = x_1, \ldots, x_r), \tag{8.4.4}$$

если $\beta = 0$ или условие, связанное с $c(x)$, отсутствует. Соотношения (8.4.3) образуют уравнения относительно неизвестных $P(x_1), \ldots, P(x_r)$. Поскольку $H_y(\mid x) = H_{y\mid x} = H$ для каналов, симметричных по входу, то количество информации $I_{xy} = H_y - H_{y\mid x}$ для таких каналов равно

$$I_{xy} = H_y - H.$$

Максимизация I_{xy} по $P(x)$ сводится поэтому к максимизации энтропии H_y по $P(x)$:

$$C = \max H_y - H.$$

Формула (8.4.4) еще более упрощается для полностью симметричных каналов. Канал называется полностью симметричным, если в матрице (8.4.1) каждый столбец также является перестановкой одних и тех же величин (скажем p'_j, $j = 1, \ldots, r'$).

Нетрудно убедиться, что равномерное распределение

$$P(x) = \text{const} = 1/r' \tag{8.4.5}$$

является решением уравнений (8.4.4) в случае полностью симметричных каналов. В самом деле, вероятности

$$P(y) = \sum_x P(x)P(y\mid x) = \frac{1}{r'}\sum_x P(y\mid x)$$

для полностью симметричных каналов не зависят от y, поскольку сумма

$$\sum_x P(y\mid x) = \sum_j p'_j$$

для всех столбцов одна и та же. Следовательно, распределение по y является равномерным

$$P(y) = \text{const} = 1/r. \tag{8.4.6}$$

Отсюда вытекает, что $H_y = \ln r$ и равенства (8.4.4) выполняются, если

$$C = \ln r - H. \tag{8.4.7}$$

Поскольку

$$r = \sum_{j=1}^{r} \frac{p_j}{p_j} = \mathbf{M} \frac{1}{p_j}, \tag{8.4.8}$$

то результат (8.4.7) можно записать так:

$$C = \ln \mathbf{M}\, 1/p_j - \mathbf{M} \ln 1/p_j. \tag{8.4.9}$$

Здесь $\mathbf{M}$ обозначает усреднение по j с весом p_j. Приведенное рассмотрение можно обобщить и на тот случай, когда переменные x и y носят непрерывный характер. При определении симметричных каналов в этом случае вместо перестановки элементов строки или столбца матрицы $P(x, y)$ нужно говорить о перестановке соответствующих интервалов. Формула (8.4.7) будет иметь вид

$$C = \ln r' + \int \ln \frac{P(dy)}{Q(dy)} P(dy) = \ln r' - H^{P/Q},$$

где

$$r' = \int \frac{Q(dy)}{P(dy)} P(dy) = \int Q(dy).$$

Вместо (8.4.9) будем иметь

$$C = \ln \mathbf{M} \frac{Q(dy)}{P(dy)} - \mathbf{M} \ln \frac{Q(dy)}{P(dy)}.$$

Усреднение здесь соответствует весу $P(dx)$; $Q(dy)$ — вспомогательная мера, на основе которой производится перестановка интервалов (меняются местами равные интервалы в шкале, при которой мера Q является равномерной).

8.5. Двоичные каналы

1. Простейшим частным случаем полностью симметричного канала является двоичный симметричный канал, имеющий два состояния на входе и на выходе, и матрицу вероятностей перехода

$$\|P(y|x)\| = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}, \qquad p < \frac{1}{2}. \tag{8.5.1}$$

Согласно (8.4.5), (8.4.6) пропускная способность такого канала осуществляется при равномерных распределениях

$$P(x) = 1/2, \qquad P(y) = 1/2 \tag{8.5.2}$$

и в силу (8.4.7) оказывается равной

$$C = \ln 2 - h_2(p) = \ln 2 + p\ln p + (1-p)\ln(1-p). \tag{8.5.3}$$

Из (8.5.1), (8.5.2) имеем

$$\| P(x, y) \| = \begin{pmatrix} (1-p)/2 & p/2 \\ p/2 & (1-p)/2 \end{pmatrix}. \tag{8.5.4}$$

Вычислим для двоичного симметричного канала термодинамический потенциал $\mu(t)$ (7.4.3), позволяющий оценить вероятность ошибки декодирования (7.4.2) в соответствии с теоремой 7.3.

Подставляя (8.5.2), (8.5.4) в (7.4.3), получаем

$$\begin{aligned} \mu(t) &= \ln\{[(1-p)^{1-t} + p^{1-t}]\, 2^{-t}\} = \\ &= \ln[p^{1-t} + (1-p)^{1-t}] - t\ln 2. \end{aligned} \tag{8.5.5}$$

Уравнение (7.4.4) согласно (8.5.5) принимает вид

$$\frac{p^{1-s}\ln p + (1-p)^{1-s}\ln(1-p)}{p^{1-s} + (1-p)^{1-s}} + \ln 2 = R, \tag{8.5.6}$$

а коэффициент α в экспоненте (7.4.2) можно записать

$$\begin{aligned} \alpha = &-sR - \ln[p^{1-s} + (1-p)^{1-s}] + s\ln 2 = \\ &-(1-s)\frac{p^{1-s}\ln p + (1-p)^{1-s}\ln(1-p)}{p^{1-s} + (1-p)^{1-s}} - \\ &-\ln[p^{1-s} + (1-p)^{1-s}] + \ln 2 - R. \end{aligned} \tag{8.5.7}$$

Формулам (8.5.7), (8.5.6) можно придать более удобную форму, вводя вместо s новую переменную

$$\rho = \frac{p^{1-s}}{p^{1-s} + (1-p)^{1-s}} \qquad \left(1 - \rho = \frac{(1-p)^{1-s}}{p^{1-s} + (1-p)^{1-s}}\right).$$

Тогда, учитывая, что

$$(1-s)\ln p = \ln p^{1-s} = \ln\rho + \ln[p^{1-s} + (1-p)^{1-s}],$$

$$(1-s)\ln(1-p) = \ln(1-p)^{1-s} = \ln(1-\rho) + \ln[p^{1-s} + (1-p)^{1-s}],$$

из (8.5.7) будем иметь

$$\alpha = \rho\ln\rho + (1-\rho)\ln(1-\rho) + \ln 2 - R = \ln 2 - h_2(\rho) - R. \tag{8.5.8}$$

Аналогичным образом упрощаем уравнение (8.5.6) и получаем уравнение

$$\rho\ln p + (1-\rho)\ln(1-p) = R - \ln 2. \tag{8.5.9}$$

Учитывая (8.5.9), (8.5.3), видим, что условие $R < C$ означает неравенство

$$\rho \ln p + (1-\rho)\ln(1-p) < p \ln p + (1-p)\ln(1-p)$$

или $(\rho - p)\ln \frac{p}{1-p} < 0$,

т. е. $\rho > p$, так как $\frac{p}{1-p} < 1$ $\left(\text{поскольку } p < \frac{1}{2}\right)$. Значению $R = 0$, как видно из (8.5.9), соответствует значение

$$\rho = \frac{\ln[4p(1-p)]}{\ln[(1-p)/p]}.$$

Когда R возрастает до C, значение ρ уменьшается до p.

Более сильные результаты получаются при помощи теоремы 7.5, на чем, однако, мы не будем останавливаться.

2. Рассмотрим теперь двоичный, но не симметричный канал $[P(y|x),\ c(x)]$, характеризуемый вероятностями перехода

$$\|P(y|x)\| = \begin{pmatrix} P(y_1|x_1) & P(y_2|x_1) \\ P(y_1|x_2) & P(y_2|x_2) \end{pmatrix} = \begin{pmatrix} 1-\alpha, & \alpha \\ \alpha', & 1-\alpha' \end{pmatrix} \quad (8.5.10)$$

и функцией штрафа

$$c(x) = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}.$$

Преобразование (8.3.2) в этом случае имеет вид

$$Lf = \begin{pmatrix} (1-\alpha) f_1 + \alpha f_2 \\ \alpha' f_1 + (1-\alpha') f_2 \end{pmatrix}.$$

Ему соответствует обратное преобразование

$$L^{-1} g = \frac{1}{D}\begin{pmatrix} 1-\alpha' & -\alpha \\ -\alpha' & 1-\alpha \end{pmatrix}\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \frac{1}{D}\begin{pmatrix} (1-\alpha) g_1 - \alpha g_l \\ -\alpha' g_1 + (1-\alpha) g_2 \end{pmatrix}.$$

$(D = 1-\alpha-\alpha')$. Далее, согласно (8.5.10),

$$H_y(|1) = h_2(\alpha),\ H_y(|2) = h_2(\alpha').$$

Поэтому

$$L^{-1}\|\beta c(x) + H_y(|x)\| =$$

$$= \frac{1}{D}\begin{pmatrix} (1-\alpha')[\beta c_1 + h_2(\alpha)] - \alpha[\beta c_2 + h_2(\alpha')] \\ -\alpha'[\beta c_1 + h_2(\alpha)] + (1-\alpha)[\beta c_2 + h_2(\alpha')] \end{pmatrix},$$

и формула (8.3.11) дает

$$Z = \exp\left\{\frac{\alpha'}{D}[\beta c_1 + h_2(\alpha)] + \frac{\alpha}{D}[\beta c_2 + h_2(\alpha')]\right\} \quad (\xi_1 + \xi_2), \tag{8.5.11}$$

где

$$\xi_1 = \exp\left\{-\frac{1}{D}[\beta c_1 + h_2(\alpha)]\right\}, \quad \xi_2 = \exp\left(-\frac{1}{D}[\beta c_2 + h_2(\alpha')]\right).$$

Используя далее (8.3.12), получаем

$$a = -\frac{\alpha' c_1 + \alpha c_2}{D} + \frac{1}{D}\frac{c_1\xi_1 + c_2\xi_2}{\xi_1 + \xi_2}. \tag{8.5.12}$$

Далее, поскольку $C = \beta a + \Gamma = \beta a + \ln Z$ [см. (8.2.6), (8.2.17)], из (8.5.11), (8.5.12) получаем

$$C = \frac{\alpha' h_2(\alpha) + \alpha h_2(\alpha')}{D} + \ln(\xi_1 + \xi_2) + \frac{\beta}{D}\frac{c_1\xi_1 + c_2\xi_2}{\xi_1 + \xi_2}. \tag{8.5.13}$$

В частности, если условие с $c(x)$ отсутствует, то, полагая $\beta = 0$, находим

$$C = \frac{\alpha' h_2(\alpha) + \alpha h_2(\alpha')}{1-\alpha-\alpha'} + \ln\left[e^{-h_2(\alpha)/(1-\alpha-\alpha')} + e^{-h_2(\alpha')/(1-\alpha-\alpha')}\right].$$

В другом частном случае, когда $\alpha = \alpha'$; $c_1 = c_2$ (симметричный канал), из (8.5.13) будем иметь

$$C = \frac{2\alpha h_2(\alpha)}{1-2\alpha} + \ln 2 - \frac{\beta c_1 + h_2(\alpha)}{1-2\alpha} + \frac{\beta c_1}{1-2\alpha} = \ln 2 - h_2(\alpha),$$

что естественно совпадает с (8.5.3).

8.6. Гауссовы каналы

1. Пусть x и y являются точками многомерных эвклидовых пространств X, Y размерности r и q соответственно.

Назовем канал $[p(y|x), c(x)]$ гауссовым, если:

1) переходная плотность вероятности гауссова

$$p(y|x) = (2\pi)^{-s/2}\det{}^{1/2}\|A_{ij}\| \times \\ \times \exp\left\{-\frac{1}{2}\sum_{ij} A_{ij}(y_i - m_i)(y_j - m_j)\right\}, \tag{8.6.1}$$

причем

2) A_{ij} не зависят от x, а средние значения m_i линейно зависят от x:

$$m_i = m_i^0 + \sum_k d_{ik} x_k, \tag{8.6.2}$$

3) функция штрафов $c(x)$ есть степенная функция от x степени не выше второй:

$$c(x) = c^{(0)} + \sum_k c_k^{(1)} x_k + \frac{1}{2}\sum_{k,l} c_{kl} x_k x_l. \tag{8.6.3}$$

Очевидно, можно так выбрать начало координат в пространствах X и Y (сделав замену $x_k + c_{kl}^{-1} c_l^{(1)} \to x_k$, $y_i - m_i^0 \to y_i$), чтобы линейные по x, y члены (8.6.3) и в экспоненте (8.6.1) обратились в нуль. Далее, постоянное слагаемое в (8.6.3) является несущественным и его можно опустить. Поэтому без ограничения общности в (8.6.1), (8.6.3) можно писать лишь билинейные члены. Применяя матричный способ записи, запишем (8.6.1), (8.6.3) в виде

$$p(y \mid x) = \det{}^{1/2} \frac{\mathrm{A}}{2\pi} \exp\left\{-\frac{1}{2}(\mathrm{y}^T - \mathrm{x}^T \mathrm{d}^T) \mathrm{A} (\mathrm{y} - \mathrm{dx})\right\}, \tag{8.6.4}$$

$$c(x) = \frac{1}{2} \mathrm{x}^T \mathrm{c} \mathrm{x}. \tag{8.6.5}$$

Здесь подразумевается матричное произведение рядом стоящих матриц. Символ T обозначает транспонирование; x, y — матрицы-столбцы, а x^T, y^T, соответственно — строки:

$$\mathrm{x}^T = (x_1, \ldots, x_r), \qquad \mathrm{y}^T = (y_1, \ldots, y_s).$$

Матрица A является, разумеется, невырожденной положительно определенной матрицей, которая обратна корреляционной матрице $\mathrm{K} = \mathrm{A}^{-1}$. Матрицу c также будем считать невырожденной и положительно определенной.

Как видно из (8.6.1), действие помех в канале сводится к прибавлению

$$\mathrm{y} = \mathrm{dx} + \mathrm{z} \tag{8.6.6}$$

шумов $\mathrm{z}^T = (z_1, \ldots, z_s)$, имеющих гауссов закон распределения с нулевым средним значением и корреляционной матрицей K:

$$\mathbf{M}\mathrm{z} = 0, \qquad \mathbf{M}\mathrm{z}\mathrm{z}^T = \mathrm{K} \tag{8.6.7}$$

(здесь снова применена матричная форма записи).

2. Перейдем к вычислению пропускной способности C и плотностей вероятности $p(x)$, $p(y)$ для рассматриваемого канала. Для этого обратимся к уравнению (8.2.9), которое, как указывалось в § 8.2, выполняется в подпространстве $\widetilde{X}$, где имеются ненулевые вероятности $p(x)\Delta x$. В нашем случае $\widetilde{X}$ будет эвклидовым подпространством исходного r-мерного эвклидова пространства X. В этом подпространстве матрица c (при помощи которой можно определить скалярное произведение), конечно, также будет невырожденной и положительно определенной.

Плотность распределения $p(x)$ будем искать в виде гауссового распределения

$$p(x) = (2\pi)^{-\tilde{r}/2} \det{}^{-1/2} K_x \exp\left[-\frac{1}{2} \mathrm{x}^T \mathrm{K}_x^{-1} \mathrm{x}\right], \quad x \in \widetilde{X} \tag{8.6.8}$$

$$(\text{так что} \quad \mathbf{M}\mathrm{x} = 0, \quad \mathbf{M}\mathrm{x}\mathrm{x}^T = \mathrm{K}_x). \tag{8.6.9}$$

Здесь $\tilde{r}$ — размерность пространства $\widetilde{X}$; K_x — неизвестная (невырожденная в $\widetilde{X}$) положительно определенная корреляционная матрица. Средние значения $\mathbf{M}\mathrm{x}$ выбраны нулевыми в соответствии с (8.6.4), (8.6.5) на основе соображений симметрии.

Из гауссового характера случайных величин x и z, если учесть (8.6.6), вытекает, что y также являются гауссовыми случайными величинами. При этом, усредняя (8.6.6) и учитывая (8.6.7), (8.6.9), нетрудно найти их среднее значение и корреляционную матрицу

$$\mathbf{M}\mathrm{y}=0, \qquad \mathbf{M}\mathrm{y}\mathrm{y}^T=\mathrm{K}_y=\mathrm{K}+\mathrm{d}\mathrm{K}_x\,\mathrm{d}^T. \tag{8.6.10}$$

Следовательно,

$$p(y)=\det{}^{-1/2}(2\pi\mathrm{K}_y)\exp\left\{-\frac{1}{2}\,\mathrm{y}^T\mathrm{K}_y^{-1}\mathrm{y}\right\}=$$
$$=\det{}^{-1/2}[2\pi(\mathrm{K}+\mathrm{d}\mathrm{K}_x\,\mathrm{d}^T)]\exp\left\{-\frac{1}{2}\,\mathrm{y}^T(\mathrm{K}+\mathrm{d}\mathrm{K}_h\,\mathrm{d}^T)^{-1}\mathrm{y}\right\}. \tag{8.6.11}$$

Подставляя (8.6.5), (8.6.11) в (8.2.9), получаем

$$-\frac{1}{2}\ln[\det\mathrm{A}\det\mathrm{K}_y]+\frac{1}{2}\mathbf{M}[(\mathrm{y}^T-\mathrm{x}^T\mathrm{d}^T)\mathrm{A}(\mathrm{y}-\mathrm{d}\mathrm{x})-$$
$$-\mathrm{y}^T\mathrm{K}_y^{-1}\mathrm{y}\mid x]=\beta a-\frac{\beta}{2}\mathrm{x}^T\mathrm{c}\mathrm{x}-C,\qquad x\in\widetilde{X}. \tag{8.6.12}$$

При взятии условного математического ожидания учтем, что

$$\mathbf{M}[(\mathrm{y}-\mathrm{d}\mathrm{x})(\mathrm{y}^T-\mathrm{x}^T\mathrm{d}^T)\mid x]=\mathbf{M}\mathrm{z}\mathrm{z}^T=\mathrm{K},\quad \mathbf{M}[\mathrm{y}\mathrm{y}^T\mid x]=$$
$$=\mathbf{M}[(\mathrm{d}\mathrm{x}+\mathrm{z})(\mathrm{x}^T\mathrm{d}^T+\mathrm{z}^T)\mid x]=\mathrm{d}\mathrm{x}\mathrm{x}^T\mathrm{d}^T+\mathbf{M}\mathrm{z}\mathrm{z}^T=\mathrm{d}\mathrm{x}\mathrm{x}^T\mathrm{d}^T+\mathrm{K}$$

в силу (8.6.6), (8.6.7). Поэтому (8.6.12) принимает вид

$$-\ln[\det\mathrm{K}_y\,\mathrm{A}]+\mathrm{Sp}\,\mathrm{K}\mathrm{A}-\mathrm{Sp}\,\mathrm{K}_y^{-1}\mathrm{K}-\mathrm{x}^T\mathrm{d}^T\mathrm{K}_y^{-1}\mathrm{d}\mathrm{x}=$$
$$=2\beta a-\beta\mathrm{x}^T\mathrm{c}\mathrm{x}-2C,\qquad x\in\widetilde{X}. \tag{8.6.13}$$

Полагая, в частности, $x=0$, имеем

$$-\ln[\det\mathrm{K}_y\,\mathrm{A}]+\mathrm{Sp}(\mathrm{A}-\mathrm{K}_y^{-1})\mathrm{K}=2\beta a-2C \tag{8.6.14}$$

и [после сравнения этого равенства с (8.6.13)]

$$\mathrm{x}^T\mathrm{d}^T\mathrm{K}_y^{-1}\mathrm{d}\mathrm{x}=\beta\mathrm{x}^T\mathrm{c}\mathrm{x},\qquad x\in\widetilde{X}. \tag{8.6.15}$$

Начиная с этого места под операторами c, $\mathrm{d}^T\mathrm{K}_y^{-1}\mathrm{d}$ и другими мы будем понимать операторы, действующие на векторы x из подпространства $\widetilde{X}$ и переводящие их в векторы из этого же подпространства (т. е. под x-операторами будем понимать соответствующие проекции исходных x-операторов). Тогда равенство (8.6.15) в силу произвольности x из $\widetilde{X}$ даст

$$\mathrm{d}^T\mathrm{K}_y^{-1}\mathrm{d}=\beta\mathrm{c}.$$

Подставляя сюда равенство

$$\mathrm{K}_y^{-1}=[\mathrm{K}(1_y+\mathrm{A}\mathrm{d}\mathrm{K}_x\,\mathrm{d}^T)]^{-1}=(1_y+\mathrm{A}\mathrm{d}\mathrm{K}_x\,\mathrm{d}^T)^{-1}\mathrm{A}, \tag{8.6.16}$$

следующее из (8.6.10) (поскольку $K = A^{-1}$), получаем

$$d^T (1_y + AdK_x d^T)^{-1} Ad = \beta c, \tag{8.6.17}$$

где 1_y — единичный y-оператор.

Используя при $A^* = d^T$, $B^* = AdK_x$ операторное тождество

$$A^* (1 + B^* A^*)^{-1} = (1 + A^* B^*)^{-1} A^* \tag{8.6.18}$$

[см. формулу (П.1.1) приложения], приводим уравнение (8.6.17) к виду

$$(1_x + \widetilde{A}K_x)^{-1} \widetilde{A} = \beta c. \tag{8.6.19}$$

Здесь

$$\widetilde{A} = d^T Ad \tag{8.6.20}$$

является x-оператором, а 1_x — единичный x-оператор.

Оператор c (как исходный r-мерный оператор, так и его $\tilde{r}$-мерная проекция) является согласно вышеуказанному невырожденным положительно определенным. Из (8.6.19) следует поэтому невырожденность оператора

$$(1_x + \widetilde{A}Kx)^{-1} \widetilde{A}.$$

Отсюда нетрудно заключить, что каждый из операторов $(1_x + \widetilde{A}K_x)^{-1}$, $\widetilde{A}$ является невырожденным. В самом деле, детерминант произведения $(1_x + \widetilde{A}K_x)^{-1} \widetilde{A}$ матриц, равный произведению детерминантов матриц-сомножителей, не мог бы быть отличным от нуля, если хотя бы один детерминант-сомножитель равнялся бы нулю. Из неравенства же $\det \widetilde{A} \neq 0$ вытекает невырожденность $\widetilde{A}$. Итак, существует обратный оператор $\widetilde{A}^{-1} = (d^T Ad)^{-1}$.

Используя этот оператор, решаем уравнение (8.6.19):

$$(1_x + \widetilde{A}K_x)^{-1} = \beta c \widetilde{A}^{-1}, \qquad 1_x + \widetilde{A}K_x = \widetilde{A}(\beta c)^{-1}. \tag{8.6.21}$$

Отсюда получаем

$$K_x = \frac{1}{\beta} c^{-1} - \widetilde{A}^{-1} = \frac{1}{\beta} c^{-1} - (d^T Ad)^{-1}. \tag{8.6.22}$$

Далее, учитывая (8.6.10), имеем

$$K_y = A^{-1} - d(d^T Ad)^{-1} d^T + \frac{1}{\beta} dc^{-1} d^T. \tag{8.6.23}$$

Тем самым распределения $p(x)$ (8.6.8), $p(y)$ (8.6.11) найдены.

3. Остается определить пропускную способность C и среднюю энергию (штрафы) $a = \mathbf{M}c(x)$. В силу (8.6.5) последняя, очевидно, равна

$$a = \frac{1}{2} \mathrm{M} x^T cx = \frac{1}{2} \operatorname{Sp} cK_x \tag{8.6.24}$$

или, если подставить (8.6.22),

$$a=\frac{1}{2}\operatorname{Sp}\left[\frac{1}{\beta}\,1_x-\mathrm{c}\tilde{\mathrm{A}}^{-1}\right]=\frac{\tilde{r}}{2}\,T-\frac{1}{2}\operatorname{Sp}\mathrm{c}\tilde{\mathrm{A}}^{-1}. \tag{8.6.25}$$

Здесь мы учли, что след единичного x-оператора равен размерности $\tilde{r}$ пространства $\tilde{X}$, т. е. «числу активных степеней свободы» случайной величины x. Соответствующая «теплоемкость» равна

$$da/dT=\tilde{r}/2 \tag{8.6.26}$$

(если $\tilde{r}$ не меняется при изменении dT). На каждую степень свободы приходится, таким образом, средняя энергия $T/2$ в соответствии с законами классической статистической термодинамики.

Чтобы определить пропускную способность можно использовать формулы (8.6.14), (8.6.25) [при использовании (8.6.16), (8.6.21)] или обычные формулы для информации связи гауссовых переменных, которые дают

$$C=\frac{1}{2}\ln\det\mathrm{K_y}\,\mathrm{A}=\frac{1}{2}\operatorname{Sp}\ln\mathrm{K_y}\,\mathrm{A}.$$

Подставим сюда (8.6.23). Легко доказать (перенося при помощи формулы (8.6.41) d слева направо), что

$$C=\frac{1}{2}\operatorname{Sp}\ln\left[1-\mathrm{d}\,(\mathrm{d}^T\mathrm{A}\mathrm{d})^{-1}\,\mathrm{d}^T\mathrm{A}+T\mathrm{d}\mathrm{c}^{-1}\,\mathrm{d}^T\mathrm{A}\right]=$$

$$=\frac{1}{2}\operatorname{Sp}\ln(T\tilde{\mathrm{A}}\mathrm{c}^{-1}),\qquad(\tilde{\mathrm{A}}=\mathrm{d}^T\mathrm{A}\mathrm{d}).$$

Иначе,

$$C=\frac{1}{2}\operatorname{Sp}\ln(\tilde{\mathrm{A}}\mathrm{c}^{-1})+\frac{\ln T}{2}\operatorname{Sp}1_x=\frac{1}{2}\operatorname{Sp}\ln(\tilde{\mathrm{A}}\mathrm{c}^{-1})+\frac{\tilde{r}}{2}\ln T. \tag{8.6.27}$$

Приведенную логарифмическую зависимость C от температуры T можно получить уже из формулы (8.6.25), если учесть общее термодинамическое соотношение (8.2.20). Оно в данном случае в силу (8.6.26) принимает вид

$$dC=da/T=\tilde{r}dT/2T.$$

Справа стоит дифференциал от $(r/2)\ln T+\text{const}$.

Найдем теперь термодинамические функции $\varphi(T)$, $C(a)$ для рассматриваемого канала. Используя (8.2.6), (8.6.25), (8.6.27), (8.2.17) имеем

$$-\varphi(T)=TC-a=\left[\frac{\tilde{r}}{2}\ln T+\frac{1}{2}\operatorname{Sp}\ln(\tilde{\mathrm{A}}\mathrm{c}^{-1})\right]T-$$

$$-\frac{\tilde{r}}{2}\,T+\frac{1}{2}\operatorname{Sp}\tilde{\mathrm{A}}^{-1}\mathrm{c}, \tag{8.6.28}$$

$$\Gamma(\beta) = \frac{1}{2} \operatorname{Sp} \ln\left(\frac{\widetilde{A}c^{-1}}{\beta e}\right) + \frac{1}{2} \operatorname{Sp} \widetilde{A}^{-1}c.$$

Далее, исключая параметр T из (8.6.25), (8.6.27), получаем

$$C = \frac{\widetilde{r}}{2} \ln\left(\frac{1}{\widetilde{r}} \operatorname{Sp} c\widetilde{A}^{-1} + \frac{2}{\widetilde{r}} a\right) + \frac{1}{2} \operatorname{Sp} \ln(\widetilde{A}c^{-1}). \qquad (8.6.29)$$

Эти функции, как указывалось в § 8.2, облегчают использование условия (8.1.1), (8.1.2).

Величина C возрастает с ростом a, зависимость между C и a в соответствии с терминологией § 8.2 является в данном случае нормальной. Если условие имеет вид (8.1.1), то пропускная способность канала получается из (8.6.29) подстановкой $a = a_2$.

4. Примеры. Простейшим является тот частный случай, когда матрица $\widetilde{A}c^{-1}$ кратна единичной:

$$\widetilde{A}c^{-1} - 1_x/2N. \qquad (8.6.30)$$

Это имеет место в том случае, когда, скажем,

$$c = 2 \cdot 1_x, \quad K = N \cdot 1_y, \quad d_{il} = \begin{cases} 1 & \text{при } l = 1, \ldots, \widetilde{r}; \\ 0 & \text{при } l > \widetilde{r}. \end{cases} \qquad (8.6.31)$$

Подставляя (8.6.30) в (8.6.29), в этом случае получаем

$$C = (\widetilde{r}/2) \ln(1 + a/\widetilde{r}N) \quad (T = 2N + 2a/\widetilde{r}), \qquad (8.6.32)$$

так как

$$\operatorname{Sp} c\widetilde{A}^{-1} - 2N \operatorname{Sp} 1_x - 2N\widetilde{r}, \quad \operatorname{Sp} \ln(\widetilde{A}c^{-1}) = -\widetilde{r} \ln(2N).$$

В соответствии с формулами (8.6.22), (8.6.25) корреляционная матрица на входе имеет вид

$$K_x = (T/2 - N) 1_x = 1_x a/\widetilde{r}.$$

Рассмотрим несколько более сложный пример. Пусть пространства X, Y совпадают друг с другом, матрицы $\frac{1}{2}c$, d являются единичными, а матрица K диагональна, но не кратна единичной:

$$K = \| N_i \delta_{ij} \|.$$

Помехи, следовательно, являются независимыми, но не одинаковыми. В этом случае

$$\widetilde{A}c^{-1} = K^{-1} c^{-1} = \| \delta_{ij} / 2N_i \|. \qquad (8.6.33)$$

Подпространство $\widetilde{X}$ есть пространство меньшей размерности. Для него отличны от нуля не все компоненты $(x_1, \ldots, x_r)$, а лишь их часть, скажем, компоненты x_i, $i \in L$. Это значит, что если i не принадлежит множеству L, то $x_i = 0$ для $(x_1, \ldots, x_r) \in \widetilde{X}$. При этих обозначениях матрицу (8.6.33) правильнее было бы записать так:

$$\delta_{ij}/2N_i, \qquad i \in L, \quad j \in L.$$

Приведем ряд других соотношений, используя введенное множество $\widetilde{X}$. Формула (8.6.22) для рассматриваемого примера принимает вид

$$(\mathrm{K}_x)_{ij} = (T/2 - N_i)\,\delta_{ij}, \qquad i, j \in L, \tag{8.6.34}$$

а равенства (8.6.27), (8.6.24) в силу (8.6.33) дают

$$\begin{gathered} C = \frac{1}{2}\sum_{i \in L} \ln \frac{T}{2N_i}, \\ a = \sum_{i \in L} (\mathrm{K}_x)_{ii} = \sum_{i \in L}\left(\frac{T}{2} - N_i\right). \end{gathered} \tag{8.6.35}$$

Приведенные формулы полностью решают задачу, если известно множество L индексов i, соответствующих ненулевым компонентам векторов x из $\widetilde{X}$. Укажем, из каких соображений оно определяется.

В случае гауссовых распределений плотности вероятности $p(x)$, $p(y)$, конечно, не могут быть отрицательными, поэтому подпространство $\widetilde{X}$ определяется не из условия положительности вероятностей (см. § 8.2). Может нарушиться, однако, условие положительной определенности матрицы K_x, которая задается формулой (8.2.22), и нужно особо проверить его выполнение. Кроме того должно выполняться условие невырожденности оператора (8.6.20). В данном примере об этом условии заботиться не приходится, поскольку $\mathrm{d} = 1$, но условие положительной определенности матрицы K_x, т. е. согласно (8.6.34) условие

$$N_i < T/2, \tag{8.6.36}$$

является весьма существенным. Для каждого фиксированного T множество индексов L определяется именно из условия (8.6.36). Поэтому в формулах (8.6.35) под знаком суммы вместо $i \in L$ можно записать $N_i < T/2$.

Полученные соотношения удобно проследить по рис. 8.2. По оси абсцисс откладывается индекс i, а по оси ординат — дисперсии помех N_i (ступенчатая линия). Фиксированной температуре $T/2$ соответствует горизонтальная линия. Под ней (между ней и ступенчатой линией) лежат отрезки $(\mathrm{K}_x)_{ii}$ в соответствии с (8.6.34). Точки пересечения указанных двух линий определяют границы множества L. Площадь, лежащая между горизонтальной и ступенчатой линией, равна суммарной полезной энергии a, а заштрихованная площадь, лежащая между ступенчатой линией и осью абсцисс — суммарной энергии помех $\sum_{i \in L} N_i$.

Аналогичными методами пространство $\tilde{X}$ (как функция от T) определяется и в более сложных случаях.

5. Вычислим для гауссовых каналов термодинамический потенциал (7.4.3), определяющий оценку (7.4.2) вероятности ошибки декодирования.

Информация

$$I(x, y) = \ln \frac{p(y \mid x)}{p(y)} = \frac{1}{2} \ln \det (\mathrm{K}_y \mathrm{A}) -$$
$$-\frac{1}{2} (\mathrm{y}^T - \mathrm{x}^T \mathrm{d}^T) \mathrm{A} (\mathrm{y} - \mathrm{dx}) + \mathrm{y}^T \mathrm{K}_y^{-1} \mathrm{y}$$

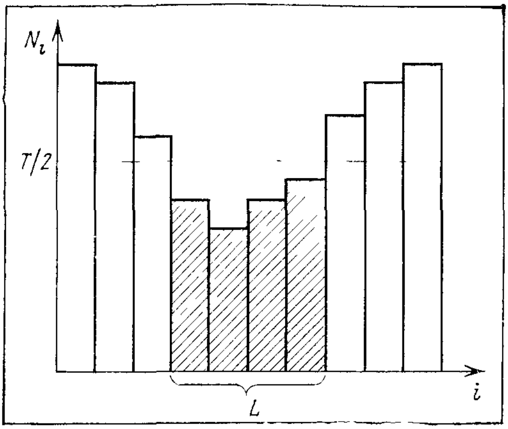


Рис. 8.2. Определение пропускной способности для случая независимых (спектральных) составляющих полезного сигнала и аддитивных помех с равнораспределенными составляющими.

уже входила в формулу (8.6.12). Подставляя ее в равенство

$$e^{\mu(s)} = \int e^{-sI(x, y)} p(x) p(y \mid x) \, dx \, dy,$$

при учете (8.6.4), (8.6.8), получаем

$$e^{\mu(s)} = (2\pi)^{-(q+\tilde{r})/2} \det{}^{-1/2} \mathrm{K}_x \det{}^{1/2} \mathrm{A} \det{}^{-s/2} \mathrm{K}_y \mathrm{A} \times$$
$$\times \int \exp\Big\{-\frac{1}{2} \mathrm{x}^T [\mathrm{K}_x^{-1} + (1-s) \mathrm{d}^T \mathrm{Ad}] \mathrm{x} +$$
$$+\frac{1-s}{2} (\mathrm{y}^T \mathrm{Adx} + \mathrm{x}^T \mathrm{d}^T \mathrm{Ay}) - \frac{1}{2} \mathrm{y}^T [(1-s) \mathrm{A} +$$
$$+ s\mathrm{K}_y^{-1}] \mathrm{y}\Big\} dx \, dy. \qquad (8.6.37)$$

Этот интеграл легко вычислить при помощи известной общей матричной формулы (5.4.19). Он, очевидно, оказывается равным

$$e^{\mu(s)} = \det{}^{-s/2} (\mathrm{K}_y \mathrm{A}) \det{}^{-1/2} (\mathrm{K}\mathrm{K}_x) \det{}^{-1/2} \mathrm{B},$$

где

$$\mathrm{B} = \begin{pmatrix} \mathrm{K}_x^{-1} + (1-s) \mathrm{d}^T \mathrm{Ad} - & (1-s) \mathrm{d}^T \mathrm{A} \\ -(1-s) \mathrm{Ad} & (1-s) \mathrm{A} + s\mathrm{K}_y^{-1} \end{pmatrix}. \qquad (8.6.38)$$

Чтобы вычислить $\det B$, воспользуемся формулой

$$\det \begin{pmatrix} b & c \\ c^T & d \end{pmatrix} = \det d \det (b - cd^{-1} c^T), \tag{8.6.39}$$

которая следует из формулы (П.2.4) приложения. Ее применение дает

$$e^{\mu(s)} = \det^{-s/2}(K_y A) \det^{-1/2}(KK_x) \det^{-1/2}[(1-s)A + sK_y^{-1}] \times$$
$$\times \det^{-1/2}[K_x^{-1} + (1-s) d^T A d - (1-s)^2 d^T A (A - sA +$$
$$+ sK_y^{-1})^{-1} A d] = \det^{-s/2}(K_y A) \det^{-1/2}[1 - s + sKK_y^{-1}] \times$$
$$\times \det^{-1/2}[1 + (1-s) K_x d^T A d - (1-s)^2 d^T A (A - sA +$$
$$+ sK_y^{-1})^{-1} A d].$$

Логарифмируя это выражение и учитывая формулу (6.5.4), находим

$$\mu(s) = -\frac{s}{2} \operatorname{Sp} \ln K_y A - \frac{1}{2} \operatorname{Sp} \ln (1 - s + sKK_y^{-1}) -$$
$$- \frac{1}{2} \operatorname{Sp} \ln [1 + (1-s) sK_x d^T K_y^{-1} (1 - s + sKK_y^{-1})^{-1} d). \tag{8.6.40}$$

Последний член мы здесь несколько упростили, приняв во внимание, что

$$d^T A d - (1-s) d^T A (A - sA + sK_y^{-1})^{-1} A d =$$
$$= d^T A \left[1 - \left(1 + \frac{s}{1-s} A^{-1} K_y^{-1} \right)^{-1} \right] d = d^T A \frac{s}{s-1} A^{-1} K_y^{-1} \times$$
$$\times \left(1 + \frac{s}{1-s} A^{-1} K_y^{-1} \right)^{-1} d.$$

Если воспользоваться формулой

$$\operatorname{Sp} f(AB) = \operatorname{Sp} f(BA) \tag{8.6.41}$$

при $f(z) = \ln(1+z)$ [см. (П.1.5), (П.1.6)], то в последнем члене в (8.6.40) можно перенести d справа налево, после чего два последних члена удобно объединить в один:

$$\operatorname{Sp} \ln (1 - s + sKK_y^{-1}) + \operatorname{Sp} \ln [1 + (1-s) s d K_x d^T K_y^{-1} (1 - s +$$
$$+ sKK_y^{-1})^{-1}] = \operatorname{Sp} \ln [1 - s + sKK_y^{-1} + s(1-s) dK_x d^T K_y^{-1}].$$

Здесь, кроме того, полезно заменить $dK_x d^T$ на $K_y - K$ [см. (8.6.10)]. Тогда равенство (8.6.40) примет вид

$$\mu(s) = -\frac{s}{2} \operatorname{Sp} \ln K_y A - \frac{1}{2} \operatorname{Sp} \ln [1 - s^2 + s^2 KK_y^{-1}]. \tag{8.6.42}$$

Поскольку $KK_y^{-1} = K_y^{-1} A$, то, как мы видим, функция $\mu(s)$ оказалась выраженной через единственную матрицу $K_y A$. Принимая во внимание (8.6.23), (8.6.22), можно преобразовать формулу

(8.6.42) таким образом, чтобы $\boldsymbol{\mu}(s)$ выражалась через единственную матрицу $c\tilde{A}^{-1}$. В самом деле, подставляя (8.6.23), имеем

$$\mu(s) = -\frac{s}{2}\operatorname{Sp}\ln(1 - d\tilde{A}^{-1}d^T A + T d c^{-1} d^T d^T A) -$$
$$-\frac{1}{2}\operatorname{Sp}\ln[(1-s)^2 + s^2(1 - d\tilde{A}^{-1}d^T A + T d c^{-1} d^T A)^{-1}].$$

Перенося $d^T A$ справа налево в соответствии с формулой (8.6.41) [во втором члене при $f(z) = \ln[1 + (1-s^2)z] - \ln(1+z)$], получаем

$$\mu(s) = -\frac{s}{2}\operatorname{Sp}\ln(T\tilde{A}c^{-1}) - \frac{1}{2}\operatorname{Sp}\ln\left[1 - s^2 + \frac{s^2}{T}c\tilde{A}^{-1}\right].$$

Этой формуле можно придать также вид

$$\mu(s) = \frac{s}{2}\operatorname{Sp}\ln(\beta c\tilde{A}^{-1}) - \frac{1}{2}\operatorname{Sp}\ln[(-s^2 + s^2\beta c\tilde{A}^{-1}] \qquad (8.6.43)$$

или

$$\mu(s) = \frac{1-s}{2}\operatorname{Sp}\ln(T\tilde{A}c^{-1}) - \frac{1}{2}\operatorname{Sp}\ln[s^2 + (1-s^2)T\tilde{A}c^{-1}]. \qquad (8.6.44)$$

Отсюда получаем, в частности, дисперсию информации

$$\mu''(0) = \operatorname{Sp}\left(1 - \frac{1}{T}c\tilde{A}^{-1}\right) = \tilde{r} - \frac{1}{T}\operatorname{Sp}c\tilde{A}^{-1}. \qquad (8.6.45)$$

Если применить (8.6.43) к частному случаю (8.6.31), то будем иметь

$$\mu(s) = -\frac{s}{2}\tilde{r}\ln\frac{T}{2N} - \frac{\tilde{r}}{2}\ln\left(1 - s^2 + \frac{2s^2N}{T}\right) =$$
$$= -\frac{s\tilde{r}}{2}\ln\left(1 + \frac{a}{\tilde{r}N}\right) - \frac{\tilde{r}}{2}\ln\left(1 - s^2\frac{a}{a + \tilde{r}N}\right).$$

Во втором же рассмотренном примере (8.6.33) имеем

$$\mu(s) = \frac{1}{2}\sum_{2N_i<T}\left[s\ln\frac{2N_i}{T} - \ln\left(1 - s^2 + 2s^2\frac{N_i}{T}\right)\right].$$

Используем (8.6.43), чтобы вычислить коэффициент (7.4.19) в формуле (7.4.2) для вероятности ошибки. Поскольку в (8.6.43) входит лишь одна матрица, производную $\boldsymbol{\mu}'(s)$ можно вычислить простым способом не принимая во внимание матричную некоммутативность. Это дает

$$\mu'(s) = -C + \operatorname{Sp}\frac{s(1 - \beta c\tilde{A}^{-1})}{1 - s^2 + s^2\beta c\tilde{A}^{-1}}.$$

Здесь учтено, что $\boldsymbol{\mu}'(0) = -C$.

В итоге согласно (7.4.2) находим, что коэффициент α в оценке

$$P_{\text{ош}} < 2e^{-\alpha n} \qquad (8.6.46)$$

находится путем решения системы уравнений

$$\alpha = s^2 \operatorname{Sp} \frac{1-\beta c\widetilde{A}^{-1}}{1-s^2+s^2\beta c\widetilde{A}^{-1}} + \frac{1}{2} \operatorname{Sp} \ln (1-s^2+s^2\beta c\widetilde{A}^{-1}),$$

$$\operatorname{Sp} \frac{s(1-\beta c\widetilde{A}^{-1})}{1-s^2+s^2\beta c\widetilde{A}} = C-R.$$

Используя тождество

$$s^2 \frac{1-\beta c\widetilde{A}^{-1}}{1-s^2+s^2\beta c\widetilde{A}^{-1}} = \frac{1}{1-s^2+s^2\beta c\widetilde{A}^{-1}} - 1,$$

эту систему можно привести к виду

$$\alpha = s(C-R) + \frac{1}{2} \operatorname{Sp} \ln (1-s^2+s^2\beta c\widetilde{A}^{-1}), \tag{8.6.47}$$

$$C-R = \operatorname{Sp}(1-s^2+s^2\beta c\widetilde{A}^{-1})^{-1} - \tilde{r}.$$

При малых отклонениях $C-R$ можно воспользоваться разложением (7.4.24). Так, удерживая лишь первый член и учитывая (8.6.45), имеем

$$\alpha = \frac{1}{2} \frac{(C-R)^2}{\tilde{r} - \beta \operatorname{Sp} c\widetilde{A}^{-1}} + \dots . \tag{8.6.48}$$

8.7. Стационарные гауссовы каналы

1. Стационарным каналам свойственна инвариантность относительно преобразования сдвига по индексу (времени):

$$(x_1, \dots, x_m) \to (x_{1+n}, \dots, x_{m+n}), \qquad (y_1, \dots, y_m) \to (y_{1+n}, \dots, y_{m+n}),$$

где n — произвольно. Пространства X, Y предполагаются имеющими одинаковую размерность. В таких каналах матрицы c_{ij}, A_{ij}, d_{ij} зависят лишь от разности индексов

$$c_{ij} = c_{i-j}, \qquad A_{ij} = A_{i-j}, \qquad d_{ij} = d_{i-j}. \tag{8.7.1}$$

Рассмотрим сначала пространства $X = Y$ конечной размерности m. Для строгой стационарности в этом случае функции c_l, a_l, d_l, входящие в (8.7.1), должны быть периодичны по l с периодом, равным m:

$$c_{l+m} = c_l, \qquad a_{l+m} = a_l, \qquad d_{l+m} = d_l. \tag{8.7.2}$$

Матрицы (8.7.1) удобно привести к диагональному виду унитарным преобразованием

$$\bar{x} = U^+ x, \tag{8.7.3}$$

где

$$U = \| U_{jl} \|, \quad U_{jl} = \frac{1}{\sqrt{m}} e^{2\pi i j l/m}, \quad j,\ l = 1, \ \dots, \ m \tag{8.7.4}$$

[см. (5.5.8)]. Его унитарность легко проверить. Эрмитово-сопряженный оператор

$$\mathrm{U}^{+}=\|U_{lk}^{+}\|=\left\|\frac{1}{\sqrt{m}}\,\mathrm{e}^{-2\pi ikl/m}\right\|, \tag{8.7.5}$$

как показано в § 5.4, действительно совпадает с обратным оператором U^{-1}. Преобразованная матрица

$$(\mathrm{U}^{+}\mathrm{cU})_{jk}=\frac{1}{m}\sum_{l'l''}\mathrm{e}^{-2\pi i\,(jl'-kl'')/m}\,c_{l'-c''}=$$

$$=\frac{1}{m}\sum_{l'}\mathrm{e}^{-2\pi i(j-k)l'/m}\sum_{l}\mathrm{e}^{-2\pi ikl/m}\,c_{l}=\delta_{jk}\,c_{k}$$

и прочие будут диагональны:

$$\mathrm{U}^{+}\mathrm{cU}=\|\bar{c}_j\,\delta_{jk}\|,\quad \mathrm{U}^{+}\mathrm{AU}=\|\bar{A}_j\,\delta_{jk}\|,$$
$$\mathrm{U}^{+}\mathrm{dU}=\|\bar{d}_j\,\delta_{jk}\|, \tag{8.7.6}$$

причем

$$\bar{c}_j=\sum_{l=1}^{m}c_l\,\mathrm{e}^{-2\pi ijl/m},\qquad \bar{A}_j=\sum_{l=1}^{m}\mathrm{e}^{-2\pi ijl/m}a_l,$$
$$\bar{d}_j=\sum_{l=1}^{m}\mathrm{e}^{-2\pi ijl/m}\,d_l. \tag{8.7.7}$$

После указанного преобразования помехи будут независимыми, и данный канал будет напоминать канал, рассмотренный в п. 4 предыдущего параграфа.

Матрица (8.6.20) будет иметь вид

$$\widetilde{\mathrm{A}}=\|\,|\bar{d}_j|^2\,\bar{A}_j\,\delta_{jk}\|=\|\,|\bar{d}_j|^2\,\delta_{jk}/N_j\|\qquad (N_j=1/\bar{A}_j).$$

Поскольку она должна быть невырожденна в пространстве $\widetilde{X}$, то пространство $\widetilde{X}$ должно строиться лишь на тех компонентах $\bar{x}_j$ ($\bar{\mathrm{x}}=\overset{+}{\mathrm{U}}\mathrm{x}$), для которых $\bar{d}_j\neq 0$. Прочие компоненты не следует рассматривать.

Матрица (8.6.22) принимает вид

$$\mathrm{K}_{xj}\,\delta_{jl}=(T/\bar{c}_j-N_j/|\bar{d}_j|^2)\,\delta_{jl}.$$

Условие ее положительной определенности еще более суживает подпространство $\widetilde{X}=\widetilde{X}\,(T)$ и оставляет лишь те компоненты, для которых $N_j\bar{c}_j<|\bar{d}_j|^2T$.

Вследствие (8.7.6) формулы (8.6.27), (8.6.25) дают

$$C=\frac{1}{2}\sum_{\bar{c}_jN_j<|\bar{d}_j|^2T}\ln\frac{|\bar{d}_j|^2\,T}{\bar{c}_j\,N_j}, \tag{8.7.8}$$

$$a=\frac{1}{2}\sum_{\bar{c}_jN_j<|\bar{d}_j|^2\,T}\left[T-\frac{\bar{c}_jN_j}{|\bar{d}_j|^2}\right]. \tag{8.7.9}$$

Не представляет труда также записать функцию (8.6.43):

$$-\mu(s)=\frac{1}{2}\sum_{\bar{c}_j N_j<|\bar{d}_j|^2 T}\left[s\ln\frac{T|\bar{d}_j|^2}{\bar{c}_j N_j}+\ln\left(1-s^2+\frac{s^2\bar{c}_j N_j}{T|\bar{d}_j|^2}\right)\right], \quad (8.7.10)$$

позволяющую оценить вероятность ошибки.

2. Приведенные формулы допускают обобщение также на случай бесконечномерного пространства X. Пусть, например, x есть случайная функция на отрезке $[0, T_0]$.

Канал, как и раньше, предполагается строго стационарным, т. е. матрицы c, A, d, которые теперь имеют вид c $(t'-t'')$, A $(t'-t'')$, d $(t'-t'')$, $(t', t''\in[0, T_0])$ и по аналогии с (8.7.2) удовлетворяют соотношениям

$$c(\tau+T_0)=c(\tau), \quad A(\tau+T_0)=A(\tau), \quad d(\tau+T_0)=d(\tau). \quad (8.7.11)$$

Рассматривая Δ-разбиение ($\Delta=T_0/m$) интервала $(0, T_0)$ точками $t_j=j\Delta$, мы можем положить

$$x_i=x(t_i), \quad c_{ij}=c(t_i-t_j), \quad A_{ij}=A(t_i-t_j), \quad d_{ij}=d(t_i-t_j)$$

и применить предыдущую теорию.

Формулы (8.7.7) теперь можно записать

$$\bar{c}_j\Delta=\sum_l e^{-2\pi ijl/m}c(l\Delta)\Delta=\sum_l e^{-2\pi ijt_l/T_0}c(t_l)\Delta$$

и аналогично для $\bar{A}_j$, $\bar{d}_j$. Выражение в правой части при $\Delta\to 0$ стремится к пределу

$$\bar{c}\left(\frac{2\pi j}{T_0}\right)=\int_0^{T_0}e^{-i2\pi jt/T_0}c(t)\,dt=\int_{-T_0/2}^{T_0/2}e^{-i2\pi jt/T_0}c(t)\,dt, \quad (8.7.12)$$

поэтому $\bar{c}_j\Delta$ приближенно равно $\bar{c}(2\pi j/T_0)$:

$$c_j\Delta\to\bar{c}(2\pi j/T_0)=\bar{c}(\omega_j) \qquad \text{при } \Delta\to 0 \quad (2\pi j/T_0=\omega_j). \quad (8.7.13)$$

Аналогично

$$\bar{A}_j\Delta\to\bar{A}(\omega_j), \qquad \bar{d}_j\Delta\to\bar{d}(\omega_j),$$

где

$$\bar{A}(\omega)=\int_0^{T_0}e^{-i\omega t}A(t)\,dt\equiv\frac{1}{N(\omega)}, \quad \bar{d}(\omega)=\int_0^{T_0}e^{-i\omega t}d(t)\,dt. \quad (8.7.14)$$

При этом

$$\frac{|\bar{d}_j|^2}{\bar{c}_j N_j}\to\frac{\bar{d}(\omega_j)}{\bar{c}(\omega_j)N(\omega_j)}. \quad (8.7.15)$$

Учитывая (8.7.15) в пределе $\Delta \to 0$ получаем

$$C = \frac{1}{2} \sum_{j=0}^{\infty} \ln \frac{T\,|\bar{d}(\omega_j)|^2}{\bar{c}(\omega_j)\,N(\omega_j)} \tag{8.7.16}$$

и аналогично

$$a = \frac{1}{2} \sum_{j=0}^{\infty} \left[T - \frac{\bar{c}(\omega_j)\,N(\omega_j)}{|\bar{d}(\omega_j)|^2} \right], \tag{8.7.17}$$

где суммирование идет по области $\bar{c}(\omega_j)\,N(\omega_j) < T\,|\bar{d}(\omega_j)|^2$. Эти формулы можно было бы получить, минуя предельный переход и непосредственно рассматривая интегральное унитарное преобразование

$$\bar{x}(\omega_j) = \frac{1}{\sqrt{T_0}} \int_0^{T_0} e^{-i\omega_j t} x(t)\,dt, \qquad j = 0, 1, 2, \dots . \tag{8.7.18}$$

Эти величины образуют компоненты Фурье исходной функции и оказываются в стационарном случае независимыми.

3. Другим случаем бесконечномерного пространства является тот случай, когда $x = (\dots, x_1, x_2, \dots)$ представляет собой процесс в дискретном времени, но на бесконечном интервале. Тогда условия периодичности (8.7.2) отпадают. Соответствующие результаты могут быть получены из формул п. 1 предельным переходом $m \to \infty$.

Преобразование (8.7.4) теперь удобно брать в форме

$$\bar{\bar{x}}(\lambda_j) = \frac{1}{\sqrt{2\pi}} \sum_l e^{-2\pi i j l/m} \quad x_l = \frac{1}{\sqrt{2\pi}} \sum_l e^{-i\lambda_j l} x_l \qquad \left(\lambda_j = \frac{2\pi j}{m}\right), \tag{8.7.19}$$

а вместо (8.7.6) писать

$$\|\bar{\bar{c}}(\lambda_r)\,\delta(\lambda_r - \lambda_k)\|, \dots ,$$

полагая $\bar{c}_j = \bar{\bar{c}}(\lambda_j)$, ... Формула (8.7.7) при этом запишется в виде

$$\bar{\bar{c}}(\lambda_j) = \sum_{l=1}^{m} e^{-i\lambda_j l} c_l, \dots \tag{8.7.20}$$

и сохранит значение при $m = \infty$.

