

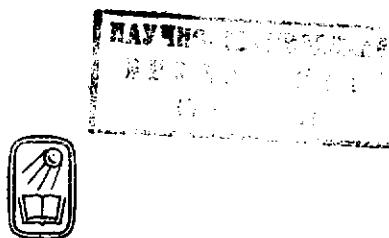
621.391
К 60

В.Д. КОЛЕСНИК
Г.Щ. ПОЛТЫРЕВ

КУРС ТЕОРИИ ИНФОРМАЦИИ

Допущено Министерством
высшего и среднего специального образования СССР
в качестве учебного пособия для студентов
высших технических учебных заведений

25.6.19



МОСКВА «НАУКА»
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
1982

32.81

К60

УДК 62-50

Курс теории информации. Колесник В. Д., Полтырев Г. Ш. — М.: Наука. Главная редакция физико-математической литературы. 1982. — 416 с.

Теория информации представляет собой ветвь статистической теории связи, круг проблем которой можно охарактеризовать как исследования кодирования для обработки и передачи сообщений. Книга состоит из следующих пяти разделов: кодирование дискретных источников, кодирование в дискретных каналах, кодирование в непрерывных каналах, кодирование непрерывных источников и кодирование в системах с многими пользователями. Основные параграфы книги задуманы как пособие для студентов, впервые знакомящихся с теорией информации. Дополнительные параграфы, отмеченные звездочкой, предназначены для углубленного изучения «традиционной» теории информации и могут быть полезны аспирантам. Особое место в книге занимает глава, посвященная кодированию в системах с многими пользователями, содержащая наиболее поздние результаты теории информации. Для чтения этой части нужна определенная теоретико-информационная эрудиция. Каждая глава снабжена рядом задач и упражнений.

Табл. 8, илл. 56, библ. 70 назв.

Виктор Дмитриевич Колесник, Григорий Шоулович Полтырев

Курс теории информации

Редактор Г. Л. Кацман

Техн. редактор И. Ш. Аксельрод

Корректоры О. А. Бутусова, Л. С. Сомова

ИБ № 11892

Сдано в набор 27.01.82. Подписано к печати 11.11.82. Т-20915.

Формат 60×90 $\frac{1}{16}$. Бумага типографская № 1. Гарнитура литературная.

Печать высокая. Усл. печ. л. 26. Уч.-изд. л. 28,99.

Тираж 14 000 экз. Заказ 119. Цена 1 р. 20 к.

Издательство «Наука»

**Главная редакция физико-математической литературы
117071, Москва, В-71, Ленинский проспект, 15**

**Ленинградская типография № 6 ордена Трудового Красного Знамени
Ленинградского объединения «Техническая книга» им. Евгении Соколовой
Союзполиграфпрома при Государственном комитете СССР
по делам издательств, полиграфии и книжной торговли.
193144, г. Ленинград, ул. Монсекро, 10.**

**К 1502000000—157
053(02)-82 119-82**

**© Издательство «Наука».
Главная редакция
физико-математической
литературы, 1982**

ОГЛАВЛЕНИЕ

Предисловие	6
Г л а в а 1. Кодирование дискретных источников	11
§ 1.1. Дискретные ансамбли и источники	11
§ 1.2. Случайные величины. Закон больших чисел	19
§ 1.3. Количество информации в сообщении. Энтропия	24
§ 1.4. Условная информация. Условная энтропия	27
§ 1.5. Энтропия на сообщение дискретного стационарного источника	31
§ 1.6. Постановка задачи кодирования дискретных источников равномерными кодами	34
§ 1.7. Теорема о высоковероятных множествах дискретного источника без памяти	39
§ 1.8. Скорость создания информации дискретным источником без памяти при равномерном кодировании	41
§ 1.9. Эргодические дискретные источники	44
§ 1.10. Постановка задачи неравномерного кодирования дискретных источников. Коды с однозначным декодированием	54
§ 1.11. Кодовые деревья. Неравенство Крафта	58
§ 1.12. Неравномерное кодирование дискретных стационарных источников	60
§ 1.13. Оптимальные неравномерные коды	64
§ 1.14. Обсуждение основных результатов	68
Задачи, упражнения и дополнения	72
Краткий исторический комментарий и литература	78
Г л а в а 2. Взаимная информация и ее свойства	80
§ 2.1. Количество информации между дискретными ансамблями	80
§ 2.2. Непрерывные ансамбли и источники. Обобщение понятия количества информации	88
§ 2.3. Относительная энтропия и ее свойства	100
§ 2.4*. Ортогональные преобразования случайных векторов	107
§ 2.5*. Выпукłość средней взаимной информации	112
§ 2.6*. Случайные процессы непрерывного времени	119
§ 2.7*. Средняя взаимная информация между случайными процессами	128
§ 2.8*. Поиск экстремумов	130
2.8.1. Метод неопределенных множителей Лагранжа (130). 2.8.2. Необходимые условия Куна—Таккера (132). 2.8.3. Достаточность условий Куна—Таккера для выпуклых функций (136). 2.8.4. Поиск экстремумов на множестве вероятностных векторов (137).	
Задачи, упражнения и дополнения	140
Краткий исторический комментарий и литература	151
Г л а в а 3. Кодирование в дискретных каналах	153
§ 3.1. Классификация каналов связи	153
§ 3.2. Постановка задачи кодирования в дискретном канале	158

§ 3.3.	Неравенство Фано	163
§ 3.4.	Общая обратная теорема кодирования для дискретных каналов	167
§ 3.5.	Информационная емкость дискретных каналов без памяти	169
	3.5.1. Упрощение формулы (3.4.3) (169). 3.5.2*. Вычисление информационной емкости дискретного канала без памяти (171).	
§ 3.6.	Симметричные дискретные каналы без памяти	179
§ 3.7.	Дискретные стационарные каналы с аддитивным по модулю L шумом	182
§ 3.8.	Неравенство Файнштейна	185
§ 3.9.	Прямая теорема кодирования для дискретных каналов без памяти	190
§ 3.10.	Прямая теорема кодирования для дискретных стационарных каналов с аддитивным эргодическим шумом	192
§ 3.11.	Декодирование для кодов с заданным множеством кодовых слов	194
§ 3.12.	Верхняя граница вероятности ошибки декодирования для дискретных каналов без памяти	197
	3.12.1. Метод случайного кодирования (197). 3.12.2. Оценка средней по ансамблю кодов вероятности ошибки декодирования для произвольного дискретного канала (197). 3.12.3. Оценка средней по ансамблю кодов вероятности ошибки декодирования для дискретных каналов без памяти (200). 3.12.4*. Свойства функции $E_0(p, Q)$ и построение экспоненты случайного кодирования (203). 3.12.5*. Показатель экспоненты случайного кодирования для симметричных каналов без памяти (211).	
§ 3.13*.	Нижняя граница вероятности ошибки декодирования для дискретных каналов без памяти (граница сферической упаковки)	215
	3.13.1. Нижняя граница вероятности ошибки для ДСК (215). 3.13.2. Коды с фиксированной композицией (219). 3.13.3. Нижняя граница для вероятности ошибки декодирования кода с фиксированной композицией (221). 3.13.4. Совместное рассмотрение экспонент случайного кодирования и сферической упаковки (227).	
	Задачи, упражнения и дополнения	234
	Краткий исторический комментарий и литература	246
Г л а в а 4. Кодирование в непрерывных каналах		248
§ 4.1.	Непрерывные каналы с дискретным временем. Обратная теорема кодирования	249
§ 4.2.	Непрерывные каналы без памяти с дискретным временем	254
§ 4.3*.	Каналы с непрерывным временем. Обратная теорема кодирования	261
§ 4.4*.	Прямая теорема кодирования для непрерывных каналов с аддитивным белым гауссовским шумом	270
	Задачи, упражнения и дополнения	275
	Краткий исторический комментарий и литература	281
Г л а в а 5. Кодирование источников с заданным критерием качества		282
§ 5.1.	Критерии качества. Постановка задачи кодирования с заданным критерием качества	283
§ 5.2.	Эпсилон-энтропия и ее свойства	290
§ 5.3.	Обратная теорема кодирования непрерывных источников при заданном критерию качества	295
§ 5.4.	Эпсилон-энтропия гауссовского источника без памяти	297

§ 5.5.	Прямая теорема кодирования стационарного источника независимых гауссовских сообщений при квадратическом критерии качества	300
	5.5.1. Закон больших чисел и принцип «затвердевания сферы» (300). 5.5.2. Аппроксимация векторов, лежащих на поверхности n -мерной сферы (302). 5.5.3. Аппроксимация последовательностей сообщений источника с помощью ε -сети на n -мерной сфере (304). 5.5.4. Прямая теорема кодирования (307). 5.5.5. Обсуждение (311).	
§ 5.6*.	Эпсилон-энтропия гауссовского случайного вектора	313
	5.6.1. Эпсилон-энтропия системы независимых гауссовских случайных величин (314). 5.6.2. Эпсилон-энтропия системы зависимых гауссовских случайных величин (317).	
§ 5.7*.	Эпсилон-энтропия стационарного гауссовского процесса дискретного времени	319
§ 5.8*.	Формулировка прямой теоремы кодирования для стационарного гауссовского источника с дискретным временем	323
	Задачи, упражнения, дополнения	324
	Краткий исторический комментарий и литература	332
Г л а в а 6*. Кодирование в системах с многими пользователями		333
§ 6.1.	Кодирование зависимых источников	335
	6.1.1. Постановка задачи (335). 6.1.2. Обратная теорема кодирования (338). 6.1.3. Прямая теорема кодирования (340).	
§ 6.2.	Кодирование источников с дополнительной информацией	345
	6.2.1. Постановка задачи (345). 6.2.2. Функция $T(d)$ и ее свойства (348). 6.2.3. Обратная теорема кодирования (354). 6.2.4. Прямая теорема кодирования (356).	
§ 6.3.	Кодирование в каналах с множественным доступом	364
	6.3.1. Постановка задачи (364). 6.3.2. Двоичный суммирующий КМД (367). 6.3.3. Обратная теорема кодирования (370). 6.3.4. Прямая теорема кодирования (374).	
§ 6.4.	Кодирование в широковещательных каналах	378
	6.4.1. Постановка задачи (378). 6.4.2. Ухудшающиеся широковещательные каналы (УШК) (382). 6.4.3. Двоичный симметричный широковещательный канал (384). 6.4.4. Обратная теорема кодирования (387). 6.4.5. Прямая теорема кодирования (396).	
	Задачи, упражнения и дополнения	401
	Краткий исторический комментарий и литература	409
Приложение I		411
Приложение II		412
Предметный указатель		414

Теория информации представляет собой ветвь статистической теории связи (ее часто с нею отождествляют), основы которой были заложены классическими трудами Н. Винера, А. Н. Колмогорова, В. А. Котельникова и К. Шеннона. Круг проблем, составляющих основное содержание теории информации (проблем «шенноновской теории информации»), можно охарактеризовать как исследование методов кодирования для экономного представления сообщений различных источников и для надежной передачи сообщений по каналам связи с шумом.

В основе теории информации лежит статистическое описание (статистические модели) источников сообщений и каналов связи, а также основанное на этом описании измерение количества информации между сообщениями по Шеннону, т. е. такое, при котором количество информации определяется только вероятностными свойствами сообщений и ни от каких других их свойств не зависит. В отличие от других разделов теории связи, например, теории обнаружения, теории оценивания, теории модуляции, алгебраической теории кодирования и т. д., предметом теории информации, как правило, являются теоремы, устанавливающие предельные возможности различных методов обработки и передачи сообщений. Эти предельные возможности зависят только от статистических свойств источников и каналов.

В качестве примеров можно привести три типичные задачи теории информации.

1. Предположим, что задан источник сообщений. Требуется найти наименьшее количество символов (например, двоичных), которое необходимо для указания последовательности сообщений, порожденных источником. При этом может быть задан критерий качества восстановления сообщений источника и требоваться указание последовательности сообщений с ошибкой относительно данного критерия качества, не превосходящей заданную величину.

2. Предположим, что задан канал связи. Требуется найти наибольшую возможную скорость передачи по этому каналу, при которой вероятность ошибочного приема сообщений может быть сделана произвольно малой.

3. Предположим, что заданы источник и канал, а также задан критерий качества. Требуется определить наименьшую возмож-

ную относительно данного критерия величину ошибки, которую можно достичь, передавая сообщения данного источника по данному каналу связи.

Всякий раз, решая подобные задачи, пытаются⁷ не только найти предельные значения количества двоичных символов, скорости передачи или величины ошибки, но и найти некоторый способ обработки сообщений (некоторый способ кодирования и декодирования), который позволяет достичь указываемых пределов. Однако очень часто не удается указать наилучший способ кодирования и декодирования. Поэтому теория информации, как правило, не дает непосредственных практических рекомендаций инженерам, проектирующим аппаратуру обработки и передачи сообщений. Тем не менее, она является важным инструментом анализа различных технических систем: телеметрических систем, систем передачи речи или телевизионных изображений, систем передачи данных, банков данных, различных систем управления и т. д.

На основе теории информации можно ответить на вопросы о предельных возможностях перечисленных систем, определить, в какой мере проектируемая система уступает теоретически возможной. Следует отметить также, что в некоторых случаях логика вывода, используемая в теории информации, подсказывает путь, на котором может быть найдено конструктивное решение для данной реальной системы.

Первоначально теория информации возникла из инженерных задач радиосвязи и телеграфии. Датой ее рождения считают 1948 год, год появления двух основополагающих статей американского инженера и математика Клода Шеннона «Математическая теория связи» и «Связь при наличии шума» (см.: Шеннон К., Сборник работ по теории информации и кибернетике. — М.: ИЛ, 1963). Начиная с этого времени, теория информации бурно развивалась, главным образом благодаря работам математиков и математически образованных инженеров. Нельзя не отметить огромный вклад, который внесли в теорию информации Дж. Вольфович, Р. Галлагер, Р. Л. Добрушин, А. Н. Колмогоров, М. С. Пинскер, В. И. Сифоров, А. Файнштейн, Р. Фано, А. А. Харкевич, А. Я. Хинчин и многие другие. В результате развития теории информации основная часть теоретических работ стала носить математически сложный характер, и образовался определенный разрыв между инженерами-практиками и адресованной в первую очередь им прикладной математической теорией. К сожалению, до настоящего времени этот разрыв не имеет тенденции уменьшаться; желание его преодолеть было одним из основных стимулов при написании этой книги, которая по замыслу авторов должна помочь студенту технического вуза познакомиться с теорией информации или ее отдельными разделами.

Сегодня можно указать две основные книги, которые могут служить учебниками по теории информации. Это книга Р. Фано «Передача информации. Статистическая теория связи» (Мир, 1965) и книга Р. Галлагера «Теория информации и надежная связь» (Сов. радио, 1974). Обе эти книги написаны известными американскими учеными, внесшими существенный вклад в теорию информации. Однако обе они в значительной степени носят монографический характер и предназначены достаточно подготовленным читателям. Следует также отметить одну из первых книг на русском языке — книгу Ф. П. Тарасенко «Введение в курс теории информации» (изд. Томского ун-та, 1963) и книгу Р. Л. Стратоновича «Теория информации» (Сов. радио, 1975), посвященную нетрадиционному изложению щеняновской теории информации с позиций статистической термодинамики.

Настоящая книга состоит из следующих пяти основных частей: кодирование дискретных источников, кодирование в дискретных каналах, кодирование в непрерывных каналах, кодирование непрерывных источников и кодирование в системах с многими пользователями.

В первой главе рассматривается задача точного или сколь угодно точного кодирования дискретных источников. В ней дается ответ на вопрос, каково наименьшее количество двоичных символов на сообщение, по которому можно точно или с какой угодно малой вероятностью ошибки восстановить последовательность сообщений на выходе дискретного источника.

Во второй главе задачи кодирования не рассматриваются. Она посвящена исследованию свойств количества информации для различных вероятностных объектов. Кроме того, в этой главе приводятся математические сведения, необходимые для чтения этой и последующих глав книги.

В третьей главе рассматривается задача кодирования в дискретных каналах связи и дается ответ на вопрос, каково наибольшее количество информационных двоичных символов, которое может быть передано по каналу связи в единицу времени при условии, что вероятность ошибки при определении переданных сообщений может быть сделана сколь угодно малой величиной. Кроме того, в этой главе строятся верхняя и нижняя границы вероятности ошибки декодирования для дискретных каналов без памяти.

Четвертая глава посвящена обобщению результатов третьей главы на случай различных непрерывных каналов.

В пятой главе рассматривается задача кодирования источников при заданном критерии качества (теория эпсилон-энтропии). В ней дается ответ на вопрос, каково наименьшее количество двоичных символов на сообщение, по которым можно восстановить с заданной ошибкой относительно выбранного критерия качества

последовательность сообщений на выходе некоторого источника.

Шестая глава посвящена задачам кодирования в системах с многими пользователями (многими источниками, каналами связи и получателями сообщений). Здесь также даются ответы на вопросы о минимальном числе двоичных символов на сообщение источника или о максимальном числе информационных двоичных символов, передаваемых по каналу в единицу времени, в ситуации, когда имеется несколько источников и они зависят или когда имеется несколько каналов и передача по одному каналу мешает передаче по другим.

В книге имеются основные и дополнительные параграфы. Основные задуманы как пособие для читателя, который впервые знакомится с теорией информации и который не рискнул бы считать себя хорошо владеющим теорией вероятностей. Тем не менее, эта часть книги позволяет читателю проникнуть в проблематику теории информации и познакомиться с ее основными результатами, пройдя через все трудности доказательств. Следующие параграфы являются основными: вся гл. I, §§ 2.1—2.3, вся гл. III (кроме пп. 3.5.2, 3.12.4, 3.12.5 и § 3.13), §§ 4.1, 4.2, 5.1—5.5. Эти разделы могут составить материал для односеместрового курса лекций по теории информации. Курсы лекций примерно с таким содержанием читаются в течение ряда лет в Ленинградском институте авиационного приборостроения. При рассмотрении основных разделов сделана попытка упростить изложение за счет сужения круга рассматриваемых вопросов, использования в связи с этим более простой математической техники, более подробного обсуждения постановок задач и примеров. Кроме того, здесь даются все необходимые математические сведения, что делает эту часть книги в определенной мере самостоятельной.

Дополнительных параграфов несколько (в книге они помечены звездочкой над номером параграфа). Все, что не вошло в перечисленные выше основные параграфы (в пределах первых пяти глав), можно рассматривать как дополнительный материал, предназначенный для углубленного изучения «традиционной» теории информации. Хотя многие математические сведения здесь также приводятся, предполагается, что читатель знаком с элементами функционального анализа, с элементами теории случайных процессов и с элементами нелинейного программирования. Кроме того, для чтения этих разделов требуется несколько большая математическая тренированность.

Особое место в книге занимает шестая глава, посвященная задачам кодирования в системах с многими пользователями. В ней представлены результаты, полученные в теории информации в течение последнего десятилетия и отражающие развитие современных систем обработки и передачи информации. Содержание

шестой главы адресовано в первую очередь читателям, хорошо знакомым с традиционными вопросами теории информации, например, по книге Р. Галлагера, или хорошо овладевшим содержанием первых пяти глав настоящей книги. Для успешного чтения этой главы от читателя требуется наличие определенной теоретико-информационной эрудиции.

Каждая глава снабжена рядом задач, упражнений и дополнений. Среди задач и упражнений имеются очень простые, предназначенные только для проверки того, что читатель правильно понял формулировки определений и теорем. Имеются задачи, которые требуют овладения техникой доказательств. В ряде случаев приводятся дополнительные сведения, затрагивающие как методы, так и интересные и важные результаты теории информации, которые по тем или другим соображениям не включены в основной текст, но которые могут быть полезны при более глубоком изучении теории или при использовании теории на практике.

В заключение отметим, что хотя в книге содержатся основные математические сведения, которые используются при изложении рассматриваемых теоретико-информационных задач, их явно недостаточно для глубокого понимания математических основ теории информации. Читателю, желающему более детально познакомиться с этими основами, мы рекомендуем обратиться к следующим книгам. С теорией вероятностей лучше знакомиться по книгам Б. В. Гнеденко «Курс теории вероятностей» (Наука, 1965) и В. Феллера «Введение в теорию вероятностей и ее приложения», т. I (Мир, 1967). С элементами теории случайных процессов можно познакомиться по книге В. Б. Давенпорта и В. Л. Рута «Введение в теорию случайных сигналов и шумов» (ИЛ, 1960). Основы матричной алгебры и теория операторов в конечномерных пространствах лучше всего изложены в книгах Ф. Р. Гантмахера «Теория матриц» (Наука, 1967) или Р. Беллмана «Введение в теорию матриц» (Наука, 1969). По книге Б. З. Вулиха «Введение в функциональный анализ» (Наука, 1967) можно познакомиться с элементами функционального анализа.

Авторы выражают огромную признательность всем, кто знакомился с многочисленными вариантами рукописи этой книги и делился своими замечаниями. Особенно большое влияние на работу авторов оказали замечания и советы Ю. М. Штарккова и рецензентов — Р. Л. Добрушина и Э. М. Габидулина.

Г л а в а 1

КОДИРОВАНИЕ ДИСКРЕТНЫХ ИСТОЧНИКОВ

В этой главе будут даны основные определения: дискретного вероятностного ансамбля, дискретного источника, дискретной случайной величины на ансамбле и кода для дискретного источника. Дискретные источники представляют собой наиболее простой объект теории информации. Начинать именно с этого типа источников удобно не только потому, что здесь требуется наименьшее количество определений и вспомогательных результатов, но и потому, что на этом простом объекте можно показать методологию теории информации и продемонстрировать ее основные технические приемы.

Задача, которая рассматривается в этой главе, весьма часто встречается на практике и иногда называется задачей сжатия данных. Предположим, что некоторый источник порождает последовательность дискретных сообщений и требуется представить эту последовательность с помощью некоторых символов, скажем, с помощью нулей и единиц. Не вызывает сомнения то, что это можно сделать для любой последовательности сообщений. Вопрос может заключаться в том, как это сделать наиболее экономным образом, т. е. как затратить на это наименьшее количество двоичных символов. Ответ на этот вопрос лежит в изучении различных статистических моделей источников и определении для этих моделей величины, называемой скоростью создания информации. Будет показано, что скорость создания информации равна энтропии источника на сообщение, величине, которая определяется с помощью вводимого в этой главе понятия количества информации в сообщении.

§ 1.1. Дискретные ансамбли и источники

Основным объектом изучения в этой главе будут дискретные источники сообщений. Здесь мы дадим определения, необходимые для описания математических моделей источников.

Описание источников удобно начать с определения дискретных вероятностных ансамблей. Пусть $X = \{x_1, \dots, x_M\}$ — множество, состоящее из M элементов. Прописные латинские буквы X , Y и т. д. будут обозначать сами множества, а соответствующие строчные буквы x , y и т. д. будут обозначать элементы множеств.

Иногда элементы множеств мы будем снабжать подстрочными индексами, как это сделано выше. Такой индекс представляет собой номер элемента в множестве. Хотя для большинства случаев природа элементов несущественна, мы будем называть элементы множеств X сообщениями, подчеркивая тем самым область применения теории.

Говорят, что на конечном множестве X задано распределение вероятностей $p(x)$, если каждому элементу $x_i \in X$ сопоставлено число $p(x_i)$, причем

$$\begin{aligned} p(x_i) &\geq 0, \quad i = 1, 2, \dots, M, \\ \sum_{i=1}^M p(x_i) &= 1. \end{aligned} \quad (1.1.1)$$

Пусть A есть подмножество множества X , $A \subseteq X$. Число *)

$$\Pr(A) \triangleq \sum_{x_i \in A} p(x_i)$$

представляет собой вероятность того, что при случайном выборе сообщения из множества X в соответствии с распределением $p(x)$, будет выбрано сообщение, принадлежащее множеству A . Число $\Pr(A)$ называют также вероятностью множества A .

Пример 1.1.1. Пусть X — множество сообщений о результатах бросания правильной игральной кости. Тогда $X = \{x_1, \dots, x_6\}$, $p(x_i) = 1/6$, $i = 1, \dots, 6$, причем x_i есть сообщение о том, что выпало i очков. Если $A = \{x_2, x_4, x_6\}$, то $\Pr(A) = 3 \cdot 1/6 = 1/2$ есть вероятность того, что при бросании кости выпало четное число очков.

Сообщения $x_i \in X$ иногда называют элементарными событиями. Как показывает предыдущая формула, могут одновременно рассматриваться как элементарные события, так и более сложные события, являющиеся объединением некоторого числа элементарных. Мы используем различные обозначения для вероятностей таких событий: $p(x_i)$ — для элементарных событий и $\Pr(A)$ — для множества A , образованного элементарными событиями $x_i \in A$. Это различие не принципиально, но делает некоторые формулы наглядными.

Определение 1.1.1. Конечное множество X вместе с заданным на нем распределением вероятностей $p(x)$ называется *дискретным вероятностным ансамблем* или коротко — *дискретным ансамблем* (сообщений) и обозначается символом $\{X, p(x)\}$. В тех случаях, когда из контекста видно, о каком распределении вероят-

*) Здесь и ниже знак \triangleq используется для обозначения того, что правая и левая части равны по определению. Иногда для упрощения обозначений мы будем писать $\Pr(A) = \sum_A p(x_i)$.

ностей идет речь или когда точное описание распределения несущественно, мы будем обозначать ансамбль через X .

Пусть $X = \{x_1, \dots, x_M\}$ и $Y = \{y_1, \dots, y_N\}$ — два конечных множества. Элементы, которого представляют собой все возможные упорядоченные пары (x_i, y_j) , $x_i \in X$, $y_j \in Y$, $i = 1, \dots, M$, $j = 1, \dots, N$, называется *произведением множеств X и Y* и обозначается через XY . Согласно этому определению XY и YX суть различные множества. Если множества X и Y совпадают, $X = Y$, то произведение XY обозначается как X^2 . Аналогичным образом определяются произведения более чем двух множеств. Произведение $X_1 X_2 \dots X_n$ представляет собой множество всех последовательностей $(x^{(1)}, x^{(2)}, \dots, x^{(n)})$ длины n таких, что первый элемент $x^{(1)}$ принадлежит множеству X_1 , второй $x^{(2)}$ — множеству X_2 и т. д., n -й элемент принадлежит множеству X_n *). Если все множества совпадают между собой и с множеством X , то такое произведение обозначается как X^n . Таким образом, X^n — это множество всех последовательностей длины n , образованных из элементов множества X .

Пусть XY есть произведение двух конечных множеств X и Y и на множестве XY задано совместное распределение вероятностей $p(x, y)$, которое каждой паре (x_i, y_j) , $x_i \in X$, $y_j \in Y$, сопоставляет вероятность $p(x_i, y_j)$. Очевидно, что соотношения

$$p_1(x_i) \triangleq \sum_{y_j \in Y} p(x_i, y_j), \quad i = 1, 2, \dots, M, \quad (1.1.2)$$

$$p_2(y_j) \triangleq \sum_{x_i \in X} p(x_i, y_j), \quad j = 1, 2, \dots, N, \quad (1.1.3)$$

задают распределения вероятностей $p_1(x)$ и $p_2(y)$ на множествах X и Y соответственно. Таким образом, при задании ансамбля $\{XY, p(x, y)\}$ фактически задаются еще два ансамбля $\{X, p_1(x)\}$ и $\{Y, p_2(y)\}$. Иногда, имея в виду ансамбль $\{XY, p(x, y)\}$, мы будем говорить, что *совместно заданы* два ансамбля X и Y . Это будет означать, что распределения вероятностей на множествах X и Y определяются по формулам (1.1.2) и (1.1.3), исходя из распределения вероятностей $p(x, y)$ на множестве XY .

Если распределение вероятностей на произведении двух множеств X и Y удовлетворяет условию

$$p(x_i, y_j) = p_1(x_i) p_2(y_j) \text{ для всех } x_i \in X, y_j \in Y, \quad (1.1.4)$$

то ансамбли X и Y называются *статистически независимыми*. В противном случае говорят, что эти ансамбли *статистически зависимы*.

*) Здесь и в аналогичных обозначениях дальше надстрочный индекс обозначает номер элемента в последовательности, другими словами, $x^{(i)}$ — элемент, расположенный на i -м месте последовательности.

Пусть задан ансамбль $\{XY, p(x, y)\}$, предположим, что x_i — такой элемент множества X , для которого $p_1(x_i) \neq 0$. Число

$$p(y_j | x_i) \triangleq \frac{p(x_i, y_j)}{p_1(x_i)} \quad (1.1.5)$$

называется *условной вероятностью сообщения y_j при условии, что сообщение x_i известно* (иногда это число называют *условной вероятностью сообщения y_j относительно сообщения x_i*). Легко увидеть, что множество условных вероятностей относительно фиксированного сообщения x_i , которое получается, когда индекс j пробегает все возможные значения, удовлетворяет определению распределения вероятностей (1.1.1). Такое распределение называется *условным распределением на множестве Y относительно фиксированного сообщения x_i* ; заметим, что (1.1.3) определяет так называемое *безусловное распределение на Y* . Понятно, что аналогичные распределения могут быть определены также на множестве X . Таким образом, задание ансамбля $\{XY, p(x, y)\}$ определяет также условные ансамбли $\{X, p(x | y)\}$, $p_2(y) \neq 0$, и $\{Y, p(y | x)\}$, $p_1(x) \neq 0$.

Опишем теперь общее семейство условных ансамблей, которые образуются при совместном задании двух ансамблей. Пусть A — произвольное подмножество элементов из X такое, что $\Pr_1(A) \triangleq \sum_{x \in A} p_1(x) \neq 0$, где распределение $p_1(x)$ определено выше. Число

$$p(y_j | A) \triangleq \frac{1}{\Pr_1(A)} \sum_{x_i \in A} p(x_i, y_j) \quad (1.1.6)$$

называется *условной вероятностью сообщения y_j при условии, что сообщение x_i принадлежит множеству A (или условной вероятностью относительно множества A)*. Легко видеть, что множество условных вероятностей относительно фиксированного множества A , которое получается, когда индекс j пробегает все возможные значения, снова удовлетворяет определению распределения вероятностей. Такое распределение называется *условным распределением на множестве Y относительно фиксированного множества A* . Если выбрать $A = X$, то $p(y_j | A) = p_2(y_j)$. Таким образом, условное распределение относительно множества X — это просто безусловное распределение на Y . Если A состоит из одного элемента, скажем x_i , то $p(y_j | A)$ равно вероятности, определенной в (1.1.5).

Аналогичные условные распределения могут быть определены также для множества X . Тем самым определены условные ансамбли $\{X, p(x | B)\}$, $B \subseteq Y$, $\Pr_2(B) \neq 0$ и $\{Y, p(y | A)\}$, $A \subseteq X$, $\Pr_1(A) \neq 0$.

Рассмотрим произведение n множеств $X_1 \dots X_n$ и распределение вероятностей $p(x^{(1)}, \dots, x^{(n)})$, $x^{(i)} \in X_i$, $i = 1, \dots, n$, задан-

ное на этом произведении. Другими словами, рассмотрим вероятностный ансамбль $\{X_1 \dots X_n, p(x^{(1)}, \dots, x^{(n)})\}$. Пусть

$$\begin{aligned} p_1(x^{(1)}) &\triangleq \sum_{X_2} \sum_{X_3} \dots \sum_{X_n} p(x^{(1)}, \dots, x^{(n)}), \\ p_2(x^{(2)}) &\triangleq \sum_{X_1} \sum_{X_3} \dots \sum_{X_n} p(x^{(1)}, \dots, x^{(n)}), \\ &\dots \\ p_n(x^{(n)}) &\triangleq \sum_{X_1} \sum_{X_2} \dots \sum_{X_{n-1}} p(x^{(1)}, \dots, x^{(n)}). \end{aligned} \quad (1.1.7)$$

Соотношения (1.1.7) задают безусловные распределения вероятностей на множествах X_1, X_2, \dots, X_n соответственно. Если для любых $x^{(1)} \in X_1, \dots, x^{(n)} \in X_n$ имеет место равенство

$$p(x^{(1)}, \dots, x^{(n)}) = p_1(x^{(1)}) \dots p_n(x^{(n)}), \quad (1.1.8)$$

то ансамбли X_1, \dots, X_n называют *статистически независимыми*.

При совместном задании n ансамблей X_1, \dots, X_n оказываются совместно заданными всевозможные совокупности по $m \leq n$ ансамблей. Так, ансамбль $\{X_{i_1} \dots X_{i_m}, p(x^{(i_1)}, \dots, x^{(i_m)})\}$ определяется с помощью следующего соотношения, которое дает безусловное распределение вероятностей на произведении множеств $X_{i_1} \dots X_{i_m}$:

$$p(x^{(i_1)}, \dots, x^{(i_m)}) \triangleq \sum_{X_{j_1}} \dots \sum_{X_{j_{n-m}}} p(x^{(1)}, \dots, x^{(n)}), \quad (1.1.9)$$

где суммирование производится по всем множествам $X_{j_1}, \dots, X_{j_{n-m}}$ таким, что произведение соответствующим образом упорядоченных множеств $X_{i_1}, \dots, X_{i_m}, X_{j_1}, \dots, X_{j_{n-m}}$ есть множество $X_1 \dots X_n (\{i_1, \dots, i_m\} \cap \{j_1, \dots, j_{n-m}\} = \emptyset, \{i_1, \dots, i_m\} \cup \{j_1, \dots, j_{n-m}\} = \{1, 2, \dots, n\})$. Другими словами, множество X_j участвует в суммировании в (1.1.9), если j не содержится среди чисел i_1, \dots, i_m .

Вводя по аналогии с (1.1.5) и (1.1.6) условные распределения на множестве X_{i_1}, \dots, X_{i_m} , можно получить различные условные ансамбли.

Пример 1.1.2. Пусть $\{X, p_1(x)\}$ — ансамбль сообщений из примера 1.1.1, а $\{Y, p_2(y)\}$ — ансамбль сообщений о результате бросания монеты, $Y = \{y_1, y_2\}$ и $p_2(y_1) = p_2(y_2) = \frac{1}{2}$. Элементы множества XY (произведения множеств X и Y) представляют собой пары сообщений (x_i, y_j) , причем первый элемент пары есть сообщение x_i о результате бросания кости, а второй элемент y_j есть сообщение о результате бросания монеты. В случае независимого бросания кости и монеты все пары имеют одинаковые вероятности: $p(x_i, y_j) = \frac{1}{12}$ для всех i, j .

Можно представить себе более сложный эксперимент с костью и монетой. Предположим, что вначале бросается кость. Если выпало четное число очков, бросается неправильная монета, у которой вероятность выпадения стороны,

соответствующей y_1 , равна $1/4$ (а стороны, соответствующей y_2 , — $3/4$). Если же выпало нечетное число очков, то бросается другая неправильная монета, у которой вероятность выпадения стороны, соответствующей y_1 , равна $3/4$ (а стороны, соответствующей y_2 , — $1/4$). При таком эксперименте на множестве Y имеются два условных распределения: $p(y_1|x) = 1/4$, $p(y_2|x) = 3/4$ при четных $x \in X$ и $p(y_1|x) = 3/4$, $p(y_2|x) = 1/4$ при нечетных $x \in X$. Распределение вероятностей на множестве XY указано в следующей таблице:

$p(x_i, y_j)$	x_1	x_2	x_3	x_4	x_5	x_6
y_1	1/8	1/24	1/8	1/24	1/8	1/24
y_2	1/24	1/8	1/24	1/8	1/24	1/8

Нетрудно видеть, что безусловные распределения в этом случае такие же, как и в случае одной симметричной (правильной) монеты, но ансамбли X и Y статистически независимыми не являются.

Пример 1.1.3. Рассмотрим n последовательных бросаний некоторой монеты. Обозначим через Y_i множество сообщений о результате i -го бросания, $i = 1, 2, \dots, n$, $Y_i = \{y_1, y_2\}$. Множество $Y^n \triangleq \prod_{i=1}^n Y_i$ содержит 2^n последовательностей длины n , являющихся сообщениями о результатах n -кратного бросания монеты. Если положить $y_1 = 0$ и $y_2 = 1$, то множество Y^n будет состоять из всех двоичных последовательностей длины n . Рассмотрим теперь два случая, соответствующие независимым и зависимым бросаниям.

а) Предположим, что бросается правильная монета. Тогда все возможные исходы имеют одинаковые вероятности, равные 2^{-n} . При каждом бросании сообщения y_1 и y_2 могут появиться с одинаковыми вероятностями. Поэтому для любой последовательности $(y^{(1)}, \dots, y^{(n)})$

$$2^{-n} = p(y^{(1)}, \dots, y^{(n)}) = \prod_{i=1}^n p(y^{(i)}).$$

Следовательно, подбрасывание правильной монеты соответствует последовательности независимых испытаний.

б) Теперь предположим, что имеются две неправильные монеты, описанные в примере 1.1.2. Сначала наудачу выбирается одна из двух монет. Если результат бросания есть y_1 , то для следующего бросания выбирается монета с $p(y_1) = 3/4$ и другая монета в противном случае. Затем в зависимости от результата второго подбрасывания аналогичным образом выбирается очередная монета и этот процесс повторяется. Соответствующее такому эксперименту распределение вероятностей задается следующим образом:

$$p(y^{(1)}, \dots, y^{(n)}) \triangleq p(y^{(1)}) p(y^{(2)} | y^{(1)}) \dots p(y^{(n)} | y^{(n-1)}),$$

где

$$\begin{aligned} p(y^{(1)} = y_1 | y^{(i-1)} = y_1) &= y_1 = 3/4, & p(y^{(1)} = y_2 | y^{(i-1)} = y_1) &= 1/4, \\ p(y^{(1)} = y_1 | y^{(i-1)} = y_2) &= 1/4, & p(y^{(1)} = y_2 | y^{(i-1)} = y_2) &= 3/4. \end{aligned}$$

Распределение вероятностей для каждого очередного бросания зависит только от результата одного предыдущего бросания. Соответствующая случайная последовательность $y^{(1)}, y^{(2)}, \dots$, называется *марковской*.

Перейдем теперь к описанию дискретных источников. Под дискретным источником понимают некоторое устройство, которое в каждую единицу времени (например, каждую секунду) выбирает одно из сообщений ^{коин} множества X . Как правило, это множество одно и то же для каждого момента времени, хотя в некоторых случаях для каждого момента времени может быть свое множество сообщений.

С точки зрения теории информации источник считается заданным полностью, если имеется некоторая вероятностная схема (модель), позволяющая вычислить вероятность любого отрезка сообщений. Например, недостаточно сказать, что в качестве источника сообщений рассматривается телеграфный аппарат или передающее телевизионное устройство. Недостаточно также сказать, что известны вероятности букв, появляющихся на выходе телеграфного аппарата, или вероятности импульсов различной амплитуды на выходе телевизионного устройства. Для полного задания источника необходимо дать вероятностное описание процесса появления сообщений на выходе источника. В примере с телеграфным аппаратом нужно дать такое описание, при котором можно вычислить вероятность любого буквенного сочетания, слова, предложения и т. д. в любой момент времени. Таким образом, если источник задан, то для любых n и i и любой последовательности сообщений $(x^{(i+1)}, \dots, x^{(i+n)})$ определена вероятность $p(x^{(i+1)}, \dots, x^{(i+n)})$ этой последовательности. Верно и обратное, если для любых n и i определены вероятности всех последовательностей сообщений длины n , начинающихся с позиции $(i+1)$, то говорят, что задан источник сообщений. Отсюда следует, что все источники, которые могут иметь совершенно различную физическую природу, задаваемые одним и тем же набором вероятностей последовательностей сообщений, с позиции теории информации отождествляются. Подытожим и уточним сказанное с помощью следующего определения.

Определение 1.1.2. Пусть U_X — дискретный источник, выбирающий сообщения из множества X . Будем говорить, что источник U_X задан, если для любых $n = 1, 2, \dots$, любых $i = 0, \pm 1, \pm 2, \dots$ задано семейство распределений вероятностей $\{p(x^{(i+1)}, \dots, x^{(i+n)})\}$, $x^{(j)} \in X$, $j = i+1, \dots, i+n$, удовлетворяющих условию *согласованности*, состоящему в том, что распределение вероятностей $p(x^{(i_1)}, \dots, x^{(i_m)})$ для любого набора позиций i_1, \dots, i_m определено однозначным образом.

Замечание. Распределение вероятности $p(x^{(i_1)}, \dots, x^{(i_m)})$ на подпоследовательностях $(x^{(i_1)}, \dots, x^{(i_m)})$ может быть получено

с помощью соотношения (1.1.9), причем многими способами. Например,

$$p(x^{(2)}, x^{(3)}) = \sum_{X_1} p(x^{(1)}, x^{(2)}, x^{(3)}),$$

$$p(x^{(2)}, x^{(3)}) = \sum_{X_4} p(x^{(2)}, x^{(3)}, x^{(4)}).$$

Условие согласованности обеспечивает совпадение всех этих распределений.

В определении 1.1.2 легко усмотреть определение случайного процесса дискретного времени, который в каждый момент времени принимает значение из множества X . Таким образом, определение источника и определение случайного процесса, который генерирует источник, совпадают. Как источник, так и случайный процесс, порождаемый источником, задается своими n -мерными распределениями, $n = 1, 2, \dots$. Всякое n -мерное распределение вероятностей в общем случае является функцией $2n$ переменных, $p(x^{(i_1)}, \dots, x^{(i_n)}; i_1, \dots, i_n)$. В случае дискретных источников и дискретных случайных процессов значение этой функции есть вероятность того, что в фиксированные моменты времени i_1, \dots, i_n источник породит сообщения (или процесс примет значения) $x^{(i_1)}, \dots, x^{(i_n)}$ соответственно.

В общем случае вероятность некоторого отрезка сообщений зависит как от самого отрезка, так и от его расположения на оси времени. Имеется, однако, важный класс источников, обладающих однородностью во времени или стационарностью. Свойство *стационарности* состоит в том, что для любого целого числа j вероятности двух одинаковых последовательностей, одна из которых занимает временные позиции i_1, \dots, i_n , а другая — временные позиции $i_1 + j, \dots, i_n + j$, равны. Другими словами,

$$\begin{aligned} p(x^{(i_1+j)}, \dots, x^{(i_n+j)}; i_1 + j, \dots, i_n + j) &= \\ &= p(x^{(i_1)}, \dots, x^{(i_n)}; i_1, \dots, i_n) \end{aligned} \quad (1.1.10)$$

при $(x^{(i_1+j)}, \dots, x^{(i_n+j)}) = (x^{(i_1)}, \dots, x^{(i_n)})$. Всюду в этой книге мы используем сокращенную форму записи $p(x^{(i+1)}, \dots, x^{(i+n)})$ вместо $p(x^{(i+1)}, \dots, x^{(i+n)}; i+1, \dots, i+n)$.

Определение 1.1.3. Дискретный источник U_X называется *стационарным*, если сообщения на его выходе образуют стационарный случайный процесс.

В случае стационарных источников распределение вероятностей не зависит от сдвига по оси времени. Все последовательности, отличающиеся только положением на оси времени, имеют одинаковые вероятности. Поэтому их положение на оси времени можно не оговаривать.

Определение 1.1.4. Дискретный источник называется *источником без памяти*, если для любых $n = 1, 2, \dots$, любых $i = 0, \pm 1, \pm 2, \dots$ и любых последовательностей $(x^{(i+1)}, \dots, x^{(i+n)})$, $x^{(j)} \in X$, имеет место равенство

$$p(x^{(i+1)}, \dots, x^{(i+n)}) = \prod_{j=1}^n p_{i+j}(x^{(i+j)}). \quad (1.1.11)$$

В общем случае распределение вероятностей для сообщений на выходе источника в момент времени i зависит от i . Эта зависимость показана с помощью нижнего индекса у сомножителей в правой части соотношения (1.1.11). Однако в случае стационарных источников, как это следует из (1.1.10), все одномерные распределения ($n = 1$) одинаковы для всех моментов времени. Поэтому для стационарных источников без памяти

$$p(x^{(i+1)}, \dots, x^{(i+n)}) = \prod_{j=1}^n p(x^{(i+j)}), \quad (1.1.12)$$

где через $p(\cdot)$ обозначено общее для всех моментов времени одномерное распределение.

Всюду в этой книге мы будем рассматривать только стационарные источники. Поэтому вместо длинного сочетания слов «стационарный источник без памяти» будет употребляться более короткое выражение «источник без памяти». В этом случае стационарность будет подразумеваться. В случае стационарных источников с памятью, для которых соотношение (1.1.12) не выполняется, мы будем использовать термин «стационарный источник». Заметим, что стационарные источники без памяти иногда называются *постоянными*.

Пример 1.1.4. Пусть X и Y — ансамбли примера 1.1.2. Рассмотрим источник, выбирающий сообщения в четные моменты времени из множества X , а в нечетные — из Y . Такой источник, очевидно, нестационарен. Пусть теперь источник выбирает сообщения из ансамбля Y примера 1.1.3. Если распределение вероятностей будет такое, как в п. б), то источник будет стационарным, но с памятью. Если же распределение вероятностей такое, как в п. а), то источник будет источником без памяти.