Суммы в формулах (8.7.8)—(8.7.10) в пределе $m \to \infty$ переходят в интегралы. В самом деле, из соотношения $\lambda_j = 2\pi j/m$ следует $m(\lambda_{j+1} - \lambda_j)/2\pi = 1$. Поэтому равенство (8.7.8) можно записать

$$C = \frac{m}{4\pi} \sum_j \ln \frac{T\,|\bar{\bar{d}}(\lambda_j)|^2}{\bar{\bar{N}}(\lambda_j)\,\bar{\bar{c}}(\lambda_j)} \quad (\lambda_{j+1} - \lambda_j).$$

Поделив обе части равенства на m и перейдя к пределу, будем иметь

$$C_1 = \lim_{m\to\infty} \frac{C}{m} = \frac{1}{4\pi} \int\limits_{\overline{\overline{N}}\overline{\overline{c}}<|\overline{\overline{d}}|^2 T} \ln \frac{T|\,|\overline{\overline{d}}(\lambda)|^2}{\overline{\overline{N}}(\lambda)\,\overline{\overline{c}}(\lambda)}\, d\lambda \tag{8.7.21}$$

и аналогично

$$a_1 = \lim_{K\to\infty} \frac{a}{m} = \frac{1}{4\pi} \int\limits_{\overline{\overline{N}}\overline{\overline{c}}<|\overline{\overline{d}}|^2 T} \left[T - \frac{\overline{\overline{N}}(\lambda)\,\overline{\overline{c}}(\lambda)}{|\overline{\overline{d}}(\lambda)|^2} \right] d\lambda. \tag{8.7.22}$$

Чтобы найти C_1 (a_1), надо исключить отсюда T.

4. Рассмотрим, наконец, тот случай, когда x есть процесс в непрерывном времени и на бесконечном интервале. Формулы для такого случая можно получить из результатов п. 2 дополнительным предельным переходом $T_0 \to \infty$.

При $T_0 \to \infty$ аргументы $\omega_j = 2\pi j/T_0$ в формулах (8.7.12)—(8.7.17) все больше сгущаются, поскольку $\omega_{j+1} - \omega_j = 2\pi/T_0$. Поделив (8.7.16) на T_0, в пределе, имеем

$$C_0 = \lim_{T_0\to\infty} \frac{C}{T_0} = \lim_{T_0\to\infty} \frac{1}{4\pi} \sum \ln \frac{T\,|\bar{d}(\omega_j)|^2}{N(\omega_j)\,\bar{c}(\omega_j)} (\omega_{j+1} - \omega_j) =$$

$$= \frac{1}{4\pi} \int\limits_{\bar{c}N<T|\bar{d}|^2} \ln \frac{T|\bar{d}(\omega)|^2}{N(\omega)\,\bar{c}(\omega)}\, d\omega. \tag{8.7.23}$$

Аналогично

$$a_0 = \lim_{T_0\to\infty} \frac{a}{T_0} = \frac{1}{4\pi} \int\limits_{\bar{c}N<T|\bar{d}|^2} \left[T - \frac{\bar{c}(\omega)\,N(\omega)}{|\bar{d}(\omega)|^2} \right] d\omega.$$

Здесь в соответствии с (8.7.12), (8.7.14) $\bar{c}(\omega) = \int\limits_{-\infty}^{\infty} e^{-i\omega t} c(t)\, dt$, $\bar{d}(\omega)$ — спектры функций $c(t' - t'')$, $d(t' - t'')$, а $N(\omega)$ — спектральная плотность помех:

$$N(\omega) = \int\limits_{-\infty}^{\infty} e^{-i\omega\tau} K(\tau)\, d\tau. \tag{8.7.23a}$$

Аналогичным образом из (8.7.10) находим удельную (рассчитанную на единицу времени) функцию $\mu_0(s)$:

$$\mu_0(s) = \lim_{T_0\to\infty} \frac{\mu(s)}{T_0} = -\frac{1}{4\pi} \int\limits_{\bar{N}c<T|\bar{d}|^2} \left[s \ln \frac{T|\bar{d}(\omega)|^2}{\bar{c}(\omega)\,N(\omega)} + \right.$$

$$\left. + \ln\left(1 - s^2 + s^2\, \frac{\bar{c}(\omega)\,N(\omega)}{T|\bar{d}(\omega)|^2} \right) \right] d\omega. \tag{8.7.24}$$

Преобразование $\bar{x} = U^+ x$, диагонализирующее матрицы c, K, d, теперь имеет вид интеграла Фурье

$$\bar{x}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-i\omega t} x(t)\, dt.$$

Интегралы в полученных формулах предполагаются сходящимися.

5. Формулы, относящиеся к непрерывному и дискретному времени, имеют общую область применимости: это те случаи, когда при непрерывном времени спектры $\bar{K}_x(\omega)$, $\bar{d}(\omega)$ отличны от нуля лишь в конечной полосе частот. Как известно, при этом процесс $x(t)$, протекающий в непрерывном времени, эквивалентен последовательности значений $x(t_k)$ в дискретные моменты времени t_k (теорема Котельникова). Соответствующий интервал обозначим

$$t_{k+1} - t_k = \tau_0, \tag{8.7.25}$$

тогда $m = T_0/\tau_0$. Поскольку

$$e^{2\pi i j/m} = e^{2\pi i (j-m)/m},$$

то в (8.7.4) можно считать, что j пробегает значения не от 1 до m, а от $-m_0$ до $m - m_0 - 1$, где m_0 — любое целое число. В частности, можно взять $m_0 = m/2$ или $(m-1)/2$. Соответственно этому можно считать, что $\lambda_j = 2\pi j/m$ в формулах п. 3 пробегает значения не от 0 до 2π, а, скажем, от $-\pi$ до π.

Сравним теперь формулы $\omega_j = 2\pi j/T_0$ [см. (8.7.13)] и $\lambda_j = 2\pi j/m$ (8.7.19). Очевидно, что ω и λ связаны соотношением

$$\omega = m\lambda/T_0 = \lambda/\tau_0. \tag{8.7.26}$$

При этом интервал $-\pi < \lambda < \pi$ эквивалентен частотному интервалу

$$-\pi/\tau_0 < \omega < \pi/\tau_0. \tag{8.7.27}$$

Если спектры сосредоточены лишь в этом интервале, то, учитывая (8.7.26), видим, что формулы (8.7.21), (8.7.23) эквивалентны друг другу, поскольку $C_0 = mC_1/T_0 = C_1/\tau_0$ в силу (8.7.25), а также

$$|\bar{\bar{d}}(\lambda)|^2 / \bar{\bar{N}}(\lambda)\bar{\bar{c}}(\lambda) = |\bar{d}(\omega)|^2 / N(\omega)\bar{c}(\omega) \quad \text{при } \lambda = \omega\tau_0$$

согласно (8.7.7), (8.7.13), (8.7.20), (8.7.23а).

Указанное совпадение имеет место и для других формул. Неравенство (8.7.27) дает связь между шириной частотного интервала и временным интервалом τ_0.

6. В заключение параграфа обсудим формулу (8.6.46) при больших m или T_0. Очевидно можно ввести удельный коэффициент $\alpha_1 = \lim\limits_{m\to\infty} \alpha/m$ при дискретном времени или $\alpha_0 = \lim\limits_{T_0\to\infty} \alpha/T_0$ при непрерывном. Тогда вместо (8.6.46) будет иметь место асимптотическая оценка

$$P_{\text{ош}} < e^{-\alpha_1 m n}, \qquad P_{\text{ош}} < e^{-\alpha_0 T_0 n},$$

точнее

$$\lim_{mn\to\infty}\frac{-\ln P_{\text{ош}}}{mn}\to\alpha_1,\qquad \lim_{T_0n\to\infty}\frac{-\ln P_{\text{ош}}}{T_0n}\to\alpha_0.$$

Вероятность ошибки $P_{\text{ош}}$ будет стремиться к нулю при $m\to\infty$ или $T_0\to\infty$, даже если $n=1$. Соответствующий удельный коэффициент α_1 или α_0 будет находиться из уравнений, аналогичных, скажем, (8.6.47), если в них перейти к удельным величинам

$$C_1=\lim_{m\to\infty}C/m,\qquad R_1=\lim_{m\to\infty}R/m,\qquad C_0=\lim_{T_0\to\infty}C/T_0,$$
$$R_0=\lim_{T_0\to\infty}R/T_0.$$

Так, в случае непрерывного времени вместо (8.6.47) будем иметь

$$\alpha_0=s(C_0-R_0)+\frac{1}{4\pi}\int\limits_{\bar{c}N<T|\bar{d}|^2}\ln\left(1-s^2+s^2\frac{\bar{c}(\omega)\,\overline{N}(\omega)}{T|\bar{d}(\omega)|^2}\right)d\omega,$$

$$C_0-R_0=\frac{1}{2\pi}\int\limits_{\bar{c}N<T|\bar{d}|^2}\left\{\left[1-s^2+s^2\frac{\bar{c}(\omega)\,N(\omega)}{T|\bar{d}(\omega)|^2}\right]^{-1}-1\right\}d\omega.$$

Формула (8.6.48), очевидно, примет вид

$$\alpha_0=\pi\left\{\int\limits_{\bar{c}N<T|\bar{d}|^2}\left[1-\frac{\bar{c}(\omega)\,N(\omega)}{T|\bar{d}(\omega)|^2}\right]d\omega\right\}^{-1}(C_0-R_0)^2+\ldots.$$

Существенным является условие конечности входящего сюда интеграла.

В приведенном выше рассмотрении конечным m, T_0 соответствовал строго стационарный периодический (с периодом m или T_0) процесс. Полученные нами предельные формулы (при $m\to\infty$, $T_0\to\infty$) будут теми же самыми, если для конечных m, T_0 рассматривать отрезки процесса, стационарного лишь для бесконечных m, T_0 подобно тому, как это делалось в § 5.5, 5.7.

8.8. Аддитивные каналы

1. Пусть X и Y одинаковые линейные пространства. Канал $[p(y|x),\ c(x)]$ называется аддитивным, если $\widetilde{X}=X$ и если плотность $p(y|x)$ зависит лишь от разности $p(y|x)=p_0(y-x)$. Это значит, что $y=x+z$ получается прибавлением к x случайной величины z, имеющей плотность распределения $p_0(z)$. Предполагая, что $\widetilde{X}=X$, рассмотрим те упрощения, которые вносит предположение об аддитивности в теорию, изложенную в § 8.2, 8.3.

Уравнение (8.3.5) при этом записывается в форме

$$\int p_0(y-x)\ln p(y)\,dy=\beta a-\beta c(x)-C-H_z, \tag{8.8.1}$$

где $H_z = H_y(\,|\,x)$ теперь не зависит от x. Ему можно также придать вид

$$\int p_0(y-x)\,[\ln p(y)+C+H_z-\beta a]\,dy = -\,\beta c(x). \qquad (8.8.2)$$

Здесь применен непрерывный вариант записи: если рассматриваемое линейное пространство дискретно, то под плотностями распределения нужно понимать сумму дельта-функций, сосредоточенных в точках решетки. Указанное уравнение можно разрешить относительно функции

$$f(y) = \ln p(y) + C + H_z - \beta a,$$

используя метод преобразования Фурье. Изложим эквивалентный этому методу операторный метод.

Введем характеристическую функцию

$$e^{\mu(s)} = \int e^{sz} p_0(z)\,dz \qquad (8.8.3)$$

(вообще s — вектор).

Поскольку в силу (8.8.3)

$$e^{\mu(d/dx)} f(x) = \int p_0(z)\, e^{z(d/dx)} f(x)\,dz = \int p_0(z) f(x+z)\,dz,$$

то преобразование

$$G(x) = \int p_0(y-x) f(y)\,dy = \int p_0(z) f(x+z)\,dz,$$

очевидно, можно записать в операторной форме

$$G(x) = e^{\mu(d/dx)} f(x). \qquad (8.8.4)$$

Сравнивая (8.8.4) с (8.3.2), видим, что в данном случае оператор L имеет вид

$$L = e^{\mu(d/dx)}.$$

Отсюда находим обратный оператор

$$L^{-1} = e^{-\mu(d/dx)}. \qquad (8.8.5)$$

При помощи этого оператора задача отыскания пропускной способности и экстремального распределения решается в соответствии с формулами § 8.3.

Оператор, транспонированный по отношению к (8.8.5), таков:

$$(L^{-1})^T = e^{-\mu(-d/dx)}. \qquad (8.8.6)$$

Поэтому

$$F(y)\,L^{-1} = (L^{-1})^T F(y) = e^{-\mu(-d/dx)} F(y).$$

Формула (8.3.7) при помощи операторов (8.8.5), (8.8.6) записывается в виде

$$p(x) = e^{-\mu(-d/dx)} \exp\{\beta a - C - H_z - \beta e^{-\mu(d/dx)} c(x)\}. \qquad (8.8.7)$$

Конструируем также некоторые другие формулы § 8.3. Соотношения (8.3.10) принимают вид $b(y) = [e^{-\mu(d/dx)} c(x)]_{y=x}$, $\nu(y) = e^{-H_z}$. Поэтому из (8.3.11) получаем потенциалы

$$\varphi(T) = TH_z - T\ln\int \exp\left[-\frac{1}{T} e^{-\mu(d/dx)} c(x)\right] dx,$$
$$\Gamma(\beta) = -H_z + \ln\int \exp\left[-\beta e^{-\mu(d/dx)} c(x)\right] dx. \tag{8.8.8}$$

Отсюда в соответствии с (8.2.13), (8.2.14) имеем

$$\begin{aligned} C &= -H_z + \ln\int \exp\left[-\beta e^{-\mu(d/dx)} c(x)\right] dx + \\ &\quad + \beta\int e^{-\mu(d/dx)} c(x) \exp\left[-\beta e^{-\mu(d/dx)} c(x)\right] dx, \\ a &= \int e^{-\mu(d/dx)} c(x) \exp\left[-\beta e^{-\mu(d/dx)} c(x)\right] dx. \end{aligned} \tag{8.8.9}$$

Эти формулы в принципе дают решение задачи.

2. **Пример 1.** Пусть x — одномерное непрерывное пространство, $c(x)$ — функция четвертой степени:

$$c(x) = x^4,$$

а распределение $p_0(z)$ — гауссово:

$$p_0(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-z^2/2\sigma^2}.$$

Тогда

$$e^{\mu(s)} = e^{\sigma^2 s^2/2},$$

и, следовательно,

$$e^{-\mu\left(\frac{d}{dy}\right)} c(y) = e^{-\frac{\sigma^2}{2}\frac{d^2}{dy^2}} y^4 = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!}\frac{\sigma^{2k}}{2^k}\frac{d^2k}{dy^{2k}} y^4 =$$
$$= y^4 - 6\sigma^2 y^2 + 3\sigma^4, \tag{8.8.10}$$

так как отличны от нуля лишь производные

$$\frac{d^2}{dy^2} y^4 = 12y^2, \qquad \frac{d^4}{dy^4} y^4 = 24.$$

Формулы (8.3.11), (8.8.8) с учетом (8.8.10) дают:

$$Z = e^{-\beta\varphi} = e^{-H_z - 3\beta\sigma^4}\int_{-\infty}^{+\infty} e^{-\beta y^4 + 6\beta\sigma^2 y^2} dy = 2\beta^{-1/4} e^{-H_z - 3\beta\sigma^4} \times$$
$$\times \int_0^{\infty} e^{-t^4 + 6\sqrt{\beta}\sigma^2 t^2} dt = e^{-H_z - 3\beta\sigma^4}\left[\frac{3}{2}\sigma^2\right]^{1/2} e^{(9/2)\beta\sigma^4} \times$$
$$\times\left[K_{1/4}\left(\frac{9}{2}\beta\sigma^4\right) + \pi\sqrt{2}\, I_{1/4}\left(\frac{9}{2}\beta\sigma^4\right)\right], \tag{8.8.11}$$
$$\Gamma = -H_z + \frac{1}{2}\ln\frac{3\sigma^2}{2} + \frac{3}{2}\beta\sigma^4 + \ln\left[K_{1/4}\left(\frac{9}{2}\beta\sigma^4\right) + \right.$$

$$+\pi\sqrt{2}\,I_{1/4}\left(\frac{9}{2}\beta\sigma^4\right)\Big].$$

Учитывая, что в данном случае

$$H_z=-\int p_0(z)\ln p_0(z)\,dz=\frac{1}{2}\ln(2\pi\sigma_0^2)+\frac{1}{2}=\frac{1}{2}\ln(2\pi\sigma^2 e),$$

и вводя обозначения

$$\frac{9}{2}\beta\sigma^4=\lambda,\quad K_{1/4}(\lambda)+\pi\sqrt{2}\,I_{1/4}(\lambda)=F(\lambda),$$

при помощи формул (8.2.18), (8.2.19), (8.8.11) или при помощи (8.8.9) получаем

$$a=-\frac{3}{2}\sigma^4-\frac{9}{2}\sigma^4\frac{F'(\lambda)}{F(\lambda)},$$

$$C=\frac{1}{2}\ln\left(\frac{3}{4\pi e}\right)-\frac{\lambda F'(\lambda)}{F(\lambda)}+\ln F(\lambda).$$

Пример 2. Пусть теперь

$$c(x)=\operatorname{ch}(\alpha x),$$

а $p_0(z)$ — распределение, имеющее характеристическую функцию

$$\theta(s)=e^{\mu(s)},$$

которое необязательно гауссово, но симметрично: $\theta(s)=\theta(-s)$.

Поскольку

$$\theta\left(\frac{d}{dx}\right)e^{\pm\alpha x}=\theta(\pm\alpha)\,e^{\pm\alpha x},$$

теперь имеем

$$\frac{1}{2}\theta^{-1}\left(\frac{d}{dx}\right)(e^{\alpha x}+e^{-\alpha x})=\frac{1}{2}\theta^{-1}(\alpha)\,e^{\alpha x}+\frac{1}{2}\theta^{-1}(-\alpha)\,e^{-\alpha x}=$$
$$=\theta^{-1}(\alpha)\operatorname{ch}\alpha x$$

вследствие указанной симметрии. Вычисляем для этого случая статистический интеграл:

$$Z=e^{\Gamma}=\int_{-\infty}^{+\infty}\exp\left(-H_z-\frac{\beta}{\theta(\alpha)}\operatorname{ch}\alpha y\right)dy=\frac{2}{\alpha}K_0\left(\frac{\beta}{\theta(\alpha)}\right)e^{-H_z}.$$

Формулы (8.2.18), (8.2.19) при этом дают

$$a=K_1(\beta')/[K_0(\beta')\,\theta(\alpha)],$$

$$C=-H_z+\ln\frac{2}{\alpha}+\ln K_0(\beta')+K_1(\beta')/K_0(\beta'). \qquad (8.8.12)$$

Здесь $\beta'=\beta/\theta(\alpha)$ — новый параметр, заменяющий β.

3. Гауссовы каналы, исследованные в § 8.6, являются частными случаями аддитивных каналов, если в качестве фигурирующего

там пространства X рассматривать пространство, образованное точками $x' = dx$, или если преобразование d тождественное (будем предполагать последнее). В этом частном случае функции $c(x)$ и $\mu(s)$ являются квадратичными:

$$c(x) = \frac{1}{2} x^T cx, \qquad \mu(s) = \frac{1}{2} s^T Ks$$

(матричная запись), и преобразование $e^{-\mu\left(\frac{d}{dx}\right)} c(x)$ сводится к следующему:

$$e^{-\mu(d/dx)} c(x) = \sum_{n=0} \frac{(-1)^n}{n!\,2^n} \left(\frac{\partial^T}{\partial x} K \frac{\partial}{\partial x}\right)^n \frac{1}{2} x^T cx = \frac{1}{2} x^T cx - \frac{1}{2} \operatorname{Sp} Kc.$$

Действительно, $K \frac{d}{dx}\left(\frac{1}{2} x^T cx\right) = Kcx$, так что

$$\left(\frac{d^T}{dx} K \frac{d}{dx}\right)\left(\frac{1}{2} x^T cx\right) = \operatorname{Sp}(Kc),$$

а прочие более высокие производные обращаются в нуль. Потенциалы (8.8.8) при этом легко вычислить, пользуясь формулой (5.3.19), что дает

$$\Gamma = -H_z + \frac{\beta}{2} \operatorname{Sp}(Kc) - \frac{1}{2} \operatorname{Sp} \ln\left(\frac{\beta c}{2\pi}\right). \tag{8.8.13}$$

Поскольку

$$H_z = \frac{1}{2} \operatorname{Sp} \ln(2\pi K) + \frac{1}{2\beta} \operatorname{Sp} 1,$$

то формулы (8.2.18), (8.2.19) приводят к результату

$$\begin{aligned} a &= -\frac{1}{2} \operatorname{Sp} Kc + \frac{1}{2} \operatorname{Sp} 1, \\ C &= -\frac{1}{2} \operatorname{Sp} \ln\left(\frac{\beta c}{2\pi}\right) - \frac{1}{2} \operatorname{Sp} \ln(2\pi K) = -\frac{1}{2} \operatorname{Sp} \ln cK - \\ &\quad - \frac{1}{2} \ln\beta \operatorname{Sp} 1. \end{aligned} \tag{8.8.14}$$

Ввиду того, что $\operatorname{Sp} 1 = \tilde{r} = r$ совпадает с размерностью пространства, а матрица $K = A^{-1}$ в данном случае не отличается от $\tilde{A}^{-1}$ (так как $d = 1$), то равенства (8.8.14) согласуются с найденными ранее формулами (8.6.25), (8.6.27).

Глава 9

ОПРЕДЕЛЕНИЕ ЦЕННОСТИ ИНФОРМАЦИИ

Понятие ценности информации, вводимое в настоящей главе, связывает шенноновскую теорию информации с теорией стастистических решений. В последней теории основным является понятие средних потерь или риска, которое характеризирует качество принимаемых решений. Ценность информации специализируется как та максимальная польза, которую данное количество информации способно принести в деле уменьшения средних потерь. Такое определение ценности информации оказывается связанным с формулировкой и решением определенных условных вариационных задач.

Ввести понятие ценности информации можно тремя родственными способами, выбирая за основу хартлиевское, больцмановское или шенноновское количество информации. При выборе шенноновского количества информации нужно решать третью вариационную задачу. Между указанными определениями существует известная связь, и одно понятие может служить удобной заменой другого. Все эти понятия характеризуют определенный объект — бейесовскую систему, который наряду с каналом является важнейшим объектом исследования теории информации.

Прежде чем теория ценности информации превратилась в самостоятельный раздел теории информации, существующий и развивающийся независимо от теории передачи сообщений по каналам, некоторые ее элементы и результаты вызревали в недрах традиционной теории, исследующей каналы. К. Шенноном в 1948 г. [1] была рассмотрена третья вариационная задача, взятая в форме минимизации энтропии при заданном уровне потерь или, в шенноновской терминологии, при заданной точности воспроизведения. Использованная при этом терминология далека от терминологии статистических решений, но это, разумеется, не меняет математической сущности. Позднее (Колмогоров [1]) было введено основанное на указанной вариационной задаче понятие ε-энтропии и получен ряд относящихся к нему результатов. Вместо термина ε-энтропия мы используем здесь термин a-информация, поскольку рассматривается все-таки не энтропия, а шенноновская информация.

В настоящее время в работах американских авторов вслед за работами Шеннона [1, 3] данная теория (в первоначальной шенноновской интерпретации) получила значительное развитие (см. особенно монографию Бергера [1]). Мы придерживаемся, однако, другой интерпретации и другой терминологии.

В принятом нами способе изложения понятия и результаты теории рассматриваются независимо от понятия каналов с помехами, их пропускной способности. Мы хотим подчеркнуть, что круг вопросов, связанный с третьей вариационной задачей, равноправен кругу вопросов, относящихся ко второй или первой вариационной задаче (это, конечно, не исключает возможности объединенного рассмотрения, например, формулировки обобщенной теоремы Шеннона, см. § 11.5).

9.1. Уменьшение средних штрафов при уменьшении неопределенности

Польза, приносимая информацией, заключается в том, что она позволяет уменьшить потери, связанные со средними штрафами. Предполагается, что в условии задачи указана функция штрафов, которая по-разному штрафует различные действия и решения, принимаемые действующим лицом. За более удачные действия назначаются меньшие штрафы или большие награды, чем за менее удачные. Цель заключается в минимизации средних штрафов. Имеющаяся в распоряжении информация позволяет добиться меньшего уровня средних потерь.

Прежде чем переходить к математической формулировке сказанного, рассмотрим в этом параграфе, носящем подготовительный характер, более простую задачу (типа первой вариационной задачи), иллюстрирующую тот факт, что высокая неопределенность в системе (негинформация), действительно, повышает уровень потерь.

Пусть имеется система с дискретными возможными состояниями. В действительности осуществляется одно из возможных состояний и величина ξ, описывающая состояние, принимает одно определенное значение. Пусть в соответствии с назначением системы указана функция штрафа $c(\xi)$. Если, например, требуется, чтобы система придерживалась вблизи нулевого состояния $\xi = 0$ (задача стабилизации), то может быть взята, скажем, функция штрафа $c(\xi) = |\xi|$.

По каким-либо причинам пусть в данной задаче невозможно обеспечить идеальное равенство $\xi = 0$. Например, вследствие неизбежных флюктуаций в составных частях системы, в ней присутствует статистический разброс, т. е. имеет место неопределенность — негинформация. При этом величина ξ будет случайной и будет описываться некоторыми вероятностями $P(\xi)$. Мерой неопределенности, как известно, является энтропия

$$H_\xi = -\sum_\xi P(\xi) \ln P(\xi). \tag{9.1.1}$$

Будем предполагать, что количество неопределенности H_ξ зафиксировано, и рассмотрим, какие при этом возможны средние штрафы

$Mc(\xi)$. Существует некоторый нижний предел для этих штрафов, который может быть найден теми же методами, что и в § 3.2, 3.3, 3.6. В самом деле, там уже решалась задача на экстремум средних штрафов при условии фиксированной энтропии (первая экстремальная задача). Напомним решение этой задачи. Оптимальное распределение вероятностей имеет вид

$$P(\xi) = e^{\beta F_0 - \beta c(\xi)}, \tag{9.1.2}$$

где

$$e^{-\beta F_0} = \sum_{\xi} e^{-\beta c(\xi)} \tag{9.1.3}$$

[см. (3.3.5)]. Параметр $\beta = 1/T$ определяется из условия фиксированной энтропии (9.1.1). Согласно (3.3.15) имеем

$$-\frac{dF_0(T)}{dT} = H_\xi. \tag{9.1.4}$$

После определения параметра β или T минимальные средние штрафы определяются по формуле

$$R_0(H_\xi) = \frac{d}{d\beta}(\beta F_0) = F_0 - T\frac{\partial F_0}{\partial T}. \tag{9.1.5}$$

Указанные формулы позволяют проследить, как минимальные средние штрафы зависят от неопределенности H_ξ в системе. Согласно теореме 3.4 потери $R_0(H_\xi)$ при $T > 0$ возрастают с ростом энтропии H_ξ. Пусть теперь имеется приток информации, уменьшающий энтропию согласно (1.1.2). Если сначала в системе была негинформация H_ξ и вследствие притока информации I она уменьшилась до величины $H_\xi - I = H_{ps}$, то, очевидно, это привело к уменьшению потерь

$$\Delta R_0 = R_0(H_\xi) - R_0(H_\xi - I). \tag{9.1.6}$$

Указанная разность говорит о той пользе, которую принесла информация I. Она есть количественная мера ценности информации.

Будем предполагать $I = \Delta H_\xi$ малой величиной. Тогда из (9.1.6) будем иметь

$$\Delta R_0 \approx \frac{dR_0}{dH_\xi} I = T\,\Delta H_\xi.$$

Таким образом, производную $dR_0/dH_\xi = T$ можно считать дифференциальной ценностью уменьшения энтропии (дифференциальной ценностью информации).

Пример 1. Пусть ξ может принимать целые значения ..., —1, 0, 1, 2, ..., а штрафом служит функция $c(\xi) = |\xi|$. Обозначая $e^{-\beta} = z$, находим для этой задачи статистическую сумму

$$e^{-\beta F_0} = \sum_{\xi=-\infty}^{\infty} e^{-\beta|\xi|} = 1 + 2\sum_{\xi=1}^{\infty} z^\xi = 1 + \frac{2z}{1-z} = \frac{1+z}{1-z}.$$

Следовательно,

$$\beta F_0 = \ln(1-z) - \ln(1+z); \tag{9.1.7}$$

$$R_0 = \frac{d(\beta F_0)}{dz}\frac{dz}{d\beta} = -z\frac{d(\beta F_0)}{dz} = \frac{z}{1-z} + \frac{z}{1+z} = \frac{2z}{1-z^2}. \tag{9.1.8}$$

Принимая во внимание формулу $H_\xi = \beta R_0 - \beta F_0$, находим энтропию

$$H_\xi = -\frac{2z\ln z}{1-z^2} + \ln\frac{1+z}{1-z}. \tag{9.1.9}$$

Это уравнение позволяет определить $z(H_\xi)$ и, следовательно, параметры β и T. Используя выражения (9.1.8), (9.1.9), дающие параметрическое представление зависимости $R_0(H_\xi)$, нетрудно построить график этой зависимости. Используя (9.1.9), нетрудно найти и дифференциальную ценность

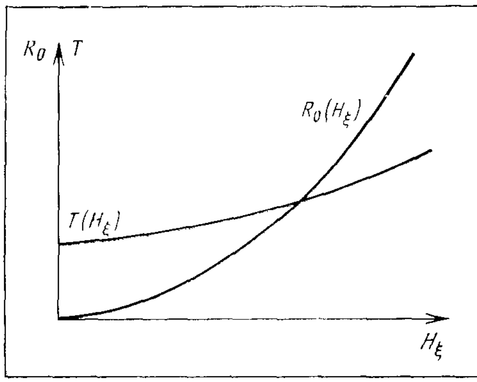


Рис. 9.1. Средние потери и дифференциальная ценность информации как функции от энтропии (пример 1).

$$T(H_\xi) = -\frac{1}{\ln z(H_\xi)}.$$

Поведение функции $R_0(H_\xi)$ и дифференциальной ценности $T(H_\xi)$ представлено на рис. 9.1.

Пример 2. Предположим, что ξ принимает одно из 8 значений $-3, -2, -1, 0, 1, 2, 3, 4$, а функция штрафа $c(\xi) = |\xi|$ такая же, что и раньше. Тогда

$$e^{-\beta F_0} = 1 + 2z + 2z^2 + 2z^3 + z^4 = (1+z)(1+z+z^2+z^3) = \\ = \frac{(1+z)(1-z^4)}{1-z}; \tag{9.1.10}$$

$$\beta F_0 = \ln\frac{1-z}{1+z} - \ln(1-z^4),$$

и поэтому

$$R_0 = -z\frac{d(\beta F_0)}{dz} = \frac{2z}{1-z^2} - \frac{4z^4}{1-z^4}. \tag{9.1.11}$$

Вместо формулы (9.1.9) теперь получаем

$$H_\xi = -2z \ln z \frac{1+z^2-2z^3}{1-z^4} + \ln \frac{1+z}{1-z} + \ln(1-z^4). \qquad (9.1.12)$$

Зависимость R_0 от H_ξ, соответствующая формулам (9.1.11), (9.1.12), представлена на рис. 9.2. Нулевой энтропии $H_\xi = 0$ соответствуют значения: $R_0 = 0$, $z = 0$, $\beta = \infty$, $T = 0$.

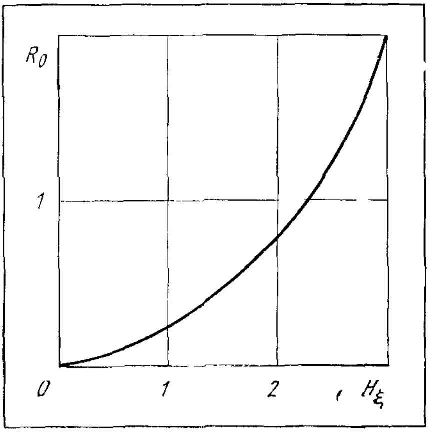


Рис. 9.2.
Средние потери для примера 2 (H_ξ в битах).

С ростом температуры T энтропия H_ξ и потери R_0 монотонно возрастают. В пределе $T \to \infty$, $\beta \to 0$, $z \to 1$ мы имеем максимально возможную энтропию

$$H_\xi = \lim_{z \to 1} \ln \frac{(1+z)(1-z^4)}{1-z} = \ln 8 = 3 \ln 2 = 3 \text{ бита}$$

и средние потери

$$R_0 = \lim_{z \to 1} \frac{2z(1+z^2-2z^3)}{1-z^4} = 2,$$

которые соответствуют равномерному распределению $P(\xi) = 1/8$.

Как было отмечено, уменьшение неопределенности в системе может быть достигнуто приобретением информации. При этом количество информации мыслилось просто как разность двух энтропий H_ξ одной переменной ξ. Между тем, согласно сказанному в гл. 6, количество информации $I = I_{xy}$ является более сложным понятием, предполагающим существование двух случайных величин x, y (а не одной ξ). Должна быть случайная величина x, о которой передается информация, и случайная величина y, которая несет эту информацию. Это заставляет усложнить приведенные в настоящем параграфе рассуждения, перейдя от простой (первой) вариационной задачи к усложненной вариационной задаче, которую будем называть третьей вариационной задачей теории информации.

9.2. Ценность хартлиевского количества информации. Пример

1. Рассмотрим следующий пример, близкий к примерам предыдущего параграфа. Пусть x есть внутренняя координата системы, принимающая значения -3, -2, -1, 0, 1, 2, 3, 4, а u является оценочной переменной, выбираемой из тех же значений, подобно переменной ξ примера 2 предыдущего параграфа.

Будем считать, что указанные восемь точек расположены по кругу (рис. 9.3). Желательно, чтобы оценочная переменная u

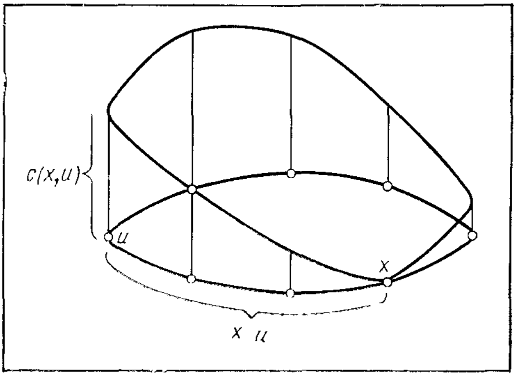


Рис. 9.3. Вид функции штрафов для рассматриваемого примера.

была как можно ближе к внутренней переменной x. Это описывается, например, введением функции штрафа

$$c(x, u) = \min [|x - u|,\ 8 - |x - u|]. \tag{9.2.1}$$

Последняя изображена на рис. 9.3 и обладает симметрией относительно отражения и поворота системы восьми точек на угол $\pi/4$.

Априори переменную x будем считать равномерно распределенной: $P(x) = {}^1/_8$. Если мы не получаем никаких сведений о значении x, то мы в качестве u можем взять наугад любое значение, скажем 3. Нетрудно подсчитать, что при этом будет иметь место средний штраф

$$\mathbf{M}c(x, u) = 0 \cdot \frac{1}{8} + 1 \cdot 2 \cdot \frac{1}{8} + 2 \cdot 2 \cdot \frac{1}{8} + 3 \cdot 2 \cdot \frac{1}{8} + 4 \cdot \frac{1}{8} = 2.$$

Уменьшить средний штраф можно только путем предварительного уточнения переменной x, то есть путем приобретения информации. Пусть, скажем, количество информации ограничено 1 битом. Мы узнаем, что осуществляется какая-то одна возможность из двух, например x, принадлежит или не принадлежит некоторому подмножеству E_1 исходного множества E восьми точек. Обозначим это сообщение $y = y(x)$. Приняв сообщение, что $x \in E_1$, мы, естественно, выберем u из E_1. Аналогично выберем u из остального множества $E - E_1 = E_0$, если узнаем, что x не принадлежит E_1. Такой выбор можно сделать оптимальным образом, выбрав то значение u, которое минимизирует выражение

$$\mathbf{M}[c(x, u) \mid y] = \sum_{x \in E_y} c(x, u) P(x) \Big/ \sum_{x \in E_y} P(x).$$

Способ разбиения восьми точек на две части также можно выбрать оптимальным, т. е. таким, который дает наименьшие средние потери

$$\mathbf{M} \min_u \mathbf{M}[c(x, u) \mid y].$$

Как нетрудно проверить, наилучшим способом разбиения является разбиение круга из восьми точек на два одинаковых полукруга

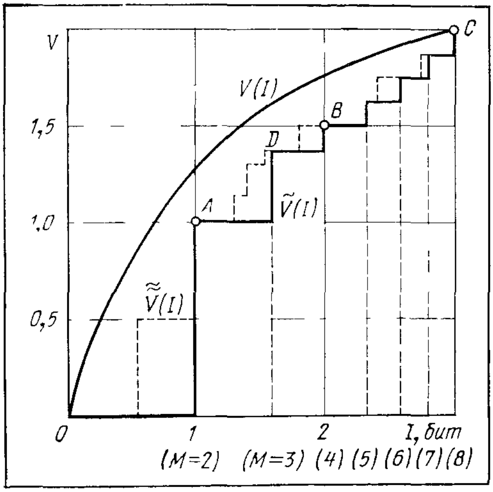


Рис. 9.4. Значения ценности информации для одного примера.

по четыре точки в каждом. Внутри каждого полукруга целесообразно выбрать любую из двух внутренних точек. Такой выбор приводит к средним штрафам

$$\mathbf{M}c(x, u) = 0 \cdot \frac{1}{4} + 2 \cdot 1 \cdot \frac{1}{4} + 1 \cdot 2 \cdot \frac{1}{4} = 1.$$

Таким образом, в оптимальном случае 1 бит информации приносит ту пользу, что уменьшает средние потери с двух до единицы. Следовательно, ценность 1 бита информации равна единице: $\tilde{V}(1) = 1$, если описанную ценность обозначить через $\tilde{V}$.

Аналогичное рассмотрение можно провести и для 2 битов информации. Теперь уже множество E разбивается на четыре части, лучше всего на четыре пары соседних точек. Узнав, в какой паре находится x, выбираем в качестве u любую точку из этой пары. С вероятностью $^1/_2$ она совпадает с x, и с вероятностью $^1/_2$ она отличается от x на единицу. Средний штраф оказывается равным $^1/_2$. Следовательно, информация уменьшила потери с 2 до $^1/_2$. Ценность 2 битов информации оказалась равной $^3/_2$: $\widetilde{V}(2) = {^3/_2}$.

Приняв информацию в 3 бита, мы можем точно узнать значение x и положить $u = x$, после чего штраф обратится в нуль. Следовательно, ценность битов информации равна в данной модели двум: $\widetilde{V}(3) = 2$. Указанные выше значения ценности соответствуют точкам A, B, C на рис. 9.4.

Если $M = e^I = 2^I_{\text{бит}}$ равняется трем, то следует производить разбиение восьми точек на три области: E_1, E_2, E_3. Для каждого из таких разбиений нужно подсчитать выгоду $2 - \mathrm{M} \min \mathrm{M}\, [c(x, u) \mid \mid E_k]$ и выбрать оптимальное разбиение. Подсчет показывает, что оптимальным является разбиение, в котором E_1 состоит из двух точек, а E_2 и E_3 — из трех. Ему соответствует выгода 1,375. Полученная точка ($\log 3$; 1,375) откладывается на плоскости ($I_{\text{бит}}$, V) (точка D на рис. 9.4). Аналогично строятся и точки, соответствующие $M = 5$; 6; 7. Перечислим оптимальные разбиения и соответствующие им координаты:

Состав разбиений	(4,4)	(2,3,3)	(2,2,2,2)	(1,1,2,2,2) (1,1,1,2,3)	(1,1,1,1,2,2) (1,1,1,1,1,3)	(1,1,1,1,1,1,2)
$I_{\text{бит}}$	1	1,59	2	2,32	2,59	2,81
$\widetilde{V}(I)$	1	1,375	1,5	1,625	1,75	1,875

Значениям $M = 5$ и 6 соответствуют два равнокачественных оптимальных разбиения. Найденные точки A, B, C, D, ..., соответствующие оптимальным разбиениям, соединяются ступенчатой линией $\widetilde{V}(I)$ (рис. 9.4), изображающей ценность хартлиевской информации.

Из сказанного ясно, что ценность информации $\widetilde{V}(I)$ определяется как максимально возможное уменьшение средних штрафов, которого можно добиться, получив данное количество информации, касающейся принадлежности величины x некоторой области из ограниченного числа $M(I)$ областей [$M(I)$ — максимальное целое число, содержащееся в e^I].

2. Описанный выше способ определения ценности информации можно распространить на общий случай. Предполагается, что задана случайная величина x и функция штрафов $c(x, u)$ как функция от двух переменных: от значения x и от оценки u.

Лицо, выбирающее оценку или управление u, получает предварительно некоторую информацию о значении x. Принимаемое им сообщение ограничено M значениями, где M — целое число, другими словами, оно должно иметь вид одного из M высказываний. Таким образом, количество получаемой информации является ограниченным, оно равно хартлиевскому количеству информации $I = \ln M$ (см. § 1.1).

Обозначим через $y = y(x)$ величину, наблюдаемую действующим лицом. Согласно сказанному выше она принимает одно из M значений. После наблюдения действующее лицо, естественно, выбирает оптимальную оценку $u = d(y)$, минимизирующую условные средние штрафы. Она определяется равенством

$$\mathbf{M}[c(x, d(y)) \mid y] = \min_u \mathbf{M}[c(x, u) \mid y]. \tag{9.2.2}$$

Каждому значению $y = y_k$ можно сопоставить тем самым одно оптимальное значение $u = u_k$. Поэтому функция $u = d(y(x))$ подобно $y(x)$ будет функцией не более чем с M значениями.

Усредняя условные средние штрафы (9.2.2), получаем полные средние штрафы

$$R = \mathbf{M} \min_u \mathbf{M}[c(x, u) \mid y]. \tag{9.2.3}$$

Если никакой информации о значении неизвестной случайной величины x не поступает, то при выборе оптимальной оценки u не остается ничего другого, как минимизировать по u средние штрафы $\mathbf{M}c(x, u) = \sum_x c(x, u) P(x)$, что дает уровень потерь

$$R_0 = \min_u \sum_x c(x, u) P(x) = \min_u \mathbf{M}c(x, u). \tag{9.2.4}$$

Здесь $\mathbf{M}c(x, u) = \mathbf{M}[c(x, u) \mid u]$, т. е. по u усреднение не производится. Пользу, приносимую информацией, естественно связать с разностью потерь (9.2.3) и (9.2.4).

Определим ценность информации $I = \ln M$ как максимальную выгоду, которую можно получить от хартлиевского количества информации $I = \ln M$:

$$\widetilde{V}(\ln M) = \min_u \mathbf{M}c(x, u) - \inf_{y(x)} \mathbf{M} \min_u \mathbf{M}[c(x, u) \mid y]. \tag{9.2.5}$$

Здесь помимо минимизации по u производится минимизация по всевозможным функциям $y(x)$ с M значениями.

Т е о р е м а 9. 1. *При минимизации* (9.2.5) *по всевозможным функциям* $y(x)$ с M *значениями можно ограничиться лишь нерандомизированными зависимостями* y от x, *т. е. включение в рассмотрение случайных зависимостей* $y(x)$ (когда y случайно при фиксированном x) *не изменяет экстремума.*

Для нерандомизированных зависимостей наблюдение конкретного значения $y = y_k$ и M значений $y_1, \ldots, y_M$ эквивалентно указанию области E_k (из M взаимно исключающих областей $E_1, \ldots, E_M$), в которой находится x. Минимизация по $y(x)$ в (9.2.5) сведется тогда к миннмизации

$$\widetilde{V}(I) = \min_u \mathbf{M}c(x, u) - \inf_{\sum_k E_k = X} \mathbf{M} \min_u \mathbf{M}[c(x, u) \mid E_k], \tag{9.2.6}$$

по всевозможным разбиениям $E_1 + \ldots + E_M$ пространства X значений x на области $E_1, \ldots, E_M$ при $M \leqslant e^I$. Подробнее (9.2.6) записывается так:

$$\widetilde{V}(I) = \min_u \int c(x, u) P(dx) - \inf_{\sum_k E_k = X} \sum_k P(E_k) \min_u \int c(x, u) P(dx|E_k). \tag{9.2.6a}$$

Доказательство теоремы 9.1. Пусть $y = y_r(x)$ зависит от x рандомизированным образом и пробегает значения $y_1, \ldots, y_M$. Это значит, что при фиксированном x величина y является случайной и описывается некоторым распределением вероятностей $P_r(y \mid x)$. Пусть $d_r(y)$ — оптимальное решающее правило для указанной зависимости, определяемое из (9.2.2). Запишем для него потери (9.2.3)

$$R_r = \sum_x \sum_y c(x, d_r(y)) P(x) P_r(y \mid x). \tag{9.2.7}$$

Построим новую, уже не рандомизированную зависимость $y = y_n(x)$ следующим образом. Пусть при фиксированном x значение $y_n(x)$ есть то значение y (из $y_1, \ldots, y_M$), которое минимизирует функцию $c(x, d_r(y))$. Тогда

$$c(x, d_r(y)) \geqslant c(x, d_r(y_n(x))), \quad (y = y_1, \ldots, y_M),$$

и в силу (9.2.7)

$$\begin{aligned} R_r &\geqslant \sum_x \sum_y c(x, d_r(y_n(x))) P(x) P_n(y \mid x) = \\ &= \sum_x c(x, d_r(y_n(x))) P(x). \end{aligned} \tag{9.2.8}$$

Сумму, стоящую справа, можно записать

$$\begin{aligned} \sum_x c(x, d_r(y_n(x))) P(x) &= \sum_{x, y_n} c(x, d_r(y_n)) P(x) P_n(y_n \mid x) = \\ &= \sum_{y_n} \sum_x c(x, d_r(y_n)) P_n(x \mid y_n) P_n(y_n), \end{aligned}$$

где индекс n отмечает нерандомизированный случай. Очевидно, что

$$\sum_x c(x, d_r(y_n)) P_n(x \mid y_n) \geqslant \min_u \sum_x c(x, u) P_n(x \mid y_n).$$

Поэтому

$$\begin{aligned} &\sum_{y_n} P_n(y_n) \sum_x c(x, d_r(y_n)) P_n(x \mid y_n) \geqslant \\ &\qquad \geqslant \sum_{y_n} P_n(y_n) \min_u \sum_x c(x, u) P_n(x \mid y_n). \end{aligned} \tag{9.2.9}$$

Выражение, стоящее справа, обозначаем через R_n. Сопоставление неравенств (9.2.8), (9.2.9) доказывает, что

$$R_r \geqslant R_n \equiv \sum_{y_n} P_n(y_n) \min_u \sum_x c(x, u) P_n(x \mid y_n),$$

т. е. нерандомизированная зависимость $y_n(x)$ не хуже с точки зрения величины средних штрафов рандомизированной зависимости $y_r(x)$. Доказательство закончено.

3. Из приведенного выше определения немедленно следует одно простое применение понятия ценности хартлиевского количества информации, а именно применение к конструированию измерительно-передающей системы, содержащей информационное ограничение.

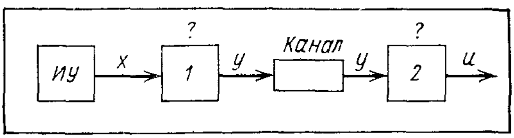


Рис. 9.5.
Система передачи максимально ценной информации. Канал без помех.
ИУ — измерительное устройство.

Пусть имеется измерительное устройство (ИУ) (рис. 9.5), выходной сигнал x которого совпадает со значением измеряемой величины, скажем непрерывной. Точное значение x, однако, не может быть сообщено потребителю, так как на пути стоит канал без помех с ограниченной пропускной способностью или записывающее устройство ограниченной информационной емкости, так что значение y может принимать лишь одно из $M = [e^I]$ значений. Целью является получение значений u, «наиболее близких» к x в смысле функции штрафов $c(x, u)$. Требуется сконструировать такие блоки *1* и *2*, помеченные на рис. 9.5 знаком вопроса, которые давали бы минимальные средние штрафы. Поскольку общее количество информации ограничено, через канал нужно передавать наиболее ценную информацию.

Принимая во внимание определение ценности хартлиевского количества информации, легко понять, как решается данная задача. Блок *1* должен осуществлять разбиение пространства значений x на оптимальные области $E_1, \ldots, E_M$, которые фигурируют в (9.2.6), и выдавать номер области, т. е. $y = k$. Блок *2* должен после получения одного из возможных сигналов k выдавать значение u_k, при котором обращается в минимум условное математическое ожидание $\mathbf{M}[c(x, u) \mid E_k]$.

Сказанное в основном справедливо и в том случае, когда в канале имеются помехи, т. е. когда его выходной сигнал y не обязан совпадать с входным y'. Этот случай асимптотически сводится к предыдущему, если систему заставлять работать многократно, заменив x на $\xi = (x_1, \ldots, x_n)$, u — на $\zeta = (u_1, \ldots, u_n)$ (n достаточно больше), и воспользоваться теоремой Шеннона об асимптотической безошибочности декодирования. При этом, как показано на рис. 9.6, потребуется поставить на входе канала кодировщик канала (КК), а на выходе — декодировщик канала (ДК), работа которых определяется по теории оптимального кодирования и декодирования (гл. 7). Структура же блоков *1* и *2* такая же, что и на рис. 9.5.

Рассмотренный выше способ определения ценности информации $\widetilde{V}(I)$ имеет ряд недостатков. Во-первых, он определяет ценность информации лишь для целого числа $M = e^I$. Остается неясным вопрос, какую пользу приносит дробная часть величины e^I. Во-вторых, если в рассмотренном примере менять число точек, беря, скажем, 9, 10 и т. д. точек, то ценность информации будет

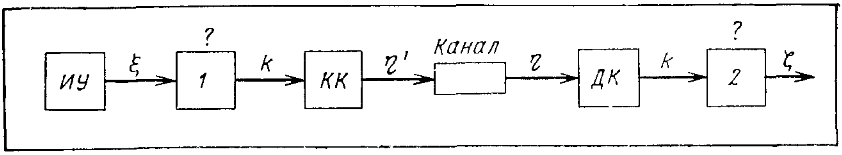


Рис. 9.6.
Система передачи максимально ценной информации. Канал с помехами.
КК — кодировщик канала; ДК — декодировщик канала; $\eta = (y_1, \ldots, y_n)$.

испытывать нерегулярные скачки. В-третьих, хартлиевское количество информации $I = \ln M$, являющееся аргументом функции $\widetilde{V}(I)$, не имеет характера разности двух энтропий, как это было, скажем, в § 9.1. Поэтому приведенное здесь определение плохо согласуется с идеями § 9.1. Согласование будет лучшим, если в качестве аргумента взять шенновское количество информации I_{xy}, поскольку оно равно разности $H_x - H_{x|y}$ энтропий до сообщения и после сообщения.

Определение ценности информации, даваемое ниже, имеет черты обоих приведенных выше определений (§ 9.1 и 9.2).

9.3. Определение ценности шенноновского количества информации и a-информации

Как и в п. 2 предыдущего параграфа, будем считать, что задана случайная величина x, описываемая распределением $P(dx)$ и (измеримая) функция штрафов $c(x, u)$ от x и от оценки u. Значения x и u являются точками заданных измеримых пространств X и U соответственно. Тем самым задана система $[P(dx), c(x, u)]$, которую будем называть *бейесовской системой*.

Для любого условного распределения $P(du \mid x)$ можно вычислить *средние штрафы* или *риск*

$$\mathbf{M}c(x, u) = \int c(x, u)\, P(du \mid x)\, P(dx) \tag{9.3.1}$$

и шенноновское количество информации

$$I_{xu} \equiv \int \ln \left[\frac{P(du \mid x)}{\int\limits_X P(du \mid x) P(dx)}\right] P(du \mid x) P(dx). \tag{9.3.2}$$

Сформулируем следующую *третью экстремальную задачу*. Назовем условное распределение $P(du \mid x)$ *экстремальным*, если оно обращает в экстремум средние штрафы (9.3.1) при фиксированном значении количества информации (9.3.2):

$$I_{xu} = I, \tag{9.3.3}$$

где I — независимо задаваемое число. Как показывает анализ, это же распределение обращает в экстремум, а именно в минимум, информацию I_{xu} при фиксированных средних штрафах:

$$\begin{gathered} I_{xu} = \min, \\ \int c(x, u) P(du \mid x) P(dx) = a = \text{fix}. \end{gathered} \tag{9.3.4}$$

Средние штрафы (риск) (9.3.1) экстремального распределения будем обозначать буквой R. Вследствие условия (9.3.3) они являются функцией от I:

$$R(I) = \int c(x, u) P(du \mid x) P(dx). \tag{9.3.5}$$

Наряду с $R(I)$ можно рассматривать и обратную зависимость $I_{xu}(a)$. Значение $I(a)$ называем информацией, соответствующей уровню потерь $R = a$ или, коротко, *a-информацией*. Как видно из последующего (теорема 9.6), функция $I_{xu}(a)$ является вогнутой (рис. 9.7). Поэтому функция $R(I)$ является, вообще говоря, двузначной. В общем случае функция $R(I)$ принимает минимальное значение, равное нулю, на некотором интервале

$$R_0 \leqslant R \leqslant R_0'. \tag{9.3.6}$$

Функцию $R(I) = R_+(I)$, обратную функции $I(R)$, $R \leqslant R_0$, назовем *нормальной ветвью*, а функцию $R(I) = R_-(I)$, обратную к $I(R)$, $R \geqslant R_0'$ — *аномальной ветвью*. Для нормальной ветви определяем *ценность шенноновской информации*

$$V(I) = R_0 - R_+(I). \tag{9.3.7}$$

Для аномальной ветви ценность информации определяем формулой

$$V(I) = R_-(I) - R_0'. \tag{9.3.8}$$

При таком определении ценность информации всегда является неотрицательной. В конкретных случаях та или иная ветвь может уйти в бесконечность, т. е. отсутствовать.

Чтобы пояснить смысл определений (9.3.7), (9.3.8), рассмотрим сначала, что представляет собой R_0, R_0'. Диапазон (9.3.6) соответствует нулевой информации связи $I_{xu}=0$. Это значит, что для него в (9.3.5) распределение $P(du\,|\,x)$ не зависит от x, так что

$$R(0)=\int P(du)\,\mathbf{M}[c(x,u)\,|\,u],$$

где $\mathbf{M}[c(x,u)\,|\,u]=\int c(x,u)\,P(dx)$. Отсюда видно, что, меняя $P(du)$, мы можем получить диапазон изменения $R(0)$:

$$\min_u \mathbf{M}[c(x,u)\,|\,u]\leqslant R(0)\leqslant \max_u \mathbf{M}[c(x,u)\,|\,u].$$

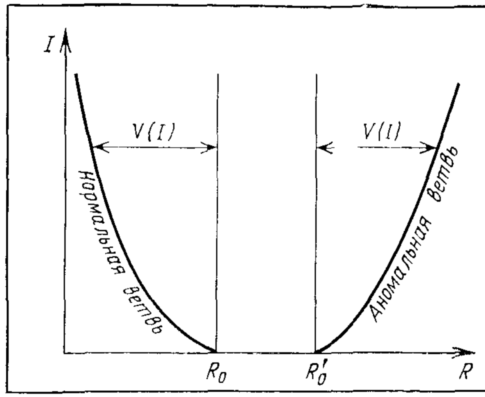


Рис. 9.7. Типичный ход зависимости $V(I)$.

Сравнивая его с (9.3.6), получаем

$$R_0=\min_u \mathbf{M}[c(x,u)\,|\,u];\quad R_0'=\max_u \mathbf{M}[c(x,u)\,|\,u]. \tag{9.3.9}$$

Далее, как будет видно из теоремы 9.3 (см. § 9.4. п. 2), нормальная ветвь соответствует минимальным штрафам

$$R_+(I)=\min_{P(du\,|\,x)} \mathbf{M}\,c(x,u), \tag{9.3.10}$$

а аномальная — максимальным

$$R_-(I)=\max_{P(du\,|\,x)} \mathbf{M}\,c(x,u) \tag{9.3.11}$$

при выполнении условия (9.3.3). Поэтому формулу (9.3.7) можно записать так:

$$V(I)=\min_u \mathbf{M}[c(x,u)\,|\,u]-\min_{P(du\,|\,x)\in G} \mathbf{M}\,c(x,u), \tag{9.3.12}$$

где множество G определяется условием

$$I_{xu}\leqslant I. \tag{9.3.13}$$

Сравнивая (9.3.13) с (9.2.5), замечаем аналогию этих двух определений ценности информации. В обоих случаях $V(I)$ имеет смысл максимально возможного (при условии фиксированного количества I) уменьшения средних штрафов.

Формула же (9.3.8) в силу (9.3.9), (9.3.11) принимает вид

$$V(I)=\max_{P(du\mid x)}\mathbf{M}c(x,u)-\max_{u}\mathbf{M}[c(x,u)\mid u]. \tag{9.3.14}$$

При этом функцию $c(x, u)$ следует интерпретировать не как штрафы, а как поощрения (выигрыши). Ценность информации имеет смысл максимально возможного среднего выигрыша, даваемого заданным количеством информации I. Разумеется, нетрудно записать и соответствующий этому случаю вариант определения ценности хартлиевской информации. Вместо формулы (9.2.5) будем иметь

$$\tilde{V}(\ln M)=\sup_{y(x)}\max_{u}\mathbf{M}[c(x,u)\mid y]-\max_{u}\mathbf{M}[c(x,u)\mid u]. \tag{9.3.15}$$

Таким образом, теории ценности информации свойственна естественная симметрия относительно замены операций минимизации и максимизации, что соответствует замене знака у функции $c(x, u)$. Коснемся вопроса о диапазоне изменения функции $R(I)$ или, что то же, функции ценности $V(I)$. Нулевому количеству информации, как уже отмечалось, соответствует интервал (9.3.6). Если количество информации, напротив, устремить к максимальному (конечному или бесконечному) значению, то будем иметь совершенно точное значение величины x. Оптимальную оценку при этом можно получить непосредственно минимизацией штрафов $c(x, u)$. При усреднении получим нижнее предельное значение

$$R_{\text{н}}=\sum_{x}P(x)\min_{u}c(x,u).$$

Аналогично для аномальной ветви максимизацией штрафов $c(x, u)$ и усреднением получаем верхнее предельное значение

$$R_{\text{в}}=\sum_{x}P(x)\max_{u}c(x,u).$$

Таким образом, функция $R(I)$ лежит в диапазоне

$$R_{\text{н}}<R(I)<R_{\text{в}}. \tag{9.3.16}$$

Соответственно этому функция ценности лежит в диапазоне $0<V(I)<R_0-R_{\text{н}}$, (нормальная ветвь) и в диапазоне $0<V(I)<R_{\text{в}}-R_0'$ (аномальная ветвь).

Непосредственно из определения (9.3.12), (9.3.13) ценности шенноновского количества информации вытекают следствия, связывающие теорию ценности информации с теорией оптимальных статистических решений. Приведем следующий несложно доказываемый результат.

Теорема 9.2. *Пусть заданы бейесовская система* $c(x, u)$, $P(x)$ *и наблюдаемая функция* $y(x)$, *описываемая условным распределением* $P(y\mid x)$. *Каков бы ни был решающий алгоритм* $u=d(y)$ *(рандомизированный или нерандомизированный), уровень потерь удовлетворяет неравенству*

$$\mathbf{M}c(x,d(y))\geqslant R(I_{xy})\equiv R_0-V(I_{xy}). \tag{9.3.17}$$

Д о к а з а т е л ь с т в о. Нетрудно видеть, что каков бы ни был решающий алгоритм $u = d(y)$, количество информации о неизвестном значении x не может увеличиться при вынесении оценки, т. е. выполняется неравенство

$$I_{xu} \leqslant I_{xy}. \tag{9.3.18}$$

Из определения (9.3.12), (9.3.13) имеем

$$\mathbf{M} c(x, d(y)) \geqslant R_0 - V(I_{xu}), \tag{9.3.19}$$

поскольку распределение

$$P_d(u \mid x) = \sum_y P(y \mid x) \mathbf{P}(u = d(y) \mid y)$$

относится к множеству распределений, перебираемых при минимизации $\min\limits_{P(u \mid x)} \mathbf{M} c(x, u)$. Зависимость $V(I)$, как указывалось выше, является неубывающей. Поэтому из (9.3.18), (9.3.19) получаем (9.3.17). Доказательство закончено.

Приведенная теорема свидетельствует о плодотворности введенного понятия ценности информации. Вопрос о том, как фактически достигать предельно малых средних штрафов, указываемых теорией ценности информации, будет разбираться в гл. 11.

Понятие информации, соответствующей заданному уровню потерь, было введено (под названием скорости создания сообщений) Шенноном [1] и (под названием W-энтропии или ε-энтропии) Колмогоровым [1], а понятие ценности информации — Стратоновичем [3].

Количества ценности хартлиевского и шенноновского количеств информации не совпадают, причем из их определения следует неравенство $V(I) \geqslant \tilde{V}(I)$. Важным асимптотическим результатом теории информации является их количественное совпадение $V(I) \approx \approx \tilde{V}(I)$ в асимптотическом смысле (гл. 11). Этот результат по глубине и важности можно сравнить с рассмотренным в гл. 7 другим асимптотическим результатом — с асимптотической безошибочностью передачи сообщений по каналу с помехами.

9.4. Решение третьей вариационной задачи. Соответствующие ей потенциалы

1. Задачу отыскания описанного в предыдущем параграфе экстремального распределения и зависимости $R(I)$ или $V(I)$ мы называем *третьей вариационной задачей* теории информации. Как и при решении второй вариационной задачи, для простоты сначала будем считать, что x и u представляют собой дискретные величины, а затем полученные результаты обобщим на произвольный случай.

Будем варьировать вероятности $P(x, u)$ совместного распределения величин x, u так, чтобы обратить в экстремум средние штрафы

$$\sum c(x, u) P(x, u) = \text{extr} \tag{9.4.1}$$

при дополнительном условии

$$\sum_{x,u} \left\{\ln P(x, u) - \ln \left[\sum_x P(x, u)\right] - \ln P(x)\right\} P(x, u) = I \tag{9.4.2}$$

[см. (9.3.3)]. Поскольку априорное распределение $P(x)$ не должно подвергаться изменениям, нужно к (9.4.2) присоединить условие

$$\sum_u P(x, u) = P(x). \tag{9.4.3}$$

Полное условие нормировки

$$\sum_{x,u} P(x, u) = 1 \tag{9.4.4}$$

можно не учитывать, так как оно является следствием из (9.4.3). Не будем вводить также условие неотрицательности $P(x, u) \geqslant 0$, так как найденное без него решение оказывается ему удовлетворяющим (см. ниже). Условную экстремальную задачу (9.4.1)—(9.4.3) будем решать при помощи неопределенных множителей Лагранжа $1/\beta$, $\gamma(x)$, отыскивая экстремум выражения

$$K = \beta \sum_{x,u} c(x, u) P(x, u) + \sum_{x,u} [\ln P(x, u) - \ln P(x) - \\ - \ln \sum_x P(x, u)] P(x, y) + \sum_{x,u} \gamma(x) P(x, u). \tag{9.4.5}$$

Обозначим через Z то множество пар (x, u) (т. е. ту область пространства $X \times U$), которым соответствуют положительные вероятности $P(x, u) > 0$ в экстремальном распределении. Производная от (9.4.5) по $P(x, u)$ должна обращаться в нуль. При этом в силу (9.4.3) не будем дифференцировать $\ln P(x)$. После дифференцирования K получаем (являющееся необходимым условием экстремальности) равенство

$$\beta c(x, u) - \ln P(x) - \ln \sum_x P(x, u) + \ln P(x, u) + \gamma(x) = 0 \tag{9.4.6}$$

или

$$P(x, u) = P(x) P(u) e^{-\gamma(x) - \beta c(x, u)} \quad \text{при } (x, y) \in Z. \tag{9.4.7}$$

Здесь обозначено

$$\sum_x P(x, u) = P(u). \tag{9.4.8}$$

Величины β, $\gamma(x)$ определяются из условий (9.4.2), (9.4.3). Умножая (9.4.6) на $P(x, u)$ и суммируя по x, u, получаем

$$\beta R + \Gamma = -I_{xu}, \tag{9.4.9}$$

где

$$\Gamma = \sum_x \gamma(x) P(x). \tag{9.4.10}$$

Вследствие (9.3.3)

$$\beta R + \Gamma = -I. \tag{9.4.11}$$

Подставляя далее (9.4.7) в (9.4.3) и (9.4.8), получаем равенства

$$\sum_{x \in Z_u} P(x)\, \mathrm{e}^{-\gamma(x) - \beta c(x, u)} = 1, \tag{9.4.12}$$

$$\sum_{u \in Z_x} P(u)\, \mathrm{e}^{-\beta c(x, u)} = \mathrm{e}^{\gamma(x)}. \tag{9.4.13}$$

Здесь условия $x \in Z_u$ и $u \in Z_x$ соответствуют условию $(x, u) \in Z$.

2. Теорема 9.3. *Если рассматривать лишь вариации $dP(x, u)$ или $dP(u \mid x)$, оставляющие область Z неизменной, то распределение* (9.4.7) *соответствует минимальной информации I_{xu} при фиксированном условии* $\mathsf{M}c(x, u) = a$. *Оно соответствует также минимальным средним штрафам* (9.3.1) *при условии* (9.3.3), *если* $\beta > 0$, *и максимальным средним штрафам, если* $\beta < 0$.

Доказательство. Дважды дифференцируя функцию (9.4.5) по переменным $P(x, u)$, $(x, u) \in Z$, находим матрицу вторых производных

$$\frac{\partial^2 K}{\partial P(x, u)\, \partial P(x', u')} = \frac{\partial^2 I}{\partial P(x, u)\, \partial P(x', u')} = \frac{\delta_{xx'}\, \delta_{uu'}}{P(x, u)} - \\ - \frac{\delta_{uu'}}{P(u)}, \quad (x, u) \in Z, (x', u') \in Z. \tag{9.4.14}$$

Докажем, что эта матрица является неотрицательно определенной. Для этого достаточно показать, что таковой является x-матрица

$$L_{xx'} = \frac{\delta_{xx'}}{P(x, u)} - \frac{1}{P(u)}$$

(u фиксировано) или матрица $\dfrac{\delta_{xx'}}{a_x} - \dfrac{1}{\sum_x a_x}$ при любом положительном a_x.

Образуем квадратичную форму

$$\sum_{xx'} L_{xx'} v_x v_{x'} = \sum_x \frac{v_x^2}{P(x, u)} - \frac{\left(\sum_x v_x\right)^2}{P(u)},$$

где v_x — произвольный вектор. Введем новые переменные w_x равенством $v_x = P(x \mid u)\, w_x$ и будем иметь

$$\sum_{xx'} L_{xx'} v_x v_{x'} = \frac{1}{P(u)} \left\{ \sum_x P(x \mid u)\, w_x^2 - \left(\sum_x P(x \mid u)\, w_x \right)^2 \right\}.$$

Но выражение в фигурных скобках есть не что иное, как дисперсия величины w_x, соответствующая распределению $P(x \mid u)$. Ее неотрицательность доказывает неотрицательную определенность матрицы $L_{xx'}$, а, следовательно, и матрицы (9.4.14).

Вследствие этого функции K и I_{xu} являются вогнутыми функциями переменных $P(x, u)$, $(x, u) \in Z$. Разлагая эти функции в ряд Тейлора в точке, соответствующей экстремальному распределению (9.4.7), и учитывая, что линейные члены разложения для этого распределения исчезают, получаем в силу указанной неотрицательной определенности

$$dK \geqslant 0; \qquad dI_{xu} \geqslant 0. \tag{9.4.15}$$

Раз эти соотношения справедливы для произвольных вариаций переменных $P(x, u)$, $(x, u) \in Z$, то они справедливы, в том числе и для вариаций, совместимых с дополнительными условиями (9.4.3) и др. При фиксированном условии $\mathbf{M}c(x, u) = a$ согласно (9.4.15) имеем $dI_{xu} \geqslant 0$, что доказывает первое утверждение теоремы. Для доказательства второго утверждения нужно учесть, что соотношение

$$dK = \beta d[\mathbf{M}\, c(x, u)] + dI_{xu} + \sum_x \gamma(x)\, d\sum_u P(x, u) \geqslant 0, \tag{9.4.15a}$$

взятое в сочетании с условиями $I_{xu} = I$ и (9.4.3) дает $\beta d\,[\mathbf{M}c(x, u)] \geqslant 0$, т. е. $d\,[\mathbf{M}c] \geqslant 0$ при $\beta > 0$ и $d\,(\mathbf{M}c) \leqslant 0$ при $\beta < 0$. Доказательство закончено.

Теорема 9.4. *«Активная» область Z, где экстремальные вероятности отличны от нуля: $P(x, u) > 0$, имеет цилиндрический вид*

$$Z = \widetilde{X} \times \widetilde{U}. \tag{9.4.16}$$

Здесь $\widetilde{X}$ — область, где $P(x) > 0$, а $\widetilde{U}$ — область, где $P(u) > 0$.

Доказательство. Предположим, что $Z = Z_1$ не совпадает с $X \times \widetilde{U}$. Тогда, очевидно, Z_1 должна быть частью области $\widetilde{X} \times \widetilde{U}$. Экстремальное распределение для области Z_1 согласно (9.4.7) представляется в виде

$$P_1(x, u) = P(x) P_1(u)\, e^{-\gamma_1(x) - \beta c(x, u)}, \qquad (x, u) \in Z_1, \tag{9.4.17}$$

где $\gamma_1(x)$, $P_1(u)$ удовлетворяют всем необходимым условиям. Используя (9.4.17), построим вспомогательное распределение $P_2(x, u)$, с вероятностями, отличными от нуля в более широкой области $Z_2 = \widetilde{X} \times \widetilde{U}$.

Положим

$$P_2(x, u) = P(x) P_2(u \mid x), \qquad (x, u) \in Z_2, \tag{9.4.18}$$

где

$$P_2(u \mid x) = \frac{P(u)\, e^{-\gamma_1(x) - \beta c(x, u)}}{\sum\limits_{u \in \widetilde{U}} P(u)\, e^{-\gamma_1(x) - \beta c(x, u)}} \equiv P(u)\, e^{-\gamma_2(x) - \beta c(x, u)}.$$

Учитывая равенство (9.4.13), принимающее вид

$$\sum_{u \in Z_x} P(u)\, \mathrm{e}^{-\beta c(x,u)} = \mathrm{e}^{\gamma_1(x)}, \quad x \in \widetilde{X},$$

нетрудно видеть, что

$$\sum_{u \in \widetilde{U}} P(u)\, \mathrm{e}^{-\beta c(x,u)} \geqslant \mathrm{e}^{\gamma_1(x)},$$

поскольку область суммирования расширена. Поэтому

$$\gamma_2(x) = \ln \sum_{u \in \widetilde{U}} P(u)\, \mathrm{e}^{-\beta c(x,u)} \geqslant \gamma_1(x).$$

Отсюда усреднением получаем

$$\Gamma_2 \geqslant \Gamma_1. \qquad (9.4.19)$$

Для обоих распределений (9.4.17), (9.4.18) справедлива формула (9.4.9), т. е. $\beta R_1 + I_1 = -\Gamma_1$; $\beta R_2 + I_2 = -\Gamma_2$. Из (9.4.19) имеем

$$\beta R_2 + I_2 \leqslant \beta R_1 + I. \qquad (9.4.20)$$

Ранее параметр β предполагается фиксированным. Рассмотрим теперь все семейство $Z(\beta)$ зависящих от β активных областей и проведем описанное расширение областей $Z_1(\beta) \to Z_2(\beta)$ для каждого β. Тогда из соотношения (9.4.20), справедливого при каждом β, будет вытекать, что $I_2(R) \leqslant I_1(R)$ или $R_2(I) \leqslant R_1(I)$ при $\beta > 0$. Следовательно, рассматривая лишь цилиндрические «активные» области (9.4.15), мы не проиграем в оптимальности. Доказательство закончено.

Как выражение, стоящее в правой части (9.4.7), так и выражение, стоящее в левой части, вне области $\widetilde{X} \times \widetilde{U}$, равны нулю. Поэтому для экстремального распределения равенство

$$P(x,u) = P(x) P(u)\, \mathrm{e}^{-\gamma(x) - \beta c(x,u)} \qquad (9.4.21)$$

выполняется повсеместно. Вследствие теоремы *9.4* равенства (9.4.12), (9.4.13) записываются в виде

$$\sum_{x \in \widetilde{X}} P(x)\, \mathrm{e}^{-\gamma(x) - \beta c(x,u)} = 1, \quad u \in \widetilde{U}, \qquad (9.4.22)$$

$$\sum_{u \in \widetilde{U}} P(u)\, \mathrm{e}^{-\beta c(x,u)} = \mathrm{e}^{\gamma(x)}, \quad x \in \widetilde{X}. \qquad (9.4.23)$$

В этих формулах суммирование можно проводить также по всему пространству значений x и u.

Уравнения (9.4.22), (9.4.23) с соответствующим условием (9.4.2) при фиксированной области $\widetilde{U}$ позволяют найти оптимальное распределение (9.4.21). Каждой области $\widetilde{U}$, удовлетворяющей ряду условий, будет соответствовать, таким образом, экстремальная зависимость $R(I)$. Для полного решения задачи остается еще решить вопрос о том, как выбрать активную область $\widetilde{U}$ из множества D

допустимых областей. Для этого выбора естественно использовать условие экстремальности

$$R(I) = \underset{\tilde{U} \in D}{\mathrm{extr}}, \tag{9.4.24}$$

вытекающее из (9.4.1). Если множество D допускает непрерывные изменения $\delta\tilde{U}$ области $\tilde{U}$, то условие (9.4.24), как правило, можно заменить условием стационарности

$$\delta R = 0. \tag{9.4.25}$$

Здесь вариация δR соответствует вариации $\delta\tilde{U}$ области $\tilde{U}$, причем информация I остается постоянной:

$$\delta I = 0. \tag{9.4.26}$$

Вариация $\delta\tilde{U}$, несмотря на отсутствие вариаций (9.4.25), (9.4.26), может сопровождаться ненулевыми вариациями $\delta\beta$, $\delta\Gamma$. Из соотношения (9.4.9) в силу (9.4.25), (9.4.26) получаем условие

$$R\delta\beta = -\delta\Gamma, \tag{9.4.27}$$

которое будет использовано в дальнейшем. Как будет показано в дальнейшем, это условие может быть приведено также к виду (10.1.8), что означает экстремальность потенциала Γ при фиксированном β.

3. Перейдем к выводу «термодинамических» соотношений, связанных с третьей вариационной задачей. Оказывается, что они во многом аналогичны соотношениям первой и второй вариационных задач (§ 3.3, 3.6 и 8.2).

Обратимся сначала к соотношению (9.4.11). Потенциал (9.4.10) естественно интерпретировать как $-\beta F$, где F примет вид

$$F = R + TI \tag{9.4.28}$$

— аналог свободной энергии, а $\beta = 1/T$. Тогда (9.4.28) будет напоминать известное в термодинамике соотношение $F = U - TH$ (U — внутренняя энергия: H — энтропия). Разница между этими соотношениями лишь в том, что член TI входит с противоположным знаком.

Перейдем к выводу прочих соотношений, напоминающих обычные соотношения термодинамики. Начнем с того варианта, в котором берется потенциал Γ (β), а не свободная энергия F $(T) = -T\Gamma$ $(1/T)$.

Теорема 9.5. *Для третьей вариационной задачи выполняются соотношения*

$$\frac{d\Gamma}{d\beta} = -R, \tag{9.4.29}$$

$$\beta\frac{d\Gamma}{d\beta} - \Gamma = I, \tag{9.4.30}$$

$$\frac{dR}{dI} = -\frac{1}{\beta} \tag{9.4.31}$$

по аналогии с первыми двумя вариационными задачами.

Доказательство. Будем варьировать параметр β в уравнении (9.4.22). Эта вариация сопровождается, вообще говоря, вариацией функции $\gamma(x)$ и активной области $\tilde{U}$. Представим вариацию $d\gamma(x)$, а также $d\Gamma$ как сумму двух вариаций:

$$d\gamma(x) = d_1\gamma(x) + \delta\gamma(x); \quad d\Gamma = d_1\Gamma + \delta\Gamma. \tag{9.4.32}$$

Вариации $d_1\gamma(x)$, $d_1\Gamma$ соответствуют вариации $d_1\beta$ параметра β при неизменной области $\tilde{U}$, а вариации $\delta\gamma(x)$, $\delta\Gamma$, $\delta\beta = d\beta - d_1\beta$ соответствуют изменению $\delta\tilde{U}$ области $\tilde{U}$.

Дифференцируя (9.4.22) при постоянной области $\tilde{U}$, получаем

$$\sum_x \left[\frac{d_1\gamma(x)}{d\beta_1} + c(x, u)\right] P(x)\, e^{-\gamma(x) - \beta c(x, u)} = 0, \quad u \in \tilde{U}.$$

Умножая это равенство на $P(u)$ и суммируя по $u \in \tilde{U}$ при учете формулы (9.4.21), находим

$$\mathbf{M}\left[\frac{d_1\gamma(x)}{d\beta_1} + c(x, u)\right] = 0,$$

т. е. получаем

$$\frac{d_1\Gamma}{d\beta_1} = -R \tag{9.4.33}$$

вследствие (9.3.5), (9.4.10). Из (9.4.33), (9.4.11) с очевидностью следует соотношение

$$\beta\frac{d_1\Gamma}{d_1\beta} - \Gamma = I. \tag{9.4.34}$$

Требуемые соотношения (9.4.29), (9.4.30) получим, если учтем, что

$$\frac{d\Gamma}{d\beta} = \frac{d_1\Gamma + \delta\Gamma}{d_1\beta + \delta\beta} = -R$$

в силу (9.4.27), (9.4.33). Для вывода (9.4.31) достаточно взять дифференциал от обеих частей равенства (9.4.30), что дает

$$d\beta\frac{d\Gamma}{d\beta} + \beta\, d\left(\frac{d\Gamma}{d\beta}\right) - d\Gamma = dI \qquad \left(d\Gamma = \frac{d\Gamma}{d\beta}d\beta\right),$$

и учесть (9.4.29). Доказательство закончено.

Согласно теореме 9.5 для вычисления зависимости $R(I)$ удобно пользоваться потенциалом $\Gamma(\beta)$. При этом, как видно из приведенного доказательства, дифференцирование потенциала по β можно проводить, считая область $\tilde{U}$ не зависящей от β.

Аналогичные соотношения справедливы и для свободной энергии $F(T) = -T\Gamma(1/T)$. В самом деле, подставляя $\Gamma'(\beta) = -\beta F(1/\beta)$ в (9.4.29)—(9.4.31), будем иметь

$$F - T\frac{dF}{dT} = R; \tag{9.4.35}$$

$$\frac{dF}{dT} = I; \tag{9.4.36}$$

$$\frac{dR}{dI} = -T. \tag{9.4.37}$$

Из приведенных соотношений видно, что зависимость $R(I)$ является преобразованием Лежандра

$$-R(I) = -F(T(I)) + T(I)\,I \qquad \left(I = \frac{dF(T)}{dT}\right) \tag{9.4.38}$$

функции $F(T)$, а функция $I(R)$ есть преобразование Лежандра

$$I(R) = -\beta(R)\,R - \Gamma(\beta(R)) \qquad \left(R = -\frac{d\Gamma}{d\beta}(\beta)\right) \tag{9.4.39}$$

функции $\Gamma(\beta)$. Производную

$$\frac{dV(I)}{dI} \equiv v(I) \tag{9.4.40}$$

или, что то же самое, производную

$$\mp dR_{\pm}(I)/dI$$

целесообразно интерпретировать как дифференциальную ценность информации. Вследствие (9.4.30), (9.4.36) она оказалась связанной с температурным параметром простым соотношением

$$v(I) = |T| = 1/|\beta|. \tag{9.4.41}$$

Эта функция всегда неотрицательна.

4. Ряд общих утверждений можно высказать относительно свойств выпуклости функций $\Gamma(\beta)$, $I(R)$, $F(T)$, $R(I)$.

Теорема 9.6. *Функции $I(R)$, $\Gamma(\beta)$ являются вогнутыми; нормальная ветвь функции $R(I)$ является вогнутой, а аномальная ветвь — выпуклой.*

Доказательство. Докажем сначала вогнутость функции $I(R)$. Для этого возьмем два значения R', R'' из допустимого диапазона (9.3.16). Пусть им соответствуют экстремальные распределения $P'(x, u)$, $P''(x, u)$ и значения информации $I(R')$, $I(R'')$. Рассмотрим промежуточную точку

$$R_\lambda = \lambda R' + (1-\lambda)R'' \quad (0 < \lambda < 1).$$

Это значение является средними штрафами для распределения

$$P_\lambda = \lambda P'(x, u) + (1-\lambda)P''(x, u). \tag{9.4.42}$$

Пусть $I_{xu}[P_\lambda$ есть информация связи I_{xu} для распределения (9.4.42). При доказательстве теоремы 9.3 было показано, что выражение $I_{xu} = I_{xu}[P]$ является вогнутой функцией от вероятностей $P(x, u)$, поэтому

$$I_{xu}[P_\lambda] \leqslant \lambda I(R') + (1-\lambda) I(R''). \tag{9.4.43}$$

Сравним теперь значение $I_{xu}[P_\lambda]$ и значение $I(R_\lambda)$, являющееся решением экстремальной задачи (9.3.4). Распределение (9.4.42) входит в число распределений, перебираемых при минимизации, так что $I(R_\lambda) \leqslant I_{xu}[R_\lambda]$. Сопоставляя это неравенство с (9.4.41), получаем

$$I\left(\lambda R' + (1-\lambda) R''\right) \leqslant \lambda I(R') + (1-\lambda) I(R''),$$

что и доказывает вогнутость функции $I(R)$. Вогнутость функции $\Gamma(\beta)$ вытекает из вогнутости функции $I(R)$, если учесть, что эти функции связаны преобразованием Лежандра (9.4.38), а преобразование Лежандра сохраняет свойство выпуклости или вогнутости. Последний факт легче всего показать для того случая, когда рассматриваемые функции дважды дифференцируемы. Дифференцируя (9.4.29), имеем

$$\frac{d^2\Gamma}{d\beta^2} = -\frac{dR}{d\beta}. \tag{9.4.44}$$

Запишем теперь (9.4.31) в виде

$$\frac{dI}{dR} = -\beta, \tag{9.4.45}$$

и возьмем производную по R

$$\frac{d^2 I}{dR^2} = -\frac{d\beta}{dR}. \tag{9.4.46}$$

Сравнивая равенства (9.4.44), (9.4.46), получаем

$$\frac{d^2\Gamma}{d\beta^2} = 1 \Big/ \frac{d^2 I}{dR^2}.$$

Отсюда видно, что условие вогнутости $d^2I/dR^2 \geqslant 0$ приводит к условию вогнутости $d^2\Gamma/d\beta^2 \geqslant 0$.

Нормальные ветви функции $R(I)$, $F(T)$ характеризуются тем, что значения параметров β, T положительны. Вследствие (9.4.31) производная dI/dR для нормальной ветви отрицательна и из вогнутости функции $I(R)$ следует вогнутость обратной функции $I(R)$. Если производная dI/dR положительна, то имеет место изменение направления выпуклости при переходе к обратной функции $R(I)$. Доказательство закончено.

Кроме функции (9.4.38) и функции ценности $V(I) = R(0) - R_+(I)$ можно ввести соответствующие случайные функции, т. е.

функции, зависящие не только от I, но и от случайной переменной x. Принимая во внимание формулу (9.4.10), записанную в виде

$$F(T) = -T \sum_x \gamma(1/T, x) P(x),$$

нетрудно видеть, что преобразование Лежандра (9.4.38) можно произвести до усреднения по x, т. е. преобразовать по Лежандру функцию $F(T, x) = -T\gamma(1/T, x)$.

Соответствующее изображение

$$V(x, I) = -F(T, x) + TI \qquad \left(I = \frac{\partial F(T, x)}{\partial T}\right)$$

назовем *случайной ценностью* или ценностью случайной информации, поскольку $I = \partial F(T, x)/\partial T = -\gamma(\beta, x) + \beta \partial\gamma(\beta, x)/\partial\beta$ в силу (9.4.23), (9.4.21) есть не что иное, как случайная информация

$$I_u(|x) = \sum_u P(u|x) \ln \frac{P(x, u)}{P(x) P(u)} \text{ при } dP(u)/d\beta = 0.$$

9.5. Решение вариационной задачи при некоторых дополнительных предположениях

1. Приведем решение третьей вариационной задачи в более явном виде для двух частных случаев. Обозначив

$$P(x) e^{-\gamma(x)} = Q(x), \tag{9.5.1}$$

запишем уравнение (9.4.22) в форме

$$\sum_x Q(x) e^{-\beta c(x, u)} = 1, \qquad u \in \tilde{U}. \tag{9.5.2}$$

Предположим сначала, что это уравнение допускает решение относительно неизвестной функции $Q(x)$ как система линейных уравнений, или линейное интегральное уравнение. Другими словами, существует преобразование

$$L^{-1} f \equiv \sum_u b_{xu} f(u) = g(x),$$

обратное преобразованию

$$Lg \equiv \sum_x e^{-\beta c(x, u)} g(x) = f(u),$$

т. е.

$$\| b_{xu} \| = \| e^{-\beta c(x, u)} \|^{-1}. \tag{9.5.3}$$

Тогда из (9.5.2) будем иметь

$$Q(x) = \sum_{u \in \tilde{U}} b_{xu},$$

и, если учесть (9.5.1),

$$\gamma(x) = \ln P(x) - \ln \sum_{u \in \widetilde{U}} b_{xu}, \quad x \in \widetilde{X}.$$

Усреднение с весом $P(x)$ дает выражение для потенциала

$$\Gamma = -H_x - \sum_x P(x) \ln \left(\sum_{u \in \widetilde{U}} b_{xu} \right). \tag{9.5.4}$$

Используя (9.4.28), (9.4.29), отсюда получаем зависимость между R и I в параметрической форме

$$R = \sum_x \frac{\sum_u \frac{\partial b_{xu}}{\partial \beta}}{\sum_u b_{xu}} P(x), \tag{9.5.5}$$

$$I = H_x - \sum_x \left[\frac{\beta \sum_u \frac{\partial b_{xu}}{\partial \beta}}{\sum_u b_{xu}} - \ln \left(\sum_u b_{xu} \right) \right] P(x). \tag{9.5.6}$$

В последней формуле вычитаемое выражение в правой части, очевидно, имеет смысл условной энтропии $H_{x \mid u}$.

Если ввести функцию

$$-\Gamma_0(\beta) = \beta F_0\left(\frac{1}{\beta}\right) = \sum_x P(x) \ln \left(\sum_{u \in \widetilde{U}} b_{xu} \right)$$

и ее преобразование Лежандра

$$H_0(R) = \Gamma_0 - \beta \frac{d\Gamma_0}{d\beta} \quad \left(-\frac{d\Gamma_0}{d\beta} = R \right),$$

то полученные результаты (9.5.5), (9.5.6) можно записать в следующей компактной форме:

$$I = H_x - H_0(R). \tag{9.5.7}$$

2. Сделаем теперь другое предположение, а именно предположим, что функция $Q(x) = Q$ (9.5.1) является постоянной в области $\widetilde{X}$. Вынося Q за знак суммы, в этом случае из (9.5.2) будем иметь

$$\sum_{x \in \widetilde{X}} e^{-\beta c(x, u)} = \frac{1}{Q}.$$

Сумма в левой части равенства напоминает статистическую сумму (3.3.11), (3.6.2), вводимую при решении первой вариационной задачи. В принятых предположениях эта сумма не должна зависеть от u. По аналогии с (9.1.3) обозначим

$$e^{-\beta F_0(\beta)} = \sum_{x \in X} e^{-\beta c(x, u)} = \frac{1}{Q(\beta)}. \tag{9.5.8}$$

Тогда $\ln Q = \beta F_0$ и из (9.5.1) получаем

$$\gamma(x) = \ln P(x) - \beta F_0, \tag{9.5.9}$$

и (после усреднения)

$$\Gamma(\beta) = -H_x - \beta F_0. \tag{9.5.10}$$

Как указывалось в предыдущем параграфе, функция $R(I)$ может быть получена преобразованием Лежандра

$$-R(I) = -F + T\frac{dF}{dT} \quad \left(\frac{dF}{dT} = I\right) \tag{9.5.11}$$

функции

$$F(T) = -T\Gamma(1/T) = TH_x + F_0(T). \tag{9.5.12}$$

Кроме вспомогательной функции $-F_0(T)$, удобно ввести ее преобразование Лежандра

$$R_0(H) = F_0 - T\frac{dF_0}{dT} \quad \left(-\frac{dF_0}{dT}(T) = H\right) \tag{9.5.13}$$

в полной аналогии с формулами (9.1.4), (9.1.5). Тогда преобразование Лежандра (9.5.11) функции (9.5.12) запишется так:

$$R(I) = F_0(T) - T\frac{dF_0}{dT}(T) \text{ при } H_x + \frac{dF_0}{dT}(T) = I,$$

т. е.

$$R(I) = R_0(H_x - I) \qquad (H = H_x - I) \tag{9.5.14}$$

в силу (9.5.13). Формула (9.5.14) согласуется с (9.5.7), поскольку функция $R_0(H)$ является обратной к функции $H_0(R)$. Образуя функцию ценности информации

$$V(I) = R(0) - R(I) \qquad (\beta > 0),$$

получаем формулу

$$V(I) = R_0(H_x) - R_0(H_x - I), \tag{9.5.15}$$

вполне аналогичную формуле (9.1.6). Энтропия $H = H_x - I$, очевидно, имеет смысл условной энтропии $H_{x|u} = H_{\text{ps}}$.

Таким образом, в рассматриваемом случае развитая теория ценности информации полностью подтверждает предварительные соображения, изложенные в § 9.1, в которых информация рассматривалась просто как разность двух энтропий, а не как шенноновское количество информации, связывающее две случайные величины. В рассматриваемом случае вследствие (9.4.21), (9.5.1) и соотношения $Q = e^{\beta F_0}$ экстремальное распределение имеет вид

$$P(x, u) = P(u)\, e^{\beta F_0 - \beta c(x, u)}.$$

Ему, очевидно, соответствуют условные вероятности

$$P(x \mid u) = e^{\beta F_0 - \beta c(x, u)},$$

вид которых напоминает выражение (9.1.2).

Рассмотрим примеры, тесно связанные с примерами, которые были изучены в § 9.1 и 9.2.

Пример 1. Пусть имеется бейесовская система, описанная в начале § **9.2.** Для нее была подсчитана ценность хартлиевского количества информации. Найдем для сравнения ценность шенноновского количества информации.

Запишем для данного примера уравнение (9.5.2), учитывая, что функция $c(x, u)$ (9.2.1) зависит лишь от разности $x - u$. Это уравнение принимает вид

$$\sum_x Q(x)\, e^{-\beta c(x-u)} = 1.$$

Это уравнение имеет решение, когда в качестве активной области выбрана вся область U, состоящая из восьми точек. Решение соответствует тривиальной, т. е. постоянной функции $Q(x) = Q$. Статистическая сумма (9.5.8)

$$e^{-\beta F_0} = \sum_x e^{-\beta c(x-u)}$$

не зависит от u и совпадает с суммой (9.1.10). Зависимость $R_0(H)$ представлена параметрически выражениями (9.1.11), (9.1.12). Она изображена на рис. 9.2.

Равномерное распределение $P(x) = 1/8$ соответствует энтропии $H_x = 3 \ln 2$. Поэтому из формулы (9.5.14) имеем $R(I) = R_0(3\ln 2 - I)$, а функция ценности шенноновского количества информации имеет вид $V(I) = R_0(3 \ln 2) - R_0(3 \ln 2 - I)$. Эта функция представлена графически на рис. 9.4. Данная кривая является не чем иным, как перевернутой кривой рис. 9.2.

Ступенчатой линией (рис. 9.4) изображены те значения ценности хартлиевской информации, которые были найдены в § 9.2. Она лежит ниже основной кривой $V(I)$. Так ценность $V(I)$ одного бита информации, как видно из графика, равна 1,35, что превосходит значение 1, найденное по теории § 9.2. Аналогично ценность $\widetilde{V}(I)$ двух битов информации равна 1,78 и больше ранее найденного значения 1,5.

Для любого другого числа точек, располагающихся по кругу, ценность шенноновской информации может быть найдена совершенно аналогичным образом, причем будут наблюдаться монотонные зависимости без каких-либо нерегулярных скачков.

В данном примере отчетливо видна разница между ценностью хартлиевского и шенноновского количества информации. Как видно из дальнейшего (гл.11) в некоторых более сложных случаях это различие стирается, например, если рассматривать не пару описанных выше случайных величин x, u, а последовательности $x_1, \ldots, x_N$; $u_1, \ldots, u_N$ подобных величин при больших N.

Пример 2. Рассмотрим пример, в котором имеется пространство из бесконечного числа точек.

Пусть x и u могут принимать любые целые значения ..., —1, 0, 1, 2, ..., подобно переменной ξ в примере 1 § 9.1. Возьмем простую функцию штрафов

$$c(x, u) = |x - \xi|.$$

и априорные вероятности

$$P(x) = \frac{1-\nu}{1+\nu} \nu^{|x|} \quad (0<\nu<1). \tag{9.5.16}$$

В отличие от предыдущего примера априорное распределение является неравномерным. Уравнение (9.5.2), записываемое в виде

$$\sum_{x=-\infty}^{\infty} Q(x) e^{-\beta |x-u|} = 1,$$

имеет в качестве решения константу

$$Q^{-1} = \frac{1+e^{-\beta}}{1-e^{-\beta}} = e^{-\beta F_0},$$

где F_0 — функция от β, подсчитанная в (9.1.7). Исключением параметра $z = e^{-\beta}$ из выражения (9.1.8), (9.1.9) может быть найдена зависимость $R_0(H)$. Для вычисления функции ценности информации по формуле (9.5.15) остается найти энтропию H_x априорного распределения (9.5.16). По аналогии с (9.1.9) нетрудно получить

$$H_x = -\frac{2\nu \ln \nu}{1-\nu^2} + \ln \frac{1+\nu}{1-\nu}.$$

Итак, в данном примере так же, как и в предыдущем, функция ценности (9.5.15) может быть подсчитана при помощи функции $R_0(H)$, найденной в § 9.1 (пример 1) и изображенной на рис. 9.1.

Вид вероятностей $P(u)$, а также экстремального распределения $P(x, u)$ теперь находится более сложно, чем в предыдущем примере, когда распределение $P(u)$ подобно $P(x)$ было равномерным. Теперь для отыскания $P(u)$ необходимо решить уравнение (9.4.23). Поскольку в силу (9.5.1), (9.5.16)

$$e^{\gamma(x)} = \frac{P(x)}{Q} = e^{-\beta F_0} \frac{1-\nu}{1+\nu} \nu^{|x|},$$

это уравнение принимает вид

$$\sum_u e^{-\beta |x-u|} P(u) = \frac{1+z}{1-z} \frac{1-\nu}{1+\nu} \nu^{|x|}.$$

Данное уравнение допускает точное решение. Умножая обе части на τ^x ($\tau = e^{i\lambda}$) и суммируя по x, получаем

$$\left(\sum_\sigma e^{-\beta|\sigma|} \tau^\sigma\right) \sum_u \tau^u P(u) = \frac{1+z}{1-z} \frac{1-\nu}{1+\nu} \sum_x \nu^{|x|} \tau^x,$$

но

$$\sum_{\sigma} e^{-\beta|\sigma|}\tau^{\sigma} = \sum_{\sigma=-\infty}^{\infty} z^{|\sigma|}\tau^{\sigma} = \frac{1}{1-z\tau} + \frac{1}{1-\dfrac{z}{\tau}} - 1 =$$

$$= \frac{1-z^2}{1-2z\cos\lambda+z^2},$$

$$\sum_{x} \nu^{|x|}\tau^{x} = \frac{1-\nu^2}{1-2\nu\cos\lambda+\nu^2} \tag{9.5.17}$$

и, следовательно,

$$\sum_{u} \tau^{u} P(u) = \left(\frac{1-\nu}{1-z}\right)^2 \frac{1+z^2-2z\cos\lambda}{1+\nu^2-2\nu\cos\lambda}.$$

Совершая преобразование

$$\frac{1+z^2-2z\cos\lambda}{1+\nu^2-2\nu\cos\lambda} = \frac{z}{\nu}\,\frac{\cos\lambda-(1+z^2)/2z}{\cos\lambda-(1+\nu^2)/2\nu} =$$

$$= \frac{z}{\nu}\left[1+\frac{(1+\nu^2)/2\nu-(1+z^2)/2z}{\cos\lambda-(1+\nu^2)/2\nu}\right] = \frac{z}{\nu} - \frac{z(1+\nu^2)/\nu-1-z^2}{1-2z\cos\lambda+\nu^2},$$

и еще раз пользуясь формулой (9.5.17), получаем искомые вероятности

$$P(u) = \frac{z}{\nu}\left(\frac{1-\nu}{1-z}\right)^2 \delta_{u0} + \frac{(\nu-z)(1-\nu z)}{\nu(1-z)^2}\,\frac{1-\nu}{1+\nu}\,\nu^{|u|}.$$

Теперь в соответствии с (9.4.21) нетрудно подсчитать и вероятности совместного распределения

$$P(x, u) = P(u)\, e^{\beta F_0 - \beta|u-x|},$$

что дает полное решение экстремальной задачи.

9.6. Ценность больцмановского количества информации

Основной исходной мерой количества информации, без сомнения, является хартлиевское количество информации $I = \ln M$ (см. § 1.1), непосредственно связанное с числом реализаций M. На практике часто пользуются шенноновским количеством информации (6.2.1), которое удобно вычислять аналитически для сложных практических случайных величин и процессов. Однако следует иметь в виду, что обоснованием шенноновского количества информации в конечном счете является его сводимость к хартлиевскому количеству, о чем говорит теорема Шеннона (гл. 7). Ценность шенноновской информации $V(I)$ является удобной аппроксимацией ценности хартлиевской информации $\widetilde{V}(I)$.

Рассмотрим еще одно количество информации — больцмановское, которое занимает промежуточное положение между хартлиевским и шенноновским количествами, и соответствующую функцию ценности.

Как видно из сказанного в гл. 1, количество информации можно определять не только формулой Хартли (1.1.5) и формулой Шеннона (6.2.1), но и формулой Больцмана

$$I = H_y = -\sum_{y=y_1}^{y_M} P(y) \ln P(y) \tag{9.6.1}$$

[см. (1.2.3)].

Последнее количество *информации отличается от* хартлиевского количества, а именно

$$H_y < \ln M,$$

если отличные от нуля вероятности $P(y)$ не равны друг другу. В любом случае количество информации (9.6.1) занимает промежуточное положение

$$\ln M \geqslant H_y \geqslant I_{yz} \tag{9.6.2}$$

между хартлиевским количеством информации $\ln M$ и шенноновским $I_{yz} = H_y - H_{y|z}$.

Соответственно этому можно ввести функцию $\tilde{\tilde{V}}(I)$ ценности больцмановского количества информации, которая будет занимать промежуточное положение между функцией $\tilde{V}(I)$ ценности хартлиевского количества информации и функцией $V(I)$ ценности шенноновского количества.

При определении ценности $\tilde{V}(I)$ (9.2.6) рассматривалась выгода

$$\Delta = \min_u \mathbf{M}\, c(x, u) - \mathbf{M} \min_u \mathbf{M}\left[c(x, u) \mid E_k\right], \tag{9.6.3}$$

приносимая информацией $y = y_k$, носящей дискретный характер. Затем эта выгода максимизировалась по различным разбиениям $E_1 + \ldots + E_M$ пространства значений x на фиксированное количество областей. При рассмотрении количества информации (9.6.1) дискретный характер наблюдаемой величины $y = y_k$, очевидно, нужно сохранить и рассматривать ту же выгоду (9.6.3), соответствующую разбиению ΣE_k. Максимизацию же по различным разбиениям нужно проводить несколько иначе. Именно теперь нужно фиксировать не условие $M \leqslant e^I$, наложенное на число областей, а условие

$$H_k \equiv -\sum_k P(E_k) \ln P(E_k) \leqslant I, \tag{9.6.4}$$

наложенное на энтропию разбиения $\sum_k E_k$. Число же областей может быть любым. В остальном формула, определяющая $\tilde{\tilde{V}}(I)$, будет совпадать с (9.2.6):

$$\tilde{\tilde{V}}(I)=\min_u \mathbf{M}c(x, u)-\inf_{\Sigma E=X_k} \mathbf{M}\min_u \mathbf{M}\left[c(x, u) \mid E_k\right], \tag{9.6.5}$$

где минимизация по ΣE_k соответствует условию (9.6.4). Мы видим, что выражение в правой части (9.6.5) совпадает с соответствующими выражениями (9.2.6) и (9.2.6a).

Теорема 9.7. *Функция ценности $\tilde{\tilde{V}}(I)$ больцмановского количества информации лежит между функциями ценности $\tilde{V}(I)$ и $V(I)$:*

$$\tilde{V}(I) \leqslant \tilde{\tilde{V}}(I) \leqslant V(I). \tag{9.6.6}$$

Доказательство. Пусть u_k — значение, определяемое минимизацией второго члена в (9.6.3). Из (9.6.4) вытекает, что $H_{u_k} \leqslant I$, и, следовательно,

$$I_{xu_k} \equiv H_{u_k}-H_{u_k \mid x} \leqslant I.$$

Поэтому переходные вероятности $P(u_k \mid x)$ входят в множество G переходных вероятностей, перебираемых при минимизации в определении (9.3.12). Отсюда

$$\tilde{\tilde{V}}(I) \leqslant V(I).$$

С другой стороны, при минимизации в (9.2.6) перебираются разбиения $E_1+\ldots+E_M=X$. Для каждого такого разбиения условие (9.6.4) выполняется. Следовательно, класс разбиений, перебираемых в (9.6.5), является более широким, чем в (9.2.6). Отсюда вытекает, что

$$\tilde{V}(I) \leqslant \tilde{\tilde{V}}(I).$$

Доказательство закончено.

Пример. Чтобы проиллюстрировать сказанное выше, вернемся к примеру, рассмотренному в п. 1 § 9.2 и в § 9.5 (пример 1). Будем перебирать всевозможные разбиения множества из 8 точек на связные области $E_1, \ldots, E_M$ (разбиения на области, не состоящие из смежных точек, не являются экстремальными). Для каждого разбиения найдем значение больцмановской информации (9.6.1), где $P(y_k)=P(E_k)$, и разность (9.6.3), причем первый член в данном случае равен двум. Найденные таким образом точки (I, Δ) откладываем на плоскости (I, V). Далее проводим на ней ступенчатую линию $\tilde{\tilde{V}}(I)$ так, чтобы последняя имела максимально низкое и правое положение при условии, что ни одна из точек не лежит выше и левее ее. Это соответствует минимизации по разбиениям, фигурирующей в (9.6.5). Точки, лежащие на указанной ступенчатой линии, относятся к экстремальным разбиениям. Прочие точки отбрасываются как неэкстремальные. Линия $\tilde{\tilde{V}}(I)$ изображена на рис. 9.4 (в тех местах, где она не совпадает с $\tilde{V}(I)$ пунктиром. Приведем экстремальные разбиения и соответствующие координаты:

Состав разбиений	(1,7)	(2,6)	(3,5)	(1,2,5)	(1,3,4)	(2,3,3)
$I_{бит}$	0,537	0,819	0,954	1,299	1,406	1,562
$V(I)$	0,5	0,75	1	1,125	1,25	1,375

Состав разбиений	(1,1,3,3)	(1,1,1,2,3)	(1,1,1,1,1,3)	(1,1,1,1,1,1,2)
$I_{бит}$	1,81	2,16	2,41	2,75
$V(I)$	1,5	1,625	1,75	1,875

Здесь в верхней строке в скобках для каждого оптимального разбиения $E_1 + \ldots + E_M$ указано число смежных точек, входящих в $E_1, \ldots, E_M$ соответственно.

Линия $\tilde{\tilde{V}}(I)$, как видно на рис. 9.4, располагается между линиями $V(I)$ и $\tilde{V}(I)$, иногда совпадая с $\tilde{V}(I)$. Это полностью согласуется с неравенствами (9.6.6).

В заключение параграфа приведем один просто доказываемый, но важный факт, касающийся всех трех функций ценности информации.

Теорема 9.8. *Функции ценности* $\tilde{V}(I)$, $\tilde{\tilde{V}}(I)$, $V(I)$ *для бейесовской системы* $[P(dx),\ c(x,u)]$ *инвариантны относительно следующего преобразования функции штрафов*:

$$c'(x,u) = c^0(x) + c(x,u), \tag{9.6.7}$$

где $c^0(x)$ — *любая (измеримая) функция с конечным средним значением.*

Для доказательства инвариантности функций $\tilde{V}(I)$, $\tilde{\tilde{V}}(I)$ нужно учесть, что

$$\mathbf{M}\min_u \mathbf{M}\left[c'(x,u)\,|\,E_k\right] = \mathbf{M}\min_u \mathbf{M}\left[c^0(x)\,|\,E_k\right] + \\ + \mathbf{M}\min_u \mathbf{M}\left[c(x,u)\,|\,E_k\right] = \mathbf{M}c^0(x) + \mathbf{M}\min_u \mathbf{M}\left[c(x\ u)\,|\,E_k\right]$$

и (9.6.8)

$$\min_u \mathbf{M}c'(x,u) = \mathbf{M}c^0(x) + \min_u \mathbf{M}c(x,u), \tag{9.6.9}$$

вследствие чего в разности (9.6.3) члены $\mathbf{M}c^0(x)$ сокращаются. Для доказательства инвариантности $V(I)$ достаточно учесть при рассмотрении второго члена в (9.3.12) кроме (9.6.9) соотношение $\mathbf{M}c'(x,u) = \mathbf{M}c^0(x) + \mathbf{M}c(x, u)$.

При замене (9.6.7) остается неизменным также оптимальное разбиение $\sum_u E_k$ или оптимальные переходные вероятности $P(du \mid x)$. Все сказанное выше распространяется и на тот случай, когда функция $c(x, u)$ имеет смысл не штрафов, а поощрений и когда минимизация в (9.6.3) и других формулах должна быть заменена на максимизацию.

9.7. Другой подход к определению ценности шенноновской информации

1. Зададимся целью приблизить по форме определение ценности шенноновского количества информации к определениям (9.2.5), (9.2.6), которые в отличие от (9.3.12) содержат минимизацию по u во втором члене. Для этого дадим другое ее определение при помощи минимизации по u.

Введем модифицированную функцию ценности $\overline{V}(I)$ шенноновского количества информации, которая, как это будет видно из дальнейшего, иногда совпадает с $V(I)$:

$$\overline{V}(I) = V(I). \tag{9.7.1}$$

Пусть задана бейесовская система $[P(dx), c(x, u)]$. Введем вспомогательную случайную величину z, принимающую значения из некоторого пространства Z и связанную с x переходными вероятностями $P(dz \mid x)$. Будем трактовать z как наблюдаемую величину, несущую информацию об x. После наблюдения z оптимальная нерандомизированная оценка $u = d(z)$, как обычно в теории оптимальных статистических решений, определяется минимизацией

$$\mathbf{M}\left[c(x, u) \mid z\right] = \inf_u . \tag{9.7.2}$$

Добавим теперь минимизацию по различным переходным вероятностям $P(dz \mid x)$, которые совместимы с информационным ограничением

$$I_{xz} \leqslant I. \tag{9.7.3}$$

Это приводит к функции ценности информации

$$\overline{V}(I) = \inf_u \mathbf{M} c(x, u) - \inf_{P(dz \mid x) \in \overline{G}} \mathbf{M} \inf_u \mathbf{M}\left[c(x, u) \mid z\right], \tag{9.7.4}$$

где $\overline{G}$ — множество условных распределений $P(dz \mid x)$, удовлетворяющих условию (9.7.3).

Нетрудно понять, что при любом выборе пространства Z выполняется неравенство

$$\overline{V}(I) \leqslant V(I). \qquad (9.7.5)$$

В самом деле, из (9.7.3) следует, что $I_{xu} \leqslant I$, так как $I_{xu} \leqslant I_{xz}$. Поэтому распределение $\int_z P(du \mid z) P(dz \mid x)$ при $P(dz \mid x) \in \overline{G}$ входит в множество G распределений $P(du \mid x)$, перебираемых при минимизации по $P(du \mid x)$ в определении (9.3.12) функции $V(I)$. Указанный факт (9.7.5), разумеется, связан с разобранной ранее теоремой 9.2.

2. Функция ценности (9.7.4) согласно ее определению оказывается зависящей от выбора измеримого пространства Z. Докажем, что его всегда можно подобрать таким образом, чтобы $\overline{V}(I)$ достигло своей верхней границы, т. е. чтобы неравенство (9.7.5) обратилось в равенство (9.7.1).