§ 1.2. Случайные величины. Закон больших чисел

Предположим, что $\{X, p(x)\}$ — дискретный ансамбль и $\phi(x)$ — функция, определенная на $X = \{x_1, \dots, x_M\}$ и принимающая значения на числовой оси. Элементы множества X могут иметь произвольную природу, однако $\phi(x_1), \dots, \phi(x_M)$ — числа. Всякая действительная функция $\phi(x)$, заданная на произвольном дискретном ансамбле, порождает *действительную дискретную случайную величину* (или коротко — *дискретную случайную величину*).

В этой главе будут встречаться только дискретные случайные величины и поэтому мы будем использовать более короткий термин — **случайная величина**.

Пример 1.2.1. Сопоставляя каждому сообщению $x_i \in X$ число i , получим случайную величину — номер сообщения. Сопоставляя каждому сообщению $x_i \in X$ величину — $\ln p(x_i)$, получим другую случайную величину, которая далее определяется как **информация в сообщении**.

Пример 1.2.2. Пусть $X = \{x_1, \dots, x_5\}$ и $p(x)$ — распределение вероятностей на X . Положим $\varphi(x_1) = 3, \varphi(x_2) = 0, \varphi(x_3) = \varphi(x_4) = -1,8, \varphi(x_5) = 2$. Так определенная функция $\varphi(x)$ задает случайную величину, принимающую значения $-1,8; 0; 2$ и 3 , причем вероятности этих значений суть

$$\begin{aligned} p(\varphi = -1,8) &= p(x_3) + p(x_4), & p(\varphi = 0) &= p(x_2), \\ p(\varphi = 2) &= p(x_5), & p(\varphi = 3) &= p(x_1). \end{aligned}$$

Пусть $\{X, p(x)\}$ — дискретный ансамбль такой, что X — числовое множество. Функция $\varphi(x) = x$, очевидно, задает случайную величину, для которой X является множеством значений и $p(x)$ — распределением вероятностей. В этом случае случайную величину мы будем обозначать так же, как и ансамбль, а именно X . Вводя в рассмотрение некоторую случайную величину X , мы будем предполагать, что для этой величины определено или может быть определено распределение вероятностей $p(x)$ для всех значений x этой случайной величины (с. в.).

Число

$$MX \triangleq \sum_{x \in X} xp(x) \quad (1.2.1)$$

называется **математическим ожиданием** с. в. X . Число

$$\mu_k \triangleq M(X - MX)^k = \sum_{x \in X} (x - MX)^k p(x) \quad (1.2.2)$$

называется k -м **центральным моментом** с. в. X . При этом μ_2 называется **дисперсией**. В дальнейшем дисперсия с. в. X будет обозначаться через σ_X^2 . Если $Y = \varphi(X)$ есть с. в., определяемая с помощью функции $\varphi(\cdot)$ и ансамбля $\{X, p_1(x)\}$ *), то

$$MY = \sum_{y \in Y} yp_2(y) = \sum_{y \in Y} y \sum_{x: \varphi(x)=y} p_1(x) = \sum_{x \in X} \varphi(x) p_1(x), \quad (1.2.3)$$

где запись $x: \varphi(x) = y$ под знаком суммы означает, что суммирование производится по всем таким $x \in X$, что $\varphi(x) = y$.

*). Запись $Y = \varphi(X)$ означает, что если с. в. X принимает значения x , то с. в. Y принимает значение $y = \varphi(x)$. В дальнейшем для упрощения записи мы иногда будем использовать обозначение $Y = \varphi(x)$. При этом из контекста всегда будет ясно, идет ли речь о случайной величине, или ее значениях.

Основное свойство математического ожидания состоит в том, что математическое ожидание суммы некоторого числа с. в. равно сумме их математических ожиданий:

$$M[\alpha_1 X_1 + \dots + \alpha_n X_n] = \alpha_1 MX_1 + \dots + \alpha_n MX_n, \quad (1.2.4)$$

где $\alpha_1, \dots, \alpha_n$ — неслучайные числа. Для обоснования этого соотношения рассмотрим ансамбль $\{X_1, \dots, X_n, p(x^{(1)}, \dots, x^{(n)})\}$ и положим $Y = \alpha_1 X_1 + \dots + \alpha_n X_n$. Тогда, применяя формулу (1.2.3), получим

$$\begin{aligned} MY &= \sum_{x_1 \dots x_n} (\alpha_1 x^{(1)} + \dots + \alpha_n x^{(n)}) p(x^{(1)}, \dots, x^{(n)}) = \\ &= \alpha_1 \sum_{x_1 \dots x_n} x^{(1)} p(x^{(1)}, \dots, x^{(n)}) + \dots + \\ &+ \alpha_n \sum_{x_1 \dots x_n} x^{(n)} p(x^{(1)}, \dots, x^{(n)}) = \alpha_1 \sum_{x_1} x^{(1)} p_1(x^{(1)}) + \dots + \\ &+ \alpha_n \sum_{x_n} x^{(n)} p_n(x^{(n)}) = \alpha_1 MX_1 + \dots + \alpha_n MX_n, \end{aligned}$$

где предпоследнее равенство получается с помощью соотношений (1.1.7).

Предположим, что совместно заданы две с. в. X и Y , причем $p(x, y)$ — распределение вероятностей на множестве XY всех пар (x, y) , где x — значение с. в. X и y — значение с. в. Y . Случайные величины X и Y называются **статистически независимыми** (иногда — просто **независимыми**), если $p(x, y) = p_1(x) \cdot p_2(y)$ для всех значений x и y .

Число

$$K_{XY} \triangleq M(X - MX)(Y - MY) \triangleq \sum_{x, y} (x - MX)(y - MY) p(x, y), \quad (1.2.5)$$

где

$$MX = \sum_{x, y} xp(x, y), \quad MY = \sum_{x, y} yp(x, y), \quad (1.2.6)$$

называется **корреляционным моментом** случайных величин X и Y . Если с. в. X и Y независимы, то $K_{XY} = 0$. Действительно, учитывая независимость, получим

$$\begin{aligned} K_{XY} &= \sum_x \sum_y (x - MX)(y - MY) p_1(x) p_2(y) = \\ &= \left(\sum_x (x - MX) p_1(x) \right) \cdot \left(\sum_y (y - MY) p_2(y) \right) = 0. \end{aligned}$$

Заметим, что из равенства $K_{XY} = 0$ в общем случае не следует, что с. в. X и Y независимы.

Следующее простое неравенство лежит в основе закона больших чисел — основного теоретико-вероятностного инструмента теории информации.

Пусть с. в. X имеет нулевое математическое ожидание и дисперсию σ_X^2 . Тогда для произвольного положительного ϵ

$$\sigma_X^2 \triangleq \sum_{x \in X} x^2 p(x) \geq \sum_{|X| \geq \epsilon} x^2 p(x) \geq \epsilon^2 \sum_{|X| \geq \epsilon} p(x) = \epsilon^2 \Pr(|X| \geq \epsilon), \quad (1.2.7)$$

где $\Pr(|X| \geq \epsilon)$ есть вероятность того, что с. в. X будет по модулю не меньше, чем ϵ .

Первое равенство есть определение дисперсии (см. (1.2.2)). Первое неравенство получается в результате сужения области суммирования до множества таких значений x , модуль которых не меньше ϵ . Второе неравенство вытекает из того, что $x^2 \geq \epsilon^2$ для всех x из указанной области суммирования. Наконец, последнее равенство следует из определения вероятности события $|X| \geq \epsilon$.

Из (1.2.7) следует, что

$$\Pr(|X| \geq \epsilon) \leq \sigma_X^2 / \epsilon^2. \quad (1.2.8)$$

Это неравенство называется неравенством Чебышева.

Пусть X_1, \dots, X_n — независимые случайные величины, имеющие одинаковые распределения вероятностей $p(x)$. Эти с. в. соответствуют n независимым экспериментам (например, n независимым бросаниям игральной кости), причем с. в. X_i соответствует эксперименту, проводимому в i -й момент времени. Пусть Y — среднее арифметическое указанных с. в., т. е.

$$Y \triangleq \frac{1}{n} \sum_{i=1}^n X_i. \quad (1.2.9)$$

Легко определить математическое ожидание и дисперсию с. в. Y :

$$\mathbf{M}Y = \frac{1}{n} \sum_{i=1}^n \mathbf{M}X_i = m_X, \quad (1.2.10)$$

где через m_X обозначено математическое ожидание с. в. X_1, \dots, X_n (в силу одинаковой распределенности оно одинаково для всех случайных величин), и

$$\begin{aligned} \sigma_Y^2 &= \mathbf{M}(Y - \mathbf{M}Y)^2 = \mathbf{M} \left[\frac{1}{n} \sum_{i=1}^n (X_i - m_X) \right]^2 = \\ &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \mathbf{M}(X_i - m_X)(X_j - m_X) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n K_{X_i X_j}. \end{aligned}$$

Так как с. в. X_1, \dots, X_n независимы, то $K_{X_i X_j} = 0$ при $i \neq j$ и $K_{X_i X_i} = \sigma_X^2$, где σ_X^2 — дисперсия каждой из с. в. X_i . Учитывая это, имеем

$$\sigma_Y^2 = \frac{1}{n} \sigma_X^2. \quad (1.2.11)$$

Применим неравенство Чебышева к с. в. Y . В результате получим, что

$$\Pr(|Y - \mathbf{M}Y| \geq \epsilon) \leq \sigma_Y^2 / \epsilon^2$$

или

$$\Pr \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - m_X \right| \geq \epsilon \right) \leq \frac{\sigma_X^2}{n\epsilon^2}. \quad (1.2.12)$$

Полученный результат мы сформулируем в виде теоремы.

Теорема 1.2.1. (Закон больших чисел в форме Чебышева.) Пусть X_1, \dots, X_n — независимые одинаково распределенные дискретные случайные величины, имеющие конечное математическое ожидание и дисперсию. Тогда для любых положительных ϵ и δ найдется такое N , зависящее от ϵ и δ , что для всех $n > N$ вероятность того, что среднее арифметическое с. в. X_1, \dots, X_n будет отличаться от математического ожидания m_X каждой из случайных величин на величину, не меньшую чем ϵ , не превосходит δ :

$$\Pr \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - m_X \right| \geq \epsilon \right) \leq \delta. \quad (1.2.13)$$

Пример 1.2.3. Пусть X_1, \dots, X_n — независимые одинаково распределенные случайные величины, принимающие значения 0 и 1 с вероятностями $1-p$ и p соответственно. Легко видеть, что каждая из указанных с. в. имеет математическое ожидание, равное p , и дисперсию

$$\sigma_X^2 = \sum_{x \in X} (x - p)^2 p(x) = p^2(1-p) + (1-p)^2 p = p(1-p).$$

С. в.

$$Y \triangleq \frac{1}{n} \sum_{i=1}^n X_i$$

представляет собой относительное число (или долю) единиц в последовательности X_1, \dots, X_n . Применяя к с. в. Y закон больших чисел, получим, что при достаточно больших n для любого $\epsilon > 0$

$$\Pr \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - p \right| > \epsilon \right) \leq \frac{p(1-p)}{n\epsilon^2}. \quad (1.2.14)$$

Другими словами, доля единиц в достаточно длинной последовательности близка к вероятности появления единицы. Точный смысл этого утверждения проявляется с помощью неравенства (1.2.14). Конечно, не исключено, что в каком-либо эксперименте, в котором наблюдаются n случайных величин, определенных выше, доля единиц будет сильно отличаться от p . Однако вероятность этого события мала при достаточно больших n .

§ 1.3. Количество информации в сообщении. Энтропия

В этом параграфе мы сформулируем основное теоретико-информационное понятие — количество информации в сообщении. Мы увидим, что информация определяется только вероятностными свойствами сообщений. Все другие их свойства, например, полезность для тех или других действий, принадлежность тому или иному автору и др., игнорируются. Это специфика теории информации, о которой иногда забывают, что часто приводит к неправильным выводам.

Пусть $\{X, p(x)\}$ ансамбль сообщений, $X = \{x_1, \dots, x_L\}$.

Определение 1.3.1. Количеством собственной информации (или собственной информацией) в сообщении $x_i \in X$ называется число $I(x_i)$, определяемое соотношением

$$I(x_i) \triangleq -\log p(x_i), \quad i = 1, 2, \dots, L. \quad (1.3.1)$$

Основание, по которому берется логарифм в этом определении, влияет на единицу измерения количества информации. Наиболее часто употребляются логарифмы по основанию 2 и натуральные логарифмы. В первом случае единица измерения количества информации называется «бит», во втором — «нат» *).

Всюду ниже (если не оговорено противное) будем использовать только двоичные логарифмы и измерять количество информации в битах.

Рассмотрим основные свойства количества информации.

1. Собственная информация неотрицательна. Она равна нулю только в том случае, когда сообщение имеет вероятность 1. Такое сообщение можно рассматривать как неслучайное и известное заранее до проведения опыта. То сообщение имеет большую собственную информацию, которое имеет меньшую вероятность.

2. Рассмотрим ансамбль $\{XY, p(x, y)\}$. Для каждого сообщения из этого ансамбля, т. е. для каждой пары x_i, y_j , собственная информация $I(x_i, y_j) = -\log p(x_i, y_j)$. Если сообщения x_i и y_j статистически независимы, т. е. $p(x_i, y_j) = p_1(x_i) \cdot p_2(y_j)$, то

$$I(x_i, y_j) = -\log p_1(x_i) - \log p_2(y_j) = I(x_i) + I(y_j). \quad (1.3.2)$$

Это свойство называется свойством аддитивности информации.

*) От английских «binary digit» и «natural digit».

Количество информации, определяемое соотношением (1.3.1), является действительной функцией на ансамбле $\{X, p(x)\}$ и, следовательно, представляет собой случайную величину со значениями $I(x_1), \dots, I(x_L)$.

Определение 1.3.2. Математическое ожидание $H(X)$ случайной величины $I(x)$, определенной на ансамбле $\{X, p(x)\}$, называется энтропией этого ансамбля:

$$H(X) \triangleq MI(x) = \sum_{x \in X} I(x)p(x) = -\sum_{x \in X} p(x) \cdot \log p(x). \quad (1.3.3)$$

Энтропия представляет собой среднее количество собственной информации в сообщениях ансамбля X .

Замечание. Согласно определению 1.3.1 собственная информация принимает бесконечные значения для сообщений, вероятности которых равны нулю. Однако энтропия любого дискретного ансамбля конечна, так как выражение вида $z \log z$ при $z = 0$ по непрерывности доопределяется как 0. Основанием для такого доопределения является следующее соотношение:

$$\lim_{z \rightarrow 0} z \log z = \lim_{z \rightarrow 0} \frac{1/z}{-1/z^2} \log e = 0,$$

которое получается в результате применения правила Лопитала для раскрытия неопределенности типа ∞/∞ .

Пример 1.3.1. Пусть $X = \{x_1, x_2\}$ — двоичный ансамбль и $p(x_1) = p, p(x_2) = 1 - p$ — вероятности его сообщений. В этом случае энтропия является функцией одной переменной p

$$H(X) = -p \log p - (1-p) \log(1-p). \quad (1.3.4)$$

Эта функция показана на рис. 1.3.1. В точках $p = 0$ и $p = 1$ она не определена и в соответствии с предыдущим замечанием доопределяется до нуля.

Поведение энтропии как функции от p может быть исследовано с помощью вычисления производной от правой части равенства (1.3.4) (см. задачу 1.3.1).

Рассмотрим теперь свойства энтропии.

1. Энтропия всякого дискретного ансамбля неотрицательна:

$$H(X) \geq 0. \quad (1.3.5)$$

Равенство нулю возможно в том и только в том случае, когда существует некоторое сообщение $x_i \in X$, для которого $p(x_i) = 1$; при этом вероятности остальных сообщений равны нулю.

Неотрицательность следует из того, что собственная информация каждого сообщения дискретного ансамбля неотрицательна.

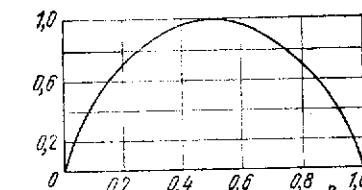


Рис. 1.3.1. Энтропия двоичного ансамбля.

Равенство нулю возможно только тогда, когда каждое слагаемое в (1.3.3) равно нулю, а это равносильно условию, указанному выше.

2. Пусть L — число сообщений в ансамбле X , тогда

$$H(X) \leq \log L, \quad (1.3.6)$$

причем равенство имеет место в том и только в том случае, когда все сообщения ансамбля имеют одинаковые вероятности.

Это неравенство можно доказывать с помощью стандартных методов поиска условного экстремума функции нескольких переменных. Однако имеется простое доказательство, основанное на следующем неравенстве для натурального логарифма:

$$\ln x \leq x - 1, \quad (1.3.7)$$

где равенство имеет место только при $x = 1$ (см. рис. 1.3.2) *).

Для доказательства (1.3.6) рассмотрим разность $H(X) - \log L$:

$$\begin{aligned} H(X) - \log L &= -\sum_x p(x) \log p(x) - \log L \sum_x p(x) = \\ &= -\sum_x p(x) [\log p(x) + \log L] = \log e \sum_x p(x) \ln \frac{1}{L p(x)}. \end{aligned}$$

Теперь используем неравенство (1.3.7). В результате получим

$$\begin{aligned} H(X) - \log L &\leq \log e \sum_x p(x) \left[\frac{1}{L p(x)} - 1 \right] = \\ &= \log e \left[\sum_x \frac{1}{L} - \sum_x p(x) \right] = 0. \end{aligned}$$

Отсюда следует неравенство (1.3.6). Поскольку равенство в (1.3.7) имеет место только тогда, когда аргумент логарифма равен единице, то равенство в (1.3.6) имеет место только тогда, когда $\frac{1}{L p(x)} = 1$ для всех $x \in X$, т. е. когда $p(x) = 1/L$ для всех $x \in X$.

Таким образом, энтропия принимает наименьшее значение 0 для ансамбля, в котором сообщения предопределены заранее — одно из них появляется всегда, остальные — никогда. Энтропия

*). Неравенство (1.3.7) может быть проверено аналитически, если заметить, что разность $\ln x - x + 1$ имеет отрицательную вторую производную и стационарную точку при $x = 1$.

принимает наибольшее значение $\log L$, когда все сообщения одинаково вероятны. Эти два свойства часто используют для того, чтобы толковать энтропию как степень или меру неопределенности (степень неопределенности знаний экспериментатора) в эксперименте с получением сообщений ансамбля. Если сообщения заранее предопределены и экспериментатор знает, какое сообщение он получит, то неопределенность и энтропия равны нулю. Если ни одно сообщение не имеет преимущества по отношению к другим и экспериментатор с одинаковыми шансами может получить любое сообщение, то как неопределенность, так и энтропия максимальны. Такое качественное толкование энтропии полезно, но не достаточно. Позже мы подробнее остановимся на содержательном смысле энтропии как скорости создания информации источником.

3. Пусть X и Y — статистически независимые ансамбли (см. 1.1.4). Тогда для каждой пары сообщений $x_i \in X$ и $y_j \in Y$ выполняется равенство (1.3.2). Усредняя левую и правую части (1.3.2) по распределению $p(x_i, y_j)$ и учитывая, что $M\{I(x_i, y_j)\} = H(XY)$ есть энтропия ансамбля $\{XY, p(x_i, y_j) = p_1(x_i) \cdot p_2(y_j)\}$, получим

$$H(XY) = M\{I(x_i) + I(y_j)\} = H(X) + H(Y). \quad (1.3.8)$$

Это свойство называется *свойством аддитивности энтропии*.

§ 1.4. Условная информация. Условная энтропия

Пусть $\{XY, p(x, y)\}$ — пара совместно заданных дискретных ансамблей $\{X, p(x)\}$ и $\{Y, p(y)\}$ *). Как указывалось выше, на каждом из множеств X и Y могут быть определены различные условные распределения. Зафиксируем некоторое сообщение $y \in Y, p(y) \neq 0$, и рассмотрим условное распределение $p(x|y)$ на X . Для каждого сообщения $x \in X$ в ансамбле $\{X, p(x|y)\}$ определена собственная информация

$$I(x|y) \triangleq -\log p(x|y), \quad (1.4.1)$$

*). Здесь и всюду, где нет опасности перепутать, мы используем один и тот же символ $p(\cdot)$ для обозначения нескольких, быть может различных, распределений вероятностей на различных множествах. Различие этих функций обозначается только использованием различных аргументов.

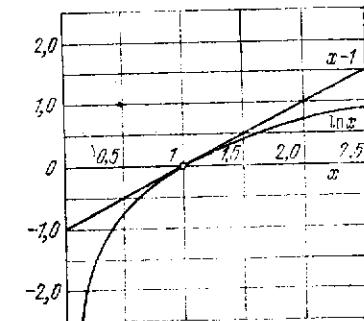


Рис. 1.3.2. К неравенству для логарифма (1.3.7).

которая называется *условной собственной информацией сообщения x при фиксированном сообщении y* . Как и раньше, $I(x|y)$ можно рассматривать как случайную величину на ансамбле $\{X; p(x|y)\}$; ее математическое ожидание

$$H(X|y) \triangleq \sum_x p(x|y) I(x|y) = -\sum_x p(x|y) \log p(x|y) \quad (1.4.2)$$

называется *условной энтропией ансамбля X относительно сообщения $y \in Y$* .

Условную энтропию $H(X|y)$ можно рассматривать как с. в. на ансамбле $\{Y, p(y)\}$.

Определение 1.4.1. Математическое ожидание $H(X|Y)$ случайной величины $H(X|y)$, определенной на ансамбле $\{Y, p(y)\}$, называется *условной энтропией ансамбля X относительно ансамбля Y* :

$$\begin{aligned} H(X|Y) &\triangleq \mathbf{M}H(X|y) = \\ &= \sum_Y p(y) H(X|y) = -\sum_X \sum_Y p(x, y) \cdot \log p(x|y). \end{aligned} \quad (1.4.3)$$

Замечание. Следует помнить, что условная вероятность $p(x|y)$, следовательно, условная собственная информация $I(x|y)$ и условная энтропия $H(X|y)$, определены только для тех сообщений $y \in Y$, вероятности которых отличны от нуля. Поэтому всюду ниже мы будем считать, что указанное условие выполнено и соответствующие вероятности не равны нулю. Если в ансамбле некоторые сообщения имеют нулевые вероятности, то мы будем исключать эти сообщения из рассмотрения, переходя тем самым к ансамблю, все сообщения которого имеют ненулевые вероятности.

Мы ввели условную энтропию $H(X|Y)$, рассматривая вначале условную энтропию $H(X|y)$ при фиксированном сообщении как случайную величину на ансамбле Y , а затем находя математическое ожидание этой с. в. Можно поступить иначе и рассмотреть условную собственную информацию $I(x|y)$ как действительную функцию, заданную на ансамбле $\{XY, p(x, y)\}$ и, следовательно, как с. в. на этом ансамбле. Тогда энтропия $H(X|Y)$ есть математическое ожидание с. в. $I(x|y)$ и может быть непосредственно определена с помощью правой части формулы (1.4.3).

Теперь мы продолжим рассмотрение свойств энтропии и условной энтропии.

1. Условная энтропия не превосходит безусловной энтропии того же ансамбля:

$$H(X|Y) \leq H(X), \quad (1.4.4)$$

причем равенство имеет место в том и только том случае, когда ансамбли X и Y статистически независимы.

Доказательство этого неравенства проводится с помощью неравенства для логарифма (1.3.7):

$$\begin{aligned} H(X|Y) - H(X) &= -\sum_X \sum_Y p(x, y) \log p(x|y) - \\ &- \sum_X p(x) \log p(x) = -\sum_X \sum_Y p(x, y) [\log p(x|y) - \log p(x)] = \\ &= \sum_{XY} p(x, y) \log \frac{p(x)}{p(x|y)} \leq \log e \cdot \sum_{XY} p(x, y) \left[\frac{p(x)}{p(x|y)} - 1 \right] = \\ &= \log e \left[\sum_{XY} p(x)p(y) - \sum_{XY} p(x, y) \right] = 0. \end{aligned} \quad (1.4.5)$$

Равенство выполняется в том и только в том случае, когда $p(x|y) = p(x)$ для всех $x \in X, y \in Y$.

2. Математическое ожидание $H(XY)$ собственной информации пары сообщений (x, y)

$$H(XY) \triangleq \sum_{XY} I(x, y) p(x, y) = -\sum_{XY} p(x, y) \log p(x, y) \quad (1.4.6)$$

по определению представляет собой энтропию ансамбля XY . Используя соотношение $p(x, y) = p(y) \cdot p(x|y)$ (определение условной вероятности), можно получить, что

$$\begin{aligned} H(XY) &= -\sum_{XY} p(x, y) \log p(x|y) - \sum_{XY} p(x, y) \log p(y) = \\ &= H(X|Y) + H(Y). \end{aligned} \quad (1.4.7)$$

Аналогично, используя соотношение $p(x, y) = p(x) p(y|x)$, можно получить, что

$$H(XY) = H(X) + H(Y|X). \quad (1.4.8)$$

Эти свойства также называются *свойствами аддитивности энтропии*. В случае независимых ансамблей соотношения (1.4.7) и (1.4.8) переходят в (1.3.8).

3. Предположим, что задан ансамбль $\{X, p(x)\}$ и на этом ансамбле определено отображение $\Phi(X)$ множества X в множество Y . Это отображение определяет ансамбль $\{Y, p(y)\}$, для которого $p(y) = \sum_{x: \Phi(x)=y} p(x)$. Пусть $H(X)$, $H(Y)$ — энтропии ансамблей X и Y соответственно. Тогда

$$H(Y) \leq H(X) \quad (1.4.9)$$

и знак равенства имеет место в том и только том случае, когда отображение $\Phi(x)$ обратимо, т. е. когда каждому элементу $y \in Y$ соответствует один и только один элемент $x \in X$.

Для доказательства неравенства (1.4.9) заметим, что совместное распределение вероятностей $p(x, y)$ на произведении множеств X и Y задается следующим образом: $p(x, y) = p(y|x)p(x)$, где $p(y|x) = 1$ для $y = \Phi(x)$ и $p(y|x) = 0$ для остальных $y \in Y$. Другими словами, каждое сообщение ансамбля X однозначно определяет сообщение ансамбля Y (в этом случае говорят, что ансамбль X однозначно определяет ансамбль Y). Тогда, как легко показать (см. задачу 1.4.1), $H(Y|X) = 0$. Из аддитивности и неотрицательности энтропии получим, что

$$H(Y) \leq H(Y) + H(X|Y) = H(X) + H(Y|X) = H(X). \quad (1.4.10)$$

Таким образом, при произвольных отображениях, задаваемых функцией $\Phi(x)$, $x \in X$, при которых ансамбль X переходит в некоторый другой ансамбль Y , энтропия не возрастает. Энтропия не изменяется тогда и только тогда, когда $H(X|Y) = 0$, т. е. когда ансамбль Y однозначно определяет ансамбль X . Другими словами, энтропия сохраняется только при обратимых преобразованиях.

4. Пусть $\{XYZ, p(x, y, z)\}$ — три совместно заданных ансамбля XYZ и

$$I(x|y, z) \triangleq -\log p(x|y, z) \quad (1.4.11)$$

— условная собственная информация сообщения x при фиксированной паре сообщений y, z , где

$$p(x|y, z) = \frac{p(x, y, z)}{\sum_X p(x, y, z)}. \quad (1.4.12)$$

Число

$$H(X|YZ) \triangleq MI(x|y, z) = -\sum_{XYZ} p(x, y, z) \log p(x|y, z) \quad (1.4.13)$$

называется *условной энтропией ансамбля X относительно пары ансамблей YZ* .

Имеет место следующее неравенство:

$$H(X|YZ) \leq H(X|Y), \quad (1.4.14)$$

которое доказывается таким же методом, как и неравенство (1.4.4). Равенство в (1.4.14) выполняется в том и только том случае, когда $p(x|y, z) = p(x|y)$ для всех $(x, y, z) \in XYZ$, т. е. когда при данном сообщении y ансамбли X и Z статистически независимы.

Это неравенство легко обобщается на случай n совместно заданных ансамблей. Рассмотрим ансамбль $\{X_1 \dots X_n, p(x^{(1)}, \dots, x^{(n)})\}$. Тогда для любых s и m , $1 \leq s \leq m \leq n$, имеет место следующее неравенство:

$$H(X_s | X_{s-1} \dots X_{1-s}) \leq H(X_m | X_{m-1} \dots X_{1-m}). \quad (1.4.15)$$

Его левая и правая части определены, как и в (1.4.13), а именно, как математические ожидания соответствующих информаций.

Для вывода еще одного важного неравенства заметим, что для любой последовательности $(x^{(1)}, \dots, x^{(n)}) \in X_1 \dots X_n$ ее вероятность может быть записана следующим образом:

$$p(x^{(1)}, \dots, x^{(n)}) = p(x^{(1)}) p(x^{(2)}|x^{(1)}) \dots p(x^{(n)}|x^{(n-1)} \dots x^{(1)}), \quad (1.4.16)$$

где

$$p(x^{(i)}|x^{(i-1)} \dots x^{(1)}) = \frac{p(x^{(1)}, \dots, x^{(i)})}{p(x^{(1)}, \dots, x^{(i-1)})} \quad (1.4.17)$$

и

$$p(x^{(1)}, \dots, x^{(n)}) = \sum_{X_{i+1} \dots X_n} p(x^{(1)}, \dots, x^{(n)}). \quad (1.4.18)$$

Тогда

$$\begin{aligned} I(x^{(1)}, \dots, x^{(n)}) &= \\ &= I(x^{(1)}) + I(x^{(2)}|x^{(1)}) + \dots + I(x^{(n)}|x^{(n-1)}, \dots, x^{(1)}), \end{aligned} \quad (1.4.19)$$

где

$$I(x^{(i)}|x^{(i-1)}, \dots, x^{(1)}) \triangleq -\log p(x^{(i)}|x^{(i-1)}, \dots, x^{(1)}) \quad (1.4.20)$$

— условная собственная информация сообщения $x^{(i)}$. Усредняя обе части соотношения (1.4.19), получим

$$\begin{aligned} H(X_1 \dots X_n) &= H(X_1) + H(X_2|X_1) + \dots + H(X_n|X_{n-1} \dots X_1) = \\ &= \sum_{i=1}^n H(X_i|X_{i-1}, \dots, X_1). \end{aligned} \quad (1.4.21)$$

Используя (1.4.15), можно записать

$$H(X_1 \dots X_n) \leq \sum_{i=1}^n H(X_i). \quad (1.4.22)$$

§ 1.5. ЭНТРОПИЯ НА СООБЩЕНИЕ ДИСКРЕТНОГО СТАЦИОНАРНОГО ИСТОЧНИКА

Рассмотрим дискретный стационарный источник, выбирающий сообщения из множества X . Для удобства мы будем обозначать через X_i ансамбль сообщений в i -й момент времени, хотя множества $\{X_i\}$ в каждый момент времени совпадают между собой и совпадают с множеством X . Для стационарных источников n -мерные распределения вероятностей $n = 1, 2, \dots$ не зависят от сдвига по оси времени. Следовательно, любые величины, зависящие только от n -мерных распределений вероятностей, одинаковы для всех сдвигов по оси времени. В частности, энтропия $H(X^n)$, где $X^n = X_{i+1} \dots X_{i+n}$, не зависит от выбора числа i ,

указывающего расположение n последовательных моментов времени на оси времени. Поэтому всюду при рассмотрении стационарных источников мы будем опускать этот индекс i из обозначения энтропии.

Пусть X^{n-1} и $X^n = X^{n-1}X_n$ — два ансамбля последовательностей длины $n-1$ и n соответственно и

$$H(X_n | X^{n-1}) \triangleq -\sum_{x^n} p(x^{(1)}, \dots, x^{(n)}) \log p(x^{(n)} | x^{(1)}, \dots, x^{(n-1)}), \quad (1.5.1)$$

— условная энтропия ансамбля X_n относительно ансамбля X^{n-1} .

Теорема 1.5.1. Для всякого дискретного стационарного источника последовательность $H(X_n | X^{n-1})$, $n = 1, 2, \dots$, имеет предел:

$$\lim_{n \rightarrow \infty} H(X_n | X^{n-1}) \triangleq H(X | X^\infty). \quad (1.5.2)$$

Доказательство. Покажем вначале, что последовательность

$$H(X_1), H(X_2 | X_1), \dots, H(X_n | X_{n-1} \dots X_1), \dots \quad (1.5.3)$$

не возрастает. Хотя эта последовательность есть последовательность энтропий с возрастающим числом условий, непосредственно применить неравенство (1.4.15) нельзя, так как (1.5.3) есть последовательность энтропий различных ансамблей, а не одного ансамбля, как в (1.4.15). Однако в случае стационарных источников для всех целых i, j, n

$$H(X_j | X_{j-1} \dots X_{j-i}) = H(X_n | X_{n-1} \dots X_{n-i}), \quad (1.5.4)$$

поскольку левая и правая части этого равенства определяются только $i+1$ -мерными распределениями вероятностей. Применяя теперь неравенство (1.4.15), получим, что последовательность (1.5.3) не возрастает.

С другой стороны, все члены этой последовательности ограничены снизу нулем. Любая невозрастающая последовательность, ограниченная снизу, имеет предел. Обозначая этот предел через $H(X | X^\infty)$, получим утверждение теоремы.

Рассмотрим теперь последовательность

$$H_n(X) \triangleq \frac{1}{n} H(X^n), \quad n = 1, 2, \dots \quad (1.5.5)$$

Если эта последовательность имеет предел при $n \rightarrow \infty$, то этот предел представляет собой среднее количество информации, порождаемое источником в единицу времени, и называется *энтропией стационарного источника на сообщение*.

Теорема 1.5.2. Для всякого дискретного стационарного источника последовательность (1.5.5) имеет предел, причем

$$\lim_{n \rightarrow \infty} H_n(X) = H(X | X^\infty). \quad (1.5.6)$$

Доказательство. Покажем вначале, что последовательность $H_n(X)$, $n = 1, 2, \dots$, не возрастает. Рассмотрим энтропию $H(X^{n+1})$, где $X^{n+1} = X^n X_{n+1}$ — ансамбль последовательностей сообщений длины $n+1$. Используя свойство аддитивности энтропии, можно записать

$$\begin{aligned} H(X^{n+1}) &= H(X^n) + H(X_{n+1} | X^n) = H(X^n) + H(X_n | X^n) \leq \\ &\leq H(X^n) + H(X_n | X^{n-1}), \end{aligned} \quad (1.5.7)$$

где второе равенство есть следствие стационарности (см. 1.5.4)), а неравенство — следствие неравенства (1.4.15). Используя (1.4.21) и (1.4.15), можно получить

$$H(X^n) = \sum_{i=1}^n H(X_i | X^{i-1}) \geq n H(X_n | X^{n-1}). \quad (1.5.8)$$

Неравенства (1.5.7) и (1.5.8) совместно дают

$$H(X^{n+1}) \leq \frac{n+1}{n} H(X^n).$$

Деля обе части этого неравенства на $n+1$, получим

$$H_{n+1}(X) \leq H_n(X), \quad (1.5.9)$$

т. е. рассматриваемая последовательность энтропий не возрастает.

Поскольку последовательность (1.5.5) не возрастает и каждый ее член ограничен снизу нулем, то она имеет предел. Покажем теперь, что этот предел совпадает с $H(X | X^\infty)$.

Можно записать

$$\begin{aligned} H_n(X) &= \frac{1}{n} \sum_{i=1}^n H(X_i | X^{i-1}) = \frac{1}{n} \sum_{i=1}^k H(X_i | X^{i-1}) + \\ &+ \frac{1}{n} \sum_{i=k+1}^n H(X_i | X^{i-1}) \leq \frac{k}{n} H(X) + \frac{n-k}{n} H(X_{k+1} | X^k), \end{aligned} \quad (1.5.10)$$

где $k \leq n$. Выберем k таким, чтобы для заданного положительного ε выполнялось неравенство

$$H(X_{k+1} | X^k) - H(X | X^\infty) \leq \frac{\varepsilon}{2}. \quad (1.5.11)$$

Это всегда можно сделать, так как последовательность $H(X_{k+1} | X^k)$, $k = 0, 1, \dots$, имеет предел $H(X | X^\infty)$ и стремится

к нему, не возрастаю. По выбранному k определим N так, чтобы для всех $n > N$

$$\frac{k}{n} H(X) \leq \frac{\epsilon}{2}. \quad (1.5.12)$$

Тогда получим, что для любого $\epsilon > 0$ всегда найдется такое N , что для всех $n > N$

$$H_n(X) \leq H(X|X^\infty) + \epsilon. \quad (1.5.13)$$

С другой стороны,

$$H_n(X) = \frac{1}{n} \sum_{i=1}^n H(X_i | X^{i-1}) \geq H(X_n | X^{n-1}) \geq H(X | X^\infty). \quad (1.5.14)$$

Так как ϵ — произвольное положительное число, то из (1.5.13) и (1.5.14) следует, что $H(X|X^\infty)$ является пределом последовательности $H_n(X)$, $n = 1, 2, \dots$. Теорема доказана.

§ 1.6. Постановка задачи кодирования дискретных источников равномерными кодами

Обозначим через A некоторое множество, состоящее из D , $D > 1$, элементов: $A = \{a_1, \dots, a_D\}$. Назовем его *алфавитом кода источника*. Элементы множества A будем называть *кодовыми символами*.

Последовательность кодовых символов называется *кодовым словом*, а любое семейство кодовых слов — *кодом над алфавитом* A .

Пример 1.6.1. Пусть A — множество букв русского алфавита (включая пробел). Любая последовательность букв заданной длины n является кодовым словом длины n над алфавитом A . Так, АБЫЙ-АБ есть слово длины 8 (-пробел). Другим примером кодового алфавита является множество десятичных цифр, включая пробел. Следующие последовательности -0001, -099 являются примерами двух кодовых слов длины 5 и 4 соответственно.

Пример 1.6.2. Пусть $A = \{0, 1\}$ — двоичный алфавит. Множество последовательностей $\{011, 1100, 11, 1\}$ является кодом объема 4 над двоичным алфавитом. Это пример кода, слова которого имеют разную длину. Множество последовательностей $\{00, 01, 10, 11\}$ также является двоичным кодом объема 4.

Определение 1.6.1. Код называется *равномерным*, если все его слова имеют одинаковую длину t , это число называется *длиной кода*. Если хотя бы два кодовых слова имеют различные длины, то код называют *неравномерным*.

Количество различных слов равномерного кода длины t не превосходит D^t — числа различных D -ичных последовательностей длины t . Количество различных слов неравномерного кода с максимальной длиной кодовых слов t не превышает $D(D^m - 1)/(D - 1)$ (см. задачу 1.6.1).

Определение 1.6.2. Кодированием сообщений ансамбля X посредством кода называется отображение (необязательно взаимно однозначное) множества сообщений в множество кодовых слов.

Пример 1.6.3. Пусть $X = \{x_1, \dots, x_6\}$. Возможны следующие способы кодирования посредством двоичных кодов:

- a) $x_1 \rightarrow 000; x_2 \rightarrow 001; x_3 \rightarrow 010; x_4 \rightarrow 011; x_5 \rightarrow 100; x_6 \rightarrow 101;$
- б) $x_1 \rightarrow 00; x_2 \rightarrow 01; x_3 \rightarrow 00; x_4 \rightarrow 10; x_5 \rightarrow 01; x_6 \rightarrow 11;$
- в) $x_1 \rightarrow 00; x_2 \rightarrow 00; x_3 \rightarrow 01; x_4 \rightarrow 01; x_5 \rightarrow 10; x_6 \rightarrow 11.$

В п. а) кодирование однозначно. В пп. б) и в) приведены два различных способа кодирования с одним и тем же множеством кодовых слов.

Рассмотрим теперь кодирование дискретного источника. Для того чтобы лучше понять основные проблемы, возникающие при кодировании, мы рассмотрим несколько подобных примеров.

Пример 1.6.4. Предположим, что на центральном телеграфе введена автоматическая обработка телеграмм, записанных в алфавите, объем которого для простоты будем считать равным 64 буквам (сюда входят сами буквы, русские и латинские, цифры, знак пробела и некоторые вспомогательные знаки). Предположим, что эффективность такой обработки определяется количеством телеграмм, которые можно ввести в запоминающее устройство (ЗУ) обрабатывающего устройства. Пусть объем ЗУ равен N двоичным ячейкам. Чтобы осуществить запись телеграмм в ЗУ, необходимо закодировать телеграммы с помощью двоичного кода.

Рассмотрим вначале так называемое побуквенное кодирование. В этом случае используют код, содержащий $2^6 = 64$ кодовых слова, и каждой букве телеграммы сопоставляют некоторое кодовое слово. Предположим, что используется равномерный код, в котором все слова имеют длину 6 двоичных символов. При этом в ЗУ можно записать $N/6$ букв. Если считать, что средняя длина телеграммы — 20 слов, а средняя длина слова 8 букв, то в ЗУ можно поместить примерно $N/960$ телеграмм.

Второй способ кодирования состоит в том, что выделяется специальный словарь, например, из $2^{13} = 8192$ слов (такого словаря достаточно для составления практически любой телеграммы). Каждое слово словаря может быть закодировано с помощью двоичной последовательности длиной 13 символов. В этом случае в ЗУ можно записать $N/260$ телеграмм, т. е. в три-четыре раза больше, чем в первом случае.

При кодировании можно условиться о том, что всякое слово, не содержащееся в словаре, кодируется так же, как слово «ошибка». При декодировании вместо этих слов будет воспроизводиться слово «ошибка».

Таким образом, второй способ является значительно более эффективным. Полученный эффект объясняется тем, что далеко не всякая последовательность букв встречается в языке, а тем более среди слов телеграмм. Имеются почти невероятные буквосочетания, например, АБЫЙ-АБ. Однако необходимо помнить о том, что вероятность такого буквосочетания хотя и мала, но не равна нулю. Последнее следует, например, из того, что оно появляется в настоящем тексте. Поэтому при кодировании возможны ошибки.

Заметим, что при побуквенном кодировании все слова телеграмм могут быть однозначно (безошибочно) закодированы.

Изложенная в примере 1.6.4 идея может быть применена для кодирования произвольных дискретных источников равномерными кодами.

Предположим, что множество всех последовательностей длиной n из сообщений источника разбито на два подмножества. Первое из них образовано всеми теми последовательностями (блоками длиной n), которые сопоставлены с кодовыми словами взаимно однозначно. Это подмножество называется *множеством однозначно кодируемых и декодируемых блоков*. Второе подмножество образовано всеми остальными блоками, каждому из которых сопоставляется одно и то же кодовое слово *). Последнее подмножество называется *множеством неоднозначно кодируемых и декодируемых блоков*. Все кодовые слова имеют одинаковую длину m , которая определяется числом M кодовых слов: $m = \lceil \log M \rceil$ — наименьшее целое, удовлетворяющее неравенству $D^m \geq M$, где D — объем кодового алфавита. При кодировании последовательность сообщений на выходе источника разбивается на блоки длиной n , каждому блоку кодер сопоставляет соответствующее кодовое слово.

Описанный метод кодирования дискретного источника мы будем называть *равномерным кодированием*. Ошибкой равномерного кодирования является событие, состоящее в появлении неоднозначного кодируемого блока.

При равномерном кодировании количество D -ичных кодовых символов, приходящихся на одно сообщение, равно $\frac{1}{n} \log M$.

Определение 1.6.3. Число

$$R \triangleq \frac{\log M}{n} \quad (1.6.1)$$

называется *скоростью равномерного кодирования источника посредством кода с M кодовыми словами при разбиении последовательности сообщений на блоки длины n* . Скорость равномерного кодирования измеряется в двоичных символах на сообщение.

Количество двоичных символов на сообщение естественно измерять числом $\frac{1}{n} \log M$, а не $\frac{m \log D}{n}$. Очевидно, что эти два числа совпадают, когда число кодовых слов равно D^n . В остальных случаях именно первое из этих чисел равно количеству двоичных символов на сообщение (см. задачу 1.6.2).

Как видно из примера 1.6.4, существенной частью задания равномерного кода является указание множества последовательностей сообщений источника, которые кодируются с помощью кодовых слов однозначно. Все коды, которые имеют одинаковое

*) Можно сопоставлять и различные кодовые слова. Это не существенно. Существенно лишь, что эти кодовые слова использованы для представления однозначно кодируемым и декодируемым блоков.

В дальнейшем мы будем употреблять более короткое выражение — однозначно (неоднозначно) кодируемыем блоки.

число кодовых слов и однозначно кодируют одно и то же множество сообщений, имеют одинаковую скорость и одинаковую вероятность ошибки. Поэтому ни сами кодовые слова, ни способ сопоставления кодовых слов и однозначно кодируемым сообщениям не имеют значения с точки зрения качества кода.