Доказательство. Допустим, что для рассматриваемой бейесовской системы $[P(dx, c(x, u)]$ экстремум третьей вариационной задачи при условии

$$I_{xu} \leqslant I \qquad (9.7.6)$$

достигается на определенном распределении $P(du \mid x)$, которое обозначим $P_э(du \mid x)$. Это экстремальное распределение задает двумерное распределение $P_э(dx \mid du) = P(dx) P_э(du \mid x)$ и условное распределение $P_э(dx \mid u)$. Почти при каждом u распределение $P_э(dx \mid u)$ есть вероятностная мера на X, т. е. точка $\pi = z$ пространства Π. Таким образом, фиксация условного распределения $P_э(dx \mid u)$ означает фиксацию (определенного почти всюду) нерандомизированного отображения пространства U значений u на Π. Пусть в схеме

$$x \to u \to \pi \qquad (9.7.7)$$

рандомизированное преобразование $x \to u$ осуществляется с вероятностями перехода $P_э(du \mid x)$, а второе преобразование — в соответствии с указанным выше отображением. Очевидно, что $I_{x\pi} \leqslant I_{xu}$ (в силу общих свойств шенноновской информации). Вследствие (9.7.6) отсюда имеем $I_{x\pi} \leqslant I$.

Таким образом, распределение $P_э(d\pi \mid x)$ описанной конкретной схемы преобразований (9.7.7) удовлетворяет условию (9.7.3) при $z = \pi$. Оно входит в множество $\overline{G}$ распределений, перебираемых в (9.7.4), откуда имеем

$$\overline{V}(I) \geqslant \inf_u \mathbf{M} c(x, u) - \mathbf{M} \inf_{u'} \mathbf{M}\left[c(x, u') \mid z = P_э(dx \mid u)\right]. \qquad (9.7.8)$$

Здесь последний член соответствует вышеуказанному конкретному распределению $P_э(dz \mid x)$. Условное математическое ожидание

$\mathbf{M}\,[c\,(x,\ u)\,|\,z=P_э\,(dx\,|\,u)]$ берется с вероятностями

$$\mathbf{P}\left[dx\,|\,z=P_э\,(dx\,|\,u)\right]\equiv\int\limits_U P\,(dx\,|\,z,\,u)\,P\,(du\,|\,z)=$$

$$=\int\limits_U P_э\,(dx\,|\,u)\,P\,(du|\,z)=P_э\,(dx\,|\,u),\qquad(9.7.9)$$

так как $\int zP\,(du\,|\,z)=z\left(\text{где } z=P_э\,(dx\,|\,u)\right)$.

Вследствие (9.4.21) имеем

$$c\,(x,\ u)=-\frac{\gamma\,(x)}{\beta}-\frac{1}{\beta}\ln\frac{P_э\,(dx\,|\,u)}{P\ (dx)},\qquad(9.7.10)$$

поэтому

$$\inf_{u'}\mathbf{M}\,\left[c\,(x,\,u')\,|\,z\right]=-\frac{1}{\beta}\,\mathbf{M}\left[\gamma\,(x)\,|\,z\right]+$$

$$+\inf_{u'}\mathbf{M}\left[-\frac{1}{\beta}\ln\frac{P_э\,(dx\,|\,u')}{P\,(dx)}\,\middle|\,z\right].\qquad(9.7.11)$$

В силу самого понятия inf имеем

$$\inf_{u'}\mathbf{M}\left[-\frac{1}{\beta}\ln\frac{P_э\,(dx\,|\,u')}{P\,(dx)}\,\middle|\,z=P_э\,(dx\,|\,u)\right]\leqslant$$

$$\leqslant\mathbf{M}\left[-\frac{1}{\beta}\ln\frac{P_э\,(dx\,|\,u'')}{P_э\,(dx)}\,\middle|\,z=P_э\,(dx\,|\,u)\right]\qquad(9.7.12)$$

при любом u''. Возьмем в качестве u'' прообраз точки z (по указанному выше отображению $u\to z$), т. е. одну из тех точек u'', для которых $P_э\,(dx\,|\,u'')=z=P_э\,(dx\,|\,u)$.

Тогда неравенство (9.7.12) примет вид

$$\inf_{u'}\mathbf{M}\left[-\frac{1}{\beta}\ln\frac{P_э\,(dx\,|\,u')}{P\,(dx)}\,\middle|\,z=P_э\,(dx\,|\,u)\right]\leqslant$$

$$\leqslant\mathbf{M}\left[-\frac{1}{\beta}\ln\frac{P_э\,(dx\,|\,u)}{P\,(dx)}\,\middle|\,z=P_э\,(dx\,|\,u)\right].\qquad(9.7.13)$$

Здесь в правую часть подставлено $P_э\,(dx\,|\,u)$ вместо $P_э\,(dx\,|\,u'')$. Подставляя далее (9.7.13) в (9.7.11) и усредняя по z, получаем

$$\mathbf{M}\inf_{u'}\mathbf{M}\left[c\,(x,\,u')\,|\,z\right]\leqslant\mathbf{M}\left[-\frac{1}{\beta}\,\gamma\,(x)\right]+$$

$$+\mathbf{M}\left[-\frac{1}{\beta}\ln\frac{P_э\,(dx\,|\,u)}{P\,(dx)}\right]=\int c\,(x,\ u)\,P_э\,(dx,\ du)$$

[опять использовано (9.7.10)]. Это соотношение позволяет преобразовать (9.7.8) к виду

$$\overline{V}\,(I)\geqslant\inf\mathbf{M}c\,(x,\ u)-\int c\,(x,\ u)\,P_э\,(dx,\ du).$$

Выражение в правой части здесь есть не что иное как ценность $V(I)$ шенноновской информации. Сравнение полученного неравенства $\overline{V}(I) \geqslant V(I)$ с (9.7.5) дает (9.7.1). Доказательство закончено.

3. Вследствие теоремы 9.9 и формулы (9.7.4) ценность шенноновского количества информации задается выражением

$$V(I)=\inf_u \int c(x, u) P(dx) - \inf_{P(d\pi \mid x) \in G} \int P(d\pi) \inf_u \int c(x, u) \times$$
$$\times P(dx \mid \pi), \tag{9.7.14}$$

где $P(d\pi)=\int_X P(dx) P(d\pi \mid x)$; $P(dx \mid \pi)=P(dx) P(d\pi \mid x) / P(d\pi)$; $\pi \equiv \pi(dx)$ — точка пространства распределений Π. При этом множество G условных распределений ограничено неравенством $I_{x\pi} \leqslant I$ (9.7.3).

Покажем теперь, что функции ценности $\widetilde{V}(I)$, $\widetilde{\widetilde{V}}(I)$, в свою очередь, могут быть определены по формуле (9.7.14), но при замене допустимого множества G на какие-то более узкие множества $\widetilde{G}$, $\widetilde{\widetilde{G}}$ соответственно, такие, что $G \supset \widetilde{\widetilde{G}} \supset \widetilde{G}$.

Для фиксированного разбиения $\sum_k E_k=X$, входящего в определения (9.2.6), (9.6.5), целесообразно положить

$$P(d\pi \mid x)=\sum_k P(E_k \mid x) \delta\left(d\pi, P(dx \mid E_k)\right), \tag{9.7.15}$$

где $\delta(d\pi, P(dx \mid E_k))$ — вероятностная мера в пространстве Π, сконцентрированная в точке $\pi(dx)=P(dx \mid E_k)$. Из (9.7.15) имеем

$$P(d\pi\, dx)=\sum_k P(E_k \mid x) P(dx) \delta\left(d\pi, P(dx \mid E_k)\right), \tag{9.7.16}$$

$$P(d\pi)=\sum_k P(E_k) \delta\left(d\pi, P(dx \mid E_k)\right), \tag{9.7.17}$$

Как видно из выражения в правой части (9.7.15), событие $x \in E_{k'}$ с вероятностью 1 влечет событие $\pi(dx)=P(dx \mid E_{k'})$ и наоборот. Поэтому

$$P(dx \mid \pi)=P(dx \mid E_{k'}) \tag{9.7.18}$$
при $\pi(dx)=P(dx \mid E_{k'})$, т. е. при $x \in E_{k'}$.

Для множества распределений типа (9.7.15) согласно (9.7.17) интеграл по π в (9.7.14) превращается в сумму и мы имеем

$$V=\inf_u \int c(x, u) P(dx) - \inf_{P(d\pi \mid x)} \sum_k P(E_k) \times$$
$$\times \inf_u \int c(x, u) P\left(dx \mid \pi=P(dx \mid E_k)\right).$$

Далее вследствие (9.7.18) это выражение принимает вид

$$V=\inf_u \int c(x, u) P(dx) - \inf_{P(d\pi \mid x)} \sum_k P(E_k) \inf_u \int c(x, u) P(dx \mid E_k).$$

При этом для множества распределений типа (9.7.15) минимизация по $P(d\pi \mid x)$ сводится к минимизации по разбиениям $\sum_k E_k = X$. Полученное выражение совпадает с выражением, стоящим в правой части формул (9.2.6а), (9.6.5). Следовательно, ценности информации $\widetilde{V}(I)$, $\widetilde{\widetilde{V}}(I)$ также могут быть определены по формуле (9.7.14), если при минимизации перебираются распределения вида (9.7.15). Множество $\widetilde{\widetilde{G}}$, соответствующее определению ценности $\widetilde{\widetilde{V}}(I)$, есть множество тех распределений типа (9.7.15), для которых выполнено неравенство (9.6.4).

Вследствие указанной эквивалентности событий $x \in E_k$ и $\pi(dx) = P(dx \mid E_k)$ преобразование $x \to \pi$ сводится к преобразованию $x \to k$. Это значит, что в данном случае

$$I_{x\pi} = H_k = -\sum_k P(E_k) \ln P(E_k).$$

Поэтому из (9.6.4) вытекает условие $I_{x\pi} \leqslant I$, т. е. условие (9.7.3). Следовательно, $\widetilde{\widetilde{G}}$ принадлежит множеству G. Принадлежность же $\widetilde{G}$ множеству $\widetilde{\widetilde{G}}$ вытекает из того факта, что условие (9.6.4) более слабое, чем условие ограниченности числа областей $E_1, \ldots, E_M$, $M = e^I$.

Глава 10

ЦЕННОСТЬ ШЕННОНОВСКОЙ ИНФОРМАЦИИ ДЛЯ ВАЖНЕЙШИХ БЕЙЕСОВСКИХ СИСТЕМ

В данной главе общая теория, касающаяся ценности шенноновского количества информации и изложенная в предыдущей главе, применяется к ряду важных частных случаев бейесовских систем. Для этих систем здесь получены явные выражения, определяющие потенциал $\Gamma(\beta)$, который позволяет в параметрической форме найти зависимость между потерями (риском) R и количеством информации I и затем найти функцию ценности $V(I)$.

Вначале рассматриваются те бейесовские системы, для которых пространство X является особенно простым, а именно состоящим из двух точек. При этом для решения третьей вариационной задачи требуется решать лишь несложное алгебраическое уравнение с одним неизвестным. В случае систем с однородной функцией штрафов (§ 10.2) для получения решения может быть применен метод преобразования Фурье или эквивалентный ему операторный метод.

Другие своеобразные (матричные) методы применяются в важном случае гауссовых бейесовских систем, характеризующихся гауссовым априорным распределением и билинейной функцией штрафов. Они позволяют получить решение задачи и рассмотреть зависимость активного подпространства $\tilde{U}$ от параметра β или I. Особо рассмотрены различные (конечномерные и бесконечномерные) стационарные гауссовы системы, для которых функция ценности информации записывается в параметрической форме.

10.1. Система с двумя состояниями

1. Рассмотрим тот простой случай, когда пространство X значений случайной величины x состоит из двух точек x_1 и x_2. Вероятности $P(x_1) = p$, $P(x_2) = 1 - p$ предполагаются заданными. Пространство U возможных оценок u предполагаем более сложным, например возьмем в качестве него действительную ось. При этом функция штрафов $c(x, u)$ сведется к двум функциям от u:

$$c(x_1, u) = c_1(u); \qquad c(x_2, u) = c_2(u).$$

Уравнение (9.5.2) для данной бейесовской системы будет иметь вид

$$\sum_{i=1}^{2} Q_i \, e^{-\beta c_i(u)} = 1, \qquad u \in \widetilde{U}. \tag{10.1.1}$$

Эти соотношения представляют собой систему уравнений относительно двух неизвестных Q_1, Q_2. В предположении невырожденности матрицы $\| e^{-\beta c_i(u_j)} \|$ указанные неизвестные определяются двумя уравнениями

$$Q_1 e^{-\beta c_1(u_1)} + Q_2 e^{-\beta c_2(u_1)} = 1, \qquad Q_1 e^{-\beta c_1(u_2)} + Q_2 e^{-\beta c_2(u_2)} = 1, \tag{10.1.2}$$

т. е. пространство $\widetilde{U}$ состоит из двух точек u_1, u_2.

Рассматриваемая бейесовская система относится к тому частному случаю, который был рассмотрен в п. 1 § 9.5. Матрица (9.5.3) при этом имеет вид

$$\| b_{ij} \| = \begin{pmatrix} e^{-\beta c_1(u_1)} & e^{-\beta c_2(u_1)} \\ e^{-\beta c_1(u_2)} & e^{-\beta c_2(u_2)} \end{pmatrix}^{-1} = [e^{-\beta c_1(u_1) - \beta c_2(u_2)} -$$

$$- e^{-\beta c_2(u_1) - \beta c_1(u_2)}]^{-1} \begin{pmatrix} e^{-\beta c_2(u_2)} & -e^{-\beta c_2(u_1)} \\ -e^{-\beta c_1(u_2)} & e^{-\beta c_1(u_1)} \end{pmatrix}.$$

Используя (9.5.4), находим потенциал

$$\Gamma(\beta) = -h_2(p) - p \ln [e^{-\beta c_2(u_2)} - e^{-\beta c_2(u_1)}] -$$
$$-(1-p) \ln [e^{-\beta c_1(u_1)} - e^{-\beta c_1(u_2)}] +$$
$$+ \ln [e^{-\beta c_1(u_1) - \beta c_2(u_2)} - e^{-\beta c_2(u_1) - \beta c_1(u_2)}]$$

или

$$\Gamma(\beta) = -h_2(p) - p \ln \operatorname{sh} \beta d_2 - (1-p) \ln \operatorname{sh} \beta d_1 +$$
$$+ \ln \operatorname{sh}(\beta d_1 + \beta d_2) - \beta M, \tag{10.1.3}$$

где

$$2d_1 = c_1(u_2) - c_1(u_1); \qquad 2d_2 = c_2(u_1) - c_2(u_2); \tag{10.1.4}$$

$$2M = p\left[c_1(u_1) + c_1(u_2)\right] + (1-p)\left[c_2(u_1) + c_2(u_2)\right]. \tag{10.1.5}$$

Дифференцируя потенциал по β (при постоянных u_1, u_2), можно найти R и I. При помощи преобразования Лежандра

$$I_2(R') = -\Gamma_2 - \beta R' \qquad \left(R' = -\frac{d\Gamma_2}{d\beta}\right)$$

функции

$$\Gamma_2(\beta) = -p \ln \operatorname{sh} \beta d_2 - (1-p) \ln \operatorname{sh} \beta d_1 + \ln\left[\operatorname{sh}(\beta d_1 + \beta d_2)\right]$$

результат можно записать

$$I = h_2(p) + I_2(R - M). \tag{10.1.6}$$

Неизвестные значения u_1, u_2 должны быть определены из дополнительных уравнений. Чтобы записать эти уравнения, используем условие (9.4.27), соответствующее вариации $\delta\widetilde{U}$ области $\widetilde{U}$, т. е. вариациям δu_1, δu_2. В силу (9.4.29) вариацию $\delta\Gamma$ можно записать

$$\delta\Gamma = \frac{d\Gamma}{d\beta}\delta\beta + \delta_0\Gamma = -R\delta\beta + \delta_0\Gamma, \tag{10.1.7}$$

где вариация $\delta_0\Gamma$ соответствует вариации области $\widetilde{U}$, проводимой без вариации параметра β. Подставляя (10.1.7) в (9.4.27), получаем условие

$$\delta_0\Gamma = 0. \tag{10.1.8}$$

Поскольку вариации δu_1, δu_2 независимы, условие (10.1.8) дает два уравнения

$$\frac{\partial\Gamma}{\partial u_1} = 0; \qquad \frac{\partial\Gamma}{\partial u_2} = 0. \tag{10.1.9}$$

Предполагая, что якобиан

$$\begin{vmatrix} \dfrac{\partial d_1}{\partial u_1} & \dfrac{\partial d_1}{\partial u_2} \\ \dfrac{\partial d_2}{\partial u_1} & \dfrac{\partial d_2}{\partial u_2} \end{vmatrix}$$

преобразования (10.1.4) отличен от нуля, уравнения (10.1.9) можно заменить уравнениями

$$\frac{\partial\Gamma}{\partial d_1} = 0; \qquad \frac{\partial\Gamma}{\partial d_2} = 0. \tag{10.1.10}$$

Дифференцируя (10.1.4), записываем их в явном виде

$$(1-p)\operatorname{cth}\beta d_1 - \operatorname{cth}\beta(d_1+d_2) + \frac{\partial M}{\partial d_1} = 0,$$
$$p\operatorname{cth}\beta d_2 - \operatorname{cth}\beta(d_1+d_2) + \frac{\partial M}{\partial d_2} = 0. \tag{10.1.11}$$

2. Рассмотрим для примера квадратичные функции штрафов

$$c_1(u) = (u-a_1)^2 + b_1; \qquad c_2(u) = (u-a_2)^2 + b_2. \tag{10.1.12}$$

Вводя центрированную переменную $\widetilde{u} = u - (a_1 + a_2)/2$, приведем их к виду

$$c_1(\widetilde{u}) = (\widetilde{u} - (a_1-a_2)/2)^2 + b_1; \qquad c_2(\widetilde{u}) = (\widetilde{u} + (a_1-a_2)/2)^2 + b_2.$$

В силу инвариантности относительно преобразования (9.6.7) эти функции можно заменить функциями

$$c_1(\widetilde{u}) = \widetilde{u}^2 - (a_1-a_2)\widetilde{u}; \qquad c_2(\widetilde{u}) = \widetilde{u}^2 + (a_1-a_2)\widetilde{u}. \tag{10.1.13}$$

В данном случае согласно (10.1.4)

$$
\begin{aligned}
&2d_1 = \tilde{u}_2^2 - \tilde{u}_1^2 - (a_1 - a_2)\left(\tilde{u}_2 - \tilde{u}_1\right),\\
&2d_2 = \tilde{u}_1^2 - \tilde{u}_2^2 + (a_1 - a_2)\left(\tilde{u}_1 - \tilde{u}_2\right),\\
&2M = \tilde{u}_1^2 + \tilde{u}_2^2 + (1 - 2p)(a_1 - a_2)\left(\tilde{u}_2 + \tilde{u}_1\right).
\end{aligned}
\tag{10.1.14}
$$

Значения $\tilde{u}_1$, $\tilde{u}_2$ должны быть определены из уравнений (10.1.9), (10.1.10). Вследствие симметрии функций (10.1.13) можно искать симметричные корни $\tilde{u}_2 = -\tilde{u}_1$, что сильно упрощает уравнения. Так функция (10.1.3) вследствие (10.1.14) принимает вид

$$
\begin{aligned}
\Gamma(\beta) &= -h_2(p) - \ln \operatorname{sh} \beta d_1 + \ln \operatorname{sh} 2\beta d_1 - \beta M =\\
&= -h_2(p) + \ln(2 \operatorname{ch} \beta d_1) - \beta M,
\end{aligned}
\tag{10.1.15}
$$

где $d_1 = d_2 = (a_1 - a_2)\tilde{u}_1$; $\quad M = \tilde{u}_1^2 = d_1^2/(a_1 - a_2)^2$.

Приравнивая нулю частную производную функции (10.1.15) по d_1, получаем уравнение для определения $\tilde{u}_1$

$$
(a_1 - a_2)^2 \operatorname{th} \beta d_1 = 2d_1. \tag{10.1.16}
$$

Это трансцендентное уравнение имеет единственное положительное решение при $(a_1 - a_2)^2\beta \geqslant 2$. После решения уравнения (10.1.16) определяется $\tilde{u}_1 = d_1/(a_1 - a_2)$.

Взяв частную производную от (10.1.15) по β, обычным способом получаем (учитывая (10.1.16))

$$
\begin{aligned}
&R = -d_1 \operatorname{th} \beta d_1 + M = -\frac{2d_1^2}{(a_1 - a_2)^2} + M = -\tilde{u}_1^2,\\
&I = h_2(p) - \ln(2 \operatorname{ch} \beta d_1) + 2\beta \tilde{u}_1^2.
\end{aligned}
$$

Удобно ввести параметр $\vartheta = 2\tilde{u}_1/(a_1 - a_2) = \operatorname{th} \beta d_1$, при помощи которого последние формулы запишутся так:

$$
R = -(a_1 - a_2)^2 \vartheta^2/4 \;\; (1/2\beta)\, \vartheta \operatorname{arth} \vartheta, \tag{10.1.17}
$$

$$
I = h_2(p) - \ln 2 + (1/2)\ln(1 - \vartheta^2) + \vartheta \operatorname{arth} \vartheta. \tag{10.1.18}
$$

Выражая $\operatorname{arth}\vartheta$ через логарифмы, последнее выражение можно представить в форме

$$
\begin{aligned}
I &= h_2(p) - \ln 2 + \frac{1}{2}(1 + \vartheta)\ln(1 + \vartheta) +\\
&+ \frac{1}{2}(1 - \vartheta)\ln(1 - \vartheta) = h_2(p) - h_2\left(\frac{1 + \vartheta}{2}\right).
\end{aligned}
\tag{10.1.19}
$$

Отсюда видно, что нулевая информация $I = 0$ имеет место при значении $(1 + \vartheta)/2 = p$, $\vartheta = 2p - 1$. Согласно (10.1.17) этому значению соответствуют потери $R(0) = -(a_1 - a_2)^2 (p - {}^1/_2)^2$, так что

$$
V(I) = R(0) - R(I) = \frac{(a_1 - a_2)^2}{4}\left[\vartheta^2 - (2p - 1)^2\right]. \tag{10.1.20}
$$

Исключая параметр ϑ из (10.1.19), (10.1.20), находим зависимость между ценностью V и количеством информации I в виде уравнения

$$I = h_2(p) - h_2\left(\frac{1}{2} + \sqrt{\frac{V}{(a_1 - a_2)^2} + \left(p - \frac{1}{2}\right)^2}\right). \qquad (10.1.21)$$

При изменении количества информации I от 0 до $h_2(p)$ ценность V меняется от 0 до $(a_1 - a_2)^2 \times [{}^1/_4 - (p - {}^1/_2)^2] = (a_1 - a_2)^2\, p\, (1-p)$ (при этом ϑ меняется от $2p - 1$ до 1).

10.2. Системы с однородной функцией штрафов

1. В этом параграфе будем предполагать, что функция штрафов $c(x, u)$ бейесовской системы $[P(x), c(x, u)]$ является однородной, т. е. зависит от x и u лишь через посредство их разности: $c(x, u) = c(x - u) = c(z)$, другими словами, является инвариантной относительно преобразования сдвига $x \to x + a$, $u \to u + a$. При этом подразумевается, что x, u, $z = x - u$, a принимают значения из одного и того же пространства X, что преобразование сдвига оставляет это пространство инвариантным. Следовательно, пространство X не должно иметь границ, но может иметь конечный «объем», например, быть периодически замкнутым.

Применим к равенству (9.5.2), принимающему вид

$$\sum_x Q(x)\, e^{-\beta c(x-u)} = 1, \qquad (10.2.1)$$

указанное преобразование сдвига, которое перейдет в равенство

$$\sum_x Q(x+a)\, e^{-\beta c(x-u)} = 1. \qquad (10.2.2)$$

Предполагая, что преобразование сдвига оставляет активное подпространство $\widetilde{U}$ инвариантным, получаем из сравнения (10.2.1) и (10.2.2), что $Q(x) = Q(x + a)$, т. е. $Q(x) = Q$ является константой. Поэтому в данном случае применима теория, изложенная в п. 2 § 9.5, и можно пользоваться формулами (9.5.8), (9.5.10), (9.5.15). Зная потенциал $\Gamma = -H_x - \beta F_0$, в соответствии с (9.4.28), (9.4.29) нетрудно найти

$$R = F_0 - T\frac{dF_0}{dT} = \frac{d}{d\beta}(\beta F_0); \quad I = H_x + \frac{dF_0}{dT}. \qquad (10.2.3)$$

Функция ценности информации получается из этих двух формул исключением параметра $T = 1/\beta$.

Поскольку в рассматриваемом случае функция $Q(x)$ является константой, экстремальное распределение (9.4.21), согласно (9.5.9) принимает вид

$$P_э(x, u) = P_э(u)\, e^{\beta F_0 - \beta c(x-u)}. \qquad (10.2.4)$$

Отсюда видно, что условное распределение

$$P_э(x \mid u) = e^{\beta F_0 - \beta c(x-u)} \tag{10.2.5}$$

не зависит от априорного распределения $P(x)$. Величина $H = -dF_0/dT = H_x - I$ в формулах (10.2.3) есть не что иное как энтропия $H_{x|u}$ этого условного распределения: $H = -\mathbf{M} \ln P(x \mid u)$.

Априорное распределение $P(x)$ оказывает влияние в (10.2.4) лишь на распределение $P_э(u)$. Уравнение (9.4.23), определяющее его, можно в силу (9.5.9) записать

$$\sum_u e^{-\beta c(x-u)} P_э(u) = e^{-\beta F_0} P(x). \tag{10.2.6}$$

Оно решается методом преобразования Фурье, поскольку Фурье-изображение интеграла свертки равно произведению изображений. Вводя характеристические функции

$$\theta(s) = e^{\mu(s)} = \sum e^{sx} P(x); \qquad \theta_э(s) = e^{\nu(s)} = \sum e^{su} P_э(u),$$

из уравнения (10.2.6) получаем

$$\theta_э(s) = e^{-\psi(\beta, s) - \beta F_0} \theta(s)$$

или

$$\nu(s) = \mu(s) - [\psi(\beta, s) - \psi(\beta, 0)], \tag{10.2.7}$$

где

$$e^{\psi(\beta, s)} = \sum_z e^{-\beta c(z) + sz}, \tag{10.2.8}$$

причем

$$\psi(\beta, 0) = -\beta F_0(1/\beta). \tag{10.2.9}$$

Соотношение (10.2.6) можно в силу (10.2.8) записать также в операторной форме

$$\exp\left[\psi\left(\beta, \frac{d}{dx}\right) - \psi(\beta, 0)\right] P_э(x) = P(x) \tag{10.2.10}$$

по аналогии с (8.8.4). Разрешая последнее уравнение, получаем

$$P_э(x) = \exp\left[-\psi\left(\beta, \frac{d}{dx}\right) + \psi(\beta, 0)\right] P(x). \tag{10.2.11}$$

Взяв разложение

$$\exp\left[-\psi\left(\beta, \frac{d}{dx}\right) + \psi(\beta, 0)\right] = \\ = 1 + \sum_{n=1}^{\infty} \frac{1}{n!}\left[\psi(\beta, 0) - \psi\left(\beta, \frac{d}{dx}\right)\right]^n,$$

а также разложение

$$\psi\left(\beta, \frac{d}{dx}\right)-\psi(\beta, 0)=\sum_{l=1}^{\infty}\frac{1}{l!}\frac{\partial^l\psi}{\partial s^l}(\beta, 0)\frac{d^l}{dx^l}$$

можно записать (10.2.11) в виде

$$P_э(x)=P(x)+\sum_{n=1}^{\infty}\sum_{l_1,\ldots,l_n=1}^{\infty}\frac{(-1)^n}{n!\,l_1!\ldots!\,l_n!},$$

$$k_{l_1}\ldots k_{l_n}\frac{d^{l_1+\ldots+l_n}}{dx^{l_1+\ldots+l_n}}P(x)=P(x)-k_1\frac{dP(x)}{dx}+$$

$$+\frac{1}{2}[k_1^2-k_2]\frac{d^2P(x)}{dx^2}+\ldots \qquad (10.2.12)$$

Здесь производные $k_l = (\partial^l\psi/\partial s^l)(\beta, 0)$ согласно (10.2.5), (10.2.8) совпадают с кумулянтами распределения $P(z) = \exp[\beta F_0 - \beta c(z)]$. Полученной формулой (10.2.12) удобно пользоваться, когда априорное распределение $P(x)$ является значительно более широким, чем условное распределение (10.2.5), так что члены разложения быстро убывают.

Особенно прост тот случай, когда распределение $P(x)$ является равномерным:

$$P(x)=e^{-H_x}.$$

Тогда из (10.2.6) вытекает, что $P_э(u)$ имеет такой же вид:

$$P_э(u)=e^{-H_x}. \qquad (10.2.13)$$

Приведенные в этом параграфе соотношения справедливы как для дискретного, так и для непрерывного пространства X. В непрерывном случае, конечно, суммы следует заменить интегралами. Может оказаться, что в случае непрерывного или неограниченного пространства некоторые члены рассмотренных выражений, например H_x, не имеют самостоятельного смысла. Однако они встречаются в сочетании с другими членами и функциями в такой форме, что сочетание в целом (например, сумма $H_x + dF_0/dT$) имеет смысл.

2. Примеры. Рассмотренные в § 9.2 и 9.5 примеры, в которых x было дискретной случайной величиной, относятся именно к случаю систем с однородной функцией штрафа.

В настоящем параграфе мы рассмотрим примеры, в которых случайная величина x принимает непрерывные значения и описывается плотностью распределения вероятностей $p(x)$.

Пусть функция штрафов $c(z)$ имеет нижеследующий вид:

$$c(z)=\begin{cases}0 & \text{при } z\in Z_0,\\ \infty & \text{вне } Z_0.\end{cases} \qquad (10.2.14)$$

Тогда, записывая формулу (9.5.8) в непрерывном варианте:

$$e^{-\beta F_0} = \int e^{-\beta c(z)}\, dz = \int\limits_{Z_0} dz,$$

будем иметь

$$\beta F_0 = -\ln \Omega,$$

где $\Omega = \int\limits_{Z_0} dz$ — «объем» той области Z_0, в которой $c(z) \neq \infty$.

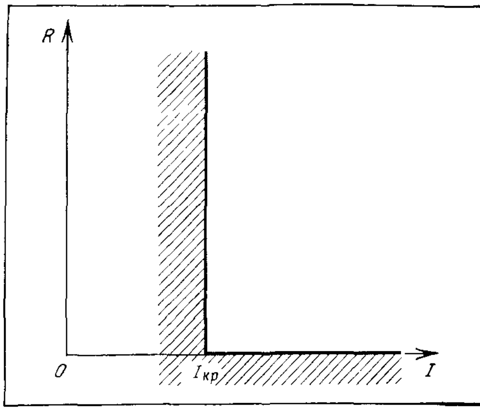


Рис. 10.1. Зависимость средних штрафов от количества информации для функции штрафов вида (10.2.14).

Применяя формулы (10.2.3) к данному случаю, находим

$$R = \frac{d}{d\beta}(\beta F) = 0; \quad I = H_x - \ln \Omega. \tag{10.2.15}$$

Отсюда видно, что для обеспечения нулевого уровня потерь необходимо количество информации, равное

$$I_{кр} = H_x - \ln \Omega = -\int \ln [\Omega p(x)]\, p(x)\, dx. \tag{10.2.16}$$

При меньшем количестве информации возникает ненулевая вероятность бесконечных штрафов и поэтому уровень потерь $R = \mathbf{M} c(x, u)$ становится равным бесконечности. Ход описанной специфической зависимости R от I показан на рис. 10.1.

В многомерном случае, когда задана точность воспроизведения (равная ε) по каждой координате, имеем

$$\Omega = \varepsilon^r,$$

где r — размерность пространства. При этом критическое значение информации (10.2.16) равно

$$I = r \ln \frac{1}{\varepsilon} + H_x. \tag{10.2.17}$$

3. Пусть теперь пространство X есть r-мерное линейное пространство, а функция $c(z)$ квадратична:

$$c(z)=\sum_{i=1}^{r} z_i^2. \tag{10.2.18}$$

Применяя (9.5.8), находим потенциал

$$\beta F_0 = -r \ln \int e^{-\beta z^2} dz = \frac{r}{2} \ln \frac{\beta}{\pi}.$$

Согласно (10.2.3)

$$R = rT/2; \quad I = H_x - (r/2)\ln(\pi e T). \tag{10.2.19}$$

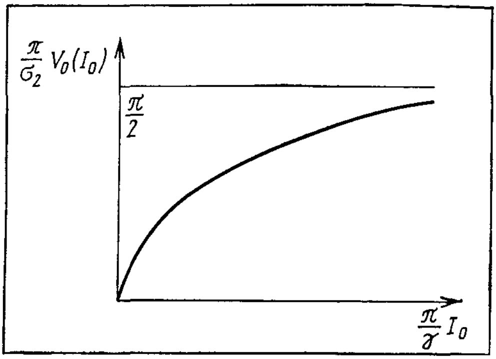


Рис. 10.2. Функции ценности информации для квадратичной функции штрафов (10.2.20).

Если $\varepsilon^2 = \mathbf{M} z_i^2$ — точность по каждой координате, то уровень потерь $R = \mathbf{M} c(z)$ можно в силу (10.2.18) приравнять $r\varepsilon^2$. Тогда формулы (10.2.19) дадут

$$I = r \ln \frac{1}{\varepsilon \sqrt{2\pi e}} + H_x,$$

что напоминает (10.2.17). Исключая параметр T из (10.2.19), находим функцию ценности информации

$$V(I) = \frac{r}{2\pi} e^{\frac{2}{r} H_x - 1} \left(1 - e^{-\frac{2}{r} I}\right), \tag{10.2.20}$$

показанную на рис. 10.2.

Запишем также плотность распределения $P_\beta(u)$, определяемую формулой (10.2.11). В данном случае распределение (10.2.5) является гауссовым

$$p_\beta(x \mid u) = \left(\frac{\beta}{\pi}\right)^r e^{-\beta \sum_i (x_i - u_i)^2},$$

и имеет характеристическую функцию

$$\theta_z(s) = \exp[\psi(\beta, s) - \psi(\beta, 0)] =$$

$$= \mathbf{M} \exp\left[\sum_i s_i (x_i - u_i)\right] = \exp\left[\frac{1}{4\beta} \sum_i s_i^2\right].$$

Поэтому (10.2.11) дает

$$p_э(x) = \exp\left[-\frac{1}{4\beta} \sum_i \frac{\partial^2}{\partial x_i^2}\right] p(x).$$

Члены разложения

$$\sum_{k=0}^{\infty} \frac{1}{k!} \left(-\frac{1}{4\beta} \sum_i \frac{\partial^2}{\partial x_i^2}\right)^k p(x) = p_э(x)$$

сходятся как степенной ряд отношения

$$\left(\frac{1}{\beta\Delta^2}\right)^2 = 4\left(\frac{\varepsilon}{\Delta}\right)^4,$$

где $\quad \Delta \sim \frac{d^n p(x)}{dx^n} \Big/ \frac{d^{n+1} p(x)}{dx^{n+1}}.$

Если это отношение мало, то можно оставить лишь несколько членов разложения, например

$$p_э(x) = p(x) - \frac{T}{4} \sum_i \frac{\partial^2}{\partial x_i^2} p(x) + O\left(\frac{\varepsilon^4}{\Delta^4}\right). \qquad (10.2.21)$$

Аналогичная формула имеет место и для предыдущего примера (10.2.14), когда

$$p_э(x \mid u) = \begin{cases} 1/\Omega & \text{при } x - u \in Z_0, \\ 0 & \text{вне } Z_0. \end{cases}$$

Именно, если

$$(k_1)_i = \int_{Z_0} z_i \, dz = 0,$$

то из разложения типа (10.2.12) будем иметь

$$p_э(x) = p(x) - \frac{1}{2} \sum_{i,j} B_{ij} \frac{\partial^2}{\partial x_i \, \partial x_j} p(x) + \ldots,$$

$$B_{ij} = \frac{1}{\Omega} \int_{Z_0} z_i z_j \, dz \qquad (B_{ij} \sim \varepsilon^2).$$

Приведенные результаты справедливы в том предположении, что вероятности (10.2.11) являются неотрицательными и энтропия H_x априорного распределения не меньше условной энтропии $H_{x|u} = H_z$ распределения (10.2.5), получаемого в результате решения экстремальной задачи.

4. В заключение этого параграфа рассмотрим функцию

$$\mu_x(t) = \ln \sum_u e^{tc(x,u)} P_э(u), \qquad (10.2.22)$$

которая, как будет видно из гл. 11, играет существенную роль при решении вопроса о том, насколько отличаются друг от друга функции ценности шенноновского и хартлиевского количества информации.

В случае однородной функции штрафа выражение

$$e^{\mu_x(t)} = \sum_u e^{tc(x-u)} P_э(u)$$

можно представить в следующей операторной форме

$$e^{\mu_x(t)} = \exp\left[\psi\left(-t, \frac{d}{dx}\right)\right] P_э(x), \tag{10.2.23}$$

что аналогично (10.2.10) (функция $\psi(\beta, s)$ определена формулой (10.2.8)). Подставляя сюда (10.2.11), будем иметь

$$e^{\mu_x(t)} = \exp\left[\psi\left(-t, \frac{d}{dx}\right) - \psi\left(\beta, \frac{d}{dx}\right) + \psi(\beta, 0)\right]. \tag{10.2.24}$$

В том простом частном случае, когда распределение $P_э(u)$ равномерное (10.2.13), формула (10.2.23) дает

$$e^{\mu_x(t)} = e^{\psi(-t, 0)} P_э = e^{\psi(-t, 0) - H_x},$$
$$\mu_x(t) = \psi(-t, 0) - H_x$$

или

$$\mu_x(t) = -H_x + tF_0\left(-\frac{1}{t}\right) = \Gamma(-t), \tag{10.2.25}$$

если учесть (10.2.9), (9.5.10). Функция $\mu_x(t)$ при этом не зависит от x.

10.3. Гауссовы бейесовские системы

1. Пусть x и u есть точки эвклидовых пространств R_r, R_s, имеющих размерность r и s соответственно. Бейесовская система $[p(x), c(x, u)]$ называется гауссовой, если распределение $p(x)$ гауссово, а $c(x, u)$ представляет собой сумму линейной и квадратичной форм от x и u. Выбирая соответствующим образом начало координат в R_s, будем, следовательно, иметь

$$c(x, u) = c^0(x) + \mathrm{x}^T \mathrm{g} \mathrm{u} + \frac{1}{2} \mathrm{u}^T \mathrm{h} \mathrm{u}, \tag{10.3.1}$$

(где $c^0(x) = c_0 + \mathrm{c}_1^T \mathrm{x} + \frac{1}{2} \mathrm{x}^T \mathrm{c}_2 \mathrm{x}$, что однако, не является обязательным). Здесь использована матричная форма записи:

$$\mathrm{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix}; \quad \mathrm{x}^T = (x_1, \dots, x_r); \quad \mathrm{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_s \end{pmatrix}; \quad \mathrm{u}^T = (u_1, \dots, u_s),$$

g, h — матрицы. Далее согласно определению гауссовой системы при подходящем выборе начала координат

$$p(x)=\det{}^{1/2}\frac{a}{2\pi}\exp\left\{-\frac{1}{2}x^T ax\right\}. \tag{10.3.2}$$

Матрица a — невырожденная, положительно определенная. Если распределение $p(x)$ является вырожденным, т. е. сосредоточено в некотором подпространстве $\widetilde{X}\subset R_r$, то можно ограничиться рассмотрением пространства $\widetilde{X}=R_{\widetilde{r}}$, заменив r на $\widetilde{r}\,(<r)$. Таким образом, без ограничения общности корреляционную матрицу $k_x=a^{-1}$ распределения (10.3.2) можно считать невырожденной.

Для гауссовых систем функцию $q(x)$ (см. (9.5.1)) можно искать в следующей гауссовой форме:

$$p(x)e^{-\gamma(x)-\beta c^0(x)}=q(x)e^{-\beta c^0(x)}=\exp\left(\sigma+x^T\rho-\frac{1}{2}x^T\varkappa x\right). \tag{10.3.3}$$

Аналогично распределение $p_э(u)$ полагаем гауссовым

$$p_э(u)=\det{}^{1/2}\left(\frac{a_u}{2\pi}\right)\exp\left\{-\frac{1}{2}(u^T-m^T)a_u(u-m)\right\},\qquad u\in\widetilde{U} \tag{10.3.4}$$

и сосредоточенным в некотором подпространстве $\widetilde{U}=R_{\widetilde{s}}$ (размерностью $\widetilde{s}$) пространства R_s. В частности, может иметь место совпадение $R_{\widetilde{s}}=R_r$. В дальнейшем, ограничиваясь рассмотрением пространства $R_{\widetilde{s}}$, мы и под матрицами h, g будем понимать $\widetilde{s}\times\widetilde{s}$- и $\widetilde{r}\times\widetilde{s}$-матрицы. Итак, рассматриваются такие пространства $R_{\widetilde{r}}$, $R_{\widetilde{s}}$ значений x, u, что матрицы $k_x=a^{-1}$, $k_u=a_u^{-1}$ являются невырожденными. Входящие в (10.3.3), (10.3.4) неизвестные σ; $\rho=\begin{pmatrix}\rho_1\\ \vdots\\ \rho_{\widetilde{r}}\end{pmatrix}$; $\varkappa=\|\varkappa_{ij}\|$; m; a_u определяются в дальнейшем.

Подставляя (10.3.3), (10.3.4) в уравнение (9.5.2), получаем

$$\int\limits_{R_{\widetilde{r}}}\exp\left\{x^T\rho-\frac{1}{2}x^T\varkappa x-\beta x^T gu\right\}dx=\exp\left[-\sigma+\frac{\beta}{2}u^T hu\right].$$

Предполагая, что матрица $\varkappa$ положительно определенная, и, следовательно, $\det\varkappa\neq 0$, возьмем этот интеграл при помощи формулы

$$\int\exp\left\{-\frac{1}{2}x^T Ax+x^T y\right\}dx=\det{}^{-1/2}\left(\frac{A}{2\pi}\right)\times$$
$$\times\exp\left\{\frac{1}{2}y^T A^{-1}y\right\}, \tag{10.3.5}$$

где А — положительно определенная матрица (эту формулу легко вывести из (5.4.19). Это приводит к результату

$$\det{}^{-1/2}\frac{\varkappa}{2\pi}\exp\left\{\frac{1}{2}(\rho^T-\beta u^T g^T)\varkappa^{-1}(\rho-\beta g u)\right\}=$$
$$=\exp\left\{-\sigma+\frac{\beta}{2}u^T h u\right\}.$$

Логарифмируем это равенство и приравниваем порознь квадратичные по u члены, линейные и постоянные. Это дает уравнения

$$\beta g^T\varkappa^{-1}g=h, \tag{10.3.6}$$
$$-g^T\varkappa^{-1}\rho=0, \tag{10.3.7}$$
$$\sigma=\frac{1}{2}\ln\det\frac{\varkappa}{2\pi}-\frac{1}{2}\rho^T\varkappa^{-1}\rho. \tag{10.3.8}$$

Чтобы получить другие необходимые соотношения, обратимся ко второму уравнению (9.4.23), которое, умножив на $p(x)$, запишем в форме

$$\int_{\widetilde{U}}p(u)\,p(x)\,e^{-\gamma(x)-\beta c(x,u)}\,du=p(x),\qquad x\in\widetilde{X}.$$

Подставляя сюда (10.3.2)—(10.3.4), будем иметь

$$\int_{\widetilde{U}}\exp\left\{-\frac{1}{2}u^T(a_u+\beta h)u+(m^T a_u-\beta x^T g)u\right\}=$$
$$=\mathrm{const}\exp\left\{-\frac{1}{2}x^T a x+\frac{1}{2}x^T\varkappa x-x^T\rho\right\}. \tag{10.3.9}$$

Матрица $a_u+\beta h$ предполагается невырожденной, положительно определенной. Из (10.3.9), еще раз используя формулу (10.3.5), получаем

$$\frac{1}{2}(m^T a_u-\beta x^T g)(a_u+\beta h)^{-1}(a_u m-\beta g^T x)=$$
$$=-\frac{1}{2}x^T(a-\varkappa)x-x^T\rho+\mathrm{const}.$$

Это равенство должно выполняться при всех x. Поэтому можно приравнять порознь квадратичные и линейные по x формулы, что дает

$$\beta^2 g(a_u+\beta h)^{-1}g^T=\varkappa-a, \tag{10.3.10}$$
$$\beta g(a_u+\beta h)^{-1}a_u m=\rho. \tag{10.3.11}$$

Уравнения (10.3.6)—(10.3.8), (10.3.10), (10.3.11) позволяют определить неизвестные a_u, $\varkappa$, m, ρ, σ, входящие в (10.3.3) и (10.3.4). Разрешая (10.3.10) относительно $\varkappa$ и подставляя в (10.3.6), получаем матричное уравнение

$$\beta g^T[a+\beta^2 g(a_u+\beta h)^{-1}g^T]^{-1}g=h, \tag{10.3.12}$$

которое, как видно из дальнейшего, полностью определяет матрицу a_u. Вводя неизвестную матрицу

$$B = (a_u + \beta h)^{-1}, \tag{10.3.13}$$

и учитывая, что $a = k_x^{-1}$, переписываем (10.3.12) в виде

$$\beta g^T k_x [1_x + \beta^2 g B g^T k_x]^{-1} g = h \tag{10.3.14}$$

(1_x — единичный оператор в $R_{\tilde{r}}$). Используя операторное тождество

$$f(AC)\, A = A f(CA) \tag{10.3.15}$$

[см. формулу (П.1.1)] при $A = g$, $C = Bg^T k_x$, $f(z) = (1 + \beta^2 z)^{-1} = \sum_{n=0}^{\infty} (-\beta^2 z)^n$, преобразуем (10.3.14) к виду

$$\beta \tilde{k}_x \left[1_u + \beta^2 B \tilde{k}_x\right]^{-1} = h, \tag{10.3.16}$$

где $\tilde{k}_x = g^T k_x g$, а 1_u — единичный оператор в $\tilde{U} = R_{\tilde{s}}$.

Нетрудно записать решение уравнения (10.3.16):

$$1_u + \beta^2 B \tilde{k}_x = \beta h^{-1} \tilde{k}_x,$$

тогда

$$\beta B = h^{-1} - (\beta \tilde{k}_x)^{-1}. \tag{10.3.17}$$

Здесь предположено, что матрицы h, $\tilde{k}_x$ являются невырожденными это условие характеризует подпространство $\tilde{U}$). Вследствие 10.3.13), (10.3.10) из (10.3.17) получаем

$$a_u = \beta \left[h^{-1} - (\beta \tilde{k}_x)^{-1}\right]^{-1} - \beta h, \tag{10.3.18}$$

$$\varkappa = a + \beta g \left[h^{-1} - (\beta \tilde{k}_x)^{-1}\right] g^T = k_x^{-1} - g \tilde{k}_x^{-1} g^T + \beta g h^{-1} g^T. \tag{10.3.19}$$

Далее в силу (10.3.11), (10.3.17), (10.3.13) имеем

$$\rho = g \left[h^{-1} - (\beta \tilde{k}_x)^{-1}\right] a_u m. \tag{10.3.20}$$

Подставляя это равенство в (10.3.7), находим

$$g^T \varkappa^{-1} g \left[h^{-1} - (\beta \tilde{k}_x)^{-1}\right] a_u m = 0,$$

и в силу (10.3.6)

$h \left[h^{-1} - (\beta \tilde{k}_x)^{-1}\right]$ [illegible] $= 0.$

Поскольку матрицы h, $h^{-1} - (\beta \tilde{k}_x)^{-1} = \beta B$ предположены невырожденными, это дает

$$a_u m = 0, \tag{10.3.21}$$

и вследствие (10.3.20)

$$\rho = 0. \tag{10.3.22}$$

Наконец, из (10.3.8) согласно (10.3.22), (10.3.19) получаем

$$\sigma = \frac{1}{2} \ln \det \frac{\varkappa}{2\pi} = \frac{1}{2} \ln \det \left[\frac{1}{2\pi} \left(a - g \widetilde{k}_x^{-1} g^T + \beta g h^{-1} g^T \right) \right]. \tag{10.3.23}$$

Тем самым функции (10.3.3), (10.3.4) являются полностью найденными.

2. Для вычисления потенциала $\Gamma(\beta)$ в силу (9.4.10) нужно разрешить (10.3.3) относительно $\gamma(x)$ и усреднить по x с весом (10.3.2):

$$\Gamma(\beta) = -\sigma + \frac{1}{2} \mathbf{M} (x^T \varkappa x - x^T a x) + \frac{1}{2} \ln \det \frac{a}{2\pi} - \\ - \beta \mathbf{M} c^0(x) = -\frac{1}{2} \ln \det (\varkappa k_x) + \\ + \frac{1}{2} \mathbf{M} x^T (\varkappa - a) x - \beta \mathbf{M} c^0(x). \tag{10.3.24}$$

Здесь учтены (10.3.22), (10.3.23). Поскольку

$$\mathbf{M} (\varkappa - a) x x^T = (\varkappa - a) k_x = \varkappa k_x - 1_x$$

в силу (10.3.2), то (10.3.24) можно записать

$$\Gamma(\beta) = -\frac{1}{2} \operatorname{Sp} \ln (\varkappa k_x) + \frac{1}{2} \operatorname{Sp} (\varkappa k_x - 1_x) - \beta \mathbf{M} c^0(x). \tag{10.3.25}$$

Сюда следует подставить равенство (10.3.19), вследствие которого

$$\varkappa k_x - 1_x = g (\beta h^{-1} - \widetilde{k}_x^{-1}) g^T k_x. \tag{10.3.26}$$

Воспользуемся матричной формулой

$$\operatorname{Sp} f(AB) = \operatorname{Sp} f(BA), \tag{10.3.27}$$

которая справедлива, если $f(0) = 0$ [см. (П.1.5), (П.1.6)].

Полагая в (10.3.27) $A = g (\beta h^{-1} - \widetilde{k}_x^{-1})$, $B = g^T k_x$, имеем

$$\operatorname{Sp} f \left(g \left(\beta h^{-1} - \widetilde{k}_x^{-1} \right) g^T k_x \right) = \operatorname{Sp} f \left(g^T k_x g (\beta h^{-1} - \widetilde{k}_x^{-1}) \right) = \\ = \operatorname{Sp} f (\beta \widetilde{k}_x h^{-1} - 1).$$

Применяя это равенство при учете (10.3.26) к (10.3.25) для функций $f(z) = \ln(1 + z)$, $f(z) = z$, будем иметь

$$\Gamma(\beta) = -\frac{1}{2} \operatorname{Sp} \ln (\beta \widetilde{k}_x h^{-1}) + \frac{1}{2} \operatorname{Sp} (\beta \widetilde{k}_x h^{-1} - 1_u) - \beta \mathbf{M} c^0(x). \tag{10.3.28}$$

Зависимость Γ от β, следовательно, можно представить формулой

$$\Gamma(\beta) = -\frac{\tilde{s}}{2}\ln\beta - \frac{1}{2}\operatorname{Sp}\ln\tilde{\mathrm{k}}_x \mathrm{h}^{-1} - \frac{\tilde{s}}{2} - \beta M. \tag{10.3.29}$$

Здесь $\tilde{s} = \operatorname{Sp} 1_u$ — размерность пространства $\tilde{U}(\beta)$, а

$$M = \mathbf{M}c^0(x) - \frac{1}{2}\operatorname{Sp}\mathrm{k}_x \mathrm{h}^{-1}.$$

Для вычисления R и I в соответствии с (9.4.29), (9.4.30) остается продифференцировать потенциал (10.3.28), (10.3.29). Из общей теории, относящейся к третьей вариационной задаче (см. доказательство теоремы 9.5) следует, что при этом активную область $\tilde{U}(\beta)$ с одинаковым успехом можно варьировать или считать постоянной. Выбирая последнюю более простую возможность из (10.3.28), получаем

$$\begin{aligned} R &= \frac{1}{2}\operatorname{Sp}\left(\frac{1_u}{\beta} - \tilde{\mathrm{k}}_x \mathrm{h}^{-1}\right) + \mathbf{M}c^0(x); \\ I &= \frac{1}{2}\operatorname{Sp}\ln(\beta\tilde{\mathrm{k}}_x \mathrm{h}^{-1}) = \frac{\tilde{s}}{2}\ln\beta + \frac{1}{2}\operatorname{Sp}\ln\left(\tilde{\mathrm{k}}_x \mathrm{h}^{-1}\right). \end{aligned} \tag{10.3.30}$$

Для получения функции ценности информации (9.3.7) следует образовать разность

$$V(I) = \frac{1}{2}\left[\operatorname{Sp}\left(\frac{1}{\beta} - \tilde{\mathrm{k}}_x \mathrm{h}^{-1}\right)\right]_{I=0} - \frac{1}{2}\operatorname{Sp}\left(\frac{1}{\beta} - \tilde{\mathrm{k}}_x \mathrm{h}^{-1}\right). \tag{10.3.31}$$

Анализ характера активных областей $\tilde{U}$, показывает, что если значение следа $\operatorname{Sp}\ln(\beta\tilde{\mathrm{k}}_x h^{-1})$, которое в силу (10.3.30) совпадает с $2I$, равно нулю, то равен нулю и след $\operatorname{Sp}(1/\beta - \tilde{\mathrm{k}}_x h^{-1})$. Поэтому формула (10.3.31) принимает более простой вид

$$V = (1/2)\operatorname{Sp}(\tilde{\mathrm{k}}_x h^{-1} - 1/\beta). \tag{10.3.32}$$

Последнее соотношение вместе со второй формулой (10.3.30) дает параметрическое представление зависимости $V(I)$.

3. Охарактеризуем пространство $\tilde{U}(\beta)$. Для того чтобы изложенная в этом параграфе теория была справедливой, как уже отмечалось, необходимо, чтобы матрицы $\varkappa$ и $\mathrm{B}^{-1} = \beta\mathrm{h} + \mathrm{a}_u$ были положительно определенными и, следовательно, невырожденными, а также чтобы матрицы h и $\tilde{\mathrm{k}}_x$ были невырожденными. Нетрудно видеть, что матрица $\tilde{\mathrm{k}}_x = \mathrm{g}^T\mathrm{k}_x\mathrm{g}$ является неотрицательно определенной, поэтому из ее невырожденности следует ее положительная определенность. Принимая далее во внимание (10.3.17), видим, что требование положительной определенности матрицы $\mathrm{B} = \mathrm{h}^{-1} - (\beta\tilde{\mathrm{k}}_x)^{-1}$ в сочетании с требованием положительной определенно-

сти $\widetilde{\mathrm{k}}_x$ приводит при $\beta > 0$ к положительной определенности и, следовательно, невырожденности матрицы $\mathrm{h}^{-1} = \mathrm{B} + (\beta\widetilde{\mathrm{k}}_x)^{-1}$. Наконец, учитывая (10.3.19), заключаем, что матрица $\varkappa = \mathrm{a} + \beta^2 \mathrm{g}\mathrm{B}\mathrm{g}^T$ является положительно определенной, коль скоро матрица a является положительно определенной, а матрица $\mathrm{g}\mathrm{B}\mathrm{g}^T$ неотрицательно определенной. Таким образом, для выполнения всех нужных требований достаточно выполнения двух требований: 1) чтобы матрица $\widetilde{\mathrm{k}}_x = \mathrm{g}^T \mathrm{k}_x \mathrm{g}$ была положительно определенной и 2) чтобы разность $\mathrm{h}^{-1} - (\beta\widetilde{\mathrm{k}}_x)^{-1}$ была положительно определенной:

$$\mathrm{h}^{-1} - (\beta\widetilde{\mathrm{k}}_x)^{-1} = \text{полож. опр.} \tag{10.3.33}$$

Положительная определенность a_u вытекает из (10.3.33) и (10.3.18).

Если x пробегает значения из $\widetilde{X} = R_{\widetilde{r}}$, то $\mathrm{u} = \mathrm{g}^T \mathrm{x}$ пробегает значения из некоторого подпространства, которое мы обозначим U_0 ($\subset R_s$). Можно утверждать, что пространство $\widetilde{U}(\beta)$ есть максимальное линейное подпространство пространства U_0, в котором выполняются указанные требования 1) и 2).

4. Примеры. Рассмотрим сначала одномерный пример, когда $\widetilde{r} = \widetilde{s} = 1$. В этом случае $\widetilde{\mathrm{k}}_x = \mathrm{g}^2 \mathrm{k}_x$ и из формулы (10.3.32) имеем

$$V(I) = \frac{\mathrm{g}^2 \mathrm{k}_x}{2h} (1 - \mathrm{e}^{-2I}) \tag{10.3.34}$$

(матрицы совпадают с числами).

Перейдем к двумерному случаю, когда имеются две независимые гауссовы случайные величины с дисперсиями k_1, k_2 и нулевыми средними значениями. Пусть матрицы g, h имеют диагональный вид:

$$\mathrm{g} = \begin{pmatrix} g_1 & 0 \\ 0 & g_2 \end{pmatrix}; \qquad \mathrm{h} = \begin{pmatrix} h_1 & 0 \\ 0 & h_2 \end{pmatrix}.$$

Для определенности будем предполагать, что $k_1 g_1^2/h_1 > k_2 g_2^2/h_2$. В соответствии с условиями 1) и 2) пространство $\widetilde{U}(\beta)$ будет состоять из точек $(u_1, u_2) = (u_1, 0)$ прямой $u_2 = 0$ при $h_1/k_1 g_1^2 < \beta < h_2/k_2 g_2^2$, и совпадать со всем двумерным пространством $U = R_2$ при $\beta > h_2/k_2 g_2^2$.

В первом случае, когда $\widetilde{U}(\beta)$ одномерно, имеем

$$\mathrm{g} = \begin{pmatrix} g_1 \\ 0 \end{pmatrix}; \qquad \mathrm{g}^T = (g_1, \ 0); \qquad \widetilde{\mathrm{k}}_x = \mathrm{g}^T \mathrm{k}_x \mathrm{g} = (k_1 g_1^2).$$

В двумерном случае

$$\widetilde{\mathrm{k}}_x = \begin{pmatrix} k_1 g_1^2 & 0 \\ 0 & k_2 g_2^2 \end{pmatrix}.$$

Записывая формулы (10.3.30) для данного примера при учете (10.3.10), получаем

$$I = \begin{cases} \frac{1}{2}\ln\beta + \frac{1}{2}\ln(k_1 g_1^2/h_1) & \text{при} \quad h_1/k_1 g_1^2 < \beta < h_2/k_2 g_2^2 \\ \ln\beta + \frac{1}{2}\ln(k_1 g_1^2/h_1) + \frac{1}{2}\ln k_2 g_2^2/h_2 & \text{при} \quad \beta > h_2/k_2 g_2^2, \end{cases}$$

$$R = \begin{cases} 1/2\beta - k_1 g_1^2/2h_1 + \mathbf{M}c_0(x) & \text{при} \quad h_1/k_1 g_1^2 < \beta < h_2/k_2 g_2^2 \\ 1/\beta - k_1 g_1^2/2h_1 - \frac{1}{2} k_2 g_2^2/h_2 + \mathbf{M}c_0(x) & \text{при} \quad \beta > h_2/k_2 g_2^2. \end{cases}$$

Исключая β, находим ценность информации

$$V(I) = \begin{cases} \frac{1}{2}\frac{k_1 g_1^2}{h_1}(1 - e^{-2I}) \quad \text{при } 0 \leqslant I \leqslant I_2 = \frac{1}{2}\ln\frac{k_1 g_1^2 h_2}{k_2 g_2^2 h_1}, \\ \frac{1}{2}\left(\frac{k_1 g_1^2}{h_1} + \frac{k_2 g_2^2}{h_2}\right) - \sqrt{\frac{k_1 g_1^2 k_2 g_2^2}{h_1 h_2}}\, e^{-I} \\ \qquad\qquad \text{при } I \geqslant I_2. \end{cases} \tag{10.3.35}$$

В точке $I = I_2$, $\beta = h_2/k_2 g_2$ вторая производная

$$\frac{d^2 V}{dI^2} = \frac{d}{dI}\frac{1}{\beta} = \frac{dT}{dI} \tag{10.3.36}$$

претерпевает скачок. Она равна $V''(I_2 - 0) = -2V' = -2k_2 g_2^2/h_2$ слева от точки $I = I_2 - 0$ и $V''(I_2 + 0) = -V' = -k_2 g_2^2/h_2$ справа от точки $I = I_2$.

Ход найденной зависимости $V(I)$ изображен на рис. 10.3. Размерность $\tilde{r}$ пространства $\tilde{U}(\beta)$ можно интерпретировать как число активных степеней свободы, которое может меняться при изменении температуры. Это приводит к скачкообразному изменению второй производной (10.3.36), что аналогично скачкообразному изменению теплоемкости в термодинамике (фазовый переход второго рода).

5. В заключение вычислим для гауссовых систем функцию (10.2.22). Учитывая (10.3.1), (10.3.4), (10.3.21), имеем

$$\mu_x(t) = \frac{1}{2}\ln\det(\mathrm{a}_u/2\pi) + tc^0(x) + \\ + \ln\int_{\tilde{U}} \exp\left(t\mathrm{x}^T \mathrm{g}\mathrm{u} + \frac{t}{2}\mathrm{u}^T\mathrm{h}\mathrm{u} - \frac{1}{2}u^T\mathrm{a}_u\mathrm{u}\right)du.$$

Применяя формулу (10.3.5), получаем

$$\mu_x(t) = tc^0(x) - \frac{1}{2}\ln\det(1 - t\mathrm{h}\,\mathrm{a}_u^{-1}) + \\ + \frac{1}{2}t^2\mathrm{x}^T\,\mathrm{g}(\mathrm{a}_u - t\mathrm{h})^{-1}\mathrm{g}^T\,\mathrm{x} = tc^0(x) -$$

$$-\frac{1}{2}\operatorname{Sp}\ln(1-t\mathrm{h}\mathrm{a}_u^{-1})+\frac{1}{2}t^2\mathrm{x}^T\mathrm{g}\mathrm{a}_u^{-1}(1-t\mathrm{h}\mathrm{a}_u^{-1})^{-1}\mathrm{g}^T\mathrm{x},\tag{10.3.37}$$

где a_u^{-1} определяется формулой (10.3.18). Отсюда, в частности, можно найти математическое ожидание

$$\mathbf{M}\mu_x(t)=t\mathbf{M}c^0(x)-\frac{1}{2}\operatorname{Sp}\ln(1-t\mathrm{h}\mathrm{a}_u^{-1})+$$
$$+\frac{1}{2}t^2\operatorname{Sp}\mathrm{a}_u^{-1}(1-t\,\mathrm{h}\mathrm{a}_u^{-1})^{-1}\widetilde{\mathrm{k}}_x\tag{10.3.38}$$

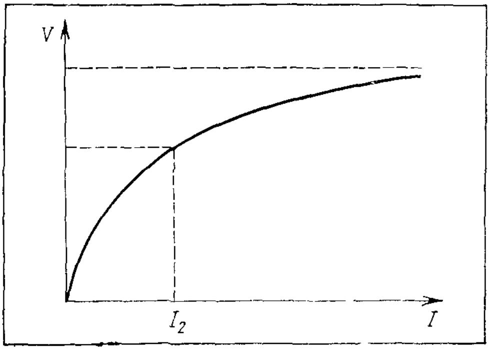


Рис. 10.3.
Функция ценности для гауссовой системы с «фазовым переходом», соответствующая формуле (10.3.35).

и производные

$$\mu_x'(t)=c^0(x)+\frac{1}{2}\operatorname{Sp}\frac{\mathrm{h}\mathrm{a}_u^{-1}}{1-t\mathrm{h}\mathrm{a}_u^{-1}}+$$
$$+\frac{1}{2}t\mathrm{x}^T\mathrm{g}\mathrm{a}_u^{-1}\,\frac{2-t\mathrm{h}\mathrm{a}_u^{-1}}{(1-t\mathrm{h}\mathrm{a}_u^{-1})^2}\,\mathrm{g}^T\,\mathrm{x},\tag{10.3.39}$$

$$\mathbf{M}\mu_x''(t)=\frac{1}{2}\operatorname{Sp}\frac{(\mathrm{h}\mathrm{a}_u^{-1})^2}{(1-t\mathrm{h}\mathrm{a}_u^{-1})^2}+\operatorname{Sp}\mathrm{a}_u^{-1}(1-t\mathrm{h}\mathrm{a}_u^{-1})^{-3}\widetilde{\mathrm{k}}_x.\tag{10.3.40}$$

Если скомбинировать (10.3.37) с (10.3.39) и положить $t=-\beta$, то будем иметь

$$\beta\mu_x'(-\beta)+\mu_x(-\beta)=\frac{1}{2}\operatorname{Sp}\left[\frac{\beta\mathrm{h}\mathrm{a}_u^{-1}}{1+\beta\mathrm{h}\mathrm{a}_u^{-1}}-\ln(1+\beta\mathrm{h}\mathrm{a}_u^{-1})\right]-$$
$$-\frac{\beta^2}{2}\mathrm{x}^T\mathrm{g}\mathrm{a}_u^{-1}(1+\beta\mathrm{h}\mathrm{a}_u^{-1})^{-2}\mathrm{g}^T\,\mathrm{x}.\tag{10.3.41}$$

Вытекающие из (10.3.18) соотношения

$$1+\frac{1}{\beta}\mathrm{a}_u\mathrm{h}^{-1}=\left(1-\frac{1}{\beta}\mathrm{h}\widetilde{\mathrm{k}}_x^{-1}\right)^{-1};\qquad \beta\mathrm{h}\mathrm{a}_u^{-1}=\beta\widetilde{\mathrm{k}}_x\mathrm{h}^{-1}-1$$

позволяют в полученных результатах перейти от матрицы $\mathrm{h}a_{u}^{-1}$ к матрице $\widetilde{\mathrm{k}}_x \mathrm{h}^{-1}$. Так из (10.3.40), (10.3.41) будем иметь

$$\beta^2 \mathbf{M}\mu_x''(-\beta) = \frac{1}{2}\operatorname{Sp}\left(1 - \frac{1}{\beta}\mathrm{h}\widetilde{\mathrm{k}}_x^{-1}\right)^2 + \operatorname{Sp}\frac{\beta\widetilde{\mathrm{k}}_x\mathrm{h}^{-1} - 1}{(\beta\widetilde{\mathrm{k}}_x\mathrm{h}^{-1})^2}, \tag{10.3.42}$$

$$\beta\mu_x'(-\beta) + \mu_x(-\beta) = \frac{1}{2}\operatorname{Sp}\left[1 - \frac{1}{\beta}\mathrm{h}\widetilde{\mathrm{k}}_x^{-1} - \ln(\beta\widetilde{\mathrm{k}}_x\mathrm{h}^{-1})\right] - \\ - \frac{\beta}{2}\mathrm{x}^T \mathrm{g}\mathrm{h}^{-1}\frac{\beta\widetilde{\mathrm{k}}_x\mathrm{h}^{-1} - 1}{(\beta\widetilde{\mathrm{k}}_x\mathrm{h}^{-1})^2}\mathrm{g}^T\mathrm{x}. \tag{10.3.43}$$

Эти формулы понадобятся в § 11.4.

10.4. Стационарные гауссовы системы

1. Пусть x и u состоят из многих компонент:

$$x = (x_1, \ldots, x_r), \qquad u = (u_1, \ldots, u_r).$$

Бейесовская система является стационарной, если ей свойственна инвариантность относительно преобразования сдвига

$$(x_1, \ldots, x_r) \to (x_{1+n}, \ldots, x_{r+n}); \qquad (u_1, \ldots, u_r) \to \\ \to (u_{1+n}, \ldots, u_{r+n}) \quad (x_{h+r} = x_h,\ u_{h+r} = u_h).$$

Для этого случайный процесс x должен быть стационарным, т. е. распределение $p(x)$ должно удовлетворять условию

$$p(x_1, \ldots, x_r) = p(x_{1+n}, \ldots, x_{r+n}),$$

а также функция штрафов должна удовлетворять условию стационарности

$$c(x_1, \ldots, x_r;\ u_1, \ldots, u_r) = c(x_{1+n}, \ldots, x_{r+n};\ u_{1+n}, \ldots, u_{r+n}).$$

Применительно к гауссовым системам указанные условия стационарности приводят к требованию, чтобы матрицы

$$\|h_{ij}\|, \qquad \|g_{ij}\|, \qquad \|a_{ij}\|, \tag{10.4.1}$$

входящие в выражения (10.3.1), (10.3.2), были стационарными, т. е. чтобы их элементы зависели лишь от разности индексов:

$$h_{ij} = h_{i-j}, \qquad g_{ij} = g_{i-j}; \qquad a_{ij} = a_{i-j}, \qquad (k_x)_{ij} = (k_x)_{i-j}.$$

Рассмотрим сначала тот случай, когда пространства имеют конечную размерность r. Матрицы (10.4.1) можно привести к диагональному виду унитарным преобразованием

$$\mathrm{x}' = \mathrm{U}^+\mathrm{x} \tag{10.4.2}$$

с матрицей

$$\mathrm{U} = \| U_{lj} \|; \qquad U_{lj} = \frac{1}{r} \mathrm{e}^{2\pi i l j / r}; \qquad l,\ j = 1, \ldots, r. \tag{10.4.3}$$

при этом

$$\mathrm{U}^+ \mathrm{hU} = \| \overline{h}_l \, \delta_{lm} \|; \qquad \mathrm{U}^+ \mathrm{gU} = \| \overline{g}_l \, \delta_{lm} \|; \qquad \mathrm{U}^+ \mathrm{k}_x \mathrm{U} = \| \overline{k}_l \, \delta_{lm} \|, \tag{10.4.4}$$

где

$$\overline{h}_l = \sum_{k=1}^{r} \mathrm{e}^{-2\pi i l k / r} h_k \tag{10.4.5}$$

по аналогии с (5.5.9), (8.7.6).

После указанной диагонализации формулы (10.3.30) примут вид

$$I = \frac{1}{2} \sum_l \ln [\beta \overline{k}_l \, | \overline{g}_l |^2 / \overline{h}_l]; \tag{10.4.6}$$

$$R = \frac{1}{2} \sum_l (1/\beta - \overline{k}_l \, | \overline{g}_l |^2 / \overline{h}_l) + \mathbf{M} c_0 (x).$$

Унитарное преобразование (10.4.2) привело к тому, что новые переменные

$$\begin{pmatrix} x_1' \\ \vdots \\ x_r' \end{pmatrix} = \mathrm{U}^+ \begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix}$$

стали некоррелированными, т. е. независимыми. Рассматривая координаты

$$\begin{pmatrix} u_1' \\ \vdots \\ u_r' \end{pmatrix} = \mathrm{U}^+ \begin{pmatrix} u_1 \\ \vdots \\ u_r \end{pmatrix},$$

удобно охарактеризовать активное подпространство $\widetilde{U}$. В нем часть этих координат равна нулю. Неравными нулю являются согласно (10.3.33) лишь те координаты, для которых

$$\beta \overline{\mathrm{k}}_l \, | \overline{g}_l |^2 / \overline{h}_l > 1. \tag{10.4.7}$$

Поэтому в (10.4.6) суммирование нужно проводить лишь по тем l, для которых последнее неравенство выполняется. Число таких значений индекса и есть $\widetilde{r}$.

2. Пусть теперь $x = \{x(t)\}$ есть случайная функция на отрезке $[0, T_0]$, т. е. пространство x является бесконечномерным (функциональным).

Бейесовскую систему будем предполагать строго стационарной, считая, что матрицы h, g, a имеют вид

$$\mathrm{h} = \| h(t' - t'') \|; \qquad \mathrm{g} = \| g(t' - t'') \|. \tag{10.4.8}$$

и аналогично для a, где $h(\tau)$, $g(\tau)$, $a(\tau)$ — периодические функции с периодом T_0. Унитарное преобразование, диагонализирующее эти матрицы, имеет вид преобразования Фурье

$$(\mathrm{Ux})_l = \mathrm{x}'_l = \mathrm{x}'(\omega_l) = \frac{1}{\sqrt{T_0}} \int_0^{T_0} \mathrm{e}^{-2\pi i l t/T_0} x(t)\,dt,$$

$$(\omega_l = 2\pi l/T_0; \quad l = 0,\ 1,\ \ldots).$$

При этом будем иметь (см. (8.7.12))

$$\mathrm{U}^+ \mathrm{hU} = \| \bar{h}(\omega_l)\, \delta_{lm} \|,$$

$$\bar{h}(\omega_l) = \int_0^{T_0} \mathrm{e}^{-i\omega_l \tau} h(\tau)\,d\tau = \int_{-T_0/2}^{T_0/2} \mathrm{e}^{i\omega_l \tau} h(\tau)\,d\tau, \tag{10.4.9}$$

и т. п. для других матриц. Формулы (10.3.30) в этом случае будут иметь вид, аналогичный (10.4.6),

$$I_0 = \frac{I}{T_0} = \frac{1}{2T_0} \sum_{l \in L} \ln\left[\beta \frac{\bar{k}(\omega_l)\,|\,\bar{g}(\omega_l)\,|^2}{\bar{h}(\omega_l)}\right],$$

$$R_0 = \frac{R}{T_0} = -\frac{1}{2T_0} \sum_{l \in L} \left[\frac{\bar{k}(\omega_l)\,|\,\bar{g}(\omega_l)\,|^2}{\bar{h}(\omega_l)} - \frac{1}{\beta}\right] + \frac{1}{T_0} \mathbf{M} c_0(x), \tag{10.4.10}$$

с той лишь разницей, что индекс l может теперь пробегать всевозможные целые значения ... —1, 0, 1, 2, ..., для которых выполняются условия

$$\beta \bar{k}(\omega_l)\,|\,\bar{g}(\omega_l)\,|^2 > \bar{h}(\omega_l) \tag{10.4.11}$$

типа (10.4.7). Пусть $\Phi_{\max}$ есть максимальная величина

$$\Phi_{\max} = \max\left[\frac{\bar{k}(\omega_l)\,|\,\bar{g}(\omega_l)\,|^2}{\bar{h}(\omega_l)} \quad l = \ldots, -1,\ 0,\ 1,\ 2,\ \ldots\right].$$

Тогда при $\beta < 1/\Phi_{\max}$ не будет иметься индексов l, для которых выполнялось бы условие (10.4.11) и суммы в (10.4.10) будут отсутствовать, так что I будет равно нулю, а

$$(R_0)_{I_0 = 0} = \frac{1}{T_0} \mathbf{M} c^0(x). \tag{10.4.12}$$

Ненулевое значение I_0 появится как только β достигнет значения $1/\Phi_{\max}$ и превзойдет его. Учитывая (10.4.12) по обычной формуле

$$V_0(I_0) = (R_0)_{I_0 = 0} - R_0(I_0),$$

из (10.4.10) получаем выражение для ценности информации

$$V_0(I_0)=\frac{1}{2T_0}\sum_{l\in L}\left[\frac{\overline{k}(\omega_l)\,|g(\omega_l)|^2}{\overline{h}(\omega_l)}-\frac{1}{\beta}\right]. \qquad (10.4.13)$$

Зависимость $V_0(I_0)$ получена согласно (10.4.10), (10.4.13) в параметрической форме.

3. Рассмотрим тот случай, когда $x=\{\ldots, x_{-1}, x_0, x_1, x_2, \ldots\}$ есть бесконечная стационарная последовательность и элементы матриц

$$\mathrm{h}=\|h_{i-j}\|, \qquad \mathrm{g}=\|g_{i-j}\|$$

зависят лишь от разности $i-j$. Тогда эти матрицы диагонализируются унитарным преобразованием

$$(\mathrm{U}^{+}x)(\lambda)=\frac{1}{\sqrt{2\pi}}\sum_{j}\mathrm{e}^{-i\lambda j}x_j, \qquad (10.4.14)$$

где $-\pi\leqslant\lambda\leqslant\pi$. По аналогии с (10.4.4), (10.4.5) в этом случае нетрудно получить

$$\mathrm{U}^{+}\mathrm{hU}=\|h(\lambda)\,\delta(\lambda-\lambda')\|, \ldots, \quad h(\lambda)=\sum_k \mathrm{e}^{-i\lambda k}h_k. \qquad (10.4.15)$$

Преобразование (10.4.14) можно рассматривать как предельный случай (при $r\to\infty$) преобразования (10.4.2), (10.4.3), причем $\lambda=\lambda_l=2\pi r/l$. Из сравнения (10.4.5) и (10.4.15) имеем $\overline{h}_l=h\,(2\pi l/r),\ldots$ Поэтому формулы (10.4.6) дают

$$I_1=\frac{I}{r}=\frac{1}{2r}\sum_l \ln\left[\beta\,\frac{k\,(2\pi l/r)\,|g(2\pi l/r)|^2}{h\,(2\pi l/r)}\right],$$

$$R_1=\frac{R}{r}=-\frac{1}{2r}\sum_l\left[\frac{k\,(2\pi l/r)\,|g\,(2\pi l/r)|^2}{h\,(2\pi l/r)}-\frac{1}{\beta}\right]+\frac{1}{r}\mathrm{M}c^0(x).$$

Совершая здесь предельный переход $r\to\infty$, поскольку $\Delta\lambda=2\pi/r$, получаем отсюда интегралы

$$I_1=\frac{1}{4\pi}\int\limits_{\lambda\in L(\beta)}\ln\left[\beta\frac{k(\lambda)\,|g(\lambda)|^2}{h(\lambda)}\right]d\lambda,$$

$$R_1=-\frac{1}{4\pi}\int\limits_{\lambda\in L(\beta)}\left[\frac{k(\lambda)\,|g(\lambda)|^2}{h(\lambda)}-\frac{1}{\beta}\right]d\lambda+\mathrm{const}. \qquad (10.4.16)$$

Интегрирование проводится по тому подынтервалу $L(\beta)$ интервала $(-\pi, \pi)$, где $\beta k(\lambda)\,|g(\lambda)|^2>h(\lambda)$. Найденные формулы определяют удельные величины I_1, R_1, приходящиеся в среднем на один элемент последовательностей $\{\ldots, x_1, x_2, \ldots\}$, $\{\ldots, u_1, u_2, \ldots\}$. Обозначая через $|L|$ длину подынтервала $L(\beta)$, из (10.4.16) имеем

$$\frac{4\pi R_1}{|L|}=-\Phi_1+\Phi_2\,\mathrm{e}^{-4\pi I_1/|L|}+\mathrm{const}, \qquad (10.4.17)$$

где

$$\Phi_1 = \frac{1}{|L|}\int\limits_L \Phi(\lambda)\,d\lambda; \qquad \Phi_2 = \exp\left\{\frac{1}{|L|}\int\limits_L \ln\Phi(\lambda)\,d\lambda\right\}$$

— некие средние (в $L(\beta)$) значения функций $\Phi(\lambda) = k(\lambda)\,|g(\lambda)|^2/h(\lambda)$, $\ln\Phi(\lambda)$.

4. Пусть, наконец, имеется стационарный процесс на бесконечной непрерывной временной оси. Функции h, g при этом есть функции лишь разности времен, подобно (10.4.8). Этот случай можно рассматривать как предельный случай системы, рассмотренной в п. 2 или в п. 3. Так в формулах п. 2 требуется совершить предельный переход $T_0 \to \infty$, в процессе которого точки $\omega_l = 2\pi l/T_0$ на оси ω неограниченно уплотняются ($\Delta\omega = 2\pi/T_0$) и суммы в (10.4.10), (10.4.13) переходят в интегралы:

$$I_0 = \frac{1}{4\pi}\int\limits_{L(\beta)} \ln\left[\beta\,\frac{\overline{k}(\omega)\,|\overline{g}(\omega)|^2}{\overline{h}(\omega)}\right]d\omega,$$

$$V_0 = \frac{1}{4\pi}\int\limits_{L(\beta)} \left[\frac{\overline{k}(\omega)\,|\overline{g}(\omega)|^2}{\overline{h}(\omega)} - \frac{1}{\beta}\right]. \tag{10.4.18}$$

Входящие сюда функции $\overline{h}(\omega)$, $\overline{g}(\omega)$, ..., определены равенством (10.4.9)

$$\overline{h}(\omega) = \int\limits_{-\infty}^{\infty} e^{-i\omega\tau} h(\tau)\,d\tau, \ \ldots\,. \tag{10.4.19}$$

Интегрирование в (10.4.18) проводится по той области $L(\beta)$ оси ω, где

$$\beta\overline{k}(\omega)\,|\overline{g}(\omega)|^2 > \overline{h}(\omega). \tag{10.4.20}$$

Обозначим через $|L|$ суммарную длину этой области, тогда из последних формул (10.4.18) аналогично (10.4.17) будем иметь

$$\frac{4\pi V_0}{|L|} = \Phi_1 - \Phi_2\, e^{-4\pi I_0/|L|},$$

где

$$\Phi_0 = \frac{1}{|L|}\int\limits_L \Phi(\omega)\,d\omega; \qquad \Phi_2 = \exp\left\{\frac{1}{|L|}\int\limits_L \ln\Phi(\omega)\,d\omega\right\},$$

$$\Phi(\omega) = \frac{\overline{k}(\omega)\,|\overline{g}(\omega)|^2}{\overline{h}(\omega)}. \tag{10.4.21}$$

Эти средние оказываются слабо зависящими от β.

5. Рассмотрим в качестве примера стационарный гауссов процесс $x = \{x(t)\}$, имеющий корреляционную функцию

$$k(\tau) = \sigma^2 e^{-\nu|\tau|},$$

так что в соответствии с (10.4.19)

$$\bar{k}(\omega) = 2\gamma\sigma^2/(\gamma^2 + \omega^2). \tag{10.4.22}$$

В качестве функции штрафа возьмем квадратичную функцию

$$c(x, u) = \frac{1}{2}\int [x(t) - u(t)]^2 \, dt \tag{10.4.23}$$

или в матричной форме записи

$$c(x, u) = \frac{1}{2}\mathrm{x}^T\mathrm{x} - \mathrm{x}^T\mathrm{u} + \frac{1}{2}\mathrm{u}^T\mathrm{u}.$$

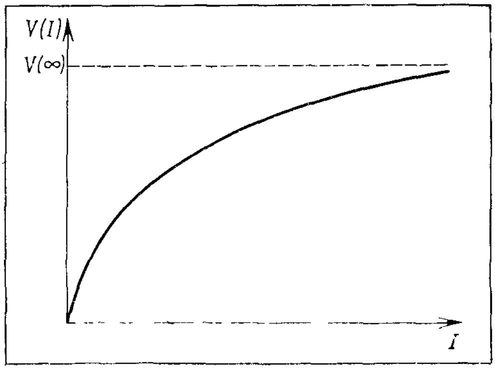


Рис. 10.4. Удельная функция ценности информации для примера с функцией штрафов (10.4.23).

При этом матрицы h, g, как видно из сравнения с (10.3.1), имеют вид $\mathrm{h} = \|\delta(t - t')\|$; $\mathrm{g} = -\|\delta(t - t')\|$ и поэтому

$$\bar{h}(\omega) = 1; \qquad \bar{g}(\omega) = -1. \tag{10.4.24}$$

Функция (10.4.21) в данном случае в силу (10.4.22), (10.4.24) выглядит так:

$$\Phi(\omega) = 2\gamma\sigma^2/(\gamma^2 + \omega^2).$$

Условие (10.4.20) принимает вид

$$2\beta\gamma\sigma^2 > \gamma^2 + \omega^2,$$

следовательно, при фиксированном значении $\beta > \gamma/2\sigma^2$ область $L(\beta)$ является отрезком $-\sqrt{2\beta\gamma\sigma^2 - \gamma^2} < \omega < \sqrt{2\beta\gamma\sigma^2 - \gamma^2}$. Вместо параметра β будем рассматривать параметр $y = \sqrt{2\beta\sigma^2/\gamma - 1}$. Тогда $\beta = \gamma(1 + y^2)/2\sigma^2$ и указанный интервал L будет иметь вид

$$-\gamma y < \omega < \gamma y.$$

Учитывая, что

$$\frac{1}{2y}\int_{-y}^{y}\frac{dx}{1+x^2} = \frac{\operatorname{arctg} y}{y},$$

$$\frac{1}{2y}\int_{-y}^{y}\ln(1+x^2)\,dx=\ln(1+y^2)+2\left(\frac{\operatorname{arctg} y}{y}-1\right),$$

из формул (10.4.18) будем иметь

$$I_0=\frac{\gamma}{\pi}(y-\operatorname{arctg}\ y);\quad V_0=\frac{\sigma^2}{\pi}\left(\operatorname{arctg} y-\frac{y}{1+y^2}\right). \qquad (10.4.25)$$

Вследствие (10.4.23) и стационарности процесса удвоенные удельные потери $2R_0$, отнесенные к единице времени, совпадают со средним квадратом ошибки $2R_0 = \mathbf{M}\,[x\,(t) - u\,(t)]^2$, а удвоенная ценность $2V_0\,(I_0) = \sigma^2 - 2R_0\,(I_0)$ показывает величину максимального его уменьшения, которое возможно при данном удельном количестве информации I_0.

Кривая ценности, соответствующая формулам (10.4.25), приведена на рис. 10.4. Пользуясь этими формулами, можно также получить приближенные формулы для зависимости $V_0\,(I_0)$ при малых и при больших значениях параметра y (или отношения I_0/y).

При малых $y \ll 1$ воспользуемся разложениями $\operatorname{arctg} y = y - y^3/3 + y^5/5 - \ldots$, $y/(1 + y^2) = y - y^3 + y^5 + \ldots$. Подставляя их в (10.4.25), получаем

$$I_0=\frac{\gamma}{\pi}\left(\frac{y^3}{3}-\frac{y^5}{5}+\ldots\right),\qquad V_0=\frac{\sigma^2}{\pi}\left(\frac{2}{3}y^3-\frac{4}{5}y^5+\ldots\right),$$

и после исключения y:

$$V_0(I_0)=\frac{2\sigma^2}{\gamma}I_0\left(1-\frac{3}{5}y^2+\ldots\right)=$$
$$=\frac{2\sigma^2}{\gamma}I_0-\frac{6}{5}\sigma^2(3\pi)^{2/3}\left(\frac{I_0}{\gamma}\right)^{5/3}+\ldots$$

При больших $y \gg 1$ произведем разложение по обратному параметру, а именно

$$\operatorname{arctg} y=\frac{\pi}{2}-\operatorname{arctg}\frac{1}{y}=\frac{\pi}{2}-y^{-1}+\frac{1}{3}y^{-3}-\ldots,$$
$$\frac{y}{1+y^2}=y^{-1}-y^{-3}+y^{-5}+\ldots.$$

Формулы (10.4.25) дадут

$$\pi I_0/\gamma=y-\pi/2+y^{-1}+\cdots,$$
$$\pi V_0/\sigma^2=\pi/2-2y^{-1}+(4/3)\,y^{-3}.$$

Исключая отсюда y, будем иметь

$$V_0/\sigma^2={}^1/_2-(2/\pi^2)(I_0/\gamma+{}^1/_2)^{-1}+O(I_0^{-3}).$$

Для данного примера нетрудно записать также потенциал (10.3.28).

Глава 11

АСИМПТОТИЧЕСКИЕ РЕЗУЛЬТАТЫ, КАСАЮЩИЕСЯ ЦЕННОСТИ ИНФОРМАЦИИ. ТРЕТЬЯ АСИМПТОТИЧЕСКАЯ ТЕОРЕМА

Главным асимптотическим результатом, затрагивающим ценность информации, следует признать факт асимптотической равноценности различных родов информации: хартлиевской, больцмановской, шенноновской, имеющей место при весьма широких предположениях типа требований информационной устойчивости. Этот факт не сводится к факту асимптотической безошибочности передачи информации через канал с помехами, утверждаемому теоремой Шеннона (гл. 7), а является самостоятельным и не менее значимым.

Комбинация двух указанных фактов приводит к обобщенному результату, носящему название обобщенной теоремы Шеннона (§ 11.5). В последней рассматривается общий критерий качества, определяемый произвольной функцией штрафов и соответствующим ей риском. Исторически факт асимптотической равноценности информации был впервые доказан (1959 г.) именно в такой комбинированной, завуалированной форме, в сочетании со вторым фактом (асимптотической безошибочности). Он не осмысливался сначала как самостоятельный факт, а составлял по существу часть обобщенной теоремы Шеннона.

Мы в этой главе придерживаемся другого способа изложения и рассматриваем факт асимптотической равноценности различных количеств информации как совершенно особый самостоятельный факт, более элементарный, нежели обобщенная теорема Шеннона. Этот способ изложения мы считаем более предпочтительным как с принципиальной, так и с педагогической точки зрения. При этом отчетливее видна симметрия теории информации, равноправие второй и третьей вариационных задач.

Помимо самого факта асимптотической равноценности информаций, разумеется, важен и интересен вопрос о величине расхождения между ценностями различных родов информации. В § 11.3, 11.4 приводятся найденные автором первые члены асимптотического разложения для указанного расхождения. Эти члены являются точными для выбранного случайного кодирования и дают представление (как во всяком асимптотическом, полусходящемся разложении) о быстроте убывания расхождения, хотя сумма всех остальных членов разложения и не оценивается. Особо рассматривается вопрос об инвариантности результатов относительно преобразования функции штрафов $c(\xi, \zeta) \to c(\xi, \zeta) + f(\xi)$, которое не сказывается на передаче информации и на ее ценности. Предпочтение отдается тем формулам, в которых фигурируют величины и функции, инвариантные относительно указанного преобразования, например, берется отношение инвариантной разности $\tilde{R} - R = V - \tilde{V}$ к инвариантной величине V, а не к риску R, который является неинвариантным (см. теорему 11.2).

Разумеется, исследования в данном направлении могут быть дополнены и улучшены. Скажем, при рассмотрении обобщенной теоремы Шеннона в § 11.5 законно поставить вопрос о быстроте исчезновения расхождения в рисках. Этот вопрос, однако, остался нерассмотренным.

11.1. О различии между ценностями различных родов информации. Предварительные формы

Пусть имеется бейесовская система $[P(dx), c(x, u)]$, т. е. задана случайная величина x из вероятностного пространства (X, F, P) и $F \times G$-измеримая функция штрафов $c(x, u)$, где u (из измеримого пространства (U, G))— принимаемая оценка. В гл. 9 для такой системы были определены функции ценности различных количеств информации: хартлиевского, больцмановского и шенноновского. Эти функции соответствуют минимальным средним потерям $\mathbf{M}c(x, u)$, достижимым при приеме заданных количеств информации. Информация хартлиевского типа состоит в указании, какой области E_k из оптимального разбиения $E_1 + \ldots + E_M = X$ $(E_k \in F)$ принадлежит значение x. Минимальные потери

$$\tilde{R}(I) = \inf_{\Sigma E_k} \mathbf{M} \inf_{u} \mathbf{M}[c(x, u) \mid E_k] \tag{11.1.1}$$

определяются минимизацией как по оценкам u, так и по различным разбиениям. Ограничение хартлиевского количества информации значением I соответствует ограничению $M \leqslant e^I$ числа указанных областей.

При ограничении шенноновского количества информации рассматриваются минимальные штрафы

$$R(I) = \inf_{P(du|x)} \int c(x, u) P(dx) P(du \mid x), \tag{11.1.2}$$

где минимизация производится по условным распределениям $P(du \mid x)$, совместимым с неравенством $I_{xu} \leqslant I$.

Минимальные потери $\tilde{\tilde{R}}(I)$, соответствующие ограничению больцмановского количества информации, согласно (9.6.6) заключены между потерями (11.1.1) и (11.1.2):

$$R(I) \leqslant \tilde{\tilde{R}}(I) \leqslant \tilde{R}(I). \tag{11.1.3}$$

Поэтому все три функции $R(I)$, $\tilde{\tilde{R}}(I)$, $R(I)$ будут близки друг к другу, если близки $\tilde{R}(I)$ и $R(I)$.

Из определения функций $\tilde{R}(I)$, $\tilde{\tilde{R}}(I)$, $R(I)$ непосредственно следует лишь неравенство (11.1.3). Возникает вопрос, насколько сильно отличаются друг от друга функции $\tilde{R}(I)$ и $R(I)$. Если они не сильно различаются, то вместо трудно вычисляемой функции $\tilde{R}(I)$ можно рассматривать значительно легче вычисляемую функцию $R(I)$, обладающую такими удобными свойствами как дифференцируемость и пр. Исследованию различия между функциями $\tilde{R}(I)$, $R(I)$, а значит и между функциями ценности $\tilde{V}(I)$, $\tilde{\tilde{V}}(I)$, $V(I)$, посвящена настоящая глава. Оказывается, что для бейесовских систем определенного типа — систем, обладающих свойством «информационной устойчивости», имеет место асимптотическое совпадение указанных функций. Этот фундаментальный асимптотический результат — третью асимптотическую теорему, по глубине и значимости можно сравнить с соответствующими асимптотическими результатами первой и второй теорем.

Прежде чем давать определение информационно устойчивых бейесовских систем, рассмотрим составные системы, обобщением которых являются информационно устойчивые системы. Назовем бейесовскую систему $[P_n(d\xi), \; c(\xi, \zeta)]$ n-й степенью системы $[P_1(dx), c_1(x, u)]$, если случайная величина ξ является совокупностью $(x_1, \ldots, x_n)$ из n одинаково распределенных, независимых, случайных величин, являющихся копией величины x, т. е.

$$P_n(d\xi) = P_1(dx_1) \ldots P_1(dx_n); \tag{11.1.4}$$

оценка ζ является совокупностью одинаковых $u_1, \ldots, u_n$, а функция штрафа является суммой

$$c(\xi, \zeta) = \sum_{i=1}^{n} c_1(x_i, u_i). \tag{11.1.5}$$

Как к системе $[P_1, c_1]$, так и к ее степени можно применять формулы (11.1.1), (11.1.2). Количество информации $I_n = nI_1$ для составной системы естественно взять в n раз больше, чем для простой системы. Тогда экстремальное распределение $P(d\zeta \mid \xi)$, соответствующее формуле (11.1.2) для составной системы, распадается согласно (9.4.21) на произведение

$$P_{I_n}(d\zeta \mid \xi) = P_{I_1}(du_1 \mid x_1) \ldots P_{I_1}(du_n \mid x_n), \tag{11.1.6}$$

где $P_{I_1}(du \mid x)$ — аналогичное экстремальное распределение для простой системы. В соответствии с (11.1.2) имеем

$$R_1(I_1) = \int c_1(x, u) P_1(dx) P_{I_1}(du \mid x). \tag{11.1.7}$$

Для n-й степени вследствие (11.1.5), (11.1.7) будем иметь

$$R_n(I_n) = nR_1(I_1). \tag{11.1.8}$$

Сложнее обстоит дело в отношении функции (11.1.1). Функции $\widetilde{R}_n(nI_1)$, $\widetilde{R}_1(I_1)$ сложной и составной системы уже не связаны таким простым соотношением. Для составной системы число областей разбиения можно брать равным $M = [e^{nI_1}]$, тогда как для простой системы $m \leqslant [e^{I_1}]$ (скобки обозначают целую часть). Очевидно, что $M \geqslant m^n$, поэтому разбиение

$$(\Sigma E_k) \dots (\Sigma E_z) \tag{11.1.9}$$

пространства $X^n = (X \times \dots \times X)$ на m^n областей, индуцируемое экстремальным разбиением $\sum E_k = X$ малой системы, входит в число допустимых разбиений, перебираемых в формуле

$$\widetilde{R}_n(nI_1) = \inf_{\sum\limits_{k=1}^{M} G_k = X^n} \mathbf{M} \inf_{\zeta} \mathbf{M}[c(\xi, \zeta) | G_k]. \tag{11.1.10}$$

Отсюда вытекает, что

$$\widetilde{R}_n(nI_1) \leqslant n\widetilde{R}_1(I_1). \tag{11.1.11}$$

Кроме разбиений (11.1.9), однако, теперь имеется большое число допустимых разбиений другого вида. Поэтому есть основание ожидать, что $n\widetilde{R}_1(I_1)$ будет существенно больше, чем $\widetilde{R}_n(nI_1)$. Есть основание ожидать, что для некоторых систем удельные потери $R_n(nI_1)/n$, [заведомо превосходящие $R_1(I_1)$ в силу (11.1.3), (11.1.8)], убывают с ростом n и приближаются к своему предельно малому возможному значению, которое оказывается совпадающим именно с $R_n(nI_1)/n = R_1(I_1)$. Этот факт и составляет содержание основного результата (третьей асимптотической теоремы). При его доказательстве попутно получается и другой важный результат, а именно обнаруживается рецепт, как находить разбиение $\sum G_k$, близкое (в асимптотическом смысле) к оптимальному. Оказывается для этого пригодна процедура, аналогичная декодированию по случайному коду Шеннона (см. § 7.2). Берутся $\widetilde{M}$ кодовых точек $\zeta_1, \dots, \zeta_{\widetilde{M}}$ (напомним, что каждая из них представляет собой блок $(u_1, \dots, u_n)$). Эти точки являются результатом M-кратного случайного выбрасывания, совершаемого с вероятностями

$$P_{I_n}(d\zeta) = \int_{X^n} P(d\xi) P_{I_n}(d\zeta | \xi) = P_{I_1}(u_1) \dots P_{I_1}(u_n), \tag{11.1.12}$$

где $P_{I_n}(d\zeta | \xi)$ — оптимальное распределение (11.1.6). Число кодовых точек берем таким: $\widetilde{M} = [e^{\widetilde{I}_n}]$ $(\widetilde{I}_n = n\widetilde{I}_1)$ $\widetilde{I}_1$ не зависит от n), причем величину $\widetilde{I}_n$ для доказательства последующей теоремы 11.1 следует полагать несколько отличающейся от величины $I_n = nI_1$, входящей в (11.1.12). Указанные кодовые точки и «расстояние» $c(\xi, \zeta)$ от ξ до ζ определяют разбиение

$$\sum_{k=1}^{\widetilde{M}} G_k = X^n. \tag{11.1.13}$$

Область G_k содержит те точки ξ, которые «ближе» к точке ζ_k, чем к другим точкам (равноудаленные точки можно по произволу отнести к любой из конкурирующих областей). Если для указанного разбиения (11.1.13) в качестве оценки вместо точки ζ, минимизирующей выражение $\mathbf{M}[c(\xi, \zeta) \mid G_k]$ выбрать точку ζ_k, при помощи которой построена область G_k, то это будет сопряжено с некоторой оптимальностью, т. е.

$$\mathbf{M} \inf_{\zeta} \mathbf{M}[c(\xi, \zeta) \mid G_k] \leqslant \mathbf{MM}[c(\xi, \zeta_k) \mid G_k]. \tag{11.1.14}$$

Сравнивая левую часть неравенства (11.1.14) с (11.1.10), очевидно, имеем

$$\widetilde{R}_n(n\widetilde{I}_1) \leqslant \mathbf{M} \inf_{\zeta} \mathbf{M}[c(\xi, \zeta) \mid G_k] \leqslant \mathbf{MM}[c(\xi, \zeta_k) \mid G_k].$$

В выражении $\mathbf{MM}[c(\xi, \zeta_k) \mid G_k]$ точки $\zeta_1, \ldots, \zeta_{\widetilde{M}}$ предполагаются фиксированными. Записывая его подробнее как $\mathbf{M}\{\mathbf{M}[c(\xi, \zeta_k) \mid G_k] \mid \zeta_1, \ldots, \zeta_n\}$, следовательно, имеем неравенство

$$\widetilde{R}_n(n\widetilde{I}_1) \leqslant \mathbf{M}\{\mathbf{M}[c(\xi, \zeta_k) \mid G_k] \mid \zeta_1, \ldots, \zeta_M\}, \tag{11.1.15}$$

которое пригодится в следующем параграфе.

Для дальнейшего полезно напомнить также (см. гл. 9), что экстремальное распределение (11.1.12) формулой (9.4.23), т. е. формулой

$$\gamma_n(\xi) = \ln \int e^{-\beta c(\xi, \zeta)} P_{I_n}(d\zeta) \tag{11.1.16}$$

или

$$\gamma_1(\xi) = \ln \int e^{-\beta c_1(x, u)} P_{I_1}(du), \tag{11.1.17}$$

связано с функциями $\gamma_1(x)$, $\gamma_n(\xi) = \gamma_1(x_1) + \ldots + \gamma_1(x_n)$. Последние после усреднения дают потенциалы

$$\Gamma_1(\beta) = \mathbf{M}\gamma_1(x) = \int P(dx) \ln \int e^{-\beta c_1(x, u)} P_{I_1}(du), \tag{11.1.18}$$

$$\Gamma_n(\beta) = \mathbf{M}\gamma_n(\xi) = n\Gamma_1(\beta), \tag{11.1.19}$$

позволяющие вычислить средние штрафы

$$R_n = -n \frac{d\Gamma_1}{d\beta}(\beta); \qquad R_1 = -\frac{d\Gamma_1}{d\beta}(\beta) \tag{11.1.20}$$

и количество информации

$$I_n = nI_1; \qquad I_1 = \beta \frac{d\Gamma_1}{d\beta}(\beta) - \Gamma_1(\beta) \tag{11.1.21}$$

согласно (9.4.10), (9.4.29), (9.4.30).

11.2. Теорема об асимптотической равноценности различных количеств информации

1. Прежде чем переходить к основным асимптотическим результатам, касающимся равноценности различных родов информации, выведем полезные для дальнейшего неравенства (11.2.5), (11.2.13), позволяющие оценить сверху минимальные потери $\widetilde{R}_n(n\widetilde{I}_1)$, соответствующие хартлиевскому количеству информации. Воспользуемся неравенством (11.1.15), которое справедливо для любого набора кодовых точек $\zeta_1, \ldots, \zeta_M$. Оно останется справедливым, если произвести усреднение по статистическому ансамблю кодовых точек, описываемому вероятностями $P_{I_n}(d\zeta_1) \ldots P_{I_n}(d\zeta_{\widetilde{M}})$, $\widetilde{M} = [e^{n\widetilde{I}_1}]$. При этом будем иметь

$$\widetilde{R}_n(n\widetilde{I}_1) \leqslant \int \mathbf{M}\{\mathbf{M}[c(\xi, \zeta_k) \mid G_k] \mid \zeta_1, \ldots, \zeta_{\widetilde{M}}\} \times \\ \times P_{I_n}(d\zeta_1) \ldots P_{I_n}(d\zeta_{\widetilde{M}}) \equiv L. \tag{11.2.1}$$

Запишем выражение L в правой части подробнее. Поскольку G_k — область точек ξ, где «расстояние» $c(\xi, \zeta_k)$ не превосходит «расстояния» $c(\xi, \zeta_i)$ до любой точки ζ_i из $\zeta_1, \ldots, \zeta_{\widetilde{M}}$, имеем

$$\mathbf{M}[c(\xi, \zeta_k) \mid G_k] = \frac{1}{P(G_k)} \int\limits_{\substack{c(\xi, \zeta_k) \leqslant c(\xi, \zeta_1) \\ \cdots\cdots\cdots \\ c(\xi, \zeta_k) \leqslant c(\xi, \zeta_{\widetilde{M}})}} c(\xi, \zeta_k) P(d\xi).$$

и, следовательно, также

$$\mathbf{M}\{\mathbf{M}[c(\xi, \zeta_k) \mid G_k] \mid \zeta_1, \ldots, \zeta_{\widetilde{M}}\} = \sum_k P(G_k) \mathbf{M}[c(\xi, \zeta_k) \mid G_k] =$$

$$= \sum_k \int\limits_{\substack{c(\xi, \zeta_k) \leqslant c(\xi, \zeta_1) \\ \cdots\cdots\cdots \\ c(\xi, \zeta_k) \leqslant c(\xi, \zeta_{\widetilde{M}})}} c(\xi, \zeta_k) P(d\xi).$$

Усредняя по $\zeta_1, \ldots, \zeta_{\widetilde{M}}$, будем иметь

$$L = \sum_k \int\limits_{\substack{c(\xi, \zeta_k) \leqslant c(\xi, \zeta_1) \\ \cdots\cdots\cdots \\ c((\xi, \zeta_k) \leqslant c(\xi, \zeta_{\widetilde{M}})}} \!\!\cdots\!\! \int c(\xi, \zeta_k) P(d\xi) P_{I_n}(d\zeta_1) \ldots P_{I_n}(d\zeta_{\widetilde{M}}). \tag{11.2.2}$$

В каждом k-м члене удобно в первую очередь произвести интегрирование по точкам ζ_i, не совпадающим с ζ_k. Если ввести функцию распределения

$$1 - F_\xi(\lambda) = \begin{cases} \int\limits_{c(\xi, \zeta) \geqslant \lambda} P_{I_n}(d\zeta) & \text{при } \lambda \geqslant 0, \\ \int\limits_{c(\xi, \zeta) > \lambda} P_{I_n}(d\zeta) & \text{при } \lambda < 0, \end{cases} \tag{11.2.3}$$

то после $(\tilde{M}-1)$-кратного интегрирования по точкам ζ_i, не совпадающим с ζ_k, из (11.2.2) будем иметь

$$L \leqslant \sum_{k=1}^{\tilde{M}} \int_{\xi} \int_{\zeta_k} c(\xi, \zeta_k) [1-F_{\xi}(c(\xi, \zeta_k))]^{\tilde{M}-1} P(d\xi) P_{I_n}(d\zeta_k) =$$

$$= \tilde{M} \int_{\xi} \int_{\zeta} c(\xi, \zeta) [1-F_{\xi}(c(\xi, \zeta))]^{\tilde{M}-1} P(d\xi) P_{I_n}(d\zeta). \qquad (11.2.4)$$

Знак неравенства возник по той причине, что при $c(\xi, \zeta_k) \geqslant 0$ мы несколько расширили области G_i $(i \neq k)$, присоединив к ним все «спорные» точки ξ, для которых $c(\xi, \zeta_k) = c(\xi, \zeta_i)$, а при $c(\xi, \zeta_k) < 0$ сузили эти области, отбросив все подобные точки.

Нетрудно видеть также, что (11.2.4) можно записать в виде

$$L \leqslant \int P(d\xi) \int_{-\infty}^{\infty} \lambda d\{1-[1-F_{\xi}(\lambda)]^{\tilde{M}}\},$$

или в силу (11.2.1)

$$\tilde{R}_n(n\tilde{I}_n) \leqslant \int P(d\xi) \int \lambda dF_1(\lambda). \qquad (11.2.5)$$

В правой части здесь усредняется по ξ математическое ожидание, соответствующее функции распределения

$$F_1(\lambda) = 1-[1-F_{\xi}(\lambda)]^{\tilde{M}}. \qquad (11.2.6)$$

Очевидно, что

$$1-F_1 = [1-F_{\xi}]^{M} \leqslant 1-F_{\xi}, \text{ т. е. } F_1(\lambda) \geqslant F_{\xi}(\lambda), \qquad (11.2.7)$$

поскольку $1-F_{\xi}(\lambda) \leqslant 1$. Вместо $F_1(\lambda)$ удобно ввести близкую к ней (при больших n) функцию распределения

$$F_2(\lambda) = \max\{1-e^{-\tilde{M}F_{\xi}(\lambda)}, F_{\xi}(\lambda)\}. \qquad (11.2.8)$$

Нетрудно доказать, что

$$F_1 = 1-(1-F_{\xi})^{\tilde{M}} \geqslant 1-e^{-\tilde{M}F_{\xi}}. \qquad (11.2.9)$$

В самом деле, используя неравенство

$$1-F_{\xi} \leqslant e^{-F_{\xi}},$$

имеем

$$(1-F_{\xi})^{\tilde{M}} \leqslant e^{-\tilde{M}F_{\xi}}, \qquad (11.2.10)$$

что эквивалентно (11.2.9).

Используя (11.2.7), (11.2.9), получаем, что функция (11.2.8) не превосходит функции (11.2.6):

$$F_2(\lambda) \leqslant F_1(\lambda). \qquad (11.2.11)$$

Из последнего неравенства вытекает

$$\int \lambda dF_1(\lambda) \leqslant \int \lambda dF_2(\lambda). \tag{11.2.12}$$

Чтобы в этом убедиться, можно учесть, например, что из (11.2.11) следует неравенство $\lambda_1(F) \leqslant \lambda_2(F)$ для обратных функций, через которые разность $\int \lambda dF_2 - \int \lambda dF_1$ записывается в виде $\int\limits_0^1 [\lambda_2(F) - \lambda_1(F)]\, dF$, и, следовательно, является неотрицательной.

Итак, неравенство (11.2.5) только усилится, если в правой части $\int \lambda dF_1$ заменить на $\int \lambda dF_2$. Следовательно, будем иметь

$$\widetilde{R}_n(n\widetilde{I}_1) \leqslant \int P(d\xi) \int \lambda dF_2(\lambda), \tag{11.2.13}$$

где $F_2(\lambda)$ определяется формулой (11.2.8) при $\widetilde{M} = [e^{n\widetilde{I}_1}]$. Полученное неравенство будет использовано в § 11.3.

2. Теорема 11.1. *Пусть имеются бейесовские системы* $[P_n(d\xi), c(\xi, \zeta)]$, *являющиеся n-й степенью системы* $[P_1(dx), c_1(x, u)]$, *причем функция штрафа ограничена сверху*:

$$|c_1(x, u)| \leqslant K_1. \tag{11.2.14}$$

Тогда удельные потери, соответствующие различным родам информации, в пределе совпадают:

$$\frac{1}{n}\widetilde{R}_n(nI_1) \to R_1(I_1) \tag{11.2.15}$$

при $n \to \infty$. *Иначе говоря, имеет место асимптотическая равноценность*

$$\widetilde{V}_n(nI_1)/V_n(nI_1) \to 1 \tag{11.2.16}$$

шенноновского и хартлиевского количеств информации. Предполагается, что входящие в (11.2.15) *удельные потери конечны и функция* $R_1(I_1)$ *непрерывна.*

При доказательстве этой теоремы будем следовать Шеннону [4], сформулировавшему этот результат в других терминах.