При кодировании с помощью равномерных кодов основная задача состоит в определении наименьшей возможной скорости кодирования, при которой вероятность ошибки может быть сделана произвольно малой. Наименьшая достижимая скорость кодирования является характеристикой источника сообщений и называется *скоростью создания информации*.

Определение 1.6.4. *Скоростью создания информации* дискретным источником при равномерном кодировании называется наименьшее число H такое, что для любого $R > H$ и любого сколь угодно малого положительного числа δ найдется n (длина кодируемым сообщений) и равномерный код со скоростью кодирования R , для которого вероятность неправильного декодирования не превосходит δ .

В следующем примере показано, что по крайней мере для некоторых источников можно построить код с заданной малой вероятностью ошибки декодирования.

Пример 1.6.5. Рассмотрим двоичный источник без памяти, независимо выбирающий сообщения из множества $X = \{0, 1\}$, причем единица выбирается с вероятностью $p < \frac{1}{2}$. Этот источник возможно кодировать побуквенно. При этом на каждое сообщение будет затрачиваться 1 двоичный символ. При любом безошибочном кодировании с помощью равномерного кода скорость кодирования не будет меньше единицы.

Предположим, что допускается некоторая достаточно малая вероятность ошибки при кодировании. Рассмотрим множество X^n всех последовательностей длины n на выходе источника. Оно содержит 2^n последовательностей. Пусть $\varepsilon > 0$ и T_n — это подмножество множества X^n , состоящее из всех последовательностей, в которых число единиц k удовлетворяет условию

$$|k - np| \ll \varepsilon n. \quad (1.6.2)$$

Согласно закону больших чисел (см. пример 1.2.3) вероятность появления на выходе источника последовательности, принадлежащей множеству T_n , не меньше чем $1 - \frac{p(1-p)}{ne^2}$ и может быть сделана как угодно близкой к единице выбором достаточно большого n .

Рассмотрим кодирование, при котором имеется равномерный код, состоящий из M_n кодовых слов, где M_n — количество различных последовательностей в T_n , и каждой последовательности из T_n сопоставляется свое кодовое слово. Всем же последовательностям, не принадлежащим T_n , ставится в соответствие одно и то же слово, например, первое. При таком кодировании все последовательности сообщений из T_n будут воспроизводиться правильно, а при появлении на выходе источника последовательности, не принадлежащей T_n , будет происходить ошибка. Очевидно, что вероятность ошибки может быть сделана как угодно малой выбором достаточно большого n .

Оценим теперь скорость кодирования. Для этого необходимо оценить число M_n :

$$M_n = \sum_{|k-np| \leq \varepsilon n} C_n^k \ll (2ne + 1) C_n^{k_0}, \quad (1.6.3)$$

где $2ne + 1$ — количество слагаемых в сумме, а k_0 соответствует максимальному слагаемому. При $p + \varepsilon < 1/2$ максимальное слагаемое получается при $k_0 = [n(p + \varepsilon)]$, где $[x]$ — целая часть x (см. задачу 1.6.3 (в)). Используя верхнюю границу для числа сочетаний (та же задача (б)), получим, что

$$M_n \ll (2ne + 1) (2\pi n \lambda_0 (1 - \lambda_0))^{-1/2} 2^{nh(\lambda_0)}, \quad (1.6.4)$$

где $\lambda_0 = k_0/n$ и $h(\lambda_0) = -\lambda_0 \log \lambda_0 - (1 - \lambda_0) \log (1 - \lambda_0)$. При этом скорость кодирования

$$R = \frac{\log M_n}{n} \ll h(\lambda_0) + \frac{1}{n} \log \frac{2ne + 1}{\sqrt{2\pi n \lambda_0 (1 - \lambda_0)}} \quad (1.6.5)$$

стремится к $h(p + \varepsilon)$ при больших n .

Возьмем, например, $p = 0.1$. Если положить $R = 0.66$ и найти ε из уравнения $h(0.1 + \varepsilon) = 0.66$, то получим $\varepsilon = 0.074$. Вероятность ошибки P_e при кодировании последовательности сообщений длины n , т. е. вероятность появления такой последовательности, количество единиц в которой удовлетворяет неравенству $|k - np| > \varepsilon n$, можно оценить, правда весьма грубо, с помощью неравенства Чебышева (см. (1.2.14))

$$P_e = \Pr(|k - np| > \varepsilon n) = \Pr\left(\left|\frac{k}{n} - p\right| > \varepsilon\right) \ll \frac{p(1-p)}{n\varepsilon^2},$$

откуда получается $P_e \leq 17/n$. Использование для оценки P_e асимптотической формулы Муавра—Лапласа [6] дает при больших n

$$\ln P_e \sim -\frac{n\varepsilon^2}{2p(1-p)} = -0.03n.$$

Выбирая ε достаточно малым, можно приближать скорость кодирования как угодно близко к величине $h(p) = 0.47$ — энтропии двоичного ансамбля, для которого одно из сообщений имеет вероятность $p = 0.1$.

Ниже мы покажем, что энтропия определяет наименьшее возможное значение скорости кодирования и согласно определению 1.6.4 дает величину скорости создания информации при равномерном кодировании.

В последующей части этой главы мы будем заниматься определением скорости создания информации различными дискретными источниками. Для того чтобы установить, что некоторая величина, скажем H , является скоростью создания информации, необходимо доказать два утверждения:

1. Для любого $R > H$ и произвольного положительного δ найдутся n и код со скоростью R , кодирующий отрезки сообщений длины n , для которого вероятность ошибки не больше δ . Такое утверждение называется прямой теоремой кодирования.

2. Для любого $R < H$ найдется зависящее от R положительное число δ такое, что для всех n и для всех равномерных кодов со скоростью R вероятность ошибки при декодировании больше, чем это δ . Другими словами, вероятность ошибки не может быть сделана равной нулю или сколь угодно малой. Это утверждение называется обратной теоремой кодирования.

Вначале, в §§ 1.7—1.9, мы будем рассматривать задачу равномерного кодирования дискретных источников. Затем, в §§ 1.10—1.12, мы рассмотрим задачу неравномерного кодирования. В заключение этой главы будут сопоставлены эти две задачи и соответствующие методы кодирования.

§ 1.7. Теорема о высоковероятных множествах дискретного источника без памяти

Рассмотрим дискретный источник без памяти, выбирающий сообщения из конечного множества X . Согласно определению источника без памяти вероятность каждой последовательности сообщений $\mathbf{x} \triangleq (x^{(1)}, \dots, x^{(n)}) \in X^n$ на его выходе может быть записана в виде

$$p(\mathbf{x}) = \prod_{i=1}^n p(x^{(i)}), \quad (1.7.1)$$

где $p(x)$, $x \in X$, — распределение вероятностей на множестве X . Будем полагать, что $p(x) > 0$ для всех $x \in X$. В противном случае всегда можно перейти к рассмотрению суженного множества, для которого это предположение верно.

Пусть $H(X)$ — энтропия ансамбля $\{X, p(x)\}$ и ε — некоторое положительное число. Определим подмножество $T_n(\varepsilon)$ множества X^n следующим образом:

$$T_n(\varepsilon) \triangleq \left\{ \mathbf{x}: H(X) - \varepsilon \leq I(\mathbf{x}) \leq H(X) + \varepsilon \right\}, \quad (1.7.2)$$

где $I(\mathbf{x}) = -\log p(\mathbf{x})$ — собственная информация последовательности $\mathbf{x} \in X^n$.

Теорема 1.7.1. Для любых положительных чисел ε и δ найдется N такое, что для всех $n > N$ выполняются следующие неравенства:

$$\Pr(\mathbf{x} \in T_n(\varepsilon)) \geq 1 - \delta, \quad (1.7.3)$$

$$(1 - \delta) 2^{n[H(X) - \varepsilon]} \leq |T_n(\varepsilon)| \leq 2^{n[H(X) + \varepsilon]}, \quad (1.7.4)$$

где $|T_n(\varepsilon)|$ — число элементов в множестве $T_n(\varepsilon)$.

Доказательство. В силу отсутствия памяти у источника

$$I(\mathbf{x}) = -\sum_{i=1}^n \log p(x^{(i)}) = \sum_{i=1}^n I(x^{(i)}), \quad (1.7.5)$$

где $I(x^{(i)})$, $i = 1, 2, \dots, n$, — независимые одинаково распределенные случайные величины, принимающие ограниченные значения. К последовательности таких случайных величин можно

применить закон больших чисел, из которого следует, что при любых $\epsilon > 0$ и $\delta > 0$ найдется N такое, что при всех $n > N$

$$\Pr\left(\left|\frac{1}{n} \sum_{i=1}^n I(x^{(i)}) - H(X)\right| > \epsilon\right) < \delta \quad (1.7.6)$$

или

$$\Pr\left(\left|\frac{1}{n} \sum_{i=1}^n I(x^{(i)}) - H(X)\right| \leq \epsilon\right) \geq 1 - \delta. \quad (1.7.7)$$

Нетрудно увидеть, что вероятность $\Pr(x \in T_n(\epsilon))$ равна левой части неравенства (1.7.7) и, следовательно, неравенство (1.7.3) справедливо.

С другой стороны, из определения множества $T_n(\epsilon)$ следует, что для всякой последовательности $x \in T_n(\epsilon)$

$$2^{-n[H(X) + \epsilon]} \leq p(x) \leq 2^{-n[H(X) - \epsilon]}. \quad (1.7.8)$$

Суммируя обе части левого неравенства по всем элементам $T_n(\epsilon)$, получим следующую цепочку неравенств:

$$1 \geq \sum_{T_n(\epsilon)} p(x) \geq |T_n(\epsilon)| \cdot 2^{-n[H(X) + \epsilon]}, \quad (1.7.9)$$

откуда следует правое неравенство в (1.7.4). Суммируя теперь обе части правого неравенства в (1.7.8) по всем элементам $T_n(\epsilon)$, получим, что для достаточно больших n

$$(1 - \delta) \leq \sum_{T_n(\epsilon)} p(x) \leq |T_n(\epsilon)| 2^{-n[H(X) - \epsilon]}, \quad (1.7.10)$$

что дает левое неравенство в (1.7.4). Теорема доказана.

Теорема показывает, что последовательности сообщений на выходе дискретного источника без памяти могут быть подразделены на две группы: $T_n(\epsilon)$ и $\bar{T}_n(\epsilon)$, где второе множество — это дополнение первого до X^n . Последовательности из первой группы обладают тем свойством, что их вероятности достаточно близки друг к другу (см. (1.7.8)) и суммарная вероятность всех элементов этого множества весьма близка к единице (см. (1.7.3)). Тем не менее, множество $T_n(\epsilon)$ может составлять лишь очень малую долю по числу элементов от множества X^n . Действительно, если обозначить число элементов в X через L , то $|X^n| = L^n = 2^{n \log L}$. Доля α множества $T_n(\epsilon)$ в X^n

$$\alpha \triangleq \frac{|T_n(\epsilon)|}{|X^n|} \leq 2^{-n[\log L - H(X) - \epsilon]}. \quad (1.7.11)$$

Если $\log L - H(X) - \epsilon = a > 0$, т. е. если энтропия источника строго меньше, чем $\log L$, и ϵ достаточно мало, то α убывает к нулю при увеличении n как 2^{-an} .

Эти свойства множества $T_n(\epsilon)$ позволяют называть его *множеством типичных последовательностей или высоковероятным множеством дискретного источника без памяти*.

Следующие примеры показывают структуру высоковероятных множеств для двух двоичных источников.

Пример 1.7.1. Рассмотрим двоичный источник сообщений о результатах подбрасывания симметричной монеты. Вероятности сообщений одинаковы и равны $1/2$. Энтропия такого ансамбля $H(X) = 1$. Вероятность любой последовательности сообщений длины n не зависит от последовательности и равна 2^{-n} . Отсюда следует, что $I(x) = n = nH(X)$, т. е. все последовательности из X^n принадлежат множеству $T_n(\epsilon)$ при любом ϵ . Другими словами, для этого источника $X^n = T_n(\epsilon)$ и высоковероятное множество совпадает с множеством всех последовательностей.

Пример 1.7.2. Пусть двоичный источник без памяти выбирает сообщения из множества $X = \{0, 1\}$, причем 1 появляется с вероятностью $p < 1/2$. Для такого источника $H(X) = -p \log p - (1-p) \log(1-p)$.

$$p(x) = p^t (1-p)^{n-t}, \quad x \in X^n, \quad (1.7.12)$$

где t — количество единиц в последовательности x . Обозначая $\frac{t}{n} = \tau$, получим

$$\frac{1}{n} I(x) = -\tau \log p - (1-\tau) \log(1-p) \quad (1.7.13)$$

и, следовательно,

$$\begin{aligned} T_n(\epsilon) &= \{x: |(\tau - p) \log p + (p - \tau) \log(1-p)| \leq \epsilon\} = \\ &= \left\{x: \left|(\tau - p) \log \frac{1-p}{p}\right| \leq \epsilon\right\} = \left\{x: \left|\frac{t}{n} - p\right| \leq \epsilon_1\right\}, \end{aligned} \quad (1.7.14)$$

где $\epsilon_1 = \epsilon / \log \frac{1-p}{p}$. Таким образом, высоковероятное множество для рассматриваемого источника состоит из таких последовательностей $x \in X^n$, число единиц в которых близко к pr . Например, если $p = 0,1$, то $H(X) = 0,47$. Высоковероятное множество состоит из последовательностей, число единиц в которых близко к 0,1 n . Количество последовательностей в этом множестве близко к $2^{nH(X)} = 2^{n \cdot 0,47}$. Доля множества $T_n(\epsilon)$ близка к $2^{-n[1-H(X)]} = 2^{-n \cdot 0,53}$. Так, при $n = 100$ это число примерно равно 10^{-16} .

§ 1.8. Скорость создания информации дискретным источником без памяти при равномерном кодировании

В этом параграфе мы покажем, что скорость создания информации дискретным источником без памяти при равномерном кодировании равна его энтропии. Как отмечалось в § 1.6, для доказательства этого утверждения нужно доказать прямую и обратную теоремы кодирования. Каждый код, кодирующий сообщения источника, характеризуется скоростью кодирования — отношением логарифма числа кодовых слов к длине кодируемых сообщений, множеством однозначно кодируемых сообщений и вероятностью ошибки. Поэтому, говоря о том, что существует код

со скоростью R , кодирующий источник с вероятностью ошибки P_e , мы подразумеваем, что существуют n , код с 2^{nR} кодовыми словами и множество однозначно кодируемым последовательностям сообщений $T_n \subseteq X^n$, для которых вероятность ошибки равна P_e .

Предположим, что дискретный источник без памяти выбирает сообщения из множества X с распределением вероятностей $p(x)$, $p(x) \neq 0$ для всех $x \in X$, и $H(X)$ — энтропия ансамбля $\{X, p(x)\}$.

Теорема 1.8.1 (прямая теорема). Пусть $R > H(X)$, тогда для любого положительного δ существует код со скоростью R , который кодирует дискретный источник без памяти с вероятностью ошибки, не превышающей δ .

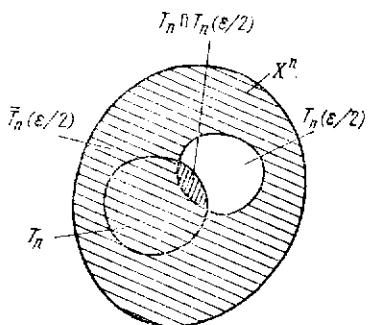


Рис. 1.8.1. К доказательству обратной теоремы кодирования.

будет превышать δ . При этом количество кодовых слов не должно быть меньше числа элементов в множестве $T_n(\epsilon)$. Это условие будет удовлетворено, если

$$2^{nR} \geq 2^{n[H(X) + \epsilon]}, \quad (1.8.1)$$

откуда следует, что $R \geq H(X) + \epsilon$. Теорема доказана.

З а м е ч а н и е. В теореме утверждается существование кода, кодирующего дискретный источник без памяти с произвольно малой вероятностью ошибки. В действительности доказано нечто большее, а именно, что для любых $\epsilon, \delta > 0$ найдется N и последовательность кодов, кодирующих отрезки сообщений длины n , $n = N + 1, N + 2, \dots$, и имеющих скорость $R = H(X) + \epsilon$, причем для каждого кода этой последовательности вероятность ошибки не превосходит δ .

Теорема 1.8.2 (обратная теорема). Пусть $R < H(X)$, тогда существует зависящее от R положительное число δ такое, что для каждого кода, кодирующего дискретный источник без памяти со скоростью R , вероятность ошибки P_e не меньше, чем δ .

Кроме того, для любой последовательности кодов со скоростью R

$$\lim_{n \rightarrow \infty} P_{en} = 1, \quad (1.8.2)$$

где P_{en} — вероятность ошибки для кода, кодирующего отрезки сообщений длины n .

Доказательство. Докажем сначала справедливость соотношения (1.8.2). Пусть $\epsilon = H(X) - R > 0$ и $T_n(\epsilon/2)$, $n = 1, 2, \dots$, — последовательность высоковероятных множеств. Обозначим через T_n , $n = 1, 2, \dots$, последовательность произвольных множеств таких, что $T_n \subseteq X^n$ и

$$|T_n| = 2^{nR} = 2^{n[H(X) - \epsilon]}. \quad (1.8.3)$$

Для каждого n множество T_n будем рассматривать как множество однозначно кодируемым последовательностям сообщений. Поэтому (см. рис. 1.8.1)

$$1 - P_{en} \triangleq \Pr(T_n) = \Pr(T_n \cap \bar{T}_n(\epsilon/2)) + \Pr(T_n \cap T_n(\epsilon/2)) \leq \Pr(\bar{T}_n(\epsilon/2)) + \Pr(T_n \cap T_n(\epsilon/2)), \quad (1.8.4)$$

где $\bar{T}_n(\epsilon/2)$ — дополнение множества $T_n(\epsilon/2)$ до X^n и \cap — символ пересечения множеств. Из теоремы о высоковероятных множествах следует, что

$$\lim_{n \rightarrow \infty} \Pr(\bar{T}_n(\epsilon/2)) = 0, \quad \epsilon > 0. \quad (1.8.5)$$

Рассмотрим теперь вероятность $\Pr(T_n \cap T_n(\epsilon/2))$. Для всякой последовательности $x \in T_n(\epsilon/2) \cap T_n$ имеем $p(x) \leq 2^{-n[H(X) - \epsilon/2]}$, так как при этом $x \in T_n(\epsilon/2)$. Число элементов в множестве $T_n \cap T_n(\epsilon/2)$ не превосходит $|T_n| = 2^{nR}$, следовательно,

$$\Pr(T_n \cap T_n(\epsilon/2)) = \sum_{x \in T_n \cap T_n(\epsilon/2)} p(x) \leq 2^{nR} \cdot 2^{-n(H(X) - \epsilon/2)} = 2^{-n\epsilon/2}. \quad (1.8.6)$$

Учитывая (1.8.5) и (1.8.6), из (1.8.4) получим (1.8.2).

Докажем теперь первое утверждение теоремы. Для этого заметим, что при $R < H(X)$ множество \bar{T}_n не пусто при каждом n . Так как $p(x) \neq 0$ для всех $x \in X$, то каждая последовательность $x \in X^n$ имеет ненулевую вероятность. Следовательно, для каждого конечного n найдется такое $\delta(n) > 0$, что $P_{en} \geq \delta(n)$. С другой стороны, в силу (1.8.2) найдется такое N , что для всех $n > N$ выполняется неравенство $P_{en} > 1/2$. Полагая

$$\delta = \min\{\delta(1), \delta(2), \dots, \delta(N), 1/2\},$$

получим, что для любого n и любого T_n

$$P_{en} \geq \delta > 0. \quad (1.8.7)$$

Теорема доказана.

Таким образом, прямая и обратная теоремы кодирования показывают, что скорость создания информации при равномерном кодировании дискретного источника без памяти равна его энтропии. Этот вывод позволяет дать энтропии следующее толкование.

Представим себе, что источник подсоединен к некоторому устройству (кодеру источника), осуществляющему кодирование со скоростью R . Если кодовый алфавит содержит D символов, то в среднем на каждое сообщение источника будет появляться $R/\log D$ символов на выходе кодера. Для наилучшего кодера это количество будет весьма близким к $H(X)/\log D$. Следовательно, энтропия источника есть наименьшее количество двоичных символов на сообщение на выходе наилучшего двоичного кодера для этого источника при условии, что сообщения источника могут быть восстановлены по выходу кодера сколь угодно точно.

§ 1.9. Эргодические дискретные источники

При рассмотрении дискретных источников без памяти мы видели, что энтропия источника представляет собой количество информации, порождаемое источником в единицу времени. Основанием для такого вывода был закон больших чисел, из которого вытекало, что собственная информация в единицу времени $\frac{1}{n} I(\mathbf{x})$ с большой вероятностью близка к энтропии $H(X)$ при больших n . В более общем случае, когда сообщения источника не являются независимыми, закон больших чисел может оказаться неприменимым. Однако существует широкий класс стационарных источников, для которых величина $\frac{1}{n} I(\mathbf{x})$ при больших n близка в вероятностном смысле к энтропии на сообщение $H(X|X^\infty)$. Это верно для так называемых эргодических источников.

Рассмотрим вначале множество из m одинаковых, не зависящих друг от друга стационарных источников. Пусть все источники рассматриваются на отрезке времени длины k и X^k — ансамбль сообщений для каждого источника. Рассмотрим функцию

$$I(x^{(k)} | x^{(k-1)}, \dots, x^{(1)}) \triangleq -\log p(x^{(k)} | x^{(k-1)}, \dots, x^{(1)}) \quad (1.9.1)$$

(см. (1.4.20)), отображающую X^k в действительную ось. Для каждого источника эта функция представляет собой случайную величину, определенную на X^k . Так как предполагается, что источники независимы и одинаковы, то случайные величины $I_i(x^{(k)} | x^{(k-1)}, \dots, x^{(1)})$, $i = 1, 2, \dots, m$, независимы и одинаково распределены.

Пусть

$$I_{\text{ср}} \triangleq \frac{1}{m} \sum_{i=1}^m I_i(x^{(k)} | x^{(k-1)}, \dots, x^{(1)}) \quad (1.9.2)$$

— среднее арифметическое условной собственной информации сообщения $x^{(k)}$ для m источников. Математическое ожидание каждого слагаемого в сумме равно $H(X | X^{k-1})$. Поэтому по закону больших чисел для любых положительных ε и δ

$$\Pr(|I_{\text{ср}} - H(X | X^{k-1})| \geq \varepsilon) \leq \delta \quad (1.9.3)$$

при достаточно больших m . Другими словами, средняя условная информация $I_{\text{ср}}$, вычисленная по множеству источников (иногда говорят — по множеству реализаций, подразумевая, что каждый источник порождает одну реализацию последовательности сообщений), с большой вероятностью близка к условной энтропии $H(X | X^{k-1})$ при больших m .

Рассмотрим теперь отрезки сообщений длины k на выходе какого-нибудь одного источника (или отрезки длины k одной реализации). Каждому отрезку $x^{(i+1)}, \dots, x^{(i+k)}$, $i = 0, 1, \dots$, можно поставить в соответствие число

$$I(x^{(i+k)} | x^{(i+k-1)}, \dots, x^{(i+1)}) \triangleq -\log p(x^{(i+k)} | x^{(i+k-1)}, \dots, x^{(i+1)}) \quad (1.9.4)$$

— условную собственную информацию сообщения $x^{(i+k)}$. Среднее арифметическое этих информаций

$$I'_{\text{ср}} \triangleq \frac{1}{m} \sum_{i=0}^{m-1} I(x^{(i+k)} | x^{(i+k-1)}, \dots, x^{(i+1)}) \quad (1.9.5)$$

есть случайная величина, определенная на произведении $k+m-1$ ансамблей $X_1 X_2 \dots X_{k+m-1}$. Величина $I'_{\text{ср}}$ есть среднее арифметическое одинаково распределенных (в силу стационарности) случайных величин, которые не являются независимыми. Поэтому, вообще говоря, нет оснований полагать, что для любых положительных ε и δ

$$\Pr(|I'_{\text{ср}} - H(X | X^{k-1})| \geq \varepsilon) \leq \delta \quad (1.9.6)$$

при достаточно больших m . Если бы (1.9.6) действительно выполнялось, то

$$\Pr(|I_{\text{ср}} - I'_{\text{ср}}| \geq 2\varepsilon) \leq \delta \quad (1.9.7)$$

при достаточно больших m (см. задачу 1.9.1). В этом случае можно было бы утверждать, что среднее значение условных информаций, вычисленное по множеству источников, и среднее значение условных информаций, вычисленное по реализации одного источника, с большой вероятностью близки друг к другу при больших m .

Это рассмотрение можно обобщить, если на ансамбле X^k рассматривать произвольные функции $\varphi(x^{(1)}, \dots, x^{(k)})$ и определять для них среднее по множеству источников аналогично (1.9.2) и среднее по одной реализации аналогично (1.9.5). Класс источников, для которых оба этих средних в вероятностном смысле близки независимо от выбора k и функции $\varphi(\cdot)$, носит название класса эргодических источников. Поэтому предположение (1.9.6) для эргодических источников выполняется.

Теперь дадим более строгое определение эргодических источников.

Пусть U_X — стационарный источник, выбирающий сообщения из множества X , и $\dots, x^{(-1)}, x^{(0)}, x^{(1)}, x^{(2)}, \dots$ — последовательность сообщений на его выходе. Пусть $\varphi(x_1, \dots, x_k)$ — произвольная функция, определенная на множестве X^k и отображающая отрезки сообщений длины k в числовую ось. Пусть

$$z^{(i)} \triangleq \varphi(x^{(i+1)}, \dots, x^{(i+k)}), \quad i = 1, 2, \dots, \quad (1.9.8)$$

— последовательность случайных величин, имеющих в силу стационарности одинаковые распределения вероятностей. Обозначим через m_z математическое ожидание случайных величин $z^{(i)}$.

Определение 1.9.1. Дискретный стационарный источник называется *эргодическим*, если для любого k , любой действительной функции $\varphi(x_1, \dots, x_k)$, $M\varphi(\cdot) < \infty$, определенной на X^k , любых положительных ϵ и δ найдется такое N , что для всех $n > N$

$$\Pr\left(\left|\frac{1}{n} \sum_{i=1}^n z^{(i)} - m_z\right| \geq \epsilon\right) \leq \delta. \quad (1.9.9)$$

Пример 1.9.1. Если выбрать $\varphi(x^{(1)}, \dots, x^{(k)}) = -\log p(x^{(k)} | x^{(k-1)}, \dots, x^{(1)})$, то в случае дискретного эргодического источника предположение (1.9.6) будет выполняться.

Свойство эргодичности является весьма полезным не только в теоретико-информационных задачах. Благодаря эргодичности, многие характеристики источников могут быть найдены экспериментально с помощью наблюдения, может быть достаточно длительного, одной реализации последовательности сообщений. Действительно, наблюдая одну реализацию и вычисляя по ней величины $z^{(1)}, \dots, z^{(n)}$, можно достаточно хорошо оценить величину m_z , если в качестве оценки этой величины брать среднее $\frac{1}{n} \sum_{i=1}^n z^{(i)}$. Неравенство (1.9.9) показывает, что с достаточно малой вероятностью, величину которой можно задать заранее, оценка и истинное значение будут отличаться более чем на ϵ . Следующий пример иллюстрирует одну из таких задач оценивания.

Пример 1.9.2. Пусть дана реализация на выходе дискретного эргодического источника и требуется по этой реализации оценить вероятность, с которой источник порождает некоторое сообщение, например, $x_1 \in X$. Интуитивно ясно, что для этого достаточно подсчитать частоту появления этого сообщения в данной реализации. Однако точное обоснование того, почему нужно делать так, основано на применении свойства эргодичности. Положим $k = 1$ и

$$\varphi(x) = \begin{cases} 1, & \text{если } x = x_1, \\ 0, & \text{если } x \neq x_1. \end{cases} \quad (1.9.10)$$

Тогда $M\varphi(x) = \Pr(x = x_1)$. Очевидно, $\frac{1}{n} \sum_{i=1}^n z^{(i)}$ есть частота появления сообщения x_1 и (1.9.9) есть обоснование того, почему частота является хорошей оценкой вероятности.

Теорема 1.9.1. Всякий дискретный стационарный источник без памяти является эргодическим.

Доказательство. Зафиксируем k и некоторую функцию $\varphi(x_1, \dots, x_k)$, заданную на X^k , положим $n = k \cdot l$,

$$z^{(i)} = \varphi(x^{(i)}, \dots, x^{(i+k-1)}), \quad i = 1, \dots, n,$$

и рассмотрим сумму

$$\sum_{i=0}^{n-1} z^{(i)} = \sum_{j=1}^k \sum_{s=1}^l z^{((s-1)k+j)} = \sum_{j=1}^k w_j, \quad (1.9.11)$$

где

$$w_j \triangleq \sum_{s=1}^l z^{((s-1)k+j)}. \quad (1.9.12)$$

Заметим, что случайные величины $z^{(i)}$, $i = 1, 2, \dots, n$, не являются независимыми, хотя сообщения рассматриваемого источника независимы. Однако случайные величины $z^{((s-1)k+j)}$, $s = 1, 2, \dots, l$, независимы, так как соответствующие им функции не имеют общих аргументов. Поэтому w_j для каждого j есть сумма независимых одинаково распределенных случайных величин.

Используя (1.9.11), можно получить следующие соотношения:

$$\begin{aligned} \Pr\left(\left|\frac{1}{n} \sum_{i=1}^n z^{(i)} - m_z\right| \geq \epsilon\right) &= \Pr\left(\left|\frac{1}{n} \sum_{j=1}^k w_j - m_z\right| \geq \epsilon\right) = \\ &= \Pr\left(\left|\frac{1}{k} \sum_{j=1}^k \left(\frac{1}{l} w_j - m_z\right)\right| \geq \epsilon\right) \leq \\ &\leq \Pr\left(\left|\frac{1}{l} w_j - m_z\right| \geq \epsilon \text{ хотя бы при одном } j\right), \quad (1.9.13) \end{aligned}$$

где неравенство следует из того, что событие, состоящее в том, что среднее арифметическое k величин больше ε , влечет событие, состоящее в том, что хотя бы одно слагаемое в сумме больше ε . Поскольку случайные величины w_j , $j = 1, \dots, k$, одинаково распределены, то вероятности $\Pr\left(\left|\frac{1}{l}w_j - m_z\right| \geq \varepsilon\right)$ одинаковы для всех j и, следовательно, правую часть (1.9.13) можно оценить сверху величиной

$$k \cdot \Pr\left(\left|\frac{1}{l}w_j - m_z\right| \geq \varepsilon\right) = k \cdot \Pr\left(\left|\frac{1}{l} \sum_{s=1}^l z^{((s-1)k+j)} - m_z\right| \geq \varepsilon\right). \quad (1.9.14)$$

Случайные величины, стоящие под знаком суммы в (1.9.14), таковы, что можно применить закон больших чисел, согласно которому при достаточно большом l , а следовательно, при фиксированном k и достаточно большом $n = kl$, правая часть (1.9.14) будет не большей, чем любое заданное наперед положительное число δ . Теорема доказана.

Замечание. В теореме показано, что стационарность и независимость являются достаточными условиями эргодичности. Метод доказательства теоремы 1.9.1 подсказывает более слабое условие эргодичности. Используя этот метод, можно показать, что эргодичным является всякий дискретный стационарный источник, сообщения на выходе которого, разделенные интервалом в i , $i \geq N$, единиц времени, независимы для некоторого ограниченного N .

Полезно разобрать пример стационарного неэргодического источника с тем, чтобы на этом примере более детально познакомиться с природой эргодичности.

Пример 1.9.3. Пусть имеются две урны. В первой урне лежат два белых и один черный шар. Во второй урне лежат два черных и один белый шар. Производится следующий эксперимент. Пусть случайно выбирается одна из двух урн, причем первая выбирается с вероятностью p . Затем из выбранной урны осуществляется последовательное извлечение шаров с возвращением. Экспериментатор подсчитывает число извлечений белого шара. Если в начале эксперимента выбрана первая урна, то частота появления белого шара, зарегистрированная экспериментатором, будет равна $\frac{2}{3}$, в противном случае — $\frac{1}{3}$. Если имеется n независимо работающих экспериментаторов, то примерно np из них зарегистрируют частоту $\frac{2}{3}$, а примерно $n(1-p)$ из них — частоту $\frac{1}{3}$. Частота v появления белого шара в среднем по множеству всех экспериментаторов равна

$$v \equiv p \cdot \frac{2}{3} + (1-p) \cdot \frac{1}{3} = \frac{1+p}{3}. \quad (1.9.15)$$

Если число p отлично от нуля или от единицы, то число v не совпадает с частотой, вычисленной по одной реализации (полученной одним экспериментатором). Следовательно, источник сообщений, соответствующий описанному эксперименту, не является эргодическим.

Структуру неэргодического источника примера 1.9.3 можно представить себе следующим образом (см. рис. 1.9.1). Имеются два эргодических (в данном примере — не имеющих памяти) источника U_1 , U_2 и переключатель P , который может находиться в одном из двух положений. Переключатель случайно устанавливается в одно из положений при включении источника и остается в этом положении на протяжении всей работы. Экспериментатор имеет дело с одной реализацией на выходе источника, которой может соответствовать только одно положение переключателя. Наблюдая эту реализацию, экспериментатор не может вынести решение о том, с каким источником, эргодическим или нет, он имеет дело, так как ему недоступны другие возможные реализации. Можно показать, что всякий неэргодический источник можно представить в виде комбинации эргодических источников (эргодических компонент), аналогичной той, которая показана на рис. 1.9.1. В общем случае число компонент может быть большим, даже бесконечно большим.

Вернемся к рассмотрению количества информации, порождаемого эргодическим источником в единицу времени. Естественно ожидать, что это количество с большой вероятностью должно быть близким к энтропии на сообщение $H(X|X^\infty)$. Для того чтобы в этом убедиться, необходимо доказать, что для любых $\varepsilon > 0$ и $\delta > 0$ найдется N такое, что при $n > N$

$$\Pr\left(\left|\frac{1}{n}I(x) - H(X|X^\infty)\right| \geq \varepsilon\right) < \delta. \quad (1.9.16)$$

На первый взгляд может показаться, что (1.9.16) прямо следует из определения эргодического источника. Однако это не так.

Информацию на сообщение можно представить следующим образом:

$$I(x) = \frac{1}{n} \sum_{i=1}^n I(x^{(i)} | x^{(i-1)}, \dots, x^{(1)}). \quad (1.9.17)$$

Как легко видеть, под знаком суммы стоят случайные величины, полученные с помощью различных функций $\Phi(\cdot)$ (первое слагаемое является функцией одной переменной, второе — двух и т. д.), тогда как в определении эргодического источника (1.9.9) суммируются случайные величины, полученные с помощью одной и той же функции. Кроме того, $H(X|X^\infty)$ не является математическим ожиданием ни одного слагаемого под знаком суммы. Вместе с тем

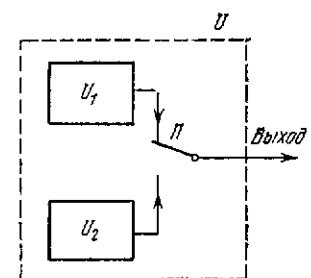


Рис. 1.9.1. Неэргодический источник с двумя эргодическими компонентами.

неравенство (1.9.16) справедливо для любого эргодического источника при достаточно большом n . Этот факт устанавливается в так называемой лемме Мак-Миллана.

Лемма 1.9.1 (Мак-Миллан). Для любого дискретного эргодического источника любых, положительных ε и δ найдется N такое, что для всех $n > N$ выполняется неравенство (1.9.16).

Доказательство. Пусть $p(\mathbf{x}) = p(x^{(1)}, \dots, x^{(n)})$, $\mathbf{x} \in X^n$, — распределение вероятностей для последовательностей сообщений дискретного стационарного эргодического источника. Определим для любого целого m , $1 \leq m \leq n$, функцию

$$Q(\mathbf{x}) \triangleq p(x^{(1)}, \dots, x^{(m)}) \prod_{i=1}^{n-m} p(x^{(m+i)} | x^{(m+i-1)}, \dots, x^{(i)}). \quad (1.9.18)$$

Согласно этому определению $Q(\mathbf{x})$ является аппроксимацией распределения вероятностей $p(\mathbf{x})$ источника, при которой учитывается статистическая зависимость каждого сообщения лишь от m предшествующих ему сообщений. Отметим, что $\sum_{\mathbf{x}^n} Q(\mathbf{x}) = 1$.

Это легко проверяется суммированием правой части (1.9.18) вначале по $x^{(n)}$, затем по $x^{(n-1)}$ и в конце — по $(x^{(1)}, \dots, x^{(m)})$.

Установим теперь простое свойство аппроксимирующего распределения $Q(\mathbf{x})$, вытекающее из эргодичности. Покажем, что для любых $\varepsilon > 0$ и $\delta > 0$ найдется N такое, что для всех $n > N$ имеет место неравенство

$$\Pr\left(\left| -\frac{1}{n} \log Q(\mathbf{x}) - H(X | X^m) \right| \geq \varepsilon \right) \leq \delta. \quad (1.9.19)$$

Для того чтобы это доказать, заметим, что для любых двух случайных величин ξ_1 и ξ_2 и любого $\varepsilon > 0$ имеет место неравенство

$$\Pr(|\xi_1 + \xi_2| \geq 2\varepsilon) \leq \Pr(|\xi_1| \geq \varepsilon) + \Pr(|\xi_2| \geq \varepsilon), \quad (1.9.20)$$

которое следует из того, что если $|\xi_1 + \xi_2| \geq 2\varepsilon$, то либо $|\xi_1| \geq \varepsilon$, либо $|\xi_2| \geq \varepsilon$. Заметим также, что

$$\begin{aligned} \frac{1}{n} \log Q(\mathbf{x}) &= \\ &= \frac{1}{n} \log p(x^{(1)}, \dots, x^{(m)}) + \frac{1}{n} \sum_{i=1}^{n-m} \log p(x^{(m+i)} | x^{(m+i-1)}, \dots, x^{(i)}). \end{aligned} \quad (1.9.21)$$

Так как $-M \log p(x^{(m+i)} | x^{(m+i-1)}, \dots, x^{(i)}) = H(X | X^m)$ для всех i , то согласно предположению об эргодичности для любых $\varepsilon > 0$ и $\delta > 0$ найдется N такое, что для всех $n > N$

$$\begin{aligned} \Pr\left(\left| \frac{1}{n} \log Q(\mathbf{x}) + H(X | X^m) \right| \geq \varepsilon \right) &= \\ &= \Pr\left(\left| \frac{1}{n} \sum_{i=1}^{n-m} \log p(x^{(m+i)} | x^{(m+i-1)}, \dots, x^{(i)}) + \right. \right. \\ &\quad \left. \left. + \frac{1}{n} \log p(x^{(1)}, \dots, x^{(m)}) + H(X | X^m) \right| \geq \varepsilon \right) \leq \\ &\leq \Pr\left(\left| \frac{1-m/n}{n-m} \sum_{i=1}^{n-m} \log p(x^{(m+i)} | x^{(m+i-1)}, \dots, x^{(i)}) + \right. \right. \\ &\quad \left. \left. + H(X | X^m) \right| \geq \frac{\varepsilon}{2} \right) + \Pr\left(\left| \frac{1}{n} \log p(x^{(1)}, \dots, x^{(m)}) \right| \geq \frac{\varepsilon}{2} \right) \leq \\ &\leq \frac{\delta}{2} + \frac{\delta}{2} = \delta, \end{aligned}$$

где первое неравенство — следствие неравенства (1.9.20), а второе — следствие эргодичности и того, что $\log p(x^{(1)}, \dots, x^{(m)})$ принимает ограниченные значения с вероятностью 1. Таким образом, утверждение, связанное с неравенством (1.9.19), доказано.

Покажем далее, что распределение $Q(\mathbf{x})$ достаточно хорошо аппроксимирует исходное распределение $p(\mathbf{x})$, а именно, что для любых $\varepsilon > 0$ и $\delta > 0$ найдется такое N , что при достаточно большом m для всех $n > N$ выполняется неравенство

$$q \triangleq \Pr\left(\left| \frac{1}{n} \log Q(\mathbf{x}) - \frac{1}{n} \log p(\mathbf{x}) \right| \geq \varepsilon \right) \leq \delta. \quad (1.9.22)$$

Для доказательства этого утверждения воспользуемся неравенством Чебышева в форме, приведенной в задаче 1.2.4. Имеем

$$q = \Pr\left(\left| \log \frac{Q(\mathbf{x})}{p(\mathbf{x})} \right| \geq n\varepsilon\right) \leq \frac{1}{n\varepsilon} M \left| \log \frac{Q(\mathbf{x})}{p(\mathbf{x})} \right|. \quad (1.9.23)$$

Чтобы оценить математическое ожидание, стоящее в правой части этого неравенства, применим неравенство для логарифма (1.3.7), согласно которому можно записать, что $\log x \leq x \log e$ для всех $x \geq 0$, и, следовательно,

$$|\log x| \leq 2x \log e - \log x. \quad (1.9.24)$$

Неравенство (1.9.24) легко проверяется посредством отдельного рассмотрения случаев $x \geq 1$ и $x < 1$. Отсюда следует, что

$$\begin{aligned} M \left| \log \frac{Q(x)}{p(x)} \right| &< 2 \log e M \frac{Q(x)}{p(x)} - M \log \frac{Q(x)}{p(x)} = \\ &= 2 \log e \sum_{X^n} p(x) \frac{Q(x)}{p(x)} - \sum_{X^n} p(x) \log Q(x) + \sum_{X^n} p(x) \log p(x) = \\ &= 2 \log e + H(X^m) + (n-m)H(X|X^m) - H(X^n), \end{aligned}$$

и из (1.9.23) вытекает, что

$$q \leq \frac{2 \log e}{\varepsilon n} + \frac{H(X^m)}{\varepsilon n} + \frac{1}{\varepsilon} \left(1 - \frac{m}{n} \right) H(X|X^m) - \frac{1}{\varepsilon n} H(X_n^e). \quad (1.9.25)$$

Из теорем 1.5.1, 1.5.2 следует, что при любых фиксированных ε и δ , при достаточно большом, но фиксированном m , правая часть неравенства (1.9.25) стремится к нулю при $n \rightarrow \infty$. Поэтому находится N такое, что при всех $n > N$ будет выполняться неравенство $q \leq \delta$.

Для завершения доказательства леммы следует показать, что для любых $\varepsilon > 0$ и $\delta > 0$ найдется N такое, что при $n > N$

$$\Pr \left(\left| -\frac{1}{n} \log p(x) - H(X|X^\infty) \right| \geq \varepsilon \right) \leq \delta.$$

Для заданных ε и δ выберем m и N такими большими, чтобы при всех $n > N$ выполнялись неравенства

$$q \leq \delta/2,$$

$$|H(X|X^m) - H(X|X^\infty)| \leq \varepsilon/3,$$

$$\Pr \left(\left| -\frac{1}{n} \log Q(x) - H(X|X^m) \right| \geq \frac{\varepsilon}{3} \right) \leq \frac{\delta}{2}.$$

Тогда из утверждений, связанных с неравенствами (1.9.19) и (1.9.22), будет следовать, что

$$\begin{aligned} \Pr \left(\left| -\frac{1}{n} \log p(x) - H(X|X^\infty) \right| \geq \varepsilon \right) &= \\ &= \Pr \left(\left| -\frac{1}{n} \log p(x) + \frac{1}{n} \log Q(x) - \frac{1}{n} \log Q(x) - \right. \right. \\ &\quad \left. \left. - H(X|X^m) + H(X|X^m) - H(X|X^\infty) \right| \geq \varepsilon \right) \leq \\ &\leq \Pr \left(\left| -\frac{1}{n} \log p(x) + \frac{1}{n} \log Q(x) \right| \geq \frac{\varepsilon}{3} \right) + \\ &\quad + \Pr \left(\left| -\frac{1}{n} \log Q(x) - H(X|X^m) \right| \geq \frac{\varepsilon}{3} \right) \leq \frac{\delta}{2} + \frac{\delta}{2} = \delta. \end{aligned}$$

Лемма доказана.

Из леммы Мак-Миллана может быть легко выведена теорема о высоковероятных множествах, а также прямая и обратная теоремы кодирования при равномерном кодировании дискретного эргодического источника.