Доказательство. 1) Вследствие (11.1.4) при экстремальном распределении (11.1.6) случайная информация

$$I(\xi, \zeta) = \sum_{i=1}^{n} I(x_i, u_i)$$

представляет собой сумму независимых случайных величин. Такой же вид имеет и функция штрафов (11.1.5). Каждая из них имеет конечное математическое ожидание. Применяя закон больших чисел (теорему Хинчина), имеем

$$\mathbf{P}\left\{\left|\frac{1}{n}c(\xi, \zeta) - \frac{1}{n}R_n(nI_1)\right| < \varepsilon\right\} \to 1,$$

$$\mathbf{P}\left\{\left|\frac{1}{n}I(\xi,\zeta)-I_1\right|<\varepsilon\right\}\to 1$$

при $n\to\infty$ при любом $\varepsilon>0$.

Следовательно, каково бы ни было $\delta>0$ при достаточно больших $n>n(\varepsilon,\ \delta)$, будем иметь

$$\mathbf{P}\{c(\xi,\zeta)/n<R_1(I_1)+\varepsilon\}>1-\delta^2/2,$$
$$\mathbf{P}\{I(\xi,\zeta)/n<I_0+\varepsilon\}>1-\delta^2/2. \tag{11.2.17}$$

Обозначим через Γ множество пар $(\xi,\ \zeta)$, для которых одновременно выполняются оба неравенства

$$c(\xi,\zeta)<n\,[R_1(I_1)+\varepsilon], \tag{11.2.18}$$
$$I(\xi,\zeta)<n\,(I_1+\varepsilon). \tag{11.2.19}$$

Тогда из (11.2.17) получаем

$$P(\Gamma)>1-\delta^2. \tag{11.2.20}$$

В самом деле, вероятность события $\overline{A}$, противоположного события (11.2.18), а также вероятность события $\overline{B}$, противоположного (11.2.19), не больше $\delta^2/2$. Поэтому для их объединения имеем

$$P(\overline{A}\cup\overline{B})\leqslant P(\overline{A})+P(\overline{B})\leqslant\delta^2.$$

Событие же, противоположное $\overline{A}\cup\overline{B}$, есть не что иное, как Γ, и, следовательно, справедливо (11.2.20).

Пусть для фиксированного ξ множество Z_ξ есть множество тех ζ, которые в паре с ξ входят в Γ, другими словами, Z_ξ — сечение множества Γ. При таком определении $P(Z_\xi\,|\,\xi)=P(\Gamma\,|\,\xi)$.

Рассмотрим множество Ξ тех элементов ξ, для которых

$$P(Z_\xi\,|\,\xi)>1-\delta. \tag{11.2.21}$$

Из (11.2.20) следует, что

$$P(\Xi)>1-\delta. \tag{11.2.22}$$

Чтобы в этом убедиться, предположим противное, т. е., что

$$P(\Xi)<1-\delta;\qquad P(\overline{\Xi})\geqslant\delta \tag{11.2.22а}$$

($\overline{\Xi}$ — дополнение к Ξ). Вероятность $P\{\overline{\Gamma}\}$ дополнения к Γ можно в этом случае оценить следующим образом. Воспользуемся представлением

$$P(\overline{\Gamma})=\int P(\overline{\Gamma}\,|\,\xi)\,P(d\xi),$$

причем разобъем этот интеграл на две части и оставим лишь один подынтеграл:

$$P(\overline{\Gamma})=\int_{\Xi}P(\overline{\Gamma}\,|\,\xi)\,P(d\xi)+\int_{\overline{\Xi}}P(\overline{\Gamma}\,|\,\xi)\,P(d\xi)\geqslant\int_{\overline{\Xi}}P(\overline{\Gamma}\,|\,\xi)\,P(d\xi). \tag{11.2.22б}$$

В дополнении $\bar{\Xi}$, очевидно, выполняется неравенство, противоположное (11.2.21):

$P(Z_\xi \mid \xi) \equiv P(\Gamma \mid \xi) \leqslant \delta$, т. е. $P(\bar{\Gamma} \mid \xi) > \delta$, при $\xi \in \bar{\Xi}$. Подставляя эту оценку в (11.2.22б) и учитывая (11.2.22а), находим

$$P(\bar{\Gamma}) > \delta \int_{\bar{\Xi}} P(d\xi) = \delta P(\bar{\Xi}) \geqslant \delta^2.$$

Это противоречит (11.2.20) и, следовательно, предположение (11.2.22а) неверно.

Неравенство (11.2.19) означает, что

$$\ln \frac{P(\zeta \mid \xi)}{P(\zeta)} < n(I_1 + \varepsilon),$$

т. е.

$$P(\zeta) > e^{-n(I_1+\varepsilon)} P(\zeta \mid \xi).$$

Суммируя по $\zeta \in Z_\xi$, отсюда получаем

$$P(Z_\xi) > e^{-n(I_1+\varepsilon)} P(Z_\xi \mid \xi),$$

или, используя (11.2.21),

$$P(Z_\xi) > e^{-n(I_1+\varepsilon)}(1-\delta) \quad (\xi \in \Xi) \text{ при } n > n(\xi, \delta). \tag{11.2.23}$$

2) Используем теперь формулу (11.2.5). Возьмем функцию распределения

$$F_3(\lambda) = \begin{cases} 0 & \text{при } \lambda < nR_1(I_1) + n\varepsilon, \\ F_1(nR_1(I_1) + n\varepsilon) & \text{при } nR_1(I_1) + n\varepsilon < \lambda < nK_1, \\ 1 & \text{при } \lambda \geqslant nK_1. \end{cases} \tag{11.2.24}$$

Вследствие условия ограниченности (11.2.14) функции штрафа, вероятность выполнения неравенства $c(\xi, \zeta) > nK_1$ равна нулю и из (11.2.3) получаем $F_\xi(nK_1) = 1$. Вследствие (11.2.6) следует, что $F_1(nK_1) = 1$. Таким образом, функции $F_1(\lambda)$ и $F_3(\lambda)$ на участке $\lambda \geqslant nK_1$ совпадают. На участке $nR_1(I_1) + n\varepsilon < \lambda < nK_1$ имеем $F_3(\lambda) \leqslant F_1(\lambda)$, поскольку

$$F_3(\lambda) = F_1(nR_1(I_1) + n\varepsilon) \leqslant F_1(\lambda) \text{ при } nR_1(I_1) + n\varepsilon \leqslant \lambda$$

в силу неубывающего характера функции $F_1(\lambda)$. Следовательно, неравенство $F_3(\lambda) \leqslant F_1(\lambda)$ справедливо при всех значениях λ. Из него следует обратное неравенство для средних:

$$\int \lambda dF_3(\lambda) \geqslant \int \lambda dF_1(\lambda),$$

подобно тому как (11.2.12) следовало из (11.2.11). Поэтому формула (11.2.5) дает

$$\tilde{R}_n(n\tilde{I}_1) \leqslant \int P(d\xi) \int \lambda dF_3(\lambda), \tag{11.2.25}$$

но

$$\int \lambda dF_3(\lambda) = [nR_1(I_1)+n\varepsilon]\, F_1(nR_1(I_1)+n\varepsilon) + \\ + nK_1[1-F_1(nR_1(I_1)+n\varepsilon)]$$

согласно (11.2.24). Следовательно, из (11.2.25) при учете (11.2.6) получаем

$$\frac{1}{n}\widetilde{R}_n(n\widetilde{I}_1) \leqslant R_1(I_1)+\varepsilon+\int [K_1-R_1(I_1)-\varepsilon]\,[1- \\ -F_\xi(nR_1(I_1)+n\varepsilon)]^{\widetilde{M}}\, P(d\xi)$$

или

$$\frac{1}{n}\widetilde{R}_n(n\widetilde{I}_1) \leqslant R_1(I_1)+\varepsilon+[K_1-R_1(I_1)-\varepsilon]\times \\ \times \int P(d\xi)\, e^{-\widetilde{M}F_\xi(nR_1(I_1)+n\varepsilon)}, \qquad (11.2.26)$$

если учесть (11.2.10). Выполнение предполагаемого при этом неравенства $K_1 - R_1(I_1) - \varepsilon > 0$ обеспечивается увеличением K_1.

3) Вследствие (11.2.3) имеем

$$F_\xi(nR_1(I_1)+n\varepsilon) = \int\limits_{c(\xi,\zeta)<nR_1(I_1)+n\varepsilon} P(d\zeta) \\ \text{при } nR_1(I_1)+n\varepsilon \geqslant 0; \qquad (11.2.26а)$$

$$F_\xi(n\,R_1(I_1)+n\varepsilon) = \int\limits_{c(\xi,\zeta)\leqslant nR_1(I_1)+n\varepsilon} P(d\zeta) \\ \text{при } nR_1(I_1)+n\varepsilon < 0,$$

где ξ фиксировано. Сравним эту величину с вероятностью

$$P(Z_\xi) = \int\limits_{Z_\xi} P(d\zeta), \qquad (11.2.26б)$$

фигурирующей в (11.2.23). Для значений ζ, входящих в множество Z_ξ согласно определению, данному ранее, выполняется неравенство (11.2.18) (а также (11.2.19)). Поэтому область интегрирования Z_ξ в (11.2.26б) является частью области интегрирования в (11.2.26а) и, следовательно,

$$F_\xi(nR_1(I_1)+n\varepsilon) \geqslant P(Z_\xi).$$

Подставляя сюда (11.2.23), получаем

$$F_\xi(nR_1(I_1)+n\varepsilon) > e^{-n(I_1+\varepsilon)}(1-\delta) \qquad (\xi\in\Xi). \qquad (11.2.27)$$

Разобьем интеграл в (11.2.26) на два: интеграл по множеству Ξ и дополнительному множеству $\bar{\Xi}$. В первом из них используем (11.2.27), а во втором заменим экспоненту на единицу. Это даст

$$\int P(d\xi)\exp[-\widetilde{M}F_\xi(nR_1(I_1)+n\varepsilon)] \leqslant$$

$$\leqslant \int_{\Xi} P(d\xi) \exp[-\tilde{M} e^{-nI_1-n\varepsilon}(1-\delta)] + 1 - P(\Xi) \leqslant$$

$$\leqslant \exp[-\tilde{M} e^{-nI_1-n\varepsilon}(1-\delta)] + \delta, \qquad (11.2.28)$$

где учтено неравенство (11.2.22). Здесь

$$\tilde{M} = [e^{n\tilde{I}_1}] \geqslant e^{n\tilde{I}_1} - 1,$$

причем целесообразно положить $\tilde{I}_1 = I_1 + 2\varepsilon$, $\varepsilon > 0$. Тогда неравенство (11.2.28) примет вид

$$\int P(d\xi) \exp[-MF_{\xi}(nR_1(I_1) + n\varepsilon)] \leqslant$$
$$\leqslant \exp\{-e^{n\varepsilon}(1-\delta)(1-e^{-n\tilde{I}_1})\} + \delta.$$

Подставляя последнее в (11.2.26), будем иметь

$$\frac{1}{n}\tilde{R}_n(n\tilde{I}_1) \leqslant R_1(\tilde{I}_1 - 2\varepsilon) + \varepsilon +$$
$$+ 2K_1\delta + 2K_1 \exp\{-e^{n\varepsilon}(1-\delta)(1-e^{-n\tilde{I}_1})\} \qquad (11.2.29)$$

(использовано, что $|R_1| < K_1$ в силу (11.2.14)).

Итак, мы получили, что при любых $\tilde{I}_1$, $\varepsilon > 0$, $\delta > 0$, не зависящих от n, найдется такое n (ε, δ), что неравенство (11.2.29) будет выполняться для всех $n > n$ (ε, δ). Это значит

$$\limsup_{n\to\infty} \frac{1}{n}\tilde{R}_n(n\tilde{I}_1) \leqslant R_1(\tilde{I}_1 - 2\varepsilon) + \varepsilon +$$
$$+ 2K_1\delta + \lim_{n\to\infty} 2K_1 \exp\{-e^{n\varepsilon}(1-\delta)(1-e^{-n\tilde{I}_1})\},$$

но

$$\lim_{n\to\infty} \exp\{-e^{n\varepsilon}(1-\delta)(1-e^{-n\tilde{I}_1})\} = 0$$

(берется $\delta < 1$). Поэтому имеем
но

$$\limsup_{n\to\infty} \frac{1}{n}\tilde{R}_n(n\tilde{I}_1) - R_1(\tilde{I}_1) \leqslant R_1(\tilde{I}_1 - 2\varepsilon) -$$
$$- R_1(\tilde{I}_1) + \varepsilon + 2K_1\delta. \qquad (11.2.30)$$

Вследствие непрерывности функции $R_1(I_1)$ подбором достаточно малых ε и δ выражение, стоящее в правой части (11.2.30), может быть сделано сколь угодно малым. Следовательно,

$$\lim_{n\to\infty} \sup \frac{1}{n}\tilde{R}_n(n\tilde{I}_1) \leqslant R_1(\tilde{I}_1).$$

Эта формула в сочетании с неравенством

$$\lim_{n\to\infty} \inf \frac{1}{n}\tilde{R}_1(n\tilde{I}_1) \geqslant R_1(\tilde{I}_1), \qquad (11.2.30a)$$

вытекающим из (11.1.3) и (11.1.8), доказывает соотношение

$$\lim_{n\to\infty}\sup\frac{1}{n}\widetilde{R}_1(n\widetilde{I}_1)=\lim_{n\to\infty}\inf\frac{1}{n}\widetilde{R}_1(\widetilde{I}_1)=R_1(\widetilde{I}_1),$$

т. е. (11.2.15). Доказательство закончено.

3. Доказанная теорема 11.1 допускает естественное обобщение на те случаи, когда рассматриваемая система не является n-й степенью какой-либо элементарной системы, но вместо этого выполнены какие-то другие более общие условия. Соответствующее обобщение аналогично обобщению, которое производится при переходе от теоремы 7.1 к теореме 7.2, при котором требование, чтобы канал был n-й степенью элементарного канала, заменяется требованием информационной устойчивости. Также и в данном случае наложим такое условие на рассматриваемые бейесовские системы, чтобы это не потребовало по существу никаких изменений в вышеприведенном доказательстве. Согласно обычному приему из доказательства надо исключить n и удельные величины I_1, R_1, ..., рассматривая вместо этого лишь комбинации $I = nI_1$, $R = nR_1$ и т. д. Потребуем, чтобы последовательность случайных величин ξ, ζ (зависящих от n или какого-либо другого параметра) была информационно устойчивой в смысле определения, данного в § 7.3. Кроме того, потребуем, чтобы имела место сходимость по вероятности

$$[c(\xi,\zeta)-R]/V(I)\to 0 \text{ при } n\to\infty. \tag{11.2.31}$$

Легко понять, что при этих условиях сохранятся неравенства типа (11.2.17), принимающие теперь вид

$$\mathbf{P}\{c(\xi,\zeta)<R+\varepsilon_1 V(I)\}>1-1/2\delta^2,$$
$$\mathbf{P}\{I(\xi,\zeta)<I+\varepsilon_2 I\}>1-1/2\delta^2$$

для $n>n(\varepsilon_1,\varepsilon_2,\delta)$ причем $\varepsilon_1 V = \varepsilon_2 I = n\varepsilon$. Вместо прежнего соотношения $\widetilde{I}_1 = I_1 + 2\varepsilon$ теперь будем иметь

$$\widetilde{I}=I+2\varepsilon_2 I.$$

Условие ограниченности (11.2.14) возьмем таким:

$$|c(\xi,\zeta)|\leqslant KV(I), \tag{11.2.32}$$

где K не зависит от n. В приведенном доказательстве потребуется лишь изменение способа записи формул. Соотношение (11.2.29) теперь примет вид

$$\widetilde{R}(\widetilde{I})\leqslant R(\widetilde{I}-2\varepsilon_2 I)+\varepsilon_1 V(I)+2KV(I)\delta+$$
$$+2KV(I)\exp\{-e^{\varepsilon_2 I}(1-\delta)(1-e^{-\widetilde{I}})\}$$

или

$$\frac{V(\widetilde{I}-2\varepsilon_2 I)-\widetilde{V}(\widetilde{I})}{V(\widetilde{I}-2\varepsilon_2 I)}\leqslant\varepsilon_1+2K\left[\delta+\exp\{-e^{\varepsilon_2 I}(1-\delta)(1-e^{-\widetilde{I}})\}\right] \tag{11.2.33}$$

при любых не зависящих от n величинах ε_1, ε_2, δ и при $n > n$ $(\varepsilon_1, \varepsilon_2, \delta)$ Вследствие условия A определения информационной устойчивости (см. § 7.3) при любом $\varepsilon_2 > 0$ выражение $\exp\{-e^{\varepsilon_2 I}(1-\delta)(1-e^{-\tilde{I}})\}$ (где $\tilde{I} = (1+2\varepsilon_2)\, I$) при $n \to \infty$ стремится к нулю, если $\delta < 1$. Поэтому в результате предельного перехода $n \to \infty$ из (11.2.33) имеем

$$1-\liminf_{n\to\infty} \frac{\tilde{V}(\tilde{I})}{V(\tilde{I})} \frac{V(\tilde{I})}{V(\tilde{I}-2\varepsilon_2 I)} \leqslant \varepsilon_1 + 2K\delta. \tag{11.2.33a}$$

Предположим, что существует предел

$$\lim \frac{V(\tilde{I})}{V(y\tilde{I})} = \varphi(y), \tag{11.2.34}$$

который представляет собой функцию, непрерывную по y. Тогда из (11.2.33а) будем иметь

$$\liminf_{n\to\infty} \frac{\tilde{V}(\tilde{I})}{V(\tilde{I})} \varphi\left(\frac{1}{1+2\varepsilon_2}\right) \geqslant 1-\varepsilon_1-2K\delta.$$

Вследствие произвольности ε_1, ε_2, δ и непрерывности функции $\varphi(y)$, отсюда получаем

$$\liminf_{n\to\infty} \frac{\tilde{V}(\tilde{I})}{V(\tilde{I})} \geqslant 1\,. \tag{11.2.34a}$$

Для доказательства сходимости

$$\frac{\tilde{V}(\tilde{I})}{V(\tilde{I})} \to 1 \tag{11.2.35}$$

остается только сравнить (11.2.34а) с неравенством

$$\limsup \frac{\tilde{V}(\tilde{I})}{V(\tilde{I})} \leqslant 1,$$

которое служит обобщением (11.2.30а).

4. В § 9.6 отмечалось, что функции ценности информации $V(I)$, $\tilde{V}(I)$ остаются инвариантными при преобразовании

$$c'(\xi, \zeta) = c(\xi, \zeta) + f(\xi) \tag{11.2.36}$$

функции штрафов (см. теорему 9.8). При этом остаются неизменными также разность рисков $\tilde{R}-R$ и области G_k, определенные в § 11.1, если не меняются кодовые точки ζ_k и распределение $P(d\zeta)$, с которым они выбрасываются. Между тем условия (11.2.31), (11.2.32) не остаются инвариантными при указанном преобразовании (11.2.36). Так (11.2.31) переходит в соотношение

$$[c'(\xi, \zeta) - \mathbf{M}c'(\xi, \zeta) - f(\xi) + \mathbf{M}f(\xi)]/V(I) \to 0,$$

которое, очевидно, вообще не совпадает с

$$\left[c'(\xi, \zeta) - \mathbf{M}c'(\xi, \zeta)\right] / V(I) \to 0.$$

Учитывая это, можно воспользоваться свободой выбора функции $f(\xi)$ в преобразовании (11.2.36), чтобы добиться выполнения условий (11.2.31), (11.2.32) в том случае, если они первоначально не выполнялись. Это расширяет круг случаев, для которых удается доказать сходимость (11.2.34), ослабляет условия асимптотической равноценности в вышеизложенной теории.

Специальным подбором функции $f(\xi)$ в (11.2.36) можно вообще устранить надобность в условии (11.2.31). В самом деле, в случае экстремального распределения $P(\zeta \mid \xi)$, как видно из (9.4.21), выполняется соотношение

$$-I(\xi, \zeta) = \beta c(\xi, \zeta) + \gamma(\xi). \qquad (11.2.37)$$

Пользуясь свободой выбора функции $f(\xi)$ в (11.2.36), положим $f(\xi) = \gamma(\xi)/\beta$. Тогда новое «расстояние» $c'(\xi, \zeta)$ совпадает с $-I(\xi, \zeta)/\beta$ и сходимость (11.2.31) для него можно будет не оговаривать, так как она будет вытекать из условия

$$\left[I(\xi, \zeta) - I_{\xi\zeta}\right] / I_{\xi\zeta} \to 0 \text{ (по вероятности)} \qquad (11.2.38)$$

информационной устойчивости, если, конечно, отношение $I/\beta V$ не возрастает до бесконечности. Назовем последовательность бейесовских систем $[P(d\xi), c(\xi, \zeta]$ *информационно устойчивой*, если для экстремальных распределений последовательность случайных величин ξ, ζ является информационно устойчивой. Тогда из вышеизложенного вытекает следующее.

Теорема 11.2 (общий вид третьей асимптотической теоремы). *Если задана информационно устойчивая последовательность бейесовских систем, для которых выполняется соотношение ограниченности*

$$|I(\xi, \zeta)| \leqslant K' V(I)/V'(I) \qquad (11.2.39)$$

(K' *не зависит от* n) *и условие непрерывности по* y *функции* (11.2.34), *то имеет место сходимость* (11.2.35). Условие (11.2.39), как нетрудно понять, возникло из (11.2.32), $K' = \beta K V'(I)$.

5. Используем изложенную выше теорию для более подробного анализа измерительно-передающей системы, рассмотренной в п. 3 § 9.2. Асимптотическая равноценность хартлиевской и шенноновской информации позволяет улучшить работу системы, схема которой дана на рис. 9.5, т. е. снизить средние штрафы от уровня

$$\widetilde{R}(I) = \min_u \mathbf{M}c(x, u) - \widetilde{V}(I), \quad I = \ln m,$$

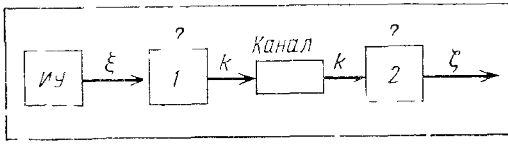
до уровня

$$R(I) = \min_u \mathbf{M}c(x, u) - V(I),$$

определяемого ценностью шенноновской информации.

Чтобы этого достичь, нужно, приводя систему в действие многократно, заменить x на блок $\xi = (x_1, \ldots, x_n)$, а u — на блок $\zeta = (u_1, \ldots, u_n)$ при достаточно большом n. В качестве штрафов должен рассматриваться суммарный штраф. После этого можно применять теорему 11.1 и конструировать блоки *1* и *2*, изображенные на рис. 11.1 по тем же принципам, как это делалось в п. 3 § 9.2. Теперь, однако, зная доказательство теоремы 11.1, мы можем конкретно разобрать их действие. Не стремясь к точной оптимальности, возьмем в качестве требуемого разбиения $\sum_k E_k = X^n$ рассмотренное выше разбиение (11.1.13). Области G_k здесь определяются по принципу близости к случайным кодовым точкам $\zeta_1, \ldots, \zeta_k$. Это раз-

Рис. 11.1.
Блочный вариант системы с информационным ограничением Пропускная способность канала в n раз больше, чем на рис 9 5



биение является, как доказано, асимптотически оптимальным. Другими словами, система (рис. 11.1), в которой блок *1* классифицирует входной сигнал ξ по принадлежности к областям G_k и выдает на выходе номер области, содержащей ξ, работает асимптотически оптимально. Легко видеть, что описанная работа блока *1* совершенно аналогична рассмотренной в гл. 7 работе декодировщика на выходе канала с помехами, только вместо «расстояния» (7.1.8) теперь учитывается «расстояние» $c(\xi, \zeta)$, а также ξ, η заменены на ζ, ξ. Эта аналогия позволяет назвать блок *1* измерительным декодировщиком, блок же *2* работает как блок оптимальных оценок.

Описанная информационная система и системы несколько более общего вида рассматривались в работах Стратоновича [3], Стратоновича и Гришанина [1], Гришанина и Стратоновича [1].

Информационные ограничения типа рассмотренных ранее (рис. 9.6, 11.1) могут учитываться в различных системах оптимальной фильтрации, автоматического регулирования, динамического программирования и даже теории игр. Иногда указанные ограничения связаны с ограниченностью притока информации, иногда с ограниченностью памяти, иногда с ограничением сложности автомата или регулятора. Учет ограничений приводит к проникновению понятий и методов теории информации в упомянутые теории, к срастанию их с теорией информации. В динамическом программировании (см., например, Беллман [1]) и часто в других теориях рассматривается ряд действий, совершаемых последовательно во времени. В связи с этим информационные ограничения в этом случае носят многократный характер. Это приводит к последовательностному обобщению вышеизложенных результатов. Ряд вопросов, относящихся к указанному направлению, изучается в работах Стратоновича [5, 7], Стратоновича и Гришанина [2].

11.3. Быстрота исчезновения различия в ценности шенноновской и хартлиевской информации

Теорема 11.1, приведенная выше, устанавливает самый факт асимптотической равноценности шенноновского и хартлиевского количеств информации. Представляет интерес вопрос о том, как быстро исчезает различие между ценностями указанных количеств. Напомним, что в гл. 7 вслед за теоремами 7.1, 7.2 устанавливающими факт асимптотического исчезновения вероятности ошибки декодирования, были приведены теоремы, в которых исследовалась быстрота исчезновения этой вероятности.

Не представляет сомнений то, что в отношении быстроты исчезновения различия $V(I)-\widetilde{V}(I)$ (аналогично проблеме асимптотической безошибочности) может быть получено большое количество результатов различной степени сложности и силы. При этом могут быть применены различные методы. Мы приведем здесь лишь некоторые сравнительно несложные результаты по данной проблеме. Подсчитаем первые члены асимптотического разложения разности $V(I)-\widetilde{V}(I)$ по степеням малого параметра n^{-1}. При этом будет выяснено, что важная для доказательства теоремы 11.1 оговорка об ограниченности (11.2.14) функции штрафов не является существенной для асимптотической равноценности информаций, а обусловлена лишь принятым методом доказательства.

Обратимся к формуле (11.2.13), которую, вводя обозначение $S=\int \lambda dF_2(\lambda)$, запишем в виде

$$\widetilde{R}_n(nI_1)\leqslant \int SP(d\xi). \tag{11.3.1}$$

Здесь и в дальнейшем мы отождествляем $\widetilde{I}$ с I, $\widetilde{I}_1$ с I_1, $\widetilde{M}$ с M, поскольку теперь нет надобности проводить различие между ними.

Займемся асимптотическим подсчетом выражения, стоящего в правой части последнего неравенства.

1. Учитывая (11.2.8), имеем

$$S=S_1+S_2+S_3, \tag{11.3.2}$$

где

$$S_1=-\int\limits_{\lambda=-\infty}^{\bar{c}} \lambda d e^{-MF_\xi(\lambda)},\quad S_2=-\int\limits_{\lambda=\bar{c}}^{\infty} \lambda d e^{-MF_\xi(\lambda)}, \tag{11.3.3}$$

$$S_3=\int\limits_{F_\xi(\lambda)>1-e^{-MF_\xi(\lambda)}} \lambda d\left[F_\xi(\lambda)-1+e^{-MF_\xi(\lambda)}\right]. \tag{11.3.4}$$

Здесь удобно положить

$$\bar{c}=\int c(\xi,\zeta)P_{I_n}(d\zeta).$$

Входящую в эти формулы функцию распределения

$$F_\xi(\lambda)=\mathbf{P}\left[c(\xi, \zeta)\leqslant\lambda\right]$$

[ξ зафиксировано, $\mathbf{P}$ соответствует $P_{I_n}(d\zeta)$, см. (11.2.3)] оценим при помощи теоремы 4.8, точнее, того ее обобщения, о котором говорится немедленно после ее доказательства. Для простоты будем предполагать, что отрезок $[s_1, s_2]$, упомянутый в теореме 4.8, достаточно велик, так что уравнение (4.4.17) (11.3.7) имеет корень s при любых значениях $\lambda=x$. Тогда

$$F_\xi(\lambda)=\left[\ 2\pi\left[\mu_\xi''(s)s^2\right]^{-1/2} e^{-s\mu_\xi'(s)+\mu_\xi(s)}\left[1+O(\mu_\xi^{-1})\right]\right. \quad (11.3.5)$$

при $\lambda<\int c(\xi, \zeta)P_{I_n}(d\zeta)=\bar{c}$, а также

$$F_\xi(\lambda)=1-\left[\ 2\pi\ \left[\mu_\xi''(s)s^2\right]^{-1/2} e^{-s\mu_\xi'(s)+\mu_\xi(s)}\times\right.$$
$$\times\left[1+O(\mu_\xi^{-1})\right] \qquad \text{при } \lambda>\bar{c}, \quad (11.3.6)$$

где s — корень уравнения

$$\mu_\xi'(s)=\lambda, \quad (11.3.7)$$

а

$$\mu_\xi(t)=\ln\int e^{tc(\xi, \zeta)}P_{I_n}(d\zeta). \quad (11.3.8)$$

Различные выражения (11.3.5), (11.3.6) соответствуют различным знакам корня s.

В силу (11.3.5) — (11.3.7) первый интеграл (11.3.3) преобразуется к виду

$$S_1=-\int\limits_{s=s_*}^{0}\mu_\xi'(s)\,d\exp\left\{-\exp\left(nI_1-s\mu_\xi'(s)+\right.\right.$$
$$\left.\left.+\mu_\xi(s)-\ldots-\frac{1}{2}\ln\left[2\pi\,\mu_\xi''(s)s^2\right]\right)\left[1+O(n^{-1})\right]\right\}, \quad (11.3.9)$$

где интегрирование ведется по s, нижний предел интегрирования s_* определяется формулой $\lim\limits_{s\to s_*}\boldsymbol{\mu}_\xi'(s)=-\infty$.

Для большей наглядности (может быть, несколько условно) обозначаем $I_1=I/n$; $\boldsymbol{\mu}_1=\boldsymbol{\mu}_\xi/n$ и пишем $O(n^{-1})$ вместо $O(\boldsymbol{\mu}_\xi^{-1})$. Чтобы оценить указанный интеграл, рассмотрим точку r оси s, определяемую уравнением

$$I_1-r\mu_1'(r)+\mu_1(r)-\frac{1}{2}\ln\left[2\pi n\mu_1''(r)r^2\right]=0, \quad (11.3.10)$$

где $\boldsymbol{\mu}_1(t)=\frac{1}{n}\boldsymbol{\mu}_\xi(t)$, и произведем разложение функций, входящих в (11.3.9) в ряд Тейлора относительно этой точки:

$$\mu_1'(s)=\boldsymbol{\mu}_1'(r)+\mu_1''(r)(s-r)+\frac{1}{2}\mu_1'''(r)(s-r)^2+\ldots,$$

$$s\mu_1'(s)-\mu_1(s)-r\mu_1'(r)-\mu_1(r)+r\mu_1''(r)(s-r)+$$
$$+\frac{1}{2}\left[r\mu_1'''(r)+\mu_1''(r)\right](s-r)^2+\ldots,$$

$$\ln\left[2\pi n\mu_1''(s)s^2\right]=\ln\left[2\pi n\mu_1''(r)r^2\right]+$$
$$+\left[\mu_1'''(r)/\mu_1''(r)+2/r\right](s-r)+\ldots$$

Подставляя эти разложения в интеграл (11.3.9), получаем

$$\frac{1}{n}S_1=\mu_1'(r)\left[1-e^{-MF_\xi(\bar{c})}\right]-$$
$$-\mu_1''(r)\int\limits_{x=s_*-r}^{-r}xd\exp\left\{-e^{n\alpha x+n\beta x^2+\cdots}\left[1+O(n^{-1})\right]\right\}-$$
$$-\frac{1}{2}\mu_1'''(r)\int\limits_{x=s_*-r}^{-r}x^2d\exp\left\{-e^{n\alpha x+n\beta x^2+\cdots}\left[1+O(n^{-1})\right]\right\}+\ldots,$$
$$(x=s-r), \qquad (11.3.11)$$

где с нужной нам точностью

$$\alpha=-r\mu_1''(r)+O(n^{-1});\ \ 2\beta=-r\mu_1'''(r)-\mu_1''(r)+O(n^{-1}). \qquad (11.3.12)$$

Выбираем отрицательный корень

$$r<0 \qquad (11.3.13)$$

уравнения (11.3.10) (если он имеется), так что $\alpha>0$. Делая замену переменной $e^{n\alpha x}=z$, приводим (11.3.11) к виду

$$\frac{1}{n}S_1-\mu_1'(r)\left[1-e^{-MF_\xi(\bar{c})}\right]-\frac{\mu_1''(r)}{n\alpha}\int\limits_{z=z_*}^{e^{-n\alpha r}}\ln z\times$$
$$\times d\exp\left\{e^{-\ln z+(\beta/n\alpha^2)\ln^2 z+\cdots}\left[1+O(n^{-1})\right]\right\}-$$
$$-\frac{1}{2}\frac{\mu_1''(r)}{n^2\alpha^2}\int\limits_{z=z_*}^{e^{-n\alpha r}}\ln^2 zd\exp\left\{e^{-\ln z+(\beta/n\alpha^2)\ln^2 z+\cdots}\times\right.$$
$$\left.\times\left[1+O(n^{-1})\right]\right\}-\ldots, \qquad (11.3.14)$$
$$(z_*=e^{na}(s_*-r)),$$

из которой видна относительная величина тех или иных членов при больших n. Разлагая экспоненту

$$\exp\left[-e^{\ln z+(\beta/n\alpha^2)\ln^2 z+\cdots}\right]\equiv\exp\left[-e^{\ln z+\varepsilon}\right]$$

в ряд Тейлора по $\varepsilon=(\beta/n\alpha^2)\ln^2 z+\ldots$, имеем

$$\exp(-ze^{\varepsilon})=\exp\left[-z-z\varepsilon-\frac{1}{2}z\varepsilon^2-\ldots\right]=$$

$$= e^{-z}\left[1 - z\varepsilon\left(1+\frac{\varepsilon}{2}+\ldots\right)+\frac{1}{2}z^2\varepsilon^2\times\right.$$

$$\left.\times\left(1+\frac{\varepsilon}{2}+\ldots\right)^2-\ldots\right]. \qquad (11.3.15)$$

Подставляя (11.3.15) в (11.3.14), удержим лишь члены порядка 1, n^{-1}, n^{-2} и получим

$$\frac{1}{n}S_1 = \mu_1'(r)\left[1-e^{-n\xi(\bar{c})}\right]-\frac{\mu_2''(r)}{n\alpha}\int\limits_{z=-z_*}^{e^{-n\alpha r}}\ln z\times$$

$$\times d\left[e^{-z}-z\frac{\beta}{n\alpha^2}\ln^2 z e^{-z}\right]-\frac{\mu_1'''(r)}{2n^2\alpha^2}\int\limits_{z=z_*}^{e^{-MF\alpha r}}\ln^2 z\, n\, de^{-z}+n^{-3}\ldots$$

Если придерживаться меньшей точности, то будет иметь место более простая формула

$$\frac{1}{n}S_1 = \mu_1'(r)-\frac{\mu_1''(r)}{n\alpha}\int\limits_0^\infty \ln z\, de^{-z}+n^{-2}\ldots$$

или, если учесть (11.3.12),

$$\frac{1}{n}S_1 - \mu_1'(r)+\frac{1}{nr}\int\limits_0^\infty \ln z\, de^{-z}+n^{-2}\ldots \qquad (11.3.16)$$

Здесь мы пренебрегли членом $-\mu_1(r)e^{-MF\xi(\bar{c})}$, который убывает с ростом n и $M = [e^{nI_1}]$ несравненно быстрее, чем n^{-2}, а также интегральными выражениями

$$\frac{\mu_1''(r)}{n\alpha}\int\limits_{e^{-n\alpha r}}^\infty \ln z e^{-z}\,dz; \qquad \frac{\mu_1''(r)}{n\alpha}\int\limits_0^{z_*}\ln z\, e^{-z}\,dz,$$

которые весьма быстро убывают с ростом n, так как $e^{-n\alpha r}$ и $z_* = e^{n\alpha(s_*-r)}$ экспоненциально стремятся к ∞ и 0 соответственно (напомним, что $r<0$, $s_*<r$). Легко оценить величину указанных интегралов, например, опустив сравнительно мало влияющие сомножители $\ln z$ в первом интеграле и e^{-z} во втором. Однако мы не будем на этом далее задерживаться.

Интеграл, входящий в (11.3.16), равен постоянной Эйлера $C = 0{,}577\ldots$ В самом деле, представляя его как предел

$$\lim_{a\to\infty}\int\limits_0^a \ln z\, d(e^{-z}-1)$$

и интегрируя по частям:

$$\int_0^a \ln z d(\mathrm{e}^{-z}-1)=(\mathrm{e}^{-z}-1)\ln z\Big|_0^a+\int_0^a (1-\mathrm{e}^{-z})\times$$

$$\times\frac{dz}{z}=(\mathrm{e}^{-a}-1)\ln a+\int_0^1 (1-\mathrm{e}^{-z})\frac{dz}{z}-\int_1^a \mathrm{e}^{-z}\frac{dz}{z}+\ln a,$$

этот интеграл можно привести к виду

$$\int_0^\infty \ln z d\mathrm{e}^{-z}=\int_0^1 (1-\mathrm{e}^{-z})\frac{dz}{z}-\int_1^\infty \mathrm{e}^{-z}\frac{dz}{z}=C$$

(см. Янке и Эмде [1], с. 97). Поэтому найденный результат (11.3.16) принимает вид

$$\frac{1}{n}S_1=\mu_1'(r)+\frac{C}{nr}+n^{-2}\ldots \tag{11.3.17}$$

Вместо корня r уравнения (11.3.10) удобно рассматривать корень q_ξ более простого уравнения

$$q_\xi\,\mu_1'(q_\xi)-\mu_1(q_\xi)=I_1. \tag{11.3.18}$$

Из сопоставления этих уравнений видно, что

$$r-q_\xi=-\frac{1}{2n}\,\frac{\ln[2\pi n\mu_1''(q_\xi)\,q_\xi^2]}{q_\xi\,\mu_1''(q_\xi)}+n^{-2}\ln^2 n\ldots, \tag{11.3.19}$$

и условие существования отрицательного корня (11.3.13) эквивалентно (при больших n) условию существования корня

$$q_\xi<0. \tag{11.3.20}$$

Дифференцированием легко убедиться, что функция $s\mu_\xi'(s)-\mu_\xi(s)$ имеет минимальное значение в точке $s=0$, которое равно нулю. Поэтому уравнение (11.3.18) хотя бы при достаточно малых I_1 имеет два корня: положительный и отрицательный, так что отрицательный корень q_ξ выбрать можно.

Согласно (11.3.19) формула (11.3.17) приводится к виду

$$\frac{1}{n}S_1=\mu_1'(q_\xi)-\frac{1}{2nq_\xi}\ln[2\pi n\mu_1''(q_\xi)\,q_\xi^2]+\frac{C}{nq_\xi}+n^{-2}\ln^2 n\ldots \tag{11.3.21}$$

2. Перейдем к оценке интеграла (11.3.4). Обозначим через $\lambda_\text{г}$ граничную точку, определяемую из уравнения

$$F_\xi(\lambda_\text{г})=1-\mathrm{e}^{-MF_\xi(\lambda_\text{г})}. \tag{11.3.22}$$

При больших M (именно, $M \gg 1$) величина $1 - F(\lambda_г) = \varepsilon$ является малой ($\varepsilon \ll 1$). Преобразуя (11.3.22) к виду

$$\varepsilon = e^{-M(1-\varepsilon)} = e^{-M}\left(1 + \varepsilon M + \frac{1}{2}\varepsilon^2 M^2 + \ldots\right),$$

находим приближенное решение этого уравнения:

$$1 - F_\xi(\lambda_г) = \varepsilon = e^{-M} + Me^{-2M} + \ldots, \qquad (11.3.22а)$$

что подтверждает указанную малость ε. В области $\lambda > \lambda_г$, по которой ведется интегрирование в (11.3.4), производная

$$\frac{d}{d\lambda}\left[F_\xi(\lambda) - 1 + e^{-MF_\xi(\lambda)}\right] = \left(1 - Me^{-M}e^{M[1-F_\xi(\lambda)]} \times\right)$$
$$\times \frac{dF_\xi(\lambda)}{d\lambda}$$

положительная при $M \gg 1$, в чем легко убедиться, учитывая, что

$$1 - F_\xi(\lambda) \leqslant \varepsilon \sim e^{-M} \qquad \text{при } \lambda > \lambda_г.$$

Поэтому интеграл (11.3.4) можно мажорировать следующим образом:

$$S_3 = \int_{\lambda_г}^{\infty} \lambda d\left[F_\xi(\lambda) - 1 + e^{-MF_\xi(\lambda)}\right] \leqslant \int_{\lambda_г}^{\infty} |\lambda| dF_\xi(\lambda) + S_3', \qquad (11.3.23)$$

где $S_3' = \int_{\lambda_г}^{\infty} |\lambda| de^{-MF_\xi(\lambda)}$ — отрицательная добавка, которую отбрасываем.

Начиная с некоторых $M > M(\bar{c})$ значение $\lambda_г$ заведомо превосходит среднее значение, так что в интеграле (11.3.23) можно воспользоваться оценкой (11.3.6). Из этой формулы нетрудно получить

$$1 - F_\xi(\lambda) = [1 - F_\xi(\lambda_г)]\, e^{-n\rho(\lambda - \lambda_г) + \cdots}, \qquad (11.3.24)$$

где

$$\rho = s_г \mu_1''(s_г) > 0 \qquad \left(n\mu_1'(s_г) = \lambda_г\right), \qquad (11.3.25)$$

а точки заменяют другие величины, вид которых для нас несуществен. Подставляя (11.3.24) в (11.3.23), получаем

$$S_3 \leqslant \varepsilon \int_{x=0}^{\infty} |\lambda_г + x|\, d\,e^{-n\rho x + \cdots} \leqslant \varepsilon \int_{x=0}^{\infty} [|\lambda_г| + x]\, d\,e^{-n\rho x + \cdots} =$$
$$= \varepsilon\left[|\lambda_г| + \frac{1}{n\rho} + \ldots\right].$$

Используя (11.3.6), выражение в правой части можно записать в таком виде:

$$S_3 \leqslant e^{-n\widetilde{\mu}(\lambda_г) + \cdots}\, [|\lambda_г| + \ldots]. \qquad (11.3.26)$$

Здесь μ — образ по Лежандру

$$\tilde{\mu}(\lambda) = s\mu_1'(s) - \mu_1(s) \qquad (n\mu_1'(s) = \lambda);$$

точки заменяют менее существенные члены.

Если бы $\lambda_г$ не зависело от n, то согласно (11.3.26) оценка интеграла S_3 исчезала бы с ростом n в основном экспоненциально. Но $\lambda_г$ в силу (11.3.22а), вообще говоря, растет с ростом n, поэтому растет и $\tilde{\mu}(\lambda_г)$ (функция $\tilde{\mu}(\lambda)$, как легко проверить, имеет при $\lambda > \bar{c}$ возрастающий характер). Это еще более увеличивает быстроту исчезновения выражения в правой части (11.3.26). Итак, интеграл S_3 исчезает с ростом n несравненно быстрее, чем члены асимптотического разложения (11.3.21). Поэтому член S_3 в (11.3.2) можно не принимать во внимание.

3. Оценим теперь второй интеграл S_2 в (11.3.3). Выбрав некоторое положительное число $b > |\bar{c}|$, представим его в виде суммы двух членов

$$S_2 = S_2' + S_2''; \tag{11.3.27}$$

$$S_2' = -\int_{\lambda=\bar{c}}^{b} \lambda\, d\, e^{-MF_\xi(\lambda)}; \qquad S_2'' = -\int_{\lambda=b}^{\infty} \lambda\, d\, e^{-MF_\xi(\lambda)}.$$

Очевидно

$$|S_2'| \leqslant b\left[e^{-MF_\xi(\bar{c})} - e^{-MF_\xi(b)}\right] \leqslant b\, e^{-MF_\xi(\bar{c})}. \tag{11.3.28}$$

Для оценки интеграла S_2'' используем (11.3.6) и формулу типа (11.3.24)

$$S_2'' = -\int_b^\infty \lambda\, d\exp\{-M[1-(1-F_\xi(b))\, e^{-n\rho_b(\lambda-b)} + \cdots]\}.$$

Здесь по аналогии с (11.3.25) $\rho_b > 0$. Но

$$-\int_b^\infty \lambda\, d\exp\{-M + N e^{-n\rho_b(\lambda-b)}\} = -e^{-M}\int_0^\infty \left(b + \frac{y}{n\rho_b}\right) \times$$

$$\times\, d\exp\{+N e^{-y}\} = e^{-M}[e^N - 1]\, b -$$

$$-\frac{1}{n\rho_b} e^{-M}\int_0^1 \ln z\, d\, e^{Nz} = T_1 + T_2, \tag{11.3.29}$$

где обозначено $N = M[1 - F_\xi(b)]$, $z = e^{-y}$.

Для первого члена в (11.3.29) имеем неравенство

$$T_1 \equiv e^{-M}[e^N - 1]\, b \leqslant b\, e^{-M+N} = b\, e^{-MF_\xi(b)}. \tag{11.3.30}$$

Вычислим второй член в (11.3.29). Из формулы

$$\int_0^1 e^{-pz} \ln z dz = -\frac{1}{p}(C + \ln p) + \frac{1}{p} \operatorname{Ei}(-p)$$

(Рыжик и Градштейн [1] формулы (3.711.2) и (3.711.3)) аналитическим продолжением получаем

$$\int_0^1 e^{Nz} \ln z dz - \frac{1}{N}(C + \ln N) - \frac{1}{N} \overline{\operatorname{Ei}}(N),$$

причем

$$\overline{\operatorname{Ei}}(N) = \frac{e^N}{N}\left[1 + \frac{1}{N} + \frac{2}{N^2} + \ldots\right]$$

(Е. Янке и Ф. Эмде [1], с. 98). Следовательно, основная зависимость второго члена T_2 в (11.3.29) от M определяется экспоненциальным множителем $e^{-M+N} = e^{-M+\xi(b)}$.

Итак, все три члена (11.3.28), (11.3.30) и T_2, составляющие S_2, убывают с ростом n весьма быстро. Так же, как и S_3, они не оказывают влияния на асимптотическое разложение типа (11.3.21) по степеням малого параметра n^{-1} (в комбинации с логарифмами $\ln n$). В формуле (11.3.2) поэтому нужно учитывать лишь один член S_1, так что в силу (11.3.21) имеем

$$S \leqslant \mu_\xi'(q_\xi) - \frac{1}{2q_\xi} \ln [(2\pi)^{-1} \mu_\xi''(q_\xi) q_\xi^2] + \\ + \frac{C}{q_\xi} + O(n^{-1} \ln^2 n). \tag{11.3.31}$$

Вследствие этого неравенство (11.3.1) принимает вид

$$\frac{1}{n} \widetilde{R}_n(nI_1) \leqslant \int \left\{\mu_1'(q_\xi) - \frac{1}{2nq_\xi} \ln [2\pi n \mu_1''(q_\xi) q_\xi^2] + \frac{C}{nq_\xi}\right\} P(d\xi) + \\ + O(n^{-2} \ln^2 n)\ldots \tag{11.3.32}$$

4. В соответствии с последней формулой проведем усреднение по ξ. Это усреднение облегчается тем, что функция (11.3.8) представляет собой сумму большого числа одинаково распределенных случайных слагаемых $\ln \int e^{tc_1(x_i, u)} P_{I_1}(du)$, а $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_\xi/n$ — их среднее арифметическое:

$$\mu_1(t) = \frac{1}{n} \sum_{i=1}^{n} \ln \int e^{tc_1(x_i, u)} P_{I_1}(du). \tag{11.3.33}$$

В силу закона больших чисел оно сходится к математическому ожиданию

$$\nu_1(-t, \beta) \equiv \mathbf{M}\mu_1(t) = \int P(dx) \ln \int e^{tc_1(x, u)} P_{I_1}(du), \tag{11.3.34}$$

которое просто связано с потенциалом (11.1.18):

$$\nu_1(-t, \beta) \equiv \nu_1(-t) = \Gamma_1(-t). \quad (11.3.35)$$

При каждом фиксированном t согласно указанному закону имеет место сходимость по вероятности:

$$\boldsymbol{\mu}_1(t) \to \nu_1(-t), \quad (11.3.36)$$

а также для производных

$$\mu_1^{(k)}(t) \to (-1)^k \nu_1^{(k)}(-t), \quad (11.37)$$

если последние существуют.

Вследствие сходимости (11.3.36), (11.3.37) корень q_ξ уравнения (11.3.18) будет сходиться по вероятности к корню $q = -z$ уравнения

$$-q\nu_1'(-q) - \nu_1(-q) = I_1 \quad \text{или} \quad z\nu_1'(z) - \nu_1(z) = I_1. \quad (11.3.38)$$

Последнее уравнение совпадает с (11.1.21). Поэтому $z = \beta$, так что

$$q_\xi \to -\beta \quad \text{при} \quad n \to \infty. \quad (11.3.39)$$

Согласно (9.4.37) имеем

$$\beta = -dI/dR,$$

причем для нормальной ветви R_+ (9.3.10) параметр β положителен (см. § 9.3). Отсюда получаем неравенство $q_\xi < 0$, которое согласуется с условием (11.3.20).

Вследствие сходимости (11.3.37), (11.3.39) и равенств (11.3.35), (11.1.20) первый член $\mathbf{M}\boldsymbol{\mu}_1'(q_\xi)$ в (11.3.32) стремится к $R_1(I_1)$. Это доказывает (если учесть еще исчезновение прочих членов) сходимость

$$\frac{1}{n}\widetilde{R}(nI_1) \to R_1(I_1) \qquad (n \to \infty),$$

о которой уже шла речь в теореме (11.1).

Чтобы исследовать быстроту сходимости, рассмотрим отклонение случайных величин, входящих в усредняемое выражение в (11.3.32) от их предельных неслучайных значений. Введем случайное отклонение $\delta\boldsymbol{\mu}_1 = \boldsymbol{\mu}_1(-\beta) - \nu_1(\beta)$. Оно, как и случайное отклонение $\delta\mu_1' = \boldsymbol{\mu}_1'(-\beta) + \nu_1'(\beta)$, в силу (11.3.33) имеет нулевое математическое ожидание и дисперсию порядка n^{-1}

$$\mathbf{M}[\delta\mu_1]^2 = \mathbf{M}\mu_1^2(-\beta) - \nu_1^2(\beta) =$$
$$= \frac{1}{n}\int P(dx) \ln^2 \int e^{-\beta c_1(x, u)} P_{I_1}(du) -$$
$$- \frac{1}{n}\left\{\int P(dx) \ln \int e^{-\beta c_1(x, u)} P_{I_1}(du)\right\}^2$$

и аналогично для $\delta\boldsymbol{\mu}_1'$.

Чтобы при усреднении величины $\mu_1'(q_\xi)$ учесть вклад случайных отклонений, имеющих порядок n^{-1}, представим эту величину как разложение по $\delta\boldsymbol{\mu}$, $\delta\boldsymbol{\mu}'$ с включением квадратичных членов.

Обозначая $q_\xi + \beta = \delta q$ и разлагая правую часть уравнений (11.3.18) по δq с учетом квадратичных членов, будем иметь

$$-\beta\mu_1'(-\beta)-\mu_1(-\beta)-\beta\mu_1''(-\beta)\,\delta q+$$
$$+\frac{1}{2}\left[\mu_1''(-\beta)-\beta\mu_1'''(-\beta)\right]\delta q^2+\ldots=I_1.$$

Подставляя сюда

$$\mu_1(-\beta)=\nu_1(\beta)+\delta\mu_1;\qquad \mu_1'(-\beta)=-\nu_1'(\beta)+\delta\mu_1';$$
$$\mu_1''(-\beta)=\nu_1''(\beta)+\delta\mu_1'',$$

и отбрасывая члены более высокого порядка, чем квадратичные получаем

$$\beta\nu_1'(\beta)-\nu_1(\beta)-\beta\delta\mu_1'-\delta\mu_1-\beta\delta\mu_1''\,\delta q-\beta\nu_1''(\beta)\,\delta q+$$
$$+\frac{1}{2}\left[\nu_1''(\beta)+\beta\nu_1'''(\beta)\right]\delta q^2+\ldots=I_1.$$

Члены нулевого порядка здесь сокращаются в силу (11.3.38) где $z=\beta$, и мы имеем

$$\delta q=\frac{1}{\beta\nu_1''(\beta)}\Big\{-\beta\delta\mu_1'-\delta\mu_1-\beta\delta\mu_1''\,\delta q+$$
$$+\frac{1}{2}\left[\nu_1''(\beta)+\beta\nu_1'''(\beta)\right]\delta q^2\Big\}+\ldots,$$

откуда

$$\delta q=-\frac{\beta\delta\mu_1'+\delta\mu_1}{\beta\nu_1''(\beta)}+\frac{\delta\mu_1''(\beta\delta\mu_1'+\delta\mu_1)}{\beta\left[\nu_1''(\beta)\right]^2}+$$
$$+\frac{\nu_1''(\beta)+\beta\nu_1'''(\beta)}{2\beta^3\left[\nu_1''(\beta)\right]^3}(\beta\delta\mu_1'+\delta\mu_1)^2+\ldots$$

Подставляя этот результат в разложение

$$\mu_1'(q_\xi)=\mu_1'(-\beta)+\mu_1''(-\beta)\,\delta q+\frac{1}{2}\mu_1'''(-\beta)\,\delta q^2+\ldots=$$
$$=-\nu_1'(\beta)+\delta\mu_1'+\nu_1''(\beta)\,\delta q+\delta\mu_1''\,\delta q-\frac{1}{2}\nu_1'''(\beta)\,\delta q^2,$$

получаем после сокращений

$$\mathbf{M}\mu_1'(q_\xi)=-\nu_1'(\beta)+\frac{\mathbf{M}(\beta\delta\mu_1'+\delta\mu_1)^2}{2\beta^3\nu_1''(\beta)}+\mathbf{M}\delta\mu_1^3\ldots$$

Здесь учтено, что $\mathbf{M}\delta\boldsymbol{\mu}_1=\mathbf{M}\delta\boldsymbol{\mu}_1'=0$.

Прочие члены, входящие в подынтегральное выражение (11.3.32), требуют меньшей точности вычислений. Для них достаточно просто подставить, вместо случайных значений, предельные значения.

В итоге будем иметь

$$0 \leqslant \frac{1}{n} \widetilde{R}_n (nI_1) - R_1 (I_1) \leqslant \frac{1}{2n\beta} \ln [2\pi n \nu_1'' (\beta) \beta^2] -$$
$$- \frac{C}{n\beta} + \mathbf{M} (\beta \delta \mu_1' + \delta \mu_1)^2 / 2\beta^3 \nu_1'' (\beta) + o (n^{-1}). \qquad (11.3.40)$$

Теми же самыми методами, производя расчеты с большей степенью точности, можно найти и более высокие члены этого асимптотического разложения. При этом можно получить подтверждение тех пунктов вышеизложенного вывода, которые представляются недостаточно обоснованными.

В приведенном рассмотрении предполагалось, что отрезок (s_1, s_2) определения и дифференцируемости потенциала $\mu_\xi (t)$, упоминаемый в теореме 4.8, достаточно велик, между тем существенной в действительности является лишь некоторая окрестность точки $s = -\beta$. Прочие участки прямой s оказывают влияние лишь на экспоненциальные члены типа (11.3.26), (11.3.28), (11.3.30) и не сказываются на асимптотическом разложении (11.3.40). Правда, аномалии в поведении функции $\mu_1 (s)$ на этих прочих участках осложняют доказательство.

Для справедливости формулы (11.3.40) не является необходимым выполнение условия (11.2.14) ограниченности функции штрафов. Однако ее вывод несколько упростится, если принять это условие. Тогда не потребуется детальная оценка интегралов S_2, S_3 и интеграла по области $\bar{\Xi}$. Вместо этого можно будет ограничиться доказательством достаточно быстрого (экспоненциального) исчезновения вероятности соответствующих областей интегрирования. Значение постоянной K при этом будет несущественным, так как не войдет в окончательный результат.

Как видно из приведенного вывода, члены асимптотического разложения (11.3.40) являются точными для выбранного случайного кодирования. Могут быть найдены более высокие члены, но уже выписанные члены не могут быть улучшены, если не отказаться от принятого способа кодирования. Представляет интерес вопрос, насколько приведенная оценка (11.3.40) близка к действительному значению разности $(1/n) \widetilde{R} (nI_1) - R_1 (I_1)$ и насколько она может быть улучшена, если использовать более совершенные приемы кодирования.

11.4. Другие способы записи основного результата. Обобщения и частные случаи

1. В предыдущем параграфе вместо функции (11.3.8) была введена функция $\mu_1 (t) = (1/n) \mu_\xi (t)$. Это было вызвано по существу лишь соображениями удобства и наглядности, чтобы подчеркнуть относительную величину членов. Изложение почти не изменится, если удельные величины μ_1, ν_1, R_1, I_1 и др. рассматривать лишь в произведении с n, иначе говоря, если вообще не вводить удельных величин. Вместо основной результирующей формулы (11.3.40), умножив ее на n, будем иметь формулу

$$0 \leqslant \tilde{R}(I) - R(I) = V(I) - \tilde{V}(I) \leqslant \frac{1}{2\beta} \ln\left[\frac{2\pi}{\gamma^2} \nu''(\beta)\beta^2\right] + \\ + \frac{\mathbf{M}(\beta\delta\mu' + \delta\mu)^2}{2\beta^3 \nu''(\beta)} + o(1), \tag{11.4.1}$$

где в соответствии с (11.3.34), (11.3.8)

$$\nu(-t) = \mathbf{M}\mu_\xi(t) = \int P(d\xi) \ln \int e^{tc(\xi,\zeta)} P_I(d\zeta), \tag{11.4.2}$$

$$\delta\mu = \delta\mu(-\beta); \qquad \delta\mu' = \delta\mu'(-\beta); \qquad \delta\mu(t) = \mu_\xi(t) - \nu(-t) \tag{11.4.3}$$

(индекс n опущен, член C/β введен под знак логарифма; $\gamma = e^C = 1{,}781$). При указанной модификации изложения, как и в п. 3 § 11.2, становится не нужным, чтобы бейесовская система была обязательно n-й степенью какой-то элементарной бейесовской системы.

Дважды дифференцируя (11.4.2) в точке $t = -\beta$ и учитывая (11.1.16), получаем

$$\nu''(\beta) = \int P(d\xi) \left\{\int c^2(\xi, \zeta) e^{-\gamma(\xi) - \beta c(\xi,\zeta)} P_I(d\zeta) - \right. \\ \left. - \left[\int c(\xi, \zeta) e^{-\gamma(\xi) - \beta c(\xi,\zeta)} P_I(d\zeta)\right]^2\right\}.$$

В силу (9.4.21) входящие сюда интегралы по ζ являются интегралами условного математического ожидания с условными вероятностями $P(d\zeta \mid \xi)$. Поэтому

$$\nu''(\beta) = \mathbf{M}\{\mathbf{M}[c^2(\xi,\zeta) \mid \xi] - (\mathbf{M}[c(\xi,\zeta) \mid \xi])^2\} \equiv \\ \equiv \mathbf{MD}[c(\xi, \zeta) \mid \xi], \tag{11.4.4}$$

где $\mathbf{D}[\ldots \mid \xi]$ означает условную дисперсию. Средняя условная дисперсия (11.4.4), как легко убедиться, не меняется при преобразовании (11.2.36). Совершая преобразование (11.2.37), следовательно, будем иметь

$$\beta^2\nu''(\beta) = \mathbf{MD}[I(\xi, \zeta) \mid \xi]. \tag{11.4.5}$$

Рассмотрим теперь, что представляет собой входящий в (11.4.1) средний квадрат $\mathbf{M}(\beta\delta\mu' + \delta\mu)^2$, который в силу (11.4.3) совпадает с дисперсией:

$$\mathbf{M}(\beta\delta\mu' + \delta\mu)^2 = \mathbf{D}[\beta\mu'_\xi(-\beta) + \mu_\xi(-\beta)] \tag{11.4.6}$$

случайной величины $\beta\mu'_\xi(-\beta) + \mu_\xi(-\beta)$.

Полагая $t = -\beta$ в (11.3.8) и сопоставляя это выражение с (9.4.23), видим, что

$$\mu_\xi(-\beta) = \gamma(\xi). \tag{11.4.7}$$

Далее, дифференцируя (11.3.8) в точке $t = -\beta$, находим

$$\mu'_\xi(-\beta) = \int c(\xi, \zeta)\, e^{-\gamma(\xi) - \beta c(\xi, \zeta)} P_I(d\zeta).$$

Вследствие (9.4.21) этот интеграл есть не что иное, как интеграл усреднения с условным распределением $P(d\zeta \mid \xi)$, т. е.

$$\mu'_\xi(-\beta) = \mathbf{M}[c(\xi, \zeta) \mid \xi].$$

Отсюда и из (11.4.7) имеем

$$\beta\mu'_\xi(-\beta) + \mu_\xi(-\beta) = \mathbf{M}[\beta c(\xi, \zeta) + \gamma(\xi) \mid \xi].$$

Но $-\beta c(\xi, \zeta) - \gamma(\xi)$ в силу (9.4.21) совпадает со случайной информацией $I(\xi, \zeta)$, так что

$$\beta\mu'_\xi(-\beta) + \mu_\xi(-\beta) = -\mathbf{M}[I(\xi, \zeta) \mid \xi] \equiv -I_{\zeta|}(|\xi). \tag{11.4.8}$$

Итак, мы видим, что выражение (11.4.6) есть дисперсия частично усредненной случайной информации:

$$\mathbf{M}(\beta\delta\mu' + \delta\mu)^2 = \mathbf{D}\mathbf{D}[\beta\mu'_\xi(-\beta) + \mu_\xi(-\beta)] = [I_{\zeta|}(\,|\,\xi)]. \tag{11.4.9}$$

Вследствие (11.4.5), (11.4.9) основной формуле (11.4.1) можно придать вид

$$0 \leqslant 2\beta[V(I) - \widetilde{V}(I)] \leqslant \ln\left\{\frac{2\pi}{\gamma^2}\mathbf{MD}[I(\xi, \zeta) \mid \xi]\right\} + \\ + \mathbf{D}[I_\zeta(\,|\,\xi)]/\mathbf{MD}[I(\xi, \zeta) \mid \xi] + o(1). \tag{11.4.10}$$

2. В некоторых частных случаях экстремальное распределение $P_I(d\zeta)$ не зависит от β, т. е. от I. Тогда, как видно из (9.4.23), (11.3.8), функция $\boldsymbol{\mu}_\xi(t) = \boldsymbol{\mu}(\xi, t, \beta)$, зависящая, вообще говоря, от ξ, t, β, оказывается не зависящей от β, а зависимость ее от ξ и t совпадает с зависимостью функции $\gamma(\xi) = \gamma(\xi, \beta)$ от ξ и от $-\beta$:

$$\boldsymbol{\mu}(\xi, t, \beta) = \gamma(\xi, -t) \text{ при всех } t. \tag{11.4.11}$$

Усреднением этого равенства по ξ в силу (9.4.10), (11.4.2), получаем

$$\nu(-t) = \Gamma(-t) \text{ при всех } t. \tag{11.4.12}$$

Поэтому в формуле (11.4.1) можно заменить $\nu''(\beta)$ на $\Gamma''(\beta)$. Далее полезно вспомнить, что согласно (9.4.31) значение β связано с производными функций $R(I)$, $V(I)$:

$$\frac{1}{\beta} = -\frac{dR}{dI} = V'(I), \tag{11.4.13}$$

так что

$$\beta = 1/V'(I). \tag{11.4.14}$$

Вторую производную $\Gamma''(\beta)$ также нетрудно выразить через функцию $I(R)$ (или $V(I)$), так как эти функции связаны между собой преобразованием Лежандра [см. (9.4.30) и (9.4.29)]. Дифференцируя (9.4.30), имеем

$$\beta\Gamma''(\beta) = \frac{dI}{d\beta}.$$

Отсюда

$$\frac{1}{\beta^2\Gamma''(\beta)} = \frac{1}{\beta}\frac{d\beta}{dI},$$

или, если продифференцировать (11.4.14) и учесть (11.4.13),

$$1/(\beta^2\Gamma''(\beta)) = -V''(I)/V'(I). \tag{11.4.15}$$

В силу (11.4.14), (11.4.15), (11.4.9) основную формулу (11.4.1) в данном случае можно записать

$$0 \leqslant V(I) - \widetilde{V}(I) \leqslant \frac{1}{2}V'(I)\ln\left[-\frac{2\pi}{\gamma^2}\frac{V'(I)}{V''(I)}\right] - \\ -\frac{1}{2}V''(I)\mathbf{D}[I_{\zeta|}(|\xi)] + o(1). \tag{11.4.16}$$

Иногда, кроме того, функция (11.4.11) оказывается не зависящей от ξ. С таким положением мы встречались в § 10.2, где для частного случая была получена формула (10.2.25), Согласно последней $\mu_\xi(t) = \Gamma(-t)$, так что $\mu_\xi(t) = \nu(-t) = \Gamma(-t)$; $\delta\mu(t) = 0$, и усреднение по ξ становится излишним. В этом случае дисперсия (11.4.6), (11.4.9) обращается в нуль и формула (11.4.16) несколько упрощается, принимая вид

$$V(I) \geqslant \tilde{V}(I) \geqslant V(I) - \frac{1}{2}V'(I)\ln\left[-\frac{2\pi}{\gamma^2}\frac{V'(I)}{V''(I)}\right] + o(1). \tag{11.4.17}$$

При этом становится излишним рассмотрение, проведенное в п. 4 предыдущего параграфа.

3. В некоторых важных случаях последовательность значений I и последовательность бейесовских систем $[P(d\xi), c(\xi, \zeta)]$ (зависящих от n или другого параметра) такова, что для экстремального распределения

А. $I = I_{\xi\zeta} \to \infty$ (11.4.18)

Б. Существуют конечные ненулевые пределы

$$\lim\frac{\mathbf{MD}[I(\xi,\zeta)\mid\xi]}{I}, \quad \lim\frac{\mathbf{D}[I_{\zeta|}(|\xi)]}{I}; \quad \lim\frac{V(yI)}{I}, \quad \lim\frac{dV(I)}{dI} \tag{11.4.19}$$

(y произвольно и не зависит от I).

Нетрудно убедиться, что сумма

$$\mathbf{MD}[I(\xi,\zeta)\mid\xi] + \mathbf{D}[I_{\zeta|}(|\xi)] = \mathbf{M}\{\mathbf{M}[I^2(\xi,\zeta)\mid\xi] - \\ -(\mathbf{M}[I(\xi,\zeta)\mid\xi])^2\} + \mathbf{M}(\mathbf{M}[I(\xi,\zeta)\mid\xi])^2 - (\mathbf{MM}[I(\xi,\zeta)\mid\xi])^2$$

совпадает с полной дисперсией $\mathbf{D}[I(\xi, \zeta)]$. Следовательно, из существования первых двух пределов (11.4.19) вытекает существование конечного предела

$$\lim \frac{1}{I} \mathbf{D}[I(\xi, \zeta)].$$

Поэтому, с одной стороны, из А, Б вытекают условия информационной устойчивости для ξ и ζ, указанные в § 7.3, как это стандартно доказывается путем применения неравенства Чебышева, из Б следует также (11.2.34). Таким образом, если еще выполнено условие ограниченности (11.2.39) и непрерывности функции (11.2.34) по y, то в данном случае будет иметь место сходимость (11.2.35) по теореме 11.2. С другой стороны, из (11.4.18) и конечности пределов (11.4.19) при учете (11.4.14) вытекает, что

$$\frac{1}{2\beta V(I)} \frac{\mathbf{D}[I_{\zeta|}(|\xi)]}{\mathbf{MD}[I(\xi, \zeta)|\xi]} =$$

$$= \frac{1}{2I} \frac{dV}{dI} \frac{I}{V(I)} \frac{\mathbf{D}[I_{\zeta|}(\xi)]}{I} \frac{I}{\mathbf{MD}[I(\xi, \zeta)|\xi]} \to 0. \tag{11.4.20}$$

Далее

$$\frac{1}{I} \ln\left\{\frac{2\pi}{\gamma^2}\mathbf{MD}[I(\xi, \zeta)|\xi]\right\} = \frac{\ln I}{I} + \frac{1}{I} \ln\left\{\frac{2\pi}{\gamma^2}\left\{\frac{\mathbf{MD}[I(\xi, \zeta)|\xi]}{I}\right\}\right\} \to 0$$

при $I \to \infty$. Поэтому и для логарифмического члена в (11.4.10 имеем

$$\frac{1}{2\beta V(I)} \ln\left\{\frac{2\pi}{\gamma^2} \mathbf{MD}[I(\xi, \zeta)|\xi]\right\} \to 0. \tag{11.4.21}$$

Вследствие (11.4.20), (11.4.21) из (11.4.10) в данном случае следует сходимость (11.2.35), прочие члены, замененные на $o(1)$ в правой части (11.4.10), убывают еще быстрее. При этом условие (11.2.39), как мы видим, является необязательным.

4. Остановимся особо на одном частном случае — случае гауссовых бейесовских систем, которым были посвящены § 10.3, 10.4. При этом, вообще говоря, нельзя пользоваться упрощениями п. 2 и приходится обращаться к формуле (11.4.10). Значение $\beta^2\nu''(\beta)$, входящее в (11.4.5), уже было найдено ранее в гл. 10; оно определяется равенством (10.3.42), которое легко преобразовать к виду

$$\beta^2 \nu''(\beta) = \frac{1}{2} \operatorname{Sp}\left\{1_u - \left(\frac{1}{\beta} \mathrm{h} \widetilde{\mathrm{k}}_x^{-1}\right)^2\right\}. \tag{11.4.22}$$

Далее, чтобы вычислить дисперсию (11.4.9), следует воспользоваться формулой (10.3.43). Дисперсия выражений, квадратичных по гауссовым переменным, уже вычислялась ранее в § 5.4. Применяя тот же способ вычисления [основанный на формулах (5.4.15), (5.4.16)] к соотношению (10.3.43), вместо (5.4.14), как нетрудно видеть, получим

$$\mathbf{D}\,[\beta\mu'_x(-\beta)+\mu_x(-\beta)] =$$
$$=\frac{\beta^2}{2}\,\mathrm{Sp}\,\mathrm{g}\mathrm{h}^{-1}\,\frac{\beta\tilde{\mathrm{k}}_x\mathrm{h}^{-1}-1_u}{(\beta\tilde{\mathrm{k}}_x\mathrm{h}^{-1})^2}\,\mathrm{g}^T\mathrm{k}_x\,\mathrm{g}\mathrm{h}^{-1}\,\frac{\beta\tilde{\mathrm{k}}_x\mathrm{h}^{-1}-1_u}{(\beta\tilde{\mathrm{k}}_x\mathrm{h}^{-1})^2}\,\mathrm{g}^T\mathrm{k}_x=$$
$$=\frac{1}{2}\,\mathrm{Sp}\left(1_u-\frac{1}{\beta}\,\mathrm{h}\tilde{\mathrm{k}}_x^{-1}\right) \qquad (11.4.23)$$

(так как $\mathrm{g}^T\mathrm{k}_x\mathrm{g}=\tilde{\mathrm{k}}_x$).

После подстановки (11.4.22), (11.4.23) в (11.4.1), (11.4.6) будем иметь

$$0\leqslant 2\beta\,[V(I)-\tilde{V}(I)]\leqslant \ln\left\{\frac{\pi}{\gamma^2}\,\mathrm{Sp}\left[1_u-\left(\frac{1}{\beta}\,\mathrm{h}\tilde{\mathrm{k}}_x^{-1}\right)^2\right]\right\}+$$
$$+\,\mathrm{Sp}\left(1_u-\frac{1}{\beta}\,\mathrm{h}\tilde{\mathrm{k}}_x^{-1}\right)^2\Big/\,\mathrm{Sp}\left[1_u-\left(\frac{1}{\beta}\,\mathrm{h}\mathrm{k}_x^{-1}\right)^2\right]+o(1) \qquad (11.4.24)$$

Если, помимо приведенных выражений, принять во внимание формулу (10.3.30), то условие А предыдущего пункта примет вид

$$\mathrm{Sp}\,\ln(\beta\tilde{\mathrm{k}}_x\mathrm{h}^{-1})\to\infty \qquad (11.4.25)$$

Требование существования предела $\lim\,(dV/dI)$, входящего в условие Б (11.4.19), эквивалентно в силу (11.4.14) требованию существования предела

$$\lim\beta=\beta_0. \qquad (11.4.26)$$

Два первых предела в (11.4.19) записываются в виде

$$\lim\frac{\mathrm{Sp}\,[1_u-(\beta^{-1}\,\mathrm{h}\tilde{\mathrm{k}}_x^{-1})^2]}{\mathrm{Sp}\,\ln(\beta\tilde{\mathrm{k}}_x\mathrm{h}^{-1})} \qquad \lim\frac{\mathrm{Sp}\,(1_u-\beta^{-1}\,\mathrm{h}\tilde{\mathrm{k}}_x^{-1})^2}{\mathrm{Sp}\,\ln(\beta\tilde{\mathrm{k}}_x\mathrm{h}^{-1})}. \qquad (11.4.27)$$

Наконец, условие существования предела $V(yI)/I$ вследствие (10.3.32) принимает вид

$$\lim\mathrm{Sp}\,(\tilde{\mathrm{k}}_x\mathrm{h}^{-1}-\beta_y^{-1}\,1_u)/\mathrm{Sp}\,\ln(\beta\tilde{\mathrm{k}}_x\mathrm{h}^{-1}), \qquad (11.4.28)$$

где β_y определяется из условия

$$\mathrm{Sp}\,\ln(\beta_y\tilde{\mathrm{k}}_x\mathrm{h}^{-1})=2yI=y\,\mathrm{Sp}\,\ln(\beta\tilde{\mathrm{k}}_x\mathrm{h}^{-1}). \qquad (11.4.29)$$

Рассмотрим для примера гауссов случайный процесс в непрерывном времени, периодический на отрезке $[0,\,T_0]$. Он может быть получен из непериодического стационарного процесса описанной в начале § 5.7 периодизацией (5.7.1). Бейесовская система с таким периодическим процессом рассматривалась в п. 2 § 10.4. Для нее следы, входящие в (10.3.30), сводились к суммам (10.4.10). Подобным же образом выражаются и следы в (11.4.24). Если к тому же суммы заменить интегралами, что с некоторым приближением справедливо при больших T_0, и устремить $T_0\to\infty$ [см. вывод формул (10.4.18)], в отходящие в (11.4.27)—(11.4.29) следы Sp будут приблизительно пропорциональны T_0:

$$\operatorname{Sp} f(\tilde{k}_x h^{-1}) \approx \frac{T_0}{2\pi} \int\limits_{L(\omega)} f(\Phi(\omega))\, d\omega \tag{11.4.30}$$

при выполнении условия сходимости интегралов. Здесь

$$\Phi(\omega) = \frac{\bar{k}(\omega)}{\bar{h}(\omega)} |\bar{g}(\omega)|^2;$$

$\bar{k}(\omega)$, $\bar{h}(\omega)$ и прочие имеют тот же смысл, что и в § 10.4 [см. (10.4.19), (10.4.21)]. Область интегрирования $L(\omega)$ определена неравенством (10.4.20). В этом случае пределы (11.4.27), (11.4.28) действительно существуют и равны отношению соответствующих интегралов. Например, первый предел (11.4.27) равен

$$\int [1 - \beta_0^{-2} \Phi^{-2}(\omega)]\, d\omega \Big/ \int \ln(\beta_0 \Phi(\omega))\, d\omega \quad (L = L(\beta_0)).$$

В силу (11.4.30) разность $1 - \tilde{V}/V$, как легко видеть из (11.4.24), убывает с ростом T_0 по закону

$$\mathrm{const}\, T_0^{-1} \ln T_0 + \mathrm{const}\, T_0^{-1} + o(T_0^{-1}).$$

11.5. Обобщенная теорема Шеннона

Здесь будет рассмотрено обобщение результатов § 7.3 (теоремы 7.1 и 7.2) и § 8.1 (теорема 8.1) на случай произвольного критерия качества. Напомним, что в гл. 7 был взят единственный критерий качества, именно, качество информационной системы характеризовалось средней вероятностью принятия ложного сообщения. Работами Колмогорова [1], Шеннона [4], Добрушина [1] начато распространение указанных результатов на случай более общего критерия качества, характеризуемого произвольной функцией штрафов $c(\xi, \zeta)$. Соответствующую теорему, сформулированную для произвольной функции $c(\xi, \zeta)$ и превращающуюся для частного вида функции

$$c(\xi, \zeta) = -\delta_{\xi\zeta} = \begin{cases} -1 & \text{при } \xi = \zeta, \\ 0 & \text{при } \xi \neq \zeta \end{cases} \tag{11.5.1}$$

(ξ, ζ — дискретные величины) в указанные обычные результаты, естественно называть обобщенной теоремой Шеннона.

Описанное направление исследований, поскольку в нем явно вводится функция штрафов $c(\xi, \zeta)$ и считается заданным априорно распределение $P(d\xi)$, тесно связано с третьей вариационной задачей и с материалом, изложенным в гл. 9—11. В этом направлении могут быть получены результаты различной силы и детальности. Мы приведем здесь лишь одну естественную теорему, которая почти немедленно вытекает из обычной теоремы Шеннона и результатов § 11.2 и не потребует, следовательно, нового доказательства.

Пусть заданы случайные величины ξ, ζ, описываемые совместным распределением $P(d\xi\, d\zeta)$. Имеется также канал $[P(d\tilde{\eta} \mid \eta)]$, выходная переменная (или переменные) $\tilde{\eta}$ которого связана с входной переменной η в соответствии с условными вероятностями $P(d\tilde{\eta} \mid \eta)$. Требуется осуществить кодирование $\xi \to \eta$, подобрав вероятности $P(d\eta \mid \xi)$, и декодирование $\tilde{\eta} \to \tilde{\zeta}$, подобрав вероятности $P(d\tilde{\zeta} \mid \tilde{\eta})$, таким способом, чтобы распределение $P(d\xi, d\tilde{\zeta})$, индуцируемое распределениями $P(d\xi)$, $P(d\eta \mid \xi)$, $P(d\tilde{\eta} \mid \eta)$, $P(d\tilde{\zeta} \mid \tilde{\eta})$, совпадало с исходным распределением $P(d\xi\, d\zeta)$. Легко видеть, что величины $\xi, \eta, \tilde{\eta}, \tilde{\zeta}$, связанные схемой

$$\xi \to \eta \to \tilde{\eta} \to \tilde{\zeta}, \tag{11.5.2}$$

представляют собой цепь Маркова с указанными переходными вероятностями. Применяя формулу (6.3.7), имеем

$$I_{(\xi\eta)(\tilde{\eta}\tilde{\zeta})} = I_{(\zeta\eta)(\tilde{\zeta}\tilde{\eta})} = I_{\xi\tilde{\zeta}} + I_{\eta\tilde{\zeta} \mid \xi} + I_{\xi\tilde{\eta} \mid \tilde{\zeta}} + I_{\eta\tilde{\eta} \mid \xi\tilde{\zeta}}. \tag{11.5.3}$$

В то же время по той же формуле можно получить

$$I_{(\eta\xi)(\tilde{\eta}\tilde{\zeta})} = I_{\eta\tilde{\eta}} + I_{\xi\tilde{\eta}/\eta} + I_{\eta\tilde{\zeta}/\tilde{\eta}} + I_{\xi\tilde{\zeta}/\eta\tilde{\eta}}. \tag{11.5.4}$$

В силу условия Маркова будущее не зависит от прошлого при фиксированном настоящем, следовательно, информация связи между прошлым (ξ) и будущим ($\tilde{\eta}$) при фиксированном настоящем (η) равна нулю: $I_{\xi\tilde{\eta} \mid \eta} = 0$. По той же причине

$$I_{\eta\tilde{\zeta} \mid \tilde{\eta}} = 0; \qquad I_{\xi\tilde{\zeta} \mid \eta\tilde{\eta}} = 0. \tag{11.5.5}$$

Приравнивая между собой (11.5.4), (11.5.3) и учитывая (11.5.5), имеем

$$I_{\xi\tilde{\zeta}} + I_{\eta\tilde{\zeta} \mid \xi} + I_{\xi\tilde{\eta} \mid \tilde{\zeta}} + I_{\eta\tilde{\eta} \mid \xi\tilde{\zeta}} = I_{\eta\tilde{\eta}}. \tag{11.5.6}$$

Но информации $I_{\eta\tilde{\zeta} \mid \xi}$, $I_{\xi\tilde{\eta} \mid \tilde{\zeta}}$, $I_{\eta\tilde{\eta} \mid \xi\tilde{\zeta}}$ (как и всякое шенноновское количество информации) неотрицательны. Поэтому из (11.5.6) получаем неравенство

$$I_{\xi\tilde{\zeta}} \leqslant I_{\eta\tilde{\eta}} \text{ или } I_{\xi\tilde{\zeta}} \leqslant C, \tag{11.5.7}$$

если учесть определение пропускной способности $C = \sup I_{\eta\tilde{\eta}}$ (см. (8.1.3)). Отсюда видно, что распределение $P(d\xi \cdot d\tilde{\zeta})$ может копировать исходное распределение $P(d\xi\, d\zeta)$ только при том необходимом условии, что

$$I_{\xi\zeta} \leqslant C. \tag{11.5.8}$$

В противном случае не существует таких способов кодирования и декодирования [т. е. таких $P(d\eta\,|\,\xi)$, $P(d\zeta\,|\,\eta)$], чтобы $P(d\xi\,d\zeta)$ совпадало с $P(d\xi\,d\zeta)$.

Этот вывод, как мы видим, следует из самых общих свойств рассматриваемых понятий. Менее тривиальным является тот факт, что условие (11.5.8), а точнее условие

$$\lim\sup I_{\xi\zeta}/C < 1, \tag{11.5.9}$$

является в асимптотическом смысле достаточным, если не для совпадения распределения $P(d\xi\,d\zeta)$ с $P(d\xi\,d\zeta)$, то во всяком случае для достаточно хорошего в определенном смысле качества этого распределения. Чтобы сформулировать указанный факт, нужно ввести критерий качества — функцию штрафов $c(\xi,\,\zeta)$. Условие совпадения $P(d\xi\,d\tilde{\zeta})$ с $P(d\xi d\zeta)$ заменяем на более слабое условие

$$|\,\mathbf{M}c(\xi,\,\zeta)\,|^{-1}\,\vartheta_+(\mathbf{M}c(\xi,\,\tilde{\zeta})-\mathbf{M}c(\xi,\,\zeta))\to 0 \tag{11.5.10}$$

$$(2\vartheta_+(z)=z+|\,z\,|),$$

которое говорит о том, что качество распределения $P(d\xi d\tilde{\zeta})$ асимптотически не хуже, чем $P(d\xi d\zeta)$. Основное утверждение формулируется для последовательности схем (11.5.2), зависящих, скажем, от n.

Теорема 11.3. *Пусть*: 1) *последовательность пар* ξ, ζ *случайных величин является информационно устойчивой* (см. § 7.3);

2) *имеет место сходимость*

$$\frac{c(\xi,\,\zeta)-\mathbf{M}c(\xi,\,\zeta)]}{\mathbf{M}c(\xi,\,\zeta)}\to 0 \tag{11.5.11}$$

по вероятности $P(d\xi d\zeta)$ [ср. (11.2.31)];

3) *последовательность функций штрафа удовлетворяет условию ограниченности*

$$|\,c(\xi,\,\zeta)\,|\leqslant K\,|\,\mathbf{M}c(\xi,\,\zeta)\,| \tag{11.5.12}$$

[ср. (11.2.32); *K не зависит от n*];

4) *последовательность каналов* $[P(d\tilde{\eta}\,|\,\eta)]$ *является информационно устойчивой* (т. е. $(\eta,\,\tilde{\eta})$ при экстремальном распределении являются информационно устойчивыми);

5) *выполняется условие* (11.5.9).

Тогда существуют способы кодирования и декодирования, обеспечивающие асимптотическую равнокачественность в смысле сходимости (11.5.10).

Доказательство использует полученные ранее результаты и указывает способы кодирования и декодирования, т. е. конструктивно определяет $\tilde{\zeta}$.

Условия 1—3 теоремы позволяют доказать неравенство

$$\limsup_{n\to\infty} |R|^{-1}[\widetilde{R}(I_{\xi\zeta}+2\varepsilon_2 I_{\xi\zeta})-R]\leqslant \varepsilon_1+4K\delta$$
$$(R=\mathbf{M}c(\xi,\zeta);\quad \varepsilon_1=\varepsilon_2 I/R) \tag{11.5.13}$$

совершенно тем же способом, как и в § 11.2 были доказаны соотношения (11.2.33а) на основе условий информационной устойчивости и условий (11.2.31), (11.2.32) (см. также вывод (11.2.30)). В (11.5.13) ε, δ — сколь угодно малые, не зависящие от n положительные величины, а $\widetilde{R}(I_{\xi\zeta}+2\varepsilon_2 I_{\xi\zeta})$ — средние штрафы:

$$\widetilde{R}(I_{\xi\zeta}+2\varepsilon_2 I_{\xi\zeta})=\mathbf{M}\inf_{\zeta}\mathbf{M}[c(\xi,\zeta)\mid E_k], \tag{11.5.14}$$

соответствующие некоторому разбиению $\sum E_k$ пространства значений ξ на $M=[e^{I_{\xi\zeta}+2\varepsilon_2 I_{\xi\zeta}}]$ областей. Эти области могут быть построены при помощи случайных кодовых точек ζ_k, выбрасываемых с вероятностями $P(d\zeta)$ (см. § 11.1, 11.2).

Из (11.5.14) следует, что существуют такие точки ζ_k, что

$$\widetilde{R}(I_{\xi\zeta}+2\varepsilon_2 I_{\xi\zeta})\geqslant\sum_k P(E_k)\mathbf{M}c(\xi,\zeta_k)-\varepsilon_3$$

($\varepsilon_3>0$ сколь угодно мало) и в силу (11.5.13)

$$\limsup_{n\to\infty}|R|^{-1}\left[\sum_k P(E_k)\mathbf{M}[c(\xi,\zeta_k)\mid E_k]-R\right]\leqslant\varepsilon_1+4K\delta+|R|^{-1}\varepsilon_3. \tag{11.5.15}$$

Будем передавать по каналу $P(d\widetilde{\eta}\mid\eta)$ сообщение о том, какой кодовой области E_k принадлежит значение ξ.

Вследствие (11.5.9) можно выбрать такое ε_2, чтобы выполнялось неравенство $[I_{\xi\zeta}+\varepsilon_2(2I_{\xi\zeta}+C)]/C<1$, начиная с некоторого n. Поскольку $M=[\exp(I_{\xi\zeta}+2\varepsilon_2 I_{\xi\zeta})]$, это означает выполнение неравенства $\ln M/C<1-\varepsilon_2$, т. е. условия (8.1.5). Последнее вместе с требованием 4) теоремы 11.3 обеспечивает применение теоремы 8.1 (обобщенной по типу теоремы 7.2), согласно которой вероятность ошибки при приеме сообщения по каналу с увеличением n может быть сделана сколь угодно малой:

$$P_{\text{ош}}<\varepsilon_4, \tag{11.5.16}$$

начиная с некоторого n (ε_4 любое). Пусть l означает номер принятого сообщения, имеющего вид $\xi\in E_l$. Декодирующее устройство, получив это сообщение, дает на выходе сигнал $\widetilde{\zeta}=\zeta_l$. Таким образом, выходная переменная $\widetilde{\zeta}$ имеет плотность распределения

$$P(\widetilde{\zeta})=\sum_l P(l)\,\delta(\widetilde{\zeta}-\zeta_l).$$

Совместное распределение $P(d\xi d\widetilde{\zeta})$ имеет вид

$$P(d\xi d\widetilde{\zeta})=\sum_{k,l}P(d\xi)\,\vartheta_{E_k}(\xi)\,P(l\mid k)\,\delta(\widetilde{\zeta}-\zeta_l)\,d\widetilde{\zeta}, \tag{11.5.17}$$

$$(\vartheta_{E_k}(\xi)=1 \text{ при } \xi\in E_k;\quad \vartheta_{E_k}(\xi)=0 \text{ вне } E_k),$$

где $P(l\,|\,k)$ — вероятность принять сообщение $\xi\in E_l$, если было передано сообщение $\xi\in E_k$.

Распределению (11.5.17) соответствуют средние штрафы

$$\begin{aligned}\mathbf{M}c(\xi,\tilde\zeta)&=\sum_{k,l}P(E_k)P(l\,|\,k)\mathbf{M}[c(\xi,\zeta_l)\,|\,E_k]=\\&=\sum_k P(E_k)\mathbf{M}[c(\xi,\zeta_k)\,|\,E_k]+\\&+\sum_{k\neq l}P(E_k)P(l\,|\,k)\{\mathbf{M}[c(\xi,\zeta_l)\,|\,E_k]-\mathbf{M}[c(\xi,\zeta_k)\,|\,E_k]\}.\end{aligned}$$

Мажорируем последнюю сумму, используя (11.5.12). Это дает

$$\begin{aligned}&|R|^{-1}|\mathbf{M}c(\xi,\tilde\zeta)-\sum_k P(E_k)\mathbf{M}[c(\xi,\zeta_k)\,|\,E_k]|\leqslant\\&\leqslant 2K\sum_k P(E_k)\,P_{\text{ош}}(|\,k)\qquad \Big(P_{\text{ош}}(|\,k)=\sum_{l,\,l\neq k}P(l\,|\,k)\Big).\end{aligned}\tag{11.5.18}$$

Усредняя по ансамблю случайных кодов и учитывая (11.5.16), имеем

$$\mathbf{M}\sum_k P(E_k)P_{\text{ош}}(|\,k)=\mathbf{M}P_{\text{ош}}(|\,k)=P_{\text{ош}}<\varepsilon_4.$$

Отсюда можно сделать вывод, что существует некоторый неслучайный код, который не хуже первого в смысле неравенства

$$\sum_k P(E_k)P_{\text{ош}}(|\,k)<\varepsilon_4,\tag{11.5.19}$$

и поэтому в силу (11.5.18)

$$\mathbf{M}c(\xi,\tilde\zeta)-\sum_k P(E_k)\mathbf{M}[c(\xi,\zeta_k)\,|\,E_k]\leqslant 2K\varepsilon_4|R|.$$

Комбинируя это неравенство с (11.5.15), получаем

$$\limsup|R|^{-1}[\mathbf{M}c(\xi,\tilde\zeta)-\mathbf{M}c(\xi,\zeta)]\leqslant\varepsilon_2+4K\delta+|R|^{-1}\varepsilon_3+2K\varepsilon_4.$$

Поскольку ε_2, δ, ε_3, ε_4 могут быть взяты сколь угодно малыми, отсюда следует (11.5.10). Доказательство закончено.

Выше предполагалось, что задано исходное распределение $P(d\xi d\zeta)$. Проведенное рассмотрение без труда распространяется на те случаи, когда задано некоторое допустимое множество подобных распределений или бейесовская система $[P(d\xi),\ c(\xi,\zeta)]$ с фиксированным уровнем потерь

$$\mathbf{M}c(\xi,\zeta)\leqslant a.$$

В этом последнем случае доказывается сходимость

$$\frac{1}{V}\vartheta_+(\mathbf{M}c(\xi,\tilde\zeta)-a)\to 0;$$

условие (11.5.9) заменяется на

$$\limsup I(a)/C<1;$$

условие (11.5.12) — на (11.2.39), а условия 1), 2) теоремы 11.3 нужно заменить требованиями информационной устойчивости бейесовской системы (§ 11.2, п. 4), чтобы можно было применить теорему 11.2.

В заключение убедимся, что обычная формулировка теоремы Шеннона следует из приведенных результатов. Для этого возьмем в качестве ξ и ζ одинаковые дискретные величины с M значениями, а распределение $P(\xi, \zeta)$ выберем таким

$$P(\xi, \zeta) = \frac{1}{M} \delta_{\xi\zeta}.$$

При этом, очевидно,

$$I_{\xi\zeta} = H_{\xi} = H_{\zeta} = \ln M, \tag{11.5.20}$$

так что условие $I_{\xi\zeta} \to \infty$ выполняется, если $M \to \infty$. Далее $I(\xi, \zeta) = \ln M$ при $\xi = \zeta$; следовательно, второе условие Б информационной устойчивости § 7.3 тривиальным образом выполняется. Требование 1) теоремы 11.3 поэтому выполняется, если $M \to \infty$. Выбирая функцию штрафов (11.5.1), убеждаемся, что выполняются также требования 2), 3). Неравенство (11.5.9) согласно (11.5.20) принимает вид $\lim\sup \ln M/C < 1$, что эквивалентно соотношению (8.1.5). При функции штрафов (11.5.1) имеем $\mathbf{M}c(\xi, \zeta) = -1$ и соотношение (11.5.10) принимает вид

$$1 - \mathbf{M}\delta_{\xi\tilde{\zeta}} = \sum_{\xi - \zeta} P(\xi, \tilde{\zeta}) \to 0, \tag{11.5.21}$$

но

$$\sum_{\xi \neq \zeta} P(\xi, \tilde{\zeta}) = \frac{1}{M} \sum_{\xi} \mathbf{P}(\tilde{\zeta} \neq \xi \mid \xi)$$

есть не что иное, как средняя вероятность ошибки (7.1.11), так что сходимость (11.5.21) совпадает с (7.3.7). Итак, мы получили, что в данном частном случае теорема 11.3 действительно дублирует теорему 7.2.

Глава 12

ТЕОРИЯ ИНФОРМАЦИИ И ВТОРОЙ ЗАКОН ТЕРМОДИНАМИКИ

В этой главе обсуждается связь понятий количества информации и физической энтропии. Как известно, последнее понятие позволяет дать количественную формулировку второго закона термодинамики, который запрещает в изолированной системе процессы, сопровождающиеся увеличением энтропии. Если возможен приток информации dI о системе, т. е. если физическая система является изолированной *лишь* в тепловом, но не в информационном отношении, то указанный закон следует обобщить, заменив неравенство $dH \geqslant 0$ неравенством $dH + dI \geqslant 0$. Следовательно, если имеется приток информации, то можно тепловую энергию системы (без помощи холодильника) превратить в механическую. Другими словами, возможен вечный двигатель второго рода, питающийся информацией.

В § 12.1 и 12.2 процесс превращения тепловой энергии в механическую за счет информации рассматривается количественно, причем описывается конкретный механизм, позволяющий его осуществить. Этот механизм заключается в установке и передвижении определенным образом непроницаемых стенок внутри физической системы. Тем самым известные относящиеся к этому вопросу полукачественные рассуждения, содержащиеся, например, в книге Бриллюэна [1], получают точное количественное подтверждение.

Обобщение второго закона термодинамики никоим образом не отменяет его первоначальной формулировки. Отсюда в § 12.3 делается заключение о необходимости энергетических затрат при фактическом измерении координат физической системы и записи этой информации. Если система находится при температуре T, то для получения и записи количества информации dI о ней необходимо потратить как минимум TdI энергии. В противном случае соединение автоматического измерителя и информационного преобразователя тепловой энергии в механическую дало бы вечный двигатель второго рода. Указанная общая закономерность подтверждается для конкретной модели измерителя, описанной в § 12.3.

Вывод о необходимости минимальных энергетических затрат распространяется также на физические каналы с шумом, соответствующим заданной температуре T (§ 12.5). Таким образом, второй закон термодинамики накладывает некоторые ограничения на возможности физической реализации инфомационных систем: автоматов-измерителей и каналов.

12.1. Информация о физической системе, находящейся в состоянии термодинамического равновесия. Обобщенный второй закон термодинамики

В теории ценности информации (гл. 9) рассматривается информация о координате x, которая является случайной величиной, имеющей закон распределения $p(x)\,dx$. В настоящем параграфе, чтобы выявить связь теории информации с законами термодинамики, будем предполагать, что x является непрерывной координатой физической системы, находящейся в состоянии термодинамического равновесия. Энергия системы предполагается известной функцией $E(x)$ от этой координаты. Будем рассматривать состояние, соответствующее температуре T. В этом случае распределение определяется формулой Больцмана—Гиббса

$$p(x) = \exp\{[F - E(x)]/T\}, \tag{12.1.1}$$

где

$$F = -T\ln\int e^{-E(x)/T}\,dx \tag{12.1.2}$$

— свободная энергия системы. Температура T берется в энергетических единицах, при которых постоянная Больцмана равна 1.

Удобно предполагать, что имеется термостат с температурой T и указанное распределение устанавливается в результате длительного контакта с этим термостатом.

С точки зрения общей теории ценности информации распределение (12.1.1) является специальным случаем распределения вероятностей, входящих в определение бейесовской системы. На него, естественно, распространяются общие результаты, полученные в гл. 9 и 10 для произвольных бейесовских систем. Кроме этого, физический характер рассматриваемой системы позволяет исследовать особые явления, связанные со вторым законом термодинамики. Нас будет интересовать здесь возможность превращения тепловой энергии в механическую, обусловленная приходом информации о координате x.

При определении ценности хартлиевского и больцмановского количества информации (§ 9.2 и 9.6), предполагалось, что поступающая информация о значении x имеет простой вид. Указывается, какой именно области E_k из заданного разбиения $\sum_k E_k = X$ пространства X значений x принадлежит точка x. При этом поступившая информация эквивалентна указанию номера области E_k. Покажем, что такая информация действительно позволяет переводить тепловую энергию в механическую. При указании области E_k априорное распределение (12.1.1) переходит в апостериорное распределение

$$p(x|E_k)=\begin{cases}\exp\{[F(E_k)-E(x)]/T\} & \text{при } x\in E_k,\\ 0 & \text{вне } E_k,\end{cases} \tag{12.1.3}$$

где

$$F(E_k)=-T\ln\int_{E_k} e^{-E(x)/T}\,dx. \tag{12.1.4}$$

— условная свободная энергия. Поскольку известно, что x находится в области E_k, эту область можно окружить непроницаемыми стенками, вводя вместо энергетической функции $E(x)$ функцию

$$E(x|k)=\begin{cases}E(x) & \text{при } x\in E_k,\\ \infty & \text{вне } E_k.\end{cases} \tag{12.1.5}$$

Распределение (12.1.3) есть как раз равновесное распределение типа (12.1.1) для такой функции.

Затем будем медленно раздвигать стенки, окружающие область E_k, до тех пор, пока они не уйдут в бесконечность. На стенки действует давление, сила давления при раздвижении стенок будет совершать работу. Энергия в форме механической работы будет поступать ко внешним телам, механически связанным со стенками. Эта энергия равна хорошо известному в термодинамике интегралу типа

$$A=\int p\,dv \qquad \left(p=-\frac{\partial F}{\partial v}\text{ — давление}\right).$$

Дифференциал работы dA можно получить, варьируя область E_k в выражении (12.1.4). При расширении области E'_k до области $E''_k=E'_k+dE_k$ ко внешним телам переходит механическая энергия

$$dA=F(E'_k)-F(E'_k+dE_k)=T\int_{dE_k} e^{\frac{F(E'_k)-E(x)}{T}}\,dx. \tag{12.1.6}$$

В силу медленности раздвижения указанный энергетический переход совершается без изменения температуры системы. Это имеет место вследствие притока тепловой энергии из термостата, контакт с которым не должен прерываться. Тогда источником уходящей из системы механической энергии будет тепловая энергия термостата, которая будет превращаться в механическую работу. Чтобы подсчитать полную работу A_k, нужно просуммировать дифференциалы (12.1.6). При уходе стенок в бесконечность область E'_k совпадает со всем пространством X, а свободная энергия $F(E'_k)$ — с (12.1.2). Поэтому полная работа равна разности свободных энергий (12.1.2) и (12.1.4)

$$A_k=F(E_k)-F. \tag{12.1.7}$$

Если проинтегрировать (12.1.1) по E_k, то получим $P(E_k)$, аналогичный интеграл для (12.1.3) равен единице. Отсюда имеем формулу

$$e^{\frac{F-F(E_k)}{T}} = P(E_k). \qquad (12.1.8)$$

Учитывая (12.1.8), из (12.1.7) получаем

$$A_k = -T \ln P(E_k) = TH(|E_k) \qquad (12.1.9)$$

$H(|E_k)$ — условная энтропия).

Найденная формула соответствует тому случаю, когда точка x оказывается в области E_k, что происходит с вероятностью $P(E_k)$. Усредняя (12.1.9) по различным областям, нетрудно подсчитать среднюю энергию, превратившуюся из тепловой формы в механическую:

$$A = \sum_k A_k P(E_k) = TH_{E_k}. \qquad (12.1.9а)$$

Дополняя найденное соотношение знаком неравенства, относящегося к неравновесному (протекающему недостаточно медленно) процессу, имеем

$$A \leqslant TH_{E_k}. \qquad (12.1.10)$$

Итак, мы получили, что максимальное количество тепловой энергии, переходящее в работу, равно произведению абсолютной температуры на больцмановское количество приходящей информации. Приток информации о физической системе позволяет переводить тепловую энергию в работу без передачи части тепловой энергии холодильнику. Утверждение второго закона термодинамики о невозможности такого процесса справедливы лишь при отсутствии притока информации. При наличии притока информации dI обычная формулировка второго закона, разрешающая в изолированной системе лишь те процессы, в которых суммарная энтропия не уменьшается:

$$dH \geqslant 0, \qquad (12.1.11)$$

становится недостаточной. Условие (12.1.11) должно быть заменено условием неубывания суммы энтропии и информации:

$$dH + dI \geqslant 0. \qquad (12.1.12)$$

В рассмотренном выше процессе превращения теплоты в работу имел место приток информации $\Delta I = H_{E_k}$. Энтропия термостата уменьшалась на $\Delta H = -A/T$, энтропия рабочей системы в итоге не изменилась. Следовательно, для указанного процесса условие (12.1.12) справедливо со знаком равенства. Знак равенства обусловлен идеальным характером рассмотренного процесса, который был подобран специальным образом. Если бы стенками ограничивалась не область E_k, а более крупная область или если бы раздвижение стенок было не бесконечно медленным и т. п., то в условии (12.1.12) имел бы место знак неравенства. Полученное количество работы было бы меньше, чем (12.1.9а). Вследствие (12.1.12) большего количества работы, чем (12.1.9а), получить из теплоты невозможно.

Мысль о возможности описанного выше обобщения второго закона термодинамики на случай систем с притоком информации возникла давно в связи с обсуждением «демона Максвелла». Последний, открывая или закрывая дверцу в стенке между двумя сосудами (в зависимости от того, с какой скоростью подлетает молекула к дверце), может создать разность температур или разность давлений, не совершая работы, вопреки второму закону термодинамики. Для такой деятельности «демону» необходим приток информации. Пределы нарушения «демоном» второго закона термодинамики ограничены величиной приходящей информации. Согласно сказанному выше это можно утверждать не только качественно, но и формулировать в виде точного количественного закона (12.1.12).

В процессах, не связанных с притоком информации, второй закон термодинамики, взятый в обычном виде *(12.1.11), конечно*, остается незыблемым. Более того, даже формулу (12.1.10), соответствующую обобщенному закону (12.1.12), мы получили, по существу, опираясь на второй закон (12.1.11) при рассмотрении расширяющейся области E_k. Именно, формулы (12.1.6), (12.1.7) являются следствием второго закона (12.1.11). Покажем это. Обозначая через dQ количество теплоты, пришедшее из термостата, записываем изменение энтропии термостата в виде

$$dH_T = -dQ/T. \tag{12.1.13}$$

По первому закону термодинамики

$$dA = dQ - dU, \tag{12.1.14}$$

где $U = \mathbf{M}E(x)$ — внутренняя энергия системы, связанная со свободной энергией F известным соотношением $U = F + TH_x$. Дифференцируя последнее, имеем

$$dF = dU - TdH_x. \tag{12.1.15}$$

Второй закон (12.1.11) в данном случае имеет вид $dH_T + dH_x \geqslant 0$, $TdH_x - dQ \geqslant 0$, что в силу (12.1.14), (12.1.15), эквивалентно соотношению

$$dA \leqslant -dF. \tag{12.1.16}$$

Взяв это соотношение со знаком равенства, что соответствует идеальному процессу, получим первое соотношение (12.1.6).

12.2. Приток шенноновской информации и превращение теплоты в работу

Сказанное выше о возможности превращения тепловой энергии в механическую вследствие притока информации распространяется и на тот случай, когда информация носит более сложный характер, когда она не сводится к указанию области принадлежности E_k. Такой более сложный случай мы будем иметь, если область E_k указывается не безошибочно. Пусть $\widetilde{k}$ — номер области, сопряженный с возможной ошибкой, а k — истинный номер области, содержащей x. В данном случае количество поступающей информации определяется шенноновской формулой $I = H_{E_k} - H_{E_k \mid \widetilde{k}}$. Оно меньше, чем энтропия H_{E_k}, рассматривавшаяся в предыдущем параграфе.

Далее в более общем случае информация о значении x может поступать не в виде номера области, а в виде какой-то другой случайной величины y, статистически связанной с x. В этом случае количество информации также определяется шенноновской формулой [см. (6 2.1)]:

$$I = H_x - H_{x \mid y}. \tag{12.2.1}$$

Апостериорное распределение $p(x \mid y)$ теперь будет иметь более сложную форму, чем (12.1.3). Тем не менее обобщенный второй закон термодинамики будет иметь прежний вид (12.1.12), если под I понимать шенноновское количество информации (12.2.1). Формула (12.1.10) теперь заменится формулой

$$A \leqslant TI. \tag{12.2.2}$$

Чтобы в этом убедиться, следует рассмотреть бесконечно медленный изотермический переход от состояния, соответствующего апостериорному распределению $p(x \mid y)$ и имеющего энтропию $H_x(\mid y)$, к первоначальному (априорному) состоянию с заданным распределением $p(x)$. Этот переход должен происходить в соответствии со вторым законом термодинамики (12.1.11), т. е. в соответствии с формулами (12.1.13), (12.1.16). Суммируя элементарные работы (12.1.16), получаем, что каждому наблюденному значению y соответствует работа

$$A_y \leqslant -F + F(y). \tag{12.2.3}$$

Здесь F — свободная энергия (12.1.2), а $F(y)$ — свободная энергия

$$F(y) = \mathbf{M}[E(x) \mid y] - TH_x(\mid y), \tag{12.2.4}$$

(соответствующая апостериорному распределению $p(x \mid y)$, которая мыслится как равновесная при поставленных стенках (см. ниже). Подставляя (12.2.4) в (12.2.3) и усредняя по y при учете соотношения $F = \mathbf{M}E(x) - TH_x$, получаем

$$A \leqslant TH_x - TH_{x \mid y}, \tag{12.2.5}$$

т. е. неравенство (12.2.2).

Однако в данном более общем случае конкретный идеальный термодинамический процесс, при котором соотношение (12.2.2) справедливо со знаком равенства, сложнее, чем в § 12.1. Поскольку теперь апостериорная вероятность $p(x \mid y)$ не сконцентрирована в одной области E_k, мы не можем удалять в бесконечность **стенки,** поставленные вокруг этой области (так можно делать только, если, используя условие информационной устойчивости, перейти к рассмотрению безошибочной информации). Тем не менее использованный прием установки, передвижения и убирания стенок применим и теперь. Как и в § 12.1, установка и убирание стенок должны осуществляться мгновенно, без изменения энергии и энтропии; поэтому-то физическая равновесная свободная энергия после установки стенок оказывается равной выражению (12.2.4).

Рассмотрим несколько подробнее, как осуществить термодинамический процесс, близкий к идеальному. Возьмем апостериорное распределение $p(x \mid y)$ и установим в пространстве X систему стенок, которая разбивает это пространство на ячейки E_k. Если бы этих стенок не было, то в процессе релаксации распределение $p(x \mid y)$ само собой перешло в равновесное распределение $p(x)$ (12.1.1). Система стенок удерживает систему в состоянии с распределением $p(x \mid y)$. На стенки действуют со стороны физической системы некоторые механические силы. Будем теперь медленно перемещать стенки таким образом, чтобы фактическое распределение непрерывно перешло от неравновесного $p(x \mid y)$ к равновесному (12.1.1). Все промежуточные состояния при этом должны быть состояниями термодинамического равновесия для данной конфигурации стенок. Конечное состояние является состоянием равновесия не только при наличии стенок, но и при их отсутствии. Поэтому стенки под конец можно незаметно убрать. В процессе перемещения стенок механические силы, действующие на стенку, совершают механическую работу, среднее значение которой подсчитывается термодинамическими методами наподобие (12.2.3)—(12.2.5). Мы будем иметь в точности равенство $A = TI$, если удастся подобрать конфигурации стенок так, чтобы начальное распределение $p(x \mid y)$ и конечное (12.1.1) были точно равновесными для начальной и конечной конфигурации. Когда распределения $p(x)$, $p(x \mid y)$ неравномерные, чтобы приблизиться к пределу TI, может потребоваться специальный предельный переход, при котором размеры отдельных ячеек между стенками должны стремиться к нулю.