Теорема 1.9.2 (теорема о высоковероятных множествах). Пусть U_X — дискретный эргодический источник и $\{X^n, p(x)\}$, $n = 1, 2, \dots$, — ансамбли последовательностей длины n на его выходе. Пусть $H(X|X^\infty)$ — энтропия на сообщение и ε — некоторое положительное число. Пусть $T_n(\varepsilon)$ — подмножество множества X^n , определенное для каждого n следующим образом:

$$T_n(\varepsilon) = \left\{ x: H(X|X^\infty) - \varepsilon \leq \frac{1}{n} I(x) \leq H(X|X^\infty) + \varepsilon \right\}, \quad (1.9.26)$$

где $I(x) = -\log p(x)$ — собственная информация последовательности $x \in X^n$. Тогда для любых $\varepsilon > 0$ и $\delta > 0$ найдется N такое, что для всех $n > N$ выполняются следующие неравенства:

$$\Pr(x \in T_n(\varepsilon)) \geq 1 - \delta, \quad (1.9.27)$$

$$(1 - \delta) 2^n (H(x|X^\infty) - \varepsilon) \leq |T_n(\varepsilon)| \leq 2^n (H(x|X^\infty) + \varepsilon). \quad (1.9.28)$$

Доказательство этой теоремы в точности повторяет доказательство теоремы 1.7.1 и поэтому опускается.

Теорема 1.9.3 (прямая теорема кодирования). Пусть $R > H(X|X^\infty)$, тогда для любого положительного δ существует код со скоростью R , который кодирует дискретный эргодический источник с вероятностью ошибки, не превышающей δ .

Теорема 1.9.4 (обратная теорема кодирования). Пусть $R < H(X|X^\infty)$, тогда существует зависящее от R положительное число δ такое, что для каждого кода со скоростью R , кодирующего дискретный эргодический источник, вероятность ошибки не меньше δ . Более того, для любой последовательности кодов со скоростью R

$$\lim_{n \rightarrow \infty} P_{en} = 1, \quad (1.9.29)$$

где P_{en} — вероятность ошибки для кода, кодирующего отрезки сообщений длины n .

Доказательства прямой и обратной теорем практически не отличаются от доказательств соответствующих теорем для дискретных источников без памяти.

Из этих теорем следует, что скорость создания информации дискретным эргодическим источником при равномерном кодировании равна его энтропии на сообщение $H(X|X^\infty)$. Это утверждение позволяет дать такое же толкование энтропии на сообщение для эргодических источников, как энтропии для постоянных

источников: и то и другое есть наименьшее количество двоичных символов, приходящихся на одно сообщение на выходе наилучшего двоичного кодера, при условии, что сообщения могут быть восстановлены по выходу кодера сколь угодно точно.

§ 1.10. Постановка задачи неравномерного кодирования дискретных источников. Коды с однозначным декодированием

В предыдущих параграфах мы познакомились с задачей равномерного кодирования дискретных источников и показали, что при таком кодировании скорость создания информации равна энтропии источника на сообщение. Здесь будет рассмотрен другой подход к задаче кодирования.

Для того чтобы его пояснить, вернемся к примеру 1.6.4. В этом примере были разобраны два метода кодирования телеграмм. Один метод, так называемый метод побуквенного кодирования, позволяет закодировать и затем безошибочно декодировать любую телеграмму, затрачивая на одну букву 6 двоичных символов. Второй метод оказался ключевым для задачи равномерного кодирования. Он позволяет однозначно кодировать не все телеграммы, а только некоторые типичные для данного источника. При таком кодировании на одну букву будет затрачиваться примерно 13/8 двоичных символов — существенно меньше, чем при побуквенном кодировании.

При кодировании телеграмм есть еще один метод кодирования, о котором не было ничего сказано раньше. Предположим, что известно, с какими вероятностями появляются те или другие буквы в телеграммах. Тогда можно использовать неравномерный код, в котором длины кодовых слов подобраны так, что часто встречающиеся буквы кодируются относительно короткими двоичными словами, а редкие — длинными. При таком способе побуквенного кодирования, использующем неравномерные коды, любая телеграмма может быть однозначно закодирована и однозначно декодирована, а среднее количество двоичных символов на букву может быть сделано меньшим, чем в случае равномерного побуквенного кодирования.

Как и для любого метода кодирования источника, основной характеристикой неравномерного кодирования является количество символов, затрачиваемых на кодирование одного сообщения. В случае равномерного кодирования это количество одно и то же для любого сообщения источника. Действительно, каждый отрезок из n сообщений при равномерном кодировании отображается в последовательность $\frac{\log M}{\log D} D$ -ичных кодовых символов, представляющую собой D -ичную запись номера кодируемого отрезка. При неравномерном кодировании это не так. Количество символов,

затрачиваемых при кодировании, зависит от сообщений источника и поэтому представляет собой случайную величину. В этом случае разумной мерой качества кодирования является среднее количество символов на сообщение.

Обозначим через m_i длину слова, кодирующего сообщение $x_i \in X$. Пусть $p(x_i)$ — вероятность этого сообщения, тогда

$$\bar{m}(X) \triangleq \sum_{x_i \in X} m_i p(x_i) \quad (1.10.1)$$

будет средней длиной кодовых слов, кодирующих ансамбль сообщений $\{X, p(x)\}$. Предположим, что неравномерный код используется для кодирования отрезков сообщений длины n , т. е. для кодирования ансамбля $\{X^n, p(x)\}$, где $p(x)$ — распределение вероятностей на X^n , задаваемое с помощью задания источника. Пусть $\bar{m}(X^n)$ — средняя длина кодовых слов.

Определение 1.10.1. Число

$$R \triangleq \frac{\bar{m}(X^n) \log D}{n} \quad (1.10.2)$$

называется *средней скоростью неравномерного кодирования посредством D -ичного кода при разбиении последовательности сообщений на блоки длины n* : Средняя скорость неравномерного кодирования измеряется в двоичных символах на сообщение.

Пример 1.10.1. Предположим, что $n = 1$ и источник порождает в каждый момент времени одно из восьми сообщений x_1, \dots, x_8 . Вероятности сообщений приведены во втором столбце табл. 1.10.1.

Таблица 1.10.1

x_i	$p(x_i)$	Равномерный код	Неравномерный код	m_i
x_1	1/4	0 0 0	0 0	2
x_2	1/4	0 0 1	0 1	2
x_3	1/8	0 1 0	1 0 0	3
x_4	1/8	0 1 1	1 0 1	3
x_5	1/16	1 0 0	1 1 0 0	4
x_6	1/16	1 0 1	1 1 0 1	4
x_7	1/16	1 1 0	1 1 1 0	4
x_8	1/16	1 1 1	1 1 1 1	4

Заметим, что энтропия ансамбля сообщений этого примера равна 2,75 бит. В таблице приведены два двоичных кода — равномерный и неравномерный, которые однозначно кодируют указанное множество сообщений. Для первого кода все слова имеют одинаковую длину, равную 3. Для второго — слова имеют различные длины и, как нетрудно подсчитать, средняя длина равна 2,75 символов (совпадение с энтропией неслучайно). Оба кода осуществляют побуквенное кодирование. Скорость кодирования в первом случае равна 3 бит/сообщение, а во втором — 2,75 бит/сообщение.

Средняя скорость неравномерного кодирования зависит от выбора n и множества кодовых слов. Наша цель состоит в определении наименьшей достижимой средней скорости. Прежде чем перейти к решению этой задачи, необходимо ответить на следующий вопрос: любой ли неравномерный код пригоден для кодирования сообщений источника, и если нет, то какие ограничения на него должны быть наложены? Покажем существо этого вопроса следующим примером.

Пример 1.10.2. Предположим, что источник порождает сообщения x_1, x_2, x_3, x_4 и эти сообщения кодируются следующим образом: $x_1 \rightarrow 0, x_2 \rightarrow 01, x_3 \rightarrow 10, x_4 \rightarrow 011$. Кодовый алфавит состоит из двух символов 0 и 1. Пусть на выходе источника появилась следующая последовательность сообщений $x_2 x_3 x_2 x_1 \dots$. Кодер вырабатывает для каждого сообщения свое кодовое слово. На его выходе возникает последовательность 0110100... Нет специального разделительного знака пробела между словами. Если бы он был, то следовало бы рассматривать кодовый алфавит, состоящий из 0, 1 и пробела, что привело бы к задаче кодирования с кодовым алфавитом из трех символов. Декодеру известно лишь начало указанной выше двоичной последовательности, но неизвестно, с какого символа начинается второе кодовое слово, третье и т. д. Оказывается, что эта последовательность допускает несколько различных способов декодирования: $(x_2, x_3, x_2, x_1, \dots)$ — правильное декодирование, $(x_4, x_2, x_1, x_1, \dots), (x_4, x_1, x_3, x_1, \dots)$ — неправильные декодирования. В этом примере причиной неоднозначности декодирования является то, что кодовое слово 0 является началом слова 01, а это слово в свою очередь является началом слова 011. Коды примера 1.10.1 позволяют декодировать однозначно. Для равномерного кода это так потому, что все слова имеют одинаковую длину; последовательно отсчитывая по три символа, мы будем получать кодовые слова. Для неравномерного кода этого примера ни одно слово не является началом другого, и поэтому разбиение на кодовые слова происходит однозначно.

Коды, в которых ни одно слово не является началом другого, называются *префиксными*. Коды, в которых любая последовательность кодовых слов допускает однозначное разбиение на кодовые слова, называются *кодами со свойством однозначного декодирования*. Префиксные коды являются кодами со свойством однозначного декодирования, но не наоборот.

Пример 1.10.3. Код, образованный словами 0, 01, 011, не является префиксным, но является однозначно декодируемым. Покажите это самостоятельно.

Таким образом, для кодирования сообщений источника пригодны только те коды, которые допускают однозначное декодирование.

Определение 1.10.2. Скоростью создания информации дискретным источником при неравномерном кодировании называется наименьшее число H такое, что для любого $R > H$ найдется n (длина кодируемых сообщений) и неравномерный код со средней скоростью кодирования R , который допускает однозначное декодирование.

Ниже мы покажем, что скорость создания информации при неравномерном кодировании совпадает со скоростью создания

информации при равномерном кодировании и равна энтропии источника на сообщение.

Для того чтобы доказать, что число H есть скорость создания информации при неравномерном кодировании, нужно, как и раньше, доказывать прямую и обратную теоремы. Первая из них будет утверждать, что для всех $R > H$ найдется n и однозначно декодируемый неравномерный код со средней скоростью R , а вторая будет утверждать, что для любого $R < H$ не существует однозначно декодируемого кода ни при каком n .

Прежде чем приступить к детальному изучению задачи неравномерного кодирования, докажем теорему, в которой формулируется необходимое условие того, чтобы код был однозначно декодируем.

Теорема 1.10.1. Предположим, что однозначно декодируемый код состоит из M слов, длины которых равны m_1, \dots, m_M , и кодовый алфавит содержит D символов. Тогда

$$\sum_{i=1}^M D^{-m_i} \leq 1. \quad (1.10.3)$$

Доказательство. Пусть L — произвольное положительное целое число. Имеет место следующее равенство:

$$\left(\sum_{i=1}^M D^{-m_i} \right)^L = \sum_{i_1=1}^M \dots \sum_{i_L=1}^M D^{-(m_{i_1} + \dots + m_{i_L})}. \quad (1.10.4)$$

Заметим, что в выражении, стоящем в правой части равенства (1.10.4), каждое слагаемое соответствует каждой возможной последовательности из L кодовых слов. Сумма $m_{i_1} + \dots + m_{i_L}$ равна суммарной длине соответствующей последовательности кодовых слов. Если через A_j обозначить число последовательностей из L кодовых слов, имеющих суммарную длину j , то (1.10.4) можно переписать следующим образом:

$$\left(\sum_{i=1}^M D^{-m_i} \right)^L = \sum_{j=1}^{Lm} A_j D^{-j}, \quad (1.10.5)$$

где m — максимальное из чисел m_1, \dots, m_M . Код однозначно декодируем, если при любом L и любом j имеется единственная последовательность кодовых символов длины j , образованная L кодовыми словами. Так как D^j — максимальное количество различных последовательностей длины j , то $A_j \leq D^j$. Подставляя это неравенство в (1.10.5), получим

$$\left(\sum_{i=1}^M D^{-m_i} \right)^L \leq L \cdot m^L \quad (1.10.6)$$

или

$$\sum_{i=1}^M D^{-m_i} \leq (Lm)^{1/L} = 2^{\frac{1}{L} \log m L}. \quad (1.10.7)$$

Неравенство (1.10.7) справедливо при всех положительных целых L . Переходя к пределу при $L \rightarrow \infty$, получим утверждение теоремы.

Таким образом, мы получили необходимое условие того, чтобы код обладал свойством однозначного декодирования. Ниже будут изучены коды, для которых легко показать, что это условие является также достаточным.

§ 1.11. Кодовые деревья. Неравенство Крафта

Как было сказано выше, префиксные коды обладают свойством однозначного декодирования. В таких кодах ни одно кодовое слово не является началом другого. Удобное описание префиксных кодов дают специальные графы, называемые деревьями (или кодовыми деревьями). D -ичным деревом называется граф, т. е. такая система узлов и связывающих их ребер, в котором нет петель или замкнутых путей и в котором из каждого узла выходит не более D ребер и в каждый узел, кроме одного (корня дерева), входит точно одно ребро. Каждому из ребер, выходящих из узла, сопоставляется один символ кодового алфавита, содержащего D символов, причем различным ребрам, выходящим из одного узла, сопоставляются различные символы.

На рис. 1.11.1 и 1.11.2 показаны двоичные кодовые деревья, соответствующие кодам примеров 1.10.1 и 1.10.2. Под рисунком

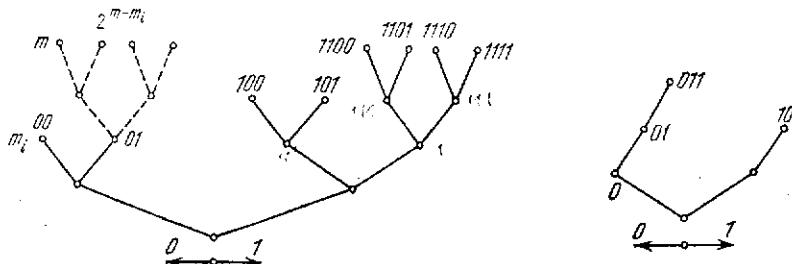


Рис. 1.11.1. Двоичное кодовое дерево для кода примера 1.10.1.

Рис. 1.11.2. Двоичное кодовое дерево для кода примера 1.10.2.

дерева показано правило сопоставления ребер и двоичных символов: каждому правому ребру сопоставляется 1, каждому левому — 0. На рис. 1.11.3 приведен пример троичного дерева, соответствующего троичному коду 00, 01, 02, 10, 11, 12, 20, 210, 220, 221, 222.

Узлы дерева, отстоящие от корня на i ребер, образуют ярус порядка i . Порядком узла называют номер его яруса. Порядком дерева называют максимальный из порядков его узлов. Узел, из которого не выходит ни одного ребра, называется концевым.

Код является префиксным, если кодовые слова соответствуют только концевым узлам дерева. В противном случае код не является префиксным.

Теорема 1.11.1 (неравенство Крафта). Для того чтобы существовал префиксный код в алфавите объема D с длинами кодовых слов m_1, \dots, m_M , необходимо и достаточно, чтобы

$$\sum_{i=1}^M D^{m-m_i} < 1. \quad (1.11.1)$$

Доказательство. Для доказательства следует показать, что условие (1.11.1) является необходимым и достаточным условием существования дерева, концевые узлы которого имеют порядки m_1, \dots, m_M .

Рассмотрим вначале необходимость. Заметим, что необходимость условия 1.11.1 для всех кодов, а не только для префиксных, была установлена в теореме 1.10.1.

Тем не менее, мы снова докажем необходимость для префиксных кодов, поскольку доказательство является очень простым, но поучительным. Предположим, что существует кодовое дерево, концевые узлы которого имеют указанные порядки. Покажем, что при этом выполняется неравенство (1.11.1). Заметим, что максимальное количество узлов на ярусе j равно D^j . Пусть $m = \max\{m_1, \dots, m_M\}$. Рассмотрим концевой узел порядка m_i . Этот узел отстоит от яруса m на $m-m_i$ ребер и, следовательно, исключает из этого яруса D^{m-m_i} возможных узлов (см. рис. 1.11.1).

Так как количество узлов, исключаемых из яруса m всеми концевыми узлами порядков m_1, \dots, m_M , не может превосходить максимального количества узлов на этом ярусе, то

$$\sum_{i=1}^M D^{m-m_i} \leq D^m. \quad (1.11.2)$$

Отсюда после деления обеих частей неравенства на D^m следует (1.11.1).

Докажем теперь достаточность. Для этого следует доказать, что при выполнении (1.11.1) дерево с концевыми узлами порядков m_1, \dots, m_M может быть построено. Предположим, что среди

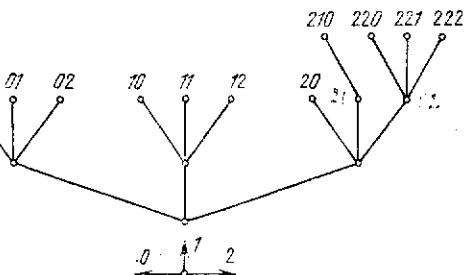


Рис. 1.11.3. Троичное кодовое дерево.

этого набора порядков число s встречается точно α_s раз, $s = 1, 2, \dots, m$. Тогда

$$\sum_{i=1}^M D^{-m_i} = \sum_{s=1}^m \alpha_s D^{-s} < 1. \quad (1.11.3)$$

Перепишем это неравенство следующим образом:

$$\sum_{s=1}^{i-1} \alpha_s D^{-s} + \alpha_i D^{-i} + \sum_{s=i+1}^m \alpha_s D^{-s} < 1. \quad (1.11.4)$$

Тогда

$$\alpha_i < D^i - \sum_{s=1}^{i-1} \alpha_s D^{i-s} - \sum_{s=i+1}^m \alpha_s D^{i-s} < D^i - \sum_{s=1}^{i-1} \alpha_s D^{i-s}. \quad (1.11.5)$$

Применим метод полной индукции. Дерево, содержащее α_1 концевых узлов порядка 1, может быть построено. Действительно, из (1.11.3) следует, что $\alpha_1 D^{-1} < 1$, т. е. $\alpha_1 < D$. Так как максимальное возможное количество концевых узлов порядка 1 есть D , а $\alpha_1 < D$, то дерево с α_1 концевыми узлами порядка 1 может быть построено.

Предположим, что дерево с α_s концевыми узлами порядка s , $s = 1, 2, \dots, i-1$, может быть построено. Докажем, что к этому дереву можно добавить еще α_i концевых узлов порядка i . Если верно предположение индукции, то из яруса порядка i исключаются $\sum_{s=1}^{i-1} \alpha_s D^{i-s}$ возможных концевых узлов. Так как максимальное количество возможных концевых узлов в этом ярусе равно D^i , то $D^i - \sum_{s=1}^{i-1} \alpha_s D^{i-s}$ есть количество свободных узлов на ярусе i . Из (1.11.5) следует, что количество α_i узлов на ярусе i , которые должны быть добавлены, не превосходит количества свободных узлов. Следовательно, к дереву с α_s концевыми узлами порядка s , $s = 1, 2, \dots, i-1$, могут быть добавлены α_i концевых узлов порядка i . Теорема доказана.

§ 1.12. Неравномерное кодирование дискретных стационарных источников

В этом параграфе мы получим теоремы кодирования дискретного стационарного источника при неравномерном кодировании. Вначале будут доказаны вспомогательные утверждения, относящиеся к побуквенному кодированию произвольных ансамблей сообщений. В дальнейшем они будут применены к ансамблю последовательностей сообщений.

Пусть $\{X, p(X)\}$, $X = \{x_1, \dots, x_M\}$, — произвольный дискретный ансамбль сообщений и $H(X)$ — его энтропия. Обозна-

чим: $\bar{m}(X)$ — средняя длина D -ичного кода, $\bar{m}(X) = \sum_{i=1}^M m_i p(x_i)$, слова которого сопоставляются сообщениям ансамбля X .

Теорема 1.12.1. Для любого кода со свойством однозначного декодирования

$$\bar{m}(X) \geq \frac{H(X)}{\log D}. \quad (1.12.1)$$

Доказательство. Из определения средней длины кодовых слов имеем

$$\bar{m}(X) \log D = \sum_{i=1}^M p(x_i) m_i \log D = \sum_{i=1}^M p(x_i) \log D^{m_i}. \quad (1.12.2)$$

Рассмотрим разность $H(X) - \bar{m}(X) \log D$:

$$\begin{aligned} H(X) - \bar{m}(X) \log D &= - \sum_{i=1}^M p(x_i) \log p(x_i) + \sum_{i=1}^M p(x_i) \log D^{-m_i} = \\ &= \sum_{i=1}^M p(x_i) \log \frac{D^{-m_i}}{p(x_i)}. \end{aligned} \quad (1.12.3)$$

Воспользуемся неравенством (1.3.7). В результате получим, что $H(X) - \bar{m}(X) \log D <$

$$\leq \log e \sum_{i=1}^M p(x_i) \left(\frac{D^{-m_i}}{p(x_i)} - 1 \right) = \log e \left(\sum_{i=1}^M D^{-m_i} - 1 \right) \leq 0, \quad (1.12.4)$$

где последнее неравенство является следствием того, что код обладает свойством однозначного декодирования (см. теорему 1.10.1). Теорема доказана.

Замечание. Первое неравенство в (1.12.4), так же как и второе, обращается в точное равенство, если

$$p(x_i) = D^{-m_i}, \quad i = 1, 2, \dots, M. \quad (1.12.5)$$

Таким образом, если вероятности сообщений являются целыми отрицательными степенями числа D (1.12.5), то существует D -ичное дерево с концевыми узлами порядков m_1, \dots, m_M (теорема 1.11.1) и соответствующий D -ичный код будет иметь среднюю длину

$$\bar{m}(X) = \frac{H(X)}{\log D}, \quad (1.12.6)$$

т. е. равную нижней границе. Коды, для которых средняя длина кодовых слов (и соответственно скорость кодирования) равна наименьшему возможному значению, называются оптимальными.

Этим рассуждением мы показали, что в случае, когда вероятности сообщений есть целые отрицательные степени числа D , средняя длина оптимального D -ичного кода совпадает со своей нижней границей. Такой код сопоставляет сообщению x_i слово длины $-\log p(x_i)/\log D$.

Теорема 1.12.2. Существует D -ичный код со свойством однозначного декодирования, для которого

$$\bar{m}(X) < \frac{H(X)}{\log D} + 1. \quad (1.12.7)$$

Доказательство. Пусть m'_i — наименьшее целое число, удовлетворяющее условию $m'_i \geq I(x_i)/\log D$, где $I(x_i) = -\log p(x_i)$ — собственная информация сообщения x_i , $i = 1, 2, \dots, M$. Очевидно,

$$\frac{I(x_i)}{\log D} \leq m'_i < \frac{I(x_i)}{\log D} + 1. \quad (1.12.8)$$

Поскольку

$$\sum_{i=1}^M D^{-m'_i} \leq \sum_{i=1}^M D^{-\frac{I(x_i)}{\log D}} = \sum_{i=1}^M p(x_i) = 1, \quad (1.12.9)$$

то по теореме 1.11.1 существует дерево с концевыми вершинами порядков m'_1, \dots, m'_M . Соответствующий код будет иметь среднюю длину

$$\bar{m}(X) = \sum_{i=1}^M m'_i p(x_i) < \frac{H(X)}{\log D} + 1. \quad (1.12.10)$$

Теорема доказана.

Легко усмотреть в теоремах 1.11.1 и 1.11.2 обратную и прямую теоремы при побуквенном кодировании источника, выбирающего сообщения из ансамбля $\{X, p(x)\}$. Теперь мы сформулируем обобщение этих теорем на случай кодирования последовательностей сообщений, другими словами, на тот случай, когда ансамбль сообщений представляет собой n -ю степень ансамбля X .

Пусть $\{X^n, p(\mathbf{x})\}$ — произвольный дискретный ансамбль последовательностей сообщений, n — длина последовательностей и $H(X^n)$ — энтропия этого ансамбля. Тогда (см. теоремы 1.12.1 и 1.12.2) для любого D -ичного кода, однозначно кодирующего последовательности из X^n , среднее количество \bar{m} символов, приходящееся на одно сообщение,

$$\bar{m} \triangleq \frac{\bar{m}(X^n)}{n} \geq \frac{H(X^n)}{n \log D}. \quad (1.12.11)$$

Кроме того, существует код, для которого

$$\bar{m} < \frac{H(X^n)}{n \log D} + \frac{1}{n}. \quad (1.12.12)$$

Рассмотрим стационарный источник U_X , выбирающий сообщения из множества X и имеющий энтропию на сообщение $H(X|X^\infty)$. Пусть этот источник кодируется с помощью D -ичного неравномерного кода, т. е. каждой последовательности сообщений $\mathbf{x} = (x^{(1)}, \dots, x^{(n)}) \in X^n$ ставится в соответствие кодовое слово длины $m(\mathbf{x})$. Средняя длина кода $\bar{m}(X^n) = \sum_{\mathbf{x} \in X^n} p(\mathbf{x}) m(\mathbf{x})$, а среднее количество кодовых символов, приходящееся на одно сообщение источника,

$$\bar{m} = \frac{\bar{m}(X^n)}{n} = \frac{1}{n} \sum_{\mathbf{x} \in X^n} p(\mathbf{x}) m(\mathbf{x}). \quad (1.12.13)$$

При этом средняя скорость кодирования

$$R \triangleq \bar{m} \log D \triangleq \frac{\log D}{n} \sum_{\mathbf{x} \in X^n} p(\mathbf{x}) m(\mathbf{x}). \quad (1.12.14)$$

Теорема 1.12.3 (обратная теорема кодирования). Для любого кода, кодирующего источник U_X однозначно, средняя скорость кодирования

$$R \geq H(X|X^\infty). \quad (1.12.15)$$

Доказательство. Из определения средней скорости (1.12.14), из неравенства (1.12.11), справедливого для всех однозначно декодирующих кодов, и из теорем 1.5.1 и 1.5.2 следует, что для всех n , $n = 1, 2, \dots$, выполняются неравенства

$$R \geq \frac{1}{n} H(X^n) \geq H(X|X^{n-1}) \geq H(X|X^\infty).$$

Теорема доказана.

Теорема 1.12.4 (прямая теорема кодирования). Пусть ε — произвольное положительное число и D — число элементов кодового алфавита. Существует однозначно декодируемый D -ичный код, кодирующий источник U_X , для которого

$$R < H(X|X^\infty) + \varepsilon. \quad (1.12.16)$$

Доказательство. Согласно теореме 1.5.2 для любого положительного ε_1 найдется такое $N(\varepsilon_1)$, что для всех $n > N(\varepsilon_1)$

$$\frac{1}{n} H(X^n) < H(X|X^\infty) + \varepsilon_1. \quad (1.12.17)$$

Отсюда и из утверждения, связанного с неравенством (1.12.12), вытекает, что для произвольного целого $D > 0$ существует одно-

значно декодируемый D -ичный код со средним количеством символов на сообщение

$$\bar{m} < \frac{H(X^n)}{n \log D} + \frac{1}{n} < \frac{H(X|X^\infty)}{\log D} + \frac{\epsilon_1}{\log D} + \frac{1}{n} \quad (1.12.18)$$

и, следовательно, со скоростью кодирования

$$R < H(X|X^\infty) + \epsilon_1 + \frac{\log D}{n}. \quad (1.12.19)$$

Полагая $\epsilon_1 = \epsilon/2$, получим, что для всех $n \geq \frac{\log D}{\epsilon - \epsilon_1} = \frac{2 \log D}{\epsilon}$ имеет место неравенство $\epsilon_1 + \frac{\log D}{n} < \epsilon$. Утверждение теоремы справедливо теперь для всех n больших, чем максимальное из чисел $N(\frac{\epsilon}{2})$ и $\frac{2 \log D}{\epsilon}$. Теорема доказана.

Из теорем 1.12.3 и 1.12.4 (обратной и прямой теорем кодирования дискретного стационарного источника при неравномерном кодировании) следует, что средняя скорость создания информации таким источником равна энтропии на сообщение $H(X|X^\infty)$, т. е. равна той же величине, что и в задаче равномерного кодирования. Это означает, что минимальное количество двоичных символов, затрачиваемое в среднем на одно сообщение источника, может быть сделано сколь угодно близким к $H(X|X^\infty)$ как при равномерном, так и при неравномерном кодировании.

§ 1.13. Оптимальные неравномерные коды

В этом параграфе будет описан метод построения оптимальных однозначно декодируемых неравномерных кодов для дискретных ансамблей сообщений $\{X, p(x)\}$. Оптимальным называется код, средняя длина кодовых слов которого равна минимально возможной.

Рассмотрим вначале простейший случай, когда вероятности сообщений ансамбля являются целыми отрицательными степенями числа D — объема кодового алфавита:

$$p(x_i) = D^{-m_i}, \quad i = 1, 2, \dots, M. \quad (1.13.1)$$

В этом случае существует оптимальный D -ичный однозначно декодируемый код, для которого средняя длина кодовых слов

$$\bar{m}(X) = \frac{H(X)}{\log D} \quad (1.13.2)$$

(см. теорему 1.12.1 и следующее за ней замечание). В таком коде сообщению x_i сопоставляется слово длины $-\log p(x_i)/\log D = m_i$. Поэтому всякое дерево с набором концевых вершин порядков m_1, \dots, m_M и указанным правилом сопоставления дает оптимальный код.

Известен следующий метод построения такого дерева (метод Шеннона—Фано):

1. Подразделить множество сообщений на D подмножеств, так, чтобы вероятности каждого из подмножеств были одинаковыми, произвольным образом перенумеровать подмножества. Для всех сообщений из i -го подмножества, $i = 1, 2, \dots, D$, положить первый символ кодовых слов равным $a_i \in A$, где

$$A = \{a_1, a_2, \dots, a_D\}$$

— кодовый алфавит.

2. Каждое из подмножеств рассматривать как некоторое новое множество сообщений. Выполнить на j -м шаге, $j = 2, 3, \dots$, действия, указанные в п. 1 для определения j -го символа кодовых слов. Считать, что действия над данным подмножеством закончены, если оно содержит одно сообщение.

Пример 1.13.1. Неравномерный код, приведенный в примере 1.10.1, получен методом Шеннона—Фано. Процесс построения кода показан в табл. 1.13.1.

Здесь первое и второе подмножества, получающиеся при разбиении, обозначены через I и II соответственно, причем всем первым подмножествам соответствует символ 0, а вторым — 1.

Таблица 1.13.1

x_i	$p(x_i)$	1-й шаг	2-й шаг	3-й шаг	4-й шаг	Кодовые слова
x_1	1/4	I	I			0 0
x_2	1/4		II			0 1
x_3	1/8	I		I		1 0 0
x_4	1/8		I	II		1 0 1
x_5	1/16	II		I	I	1 1 0 0
x_6	1/16			I	II	1 1 0 1
x_7	1/16	II		II	I	1 1 1 0
x_8	1/16			II	II	1 1 1 1

Рассмотрим теперь общий случай, когда сообщения в ансамбле X имеют производственные вероятности. Мы ограничимся рассмотрением префиксных кодов, так как любой набор длин кодовых слов однозначно декодируемого кода может быть получен и на префиксном коде. Ниже будут даны условия, которым должен удовлетворять оптимальный префиксный код, а затем будет показано, как построить код, удовлетворяющий этим условиям. Будет рассмотрен только двоичный случай ($D = 2$). Недвоичный случай легко получить как обобщение двоичного. Всюду ниже будет предполагаться, что сообщения в ансамбле упорядочены так, что $p(x_1) \geq p(x_2) \geq \dots \geq p(x_M)$.

Лемма 1.13.1. В оптимальном коде слово, соответствующее наименее вероятному сообщению, имеет наибольшую длину.

Доказательство. Пусть m_i — длина слова, кодирующего сообщение $x_i \in X$, и \bar{m} — средняя длина кодовых слов:

$$\bar{m} = \sum_{i=1}^M m_i p(x_i). \quad (1.13.3)$$

Предположим, что в оптимальном коде $m_i > m_M$ для некоторого $i < M$. Рассмотрим код, в котором i -е и M -е слова исходного кода заменены одно другим. Средняя длина \bar{m}' для этого кода

$$\begin{aligned} \bar{m}' &= \bar{m} - p(x_i)m_i - p(x_M)m_M + p(x_i)m_M + p(x_M)m_i = \\ &= \bar{m} - (m_i - m_M)(p(x_i) - p(x_M)) < \bar{m}, \end{aligned} \quad (1.13.4)$$

что противоречит предположению об оптимальности исходного кода.

Лемма 1.13.2. В оптимальном двоичном префиксном коде два наименее вероятных сообщения кодируются словами одинаковой длины, одно из которых оканчивается нулем, а другое единицей.

Доказательство. Обозначим через u_j слово, кодирующее сообщение x_j . Пусть u_M — слово наибольшей длины оптимального кода. Тогда существует еще одно слово, скажем u_i , такой же длины. В противном случае единственное слово наибольшей длины можно было бы укоротить без нарушения декодируемости и получить меньшую среднюю длину. Таким образом, слова u_M и u_i должны отличаться в последнем символе. Покажем теперь, что эти два слова кодируют наименее вероятные сообщения. Предположим противное, т. е. что $i < M - 1$. Тогда $m_i = m_M > m_{M-1}$. В этом случае среднюю длину кода можно было бы уменьшить, заменив слово u_i на u_{M-1} и u_{M-1} на u_i . Следовательно, это предположение не справедливо и наибольшую длину имеют слова u_{M-1} и u_M . Лемма доказана.

Рассмотрим новый ансамбль X' , состоящий из $M - 1$ сообщений $\{x'_1, \dots, x'_{M-1}\}$ с вероятностями

$$p(x'_i) = \begin{cases} p(x_i), & i = 1, 2, \dots, M - 2, \\ p(x_{M-1}) + p(x_M), & i = M - 1. \end{cases} \quad (1.13.5)$$

Любой декодируемый префиксный код для ансамбля X' можно превратить в декодируемый код для ансамбля X приписыванием к кодовому слову, кодирующему сообщение x'_{M-1} , символов 0, 1 для получения слов, кодирующих сообщения x_{M-1}, x_M . Теперь нетрудно указать последовательную процедуру построения оптимального префиксного кода. Для этого необходимо обосновать, что оптимизация на каждом шаге приведет к оптимизации кода. Это обоснование выполняется с помощью следующего утверждения.

Лемма 1.13.3. Если оптимален однозначно декодируемый префиксный код для ансамбля X' , то оптимален полученный из него префиксный код для ансамбля X .

Доказательство. Обозначим через \bar{m}' среднюю длину кодовых слов кода для ансамбля X' . Тогда средняя длина \bar{m} кодовых слов кода для ансамбля X

$$\begin{aligned} \bar{m} &= \sum_{i=1}^M m_i p(x_i) = \sum_{i=1}^{M-2} m_i p(x_i) + m_{M-1} p(x_{M-1}) + m_M p(x_M) = \\ &= \sum_{i=1}^{M-1} m'_i p(x'_i) - m'_{M-1} [p(x_{M-1}) + p(x_M)] + m_{M-1} p(x_{M-1}) + \\ &\quad + m_M p(x_M) = \bar{m}' + p(x_{M-1}) [m_{M-1} - m'_{M-1}] + \\ &\quad + p(x_M) [m_M - m'_{M-1}] = \bar{m}' + p(x'_{M-1}), \end{aligned} \quad (1.13.6)$$

где использовано то обстоятельство, что длины m'_i , $i = 1, 2, \dots, M - 1$, кодовых слов кода для ансамбля X' связаны с длинами m_i , $i = 1, 2, \dots, M$, кодовых слов кода для ансамбля X следующими соотношениями:

$$\begin{cases} m_i = m'_i, & i = 1, 2, \dots, M - 2, \\ m_M = m_{M-1} = m'_{M-1} + 1. \end{cases} \quad (1.13.7)$$

Из (1.13.6) следует, что \bar{m} и \bar{m}' отличаются на константу $p(x'_{M-1})$, которая не зависит от выбора кодовых слов, и строя декодируемый код для ансамбля X' с минимальным значением \bar{m}' , мы получаем декодируемый код для ансамбля X с минимальным значением \bar{m} . Лемма доказана.

Таким образом, задача построения оптимального префиксного кода сводится к задаче построения оптимального префиксного кода для ансамбля, содержащего на одно сообщение меньше. В этом ансамбле снова можно выделить два наименее вероятных

сообщения и, объединяя их, получить новый ансамбль, содержащий теперь уже на два сообщения меньше, чем исходный. Очевидно, что таким образом можно дойти до ансамбля, содержащего всего два слова, оптимальным кодом для которого является просто 0 для одного сообщения и 1 для другого. Описанный метод построения оптимального префиксного кода называется методом Хаффмена.

Пример 1.13.2. Рассмотрим ансамбль, состоящий из 7 сообщений, вероятности которых равны 0,3; 0,2; 0,15; 0,15; 0,1; 0,05; 0,05. В следующей таблице показаны 6 последовательных шагов, на каждом из которых происходит образование нового ансамбля с помощью склеивания наименее вероятных сообщений предыдущего ансамбля. Как видно из этого примера, процесс склеивания приводит к дереву и, следовательно, к однозначно декодируемому коду.

Таблица 1.13.2

Сообщения	Вероятности	1	2	3	4	5	6	Кодовые слова
x_1	0,3							11
x_2	0,2							01
x_3	0,15							101
x_4	0,15							100
x_5	0,1							001
x_6	0,05							0001
x_7	0,05							0000

Средняя длина кодовых слов равна 2,6. Нижняя граница согласно теореме 1.12.1 равна 2,354. Кода со средней длиной меньшей, чем 2,6, не существует. Движению вверх по дереву сопоставлен символ 1, движению вниз — символ 0.

§ 1.14. Обсуждение основных результатов

Мы познакомились с двумя задачами кодирования дискретных источников — с задачей равномерного кодирования и задачей неравномерного кодирования. Отличие постановок этих задач состоит в том, что в первом случае требуется восстановление сообщений источника с заданной, обычно малой, вероятностью ошибки, а во втором случае требуется точное восстановление сообщений источника.

Теперь мы обсудим работу равномерного и неравномерного кодеров. При этом мы коснемся двух вопросов. Первый — это вопрос о том, как можно охарактеризовать выходную последовательность достаточно хорошего кодера, а второй — вопрос об организации работы кодера в типичном для практики случае, когда источник порождает сообщения независимо от процесса кодирования (т. е. в случае неуправляемых источников). Как

мы увидим, и равномерные и неравномерные коды приводят к однотипному поведению соответствующих кодеров.

Из теорем 1.12.3 и 1.12.4 следует, что при неравномерном кодировании скорость создания информации дискретным стационарным источником равна энтропии на сообщение $H(X|X^\infty)$, т. е. той же величине, что и при равномерном кодировании. Это означает, что минимальное количество символов, затрачиваемое в среднем на кодирование одного сообщения D -ичным кодом, равно $H(X|X^\infty)/\log D$ как при равномерном, так и при неравномерном кодировании. Если источник эргодичен, то при равномерном кодировании это утверждение верно и для каждой отдельной достаточно длинной реализации последовательности сообщений на выходе источника. Во втором случае это не так. При неравномерном кодировании мы можем говорить только о среднем по множеству источников (по множеству реализаций) количестве символов на сообщение.

Наш интерес к величине средней скорости кодирования обусловлен прежде всего тем, что по этой величине в ряде случаев можно судить о работе кодера с каждой отдельной реализацией на выходе источника, а не только в среднем по множеству всех реализаций. В следующем примере показано, что эргодичность является необходимым условием для того, чтобы можно было судить о количестве символов на сообщение на выходе неравномерного кодера.

Пример 1.14.1. Предположим, что неэргодический источник имеет две равновероятные эргодические компоненты (см. рис. 1.9.1), причем одна компонента — источник, который в каждый момент времени порождает одно и то же сообщение, скажем 0, а другая компонента — источник, порождающий независимо и с равными вероятностями два других сообщения, скажем 1 и 2. Если кодировать последовательности сообщений длины n с помощью двоичного кода, то, как было показано в § 1.13, средняя длина $\bar{m}(X^n)$ минимизируется выбором одного слова длины 1 для первой компоненты и 2^n слов длины $n+1$ — для второй. Так как выбранная случайно один раз компонента никогда более не меняется, то либо все кодовые слова будут иметь длину 1, либо $n+1$. Для такого источника $H(X|X^\infty) = \frac{1}{2}$ (см. задачу 1.14.1). Следовательно, средняя длина наилучшего кода равна $n/2$, но не равна количеству символов на сообщение на выходе кодера.

Из этого примера следует, что для неэргодических источников количество символов на сообщение на выходе кодера не определяется средней скоростью кодирования. Покажем теперь, что одного условия эргодичности недостаточно. Действительно, пусть источник порождает последовательность из N сообщений. Для этой последовательности количество кодовых символов на сообщение m_{cp} может быть определено по формуле

$$m_{cp} \triangleq \frac{1}{N} \sum_{i=1}^N m^{(i)} = \frac{1}{L} \sum_{i=1}^L v^{(i)}, \quad (1.14.1)$$

где $L = N/n$ — количество блоков длины n , $m^{(i)}$ — длина слова, кодирующего i -й блок, и $v^{(i)} = m^{(i)}/n$ — количество символов на сообщение на i -м блоке. Величины $v^{(i)}$ случайны, но для стационарного источника $Mv^{(i)}$ одинаковы для всех $i = 1, 2, \dots$ и при больших n могут быть сделаны близкими к $H(X|X^\infty)/\log D$. Будет ли величина $m_{\text{ср}}$ близкой к $H(X|X^\infty)/\log D$, зависит от свойств источника. Это так, если при любых положительных ϵ и δ и при достаточно больших L

$$\Pr\left(\left|m_{\text{ср}} - \frac{H(X|X^\infty)}{\log D}\right| \geq \epsilon\right) < \delta. \quad (1.14.2)$$

Хотя (1.14.2) по форме напоминает определение эргодичности, но в действительности из эргодичности не выводится, так как для каждого i величина $v^{(i)}$ есть функция i -го блока длины n . Можно показать, что (1.14.2) выводится из предположения об эргодичности блоков длины n (из так называемого свойства *блочной эргодичности*). То, что из эргодичности в общем случае не следует блочная эргодичность, иллюстрируется в задаче 1.14.2.

Предположим теперь, что источник является эргодическим, а при неравномерном кодировании — и блочно эргодическим. Пусть N есть количество кодовых символов на выходе кодера, появляющееся при подаче на его вход блока из L сообщений источника. Пусть A — D -ичный кодовый алфавит, тогда можно записать следующую цепочку неравенств:

$$N \log D \geq H(A^N) \geq H(X^L) \geq LH(X|X^\infty), \quad (1.14.3)$$

где D — число элементов в множестве A . Первое неравенство обращается в равенство в том и только том случае, когда символы на выходе кодера статистически независимы и равновероятны. Второе неравенство должно выполняться вследствие требования однозначности кодирования: среднее количество информации в кодовых словах не может быть меньше, чем среднее количество информации в сообщениях, которые отображаются с помощью этих слов. Последнее неравенство — следствие предположения о стационарности источника. Для достаточно хорошего кода (и соответственно кодера) среднее число символов на сообщение должно быть близким к $H(X|X^\infty)/\log D$, т. е.

$$\frac{N}{L} \leq \frac{H(X|X^\infty) + \epsilon}{\log D}, \quad (1.14.4)$$

где ϵ — достаточно малое положительное число. Отсюда следует, что

$$L[H(X|X^\infty) + \epsilon] \geq N \log D \quad (1.14.5)$$

и, следовательно, все величины, фигурирующие в неравенствах (1.14.3), должны быть близкими друг к другу. Здесь для нас

существенным является близость величин $N \log D$ и $H(A^N)$, из которой следует, что символы на выходе кодера (равномерного или неравномерного) являются почти независимыми и почти равновероятными. Близость к независимости и равновероятности тем сильнее, чем ближе скорость кодирования к $H(X|X^\infty)$.

Таким образом, действие оптимального кодера источника можно описать следующим образом. На вход кодера поступают в общем случае зависимые и неравновероятные сообщения. На выходе кодера появляется последовательность почти независимых и почти равновероятных кодовых символов, по которым точно или сколь угодно точно можно восстановить сообщения, порождаемые источником. Этим мы закончим обсуждение первого вопроса и перейдем ко второму.

Источники сообщений можно подразделить на управляемые и неуправляемые (в литературе иногда встречаются соответствующие термины — источники с переменной и источники с постоянной скоростью). Управляемые источники имеют специальный управляющий вход. Каждое очередное сообщение появляется на выходе источника только после того, как будет подан сигнал на управляющий вход. Типичным примером такого источника является запоминающее устройство, считывание из ячеек памяти которого происходит только при подаче соответствующих управляющих сигналов. Неуправляемые источники являются более типичными для различных систем сбора и передачи информации. Такие источники порождают сообщения с некоторой фиксированной скоростью, например, по одному сообщению каждую секунду, и никакие управляющие воздействия не могут замедлить или ускорить процесс появления сообщений.

При кодировании с помощью равномерных кодов никаких дополнительных проблем не возникает. Если источник эргодический, то с высокой вероятностью для кодирования L сообщений будет использовано примерно $LH(X|X^\infty)/\log D$ кодовых символов. Если n — длина кодируемых сообщений, то на каждые n сообщений источника появляется одно кодовое слово длины $nR/\log D$, где R — скорость кодирования. Число n определяет, с одной стороны, близость R и $H(X|X^\infty)$, а с другой достижимую вероятность ошибки. Выбирая n достаточно большим, можно для сколь угодно малых положительных ϵ и δ сделать $R = H(X|X^\infty) + \epsilon$ и сделать вероятность ошибки меньшей или равной δ .

Проблема возникает при использовании неравномерных кодов для кодирования неуправляемых источников. Так как длины кодовых слов отличаются друг от друга, то и время кодирования для различных сообщений будет различным. Это приводит к появлению очередей на выходе кодера. При появлении на выходе источника сообщений, кодируемый длинными словами, кодер не

будет успевать кодировать, тогда как при появлении сообщений, кодируемыми короткими словами, он будет простаивать. Хотя в среднем на одно сообщение будет по-прежнему тратиться примерно $H(X|X^\infty)/\log D$ кодовых символов, но для каждой отдельной реализации могут возникнуть сильные колебания числа сообщений, ожидающих кодирования.

Частично эта трудность преодолевается, если между источником и кодером поставить буферное запоминающее устройство, которое теперь будет играть роль управляемого источника. Считыванием из буфера можно управлять по мере окончания кодирования каждого очередного сообщения. Однако теперь следует учитывать вероятность переполнения буфера и рассматривать эту вероятность как вероятность ошибки при кодировании.

Таким образом, ошибки присущи не только кодированию с помощью равномерных кодов, но также и кодированию управляемых источников, с помощью неравномерных кодов.

Задачи, упражнения и дополнения

1.1.1. Пусть X состоит из 3-х элементов. Сколько элементов в множестве X^{10} ?

1.1.2. Покажите, что для любого $A \subseteq X$ такого, что $\Pr_1(A) \neq 0$, имеет место равенство $\sum_{y_j \in Y} p(y_j|A) = 1$, где $p(y_j|A)$ определено в (1.1.6).

1.1.3. Пусть множества X и Y состоят из двух элементов, именно из 0 и 1. Предположим, что распределение вероятностей $p(x, y)$ на множестве XY задано следующим образом: $p(0, 0) = 3/20$, $p(0, 1) = 2/20$, $p(1, 0) = 9/20$ и $p(1, 1) = 6/20$. Являются ли ансамбли X и Y статистически независимыми?

1.1.4. Покажите, что случайные процессы, определенные в примере 1.1.3, являются стационарными. Указание: для ответа в случае п. б) найдите распределение $p(y^{(i)})$ для всех моментов времени в предположении, что в начальный момент $p(y^{(1)} = y_1) = p(y^{(1)} = y_2)$.

1.2.1. Найдите математическое ожидание и дисперсию случайной величины, определенной в примере 1.2.2, при условии, что $p(x_i) = 1/5$ для всех i .

1.2.2. Пусть $p(x|y)$ — условное распределение на X при фиксированном $y \in Y$. Число $M(X|y) \triangleq \sum_{x \in X} p(x|y) \cdot x$ называется условным математическим

ожиданием (м. о.) случайной величины X . Рассмотрите в условии примера 1.1.2 случайную величину — номер элемента в множестве X : $\phi(x_i) = i$. Найдите условные м. о. $M[\phi(X)|y_1]$ и $M[\phi(X)|y_2]$. Покажите, что безусловное м. о. $M[\phi(X)] = 1/2 \{M[\phi(X)|y_1] + M[\phi(X)|y_2]\}$.

1.2.3. Покажите, что для всякого четного $k > 0$ имеет место неравенство

$$\Pr(|X| > e) \leq \mu_k/e^k, \quad e > 0,$$

где X — случайная величина с нулевым м. о., μ_k — центральный момент порядка k (см. формулу (1.2.2)). Такое неравенство также называют неравенством Чебышева.

1.2.4. Пусть X — положительная случайная величина и m — ее м. о. Покажите, что $\Pr(X > e) \leq m/e$. Используйте этот результат для того, чтобы показать, что не более чем α -я доля всех значений случайной величины X больше, чем m , в $1/\alpha$ раз.

1.2.5. Покажите, что не более половины яблок, лежащих в вазе, вдвое тяжелее среднего по весу яблока. Покажите, что не менее двух третей яблок имеют вес меньший, чем утроенный вес среднего.

1.2.6. Оцените с помощью неравенства Чебышева, при каком n доля наблюдаемых единиц в эксперименте, описанном в примере 1.2.3, отличается на $\alpha\%$ от ожидаемой с вероятностью меньшей, чем 0,01. Как ведет себя число n с увеличением ρ ?

1.3.1. Пусть $\{X, p(x)\}$ — произвольный дискретный ансамбль. Покажите, что k -й момент μ_k случайной величины $I(x) = -\log p(x)$ конечен, $k = 1, 2, \dots$ Для этого выполните следующие шаги:

а) Покажите, что в случае, когда $p(x) > 0$, для всех $x \in X$

$$\mu_k \triangleq \sum_X I^k(x) p(x) < \infty.$$

б) Покажите, что

$$\lim_{p \rightarrow 0} p \log^k p = 0.$$

1.3.2. Покажите, что энтропия двоичного ансамбля как функция от вероятности p одного из сообщений имеет вид, изображенный на рис. 13.1. Для этого вычислите производную от правой части выражения (1.3.3) и определите область значений переменной p , где энтропия возрастает, убывает, принимает экстремальное значение. Используйте вторую производную, чтобы показать, что экстремальное значение есть максимум.

1.3.3. Пусть $X = \{x_1, x_2, \dots\}$ — множество, состоящее из бесконечного числа элементов. Приведите пример распределения вероятностей на X такого, чтобы энтропия $H(X) \triangleq -\sum_{i=1}^{\infty} p(x_i) \log p(x_i)$ была ограниченной. Каким

должно быть распределение, чтобы $H(X) = 2$?

1.3.4. Пусть $X = \{x_1, x_2, \dots, x_M\}$ и требуется найти такое распределение на X , при котором $H(X)$ максимально и $p(x_1) = e$, где e — заданное число. Покажите, используя метод неопределенных множителей Лагранжа, что $H(X) \leq -e \log e - (1-e) \log \frac{1-e}{M-1}$ и равенство достигается в том и только в том случае, когда все сообщения, кроме x_1 , имеют одинаковые вероятности. Получите этот же результат с помощью неравенства для логарифма.

1.4.1. Предположим, что ансамбль Y однозначно определяет ансамбль X . Это означает, что при каждом фиксированном сообщении $y \in Y$ имеется сообщение $x(y) \in X$, для которого $p(x(y)|y) = 1$. Покажите, что в этом случае $H(X|Y) = 0$ и наоборот, из равенства нулю $H(X|Y)$ следует, что ансамбль Y однозначно определяет ансамбль X , если $p(y) \neq 0$ для всех $y \in Y$.

1.4.2. Пусть $X = \{x_1, x_2, \dots, x_8\}$ и $p(x_1) = p(x_2) = p(x_3) = \frac{1}{4}$, $p(x_4) = p(x_5) = p(x_6) = \frac{1}{12}$. Пусть $Y = \{y_1, y_2, y_3\}$ и $p(y_1) = p(y_2) = p(y_3) = \frac{1}{3}$. Покажите, что $H(X) > H(Y)$. Указание: воспользуйтесь свойством 3 из раздела 1.4.

1.5.1. Пусть дискретный стационарный источник порождает случайный процесс, для которого $p(x^{(n)}|x^{(n-1)}, \dots, x^{(1)}) = p(x^{(n)}|x^{(n-1)}, \dots, x^{(n-s)})$ для любых $n \geq s$ и любых последовательностей $(x^{(1)}, \dots, x^{(n)})$. Такой источник называется марковским источником порядка s . Покажите, что $H(X|X^n) = H(X|X^s)$ для всех $n \geq s$, где $H(X|X^k) \triangleq H(X_i|X_{i-k}, \dots, X_{i-1})$ и индекс i опущен, ввиду стационарности.

1.5.2. Используйте предыдущую задачу, чтобы показать, что энтропия на сообщение стационарного марковского источника порядка s равна $H(X|X^s)$.

1.5.3. Как соотносятся по величине $H(X|X^k)$ и $H(X^k)/k$?

1.6.1. Покажите, что наибольшее возможное количество D -ичных последовательностей, длина которых меньше или равна m , равно $D(D^m - 1)/(D - 1)$.

1.6.2. Обозначим через $\lfloor x \rfloor$ наименьшее целое число, большее или равное x , например, $\lfloor 7/3 \rfloor = 3$. Если для кодирования отрезков сообщений длины n используется код с M словами, то количество двоичных символов на сообщение определяется следующим образом. Каждому слову можно поставить в соответствие двоичную последовательность длины $\lceil \log M \rceil$ (номер слова). Тогда $\lceil \log M \rceil$ есть число двоичных символов на сообщение. Предположим теперь, что каждой паре слов ставится в соответствие двоичная последовательность (номер пары). Так как число пар равно M^2 , то количество двоичных символов на сообщение равно $\lceil \log M^2 / 2n \rceil = \lceil 2\log M / 2n \rceil$. Если каждому набору из k кодовых слов ставится в соответствие двоичная последовательность, то количество двоичных символов на сообщение будет равно $\lceil k \log M / kn \rceil$. Покажите, что

$$\lim_{k \rightarrow \infty} \lceil k \log M / kn \rceil = \log M/n.$$

1.6.3. Известно, что факториалы хорошо аппроксимируются формулой Стирлинга (см., например, [6])

$$n! \sim \sqrt{2\pi n} n^n e^{-n}.$$

Относительная ошибка мала даже при малых n . Так, при $n = 2, 5, 10, 100$ относительные ошибки равны соответственно 0,04, 0,02, 0,008, 0,0008. Можно показать, что

$$\sqrt{2\pi n} n^n e^{-n + \frac{1}{12n+1}} < n! < \sqrt{2\pi n} n^n e^{-n + \frac{1}{12n}}.$$

а) Пользуясь формулой Стирлинга, покажите, что при больших n , r и фиксированном отношении $\rho = r/n$ число сочетаний из n по r

$$C_n^r \triangleq \frac{n!}{r!(n-r)!} \sim (2\pi n \rho(1-\rho))^{-1/2} 2^{nh(\rho)},$$

где $h(\rho) = -\rho \log \rho - (1-\rho) \log(1-\rho)$ — энтропия двоичного ансамбля.

б) Покажите, что

$$(1/\sqrt{2\pi n \rho(1-\rho)}) 2^{n(h(\rho)-o(n))} < C_n^r < (1/\sqrt{2\pi n \rho(1-\rho)}) 2^{n h(\rho)}.$$

Покажите, что остаточный член $o(n)$ положителен и убывает к нулю с ростом n .

в) Покажите, что $C_n^{\rho n}$ возрастает при увеличении ρ от нуля до $1/2$, максимально при $\rho = 1/2$ и убывает при дальнейшем увеличении ρ от $1/2$ до единицы.

1.7.1. Рассмотрим двоичный источник без памяти, описанный в примере 1.7.2. Пусть числа ε_1 и ε_2 зафиксированы. Оцените число N , начиная с которого выполняется утверждение теоремы о высоковероятных множествах для этого источника, в зависимости от ρ — вероятности появления единицы.

1.7.2. Как соотносятся числа элементов в множествах $T_{n_1}(\varepsilon_1)$ и $T_{n_2}(\varepsilon_2)$, если $\varepsilon_1 > \varepsilon_2$ при $n_1 = n_2$? Тот же вопрос, если при $n_1 > n_2$, $\varepsilon_1 = \varepsilon_2$.

1.8.1. Пусть $R < H(X)$. Покажите, используя неравенство Чебышева, что $P_{en} \geq 1 - \frac{c}{n}$ для любого равномерного кода, кодирующего дискретный источник без памяти со скоростью R . Покажите, что в случае двоичного источника, независимо порождающего сообщения с вероятностями ρ и $1-\rho$,

$$c = \left(\log \frac{1-\rho}{\rho} \right)^2 4\rho(1-\rho)/(R-H(X))^2.$$

Указание: воспользуйтесь примером 1.7.2.

1.8.2. Уточните с помощью задачи 1.6.3 предыдущее неравенство для двоичного источника, независимо порождающего сообщения с вероятностями ρ и $1-\rho$. Покажите для этого, что

$$P_{en} > 1 - a \sqrt{n \cdot 2^{-n} \left[(\rho+\varepsilon_1) \log \frac{1}{\rho} + (1-\rho-\varepsilon_1) \log \frac{1}{1-\rho} - h(\rho+\varepsilon_1) \right]},$$

где $\varepsilon_1 = [h(\rho) - R] / \log \frac{1-\rho}{\rho}$. Оцените величину a .

1.8.3. Пусть дискретный источник без памяти имеет энтропию $H(X) = 2,5$ бит. Пусть 1000 сообщений такого источника кодируются десятичным равномерным кодом. Оцените наименьшее количество кодовых символов, которое может появиться на выходе кодера. Как зависит это количество от допустимой вероятности ошибки?

1.9.1. Покажите, что из соотношений (1.9.3) и (1.9.6) следует неравенство (1.9.7).

1.9.2. Рассмотрим двоичный эргодический источник, выбирающий сообщения из множества $\{0, 1\}$.

а) Предположим, что требуется оценить вероятность появления пары сообщений $(1, 0)$. Как это можно сделать?

б) Предположим, требуется оценить вероятность появления не более k единиц в последовательности длины n . Какой вид имеет оценка? Укажите соответствующую функцию $\Phi(\cdot)$.

1.9.3. Лемма Мак-Миллана имеет очень простое доказательство для марковских источников порядка s , у которых все условные вероятности сообщений отличны от нуля. Марковский стационарный источник порядка s был определен в задаче 1.5.1. Там же показано, что энтропия на сообщение для такого источника равна $H(X|X^s)$. В этом случае неравенство (1.9.16) может быть записано с учетом (1.9.17) как

$$P_s \triangleq \Pr \left(\left| \frac{1}{n} \sum_{i=1}^n I(x^{(i)} | x^{(i-1)}, \dots, x^{(1)}) - H(X|X^s) \right| > \varepsilon \right) < \delta.$$

Пусть

$$\lambda \triangleq \sum_{i=1}^s I(x^{(i)} | x^{(i-1)}, \dots, x^{(1)}) - I(x^{(i)} | x^{(i-1)}, \dots, x^{(i-s)}).$$

Покажите, что

$$P_s = \Pr \left(\left| \frac{1}{n} \sum_{i=1}^n I(x^{(i)} | x^{(i-1)}, \dots, x^{(1)}) - H(X|X^s) \right| > \varepsilon - \frac{\lambda}{n} \right).$$

Покажите далее, что λ — ограниченная случайная величина, не зависящая от n . Покажите, что теперь применимо определение эргодического источника, откуда и следует доказываемое утверждение.

1.10.1. Покажите, что код примера 1.10.3 является однозначно декодируемым. Найдите непрефиксный код, состоящий из 4 кодовых слов и обладающий свойством однозначного декодирования.

1.10.2. Убедитесь в том, что необходимое условие однозначного декодирования для кода примера 1.10.2 не выполнено, а для кода примера 1.10.3 выполнено. Положите $L = 2$ и $j = 4$. Найдите A_j (см. теорему 1.10.1) для кода примера 1.10.3 и сравните это число с D^j .

1.11.1. Проверьте, может ли быть построено двоичное дерево с концевыми узлами порядков 1, 2, 2, 3. Тот же вопрос для троичного дерева.

1.11.2. Предположим, что требуется построить двоичный код с $M = 100$ кодовыми словами, у которого длина i -го слова равна i . Существует ли такой код? Тот же вопрос, если $M = 1000$.

1.11.3. D -ичное дерево называется полным, если из каждого неконцевого узла выходит точно D ребер. Дерево на рис. 1.11.1 — полное, а на рис. 1.11.2 и 1.11.3 — не полное. Покажите, что для полного дерева неравенство Крафта выполняется со знаком равенства.

1.12.1. Предположим, что однозначно декодируемый двоичный код для ансамбля сообщений $\{X, p(x)\}$ состоит из кодовых слов 10, 01, 000, 111. Вероятности этих кодовых слов равны соответственно $1/2$, $1/4$, $3/16$, $1/16$.

а) Не пользуясь таблицами логарифмов, оцените сверху энтропию ансамбля X .

б) Покажите, что этот код не является оптимальным. Указание: несовпадение средней длины кодовых слов и энтропии не является, конечно, признаком неоптимальности кода. Объясните, почему.

1.12.2. Пусть $X = \{x_1, x_2, \dots\}$ — множество, состоящее из бесконечного числа элементов. Пусть $p(x_1) = 1 - p_0$, а $p(x_i) = q^{i-1}$, $i > 1$, где p_0 — заданное число, а $q < 1$.

а) Найдите q как функцию от p_0 .

б) Покажите, что энтропия ансамбля $\{X, p(x)\}$

$$H(X) = -(1 - p_0) \log(1 - p_0) - (q \log q)/(1 - q^2).$$

в) Найдите однозначно декодируемый код для ансамбля $\{X, p(x)\}$, у которого средняя длина кодовых слов конечна. Найдите двоичный код со средней длиной $1 + p_0 + p_0^2$.

г) Покажите, что при $p_0 = 1$ существует оптимальный двоичный код. Покажите, что в этом случае существует неоптимальный двоичный код, все кодовые слова которого имеют ограниченные длины, а средняя длина кодовых слов отличается от $H(X)$ не более чем на ε , где ε — произвольное положительное число.

1.12.3. Пусть сообщения x_1, x_2, x_3, x_4 имеют вероятности $p(x_1) = 1/2$, $p(x_2) = 1/4$, $p(x_3) = p(x_4) = 1/8$. Постройте оптимальный двоичный код со свойством однозначной декодируемости.

1.13.1. Покажите, что в случае, когда сообщения имеют вероятности D^{-m_i} , действия, указанные в описании метода Шеннона—Фано, в действительности приводят к оптимальному однозначно декодируемому коду. Для этого выполните следующее.

а) Покажите, что на каждом шаге возможно разбиение сообщений на равновероятные подмножества. Указание: воспользуйтесь тем, что для полных деревьев $\sum_i D^{-m_i} = 1$; такому дереву соответствует D полных поддеревьев,

корнями которых служат узлы первого яруса исходного дерева, следовательно, $\sum_{i \in I_j} D^{-m_i+1} = 1$, где I_j — множество концевых узлов j -го поддерева, $j = 1, 2, \dots, D$, и эти множества задают разбиение на равновероятные подмножества для первого шага.

б) Покажите, что полученный в результате последовательных разбиений код — древовидный и сообщению x_l соответствует слово длины $-\log p(x_l)$.

1.13.2. Метод построения оптимального кода (метод Хаффмена) сводится к построению дерева (см. пример 1.13.2).

а) В общем случае на некотором шаге может оказаться несколько пар наименее вероятных сообщений. Покажите, что любое склеивание наименее вероятных сообщений приводит к дереву с одним и тем же набором порядков концевых узлов и, следовательно, с одной и той же средней длиной.

б) Пусть p'_i — вероятность наименее вероятного сообщения в ансамбле, полученном на i -м шаге ($p'_1 = 0.1$ для примера 1.13.2). Покажите, что средняя

длина оптимального двоичного кода $\bar{m} = 1 + \sum_{i=1}^{l-1} p'_i$, где l — число шагов в процедуре Хаффмена ($l = 6$ для примера 1.13.2).

1.13.3. Пусть X — ансамбль равновероятных десятичных цифр от 0 до 9. Энтропия этого ансамбля $H(X) = \log 10 \approx 3.32$ бит.

а) Найдите длину равномерного двоичного кода, однозначно кодирующего ансамбль X .

б) Найдите среднюю длину двоичного кода Хаффмена, кодирующего этот же ансамбль.

в) Покажите, что длина равномерного двоичного кода, кодирующего блоки, состоящие из n десятичных цифр, равна $\lceil n \log 10 \rceil$, где $\lceil x \rceil$ — наименьшее целое, большее или равное x . При каком n количество двоичных символов на сообщение отличается от минимально возможного не более чем на 1%?

г) Оцените среднюю длину оптимального неравномерного кода, кодирующего блоки длины n .

1.13.4. Все методы кодирования, которые рассматривались в этой главе, заключались в том, что последовательность сообщений источника разбивалась на блоки фиксированной длины n , а затем осуществлялось равномерное или неравномерное кодирование ансамбля X^n . В этой задаче описан метод, называемый кодированием длин серий, не связанный с разбиением последовательности сообщений на блоки фиксированной длины. Из этого метода следует, что помимо задачи выбора оптимального кода для данного множества сообщений интерес представляет также задача выбора хорошего множества кодируемых сообщений.

Пусть источник порождает последовательность независимых символов 0 и 1, причем вероятность появления 1 равна $p < 1/2$. Рассмотрим следующий метод кодирования. Выходная последовательность источника разбивается на сегменты вида 1, 01, 001, ..., 000...01, 000...00, где длины сегментов соответственно равны 1, 2, ..., N и количество сегментов равно $N + 1$ (два последних сегмента имеют одинаковую длину N). Очевидно, разбиение любой двоичной последовательности на сегменты однозначно. Далее сегменты рассматриваются как сообщения некоторого источника; вероятность появления i -го сегмента равна $p q^{i-1}$, $i = 1, 2, \dots, N$, $q = 1 - p$ и вероятность появления $(N + 1)$ -го сегмента равна q^N . Эти сообщения кодируются неравномерным двоичным кодом так, что последнему сегменту сопоставляется слово длины 1, а всем остальным — слова длины $k + 1$, где k — такое, что $N < 2^k < N + 1$.

а) Покажите, что такое кодирование однозначно.

б) Пусть m_1 — средняя длина сегмента, т. е. среднее количество сообщений источника в сегменте. Покажите, что

$$m_1 = \sum_{i=1}^N i p q^{i-1} + N q^N = (1 - q^N)/p.$$

в) Пусть m_2 — средняя длина кодовых слов, кодирующих сегменты, т. е. среднее количество кодовых символов, приходящихся на один сегмент. Покажите, что

$$m_2 = k(1 - q^N) + 1.$$

г) Число m_2/m_1 можно рассматривать как среднее количество кодовых символов на сообщение, т. е. как скорость кодирования при кодировании длин серий. Покажите, что

$$m_2/m_1 = kp + \frac{p}{1 - q^N}.$$

д) Пусть $p = 0,1$, $N = 8$ и, следовательно, $k = 3$. Покажите, что средняя длина кода Хаффмена, кодирующего блоки длины 4, равна 1,9702 и поэтому скорость кодирования равна 0,4925. Покажите, что скорость при кодировании длин серий равна 0,4756, что меньше, чем для кода Хаффмена. Объясните причину такого различия. Обратите внимание на то, что кодирование длин серий более просто с точки зрения технической реализации.

е) Из п. г) видно, что скорость кодирования зависит от параметра N . При слишком больших или при слишком малых значениях N скорость будет слишком большой. Поэтому для каждого p имеется оптимальный выбор числа N . Найдите подбором это оптимальное N для $p = 0,1$.

1.14.1. Покажите, что для источника примера 1.14.1 $H(X|X^\infty) = \frac{1}{2}$. Для этого выполните следующие шаги.

а) Покажите, что при любом n выполняются неравенства $H(X^n|S)/n \leq H(X^n)/n \leq H(X^n|S)/n + 1/n$, где $S = \{s_1, s_2\}$ — множество состояний переключателя эргодических компонент. Указание: воспользуйтесь тождеством $H(X^n) = H(X^n|S) + H(S|X^n)$ (обоснование тождества будет дано в следующей главе).

б) Покажите, что $H(X^n|s_1) = 0$ и $H(X^n|s_2) = n$, где s_i соответствует i -й эргодической компоненте.

в) Покажите, что $H(X^n|S) = n/2$ и $H(X|X^\infty) = \frac{1}{2}$.

1.14.2. Пусть двоичный источник порождает с одинаковыми вероятностями всего две реализации — либо ...0101..., либо ...1010... Поэтому на любом отрезке длины n может появиться либо последовательность 0101...1, либо последовательность 1010...0, причем последний символ зависит от того, четно или нечетно n . Каждый из этих отрезков имеет вероятность $\frac{1}{2}$.

а) Покажите, что рассматриваемый источник стационарен.

б) Покажите, что он эргодичен. Для этого покажите вначале, что $M\varphi = (\varphi_1 + \varphi_2)/2$ при любом k и любой функции $\varphi(x^{(1)}, \dots, x^{(k)})$, где φ_1 и φ_2 — значения функции $\varphi(\cdot)$ и $\left| \frac{1}{n} \sum_{i=1}^n \varphi(x^{(i+1)}, \dots, x^{(i+k)}) - M\varphi \right| \leq (\varphi_1 + \varphi_2)/2n$.

в) Покажите, что укрупненный источник, сообщениями которого являются последовательные пары сообщений исходного источника, не является эргодическим. Более того, этот источник не является блочно-эргодическим для всех блоков четной длины. Таким образом, эргодический источник — не обязательно блочно-эргодический.

КРАТКИЙ ИСТОРИЧЕСКИЙ КОММЕНТАРИЙ И ЛИТЕРАТУРА

Основные результаты этой главы принадлежат К. Шеннону [8]. Он получил теоремы кодирования для равномерного и неравномерного кодирования постоянных и марковских источников. Метод построения оптимальных неравномерных кодов принадлежит Д. Хаффмену [7]. Определение эргодического процесса, данное в этой главе, является удобным, как представляется авторам, для первого введения в предмет. Более глубокое изложение, связывающее эргодическую теорию и теорию информации, можно найти у П. Биллингслея [1] и у М. С. Пинскера [3].

Изложение этой главы в достаточной степени стандартно и следует книгам Р. Фано [5] и Р. Галлагера [2], а также книге Ф. П. Тарасенко [4]. С теоретико-вероятностными вопросами лучше всего познакомиться по книге В. Феллера [6].

1. Биллингслий (Billingsley P.). Ergodic Theory and Information. — New York: Wiley, 1965. [Русский перевод: Биллингслий П. Эргодическая теория и информация. — М.: Мир, 1969.]

2. Галлагер (Gallager R.). Information Theory and Reliable Communication. — New York: Wiley, 1968. [Русский перевод: Галлагер Р. Теория информации и надежная связь. — М.: Советское радио, 1974.]
3. Пинскер М. С. Информация и информационная устойчивость случайных величин и процессов. — Проблемы передачи информации, 1960, т. 7.
4. Тарасенко Ф. П. Введение в курс теории информации. — Изд. Томского университета, 1963.
5. Фано (Fano R.). Transmission of Information. A statistical theory of communication. — New York: Wiley, 1961. [Русский перевод: Фано Р. Передача информации. Статистическая теория связи. — М.: Мир, 1965.]
6. Феллер (Feller W.). An Introduction to Probability Theory and its Application. — New York: Wiley, 1966. [Русский перевод: Феллер В. Введение в теорию вероятностей и ее приложения, т. 1. — М.: Мир, 1967.]
7. Хаффмен (Huffman D. A.). A Method for the Construction of Minimum Redundancy Codes. — Proc. IRE, 1952, т. 40, с. 1098—1101. [Русский перевод: Хаффмен Д. А. Метод построения кодов с минимальной избыточностью. — Кибернетический сб., 1961, вып. 3. — М.: ИЛ, с. 79—87.]
8. Шеннон (Shannon C. E.). A Mathematical Theory of Communications. — Bell Syst. Tech. J., 1948, 27, 379—423 (Part I), 623—656 (Part II). [Русский перевод: Шеннон К. Математическая теория связи. — В сб. Работы по теории информации и кибернетике. — М.: ИЛ, 1963.]

ВЗАЙМНАЯ ИНФОРМАЦИЯ И ЕЕ СВОЙСТВА

Эта глава носит вспомогательный характер. В ней определяется информация между сообщениями и ансамблями и изучаются ее свойства. Количество информации для дискретных ансамблей вводится с помощью понятия собственной информации, которое было определено в предыдущей главе. Количество информации для непрерывных ансамблей не может быть введено таким же образом, как для дискретных, поскольку для сообщений непрерывных ансамблей собственной информации не существует. Для того чтобы избежать трудностей, связанных с использованием понятий теории мер при описании количества информации в случае произвольных непрерывных ансамблей, мы ограничимся весьма частным с позиций современной теории вероятностей, но весьма типичным для практических приложений случаем, когда распределения вероятностей на непрерывных ансамблях задаются посредством функций плотностей вероятностей. Другой чертой изложения этой главы является двойственность, при которой по существу одни и те же факты приходится давать в двух формулировках — отдельно для дискретного и непрерывного случаев. Эта черта — следствие традиционного для технических вузов изложения математического анализа, при котором в теории интеграла дается только интеграл Римана, а интегралы Стильтьеса и Лебега — Стильтьеса не определяются. В тех случаях, когда необходимо одновременно рассматривать дискретные и непрерывные ансамбли, используется техника обобщенных функций с помощью которой вводится понятие функции плотности вероятностей для дискретных случайных величин.

§ 2.1. Количество информации между дискретными ансамблями

Пусть X и Y — два дискретных множества. Рассмотрим ансамбль $\{XY, p(x, y)\}$, который образован всевозможными парами $(x, y) \in XY$. Как указывалось выше, при задании ансамбля XY определены также ансамбли $\{X, p(x)\}$ и $\{Y, p(y)\}$, где

$$p(x) = \sum_y p(x, y); \quad p(y) = \sum_x p(x, y). \quad (2.1.1)$$

Кроме того, для каждого из сообщений $y' \in Y$ и $x' \in X$, для которых $p(y') \neq 0$ и $p(x') \neq 0$, определены условные распределения вероятностей $p(x|y')$ и $p(y|x')$, а следовательно, и условные ансамбли $\{X, p(x|y')\}$ и $\{Y, p(y|x')\}$.

В соответствии с определением 1.3.1 для каждого сообщения $x \in X$ и $y \in Y$ вводится собственная информация

$$I(x) = -\log p(x); \quad I(y) = -\log p(y) \quad (2.1.2)$$

и условная собственная информация

$$I(x|y) = -\log p(x|y); \quad I(y|x) = -\log p(y|x). \quad (2.1.3)$$

Величины (2.1.2) и (2.1.3) могут принимать конечные и бесконечные значения, но для некоторых пар сообщений условная собственная информация может быть не определена. В последнем случае эта информация при необходимости доопределяется произвольным образом. Нетрудно показать, что способ доопределения не влияет на величины средних количеств информации, т. е. на энтропии $H(X|Y)$ и $H(Y|X)$.

Определение 2.1.1. Количество информации в сообщении $x \in X$ о сообщении $y \in Y$ называется величина

$$I(x; y) \triangleq I(y) - I(y|x) = \log \frac{p(y|x)}{p(y)}. \quad (2.1.4)$$

Замечание. Количество информации $I(x; y)$ может принимать различные по знаку и величине конечные и бесконечные значения, но может быть не определено для некоторых пар сообщений. Неопределенность появляется, либо когда под знаком логарифма в (2.1.4) оказывается выражение вида $0/0$, либо когда условная вероятность не определена. Нетрудно видеть, что неопределенности не возникает, если для пары $(x, y) \in XY$ выполнены условия $p(x) \neq 0$, $p(y) \neq 0$. Неопределенность можно устранить, либо произвольным образом доопределив количество информации, либо исключив из рассмотрения сообщения $x \in X$, $y \in Y$, вероятности которых равны нулю.

Так как для любых $x \in X$ и $y \in Y$ таких, что $p(x) \neq 0$ и $p(y) \neq 0$, имеют места равенства

$$p(x, y) = p(x|y)p(y) = p(y|x)p(x), \quad (2.1.5)$$

то

$$I(x; y) = I(x) - I(x|y) \triangleq I(y; x), \quad (2.1.6)$$

т. е. количество информации в сообщении x о сообщении y равно количеству информации в сообщении y о сообщении x . Это замечание показывает, что количество информации есть симметрическая функция пары сообщений. Поэтому величину $I(x; y)$ называют *количество взаимной информации между сообщениями x и y* .

или просто *взаимной информацией* между этими сообщениями. Формуле (2.1.4) можно придать симметричную форму:

$$I(x; y) = \log \frac{p(x, y)}{p(x)p(y)}. \quad (2.1.7)$$

Рассмотрим теперь ансамбль $\{XYZ, p(x, y, z)\}$, образованный всевозможными тройками $(x, y, z) \in XYZ$, где X, Y и Z — дискретные множества. Как указывалось раньше, задание такого ансамбля одновременно задает различные условные и безусловные ансамбли. Ниже мы выпишем некоторые распределения вероятностей, определяемые данным ансамблем, которые понадобятся ниже. Пусть

$$\begin{aligned} p(x, y) &= \sum_z p(x, y, z), \\ p(y, z) &= \sum_x p(x, y, z), \\ p(x, z) &= \sum_y p(x, y, z) \end{aligned} \quad (2.1.8)$$

— безусловные распределения вероятностей на парах $(x, y) \in XY$, $(y, z) \in YZ$, $(x, z) \in XZ$ соответственно и

$$\begin{aligned} p(x) &= \sum_Y p(x, y), \\ p(z) &= \sum_Y p(y, z) \end{aligned} \quad (2.1.9)$$

— безусловные распределения вероятностей на сообщениях $x \in X$ и $z \in Z$ соответственно. Пусть, далее,

$$\begin{aligned} p(x, y | z) &= \frac{p(x, y, z)}{p(z)}, \quad p(z) \neq 0, \\ p(y, z | x) &= \frac{p(x, y, z)}{p(x)}, \quad p(x) \neq 0, \end{aligned} \quad (2.1.10)$$

— условные распределения вероятностей на XY при заданном фиксированном сообщении $z \in Z$ и на YZ при заданном фиксированном сообщении $x \in X$ соответственно. Пусть, кроме того,

$$\begin{aligned} p(y | x) &= \frac{p(x, y)}{p(x)}, \quad p(x) \neq 0, \\ p(y | xz) &= \frac{p(x, y, z)}{p(x, z)}, \quad p(x, z) \neq 0, \end{aligned} \quad (2.1.11)$$

— условные распределения вероятностей на сообщениях $y \in Y$ при фиксированных $x \in X$ и $(x, z) \in XZ$ соответственно.

С помощью определения 2.1.1 может быть введена условная информация $I(y; z|x)$ между сообщениями $y \in Y$ и $z \in Z$ при данном сообщении $x \in X$:

$$I(y; z|x) \triangleq I(y|x) - I(y|xz) = \log \frac{p(y|xz)}{p(y|x)}. \quad (2.1.12)$$

С помощью того же определения может быть введена информация между парой сообщений $(x, y) \in XY$ и сообщением $z \in Z$:

$$I((x, y); z) \triangleq I(x, y) - I((x, y)|z) = \log \frac{p(x, y|z)}{p(x, y)}. \quad (2.1.13)$$

Мы хотим представить информацию между парой (x, y) и сообщением z в виде суммы, в которой фигурировали бы информации между x и z , а также между y и z . Это можно сделать, если воспользоваться свойством аддитивности собственной информации (см. § 1.3):

$$I(x, y) = I(x) + I(y|x), \quad I((x, y)|z) = I(x|z) + I(y|xz). \quad (2.1.14)$$

Отсюда, а также из (2.1.12) и (2.1.13) следует, что

$$\begin{aligned} I((x, y); z) &= I(x) + I(y|x) - I(x|z) - I(y|xz) = \\ &= I(x; z) + I(y; z|x), \end{aligned} \quad (2.1.15)$$

$$I((x, y); z) = I(y; z) + I(x; z|y).$$

Эти последние соотношения называются *свойством аддитивности взаимной информации*.

Не останавливаясь подробно, заметим, что аналогичным образом могут быть определены и другие количества информации, скажем $I(x; y|z)$, $I((x, z); y)$ и т. д.

З а м е ч а н и е. Количества информации $I(y; z|x)$, $I((x, y); z)$, $I(x; y|z)$ и др., введенные выше для сообщений ансамбля XYZ , могут быть не определены для некоторых троек $(x, y, z) \in XYZ$. Неопределенность возникает либо при появлении выражений вида $0/0$, либо при несуществовании условных вероятностей. В каждом отдельном случае легко выписать условия, при выполнении которых неопределенности не возникает. Так, например, информация $I(y; z|x)$ определена для всех сообщений (x, y, z) , для которых $p(x, y) \neq 0$ и $p(x, z) \neq 0$. В этом примере нельзя устраниТЬ неопределенность, исключая часть сообщений из множеств X, Y и Z (ср. с замечанием, сделанным вслед за определением 2.1.1). Поэтому в дальнейшем мы будем предполагать, что в случае необходимости неопределенность устраняется произвольным доопределением рассматриваемого количества информации.

* Следует различать обозначения $I(x, y)$ и $I(x; y)$. Первое есть собственная информация пары (x, y) , а второе — взаимная информация между сообщениями x и y .

В точности так же, как в случае собственной информации, взаимную информацию можно рассматривать как случайную величину на ансамбле и вводить для нее различные числовые характеристики, и, в частности, математическое ожидание.

Пусть задан дискретный ансамбль $\{XY, p(x, y)\}$. Будем рассматривать количество взаимной информации $I(x; y)$ как функцию, отображающую элементы ансамбля XY в числовую ось. Таким образом, количество взаимной информации является случайной величиной на ансамбле XY .

Определение 2.1.2. Математическое ожидание случайной величины $I(x; y)$ на ансамбле $\{XY, p(x, y)\}$ называется *средним количеством взаимной информации* или просто *средней взаимной информацией между ансамблями* $\{X, p(x)\}$ и $\{Y, p(y)\}$ ($p(x)$ и $p(y)$ определены соотношениями (2.1.1)) и обозначается через $I(X; Y)$:

$$I(X; Y) \triangleq M_I(x; y) = \sum_{XY} p(x, y) \log \frac{p(x|y)}{p(x)}. \quad (2.1.16)$$

Легко видеть, что величина математического ожидания не зависит от способа доопределения функции $I(x; y)$, поскольку вероятность всех пар сообщений, для которых количество информации доопределено, равна нулю.

Предположим теперь, что зафиксировано некоторое сообщение $y \in Y$ (или $x \in X$), причем $p(y) \neq 0$ (или $p(x) \neq 0$). Тогда количество информации $I(x; y)$ можно рассматривать как случайную величину на ансамбле $\{X, p(x|y)\}$ (или на ансамбле $\{Y, p(y|x)\}$).

Определение 2.1.3. Математическое ожидание случайной величины $I(x; y)$ на ансамбле $\{X, p(x|y)\}$, $p(y) \neq 0$, называется *средней взаимной информацией между ансамблем* X и *сообщением* $y \in Y$ и обозначается через $I(X; y)$:

$$I(X; y) \triangleq M_y I(x; y) = \sum_x p(x|y) \log \frac{p(x|y)}{p(x)}. \quad (2.1.17)$$

Аналогичным образом определяется средняя взаимная информация между ансамблем Y и сообщением $x \in X$, $p(x) \neq 0$:

$$I(x; Y) \triangleq M_x I(x; y) = \sum_y p(y|x) \log \frac{p(y|x)}{p(y)}. \quad (2.1.18)$$

Средняя взаимная информация $I(X; y)$ или $I(x; Y)$ зависит от выбора сообщения $y \in Y$ или $x \in X$ и не определена для тех

*) Символы M_x и M_y используются для обозначения условного математического ожидания при фиксированных значениях x и y случайных величин или случайных событий соответственно.

сообщений, вероятности которых равны нулю. Если для таких сообщений $I(X; y)$ или $I(x; Y)$ произвольным образом доопределить, то среднюю взаимную информацию $I(X; y)$ можно рассматривать как случайную величину на ансамбле $\{Y; p(y)\}$, а среднюю взаимную информацию $I(x; Y)$ — как случайную величину на ансамбле $\{X, p(x)\}$. Нетрудно увидеть, что независимо от способа доопределения

$$\begin{aligned} M_I(X; y) &= \sum_Y p(y) I(X; y) = I(X; Y), \\ M_I(x; Y) &= \sum_X p(x) I(x; Y) = I(X; Y). \end{aligned} \quad (2.1.19)$$

Таким образом, среднюю взаимную информацию $I(X; Y)$ между ансамблями X и Y можно определить двояким способом, либо как в определении 2.1.2, либо как повторное математическое ожидание $M M_y I(x; y)$ или $M M_x I(x; y)$.

Поскольку взаимная информация между сообщениями была определена как разность собственных информаций (безусловной и условной), а математическое ожидание собственной информации является по определению энтропией ансамбля, то можно записать

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X). \quad (2.1.20)$$

Изучение средней взаимной информации между дискретными ансамблями мы начнем с установления простейших ее свойств.

Теорема 2.1.1. Средняя взаимная информация между сообщением, вероятность которого отлична от нуля, и ансамблем, а также средняя взаимная информация между двумя ансамблями неотрицательна.

Доказательство. Покажем только, что $I(X; y) \geq 0$, если $p(y) \neq 0$. Второе утверждение теоремы будет тогда следовать из (2.1.19). Рассмотрим величину $-I(X; y)$. Так как $p(y) \neq 0$, то существует условное распределение вероятностей $p(x|y)$ и

$$\begin{aligned} -I(X; y) &= \sum_x p(x|y) \log \frac{p(x)}{p(x|y)} \leq \\ &\leq \log e \sum_x p(x|y) \left[\frac{p(x)}{p(x|y)} - 1 \right] = 0. \end{aligned} \quad (2.1.21)$$

Последнее соотношение получается в результате применения неравенства для логарифма (1.3.7). Если $p(x, y) = p(x)p(y)$ при всех $x \in X$, то $I(X; y) = 0$. Очевидно, что средняя взаимная информация $I(X; Y)$ равна нулю в том и только том случае, когда $p(x, y) = p(x)p(y)$ для всех $x \in X$ и $y \in Y$, т. е. когда ансамбли X и Y статистически независимы.

Снова будем рассматривать ансамбль троек $\{XYZ, p(x, y, z)\}$. Для этого ансамбля определена условная взаимная информация

$I(x; y|z)$, которая при фиксированном $z \in Z$ представляет собой функцию, отображающую условный ансамбль $\{XY, p(x, y|z)\}$ на числовую ось, и поэтому является случайной величиной на этом ансамбле.

Определение 2.1.4. Математическое ожидание случайной величины $I(x; y|z)$ на условном ансамбле $\{XY, p(x, y|z)\}$ называется *средней взаимной информацией между ансамблями X и Y относительно сообщения z из ансамбля Z* и обозначается через $I(X; Y|z)$:

$$I(X; Y|z) \triangleq M_I(x; y|z) = \sum_{XY} p(x, y|z) \log \frac{p(x|yz)}{p(x|z)}. \quad (2.1.22)$$

Как и раньше, средняя взаимная информация $I(X; Y|z)$ может рассматриваться как случайная величина на ансамбле $\{z, p(z)\}$.

Определение 2.1.5. Математическое ожидание случайной величины $I(X; Y|z)$ на ансамбле $\{Z, p(z)\}$ называется *средней взаимной информацией между ансамблями X и Y относительно ансамбля Z* и обозначается через $I(X; Y|Z)$:

$$I(X; Y|Z) \triangleq MI(X; Y|z) = \sum_{XYZ} p(x, y, z) \log \frac{p(x|yz)}{p(x|z)}. \quad (2.1.23)$$

Для ансамбля $\{XYZ, p(x, y, z)\}$ определено также количество взаимной информации $I((x, y); z)$ между парой сообщений (x, y) и сообщением z . Пару (x, y) можно рассматривать как элемент ансамбля XY , тогда математическое ожидание случайной величины $I((x, y); z)$ на ансамбле XYZ представляет собой среднюю взаимную информацию между парой ансамблей XY и ансамблем Z :

$$I(XY; Z) \triangleq MI((x, y); z) = \sum_{XYZ} p(x, y, z) \log \frac{p(xy|z)}{p(x, y)}. \quad (2.1.24)$$

Из свойства аддитивности (см. 2.1.15)) тогда следует, что

$$I(XY; Z) = I(X; Z) + I(Y; Z|X) = I(Y; Z) + I(X; Z|Y), \quad (2.1.25)$$

а из (2.1.20) — что

$$I(XY; Z) = H(XY) - H(XY|Z) = H(Z) - H(Z|XY). \quad (2.1.26)$$

Одно из важнейших свойств средней взаимной информации состоит в том, что она не увеличивается при преобразованиях. Для того чтобы точно сформулировать и доказать это свойство, введем в рассмотрение некоторое преобразование $\varphi(\cdot)$, отображающее элементы множества X на элементы другого множества, скажем Z . Будем предполагать, что каждый элемент множества Z является образом некоторого (возможно, не одного) элемента из X . Будем это записывать так: $Z = \varphi(X)$. Предположим также, что задан ансамбль $\{XY, p(x, y)\}$ и тем самым определена величина

средней взаимной информации $I(X; Y)$. Преобразование $\varphi(\cdot)$ определяет ансамбль $\{ZY, p(z, y)\}$, для которого

$$p(z, y) = \sum_{x: \varphi(x)=z} p(x, y). \quad (2.1.27)$$

Поэтому средняя взаимная информация $I(Z; Y)$ определена для каждого отображения φ и принимает значения, определяемые выбором φ .