В качестве примера рассмотрим более простой случай, когда идеальный термодинамический процесс возможен при небольшом числе ячеек.

Пусть $X = [0, V]$ представляет собой отрезок, а функция $E(x) \equiv 0$ является постоянной, так что распределение (12.1.1) равномерное. Пусть далее производится измерение, в какой половине отрезка $[0, V]$ находится точка x. Пусть информация об этом передается с ошибкой, вероятность которой равна p. При этом количество информации равно

$$I = \ln 2 + p \ln p + (1 - p) \ln (1 - p). \qquad (12.2.5\text{a})$$

Пусть пришло сообщение, что $y = 1$, т. е. x находится в левой половине: $x \in [0, V/2]$. Этому сообщению соответствует апостериорное распределение

$$p(x \mid y=1) = \begin{cases} 2(1-p)/V & \text{при } 0 \leqslant x \leqslant V/2, \\ 2p/V & \text{при } V/2 < x \leqslant V. \end{cases} \tag{12.2.6}$$

В точке $x = z_0 = V/2$ устанавливаем стенку, которую затем будем медленно перемещать. Чтобы найти силы, действующие на стенку, для каждого положения z стенки вычислим свободную энергию. Поскольку $E(x) \equiv 0$, вычисление свободной энергии сводится к вычислению энтропии. Если стенка из точки $x = V/2$ перемещена в точку $x = z$, то распределение (12.2.6) заменяется распределением

$$p(x \mid z, 1) = \begin{cases} \dfrac{1-p}{z} & \text{при } 0 \leqslant x < z, \\[2ex] \dfrac{p}{V-z} & \text{при } z < x \leqslant V, \end{cases} \tag{12.2.7}$$

ему соответствует энтропия

$$H_x(\mid 1, z) = (1-p)\ln\frac{z}{1-p} + p\ln\frac{V-z}{p}$$

и свободная энергия

$$F(1, z) = -T(1-p)\ln\frac{z}{1-p} - Tp\ln\frac{V-z}{p}. \tag{12.2.8}$$

Дифференцируя по z, находим действующую на стенку силу

$$-\frac{\partial F(1, z)}{\partial z} = (1-p)\frac{T}{z} - p\frac{T}{V-z}. \tag{12.2.9}$$

Если бы координата x находилась слева, то действующая сила равнялась бы T/z (по аналогии с формулой давления идеального газа z играет роль объема), если бы x была справа, то действовала бы сила $-T/(V-z)$. Формула (12.2.9) дает апостериорное математическое ожидание этих сил, так как $1-p$ есть апостериорная вероятность неравенства $x < V/2$.

Работу силы (12.2.9) на отрезке $[z_0, z_1]$ можно вычислить, взяв разность потенциалов (12.2.8):

$$A_1 = -F(1, z_1) + F(1, z_0). \tag{12.2.10}$$

Начальное положение стенки, как указывалось, лежит посередине ($z_0 = V/2$). Конечное положение таково, что распределение (12.2.7) становится равновесным. Это дает

$$(1-p)/z_1 = p/(V-z_1) = 1/V, \qquad z_1 = (1-p)V.$$

Подставляя указанные значения z_0, z_1 в (12.2.10), (12.2.8), находим искомую работу

$$A_1 = T(1-p)\ln[2(1-p)] + Tp\ln(2p). \tag{12.2.11}$$

Аналогичный результат имеет место и для второго сообщения $y = 2$. При этом стенку нужно передвигать в другую сторону. Средняя работа A определяется тем же выражением (12.2.11). Сравнение этого выражения с (12.2.5а) показывает, что выполняется соотношение (12.2.2) со знаком равенства.

В заключение нужно указать, что условие равновесности исходного распределения $p(x)$ [см. (12.1.1)] является необязательным. Если исходное состояние $p(x)$ физической системы является неравновесным, то следует мгновенно (пока координаты x не успевают измениться) «включить в действие» некоторый новый гамильтониан $\widetilde{E}(x)$, отличающийся от первоначального $E(x)$ и такой, чтобы распределение $p(x)$ для него было равновесным. После этого можно применить все предыдущие рассуждения. Когда же процесс превращения тепловой энергии в работу будет закончен, следует мгновенно «выключить» $\widetilde{E}(x)$, перейдя к $E(x)$. Энергетические затраты на «включение» и «выключение» гамильтониана $\widetilde{E}(x)$ взаимно компенсируются. Итак, независимо от условия равновесности информационное обобщение второго закона термодинамики формулируется:

Т е о р е м а 12.1. *Если физическая система S изолирована в тепловом отношении и о ней имеется количество информации I* (неважно хартлиевское, больцмановское или шенноновское), *то в ней возможны лишь такие процессы, для которых изменение суммарной энтропии превосходит $-I$: $\Delta H \geqslant -I$. При этом нижняя граница физически достижима.*

При упоминании о тепловой изоляции здесь имеется в виду, что тепловая энергия, которую при указанных процессах можно превратить в работу, берется из самой системы S, т. е. термостат включен в S.

Как известно, второй закон термодинамики является асимптотическим, не вполне точным. Он нарушается для процессов, связанных с тепловыми флюктуациями. Учитывая это обстоятельство, можно дать ему уточненную (ослабленную) формулировку: в теплоизолированной системе не могут происходить процессы, для которых приращение энтропии

$$\Delta H \ll -1. \tag{12.2.12}$$

Если энтропию брать в термодинамических единицах согласно (1.1.7), то в правой части (12.2.12), вместо 1, следует поставить постоянную Больцмана k. Тогда (12.2.12) примет вид $\Delta H_{\text{физ}} \lll -k$. Соответственно тому, как в уточненной формулировке, условие $\Delta H < 0$ заменено более сильным неравенством $\Delta H \ll -1$, можно изменить и формулировку теоремы 12.1. В уточненной формулировке она должна запрещать процессы, для которых

$$\Delta H + I \lll -1 \quad \text{или} \quad \Delta H_{\text{физ}} + kI \lll -k. \tag{12.2.13}$$

Член с I здесь существен, если $I \gg 1$.

Аналогичные уточнения можно ввести и в материал других параграфов, однако мы не будем на этом останавливаться.

12.3. Энергетические затраты на создание и запись информации. Пример

В предыдущих параграфах случайная переменная y, несущая информацию о координате (или координатах) x физической системы, предполагалась заранее заданной случайной величиной, статистически связанной (коррелированной) с x. Особого внимания заслуживает вопрос, как физически реализовать эту информационную случайную величину. Ее естественно трактовать как одну (или несколько) координату из координат (т. е. динамических переменных) некоторой другой физической системы S_0, которую назовем измерительным прибором. Второй закон термодинамики приводит к определенным утверждениям, касающимся процедуры физического создания такого информационного сигнала y, статистически связанного с x. Эти утверждения касаются энергетических затрат, необходимых для создания y, и являются своего рода обращением утверждений, изложенных в § 12.1 и 12.2.

Дело в том, что рассмотренную ранее физическую систему S (с термостатом), которая «превращает информацию и теплоту в работу», и измерительный прибор S_0 (действующий автоматически), который создает информацию, можно объединить в одну комбинированную систему и применить к ней второй закон термодинамики. К комбинированной системе уже не будет поступать информации, поэтому неравенство (12.1.12) для нее примет обычную форму (12.1.11). Если измерительный прибор или термостат, с которым он контактирует, имеют ту же температуру T, что и физическая система с координатой x, то итоговое превращение теплоты в работу должно отсутствовать по обычному второму закону термодинамики. Это значит, что механическая энергия типа (12.1.10) или (12.2.5) должна превратиться в теплоту в измерительном приборе. Мы получаем, следовательно, что всякий физический измерительный прибор, имеющий температуру T, для создания количества информации I о координате физической системы должен превращать в теплоту по меньшей мере энергию TI.

Это неотъемлемое свойство всякого физического измерителя проверим сначала на одной простой модели реального измерителя. Сконструируем измеритель таким образом, чтобы измерение координаты x физической системы не влияло на поведение этой системы. Удобно рассматривать двумерную или даже одномерную модель. Пусть x есть координата (или координаты) центра металлического шарика, который без трения движется внутри трубки из изолятора (или между параллельными пластинками). Внутрь изолятора вставлены (и зашлифованы) пары электродов — металлических пластин, на которые подается разность потенциалов. Когда шарик находится в определенном месте, он соединяет одну пару электродов и тем самым замыкает соответствующую цепь (рис. 12.1). По протеканию тока можно судить о месте положения шарика. Пространство значений x оказывается разбито на области $E_1, \ldots, E_M$, соответствующие размерам электродов. Если число пар электродов равно $M = 2^n$, т. е. целой степени от 2, то удобно выбрать n источников э. д. с. и так соединить электроды, чтобы факт протекания (или непротекания) тока от одного источника нес в себе один бит информации о номере области $E_1, \ldots, E_M$. Пример такого соединения для $n = 3$ показан на рис. 12.1 и 12.2. Когда цепь, соответствующая одному биту информации, замыкается, то возникающий ток перемагничивает магнитик, помещенный в катушке и играющий роль ячейки памяти. Если тока не возникает, магнитик остается намагниченным в другом направлении. В итоге в тройке магнитиков оказывается записанным (двоичным кодом) номер области E_k.

Размеры областей E_k, т. е. пластин, можно подбирать из соображения оптимальности, рассматриваемых в теории ценности информации. Если целью является получение максимальной работы

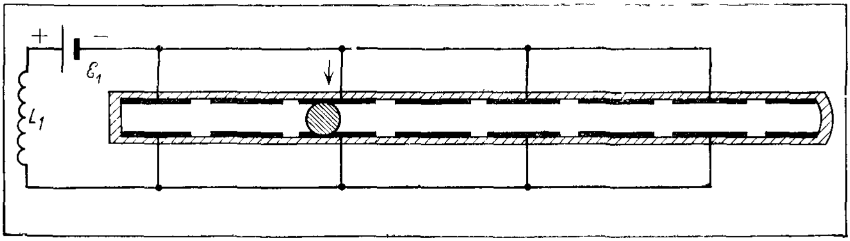


Рис. 12.1.
Расположение измерительных контактов в трубке.

(12.1.9а), то области E_k нужно выбирать так, чтобы им соответствовали одинаковые вероятности

$$p(E_1) = \ldots = p(E_M) = 1/M, \tag{12.3.1}$$

тогда больцмановское и хартлиевское количества информации будут совпадать: $H_{E_k} = \ln M$ и формула (12.1.9а) даст максимальную механическую энергию, равную $T \ln M$. Логарифм числа областей $\ln M$ можно назвать предельной информационной емкостью измерительного прибора.

В действительности количество информации, даваемое измерительным прибором, оказывается меньше, чем $\ln M$, вследствие ошибок, возникающих в приборе. Источником этих ошибок являются тепловые флюктуации, в частности флюктуации тока, протекающего через катушки L_r. Предполагаем, что температура T измерительного прибора задана (T берется в энергетических единицах). По законам статистической физики средняя энергия флюктуационного тока $i_ф$, текущего через катушку индуктивности L, определяется формулой

$$\frac{1}{2} LMi_ф^2 = \frac{1}{2} T, \tag{12.3.2}$$

т. е. пропорциональна абсолютной температуре T. Полезный ток $i_п$ от источника э. д. с. $\mathscr{E}_т$ (рис. 12.1 и 12.2) складывается с флюктуационным током $i_ф$. Средняя энергия полезного тока $Li_п^2/2$ и составляет в данном случае те самые энергетические затраты, о которых говорилось выше, которые присущи любому реальному измерителю. Найдем связь между энергетическими затратами и количеством информации, учитывая флюктуационный ток $i_ф$.

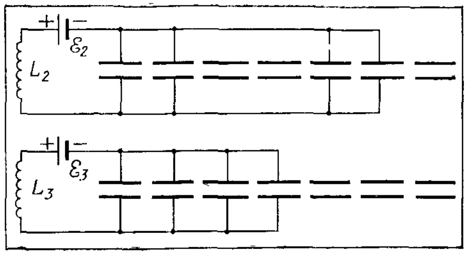


Рис. 12.2.
Другие способы соединения измерительных контактов.

При заданном полезном токе $i_п$ суммарный ток $i = i_ф + i_п$ имеет гауссово распределение с дисперсией, находимой из (12.3.2), т. е.

$$p(i \mid i_п) = \sqrt{\frac{L}{2\pi T}}\, e^{-L(i-i_п)^2/2T}. \tag{12.3.3}$$

Полезный ток [если выполняется (12.3.1)] при $M = 2^n$ с вероятностью $^1/_2$ принимает значение 0 и с вероятностью $^1/_2$ — некоторое значение i_1. Следовательно,

$$p(i) = \frac{1}{2}\sqrt{\frac{L}{2\pi T}}\,[e^{-Li^2/2T} + e^{-L(i-i_1)^2/2T}]. \tag{12.3.4}$$

Подсчитаем информацию связи $I_{i,i_п}$. Учитывая (12.3.3), (12.3.4), имеем

$$\ln\frac{p(i\mid 0)}{p(i)} = -\ln\frac{1}{2}[1 + e^{(Lii_1/T)-(Li_1^2/2T)}],\ \ln\frac{p(i\mid i_1)}{p(i)} =$$
$$= -\ln\frac{1}{2}[e^{-(Lii_1/T)+(Li_1^2/2T)} + 1].$$

Первое выражение нужно усреднить с весом $p(i|0)$, а второе — с весом $p(i|i_1)$. Вводя переменную $\xi = (i_1/2 - i)\sqrt{L/T}$ (и $\xi = (i - i_1/2)\sqrt{L/T}$ для второго интеграла), имеем для обоих интегралов одно и то же выражение

$$\int p(i\mid 0)\ln\frac{p(i\mid 0)}{p(i)}\,di = \int p(i\mid i_1)\ln\frac{p(i\mid i_1)}{p(i)}\,di =$$
$$= -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(\xi-\eta)^2}\ln\left[\frac{1}{2}+\frac{1}{2}e^{-2\xi\eta}\right]d\xi,$$

где $\eta = (i_1/2)\sqrt{L/T}$.

Следовательно, искомая информация равна этому же выражению. Приведем его к такому виду:

$$I_{i,i_{\text{п}}} = \eta^2 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(\xi-\eta)^2} \ln \operatorname{ch} \xi\eta d\xi.$$

Второй член заведомо отрицателен, поскольку $\operatorname{ch} \xi\eta > 1$, $\ln \operatorname{ch} \xi\eta > 0$, следовательно,

$$I_{i,i_{\text{п}}} \leqslant \eta^2 = Li_1^2/4T. \tag{12.3.5}$$

Но величина $Li_1^2/4$ есть не что иное, как средняя энергия полезного тока:

$$\mathbf{M}\, \frac{1}{2} Li_{\text{п}}^2 = \frac{1}{2}\left(\frac{1}{2} Li_1^2\right) + \frac{1}{2}\, 0.$$

Мы доказали тем самым, что для получения информации $I_{i,i_{\text{п}}}$ затрата средней энергии полезного тока должна превосходить $TI_{i,i_{\text{п}}}$. Сказанное относится к одной катушке. Суммируя по различным цепям, для получения суммарной информации $I = \sum_{r=1}^{n} (I_{i,i_{\text{п}}})_r$ необходимо затратить энергии в среднем не менее чем TI. Для данного измерительного прибора подтверждается, следовательно, сказанное выше об энергетических затратах, необходимых для получения информации. Учет тепловых флюктуаций в других элементах прибора еще более усиливает неравенство $A > TI$.

12.4. Энергетические затраты на создание и запись информации. Общая формулировка

В рассмотренном примере в роли информационной координаты y выступали токи i, текущие через катушки L_1, L_2, ... (рис. 12.1, 12.2). Чтобы иметь дело с более устойчивыми, лучше сохраняющимися во времени информационными переменными, целесообразно поместить в катушки магнитики, которые намагничивались бы при прохождении тока. Тогда в роли переменных y будут выступать соответствующие намагниченности m_1, m_2, ... Поскольку намагниченность m_r является некоторой функцией начальной m_{r0} (независимой от x) намагниченности и тока i ($m = f(m_0, i)$), то ясно, что количество информации I_{mx} не будет превосходить количества $I_{ix} + I_{m_0x} = I_{ix}$, т. е. рассмотренного ранее количества I_{ix}, так что неравенство $I_{ix} \leqslant A/T$ (12.3.5) при переходе к I_{mx} может только усилиться.

Тот случай, когда в роли информационного сигнала y выступают намагниченности m_1, m_2, ... запоминающих магнитиков, вполне аналогичен случаю записи информации на магнитофонной ленте, где число «элементарных магнитиков» весьма велико. Из рассмотренного примера видно, что физически процесс «создания» информации измерительным прибором неотделим от процесса физического записывания информации. Проверенное ранее для частного примера основное неравенство $IT \leqslant A$ можно доказать, исходя из общетермодинамических соображений, если ввести ряд общих и точных определений.

Пусть x — часть переменных ξ динамической системы S, а y — часть координат η системы S_0, относящихся к тому же самому моменту времени. Назовем *нормальным физическим записыванием информации* физический процесс, протекающий при взаимодействии систем S и S_0 между собой и, возможно, также с другими системами, при котором первоначальное состояние, характеризуемое мультипликативным совместным распределением $p_1(\xi)\,p_2(\eta)$ переходит в конечное состояние $p(\xi, \eta)$ с теми же парциальными распределениями $p_1(\xi)$ и $p_2(\eta)$. До начала процесса записывания и после него системы S и S_0 предполагаются невзаимодействующими.

Теперь можно дать общую формулировку основному утверждению.

Теорема 12.2 *Если нормальное физическое записывание информации протекает при контакте с термостатом, имеющим температуру T, то для его осуществления необходимо потребление и передача термостату (в виде теплоты) энергии*

$$A \geqslant IT, \tag{12.4.1}$$

где

$$I = H_x + H_y - H_{xy} \tag{12.4.2}$$

— шенноновское количество информации.

Доказательство. Обозначим через H_+ энтропию комбинированной системы $S + S_0$, а через H_T — энтропию термостата. Применяя второй закон термодинамики к процессу записывания информации, имеем

$$\Delta H_+ + \Delta H_T \geqslant 0. \tag{12.4.3}$$

При этом изменение энтропии H_+, очевидно, таково:

$$\Delta H_+ = H_{\xi\eta} - H_\xi - H_\eta = -I_{\xi\eta}. \tag{12.4.4}$$

Термостату в итоге передана энтропия $\Delta H_T \geqslant I_{\xi\eta}$ и, следовательно, отдана тепловая энергия $A \geqslant TI_{\xi\eta}$. Откуда она взялась? Согласно условиям теоремы взаимодействие систем S и S_0 в начале и в конце отсутствует, так что средняя суммарная энергия U_+ сводится к сумме средних парциальных энергий: $U_+ = \mathbf{M}E_1(\xi) + \mathbf{M}E_2(\eta)$. Они же остаются неизменными вследствие неизменности парциальных распределений $p_1(\xi)$ и $p_2(\eta)$. Итак $\Delta U_+ = 0$ и, следовательно, энергия A должна быть взята в процессе записывания информации извне от каких-то внешних источников нетепловой энергии. Для получения (12.4.1) остается лишь учесть неравенство $I_{xy} \leqslant I_{\xi\eta}$. Доказательство закончено.

В заключение этого параграфа рассмотрим еще один пример записи информации, совершенно другого типа, нежели пример из

§ 12.3. Из сравнения этих примеров видно, что источники внешней энергии могут иметь совершенно различную природу.

Пусть требуется получить и записать информацию о флюктуирующей координате x стрелки, вращающейся на оси и поддерживаемой вблизи положения равновесия пружинкой П (рис. 12.3).

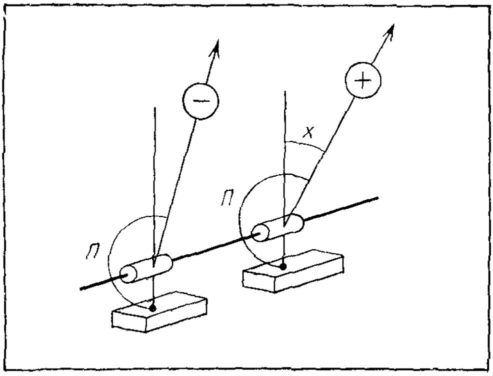


Рис. 12.3.
Пример создания и записи информации при помощи сдвигаемых и раздвигаемых взаимодействующих стрелок с пружинами П.

На конце стрелки помещен шарик, заряженный положительно. Чтобы осуществить «измерение и запись» координаты x стрелки, подносим к ней аналогичную же вторую стрелку, посаженную на ту же ось, но с шариком, заряженным отрицательно. При сближении стрелок мы воспримем некоторую энергию A_1, обусловленную притяжением шариков. Сближение предполагаем очень быстрым, так что координаты стрелок x, y в процессе него не успевают измениться. Получившееся после сближения состояние будет неравновесным. Оно перейдет в равновесное состояние, в котором между координатами стрелок x, y установится корреляция. Удобно предполагать, что сила притяжения шариков гораздо больше возвращающих сил пружинок, так что корреляция будет очень тесной. Переход в равновесное состояние будет сопровождаться уменьшением средней энергии взаимодействия шариков («спуск в потенциальную яму») и выделением в термостат (окружающую среду) тепловой энергии A. После того, как равновесное (коррелированное) распределение $p(x, y)$ установится, мы быстро раздвигаем стрелки (раздвижение, как и сближение, происходит вдоль оси вращения и не затрагивает вращательной координаты). При этом мы затрачиваем работу A_2, которая, разумеется, больше, чем воспринятая ранее работа A_1, поскольку модуль разности $|x - y|$ теперь стал в среднем меньше. Парциальные распределения $p(x)$ и $p(y)$ почти не изменятся, если среднее $\mathbf{M}(x - y)^2$ очень мало. Как показывает термодинамический анализ, подобный тому, который был приведен при доказательстве теоремы 12.2, в данном примере $A_2 - A_1 = A \geqslant IT$.

После раздвижения стрелок между ними имеется корреляция, но нет силового взаимодействия. Процесс «записи» информации закончен. Полученную «запись», разумеется, можно переработать, превратить в другую форму, скажем, записать на магнитофонную ленту. При этом количество записанной информации может только уменьшиться.

В данном примере работу, необходимую для записи информации, совершает человек или устройство, сближающее и раздвигающее стрелки. Такая работа, согласно общей теории, должна затрачиваться везде, где происходит *создание* информации, создание новых корреляций, а не просто переработка старых. Общая теория при этом указывает лишь нижний теоретический уровень этих затрат. Практические энергетические затраты, конечно, могут превосходить и даже значительно превосходить этот термодинамический уровень. Сравнение практических затрат с минимальными теоретическими позволяет судить о степени энергетического совершенства реальных устройств.

12.5. Энергетические затраты в физических каналах

Энергетические затраты необходимы не только для создания и записи информации, но и для ее передачи, если последняя происходит в условиях наличия флюктуационных возмущений, например тепловых. Как известно из статистической физики, в линейных системах на каждую степень свободы приходится средняя равновесная флюктуационная энергия, равная $T_ф/2$, где $T_ф$ — температура окружающей среды (термостата). В ряде работ (Брюллюэна [1] и др.) высказывалась мысль, что для передачи одного ната информации в этих условиях необходима энергия не меньше $T_ф$ (мы пользуемся энергетическими единицами температуры, так что постоянная Больцмана полагается равной единице). В настоящем параграфе мы пытаемся уточнить это положение и привести для него соответствующее доказательство.

Назовем *физическим каналом* канал, описываемый переходными вероятностями $p(y \mid x)$ и функцией штрафов $c(x)$, если y имеет смысл полного набора динамических переменных некоторой физической системы S. Функцию Гамильтона (энергию) последней обозначим через $E(y)$ (она неотрицательна). Зная ее, по обычным формулам можно вычислить равновесный потенциал

$$\Gamma(\beta) = -\beta F = \ln \int e^{-\beta E(y)}\, dy \tag{12.5.1}$$

и зависимость энтропии от средней энергии

$$H(R) = \Gamma + \beta R \quad \left(R = -\frac{d\Gamma}{d\beta}(\beta) \geqslant 0 \right). \tag{12.5.2}$$

Теорема 12.3 *Пропускная способность* (см. § 8.1) *физического канала* $[p(y \mid x),\ c(x)]$ *удовлетворяет неравенству*

$$T_ф C(a) \leqslant \mathbf{M}E(y) - a_ф, \tag{12.5.3}$$

где уровень $a_ф$ и «температура флюктуаций» $T_ф$ определены равенствами

$$T_ф^{-1} = \frac{dH}{dR}(a_ф); \qquad H(a_ф) = H_{y \mid x}. \tag{12.5.4}$$

Средняя энергия $\mathbf{M}E(y)$ *и условная энтропия* $H_{y \mid x}$ *подсчитываются при экстремальном распределении* $p_0(x)$, *реализующим пропускную способность, которое предполагается существующим. Предполагается также, что второе уравнение* (12.5.4) *имеет корень* $a_ф$ *в нормальной ветви, где* $T_ф > 0$.

Доказательство. Формулы (12.5.1), (12.5.2) появляются при решении первой вариационной задачи (см., например, § 3.6)— задачи на экстремум энтропии H_y при ограничении $\mathbf{M}E(y) = A$. Поэтому выполняется неравенство

$$H_y \leqslant H(\mathbf{M}E(y)) \tag{12.5.5}$$

(фиксирован уровень $\mathbf{M}E(y)$). Далее, как следует из общей теории (см. следствие из теоремы 4.1), функция $H(R)$ является выпуклой так что ее производная

$$\beta(R) = \frac{dH(R)}{dR} \tag{12.5.6}$$

— невозрастающая функция от R.

Пропускная способность канала совпадает с шенноновским количеством информации

$$C(a) = H_y - H_{y \mid x} \tag{12.5.7}$$

для экстремального распределения $p_0(x)$, обращающего его в условный максимум. Из (12.5.5) и обычного неравенства $H_y \geqslant H_{y|x}$ находим

$$H_{y \mid x} \leqslant H(\mathbf{M}E(y)). \tag{12.5.8}$$

Поскольку $a_ф$ — корень уравнения $H_{y \mid x} = H(a_ф)$ в нормальной ветви выпуклой функции $H(\cdot)$, где производная неотрицательна, то в какой бы ветви ни лежало значение $\mathbf{M}E(y)$, из (12.5.8) имеем $a_ф \leqslant \mathbf{M}E(y)$. Учитывая его и невозрастающий характер производной (12.5.6), получаем $H(\mathbf{M}E(y)) - H(a_ф) \leqslant \beta(a_ф)[\mathbf{M}E(y) - a_ф]$. Это неравенство в сочетании с $C(a) \leqslant H(\mathbf{M}E(y)) - H(a_ф)$, вытекающим из (12.5.7) и (12.5.5), дает (12.5.3). Доказательство закончено.

Вследствие положительности величины $a_ф$, вытекающей из неотрицательного характера энергии $E(y)$, в (12.5.3) член $-a_ф$ можно отбросить. От этого неравенство может только усилиться. Однако это делать не всегда целесообразно, поскольку формула (12.5.3) допускает следующую простую интерпретацию. Величину $a_ф$ можно трактовать как среднюю «энергию флюктуационного шума», а разность $\mathbf{M}E(y) - a_ф$ — как оставшуюся после прохождения через канал «энергию полезного сигнала». Тогда согласно (12.5.3) на каждый нат пропускной способности будет приходиться по меньшей мере «энергия полезного сигнала» $T_ф$. Если a — энергия полезного сигнала до прохождения через канал, которую естественно считать превосходящей оставшуюся энергию $\mathbf{M}E(y) - a_ф$, то тем более будет выполняться неравенство $a/C(a) \geqslant T_ф$.

Проще всего конкретизировать эти рассуждения для случая квадратичной энергии

$$E(y) = \sum_{j,k} e_{jk}\, y_j\, y_k \equiv y^T e_2 y \tag{12.5.9}$$

и аддитивного независимого шума ζ (с нулевым средним значением):

$$y = x + \zeta. \tag{12.5.10}$$

Подставляя (12.5.10) в (12.5.9) и усредняя, имеем

$$\mathbf{M}E(y) = \mathbf{M}\, x^T e_2 x + \mathbf{M}\, \zeta^T e_2 \zeta,$$

т. е.

$$\mathbf{M}E(y) = \mathbf{M}E(x) + \mathbf{M}E(\zeta). \tag{12.5.11}$$

Вследствие (12.5.10) и независимости шума далее имеем

$$H_{y\,|\,x} = H_\zeta. \tag{12.5.12}$$

Первую экстремальную задачу, решением которой является рассмотренная ранее функция $H(R)$ или обратная функция $R(H)$, можно, как известно (§ 3.2), трактовать как задачу минимизации средней энергии $\mathbf{M}E$ при фиксированной энтропии. Следовательно, $R(H_\zeta)$ есть минимальное значение энергии, возможное при фиксированной энтропии H_ζ, т. е.

$$\mathbf{M}E(\zeta) \geqslant R(H_\zeta), \tag{12.5.13}$$

но $R(H_\zeta)$ в силу (12.5.4), (12.5.12) есть не что иное, как $a_ф$, поэтому (12.5.13) можно записать

$$\mathbf{M}E(\zeta) \geqslant a_ф. \tag{12.5.14}$$

Из (12.5.11) и (12.5.14) получаем $\mathbf{M}E(y) - a_ф \leqslant \mathbf{M}E(x)$. Это неравенство позволяет преобразовать основное доказанное нами неравенство (12.5.3) к виду $T_ф C(a) \leqslant \mathbf{M}E(x)$ или

$$T_ф C(a) \leqslant a, \tag{12.5.15}$$

если функция штрафов $c(x)$ совпадает с энергией $E(x)$.

Приведенные здесь результаты, касающиеся физических каналов, тесно связаны с теоремой 8.5. В этой теореме, однако, «температурный» параметр $1/\beta$ имеет формально-математический смысл. Чтобы он приобрел смысл физической температуры, нужно штрафы $c(x)$ или $b(y)$ специализировать как физическую энергию.

Согласно неравенству (12.5.15) в гауссовых физических каналах для передачи одного ната информации требуется по меньшей мере энергия $T_ф$. Следует отметить, что какого-либо универсального неравенства типа (12.5.15), включающего не эффективную, а фактическую температуру канала, вывести не удается.

НЕКОТОРЫЕ МАТРИЧНЫЕ (ОПЕРАТОРНЫЕ) ТОЖДЕСТВА

П.1. Правила переноса оператора слева направо

Пусть заданы две произвольные не обязательно квадратные матрицы A и B, такие, чтобы произведения BA и AB имели смысл. На языке операторов это означает следующее: если A переводит элемент пространства X в элемент пространства Y, то B переводит элемент пространства Y в элемент пространства X, т. е. задает обратное отображение. При широких предположениях относительно функции $f(z)$ справедлива формула

$$\mathrm{A}f(\mathrm{BA}) = f(\mathrm{AB})\,\mathrm{A}. \tag{П.1.1}$$

Докажем ее в предположении, что функция f представима разложением Тейлора

$$f(\mathrm{z}) = \sum_{n=0}^{\infty} \frac{1}{n!} f^n(0)\,\mathrm{z}^n, \tag{П.1.2}$$

где $\mathrm{z} = \mathrm{AB}$ или BA. (Большое число обобщений может быть получено предельным переходом $\lim f_m = f$, где f_m — последовательность подходящих представимых функций).

Подставляя (П.1.2) в (П.1.1), убеждаемся, что как правая, так и левая часть равенства обращаются в одно и то же выражение

$$\sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(0)\,\mathrm{ABA}\ldots\mathrm{ABA}.$$

Это доказывает равенство (П.1.1).

Матрицы AB и BA имеют, вообще говоря, разную размерность, являются операторами, действующими в различных пространствах (первый в Y, второй в X). То же самое относится и к матрицам $f(\mathrm{AB})$ и $f(\mathrm{BA})$. Сравним, однако, их следы. Пользуясь разложением (П.1.2), имеем

$$Sp\, f(\mathrm{AB}) = \mathrm{Sp}\, f(0) + f'(0)\, Sp\,(\mathrm{AB}) + \frac{1}{2} f''(0)\, Sp\,(\mathrm{ABAB}) + \ldots, \tag{П.1.3}$$

$$Sp\, f(\mathrm{BA}) = Sp\, f(0) + f'(0)\, Sp\,(\mathrm{BA}) + \frac{1}{2} f''(0)\, Sp\,(\mathrm{BABA}) + \ldots \tag{П.1.4}$$

Но $Sp\,\mathrm{AB} = Sp\,\mathrm{BA} = \sum_{ij} A_{ij} B_{ji}$, а значит и $Sp\,\mathrm{A}\,[(\mathrm{BA})^k\,\mathrm{B}] = Sp\,[(\mathrm{BA})^k\,\mathrm{B}]\,\mathrm{A}$. Поэтому все члены в (П.1.3), (П.1.4), кроме первого, совпадают. Первые члены $\mathrm{Sp}\, f(0)$ в общем случае не совпадают, так как оператор $f(0)$ в разложении функции $f(\mathrm{AB})$ и тот же оператор в разложении $f(\mathrm{BA})$ кратен единичной матрице различных размерностей. Если же выполнено условие

$$f(0) = 0, \tag{П.1.5}$$

то, следовательно,

$$\operatorname{Sp} f(\mathrm{AB}) = \operatorname{Sp} f(\mathrm{BA}). \tag{П.1.6}$$

Если бы мы интересовались детерминантами, то для соответствующего равенства

$$\det F(\mathrm{AB}) = \det F(\mathrm{BA}) \tag{П.1.7}$$

вместо (П.1.5) имели бы условие $F(0) = 1$, так как $\ln \det F = \operatorname{Sp} \ln F$ $(\ln F = f)$.

П.2. Детерминант составной матрицы

Требуется вычислить детерминант матрицы

$$\mathrm{K} = \begin{pmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{C} & \mathrm{D} \end{pmatrix}, \tag{П.2.1}$$

где A, D — квадратные, а B, C не обязательно квадратные матрицы. Матрицу D предполагаем невырожденной. Обозначим

$$\mathrm{L} = \begin{pmatrix} 1 & 0 \\ 0 & \mathrm{D}^{-1} \end{pmatrix} \begin{pmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{C} & \mathrm{D} \end{pmatrix} = \begin{pmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{D}^{-1}\mathrm{C} & 1 \end{pmatrix}, \text{ так что } \mathrm{K} = \begin{pmatrix} 1 & 0 \\ 0 & \mathrm{D} \end{pmatrix} \mathrm{L}$$

(1 и 0 здесь матрицы) и, следовательно,

$$\det \mathrm{K} = \det \mathrm{D} \det \mathrm{L}. \tag{П.2.2}$$

Непосредственным перемножением легко проверить, что

$$\begin{pmatrix} A & \mathrm{B} \\ \mathrm{D}^{-1}\mathrm{C} & 1 \end{pmatrix} = \begin{pmatrix} 1 & \mathrm{B} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathrm{A} - \mathrm{BD}^{-1}\mathrm{C} & 0 \\ \mathrm{D}^{-1}\mathrm{C} & 1 \end{pmatrix}, \tag{П.2.3}$$

но $\det \begin{pmatrix} 1 & \mathrm{B} \\ 0 & 1 \end{pmatrix} = 1$; $\det \begin{pmatrix} \mathrm{A} - \mathrm{BD}^{-1}\mathrm{C} & 0 \\ \mathrm{D}^{-1}\mathrm{C} & 1 \end{pmatrix} = \det(\mathrm{A} - \mathrm{BD}^{-1}\mathrm{C})$.

Поэтому (П.2.3) дает $\det \mathrm{L} = \det(\mathrm{A} - \mathrm{BD}^{-1}\mathrm{C})$. Подставляя это равенство в (П.2.2), получаем

$$\det \begin{pmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{C} & \mathrm{D} \end{pmatrix} = \det \mathrm{D} \det(\mathrm{A} - \mathrm{BD}^{-1}\mathrm{C}). \tag{П.2.4}$$

Аналогично, если A^{-1} существует, то

$$\det \begin{pmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{C} & \mathrm{D} \end{pmatrix} = \det \mathrm{A} \det(\mathrm{D} - \mathrm{CA}^{-1}\mathrm{B}). \tag{П.2.5}$$

Согласно указанным формулам задача вычисления исходного детерминанта сводится к задаче вычисления детерминантов меньшей размерности.

Список литературы

Беллман Р. (Bellman R).

1. Динамическое программирование. Пер. с англ., М., ИЛ, 1960.

Бергер Т. (Berger T.)

1. Rate distortion theory. A Mathematical basis for data compression. Prentice-Hall, Inc. Englewood Cliffs, New Jersey, 1971.

Бриллюен Л. (Brillouin L.).

1. Наука и теория информации. Пер. с франц. М., Физматгиз, 1960.

Добрушин Р. Л.

1. Общая формулировка основной теоремы Шеннона в теории информации. — УМН, 1959, т. 14, вып. 6.

2. Математические вопросы шенноновской теории оптимального кодирования информации. — «Проблемы передачи информации», 1961, вып. 10.

Дуб Дж. Л. (Doob J. L.).

1. Вероятностные процессы. Пер. с англ., М., ИЛ, 1956.

Гнеденко Б. В.

1. Курс теории вероятностей. М., Физматгиз, 1961.

Голдман С. (Goldman S.).

1. Теория информации. Пер. с англ., М., ИЛ, 1957.

Гришанин Б. А., Стратонович Р. Л.

1. Ценность информации и достаточные статистики при наблюдении случайного процесса. — «Известия АН СССР. Техническая кибернетика», 1966, № 6, с. 4—12.

Колмогоров А. Н.

1. Теория передачи информации. — В кн.: Сессия АН СССР по научным проблемам автоматизации производства, 1956. Пленарные заседания. М., Изд. АН СССР, 1957.

Крафт Л. Г. (Kraft L. G.).

1. A device for quantizing, grouping and coding amplitude modulated pulses.—M. S. Thesis, Electrical Engeneering dept., M. I. T., 1949, March.

Кульбак С. (Kullback S.).

1. Теория информации и статистика. Пер. с англ., М., «Наука», 1967.

Леонтович М. А.

1. Статистическая физика. М. — Л., Гостехиздат, 1944.

2. Введение в термодинамику. М. — Л. Гостехиздат, 1952.

Пинскер М. С.

1. Количество информации о гауссовом случайном процессе, содержащееся во втором процессе, стационарно с ним связанном. — «ДАН СССР», 1954, т. 98, с. 213—216.

2. Вычисление скорости создания сообщений стационарным случайным процессом и пропускной способности стационарного канала. — ДАН СССР, 1956, т. 111, вып. 4, с. 753—756.

Рао С. Р. (Rao C. R.)

1. *Линейные статистические методы и их применения. Пер. с англ.,* М., ИЛ, 1968.

Рыжик И. М., Градштейн И. С.

1. Таблицы интегралов, сумм, рядов и произведений. М., Гостехиздат, 1951.

Стратонович Р. Л.

1. Энтропия систем со случайным числом частиц. — ЖЭТФ, 1955, т. 28, вып. 4.

2. О статистике намагниченности в модели Изинга. — «Физика твердого тела», 1961, т. 3, вып. 10.

3. О ценности информации. — «Известия АН СССР. Техническая кибернетика», 1965, № 5, с. 3—12.

4. Условные марковские процессы и их применение к теории оптимального управления. МГУ, 1966.

5. Ценность информации при наблюдении случайного процесса в системах, содержащих конечные автоматы. — «Известия АН СССР. Техническая кибернетика», 1966, т. 5, с. 3—13.

6. Количество информации и энтропия отрезков стационарных гауссовых процессов. — «Проблемы передачи информации», 1967, т. 3, вып. 2.

7. Экстремальные задачи теории информации и динамическое программирование. — «Известия АН СССР. Техническая кибернетика», 1967, № 5, с. 63—77.

Стратонович Р. Л., Гришанин Б. А.

1. Ценность информации при невозможности прямого наблюдения оцениваемой случайной величины. — «Известия АН СССР. Техническая кибернетика», 1966, № 3, с. 3—15.

2. Игровые задачи с ограничениями информационного типа. — «Известия АН СССР. Техническая кибернетика», 1968, № 1, с. 3—12.

Файнстейн А. (Feinstein A.)

1. Основы теории информации. Пер. с англ., М., ИЛ, 1960.

Фано Р. М. (Fano R. M.)

1. Передача информации. Статистическая теория связи. Пер. с англ., М. «Мир», 1965.

Хартли Р. В. Л. (Hartley R. V. L.)

1. Передача информации, 1928. — В кн.: Теория информации и ее приложения. Пер. с англ. Под ред. А. А. Харкевича, М., Физматгиз, 1959.

Хилл Т. Л. (Hill T. L.)

1. Статистическая механика. Пер. с англ., М., ИЛ, 1960.

Хуфман Д. А. (Huffman D. A.)

1. A method for the construction of minimum redundancy codes. — «Proc. IRE», 1952, v. 40.

Шеннон К. (Shannon C.)

1. Математическая теория связи, 1948. — В кн.: К. Шеннон. Работы по теории информации и кибернетике. Пер. с англ. Под ред. Р. Л. Добрушина и О. Б. Лупанова, М., ИЛ, 1963.

2. Связь при наличии шума, 1949. Там же.

3. Некоторые результаты теории кодирования для каналов с шумами, 1957. Там же.

4. Теоремы кодирования для дискретного источника при заданном критерии точности, 1959. Там же.

Янке Е., Эмде Ф.

1. Таблицы функций с формулами и кривыми. Пер. с нем., М. — Л., Гостехиздат, 1949.

Предметный указатель

а-информация 308
Бит 13
Больцмана формула 10
Вероятность апостериорная финальная 123
— ошибки 229
— — средняя 230
Ветвь аномальная 265, 308
— нормальная 265, 308
Гиббса распределение каноническое 70, 85, 89, 398
— теорема 89
Длина записи 45, 51
W-процесс вторичный апостериорный 124
Емкость информационная 65
Задача вариационная первая 65, 67
— — вторая 257, 258, 259
— — третья 311
Закон сохранения количества информации 43
— термодинамики второй 400, 405
— — — обобщенный 400, 405
Иенсона неравенство 15
Информация связи парная (взаимная) 184
— — случайная 187
— — тройная 193
— — удельная 206
— — условная 189
Канал абстрактный 258, 260
— аддитивный 291
— гауссов 275
— — стационарный 285
— двоичный 272, 274
— дискретный без помех 64
— симметричный 270
— физический 413
— — пропускная способность 413
Код 45, 228
— дешифруемый 49
— — Крафта 49
— оптимальный 46
— Шеннона случайный 231
Кодирование информации 43, 44
— — «блочное» 44
— — оптимальное 43
— — скользящее (текущее) 44
Количество информации больцмановское 14, 326
— — хартлиевское 12
— — шенноновское 181, 186
Котельникова теорема 290
Леви формула 104
Лежандра преобразование 80, 88, 239
Маркова условие 122
Метод неопределенных множителей Лагранжа 67
«Микросостояние» 10
«Модель Изинга» 77
Нат 13
Негинформация 297
Область «активная» 67, 260, 314
Ошибка декодирования 58, 230
Параметр канонический 85, 86
— термодинамический внешний 79
— — внутренний 79, 85
— — сопряженный 80, 86
Плотность энтропии 165
Помеха простая 183
Последовательность бейесовских систем информационно-устойчивая 374
Потенциал термодинамический 73
— характеристический 29, 86, 101, 133, 197, 203, 211
— — условный 249
Принцип аддитивности 12
Пропускная способность 61, 65, 258
Процесс вторичный апостериорный 119, 124
— дискретный 110
— — марковский 114
— — — стационарный 114
— — стационарный 110
— марковский диффузный 165
— — условный 120, 172, 215
— — —, энтропия 120
— стационарно-связанный 205
— стационарный периодический 145
— точечный случайный 152
Равноценность информации асимптотическая 367
Радона-Никодимова производная 34
Распределение каноническое 86
— экстремальное 308
Риск 307
Свободная энергия 70
Свойство иерархической аддитивности 19, 39, 190, 191
Система бейесовская 307
— — гауссовая 344
— — — стационарная 353
Случайный поток 152
Сообщение элементарное 47
Соотношение термодинамическое 70, 71, 263, 316
Стирлинга формула 158
Сумма статистическая 70, 82
Теорема асимптотическая вторая 234, 236
— — первая 99
— — третья 363, 367, 374
Условие мультипликативности 40, 42
— нормировки 67, 312
Условный марковский процесс 172, 215
Устойчивость информационная 236
— каноническая 95
— энтропийная 25, 28
— — достаточное условие 27
Фокера — Планка уравнение 165, 175
— — — стационарное 168
Функция правдоподобия 229
— семиинвариантов производящая 87
— ценности информации 303, 329
— штрафов 64, 303
Хартли формула 10, 13
Хинчина теорема 235
Ценность информации 296, 298, 303
— — больцмановской 326, 327
— — дифференциальная 298, 299
— — случайная 320
— — хартлиевской 304
— — шенноновской 308, 329
Цепь марковская 114
Чебышева неравенство 22
Чернова неравенство 102
Шеннона теорема 234, 259
— — обобщенная 360, 392, 394
Штраф средний 307
Энтропия 10, 11
— больцмановская 14
— конца отрезка 109, 113
— максимальное значение 15, 38
— непрерывной случайной величины 33, 34
— свойства 15, 16
— случайная 14
— удельная 23, 109, 111, 165
— условная 17, 39
Энергия свободная 70

Оглавление

Предисловие 3
Введение 5

Глава 1

ОПРЕДЕЛЕНИЕ ИНФОРМАЦИИ И ЭНТРОПИИ ПРИ ОТСУТСТВИИ ПОМЕХ 9

§ 1.1. Определение энтропии в случае равновероятных возможностей 11
§ 1.2. Энтропия в случае неравновероятных возможностей и ее свойства 13
§ 1.3. Условная энтропия. Свойство иерархической аддитивности 16
§ 1.4. Асимптотическая эквивалентность неравновероятных возможностей равновероятным 21
§ 1.5. Асимптотическая равновероятность и энтропийная устойчивость 25
§ 1.6. Определение энтропии непрерывной случайной величины . . . 30
§ 1.7. Свойства энтропии в обобщенной версии. Условная энтропия 37

Глава 2

КОДИРОВАНИЕ ДИСКРЕТНОЙ ИНФОРМАЦИИ ПРИ ОТСУТСТВИИ ПОМЕХ И ШТРАФОВ 43

§ 2.1. Основные принципы кодирования дискретной информации 44
§ 2.2. Основные теоремы для кодирования без помех. Равнораспределенные независимые сообщения 48
§ 2.3. Оптимальное кодирование по Хуфману. Примеры. 53
§ 2.4. Погрешности кодирования без помех при конечной длине записи 57

Глава 3

КОДИРОВАНИЕ ПРИ НАЛИЧИИ ШТРАФОВ. ПЕРВАЯ ВАРИАЦИОННАЯ ЗАДАЧА 61

§ 3.1. Прямой способ вычисления информационной емкости записи для одного примера 62
§ 3.2. Дискретный канал без помех и его пропускная способность 64
§ 3.3. Решение первой вариационной задачи. Термодинамические параметры и потенциалы 67
§ 3.4. Примеры применения общих методов вычисления пропускной способности 73
§ 3.5. Метод потенциалов в случае бóльшего числа параметров . . 79
§ 3.6. Пропускная способность канала без шумов со штрафами в обобщенной версии 82

Глава 4

ПЕРВАЯ АСИМПТОТИЧЕСКАЯ ТЕОРЕМА И СВЯЗАННЫЕ С НЕЙ РЕЗУЛЬТАТЫ 84

§ 4.1. Потенциал Г или производящая функция семиинвариантов 85
§ 4.2. Некоторые асимптотические результаты статистической термодинамики. Устойчивость канонического распределения . . 89
§ 4.3. Асимптотическая эквивалентность двух видов ограничений 96
§ 4.4. Некоторые теоремы, касающиеся характеристического потенциала 101

Глава 5

ВЫЧИСЛЕНИЕ ЭНТРОПИИ ДЛЯ ЧАСТНЫХ СЛУЧАЕВ. ЭНТРОПИЯ СЛУЧАЙНЫХ ПРОЦЕССОВ 109

§ 5.1. Энтропия отрезка стационарного дискретного процесса и удельная энтропия 110
§ 5.2. Энтропия марковской цепи 114
§ 5.3. Удельная энтропия части компонент дискретного марковского процесса и условного марковского процесса 120
§ 5.4. Энтропия гауссовых случайных величин 129
§ 5.5. Энтропия стационарной последовательности. Гауссова последовательность 134
§ 5.6. Энтропия случайных процессов в непрерывном времени. Общие понятия и соотношения 141
§ 5.7. Энтропия гауссового процесса в непрерывном времени . . . 144
§ 5.8. Энтропия точечного случайного процесса 152
§ 5.9. Энтропия дискретного марковского процесса в непрерывном времени 162
§ 5.10. Энтропия диффузионных марковских процессов 165
§ 5.11. Энтропия комбинированного марковского процесса, условного процесса и части компонент марковского процесса . . . 170

Глава 6

ИНФОРМАЦИЯ ПРИ НАЛИЧИИ ПОМЕХ. ШЕННОНОВСКОЕ КОЛИЧЕСТВО ИНФОРМАЦИИ 181

§ 6.1. Потери информации при вырожденных преобразованиях и при простых помехах 181
§ 6.2. Информация связи дискретных случайных величин 186
§ 6.3. Условная информация. Иерархическая аддитивность информации 189
§ 6.4. Количество информации связи в общем случае 195
§ 6.5. Информация связи гауссовых величин 198
§ 6.6. Удельная информация стационарных и стационарно связанных процессов. Гауссовы процессы 205
§ 6.7. Информация связи компонент марковского процесса 212

Глава 7

ПЕРЕДАЧА СООБЩЕНИЙ ПРИ НАЛИЧИИ ПОМЕХ. ВТОРАЯ АСИМПТОТИЧЕСКАЯ ТЕОРЕМА В РАЗЛИЧНЫХ ФОРМУЛИРОВКАХ 226

§ 7.1. Принципы передачи и приема информации при наличии помех 227
§ 7.2. Случайный код и средняя вероятность ошибки 230
§ 7.3. Асимптотическая безошибочность декодирования. Теорема Шеннона (вторая асимптотическая теорема) 234
§ 7.4. Асимптотическая формула для вероятности ошибки 237
§ 7.5. Усиленные оценки для оптимального декодирования 241
§ 7.6. Некоторые общие соотношения между энтропиями и взаимными информациями при кодировании и декодировании . . . 252

Глава 8

ПРОПУСКНАЯ СПОСОБНОСТЬ КАНАЛОВ. ВАЖНЫЕ ЧАСТНЫЕ СЛУЧАИ КАНАЛОВ 257

§ 8.1. Определение пропускной способности каналов 257
§ 8.2. Решение второй экстремальной задачи. Соотношения для пропускной способности и потенциала 260
§ 8.3. Вид оптимального распределения и статистическая сумма . . 267
§ 8.4. Симметричные каналы 270
§ 8.5. Двоичные каналы 272
§ 8.6. Гауссовы каналы 275
§ 8.7. Стационарные гауссовы каналы 285
§ 8.8. Аддитивные каналы 291

Глава 9

ОПРЕДЕЛЕНИЕ ЦЕННОСТИ ИНФОРМАЦИИ 296

§ 9.1. Уменьшение средних штрафов при уменьшении неопределенности 297
§ 9.2. Ценность хартлиевского количества информации. Пример . . 301
§ 9.3. Определение ценности шенноновского количества информации и a-информации 307
§ 9.4. Решение третьей вариационной задачи. Соответствующие ей потенциалы 311
§ 9.5. Решение вариационной задачи при некоторых дополнительных предположениях 320
§ 9.6. Ценность больцмановского количества информации 325
§ 9.7. Другой подход к определению ценности шенноновской информации 329

Глава 10

ЦЕННОСТЬ ШЕННОНОВСКОЙ ИНФОРМАЦИИ ДЛЯ ВАЖНЕЙШИХ БЕЙЕСОВСКИХ СИСТЕМ 334

§ 10.1. Система с двумя состояниями 334
§ 10.2. Системы с однородной функцией штрафов 338
§ 10.3. Гауссовы бейесовские системы 344
§ 10.4. Стационарные гауссовы системы 353

Глава 11

АСИМПТОТИЧЕСКИЕ РЕЗУЛЬТАТЫ, КАСАЮЩИЕСЯ ЦЕННОСТИ ИНФОРМАЦИИ. ТРЕТЬЯ АСИМПТОТИЧЕСКАЯ ТЕОРЕМА 360

§ 11.1. *О различии между ценностями различных родов информации.* Предварительные формы 361
§ 11.2. Теорема об асимптотической равноценности различных количеств информации 365
§ 11.3. Быстрота исчезновения различия в ценности шенноновской и хартлиевской информации 376
§ 11.4. Другие способы записи основного результата. Обобщения и частные случаи 386
§ 11.5. Обобщенная теорема Шеннона 392

Глава 12

ТЕОРИЯ ИНФОРМАЦИИ И ВТОРОЙ ЗАКОН ТЕРМОДИНАМИКИ 398

§ 12.1. Информация о физической системе, находящейся в состоянии термодинамического равновесия. Обобщенный второй закон термодинамики 399
§ 12.2. Приток шенноновской информации и превращение теплоты в работу 402
§ 12.3. Энергетические затраты на создание и запись информации. Пример 407
§ 12.4. Энергетические затраты на создание и запись информации. Общая формулировка 410
§ 12.5. Энергетические затраты в физических каналах 413

Приложение

НЕКОТОРЫЕ МАТРИЧНЫЕ (ОПЕРАТОРНЫЕ) ТОЖДЕСТВА 416

§ П.1. Правило переноса оператора слева направо 416
§ П.2. Детерминант составной матрицы 417

Список литературы 418
Предметный указатель 420

Стратонович Р. Л.

С83 Теория информации М, «Сов радио», 1975

424 с с ил

Книга посвящена одному из главных направлении теоретической кибернетики Дается систематическое изложение важнейших, ставших уже традиционными, результатов шенноновской теории информации а также *ряда новых вопросов, разработанных* автором

Книга рассчитана на научных работников — специалистов в области кибернетики и статистической теории связи, а также аспирантов и студентов высших учебных заведении