Теорема 2.1.2. Для любого отображения $Z = \varphi(X)$ ансамбля X в ансамбль Z

$$I(X; Y) \geq I(Z; Y), \quad (2.1.28)$$

причем равенство имеет место всегда, когда отображение обратимо, т. е. каждому элементу $z \in Z$ соответствует единственный элемент $x \in X$.

Доказательство. Рассмотрим множество XYZ . Так как при выбранном сообщении $x \in X$ сообщение $z \in Z$ однозначно определено и, следовательно, не зависит от сообщения $y \in Y$, то распределение вероятностей на тройках (x, y, z) , соответствующее описанному выше отображению, удовлетворяет условию

$$p(z|xy) = p(z|x) \quad (2.1.29)$$

или $p(x, y, z) = p(x, y)p(z|x)$ для всех $(x, y, z) \in XYZ$. Действительно, при данном x с вероятностью 1 $z = \varphi(x)$, т. е.

$$p(z|xy) = p(z|x) = \begin{cases} 1, & \text{если } z = \varphi(x), \\ 0, & \text{если } z \neq \varphi(x). \end{cases}$$

Из условия (2.1.29) следует, что

$$I(y; z|x) = \log \frac{p(z|xy)}{p(z|x)} = 0 \quad (2.1.30)$$

для всех $(x, y, z) \in XYZ$, для которых $p(x, y, z) \neq 0$, и, следовательно, $I(Y; Z|X) = 0$. Отсюда и из (2.1.25) следует, что

$$I(XZ; Y) = I(X; Y) + I(Y; Z|X) = I(X; Y). \quad (2.1.31)$$

С другой стороны, в силу неотрицательности средней взаимной информации $I(X; Y|Z)$

$$I(XZ; Y) = I(Z; Y) + I(X; Y|Z) \geq I(Z; Y), \quad (2.1.32)$$

что и доказывает (2.1.28).

Равенство в (2.1.28) имеет место в том случае, когда $I(X; Y|Z) = 0$. Очевидно, что последнее равенство выполняется, если для всех $(x, y, z) \in XYZ$

$$p(x|yz) = p(x|z). \quad (2.1.33)$$

Условие (2.1.33) означает, что при выбранном сообщении $z \in Z$ сообщение $x \in X$ статистически не зависит от $y \in Y$. Это условие всегда выполняется, если сообщение z однозначно определяет сообщение x , т. е. если сообщения x и z однозначно определяют друг друга и, следовательно, если отображение $\varphi(\cdot)$ обратимо. Теорема доказана.

Заметим, что в теореме 2.1.2 доказано нечто большее, чем утверждается. А именно, доказано, что неравенство (2.1.28) имеет место не только при детерминированных отображениях X в Z , но также и при произвольных случайных отображениях, определяемых распределением вероятностей $p(z|x)$, для которых выполнено условие (2.1.29).

Свойство невозрастания информации при преобразованиях имеет следующее физическое толкование.

Предположим, что имеются наблюдаемые события, образующие множество X . По этим наблюдениям мы хотим получить информацию о некотором объекте, возможные состояния которого образуют множество Y . Например, X — множество возможных сигналов на выходе некоторого канала связи, а Y — множество различных передаваемых сообщений. Теорема утверждает, что никакая обработка наблюдений, при которой происходит детерминированное или случайное их преобразование, не может увеличить средней информации об интересующем нас объекте. Информация сохраняется, если преобразование обратимо.

Очевидно, что теорема остается верной в том случае, когда преобразование осуществляется над ансамблем Y , а также в том случае, когда осуществляются преобразования как ансамбля X , так и ансамбля Y . Пусть $U = \varphi(X)$ и $V = \psi(Y)$ — два отображения, заданные на множествах X и Y соответственно. Тогда

$$I(X; Y) \geq I(U; V). \quad (2.1.34)$$

Если оба отображения обратимы, то имеет место знак равенства.

§ 2.2. Непрерывные ансамбли и источники. Обобщение понятия количества информации

Все предыдущее рассмотрение относилось только к случаю дискретных ансамблей и соответственно к случаю, когда дискретные ансамбли являлись моделями источников сообщений. Класс дискретных источников не исчерпывает всего многообразия источников, встречающихся на практике. Например, источник, порождающий речевые сообщения, не является дискретным, ибо в каждый момент времени выходным сигналом источника является некоторое действительное число — величина звукового давления.

В этом параграфе мы введем непрерывные ансамбли сообщений, которые могут служить моделями источников непрерывных сооб-

щений. Начнем с наиболее простого случая, а именно, ансамбля, соответствующего непрерывной случайной величине (с. в.).

Пусть на числовой оси задано некоторое распределение вероятностей, определяющее с. в. X , и $F(x)$ — функция распределения этой с. в., т. е. такая функция, что ее значение в точке x равно вероятности появления с. в. X в интервале $(-\infty, x]$:

$$F(x) \triangleq \Pr(-\infty < X \leq x).$$

Если существует функция $f(x)$ такая, что для всех x на числовой оси

$$F(x) = \int_{-\infty}^x f(x') dx', \quad (2.2.1)$$

то она называется *функцией плотности вероятностей* (ф. п. в.) (или просто *плотностью вероятностей*) с. в. X . Для любого интервала $(a, b]$ числовой оси вероятность появления с. в. в этом интервале определяется по формуле

$$\Pr(a, b) = F(b) - F(a) = \int_a^b f(x) dx. \quad (2.2.2)$$

Очевидно, функция $F(x)$ неотрицательна и монотонно неубывает, причем $F(-\infty) = 0$, $F(\infty) = 1$. Функция плотности вероятностей неотрицательна и ее интеграл в пределах от $-\infty$ до $+\infty$ равен единице. Последнее условие обычно называют условием нормировки. В зависимости от свойств распределений вероятностей может быть нескольких типов функции $F(x)$. Если эта функция ступенчатая и имеет конечное число ступенек, то распределение называется *дискретным* и соответствует *дискретной случайной величине*. В этом случае функции плотности в обычном смысле не существует. Если для $F(x)$ в каждой точке может быть определена производная, то распределение соответствует *непрерывной случайной величине*. Производная функции распределения в этом случае есть ф. п. в. Указанные два случая — наиболее часто встречающиеся в приложениях. Главным образом, этими случаями ограничивается рассмотрение настоящей книги. Смешанный тип распределения — это такой, когда $F(x)$ непрерывна (справа) в каждой точке, за исключением конечного числа точек, где функция распределения имеет ступеньки. Наконец, последний тип распределения имеет место, когда $F(x)$ непрерывна в каждой точке (справа), но ф. п. в. всюду или на каком-либо интервале не существует.

Пример 2.2.1. Рассмотрим дискретную случайную величину — число очков при бросании кости. Возможные значения для этой с. в. суть 1, 2, ..., 6. Оч-

видно, $F(x) = [x]/6$, где $[x]$ — целая часть x , $F(x) = 0$ при $x < 1$ и $F(x) = 1$ при $x \geq 6$ (см. рис. 2.2.1).

Пример 2.2.2. Рассмотрим непрерывную с. в., которая задается функцией распределения

$$F(x) = \begin{cases} \frac{1}{2} \exp \alpha x & \text{при } x \leq 0, \\ 1 - \frac{1}{2} \exp(-\alpha x) & \text{при } x > 0. \end{cases} \quad (2.2.3)$$

Нетрудно найти ф.п.в. этого распределения (см. рис. 2.2.2)

$$f(x) = \frac{\alpha}{\pi} \exp(-\alpha|x|). \quad (2.2.4)$$

Пример 2.2.3. На рис. 2.2.3 показана функция распределения смешанного типа. Все значения, кроме $x = 1$ и $x = 2$, имеют нулевые вероятности (но не

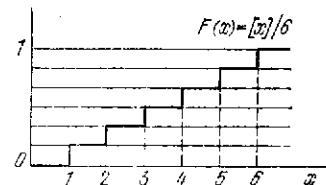


Рис. 2.2.1. Функция распределения числа очков при бросании игральной кости.

Определение 2.2.1. Непрерывным ансамблем, задаваемым ф. п. в. $f(x)$, будем называть пару $\{X, f(x)\}$, где X — числовая ось и распределение вероятностей на X задается ф. п. в. $f(x)$.

Согласно этому определению мы отождествляем понятия непрерывной действительной случайной величины, тема совместно заданных непрерывных же образом, как и в случае дискретной Y — числовые оси и XY (произведенная действительная плоскость, т. е. множество пар (x, y) , где $x \in X$ и $y \in Y$). Пусть тема распределения на множестве XY . множествах X и Y при этом опреде-

$$F_1(x) = F(x, \infty), \quad F_2(y) = F(\infty, y) \quad (2.2.5)$$

соответственно.

Определение 2.2.2. Пусть распределения вероятностей на X , Y и XY задаются ф. п. в. $f_1(x)$, $f_2(y)$ и $f(x, y)$, причем $f(x, y)$ определяется соотношением

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(x, y) dx dy, \quad (2.2.6)$$

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx. \quad (2.2.7)$$

В этом случае будем говорить, что $\{XY, f(x, y)\}$ есть система двух совместно заданных непрерывных ансамблей $\{X_1, f_1(x)\}$ и $\{Y, f_2(y)\}$.

Всякий раз, как совместно заданы два непрерывных ансамбля, определено семейство различных условных непрерывных ансамблей. Так, если фиксировано сообщение $y \in Y$, для которого $f_1(y) \neq 0$, то на множестве X определена условная ф. п. в.

$$f(x|y) \triangleq \frac{f(x,y)}{f_y(y)} \quad (2.2.8)$$

и условный непрерывный ансамбль $\{X, f(x|y)\}$. Аналогичным образом определяется условный непрерывный ансамбль $\{Y, f(y|x)\}$.

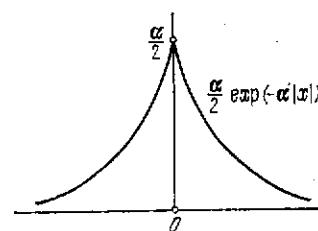


Рис. 2.2.2. Функция плотности вероятностей непрерывной с. в., примера 2.2.2.

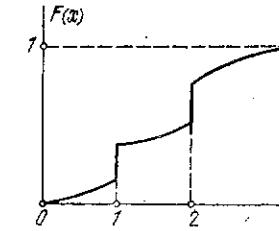


Рис. 2.2.3. Функция распределения смешанного типа.

Система более двух непрерывных ансамблей вводится в точности так же, как система двух ансамблей. Пусть $X_1 \dots X_n$ — произведение n множеств, каждое из которых является числовой прямой. Элементы множества $X_1 \dots X_n$ представляют собой действительные последовательности $(x^{(1)}, \dots, x^{(n)})$ длины n , $x^{(1)} \in X_1, \dots, x^{(n)} \in X_n$. Будем считать, что распределение вероятностей на этом множестве задается n -мерной ф. п. в. $f(x^{(1)}, \dots, x^{(n)})$. Другими словами, для любого набора интервалов $\Delta_1, \dots, \Delta_n$ вероятность попадания последовательности $(x^{(1)}, \dots, x^{(n)})$ в n -мерную область, задаваемую указанными интервалами, определяется n -кратным интегралом

$$\int_{\Delta_1} \dots \int_{\Delta_n} f(x^{(1)}, \dots, x^{(n)}) dx^{(1)} \dots dx^{(n)}$$

ПУСТЬ

$$f_1(x^{(1)}) = \int_{X_2} \dots \int_{X_n} f(x^{(1)}, \dots, x^{(n)}) dx^{(2)} \dots dx^{(n)}, \quad (2.2.9)$$

$$f_n(x^{(n)}) = \int_{X_1} \dots \int_{X_{n-1}} f(x^{(1)}, \dots, x^{(n)}) dx^{(1)} \dots dx^{(n-1)}$$

Соотношения (2.2.9) задают безусловные ф. п. в. на множествах X_1, \dots, X_n соответственно. В этом случае будем говорить, что $\{X_1, \dots, X_n, f(x^{(1)}, \dots, x^{(n)})\}$ есть система n совместно заданных непрерывных ансамблей $\{X_1, f_1(x)\}, \dots, \{X_n, f_n(x)\}$. Если

$$f(x^{(1)}, \dots, x^{(n)}) = f_1(x^{(1)}) \dots f_n(x^{(n)}) \quad (2.2.10)$$

для любых $x^{(1)} \in X_1, \dots, x^{(n)} \in X_n$, то непрерывные ансамбли X_1, \dots, X_n называются статистически независимыми.

Заметим, что при задании системы n непрерывных ансамблей фактически оказываются заданными всевозможные системы по $m \ll n$ непрерывных ансамблей.

Пусть $\{X, f_1(x)\}$ — непрерывный ансамбль и $\phi(x)$ — произвольная действительная функция на X , отображающая числовую прямую X в себя *). Всякая такая функция порождает некоторый ансамбль и называется случайной величиной на ансамбле $\{X, f_1(x)\}$. Если $\phi(x)$ — ступенчатая функция с конечным числом значений, то она порождает дискретный ансамбль $\{Y, p(y)\}$, где $Y = \{y_1, \dots, y_N\}$ — множество значений функции $\phi(x)$ и

$$p(y_j) = \int_{x: \phi(x) = y_j} f_1(x) dx.$$

Если $\phi(x)$ — непрерывная функция, то она порождает непрерывный ансамбль $\{Y, f_2^*(y)\}$, где Y — числовая ось, а ф. п. в. $f_2^*(y)$ определяется из уравнения

$$\int_{-\infty}^y f_2^*(z) dz = \int_{x: \phi(x) \leq y} f_1(x) dx.$$

Для каждой с. в. определены ее числовые характеристики, например, математическое ожидание и дисперсия. Все числовые характеристики непрерывных случайных величин определяются так же, как и в случае дискретных случайных величин (см. § 1.2) с заменой вероятностей на ф. п. в. и сумм на интегралы. Очевидно, что для непрерывных с. в. справедливо неравенство Чебышева (1.2.8) и закон больших чисел (теорема 1.2.1) (см. также задачи 2.2.8—2.2.10).

Рассмотрим теперь совместное задание непрерывного ансамбля X и дискретного ансамбля Y_d . Для этого удобно рассматривать дискретный ансамбль как результат дискретизации некоторого непрерывного ансамбля Y . Пусть $\{XY, f(x, y)\}$ — пара совместно заданных непрерывных ансамблей $\{X, f_1(x)\}$ и $\{Y, f_2(y)\}$. Пусть B_1, \dots, B_N — разбиение множества Y на непересекающиеся под-

*) Для наших целей достаточно рассматривать непрерывные функции с конечным числом разрывов первого рода. Строго говоря, рассматриваемые функции должны удовлетворять некоторым дополнительным условиям измеримости.

множества. Введем в рассмотрение дискретное множество $Y_d = \{y_1, \dots, y_N\}$ и будем говорить, что происходит событие y_j , если точка непрерывного ансамбля $\{Y, f_2(y)\}$ попадает в множество B_j . Для каждого y вероятность $p_2(y_j)$ этого события определяется формулой

$$p_2(y_j) = \int_{B_j} f_2(y) dy. \quad (2.2.11)$$

Переход от непрерывного ансамбля $\{Y, f_2(y)\}$ к дискретному называется дискретизацией. Очевидно, что любой дискретный ансамбль можно себе представлять как результат дискретизации некоторого непрерывного.

Так же, как и в дискретном случае, разбиение Y на подмножества задает семейство условных плотностей вероятностей на множестве X

$$f(x|y_j) = f(x|B_j) \triangleq \frac{1}{p_2(y_j)} \int_{B_j} f(x, y) dy, \quad p_2(y_j) \neq 0. \quad (2.2.12)$$

Отсюда и из (2.2.7) следует, что

$$f_1(x) = \sum_{j=1}^N f(x|y_j) p_2(y_j). \quad (2.2.13)$$

С другой стороны, на дискретном множестве Y_d определены условные вероятности

$$p(y_j|x) \triangleq \int_{B_j} \frac{f(x, y)}{f_1(x)} dy = \frac{f(x|y_j) p_2(y_j)}{f_1(x)}, \quad f_1(x) \neq 0. \quad (2.2.14)$$

Таким образом, ансамбль X задается ф. п. в. $f_1(x)$, ансамбль Y_d — распределением вероятностей $p_2(y_j)$, а ансамбль XY_d задается двумя эквивалентными способами: либо посредством функции

$$f_1(x) p(y_j|x), \quad x \in X, \quad y_j \in Y_d, \quad (2.2.15)$$

либо посредством функции

$$f(x|y_j) p_2(y_j), \quad x \in X, \quad y_j \in Y_d. \quad (2.2.16)$$

Соотношение (2.2.14) показывает, что обе эти функции совпадают и определяют следующую функцию распределения на множестве XY_d :

$$F(x, y) = \begin{cases} \int_{-\infty}^x \sum_{y_j \leq y} f_1(x) p(y_j|x) dx, \\ \int_{-\infty}^y \sum_{x: f(x|y_j) \neq 0} f(x|y_j) p_2(y_j) dx. \end{cases} \quad (2.2.17)$$

При рассмотрении различных функций на числовых множествах и изучении их свойств часто бывает необходимо рассматривать одновременно и дискретные и непрерывные распределения вероятностей. Такое общее рассмотрение можно осуществить с помощью функций распределений. Однако многие результаты, относящиеся к непрерывным распределениям, традиционно излагаются в терминах плотностей вероятностей и в этой форме имеют более простой вид, чем в форме, использующей функции распределения. Поэтому желательно также иметь описание дискретных распределений с помощью функций, имеющих смысл функций плотностей вероятностей. Ниже мы рассмотрим дельта-функцию Дирака (одну из так называемых обобщенных функций) и используем ее для задания плотностей вероятностей дискретных случайных величин.

Дельта-функция Дирака, $\delta(x)$, определяется следующим формальным равенством:

$$\int_{\Delta} \delta(x) \varphi(x) dx \triangleq \begin{cases} \varphi(0), & \text{если } 0 \in \Delta, \\ 0 & \text{в противном случае,} \end{cases} \quad (2.2.18)$$

где $\varphi(x)$ — произвольная непрерывная функция и Δ — произвольный интервал на числовой оси. Очевидно, что

$$\int_{\Delta} \delta(x - x_0) \varphi(x) dx = \begin{cases} \varphi(x_0), & \text{если } x_0 \in \Delta, \\ 0 & \text{в противном случае,} \end{cases} \quad (2.2.19)$$

и

$$\int_{\Delta} \delta(x - x_0) dx = \begin{cases} 1, & \text{если } x_0 \in \Delta, \\ 0 & \text{в противном случае.} \end{cases} \quad (2.2.20)$$

Дельта-функцию можно умножать на число, складывать с другими дельта-функциями и складывать с обычными интегрируемыми функциями. Отдельно аналитические свойства дельта-функции не исследуются (хотя и можно представлять функцию $\delta(x)$, например, как предел последовательности сужающихся импульсов с единичной площадью). Она имеет смысл только в выражениях вида (2.2.18).

Покажем теперь, как используется дельта-функция для описания дискретных с. в. Пусть $X = \{x_1, \dots, x_M\}$ — произвольное числовое множество и $\{p_1, \dots, p_M\}$ — распределение вероятностей на X . Очевидно, что функция распределения $F(x) = \sum p(x_i)$, где суммирование выполняется по всем таким индексам i , что $x_i \leq x$. В этом случае

$$f^*(x) \triangleq \sum_{i=1}^M p_i \delta(x - x_i) \quad (2.2.21)$$

есть формальное выражение для обобщенной ф. п. в. указанного дискретного распределения, так как

$$F(x) = \int_{-\infty}^x f^*(x') dx'. \quad (2.2.22)$$

Если на произведении XY дискретных множеств $X = \{x_1, \dots, x_M\}$, $Y = \{y_1, \dots, y_N\}$ задано распределение вероятностей $p(x_i, y_j)$, то обобщенная ф. п. в. этого распределения

$$f^*(x, y) \triangleq \sum_{i, j} p(x_i, y_j) \delta(x - x_i) \delta(y - y_j). \quad (2.2.23)$$

Безусловные ф. п. в. определяются обычным образом:

$$f_1^*(x) = \int f^*(x, y) dy = \sum_{i, j} p(x_i, y_j) \delta(x - x_i),$$

$$f_2^*(y) = \int f^*(x, y) dx = \sum_{i, j} p(x_i, y_j) \delta(y - y_j).$$

Отношение выражений, содержащих дельта-функции, вообще говоря, не определено, и поэтому условные ф. п. в., например $f^*(x|y) = f^*(x, y)/f^*(y)$, не являются ни обычными, ни обобщенными ф. п. в. Однако, если положить

$$\frac{\sum_{j=1}^N a_j \delta(y - y_j)}{\sum_{j=1}^N b_j \delta(y - y_j)} \triangleq \frac{a_i}{b_i}, \quad y = y_i, \quad i = 1, \dots, N, \quad (2.2.24)$$

для всех выражений такого вида с ненулевыми коэффициентами b_i , то

$$f^*(x|y) = \frac{\sum_{i, j} p(x_i, y_j) \delta(x - x_i) \delta(y - y_j)}{\sum_{i, j} p(x_i, y_j) \delta(y - y_j)} = \sum_{i=1}^N \frac{p(x_i, y_j)}{p_2(y_j)} \delta(x - x_i),$$

$$y = y_j, \quad j = 1, \dots, N, \quad (2.2.25)$$

где

$$p_2(y_j) = \sum_i p(x_i, y_j),$$

и функция $f^*(x|y)$ становится обобщенной ф. п. в. для всех $y = y_j$, $j = 1, 2, \dots, N$. Для остальных y эта функция не определена. Кроме того,

$$\frac{f^*(x, y)}{f_1^*(x) f_2^*(y)} = \frac{p(x_i, y_j)}{p_1(x_i) p_2(y_j)}, \quad x = x_i, \quad i = 1, \dots, M, \\ y = y_j, \quad j = 1, \dots, N, \quad (2.2.26)$$

где

$$p_1(x_i) = \sum_j p(x_i, y_j).$$

Введем теперь количество информации между сообщениями непрерывных ансамблей. В дискретном случае взаимная информация определялась через количество собственной информации в сообщении (см. определение 2.1.1). Однако в случае непрерывных ансамблей вероятность каждого отдельного сообщения равна нулю, и, следовательно, собственная информация сообщений бесконечна. С физической точки зрения бесконечно большая собственная информация соответствует тому, что всякая непрерывная с. в. принимает бесконечное число значений, каждое из которых можно рассматривать как некоторое сообщение. Хотя собственная информация сообщений непрерывного ансамбля бесконечно велика, взаимная информация между парой сообщений, как мы увидим ниже, как правило, ограничена.

Определение 2.2.3. Количество взаимной информации между сообщениями $x \in X$, $y \in Y$ непрерывного ансамбля $\{XY, f(x, y)\}$ называется величина

$$I(x; y) \triangleq \log \frac{f(x, y)}{f(x)f(y)} ^*), \quad (2.2.27)$$

определенная для всех пар (x, y) таких, что $f(x) \neq 0$, $f(y) \neq 0$. Для остальных пар сообщений количество взаимной информации $I(x; y)$ в случае необходимости доопределяется произвольным образом.

Заметим, что из (2.2.26) вытекает совпадение формул (2.2.27) и (2.1.7), определяющих взаимную информацию для непрерывного и дискретного случаев, если под $f(x, y)$, $f(x)$ и $f(y)$ понимать обобщенные ф. п. в. дискретных с. в.

Дополнительным основанием для указанного выше определения взаимной информации служит следующее рассуждение, основанное на дискретизации непрерывных ансамблей и предельном

*) Как и в дискретном случае, если нет опасности перепутать, мы используем один и тот же символ $f(\cdot)$ для обозначения нескольких, быть может, различных ф. п. в.

переходе. Пусть $(x, x + \Delta x)$ и $(y, y + \Delta y)$ — интервалы на осях X и Y соответственно и

$$P_{XY} \triangleq \int_x^{x+\Delta x} \int_y^{y+\Delta y} f(x', y') dx' dy' \approx f(x, y) \Delta x \Delta y, \\ P_X \triangleq \int_x^{x+\Delta x} f_1(x') dx' \approx f_1(x) \Delta x, \\ P_Y \triangleq \int_y^{y+\Delta y} f_2(y') dy' \approx f_2(y) \Delta y \quad (2.2.28)$$

— соответствующие этим интервалам вероятности. С каждым интервалом можно связать событие из некоторого дискретного множества. Так, предположим, что x_k и y_l — события, состоящие в том, что точка из X принадлежит интервалу $(x, x + \Delta x)$ и точка из Y принадлежит интервалу $(y, y + \Delta y)$ соответственно. Между событиями x_k и y_l определена взаимная информация

$$I(x_k; y_l) = \log \frac{P_{XY}}{P_X \cdot P_Y}. \quad (2.2.29)$$

Устремляя $\Delta x \rightarrow 0$, $\Delta y \rightarrow 0$, получим

$$\lim_{\Delta x \rightarrow 0, \Delta y \rightarrow 0} I(x_k; y_l) = \log \frac{f(x, y)}{f_1(x) \cdot f_2(y)}.$$

Этот предел и служит взаимной информацией между сообщениями (x, y) непрерывных ансамблей.

Рассмотрим теперь непрерывный ансамбль $\{XYZ, f(x, y, z)\}$. По аналогии с (2.2.27) можно определить условную информацию между парой сообщений $x \in X$ и $y \in Y$ при фиксированном сообщении $z \in Z$

$$I(x; y | z) \triangleq \log \frac{f(y | xz)}{f(y | z)} = \log \frac{f(x, y | z)}{f(x | z) f(y | z)} \quad (2.2.30)$$

и информацию между парой сообщений $(x, y) \in XY$ и третьим сообщением $z \in Z$

$$I((x, y); z) \triangleq \log \frac{f(x, y | z)}{f(x, y)}, \quad (2.2.31)$$

где условные и безусловные ф. п. в. определяются соотношениями, аналогичными (2.1.8)–(2.1.11), в которых суммы заменены на интегралы, а вероятности — на ф. п. в.

Каждое из количеств информации, определяемое соотношениями (2.2.27), (2.2.30) и (2.2.31), представляет собой случайную

величину на соответствующем непрерывном ансамбле. Математическое ожидание

$$I(X; Y) \triangleq MI(x; y) = \int_{X Y} f(x, y) \log \frac{f(x, y)}{f(x)f(y)} dx dy \quad (2.2.32)$$

называется *средней взаимной информацией между непрерывными ансамблями X и Y*. Математическое ожидание

$$I(X; Y|z) \triangleq M_z I(x; y|z) = \int_{X Y} f(x, y|z) \log \frac{f(x, y|z)}{f(x|z)f(y|z)} dx dy \quad (2.2.33)$$

называется *средней взаимной информацией между непрерывными ансамблями X и Y относительно сообщения z ансамбля Z*. Математическое ожидание

$$I(X; Y|Z) \triangleq MI(x; y|z) = \int_{X Y Z} f(x, y, z) \log \frac{f(x, y|z)}{f(x|z)f(y|z)} dx dy dz \quad (2.2.34)$$

называется *средней взаимной информацией между непрерывными ансамблями X и Y относительно непрерывного ансамбля Z*. Математическое ожидание

$$I(XY; Z) \triangleq MI((x, y); z) = \int_{X Y Z} f(x, y, z) \log \frac{f(x, y|z)}{f(x, y)} dx dy dz \quad (2.2.35)$$

называется *средней взаимной информацией между парой непрерывных ансамблей XY и непрерывным ансамблем Z*.

Пример 2.2.4. Пусть $f(x, y)$ — ф. п. в. двумерного гауссовского (нормального) закона распределения вероятностей на плоскости (см. [2]):

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x-m_X)^2}{\sigma_X^2} - \frac{2\rho(x-m_X)(y-m_Y)}{\sigma_X\sigma_Y} + \frac{(y-m_Y)^2}{\sigma_Y^2} \right] \right\}. \quad (2.2.36)$$

Эта функция зависит от пяти параметров m_X , m_Y , σ_X^2 , σ_Y^2 и ρ . Первые четыре параметра представляют собой математические ожидания и дисперсии соответствующих одномерных распределений, которые также являются гауссовскими:

$$f(x) = \int_{-\infty}^{\infty} f(x, y) dy = \frac{1}{\sigma_X\sqrt{2\pi}} \exp \left[-\frac{1}{2\sigma_X^2}(x-m_X)^2 \right], \quad (2.2.37)$$

$$f(y) = \int_{-\infty}^{\infty} f(x, y) dx = \frac{1}{\sigma_Y\sqrt{2\pi}} \exp \left[-\frac{1}{2\sigma_Y^2}(y-m_Y)^2 \right]. \quad (2.2.38)$$

Параметр ρ называется коэффициентом корреляции.

Для такого распределения вероятностей нетрудно вычислить величину средней взаимной информации. Используя в этом месте натуральные логарифмы и применяя формулу (2.2.32), получим

$$\begin{aligned} I(X; Y) &= M \ln \frac{f(x, y)}{f(x)f(y)} = \ln \frac{1}{\sqrt{1-\rho^2}} + \\ &+ M \left\{ -\frac{1}{2} \left[\frac{(x-m_X)^2}{(1-\rho^2)\sigma_X^2} - \frac{2\rho(x-m_X)(y-m_Y)}{(1-\rho^2)\sigma_X\sigma_Y} + \right. \right. \\ &\left. \left. + \frac{(y-m_Y)^2}{(1-\rho^2)\sigma_Y^2} - \frac{(x-m_X)^2}{\sigma_X^2} - \frac{(y-m_Y)^2}{\sigma_Y^2} \right] \right\} = \\ &= -\frac{1}{2} \ln (1-\rho^2) - \frac{1}{2} \left[\frac{1}{1-\rho^2} - \frac{2\rho^2}{1-\rho^2} + \frac{1}{1-\rho^2} - 1 - 1 \right] = \\ &= -\frac{1}{2} \ln (1-\rho^2) \text{ (нат)} = -\frac{1}{2} \log (1-\rho^2) \text{ (бит)}, \end{aligned} \quad (2.2.39)$$

где использованы определения дисперсии и коэффициента корреляции:

$$\begin{aligned} M(x-m_X)^2 &\triangleq \sigma_X^2, \quad M(y-m_Y)^2 \triangleq \sigma_Y^2, \\ M \frac{(x-m_X)(y-m_Y)}{\sigma_X \cdot \sigma_Y} &\triangleq \rho. \end{aligned} \quad (2.2.40)$$

Таким образом, средняя взаимная информация между двумя непрерывными совместно гауссовскими ансамблями (с. в.), имеющими коэффициент корреляции ρ , определяется выражением

$$I(X; Y) = -\frac{1}{2} \log (1-\rho^2). \quad (2.2.41)$$

Во всех приведенных выше определениях средней взаимной информации (2.2.32)–(2.2.35) некоторые ансамбли могут быть непрерывными, а другие — дискретными. В этом случае приведенные формулы сохраняются, если ф. п. в. рассматривать как обобщенные ф. п. в. Возможно также представить эти формулы и в виде, использующем дискретные распределения. Для этого интеграл по соответствующей переменной должен быть заменен суммой, а ф. п. в. — выражениями вида (2.2.15), (2.2.16) с соответствующими обобщениями на случай более двух переменных. В качестве примера приведем формулу для средней взаимной информации $I(X; Y|Z)$ для случая, когда X , Y — непрерывные ансамбли, а Z — дискретный:

$$I(X; Y|Z) = \sum_Z \int_{X Y} f(x, y|z_i) p(z_i) \log \frac{f(x, y|z_i)}{f(x|z_i)f(y|z_i)} dx dy. \quad (2.2.42)$$

Средняя взаимная информация между непрерывными или между непрерывными и дискретными ансамблями обладает мно-

гими из свойств, которые были раньше сформулированы для случая дискретных ансамблей. Так, теоремы 2.1.1 и 2.1.2 остаются справедливыми. Доказательства почти полностью повторяют соответствующие доказательства в дискретном случае. Читателю предлагается доказать указанные теоремы самостоятельно.

В заключение этого параграфа — несколько замечаний о непрерывных источниках сообщений. Определение непрерывного источника (с дискретным временем) опирается на определение последовательности непрерывных ансамблей. Все, что было сказано о задании дискретных источников, переносится без существенных изменений на случай непрерывных источников. В частности, без изменений остается определение 1.1.2, в котором для задания непрерывного источника требуется согласованное задание всех n -мерных распределений вероятностей отрезков сообщений длины n , $n = 1, 2, \dots$. Также без изменений остаются определения стационарного источника, источников без памяти и стационарного источника без памяти (определения 1.1.3, 1.1.4). Единственное отличие непрерывного случая от дискретного состоит в том, что вероятности заменяются на ф. п. в., а суммы — на интегралы.

§ 2.3. Относительная энтропия и ее свойства

Практически все свойства средней взаимной информации являются общими для дискретных и непрерывных ансамблей. Эта общность является следствием того, что и дискретный и непрерывный случаи являются частными в общей абстрактной схеме введения информационной меры на измеримых вероятностных пространствах *). Единственное отличие состоит в том, что для непрерывных ансамблей не определена собственная информация сообщений и, как следствие, не определена энтропия. Поэтому представление информации в виде разности энтропий (2.1.20) имеет место только в дискретном случае.

Однако можно ввести некоторые аналоги энтропий и в непрерывном случае и получить представление, похожее на (2.1.20). Рассмотрим среднюю взаимную информацию (2.2.32) между непрерывными ансамблями X и Y . Используя условные функции плотности вероятностей, можно записать

$$I(X; Y) = \int_{X \times Y} f(x, y) \log \frac{f(x|y)}{f(x)} dx dy = \int_{X \times Y} f(x, y) \log \frac{f(y|x)}{f(y)} dx dy. \quad (2.3.1)$$

*) Интересующийся читатель может найти такое общее изложение в статье [5] или в книге [10].

Если обозначить

$$\begin{aligned} H_0(X) &\triangleq - \int_X f(x) \log f(x) dx, \quad H_0(Y) \triangleq - \int_Y f(y) \log f(y) dy, \\ H_0(X|Y) &\triangleq - \int_X \int_Y f(x, y) \log f(x|y) dx dy, \\ H_0(Y|X) &\triangleq - \int_X \int_Y f(x, y) \log f(y|x) dx dy, \end{aligned} \quad (2.3.2)$$

то, используя (2.2.7), получим, что

$$I(X; Y) = H_0(X) - H_0(X|Y) = H_0(Y) - H_0(Y|X). \quad (2.3.3)$$

Величины $H_0(\cdot)$, если существуют соответствующие интегралы, называются *относительными* (или *дифференциальными*) *энтропиями непрерывных ансамблей*. Они имеют много общих свойств с энтропиями дискретных ансамблей.

Первое свойство, отличающее относительную энтропию от энтропии дискретных ансамблей, состоит в том, что она может принимать различные по знаку значения. Это будет показано с помощью следующих примеров, в которых вычислена относительная энтропия для некоторых простых распределений вероятностей.

Пример 2.3.1. Пусть $f(x)$ — ф. п. в. равномерного на отрезке (a, b) распределения:

$$f(x) = \begin{cases} 1/(b-a) & \text{для } x \in (a, b), \\ 0 & \text{для остальных } x. \end{cases}$$

Для ансамбля с таким распределением относительная энтропия $H_0(X) = -\log(b-a)$. Она принимает отрицательные значения, если $(b-a) < 1$.

Пример 2.3.2. Пусть $f(x)$ — ф. п. в. гауссовского распределения вероятностей с нулевым средним и дисперсией σ^2 :

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-\frac{1}{2\sigma^2} x^2 \right]. \quad (2.3.4)$$

Тогда

$$\begin{aligned} H_0(X) &= - \int_{-\infty}^{\infty} f(x) \log \sigma \sqrt{2\pi} dx + \frac{\log e}{2\sigma^2} \int_{-\infty}^{\infty} x^2 f(x) dx = \\ &= \frac{1}{2} \log 2\pi\sigma^2 + \frac{1}{2} \log e = \frac{1}{2} \log 2\pi e\sigma^2, \end{aligned} \quad (2.3.5)$$

где использовано условие нормировки $\int_{-\infty}^{\infty} f(x) dx = 1$, а также то, что математическое ожидание равно нулю и, следовательно, $\int_{-\infty}^{\infty} x^2 f(x) dx = \sigma^2$,

Пусть $\{XY, f(x, y)\}$ — непрерывный ансамбль, образованный парой совместно заданных непрерывных ансамблей $\{X, f(x)\}$ и $\{Y, f(y)\}$. Величина

$$H_0(XY) \triangleq - \int \int f(x, y) \log f(x, y) dx dy \quad (2.3.6)$$

называется *относительной энтропией ансамбля XY*. Представляя $f(x, y)$ в виде произведения условной и безусловной ф. п. в., получим

$$H_0(XY) = H_0(X) + H_0(Y | X) = H_0(Y) + H_0(X | Y), \quad (2.3.7)$$

т. е. относительная энтропия обладает свойством аддитивности.

Если $\{X_1 \dots X_n, f(\mathbf{x})\}$ совместно заданные n непрерывных ансамблей $\mathbf{x} = (x^{(1)}, \dots, x^{(n)})$, $x^{(i)} \in X_i$, $i = 1, \dots, n$, то, используя соотношение $f(\mathbf{x}) = f(x^{(1)})f(x^{(2)} | x^{(1)}) \dots f(x^{(n)} | x^{(1)}, \dots, x^{(n-1)})$, получим

$$\begin{aligned} H_0(X_1 \dots X_n) &\triangleq - \int f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x} = \\ &= H_0(X_1) + H_0(X_2 | X_1) + \dots + H_0(X_n | X_1 \dots X_{n-1}). \end{aligned} \quad (2.3.8)$$

С помощью неравенства для логарифма (1.3.7) легко доказывается, что для непрерывных ансамблей X и Y

$$H_0(Y | X) \leq H_0(Y) \quad (2.3.9)$$

с равенством в том и только том случае, когда ансамбли X и Y статистически независимы, т. е. когда выполняется соотношение (2.2.10).

Рассмотрим один важный частный случай, когда выражение для условной относительной энтропии можно упростить. Пусть X и Y — две случайные величины, связанные равенством $Y = X + Z$, где с. в. Z статистически не зависит от X . Обозначим через $f_Z(\cdot)$ ф. п. в. этой с. в. Тогда, как нетрудно увидеть,

$$f(y | x) = f_Z(y - x). \quad (2.3.10)$$

Действительно, левая часть это ф. п. в. с. в. Y при фиксированном значении $X = x$. Так как при этом с. в. Y и Z отличаются только математическим ожиданием (математическое ожидание Y равно математическому ожиданию Z плюс x), то имеет место (2.3.10). Отсюда следует, что при любом $x \in X$

$$\begin{aligned} H_0(Y | x) &\triangleq M_x(-\log f(y | x)) = M_x(-\log f_Z(y - x)) = \\ &= M_x(-\log f_Z(z)) = M(-\log f_Z(z)) \triangleq H_0(Z), \end{aligned} \quad (2.3.11)$$

где третье равенство — результат замены переменных, а четвертое — следствие независимости X и Z . Усредняя обе части (2.3.11) по всем x , получим, что для рассматриваемых с. в.

$$H_0(Y | X) \triangleq M H_0(Y | x) = H_0(Z). \quad (2.3.12)$$

Относительная энтропия определяется распределением вероятностей на ансамбле, и естественным является вопрос о том, для каких распределений она больше. Однако такой вопрос без дополнительных ограничивающих предположений лишен смысла, поскольку, как видно, например, из (2.3.4) или (2.3.5), относительная энтропия может быть сделана сколь угодно большой либо соответствующим выбором интервала (a, b) в первом случае, либо выбором параметра c^2 во втором.

Пусть \mathbb{C} — такой класс ф. п. в. на числовой прямой, что для каждой функции $f(x) \in \mathbb{C}$ выполняется условие

$$\int_{-\infty}^{\infty} x^2 f(x) dx \leq c^2. \quad (2.3.13)$$

Слева в этом неравенстве написан второй начальный момент распределения с ф. п. в. $f(x)$. Он называется *средней мощностью* с. в., ф. п. в. которой есть $f(x)$. Название связано с тем, что в случае, когда x есть напряжение, то x^2 есть мощность в единичном сопротивлении. Таким образом, \mathbb{C} — множество ф. п. в. для с. в. со средней мощностью, ограниченной числом c^2 .

Теорема 2.3.1. Для любой ф. п. в. $f(x) \in \mathbb{C}$ выполняется неравенство

$$H_0(X) \leq \frac{1}{2} \log(2\pi e c^2), \quad (2.3.14)$$

где $H_0(X)$ — относительная энтропия ансамбля $\{X, f(x)\}$. Равенство имеет место в том случае, когда

$$f(x) = \frac{1}{c\sqrt{2\pi}} \exp\left[-\frac{x^2}{2c^2}\right], \quad (2.3.15)$$

т. е. когда распределение вероятностей является гауссовским и имеет нулевое среднее и дисперсию c^2 .

Доказательство. Докажем вначале вспомогательное неравенство. Пусть $f(x)$ — произвольная функция из \mathbb{C} , тогда

$$\begin{aligned} &- \int_{-\infty}^{\infty} f(x) \log \frac{1}{c\sqrt{2\pi}} \exp\left[-\frac{x^2}{2c^2}\right] dx = \\ &= \frac{1}{2} \log 2\pi e^2 + \frac{\log c}{2c^2} \int_{-\infty}^{\infty} x^2 f(x) dx \leq \frac{1}{2} \log 2\pi e c^2. \end{aligned} \quad (2.3.16)$$

Из этого неравенства следует, что

$$\begin{aligned} H_0(X) - \frac{1}{2} \log 2\pi e^2 &\leq \\ &\leq - \int_{-\infty}^{\infty} f(x) \log f(x) dx + \int_{-\infty}^{\infty} f(x) \log \frac{1}{c\sqrt{2\pi}} \exp\left(-\frac{x^2}{2c^2}\right) dx = \\ &= \int_{-\infty}^{\infty} f(x) \log \frac{\exp\left(-\frac{x^2}{2c^2}\right)}{f(x) c\sqrt{2\pi}} dx. \end{aligned}$$

Отсюда, применяя неравенство для логарифма, получим

$$\begin{aligned} H_0(X) - \frac{1}{2} \log 2\pi e^2 &\leq \log e \int_{-\infty}^{\infty} f(x) \left[\frac{\exp\left(-\frac{x^2}{2c^2}\right)}{f(x) c\sqrt{2\pi}} - 1 \right] dx = \\ &= \log e \left[\frac{1}{c\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2c^2}\right) dx - \int_{-\infty}^{\infty} f(x) dx \right] = 0, \quad (2.3.17) \end{aligned}$$

где последнее равенство следует из того, что каждое из выражений в квадратных скобках равно единице. Таким образом, получено неравенство (2.3.14). Случай равенства получается, когда имеет место равенство в неравенстве для логарифма, т. е. когда ф. п. в. определяется формулой (2.3.15). Теорема доказана.

Таким образом, мы показали, что среди всех случайных величин с ограниченным средним квадратом наибольшей относительной энтропией обладает гауссовская случайная величина. То, что эта случайная величина имеет нулевое математическое ожидание, не является существенным требованием. Легко показать (см. задачу 2.3.2), что случайные величины, отличающиеся только математическим ожиданием, имеют одинаковые относительные энтропии. Поэтому упоминание в теореме 2.3.1 о том, что математическое ожидание равно нулю, можно опустить.

Рассмотрим теперь относительную энтропию системы случайных величин X_1, \dots, X_n с совместной ф. п. в. $f(x^{(1)}, \dots, x^{(n)})$. Обозначим через m_1, \dots, m_n математические ожидания этих величин:

$$\begin{aligned} m_i \triangleq M X_i &= \int x^{(i)} f(x^{(1)}, \dots, x^{(n)}) dx^{(1)} \dots dx^{(n)} = \\ &= \int x^{(i)} f_i(x^{(i)}) dx^{(i)}, \quad i = 1, \dots, n, \quad (2.3.18) \end{aligned}$$

где функции $f_i(x^{(i)})$, $i = 1, \dots, n$, определяются соотношениями (2.2.9). Обозначим через K_{ij} корреляционный момент с. в. X_i и X_j :

$$K_{ij} \triangleq M(X_i - m_i)(X_j - m_j) = \int (x^{(i)} - m_i)(x^{(j)} - m_j) f(x^{(1)}, \dots, x^{(n)}) dx^{(1)} \dots dx^{(n)}. \quad (2.3.19)$$

Матрица $\mathbf{K} = [K_{ij}]$, $i, j = 1, \dots, n$, элементами которой являются корреляционные моменты K_{ij} , называется *корреляционной матрицей* системы с. в. X_1, \dots, X_n .

Относительная энтропия системы с. в. X_1, \dots, X_n определяется соотношением

$$H_0(X_1, \dots, X_n) \triangleq \int f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x}, \quad (2.3.20)$$

где $f(\mathbf{x}) = f(x^{(1)}, \dots, x^{(n)})$. Мы хотим показать, что в n -мерном случае сохраняется свойство экстремальности гауссовского распределения вероятностей по отношению к относительной энтропии, доказанное для одномерного случая в теореме 2.3.1. Для этого чтобы это сделать, необходимо ввести некоторые дополнительные определения.

Обозначим через \mathbf{K}^{-1} матрицу, обратную матрице \mathbf{K} , т. е. такую, что

$$\mathbf{K}^{-1} \mathbf{K} = \mathbf{I}_n, \quad (2.3.21)$$

где \mathbf{I}_n — единичная матрица порядка n . Обратная матрица \mathbf{K}^{-1} всегда существует, если матрица \mathbf{K} обладает ненулевым определителем, что в дальнейшем и будет предполагаться:

$$\det \mathbf{K} \neq 0. \quad (2.3.22)$$

Так как $K_{ij} = K_{ji}$, то матрица \mathbf{K} и, как нетрудно доказать, матрица \mathbf{K}^{-1} являются симметрическими. Из определения обратной матрицы и определения матричного умножения следует, что

$$\sum_{i=1}^n K_{ij}^* K_{ji} = \begin{cases} 1, & \text{если } i = i', \\ 0 & \text{в противном случае,} \end{cases} \quad (2.3.23)$$

где K_{ij}^* — элементы матрицы \mathbf{K}^{-1} , а отсюда из симметричности матрицы \mathbf{K} следует, что

$$\sum_{i, j=1}^n K_{ij}^* K_{ij} = n. \quad (2.3.24)$$

Функция

$$\begin{aligned} f_G(x^{(1)}, \dots, x^{(n)}) &\triangleq \\ &\triangleq (2\pi)^{-\frac{n}{2}} (\det \mathbf{K})^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \sum_{i, j} (x^{(i)} - m_i)(x^{(j)} - m_j) K_{ij}^* \right] \end{aligned} \quad (2.3.25)$$

называется ф. п. в. n -мерного невырожденного гауссовского распределения вероятностей. Если с. в. X_1, \dots, X_n имеют ф. п. в. (2.3.25), то эти с. в. называются *согласно гауссовскими*. Можно показать, что в этом случае каждая из с. в. системы также имеет гауссовское распределение вероятностей.

Теорема 2.3.2. Пусть \mathfrak{C}_n — класс ф. п. в. n с. в. с заданными значениями математических ожиданий m_1, \dots, m_n и заданной корреляционной матрицей K , $\det K \neq 0$. Для любой функции $f(\mathbf{x}) \in \mathfrak{C}_n$ выполняется неравенство

$$H_0(X_1, \dots, X_n) \leq \frac{n}{2} \log 2\pi e + \frac{1}{2} \log \det K, \quad (2.3.26)$$

причем равенство имеет место в том случае, когда $f(\mathbf{x})$ есть ф. п. в. n -мерного гауссовского распределения вероятностей с математическими ожиданиями m_1, \dots, m_n и корреляционной матрицей K , т. е. когда $f(\mathbf{x}) = f_G(\mathbf{x})$.

Доказательство. Докажем вначале вспомогательное равенство. Пусть $f(\mathbf{x})$ — произвольная функция из \mathfrak{C}_n , тогда

$$\begin{aligned} F &\triangleq - \int f(\mathbf{x}) \log (2\pi)^{-\frac{n}{2}} (\det K)^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \sum_{i,j} (x^{(i)} - m_i) \times \right. \\ &\quad \times \left. (x^{(j)} - m_j) K_{ij}^* \right] d\mathbf{x} = \frac{n}{2} \log 2\pi + \frac{1}{2} \log \det K + \\ &\quad + \frac{\log e}{2} \int f(\mathbf{x}) \sum_{i,j} (x^{(i)} - m_i) (x^{(j)} - m_j) K_{ij}^* d\mathbf{x}. \end{aligned} \quad (2.3.27)$$

Очевидно (см. (2.3.19) и (2.3.24)), что

$$\begin{aligned} \int f(\mathbf{x}) \sum_{i,j} (x^{(i)} - m_i) (x^{(j)} - m_j) K_{ij}^* d\mathbf{x} &= \\ &= \sum_{i,j} K_{ij}^* \int f(\mathbf{x}) (x^{(i)} - m_i) (x^{(j)} - m_j) d\mathbf{x} = \sum_{i,j} K_{ij}^* K_{ij} = n, \end{aligned} \quad (2.3.28)$$

откуда

$$F = \frac{n}{2} \log 2\pi e + \frac{1}{2} \log \det K. \quad (2.3.29)$$

Используя это равенство и неравенство для логарифма, получим

$$\begin{aligned} H_0(X_1, \dots, X_n) - F &= - \int f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x} + \int f(\mathbf{x}) \log f_G(\mathbf{x}) d\mathbf{x} = \\ &= \int f(\mathbf{x}) \log \frac{f_G(\mathbf{x})}{f(\mathbf{x})} d\mathbf{x} \leq \int f(\mathbf{x}) \left[\frac{f_G(\mathbf{x})}{f(\mathbf{x})} - 1 \right] d\mathbf{x} = 0, \end{aligned} \quad (2.3.30)$$

что и доказывает теорему.

Как и в одномерном случае, легко доказать, что относительная энтропия системы с. в. не зависит от математических ожиданий, поэтому упоминание о математических ожиданиях в теореме 2.3.2 можно опустить.

Пример 2.3.3. Здесь мы дадим пример использования теоремы 2.3.2 для вычисления средней взаимной информации между двумя гауссовскими случайными векторами. Пусть $\mathbf{X} = (X_1, \dots, X_n)$ — вектор, образованный системой гауссовских с. в. X_1, \dots, X_n с нулевыми средними и корреляционной матрицей K_X . Пусть $\mathbf{Z} = (Z_1, \dots, Z_n)$ — такой же вектор с корреляционной матрицей K_Z , $\det K_Z \neq 0$; будем считать, что векторы \mathbf{X} и \mathbf{Z} статистически независимы. Нас интересует средняя взаимная информация $I(\mathbf{X}; \mathbf{Y})$ между векторами \mathbf{X} и $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$. Как было показано выше,

$$I(\mathbf{X}; \mathbf{Y}) = H_0(\mathbf{Y}) - H_0(\mathbf{Y} | \mathbf{X}). \quad (2.3.31)$$

Так как \mathbf{Y} — гауссовский вектор и корреляционная матрица этого вектора

$$K_Y = [\mathbf{M} Y_i Y_j] = [\mathbf{M} (X_i + Z_i) (X_j + Z_j)] = [\mathbf{M} X_i X_j] + [\mathbf{M} Z_i Z_j] = K_X + K_Z, \quad (2.3.32)$$

где третье равенство есть следствие независимости векторов \mathbf{X} и \mathbf{Z} , то $\det(K_X + K_Z) \neq 0$ (см. задачу 2.4.3) и

$$H_0(\mathbf{Y}) = \frac{n}{2} \log 2\pi e + \frac{1}{2} \log \det(K_X + K_Z). \quad (2.3.33)$$

Условная относительная энтропия $H_0(\mathbf{Y} | \mathbf{X}) = H_0(\mathbf{Z})$, поскольку $f(\mathbf{y} | \mathbf{x}) = f_Z(\mathbf{y} - \mathbf{x})$, где $f_Z(\cdot)$ — ф. п. в. вектора \mathbf{Z} . Аргументация здесь в точности такая же, как и при получении соотношения (2.3.12). Следовательно,

$$H_0(\mathbf{Y} | \mathbf{X}) = \frac{n}{2} \log 2\pi e + \frac{1}{2} \log \det K_Z \quad (2.3.34)$$

и

$$I(\mathbf{X}; \mathbf{Y}) = \frac{1}{2} \log \frac{\det(K_X + K_Z)}{\det K_Z}. \quad (2.3.35)$$

§ 2.4 *. Ортогональные преобразования случайных векторов

Из рассмотрения относительной энтропии n -мерных распределений вероятностей (системы n случайных величин) видно, что при $n > 1$ возникают серьезные технические трудности, если только с. в. не являются статистически независимыми. В этом параграфе будет показано, что многие трудности можно преодолеть с помощью преобразования системы координат, при котором данная система с. в. переходит в систему некоррелированных с. в.

Пусть X_1, \dots, X_n — система n с. в. Мы будем рассматривать эту систему как случайный вектор $\mathbf{X} = (X_1, \dots, X_n)$ в n -мерном евклидовом пространстве. Пусть $K = [K_{ij}]$ — корреляционная матрица вектора \mathbf{X} , где K_{ij} — корреляционные моменты, определяемые соотношением (2.3.19). Здесь мы не будем предполагать, что $\det K$ обязательно отличен от нуля, т. е. рассматриваются и вырожденные распределения.

Всякая корреляционная $n \times n$ -матрица является симметрической, т. е. $\mathbf{K}^\tau = \mathbf{K}$, где « τ » — символ транспонирования (замены местами строк и столбцов), и неотрицательно определенной. Последнее означает, что для любого ненулевого вектора $\mathbf{z} = (z^{(1)}, \dots, z^{(n)})$

$$\sum_{i,j=1}^n z^{(i)} z^{(j)} K_{ij} = \mathbf{z} \mathbf{K} \mathbf{z}^\tau \geq 0, \quad (2.4.1)$$

где равенство — следствие определения матричного произведения вектора-строки \mathbf{z} , матрицы \mathbf{K} и вектора-столбца \mathbf{z}^τ . Если в (2.4.1) имеет место строгое неравенство, то матрица \mathbf{K} называется положительно определенной. Неотрицательная определенность корреляционной матрицы следует из того, что ее можно представить в виде

$$\mathbf{K} = \mathbf{M}(\mathbf{X} - \mathbf{m})^\tau (\mathbf{X} - \mathbf{m}), \quad (2.4.2)$$

где \mathbf{m} — вектор математических ожиданий и справа написано математическое ожидание произведения вектора-столбца $(\mathbf{X} - \mathbf{m})^\tau$ на вектор-строку $(\mathbf{X} - \mathbf{m})$, т. е. математическое ожидание матрицы размера $n \times n$. Для любого неслучайного вектора \mathbf{z}

$$\mathbf{z} \mathbf{K} \mathbf{z}^\tau = \mathbf{M} \mathbf{z} (\mathbf{X} - \mathbf{m})^\tau (\mathbf{X} - \mathbf{m}) \mathbf{z}^\tau = \mathbf{M} \xi^2 \geq 0, \quad (2.4.3)$$

где использована перестановочность взятия математического ожидания и суммирования, а также введена с. в.

$$\xi \triangleq \mathbf{z} (\mathbf{X} - \mathbf{m})^\tau. \quad (2.4.4)$$

Известно (см., например, [1]), что для любой симметрической матрицы \mathbf{K} существует такая система ортонормальных векторов $\mathbf{q}_1, \dots, \mathbf{q}_n$, что

$$\mathbf{q}_i \mathbf{K} = \lambda_i \mathbf{q}_i, \quad i = 1, 2, \dots, n. \quad (2.4.5)$$

Действительные числа $\lambda_1, \dots, \lambda_n$ называются *собственными числами*, а векторы $\mathbf{q}_1, \dots, \mathbf{q}_n$ — *собственными векторами* матрицы \mathbf{K} , причем \mathbf{q}_i — собственный вектор, соответствующий собственному числу λ_i . Систему равенств (2.4.5) можно записать в матричном виде:

$$\mathbf{Q} \mathbf{K} \mathbf{Q}^\tau = \Lambda, \quad (2.4.6)$$

где

$$\mathbf{Q} \triangleq \begin{bmatrix} \mathbf{q}_1 \\ \vdots \\ \mathbf{q}_n \end{bmatrix} \quad (2.4.7)$$

— матрица, строки которой — собственные векторы матрицы \mathbf{K} , а $\Lambda = [\lambda_1, \dots, \lambda_n]$ — диагональная матрица с элементами $\lambda_1, \dots, \lambda_n$.

на главной диагонали. В силу ортонормальности собственных векторов матрица \mathbf{Q} является *ортогональной*, т. е.

$$\mathbf{Q} \mathbf{Q}^\tau = \mathbf{I}_n, \quad (2.4.8)$$

где \mathbf{I}_n — единичная $n \times n$ -матрица.

Используя хорошо известные свойства определителей и, в частности, то, что $\det \mathbf{A}^\tau = \det \mathbf{A}$, $\det \mathbf{AB} = \det \mathbf{A} \cdot \det \mathbf{B}$, можно получить, что

$$|\det \mathbf{Q}| = 1, \quad \det \mathbf{K} = \lambda_1 \lambda_2 \dots \lambda_n. \quad (2.4.9)$$

Всякая квадратная $n \times n$ -матрица \mathbf{A} задает некоторое линейное преобразование n -мерного векторного пространства в себя, при котором вектор \mathbf{x} переводится в вектор $\mathbf{y} = \mathbf{x}\mathbf{A}$. Рассмотрим линейное преобразование, задаваемое ортогональной матрицей \mathbf{Q} . Такое преобразование называется *ортогональным*. Для любого вектора \mathbf{x}

$$\mathbf{y} = \mathbf{x}\mathbf{Q}. \quad (2.4.10)$$

Заметим, что всякое линейное преобразование переводит нулевой вектор в себя, а всякое ортогональное преобразование сохраняет длину вектора. Действительно, если $\mathbf{x} = (x^{(1)}, \dots, x^{(n)})$, то

$$\left(\sum_{i=1}^n (x^{(i)})^2 \right)^{1/2} = (\mathbf{x} \mathbf{x}^\tau)^{1/2} \quad (2.4.11)$$

есть длина вектора \mathbf{x} . Очевидно,

$$\mathbf{y} \mathbf{y}^\tau = \mathbf{x} \mathbf{Q} (\mathbf{x} \mathbf{Q})^\tau = \mathbf{x} \mathbf{Q} \mathbf{Q}^\tau \mathbf{x}^\tau = \mathbf{x} \mathbf{x}^\tau, \quad (2.4.12)$$

т. е. длина преобразованного вектора \mathbf{y} равна длине исходного вектора \mathbf{x} . Это свойство ортогонального преобразования позволяет его интерпретировать как вращение системы координат n -мерного векторного пространства вокруг начала координат.

Пусть \mathbf{X} — случайный вектор с корреляционной матрицей \mathbf{K}_X и \mathbf{Q} — ортогональная матрица, образованная собственными векторами матрицы \mathbf{K}_X . Пусть $\mathbf{Y} = (Y_1, \dots, Y_n)$ — случайный вектор, получающийся в результате ортогонального преобразования с матрицей \mathbf{Q} из случайного вектора $\mathbf{X} = (X_1, \dots, X_n)$:

$$\mathbf{Y} = \mathbf{X}\mathbf{Q}. \quad (2.4.13)$$

Обозначим через \mathbf{K}_Y корреляционную матрицу вектора \mathbf{Y} . Легко видеть (см. (2.4.2) и (2.4.8)), что

$$\begin{aligned} \mathbf{K}_Y &= \mathbf{M}(\mathbf{Y} - \mathbf{m}_Y)^\tau (\mathbf{Y} - \mathbf{m}_Y) = \mathbf{M}(\mathbf{X}\mathbf{Q} - \mathbf{m}_X\mathbf{Q})^\tau (\mathbf{X}\mathbf{Q} - \mathbf{m}_X\mathbf{Q}) = \\ &= \mathbf{M}\mathbf{Q}^\tau (\mathbf{X} - \mathbf{m}_X)^\tau (\mathbf{X} - \mathbf{m}_X)\mathbf{Q} = \mathbf{Q}^\tau \mathbf{K}_X \mathbf{Q} = \mathbf{Q} \mathbf{K}_X \mathbf{Q}^\tau = \Lambda, \end{aligned} \quad (2.4.14)$$

где \mathbf{m}_X и \mathbf{m}_Y — векторы математических ожиданий.

Таким образом, корреляционная матрица вектора \mathbf{Y} является диагональной, причем элементы главной диагонали суть собствен-

ные числа матрицы K_X . Это означает, что с. в. Y_1, \dots, Y_n являются некоррелированными и дисперсия с. в. Y_i

$$M(Y_i - m_{Y_i})^2 = \lambda_i, \quad i = 1, 2, \dots, n, \quad (2.4.15)$$

т. е. равна i -му собственному числу корреляционной матрицы K_X .

Если X — гауссовский случайный вектор, то Y — также гауссовский случайный вектор. При этом случайные величины Y_1, \dots, Y_n являются независимыми и гауссовскими, т. е.

$$f(y^{(1)}, \dots, y^{(n)}) = \prod_{i=1}^n f_i(y^{(i)}), \quad (2.4.16)$$

где

$$f_i(y^{(i)}) = \frac{1}{\sqrt{2\pi\lambda_i}} \exp\left(-\frac{1}{2\lambda_i}(y^{(i)} - m_{Y_i})^2\right). \quad (2.4.17)$$

Мы показали, что любую систему гауссовских с. в. с помощью ортогонального преобразования (поворота системы координат) можно преобразовать в систему независимых гауссовских с. в. Верно и обратное, если $Y = (Y_1, \dots, Y_n)$ — система независимых гауссовских с. в. с математическими ожиданиями m_{Y_1}, \dots, m_{Y_n} и дисперсиями $\lambda_1, \dots, \lambda_n$ и Q — произвольная ортогональная матрица, то

$$X = YQ \quad (2.4.18)$$

есть гауссовский случайный вектор, компоненты которого имеют математические ожидания

$$m_X = m_Y Q \quad (2.4.19)$$

и корреляционная матрица K_X которой есть

$$K_X = Q K_Y Q^T = Q \Lambda Q^T. \quad (2.4.20)$$

При этом строки матрицы Q есть собственные векторы, а $\lambda_1, \dots, \lambda_n$ — собственные числа матрицы K_X . Если $\lambda_i \neq 0, i = 1, \dots, n$, то матрица K_X невырождена, т. е. $\det K_X \neq 0$, и система с. в. X_1, \dots, X_n имеет невырожденное n -мерное гауссовское распределение с ф. п. в. (2.3.25).

Пример 2.4.1. Пусть $X = (X_1, X_2)$ — гауссовский случайный вектор на плоскости с ф. п. в. (2.2.36). Предположим, что $\sigma_{X_1}^2 = \sigma_{X_2}^2 = \sigma^2$, тогда корреляционная матрица этого вектора есть

$$K_X = \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}, \quad (2.4.21)$$

где ρ — коэффициент корреляции с. в. X_1 и X_2 . Нетрудно проверить, что векторы

$$\begin{aligned} q_1 &= (1/\sqrt{2}, 1/\sqrt{2}), \\ q_2 &= (1/\sqrt{2}, -1/\sqrt{2}) \end{aligned} \quad (2.4.22)$$

— собственные векторы матрицы K_X , соответствующие собственным числам

$$\begin{aligned} \lambda_1 &= \sigma^2(1 + \rho), \\ \lambda_2 &= \sigma^2(1 - \rho). \end{aligned} \quad (2.4.23)$$

Ортогональное преобразование, задаваемое ортогональной матрицей

$$Q = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}, \quad (2.4.24)$$

переводит случайный вектор X в случайный вектор $Y = XQ = (Y_1, Y_2)$ с независимыми гауссовскими компонентами Y_1, Y_2 . Очевидно, что математические ожидания и дисперсии этих с. в.

$$\begin{aligned} m_{Y_1} &= \frac{m_{X_1} + m_{X_2}}{\sqrt{2}}, & m_{Y_2} &= \frac{m_{X_1} - m_{X_2}}{\sqrt{2}}, \\ \sigma_{Y_1}^2 &= \sigma^2(1 + \rho), & \sigma_{Y_2}^2 &= \sigma^2(1 - \rho). \end{aligned} \quad (2.4.25)$$

Заметим, что ортогональное преобразование есть обратимое преобразование, поскольку $\det Q \neq 0$.

В следующем примере мы покажем, как можно использовать ортогональные преобразования для вычисления средней взаимной информации между гауссовскими случайными векторами.

Пример 2.4.2. Пусть $X = (X_1, \dots, X_n)$ — гауссовский случайный вектор с корреляционной матрицей K_X , пусть $Z = (Z_1, \dots, Z_n)$ — такой же вектор с корреляционной матрицей K_Z , который будем считать статистически независимым от вектора X . В отличие от примера 2.3.3, здесь будет предполагаться, что K_Z — диагональная матрица с одинаковыми ненулевыми элементами v_1, \dots, v_1 на главной диагонали. Нас будет интересовать средняя взаимная информация $I(X; Y)$ между векторами X и $Y = X + Z$. Хотя $I(X; Y)$ вычислено в примере 2.3.3, мы слова рассмотрим вычисление этой величины с использованием ортогонального преобразования.

Пусть Q — ортогональная матрица, соответствующая матрице K_X . Тогда

$$Q K_X Q^T = \Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}, \quad (2.4.26)$$

где $\lambda_1, \dots, \lambda_n$ — собственные числа матрицы K_X и

$$Q K_Z Q^T = K_Z. \quad (2.4.27)$$

Отсюда следует, что

$$Q K_Y Q^T = Q K_X Q^T + Q K_Z Q^T = \begin{bmatrix} \lambda_1 + v_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n + v_1 \end{bmatrix}, \quad (2.4.28)$$

т. е. ортогональная матрица Q диагонализирует также корреляционную матрицу K_Y вектора Y , и, следовательно, вектор YQ имеет независимые ком-

компоненты с дисперсиями $\lambda_i + v_1$, $i = 1, 2, \dots, n$. Далее из обратимости ортогонального преобразования вытекает, что

$$\begin{aligned} I(X; Y) &= H_0(Y) - H_0(Y|X) = H_0(YQ) - H_0(Z) = \\ &= \frac{1}{2} \sum_{i=1}^n \log 2\pi e (\lambda_i + v_1) - \frac{1}{2} \sum_{i=1}^n \log 2\pi e v_1 = \frac{1}{2} \sum_{i=1}^n \log \left(1 + \frac{\lambda_i}{v_1} \right), \end{aligned} \quad (2.4.29)$$

где второе равенство есть следствие того, что ортогональное преобразование не меняет относительную энтропию (см. задачу 2.4.4) и независимости векторов X и Z , а третье равенство — следствие того, что относительная энтропия системы независимых с. в. равна сумме относительных энтропий каждой из с. в. системы. Нетрудно увидеть, что в рассматриваемом случае выражения (2.3.35) и (2.4.29) совпадают.

§ 2.5*. Выпукłość средней взаимной информации

Пусть V — линейное пространство. Область R линейного пространства V называется выпуклой, если для любого λ , $0 \leq \lambda \leq 1$, и любых двух элементов $v_1, v_2 \in R$ элемент $\lambda v_1 + (1 - \lambda) v_2$ также принадлежит R .

Хотя многое из того, что будет сказано в этом параграфе, справедливо для произвольных линейных пространств и их выпуклых областей, мы вначале будем конкретизировать линейное пространство в одной из следующих двух форм. В случае, удобном для исследования свойств выпуклости информации между дискретными ансамблями, мы будем понимать под V конечномерное векторное пространство, элементами которого являются векторы с действительными компонентами. Вектор называется вероятностным, если его компоненты неотрицательны и в сумме дают единицу. Очевидно, что множество вероятностных векторов является выпуклой областью векторного пространства *). В другом случае, удобном для исследования непрерывных ансамблей, мы будем понимать под V множество действительных функций $f(x)$ с действительным аргументом $x \in X$, X — числовая ось. Функция называется функцией плотности вероятностей, если она неотрицательна и ее интеграл по X равен единице. Очевидно, что множество ф. п. в. является выпуклой областью линейного пространства V , указанного выше.

Определение 2.5.1. Функция $\varphi(v)$ называется выпуклой ввысь в выпуклой области R , если для любого $k = 1, 2, \dots$, любых неотрицательных $\lambda_1, \dots, \lambda_k$, $\sum \lambda_i = 1$, и любых $v_1, \dots, v_k \in R$ выполняется неравенство

$$\sum_{i=1}^k \lambda_i \varphi(v_i) \leq \varphi \left(\sum_{i=1}^k \lambda_i v_i \right). \quad (2.5.1)$$

*) Действительно, если v_1 и v_2 — вероятностные векторы, то для любого λ , $0 < \lambda < 1$, $\lambda v_1 + (1 - \lambda) v_2$ — вероятностный вектор, так как его компоненты неотрицательны и в сумме дают единицу.

Если в (2.5.1) имеет место противоположное неравенство, то функция $\varphi(v)$ называется выпуклой вниз *).

Геометрически неравенство (2.5.1) при $k = 2$ означает, что хорда, соединяющая две любые точки поверхности $\varphi(v)$, лежит под этой поверхностью или на ней. Нетрудно показать (см. задачу 2.5.1), что приведенное геометрическое утверждение эквивалентно определению 2.5.1, т. е. определение достаточно давать для случая $k = 2$.

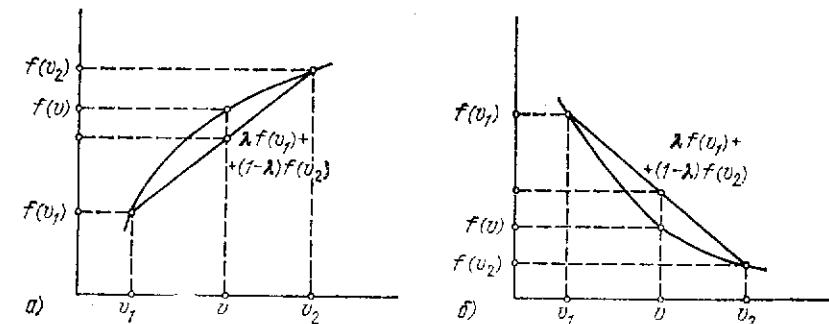


Рис. 2.5.1. Функция, выпуклая вверх (a) и выпуклая вниз (б).

На рис. 2.5.1 приведен пример функции одного аргумента, выпуклой ввысь (a) и выпуклой вниз (б). Для любых чисел v_1, v_2 и неотрицательного λ , $0 < \lambda < 1$, число $v = \lambda v_1 + (1 - \lambda) v_2$ является средним с весами λ и $1 - \lambda$ из чисел v_1, v_2 . При этом $f(v)$ — ордината функции $f(\cdot)$ в точке $\lambda v_1 + (1 - \lambda) v_2$, а $\lambda f(v_1) + (1 - \lambda) f(v_2)$ — ордината хорды, соединяющей точки $f(v_1)$, $f(v_2)$ на графике функции. Выпуклость ввысь означает, что ордината хорды не больше, чем соответствующая ордината функции.

Пусть XY — дискретный ансамбль с совместным распределением вероятностей $p(x, y)$. Средняя взаимная информация между ансамблями X и Y определяется выражением

$$I(X; Y) = \sum_X \sum_Y p(x) p(y|x) \log \frac{p(y|x)}{p(y)}. \quad (2.5.2)$$

Аналогичное выражение для непрерывного ансамбля $\{XY, f(x, y)\}$ вид

$$I(X; Y) = \int \int f(x) f(y|x) \log \frac{f(y|x)}{f(y)} dx dy. \quad (2.5.3)$$

*) Используются также другие названия — вогнутая (в первом случае) и выпуклая (во втором) функция.

Пусть считается фиксированным условное распределение вероятностей, задаваемое условными вероятностями $p(y|x)$ в дискретном и условными ф. п. в. $f(y|x)$ — в непрерывном случае. Тогда средняя взаимная информация в дискретном случае является функцией безусловного распределения вероятностей на множестве X , т. е. функцией вероятностного вектора $\mathbf{p} = (p_1, \dots, p_M)$, $p_i \triangleq \sum p(x_i)$, $i = 1, 2, \dots, M$, $\sum p_i = 1$. В непрерывном случае она является функционалом от безусловной ф. п. в. $f(x)$. И в том, и в другом случае $I(X; Y)$ можно рассматривать как функцию (функционал), определенную на элементах выпуклой области R соответствующего линейного пространства, и говорить о выпуклости функции $I(X; Y)$ в области R .

Теорема 2.5.1. Средняя взаимная информация $I(X; Y)$ является выпуклой вверх функцией в области R всех безусловных распределений вероятностей (вероятностных векторов или ф. п. в.) на множестве X при условии, что условные распределения на множестве Y (условные вероятности или условные ф. п. в.) зафиксированы.

Доказательство. Вначале мы сведем утверждение теоремы о выпуклости к некоторому неравенству между средними информациами, а затем докажем это неравенство. Метод доказательства является одним и тем же как для дискретного, так и для непрерывного случаев. Специфика состоит только в описании рассматриваемых ансамблей. Для упрощения мы будем подробно рассматривать только дискретный случай. В заключение будет намечено доказательство для непрерывного случая.

Введем в рассмотрение множество XYZ , где X и Y — данные множества, а множество Z содержит k элементов, которые обозначим через z_1, \dots, z_k . Предположим, что распределение вероятностей на тройках задано следующим образом:

$$p(x, y, z_i) \triangleq p(x|z_i)p(y|x)p(z_i), \quad (2.5.4)$$

где $p(y|x)$ — данное условное распределение. Тогда при каждом $z_i \in Z$ на множестве XY определено условное распределение вероятностей

$$p(x, y|z_i) = p(x|z_i)p(y|x). \quad (2.5.5)$$

Каждому такому распределению и, следовательно, каждому элементу z_i из Z соответствует средняя взаимная информация $I(X; Y|z_i)$, определяемая формулой (2.1.22). С другой стороны, на множестве XY определено безусловное распределение вероятностей

$$p(x, y) = \sum_Z p(x, y|z_i)p(z_i) = p(x)p(y|x), \quad (2.5.6)$$

где

$$p(x) = \sum_{i=1}^k p(x|z_i)p(z_i), \quad (2.5.7)$$

и, следовательно, определена безусловная средняя взаимная информация $I(X; Y)$.

Каждое из условных распределений $p(x|z_i)$ представляет собой вероятностный вектор, а величина $I(X; Y|z_i)$ — значение исследуемой функции — средней взаимной информации — на этом векторе. Если обозначить $p(z_i) = \lambda_i$, то левая часть неравенства (2.5.1) в нашем случае есть

$$\sum_{i=1}^k \lambda_i I(X; Y|z_i) = I(X; Y|Z), \quad (2.5.8)$$

где использовано определение 2.1.5 средней взаимной информации относительно ансамбля Z . С другой стороны, из (2.5.7) следует, что $\{p(x)\}$, $x \in X$, есть вероятностный вектор, равный среднему с весами $\lambda_1, \dots, \lambda_k$ из вероятностных векторов $\{p(x|z_i)\}$, $x \in X$, $i = 1, 2, \dots, k$. Поэтому правая часть неравенства (2.5.1) есть просто $I(X; Y)$. Таким образом, для доказательства теоремы теперь необходимо доказать неравенство

$$I(X; Y|Z) \leq I(X; Y) \quad (2.5.9)$$

в предположении, что справедливо (2.5.4).

Для доказательства этого неравенства рассмотрим среднюю взаимную информацию $I(Y; XZ)$. По свойству аддитивности имеем $I(Y; XZ) = I(X; Y) + I(Y; Z|X) = I(Y; Z) + I(X; Y|Z)$. (2.5.10)

Из (2.5.4) следует, что для всех $(x, y, z) \in XYZ$

$$\frac{p(yz|x)}{p(y|x)p(z|x)} = \frac{p(x|z)p(y|x)p(z)}{p(x)p(y|x)p(z|x)} = 1, \quad (2.5.11)$$

т. е. $I(y; z|x) = 0$. Поэтому

$$I(Y; Z|X) = 0. \quad (2.5.12)$$

Теперь, учитывая неотрицательность средней взаимной информации, из (2.5.10) и (2.5.12) получим (2.5.9). Теорема доказана для случая дискретных ансамблей.

Непрерывный случай отличается от рассмотренного только тем, что распределения на тройках из XYZ следует задать посредством функций

$$f(x|z_i)f(y|x)p(z_i). \quad (2.5.13)$$

Эта функция задает распределение смешанного типа (см. (2.2.13)), при котором ансамбли X и Y непрерывны, а ансамбль Z дискретен. В остальном доказательство полностью сохраняется.

Для рассмотрения второго случая при исследовании выпуклости средней взаимной информации необходимо снова вернуться к абстрактному линейному пространству и рассмотреть другие две конкретизации этого пространства. Мы теперь будем предполагать, что в случае дискретных ансамблей линейное пространство V

есть пространство всех действительных матриц с конечным числом строк и столбцов. Матрица называется *стохастической*, если каждая ее строка есть вероятностный вектор. Очевидно, что множество стохастических матриц есть выпуклая область пространства V . В случае непрерывных ансамблей будем предполагать, что элементами V являются все действительные функции двух переменных, которые будем обозначать символом $f(y|x)$, $x \in X$, $y \in Y$, X и Y — действительные оси. Функция $f(y|x)$ называется условной функцией плотности вероятностей, если при каждом $x \in X$ она неотрицательна и ее интеграл по Y равен единице. Множество условных ф. п. в. является выпуклой областью в описанном линейном пространстве. Действительно, если $f_1(y|x)$ и $f_2(y|x)$ — две условные ф. п. в., то для любого λ , $0 < \lambda < 1$, функция $f(y|x) \triangleq \lambda f_1(y|x) + (1 - \lambda) f_2(y|x)$ при каждом значении $x \in X$ неотрицательна и

$$\int_Y f(y|x) dy = \lambda \int_Y f_1(y|x) dy + (1 - \lambda) \int_Y f_2(y|x) dy = 1.$$

Следовательно, $f(y|x)$ — условная ф. п. в.

Рассмотрим выражения (2.5.2), (2.5.3) для средней взаимной информации и будем считать, что в них безусловные распределения вероятностей на одном из ансамблей, например X , фиксированы. Для дискретного случая это означает, что фиксированы вероятности $p(x)$, $x \in X$, а для непрерывного это означает, что фиксирована ф. п. в. $f(x)$, $x \in X$. Тогда средняя взаимная информация является функцией условных распределений вероятностей на множестве Y , т. е. функцией от стохастической матрицы с элементами $p(y|x)$, $x \in X$, $y \in Y$, в дискретном случае или функционалом от условной ф. п. в. $f(y|x)$, $x \in X$, $y \in Y$, — в непрерывном. И в том, и в другом случае среднюю взаимную информацию $I(X; Y)$ можно рассматривать как функцию, определенную на элементах выпуклой области R соответствующего линейного пространства, и говорить о выпуклости функции (функционала) в области R .

Теорема 2.5.2. Средняя взаимная информация $I(X; Y)$ является выпуклой вниз функцией в области R всех условных распределений вероятностей (стохастических матриц или условных ф. п. в.) на множестве Y при условии, что безусловное распределение вероятностей на множестве X (вероятности или ф. п. в.) зафиксировано.

Доказательство. Так же, как и при доказательстве теоремы 2.5.1, мы вначале сведем утверждение теоремы к некоторому неравенству между средними информаций, а затем докажем это неравенство. Метод доказательства этой теоремы является общим и для дискретного и для непрерывного случаев. Отличие состоит только в описании рассматриваемых ансамблей. Как и

раньше, мы подробно рассмотрим только дискретный случай и наметим доказательство для непрерывного случая.

Пусть XYZ — произведение трех множеств, где X и Y — это данные множества, а Z — дискретное множество, состоящее из k элементов z_1, \dots, z_k . Предположим, что распределение вероятностей на тройках задано следующим образом:

$$p(x, y, z_i) \triangleq p(x)p(y|xz_i)p(z_i), \quad (2.5.14)$$

где $p(x)$ — данное безусловное распределение на X . Тогда при каждом $z_i \in Z$ на множестве XY задано условное распределение вероятностей

$$p(x, y|z_i) = p(x)p(y|xz_i). \quad (2.5.15)$$

Каждому такому распределению и, следовательно, каждому элементу $z_i \in Z$ соответствует средняя взаимная информация $I(X; Y|z_i)$, определяемая по формуле (2.1.22). С другой стороны, на множестве XY задано безусловное распределение вероятностей:

$$p(x, y) = \sum_{i=1}^k p(x, y|z_i)p(z_i) = p(x)p(y|x), \quad (2.5.16)$$

где

$$p(y|x) = \sum_{i=1}^k p(y|xz_i)p(z_i) \quad (2.5.17)$$

и, следовательно, определена безусловная средняя взаимная информация $I(X; Y)$.

Каждое из условных распределений $p(y|xz_i)$ может быть записано в форме стохастической матрицы, в которой строки соответствуют сообщениям $x \in X$, а элементы в строке — сообщениям $y \in Y$. Поэтому $I(X; Y|z_i)$ можно рассматривать как значение исследуемой функции — средней взаимной информации — на этой матрице. Если обозначить $p(z_i) = \lambda_i$, то левая часть неравенства (2.5.1) в рассматриваемом случае может быть записана в форме

$$\sum_{i=1}^k \lambda_i I(X; Y|z_i) = I(X; Y|Z), \quad (2.5.18)$$

где использовано определение 2.1.5. С другой стороны, из (2.5.17) следует, что $[p(y|x)]$ является стохастической матрицей, равной средней с весами $\lambda_1, \dots, \lambda_k$ из стохастических матриц $[p(y|xz_i)]$, $i = 1, \dots, k$. Поэтому правая часть неравенства (2.5.1) есть просто $I(X; Y)$. Таким образом, для доказательства теоремы теперь необходимо доказать неравенство

$$I(X; Y|Z) \geq I(X; Y) \quad (2.5.19)$$

в предположении, что выполняется (2.5.14).

Для доказательства этого неравенства рассмотрим среднюю взаимную информацию $I(X; YZ)$. По свойству аддитивности имеем $I(X; YZ) = I(X; Y) + I(X; Z|Y) = I(X; Z) + I(X; Y|Z)$. (2.5.20)

Из (2.5.14) следует, что для всех $(x, y, z) \in XYZ$

$$\frac{p(xz)}{p(x)p(z)} = \frac{\sum_y p(x, y, z)}{p(x)p(z)} = 1, \quad (2.5.21)$$

т. е. $I(x; z) = 0$ и поэтому

$$I(X; Z) = 0. \quad (2.5.22)$$

Теперь, учитывая неотрицательность информации $I(X; Z|Y)$ из (2.5.20) и (2.5.22), получим (2.5.19). Теорема доказана для случая дискретных ансамблей.

Непрерывный случай отличается от рассмотренного только тем, что иначе задается распределение вероятностей на тройках из XYZ . Оно задается посредством функции

$$f(x)f(y|xz_i)p(z_i), \quad (2.5.23)$$

которая, как и в теореме 2.5.1, задает распределение смешанного типа: ансамбли X и Y непрерывны, а ансамбль Z дискретен. В остальном доказательство полностью сохраняется.

Пример 2.5.1. Пусть $X = \{0, 1\}$ — множество сообщений на входе канала, $Y = \{0, 1\}$ — множество сообщений на выходе канала и переходы входных сообщений в выходные задаются с помощью графа переходов, изображенного на рис. 2.5.2. Вероятность $p(y|x)$ перехода сообщения $x \in X$ в сообщение $y \in Y$ написана над соответствующим ребром графа. В этом примере будет рассматриваться так называемый *двоичный симметричный канал*, в котором вероятности неправильных переходов (ошибок) одинаковы и равны p . Вероятности входных сообщений обозначены через α и $1 - \alpha$, а выходных сообщений — через β и $1 - \beta$. Очевидно,

$$\beta = \alpha(1 - p) + (1 - \alpha)p, \quad (2.5.24)$$

поэтому среднюю взаимную информацию $I(X; Y)$ можно рассматривать как функцию двух параметров α и p и записывать ее как $I(\alpha, p)$.

Предположим вначале, что p фиксировано и $I(\alpha, p)$ рассматривается только как функция от α . Этот случай соответствует фиксации условных распределений. Вычислим среднюю взаимную информацию и убедимся на этом примере, что функция $I(\alpha, p)$ — выпуклая вверх по α .

Имеем

$$I(\alpha, p) = H(Y) - H(Y|X), \quad (2.5.25)$$

где

$$H(Y) = h(\beta) \triangleq -\beta \log \beta - (1 - \beta) \log(1 - \beta), \quad (2.5.26)$$

$$H(Y|X) = \sum_i p(x_i) H(Y|x_i) = H(Y|x_1) = -p \log p - (1 - p) \log(1 - p) \triangleq h(p). \quad (2.5.27)$$

Второе равенство в (2.5.27) — следствие того, что $H(Y|x_i)$ не зависит от x_i , так как набор условных вероятностей $p(y|x_i)$ совпадает с набором условных вероятностей $p(y|x_2)$, $y \in Y$, $x_1 = 0$, $x_2 = 1$. Таким образом, при фиксированном p

$$I(\alpha, p) = h(\beta) - h(p),$$

где β определяется через α с помощью соотношения (2.5.24). Очевидно, $I(0, p) = I(1, p) = 0$, так как $h(\beta) = h(p)$ при $\alpha = 0$ или $\alpha = 1$. $h(\beta)$ принимает максимальное значение при $\beta = 1/2$. Так как $\beta = 1/2$ только в том случае, когда $\alpha = 1/2$, то $I(\alpha, p)$ принимает максимальное значение, равное $1 - h(p)$ в точке $\alpha = 1/2$. Поскольку $h(\beta)$ — выпуклая вверх функция β (см. задачу 2.5.5), а β — линейная функция α , то $I(\alpha, p)$ — выпуклая вверх функция (см. рис. 2.5.3, а).

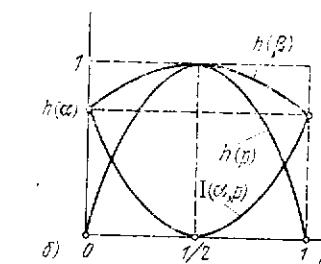
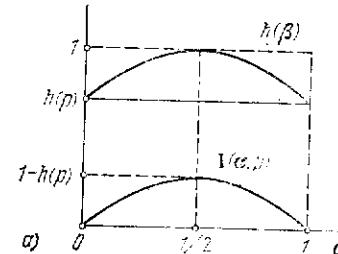


Рис. 2.5.3. Иллюстрация выпуклости средней взаимной информации вверх по распределениям на входе (а) и вниз по условным распределениям, задающим канал (б).

Пусть теперь зафиксировано число α , а p изменяется. По-прежнему будем пользоваться соотношением (2.5.25), оценивая по отдельности каждое слагаемое. Первое слагаемое есть выпуклая вверх функция, принимающая максимальное значение, равное 1, при $p = 1/2$, так как при таком значении p вероятность β (см. (2.5.24)) равна $1/2$. Очевидно, $h(\beta)|_{p=0} = h(\beta)|_{p=1} = h(\alpha)$, функция $h(p)$ равна нулю при $p = 0$ и $p = 1$, выпукла вверх и принимает максимальное значение, равное 1, при $p = 1/2$. На рис. 2.5.3, б показаны три функции: $h(\beta)$, $h(p)$ и $I(\alpha, p) = h(\beta) - h(p)$. Видно, что средняя взаимная информация является выпуклой вниз функцией от p .

§ 2.6 *. Случайные процессы непрерывного времени

Случайные процессы дискретного времени были определены в первой главе. В этом параграфе мы кратко остановимся на случайных процессах непрерывного времени — на способах их задания, интегрирования и представления в виде рядов. Целью настоящего параграфа является напоминание важных с нашей точки зрения фактов из теории случайных процессов. Многие вопросы, связанные с обоснованиями сходимости, перемены порядка интегрирования, существования пределов и др., остаются вне изложения. Более подробно с этими вопросами можно познакомиться по книгам [6, 12].

Пусть $[T_1, T_2]$ — интервал на числовой прямой. *Случайным процессом* называется система случайных величин $\{X_t\}$, индексованная параметром t , $t \in [T_1, T_2]$. В дальнейшем X_t при каждом t является действительной случайной величиной и параметр t понимается как время. Если t пробегает все множество значений интервала $[T_1, T_2]$, то случайный процесс называется *процессом непрерывного времени* и обозначается через $X(t)$. Будем также предполагать, что случайные величины, представляющие значения случайного процесса, задаются ф. п. в. Для с. в. X_t ф. п. в. обозначается через $f(x; t)$. Для системы с. в. X_{t_1}, \dots, X_{t_n} , $t_1, \dots, t_n \in [T_1, T_2]$, совместная ф. п. в. зависит от выбранных моментов времени t_1, \dots, t_n и обозначается через $f(x_1, \dots, x_n; t_1, \dots, t_n)$.

Случайный процесс $X(t)$ считается заданным на интервале $[T_1, T_2]$, если для любого $n = 1, 2, \dots$ любых $t_1, \dots, t_n \in [T_1, T_2]$ задана совместная ф. п. в. системы с. в. X_{t_1}, \dots, X_{t_n} .

Определение 2.6.1. Случайный процесс непрерывного времени $X(t)$, заданный на интервале $[-\infty, \infty]$, называется *стационарным*, если для любого n , $n = 1, 2, \dots$, любых t_1, \dots, t_n и x_1, \dots, x_n , а также для любого τ

$$f(x_1, \dots, x_n; t_1, \dots, t_n) = f(x_1, \dots, x_n; t_1 + \tau, \dots, t_n + \tau). \quad (2.6.1)$$

Другими словами, процесс $X(t)$ стационарен, если его любые n -мерные ф. п. в. не зависят от сдвига по оси времени.

Основными числовыми характеристиками случайного процесса являются его *математическое ожидание* $m(t)$, определяемое соотношением

$$m(t) \triangleq M X_t, \quad t \in [T_1, T_2], \quad (2.6.2)$$

и *корреляционная функция* $K(t_1, t_2)$ — неслучайная функция двух аргументов, определяемая соотношением

$$K(t_1, t_2) \triangleq M[X_{t_1} - m(t_1)][X_{t_2} - m(t_2)]. \quad (2.6.3)$$

Функция $\sigma^2(t) \triangleq K(t, t)$ называется *дисперсией* случайного процесса $X(t)$.

Очевидно, что для стационарного случайного процесса $m(t) = \text{const}$, $\sigma^2(t) = \text{const}$, а корреляционная функция $K(t_1, t_2)$ зависит только от разности аргументов $\tau = t_1 - t_2$: $K(t_1, t_2) = K(\tau)$.

Рассмотрим основные свойства функции корреляции случайного процесса. Имеют место следующие свойства:

$$1) \quad K(t_1, t_2) = K(t_2, t_1); \quad (2.6.4)$$

$$2) \quad K(t, t) \geq 0, \quad (2.6.5)$$

$$3) \quad |K(t_1, t_2)| \leq \sigma(t_1) \sigma(t_2); \quad (2.6.6)$$

3) для любой функции $\varphi(t)$, заданной на интервале $[T_1, T_2]$,

$$\int_{T_1}^{T_2} \int_{T_1}^{T_2} K(t_1, t_2) \varphi(t_1) \varphi(t_2) dt_1 dt_2 \geq 0, \quad (2.6.7)$$

если интеграл существует. Последнее свойство называется неотрицательной определенностью функции корреляции. Если (2.6.7) выполняется для каждой ненулевой функции со строгим неравенством, то говорят, что функция корреляции $K(t_1, t_2)$ положительно определена.

Для стационарных случайных процессов из первого свойства следует, что $K(\tau)$ — четная функция, а из второго, что $|K(\tau)| \leq \sigma^2$.

Преобразование Фурье корреляционной функции $K(\tau)$ стационарного случайного процесса $X(t)$, если оно существует, будем обозначать через $N(f)$:

$$N(f) \triangleq \int_{-\infty}^{\infty} K(\tau) \exp(-j2\pi f\tau) d\tau, \quad (2.6.8)$$

где здесь и далее $j = \sqrt{-1}$. Функцию $N(f)$ называют *спектральной плотностью мощности* процесса $X(t)$. Она является четной неотрицательной функцией *). Зная $N(f)$, можно найти корреляционную функцию с помощью обратного преобразования Фурье:

$$K(\tau) = \int_{-\infty}^{\infty} N(f) \exp(j2\pi f\tau) df. \quad (2.6.9)$$

Полагая в последнем соотношении $\tau = 0$, получим

$$\sigma^2 = K(0) = \int_{-\infty}^{\infty} N(f) df. \quad (2.6.10)$$

Дисперсия случайного процесса характеризует его среднюю мощность, и соотношение (2.6.10) показывает, что функция $N(f)$ описывает распределение мощности по частотам.

Ниже мы встретимся с интегралами от случайного процесса. Прежде чем ввести понятие интеграла, определим среднеквадратическую сходимость последовательности случайных величин.

Определение 2.6.2. Последовательность случайных величин X_1, X_2, \dots сходится в *среднеквадратическом* (с. к. сходится) к случайной величине X , если

$$\lim_{n \rightarrow \infty} M(X - X_n)^2 = 0. \quad (2.6.11)$$

*) Четность следует из четности, а неотрицательность — из неотрицательной определенности функции $K(\tau)$.

Можно показать, что условие (2.6.11) эквивалентно следующему:

$$\lim_{m, n \rightarrow \infty} M(X_m - X_n)^2 = 0, \quad (2.6.12)$$

где m и n стремятся к бесконечности независимо друг от друга.

Имеет место следующая лемма, в которой указывается условие существования с. к. предела последовательности с. в., имеющих ограниченную дисперсию.

Л е м м а 2.6.1. Для того чтобы существовал с. к. предел последовательности с. в. X_n , $n = 1, 2, \dots$, при $MX_n^2 < \infty$, необходимо и достаточно, чтобы существовал предел $\lim_{m, n \rightarrow \infty} MX_m X_n \triangleq K_0$, где m и n стремятся к бесконечности независимо друг от друга. Если это условие выполнено, то $MX^2 = K_0$ для предельной с. в. X .

Введем теперь понятие интеграла от случайного процесса. Пусть $T_1 = t_{n_0}$, $t_{n_1}, \dots, t_{n_k} = T_2$ — некоторое разбиение интервала $[T_1, T_2]$ и $\varphi(t)$ — произвольная функция, заданная на том же интервале.

Определение 2.6.3. Пусть случайный процесс $X(t)$ задан на интервале $[T_1, T_2]$. Если последовательность с. в.

$$\xi_n \triangleq \sum_{k=0}^n X_{t_{nk}} \varphi(t_{nk}) \Delta t_{nk},$$

где $\Delta t_{nk} = t_{n,k+1} - t_{nk}$, с. к. сходится независимо от выбранной последовательности разбиений к с. в. ξ при $n \rightarrow \infty$ и $\max_k \Delta t_{nk} \rightarrow 0$, то ξ называется *интегралом в среднеквадратическом смысле* или просто *интегралом от случайного процесса $X(t)$ с весом $\varphi(t)$* и обозначается символом $\int_{T_1}^{T_2} X(t) \varphi(t) dt$.

С помощью леммы 2.6.1 доказывается следующее утверждение.

Теорема 2.6.1. Пусть случайный процесс $X(t)$ имеет нулевое среднее, $m(t) = 0$, и ограниченную функцию корреляции $K(t_1, t_2)$. Необходимым и достаточным условием существования интеграла от случайного процесса с весом $\varphi(t)$ является существование и ограниченность обычного интеграла

$$\int_{T_1}^{T_2} \int_{T_1}^{T_2} K(t_1, t_2) \varphi(t_1) \varphi(t_2) dt_1 dt_2.$$

Если это условие выполнено, то для двух с. в.

$$\xi_1 \triangleq \int_{T_1}^{T_2} X(t) \varphi_1(t) dt, \quad \xi_2 \triangleq \int_{T_1}^{T_2} X(t) \varphi_2(t) dt$$

выполняются следующие соотношения: $M\xi_1 = M\xi_2 = 0$ и

$$M\xi_1 \xi_2 = \int_{T_1}^{T_2} \int_{T_1}^{T_2} K(t_1, t_2) \varphi_1(t_1) \varphi_2(t_2) dt_1 dt_2.$$

Обсуждение леммы 2.6.1 и теоремы 2.6.1 см. в задачах 2.6.2—2.6.4.

Далее мы будем рассматривать представление случайных процессов с помощью ортогональных рядов. Будем предполагать до конца этого раздела, что рассматриваемые функции определены на интервале $[0, T]$ и для каждой неслучайной функции интеграл от ее квадрата конечен. Другими словами, мы рассматриваем пространство функций $L_2[0, T]$. Две функции $\varphi_1(t)$ и $\varphi_2(t)$ называются *ортогональными*, если $\int_0^T \varphi_1(t) \varphi_2(t) dt = 0$. Функция $\varphi(t)$ называется *нормированной*, если $\int_0^T \varphi^2(t) dt = 1$. Система ортогональных нормированных функций называется *системой ортонормированных функций*. Известно, что в L_2 существует так называемая полная система ортонормированных функций, т. е. такая система, что в L_2 не существует ненулевой функции, ортогональной всем функциям из этой системы. Если $\{\varphi_i(t)\}$ — полная ортонормированная система, то любая функция $x(t)$ из L_2 представима в виде $x(t) = \sum x_i \varphi_i(t)$, где $x_i = \int x(t) \varphi_i(t) dt$, либо в виде предела последовательности таких функций.

Теорема 2.6.2. Пусть функция корреляции $K(t_1, t_2)$ случайного процесса $X(t)$, $t \in [0, T]$, непрерывна и удовлетворяет условию

$$\int_0^T K(t, t) dt < \infty. \quad (2.6.13)$$

Пусть $\{\varphi_i(t)\}$ — полная в $L_2[0, T]$ система ортонормированных функций и X_1, X_2, \dots — случайные величины, определяемые соотношениями

$$X_i \triangleq \int_0^T X(t) \varphi_i(t) dt, \quad i = 1, 2, \dots \quad (2.6.14)$$

Тогда

$$X(t) = \sum_{i=1}^{\infty} X_i \varphi_i(t), \quad (2.6.15)$$

где равенство понимается в среднеквадратическом смысле, т. е.

$$\lim_{n \rightarrow \infty} M \left[X(t) - \sum_{i=1}^n X_i \varphi_i(t) \right]^2 = 0 \quad (2.6.16)$$

для любого t из интервала $[0, T]$.

В теореме 2.6.2 утверждается, что при выполнении определенных условий случайный процесс непрерывного времени представим в виде ряда (2.6.15). Благодаря такому представлению возможно установить соответствие между случайнм процессом $X(t)$ и бесконечной последовательностью с. в. X_1, X_2, \dots — коэффициентами ряда.

Особый интерес представляет разложение Карунена—Лоэва. Прежде чем формулировать теорему об этом разложении, рассмотрим интегральное уравнение

$$\int_0^T K(t_1, t_2) \psi_i(t_1) dt_1 = \lambda_i \psi_i(t_2), \quad (2.6.17)$$

где $\psi_i(t)$, $t \in [0, T]$ — неизвестная функция.

Из теории интегральных уравнений известно, что уравнение (2.6.17) имеет ненулевые решения не более чем для счетного множества чисел $\{\lambda_i\}$. Числа λ_i , при которых уравнение (2.6.17) имеет решения, называются собственными числами, а сами решения — собственными функциями уравнения (2.6.17) или собственными функциями корреляционного ядра $K(t_1, t_2)$. Известно также, что из симметричности и неотрицательной определенности корреляционной функции $K(t_1, t_2)$ следует, что все собственные числа $\lambda_1, \lambda_2, \dots$ действительны и неотрицательны, а нормированные собственные функции $\psi_1(t)$, $\psi_2(t)$, ... образуют ортонормированную систему функций. Эта система функций полна в L_2 в том и только в том случае, когда все собственные числа положительны, т. е. когда уравнение (2.6.17) не имеет ненулевых решений при $\lambda = 0$.

Теорема 2.6.3. (Разложение Карунена—Лоэва.) Пусть случайный процесс $X(t)$, заданный на интервале $[0, T]$, имеет нулевое среднее, $m(t) = 0$, и непрерывную функцию корреляции $K(t_1, t_2)$, которая удовлетворяет условию (2.6.13). Пусть $\{\psi_i(t)\}$, $i = 1, 2, \dots$, $t \in [0, T]$ — система нормированных собственных функций корреляционного ядра $K(t_1, t_2)$, а λ_i , $i = 1, 2, \dots$ — соответствующие собственные числа. Тогда имеет место равенство в среднеквадратическом смысле

$$X(t) = \sum_{i=1}^{\infty} X_i \psi_i(t), \quad (2.6.18)$$

где

$$X_i = \int_0^T X(t) \psi_i(t) dt, \quad i = 1, 2, \dots \quad (2.6.19)$$

При этом

$$MX_i X_j = \begin{cases} \lambda_i, & \text{если } i = j, \\ 0 & \text{в противном случае.} \end{cases} \quad (2.6.20)$$

Другими словами, коэффициенты ряда Карунена—Лоэва, имеющие различные индексы, некоррелированы, а дисперсия i -го коэффициента равна i -му собственному числу корреляционного ядра.

Заметим, что в случае разложения случайного процесса $X(t)$, имеющего нулевое среднее, по произвольной системе ортогональных функций $\{\phi_i(t)\}$, корреляционный момент с. в. X_i и X_j определяется из соотношения

$$\begin{aligned} MX_i X_j &= M \int_0^T \int_0^T X(t_1) X(t_2) \phi_i(t_1) \phi_j(t_2) dt_1 dt_2 = \\ &= \int_0^T \int_0^T K(t_1, t_2) \phi_i(t_1) \phi_j(t_2) dt_1 dt_2. \end{aligned} \quad (2.6.21)$$

Ввиду некоррелированности коэффициентов разложение Карунена—Лоэва представляет особый интерес для гауссовских случайных процессов. Гауссовский случайный процесс может быть определен одним из следующих двух эквивалентных способов.

Определение 2.6.4 (а). Случайный процесс $X(t)$, заданный на интервале $[0, T]$, называется *гауссовским*, если для любого n , $n = 1, 2, \dots$, и любых $t_1, t_2, \dots, t_n \in [0, T]$, случайные величины X_{t_1}, \dots, X_{t_n} имеют гауссовскую совместную ф. п. в.

Определение 2.6.4 (б). Случайный процесс $X(t)$, заданный на интервале $[0, T]$, называется *гауссовским*, если для любого $n = 1, 2, \dots$ и любой системы функций $\{\phi_i(t)\}$, $i = 1, \dots, n$, случайные величины X_1, X_2, \dots, X_n , определяемые соотношениями (2.6.14), имеют гауссовскую совместную ф. п. в.

Так как для гауссовских случайных величин из некоррелированности следует их независимость, то в случае разложения гауссовского случайного процесса в ряд Карунена—Лоэва совместная ф. п. в. коэффициентов X_1, \dots, X_n разложения задается особенно просто, а именно

$$f(x_1, \dots, x_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\lambda_i}} \exp\left(-\frac{x_i^2}{2\lambda_i}\right), \quad n = 1, 2, \dots \quad (2.6.22)$$

В заключение этого параграфа определим процесс, называемый белым шумом. В случае дискретного времени белым шумом называют случайный процесс, значения которого в любые различные моменты времени являются независимыми случайными величинами. Аналогичным образом определяется белый шум и в случае непрерывного времени. Однако в последнем случае процесс белого шума оказывается столь «странным», что его строгое определение оказывается невозможным, если только не вводить обобщенные функции. Поэтому белый шум часто называют обобщенным случайнм процессом.

Пусть $\{\varphi_i(t)\}$ — полная в $L_2[0, T]$ система ортонормальных функций и $\{X_i\}$ — система независимых гауссовых с. в. с нулевым средним и дисперсией $M X_i^2 = \frac{N_0}{2}$, одинаковой для всех индексов i . Зафиксируем n и рассмотрим случайный процесс $X_n(t)$, представляющий собой усеченный ряд:

$$X_n(t) \triangleq \sum_{i=1}^n X_i \varphi_i(t). \quad (2.6.23)$$

Если взять систему ортонормальных гармонических функций

$$\varphi_{2i-1}(t) = \sqrt{\frac{T}{2}} \sin 2\pi \frac{i}{T} t, \quad \varphi_{2i}(t) = \sqrt{\frac{T}{2}} \cos 2\pi \frac{i}{T} t,$$

то X_i будет случайной амплитудой i -й гармонической (частотной) составляющей, а $M X_i^2$ — мощностью этой составляющей процесса $X_n(t)$. Последовательность чисел $M X_1^2, M X_2^2, \dots$ естественно называть распределением мощности по спектру частот. Для процесса $X_n(t)$ распределение мощности таково: для всех частот $2\pi i \frac{1}{T} < 2\pi \frac{n}{T}$ мощность одинакова и равна $\frac{N_0}{2}$, а для всех остальных частот мощность равна нулю.

Пусть теперь n возрастает к бесконечности. Случайный процесс $X(t)$, который получается как предельный для последовательности случайных процессов $X_n(t)$, $n = 1, 2, \dots$, называется *белым гауссовским шумом*:

$$X(t) = \sum_{i=1}^{\infty} X_i \varphi_i(t). \quad (2.6.24)$$

В этом процессе присутствуют все частотные составляющие; их мощности одинаковы и равны $N_0/2$. При этом число $N_0/2$ называется *интенсивностью* белого шума. Случайный процесс $X_n(t)$ для каждого фиксированного n называется *частотно-ограниченным белым гауссовским шумом*.

Пусть $\psi_1(t)$ и $\psi_2(t)$ — две произвольные ортонормальные функции из $L_2[0, T]$, т. е. $\int_0^T \psi_i(t) \psi_j(t) dt = \delta_{ij}^*$. Тогда

$$\psi_1(t) = \sum_{i=1}^{\infty} \mu_i \varphi_i(t), \quad \psi_2(t) = \sum_{i=1}^{\infty} v_i \varphi_i(t)$$

$$\text{и } \sum_{i=1}^{\infty} \mu_i v_i = \int_0^T \psi_1(t) \psi_2(t) dt = 0, \text{ где}$$

$$\mu_i = \int_0^T \psi_1(t) \varphi_i(t) dt, \quad v_i = \int_0^T \psi_2(t) \varphi_i(t) dt.$$

^{*}) Величина $\delta_{ij} \triangleq \begin{cases} 1, & \text{если } i = j, \\ 0, & \text{если } i \neq j, \end{cases}$ называется δ -Кронекера.

Введем в рассмотрение две с. в.

$$\xi_1 = \int_0^T \psi_1(t) X(t) dt, \quad \xi_2 = \int_0^T \psi_2(t) X(t) dt. \quad (2.6.25)$$

Из независимости с. в. $\{X_i\}$ теперь вытекает, что

$$M \xi_1 \xi_2 = M \sum_{i, j} X_i X_j \int_0^T \psi_1(t) \varphi_i(t) dt \int_0^T \psi_2(t) \varphi_j(t) dt = \frac{N_0}{2} \sum_i \mu_i v_i = 0,$$

а из гауссности этих с. в. — что ξ_1, ξ_2 — пара независимых гауссовых с. в., математическое ожидание которых равно нулю, а дисперсии равны $N_0/2$.

Основное свойство белого гауссовского шума заключается в том, что пара с. в., определяемая соотношениями (2.6.25) для любых двух ортонормальных функций $\psi_1(t)$ и $\psi_2(t)$, является парой независимых гауссовых с. в. Это свойство часто принимают за другое определение белого гауссовского шума.

Функция корреляции белого шума не существует как обычная функция. Функция корреляции для частотно-ограниченного белого шума

$$K_n(t_1, t_2) = \frac{N_0}{2} \sum_{i=1}^n \varphi_i(t_1) \varphi_i(t_2).$$

Можно показать, что в случае гармонических функций

$$K_n(t_1, t_2) \xrightarrow{n \rightarrow \infty} \frac{N_0}{2} \delta(t_1 - t_2).$$

Поэтому иногда считают, что белый шум — это процесс, функция корреляции которого определяется соотношением

$$K(t_1, t_2) = \frac{N_0}{2} \delta(t_1 - t_2).$$

Процесс белого шума имеет бесконечную мощность в каждый момент времени и поэтому не может быть задан как совокупность с. в. $X(t)$, $t \in [0, T]$, каждая из которых соответствует одному из значений параметра t . Очевидно, что белый шум не соответствует никакому физическому процессу. Однако модель белого шума часто используют в практических расчетах. Результаты расчетов хорошо согласуются с реальными характеристиками систем в тех случаях, когда физический процесс имеет постоянную спектральную плотность мощности в достаточно широкой полосе частот.

§ 2.7*. Средняя взаимная информация между случайными процессами

Пусть $X(t)$ и $Y(t)$ — два случайных процесса, заданных на интервале $[0, T]$, каждый из которых допускает ортогональное разложение (2.6.15). Рассмотрим две произвольные полные в $L_2[0, T]$ ортонормированные системы функций $\{\varphi_i(t)\}$ и $\{\psi_i(t)\}$. Каждый из процессов $X(t)$ и $Y(t)$ может быть представлен рядом:

$$X(t) = \sum_{i=1}^{\infty} X_i \varphi_i(t), \quad (2.7.1)$$

$$Y(t) = \sum_{i=1}^{\infty} Y_i \psi_i(t), \quad (2.7.2)$$

где

$$X_i = \int_0^T X(t) \varphi_i(t) dt, \quad i = 1, 2, \dots, \quad (2.7.3)$$

$$Y_i = \int_0^T Y(t) \psi_i(t) dt, \quad i = 1, 2, \dots \quad (2.7.4)$$

Таким образом, случайные процессы $X(t)$ и $Y(t)$ в среднеквадратическом смысле полностью определяются бесконечными последовательностями коэффициентов разложений X_1, X_2, \dots и Y_1, Y_2, \dots соответственно.

Будем говорить, что случайные процессы $X(t)$ и $Y(t)$ заданы совместно на интервале $[0, T]$, если для всех $n = 1, 2, \dots$ заданы совместные функции плотности вероятностей $f(x_1, \dots, x_n, y_1, \dots, y_n) \triangleq f(\mathbf{x}, \mathbf{y})$ с. в. $X_1, \dots, X_n, Y_1, \dots, Y_n$.

Пусть $\mathbf{X} = (X_1, \dots, X_n)$ и $\mathbf{Y} = (Y_1, \dots, Y_n)$ — случайные векторы, образованные первыми n коэффициентами разложений процессов $X(t)$ и $Y(t)$ в ряды (2.7.1) и (2.7.2) соответственно. Можно вычислить среднюю взаимную информацию

$$I(\mathbf{X}; \mathbf{Y}) = \int f(\mathbf{x}, \mathbf{y}) \log \frac{f(\mathbf{x}, \mathbf{y})}{f(\mathbf{x})f(\mathbf{y})} d\mathbf{x} d\mathbf{y}, \quad (2.7.5)$$

где

$$f(\mathbf{x}) = \int f(\mathbf{x}, \mathbf{y}) d\mathbf{y}, \quad (2.7.6)$$

$$f(\mathbf{y}) = \int f(\mathbf{x}, \mathbf{y}) d\mathbf{x}. \quad (2.7.7)$$

Определение 2.7.1. Средняя взаимная информация между двумя случайными процессами $X(t)$ и $Y(t)$, заданными совместно на интервале $[0, T]$, определяется соотношением

$$I_T(X(t); Y(t)) \triangleq \lim_{n \rightarrow \infty} I(\mathbf{X}; \mathbf{Y}), \quad (2.7.8)$$

если предел существует.

Замечание. Можно показать, что определение 2.7.1 корректно в том смысле, что величина $I_T(X(t); Y(t))$ не зависит от выбора полных в L_2 систем функций $\{\varphi_i(t)\}$ и $\{\psi_i(t)\}$. Доказательство этого факта заинтересованный читатель может найти в работе [5].

Рассмотрим некоторые примеры.

Пример 2.7.1. Пусть $Y(t) = X(t) + Z(t)$ и случайные процессы $X(t)$ и $Z(t)$ статистически независимы. Статистическая независимость процессов $X(t)$ и $Z(t)$ означает, что для любого $n = 1, 2, \dots$ векторы \mathbf{X} и \mathbf{Z} , образованные первыми n коэффициентами разложений ряды, статистически независимы.

Для вычисления средней взаимной информации $I_T(X(t); Y(t))$ мы будем использовать разложения процессов $X(t)$ и $Y(t)$ по одной и той же системе функций $\{\varphi_i(t)\}$. Тогда

$$Y_i = \int_0^T Y(t) \varphi_i(t) dt = \int_0^T X(t) \varphi_i(t) dt + \int_0^T Z(t) \varphi_i(t) dt = X_i + Z_i, \quad i = 1, 2, \dots \quad (2.7.9)$$

Из независимости процессов $X(t)$ и $Z(t)$ следует, что $f(y|\mathbf{x}) = f(z)$, где $f(z)$ — ф. п. в. случайного вектора $\mathbf{Z} = (Z_1, \dots, Z_n)$ (см. также пример 2.3.3).

Имеем

$$I(\mathbf{X}; \mathbf{Y}) = H_0(\mathbf{Y}) - H_0(\mathbf{Y} | \mathbf{X}) =$$

$$= H_0(\mathbf{Y}) - H_0(\mathbf{Z}) = H_0(\mathbf{Y}) + \int f(z) \log f(z) dz, \quad (2.7.10)$$

где $H_0(\mathbf{Y})$ и $H_0(\mathbf{Z})$ — относительные энтропии векторов \mathbf{Y} и \mathbf{Z} соответственно. Тогда

$$I_T(X(t); Y(t)) = \lim_{n \rightarrow \infty} [H_0(\mathbf{Y}) - H_0(\mathbf{Z})]. \quad (2.7.11)$$

Пример 2.7.2. Продолжим рассмотрение предыдущего примера. Пусть $Y(t) = X(t) + Z(t)$, причем $X(t)$ и $Z(t)$ — теперь статистически независимые гауссовские случайные процессы. Пусть $\{\varphi_i(t)\}$ — система собственных функций для корреляционного ядра $K_Z(t_1, t_2)$ процесса $Z(t)$. Предположим также, что система функций $\{\varphi_i(t)\}$ полна в L_2 . Для вычисления $I_T(X(t); Y(t))$ воспользуемся разложением по системе функций $\{\varphi_i(t)\}$. Так как в этом случае с. в. Z_i , $i = 1, 2, \dots$, — коэффициенты разложения процесса $Z(t)$ — являются статистически независимыми гауссовскими случайными величинами с дисперсиями λ_i , $i = 1, 2, \dots$, где $\{\lambda_i\}$ — собственные числа ядра $K_Z(t_1, t_2)$, то

$$H_0(\mathbf{Z}) = \sum_{i=1}^n H_0(Z_i) = \frac{1}{2} \sum_{i=1}^n \log 2\pi e \lambda_i \quad (2.7.12)$$

и

$$I(X(t); Y(t)) = \lim_{n \rightarrow \infty} H_0(\mathbf{Y}) - \frac{1}{2} \sum_{i=1}^n \log 2\pi e \lambda_i. \quad (2.7.13)$$

Предположим теперь, что $\{\varphi_i(t)\}$ является системой собственных функций также и для корреляционного ядра $K_X(t_1, t_2)$ процесса $X(t)$. Тогда коэффициенты разложения X_i , $i = 1, 2, \dots$, процесса $X(t)$ являются гауссовскими независимыми случайными величинами с дисперсиями μ_i , $i = 1, 2, \dots$, где $\{\mu_i\}$ — собственные числа ядра $K_X(t_1, t_2)$. Очевидно, что в этом случае коэффициенты

разложения Y_i , $i = 1, 2, \dots$, процесса $Y(t)$ — независимые гауссовские величины с дисперсиями $\mu_i + \lambda_i$, $i = 1, 2, \dots$. Поэтому

$$H_0(Y) = \sum_{i=1}^n H_0(Y_i) = \frac{1}{2} \sum_{i=1}^n \log 2\pi e (\mu_i + \lambda_i) \quad (2.7.14)$$

и

$$I_T(X(t); Y(t)) = \frac{1}{2} \sum_{i=1}^{\infty} \log \left(1 + \frac{\mu_i}{\lambda_i} \right). \quad (2.7.15)$$

§ 2.8*. Поиск экстремумов

В заключение этой главы мы рассмотрим задачу поиска экстремумов функций нескольких переменных при наличии ограничений. Эта задача непосредственно не относится к исследованию свойств количества информации, но весьма часто встречается в теории информации в приложении к максимизации или минимизации средней взаимной информации, а также других теоретико-информационных функций. Ниже будут сформулированы необходимые условия экстремума, так называемые условия Куна—Таккера, и показано, что эти условия являются достаточными для максимизации или минимизации выпуклых функций, заданных на выпуклых множествах.

Всюду ниже будет предполагаться, что рассматриваемые функции имеют непрерывные частные производные в области определений. Особый интерес для теории информации представляет случай, когда рассматриваемые функции определены на множестве вероятностных векторов (x_1, \dots, x_m) и, следовательно, ограничения при поиске экстремума выделяют только те векторы, для которых

$$x_i \geq 0, i = 1, \dots, m, \text{ и } \sum_{i=1}^m x_i = 1.$$

2.8.1. Метод неопределенных множителей Лагранжа. Предположим, что требуется найти экстремум функции $f(\mathbf{x})$ при условии, что $g_j(\mathbf{x}) = 0$, $j = 1, \dots, s$, где все функции определены на множестве m -мерных действительных векторов $\mathbf{x} = (x_1, \dots, x_m)$. Стандартным методом решения такой задачи, приводящим задачу поиска условного экстремума к задаче поиска безусловного, является метод неопределенных множителей Лагранжа.

Идею этого метода поясним на примере функций двух переменных: пусть требуется найти экстремум $f(x, y)$ при условии, что в экстремальной точке (x^*, y^*) выполняется равенство $g(x^*, y^*) = 0$. Предположим, что в точке (x^*, y^*) частные производные $g'_x \triangleq dg/dx$, $g'_y \triangleq dg/dy$ не равны нулю одновременно. Если последнее условие выполнено, например, $g'_y(x^*, y^*) \neq 0$, то в некоторой окрестности точки (x^*, y^*) однозначно определена дифференцируемая по x функция $y = y(x)$ и, следовательно, задача заключается

в поиске безусловного экстремума функции $f(x, y(x))$. Необходимые условия экстремума можно записать следующим образом:

$$\begin{cases} \frac{\partial f}{\partial x} = f'_x + f'_y \frac{dy}{dx} = 0, \\ \frac{\partial g}{\partial x} = g'_x + g'_y \frac{dy}{dx} = 0, \end{cases} \quad (2.8.1)$$

где второе условие есть тождественное равенство. Обозначая $\lambda \triangleq -f'_y/g'_y$ (заметим, что число λ называется *неопределенным множителем Лагранжа*), условия (2.8.1) можно привести к виду

$$f'_x + \lambda g'_x = 0, \quad f'_y + \lambda g'_y = 0.$$

Таким образом, для поиска условного экстремума необходимо найти решения относительно переменных x, y и λ следующих трех уравнений:

$$\begin{cases} \frac{\partial}{\partial x} [f(x, y) + \lambda g(x, y)] = 0, \\ \frac{\partial}{\partial y} [f(x, y) + \lambda g(x, y)] = 0, \\ \frac{\partial}{\partial \lambda} [f(x, y) + \lambda g(x, y)] = g(x, y) = 0. \end{cases} \quad (2.8.2)$$

Функция $f(x, y) + \lambda g(x, y)$ называется *функцией Лагранжа*. Всякое решение системы уравнений (2.8.2) является стационарной точкой функции Лагранжа (точкой максимума, минимума или седловой точкой).

В общем случае для функции $f(\mathbf{x})$ от m переменных при наличии $s < m$ ограничивающих условий $g_j(\mathbf{x}) = 0$, $j = 1, \dots, s$, метод множителей Лагранжа дает следующие необходимые условия экстремума

$$\begin{cases} \frac{\partial}{\partial x_i} \left[f(\mathbf{x}) + \sum_{j=1}^s \lambda_j g_j(\mathbf{x}) \right] = 0, \quad i = 1, \dots, m, \\ \frac{\partial}{\partial \lambda_j} \left[f(\mathbf{x}) + \sum_{i=1}^s \lambda_i g_i(\mathbf{x}) \right] = g_j(\mathbf{x}) = 0, \quad j = 1, \dots, s, \end{cases} \quad (2.8.3)$$

из которых определяются m значений переменных (x_1, \dots, x_m) и s неопределенных множителей Лагранжа $(\lambda_1, \dots, \lambda_s)$. Заметим, что система уравнений (2.8.2) имеет решение только в том случае, когда в стационарной точке обе частные производные g'_x и g'_y не равны нулю одновременно. Аналогичное условие для системы (2.8.3) выглядит следующим образом: в стационарной точке одновременно не равны нулю все определители

$$\left| \frac{\partial g_j}{\partial x_i} \right|, \quad j = 1, \dots, s, \quad i \in [1, m], \quad (2.8.4)$$

порядка s , образованные частными производными функций $g_j(\mathbf{x})$ по s из m переменных (x_1, \dots, x_m) .

Метод неопределенных множителей Лагранжа дает один из возможных путей поиска экстремумов функций нескольких переменных при наличии ограничений. Однако условия (2.8.4), при которых этот метод применим, могут не выполняться. С другой стороны, часто встречаются задачи поиска экстремума при ограничениях, представляющих собой неравенства, а не равенства, как в разобранном выше случае. Поэтому желательно иметь необходимые условия экстремума, не зависящие от типа ограничений и от свойств функций, задающих ограничения. Такого sorta задачи составляют предмет теории нелинейного программирования (см., например, [8]). Одним из центральных результатов нелинейного программирования, который широко используется в теории информации, является теорема Куна—Таккера.

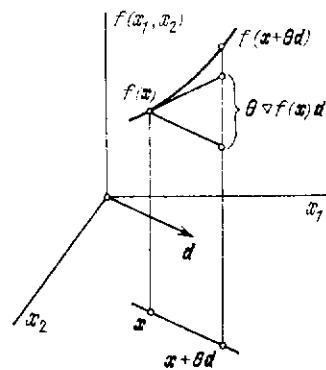


Рис. 2.8.1. Градиент и приращение функции $f(\mathbf{x})$ по направлению \mathbf{d} .

Каждый ненулевой вектор $\mathbf{d} \in E_m$ можно рассматривать как некоторое направление. Пусть \mathbf{x} — некоторая фиксированная точка из E_m , а θ — скалярная величина, изменяющаяся от 0 до бесконечности. Тогда каждая из точек вида $\mathbf{x} + \theta\mathbf{d} \in E_m$ представляет собой вектор, проходящий через точку \mathbf{x} и имеющий направление \mathbf{d} .

Пусть $f(\mathbf{x})$ — функция, определенная на E_m . Градиентом этой функции в точке \mathbf{x} называется вектор $\nabla f(\mathbf{x})$, определяемый через частные производные в точке $\mathbf{x} = (x_1, \dots, x_m)$ следующим образом:

$$\nabla f(\mathbf{x}) \triangleq (f'_{x_1}, \dots, f'_{x_m}). \quad (2.8.5)$$

Нетрудно проверить, что для произвольного направления $\mathbf{d} = (d_1, \dots, d_m)$

$$\lim_{\theta \rightarrow 0} \frac{f(\mathbf{x} + \theta\mathbf{d}) - f(\mathbf{x})}{\theta} = \sum_{i=1}^m f'_{x_i} d_i = \nabla f(\mathbf{x}) \mathbf{d}^T, \quad (2.8.6)$$

где в правой части равенства написано матричное произведение вектора-строки $\nabla f(\mathbf{x})$ на вектор-столбец \mathbf{d}^T . Градиент $\nabla f(\mathbf{x})$ определяет тем самым скорость роста функции $f(\mathbf{x})$ в направлении

вектора \mathbf{d} . Если для некоторого направления \mathbf{d} выполняется неравенство $\nabla f(\mathbf{x}) \mathbf{d}^T > 0$, то найдется окрестность точки \mathbf{x} , в которой функция $f(\mathbf{x})$ возрастает в направлении вектора \mathbf{d} (см. рис. 2.8.1). Как следует из соотношения (2.8.6), величина приращения при малых θ , $\theta > 0$, примерно равна $\theta \nabla f(\mathbf{x}) \mathbf{d}^T$.

Предположим, что задана функция $f(\mathbf{x})$, а также функции $g_j(\mathbf{x})$, $j = 1, \dots, s$, и $h_k(\mathbf{x})$, $k = 1, \dots, t$, $\mathbf{x} \in E_m$, причем все функции дифференцируемы в E_m . Рассмотрим задачу

$$\text{максимизировать } f(\mathbf{x}) \quad (2.8.7)$$

при ограничениях

$$\begin{cases} g_j(\mathbf{x}) \geq 0, & j = 1, \dots, s, \\ h_k(\mathbf{x}) = 0, & k = 1, \dots, t. \end{cases} \quad (2.8.8)$$

В этой задаче s условий имеют вид неравенств и t условий — вид равенств.

Заметим, что ограничение $g(\mathbf{x}) \leq 0$ эквивалентно ограничению $-g(\mathbf{x}) \geq 0$, а минимизация функции $f(\mathbf{x})$ при некоторых ограничениях эквивалентна максимизации функции $-f(\mathbf{x})$ при тех же ограничениях, поэтому достаточно всегда рассматривать задачу (2.8.7) при ограничениях (2.8.8).

Если ограничения в задаче поиска максимума отсутствуют, то $\nabla f(\mathbf{x}^*) = 0$ в точке максимума $\mathbf{x} = \mathbf{x}^*$. Другими словами, движение из \mathbf{x}^* в любом направлении $\mathbf{d} \in E_m$ не увеличивает значение функции $f(\mathbf{x})$ в некоторой окрестности точки \mathbf{x}^* . Если же ограничения имеются, то в точке максимума \mathbf{x}^* градиент $\nabla f(\mathbf{x}^*)$ не обязательно равен нулю и, вообще говоря, может найтись направление \mathbf{d} , в котором функция $f(\mathbf{x})$ возрастает. Однако теперь следует учитывать не все возможные направления, а лишь те, движения в направлении которых не нарушают заданных ограничений. Для каждой точки $\mathbf{x} \in E_m$, удовлетворяющей условиям (2.8.8) (такие точки будем называть *допустимыми*), и некоторого $\sigma > 0$ имеется множество $D_\sigma(\mathbf{x})$ *допустимых направлений*:

$$D_\sigma(\mathbf{x}) \triangleq \{\mathbf{d}: \mathbf{x} + \theta\mathbf{d} \text{ — допустимая точка для любого } \theta, 0 < \theta < \sigma\}. \quad (2.8.9)$$

Точка \mathbf{x}^* будет доставлять максимум функции $f(\mathbf{x})$ при заданных ограничениях, если движение в любом допустимом направлении $\mathbf{d} \in D_\sigma(\mathbf{x}^*)$ не увеличивает значение функции. Таким образом, если \mathbf{x}^* — оптимальная точка, то существует $\sigma > 0$ такое, что

$$\nabla f(\mathbf{x}^*) \mathbf{d}^T \leq 0, \quad \mathbf{d} \in D_\sigma(\mathbf{x}^*). \quad (2.8.10)$$

Условия Куна—Таккера позволяют строго сформулировать приведенные выше соображения, а также придать им более удобную форму.

Рассмотрим множество $D_\sigma(\mathbf{x})$ более подробно, чтобы описать его в явной форме через ограничения. В каждой допустимой точке \mathbf{x} ограничения типа неравенств могут быть разбиты на два множества: на те, которые *активны* в \mathbf{x} — для них $g_j(\mathbf{x}) = 0$, и на те, которые *неактивны* в \mathbf{x} — для последних $g_j(\mathbf{x}) > 0$. Пусть $J(\mathbf{x})$ — множество тех индексов $j \in [1, s]$, для которых $g_j(\mathbf{x}) = 0$, т. е. $J(\mathbf{x})$ — множество индексов активных ограничений. Для неактивных ограничений $g_j(\mathbf{x}) > 0$ и в силу непрерывности функций $g_j(\mathbf{x})$ существует некоторая окрестность точки \mathbf{x} , движение внутри которой не нарушает этих ограничений.

Определим множество $\tilde{D}(\mathbf{x})$ следующим образом:

$$\tilde{D}(\mathbf{x}) \triangleq \{\mathbf{d}: \nabla g_j(\mathbf{x}) \mathbf{d}^\tau \geq 0, j \in J(\mathbf{x}), \nabla h_k(\mathbf{x}) \mathbf{d}^\tau = 0, k = 1, \dots, t\}. \quad (2.8.11)$$

Лемма 2.8.1. Для каждой допустимой точки $\mathbf{x} \in E_m$ и любого $\sigma > 0$ имеет место включение $D_\sigma(\mathbf{x}) \subseteq \tilde{D}(\mathbf{x})$.

Доказательство. Пусть зафиксировано $\sigma > 0$ и $\mathbf{d} \in D_\sigma(\mathbf{x})$. Покажем, что $\mathbf{d} \in \tilde{D}(\mathbf{x})$. Предположим вначале, что $\nabla h_k(\mathbf{x}) \mathbf{d}^\tau > 0$ хотя бы для одного индекса $k \in [1, t]$. Тогда из (2.8.6) следует, что $h_k(\mathbf{x}) < h_k(\mathbf{x} + \theta \mathbf{d}) = 0$, $0 < \theta < \sigma$. Это противоречит предположению о том, что \mathbf{x} — допустимая точка. Следовательно, $\nabla h_k(\mathbf{x}) \mathbf{d}^\tau \leq 0$. Аналогично доказывается, что $\nabla h_k(\mathbf{x}) \mathbf{d}^\tau \geq 0$. Окончательно имеем $\nabla h_k(\mathbf{x}) \mathbf{d}^\tau = 0$ для всех k . Предположим теперь, что $\nabla g_j(\mathbf{x}) \mathbf{d}^\tau < 0$ для некоторого $j \in J(\mathbf{x})$. Тогда из (2.8.6) следует, что $g_j(\mathbf{x} + \theta \mathbf{d}) < g_j(\mathbf{x}) = 0$. Это противоречит предположению, что $\mathbf{d} \in D_\sigma(\mathbf{x})$. Следовательно, $\nabla g_j(\mathbf{x}) \mathbf{d}^\tau \geq 0$ для всех $j \in J(\mathbf{x})$. Лемма доказана.

Следует сказать, что в общем случае имеются направления из $\tilde{D}(\mathbf{x})$, которые не принадлежат $D_\sigma(\mathbf{x})$ ни при каком $\sigma > 0$ (см. задачу 2.8.1). Вместе с тем для широкого класса функций $f(\mathbf{x})$, $g_j(\mathbf{x})$, $h_k(\mathbf{x})$ существует $\sigma > 0$, для которого $D_\sigma(\mathbf{x}) = \tilde{D}(\mathbf{x})$ в оптимальной точке $\mathbf{x} = \mathbf{x}^*$.

Говорят, что выполнено *условие регулярности*, если для некоторого $\sigma > 0$ в оптимальной точке \mathbf{x}^* имеет место равенство

$$D_\sigma(\mathbf{x}^*) = \tilde{D}(\mathbf{x}^*). \quad (2.8.12)$$

В дальнейшем мы будем предполагать, что условие регулярности выполняется. Заметим, что в случае, когда (2.8.8) представляют собой ограничения, которым удовлетворяют только вероятностные векторы, выполнение условия регулярности (2.8.12) может быть легко показано (см. задачу 2.8.2).

Таким образом, из (2.8.10) и условия регулярности вытекает, что для оптимальной точки \mathbf{x}^*

$$\nabla f(\mathbf{x}^*) \mathbf{d}^\tau \leq 0, \quad \mathbf{d} \in \tilde{D}(\mathbf{x}^*). \quad (2.8.13)$$

Теперь будет приведено без доказательства утверждение, известное под названием леммы Фаркаша.

Лемма 2.8.2. Пусть \mathbf{a}_k , $k = 0, 1, \dots, n$, — множество векторов из E_m . Для того чтобы существовали n неотрицательных чисел $\lambda_1, \dots, \lambda_n$ таких, что $\sum_{i=1}^n \lambda_i \mathbf{a}_i = \mathbf{a}_0$, необходимо и достаточно, чтобы для любого вектора $\mathbf{z} \in E_m$, для которого $\mathbf{a}_k \mathbf{z}^\tau \geq 0$, $k = 1, \dots, n$, выполнялось неравенство $\mathbf{a}_0 \mathbf{z}^\tau \geq 0$.

Из этой леммы и определения множества $\tilde{D}(\mathbf{x})$ непосредственно следует теорема Куна—Таккера, дающая необходимые условия решения задачи (2.8.7) при ограничениях (2.8.8).

Теорема 2.8.1 (Куна—Таккера). Пусть \mathbf{x}^* — решение задачи (2.8.7) при ограничениях (2.8.8) и выполнено условие регулярности. Тогда существуют числа λ_j , $j = 1, \dots, s$, μ_k , $k = 1, \dots, t$ (неопределенные множители Лагранжа), такие, что

$$1) \lambda_j g_j(\mathbf{x}^*) = 0, \quad \lambda_j \geq 0, \quad j = 1, \dots, s,$$

2) имеет место равенство

$$\nabla f(\mathbf{x}^*) + \sum_{j=1}^s \lambda_j \nabla g_j(\mathbf{x}^*) + \sum_{k=1}^t \mu_k \nabla h_k(\mathbf{x}^*) = 0. \quad (2.8.14)$$

Доказательство. Заметим, что ограничение $h(\mathbf{x}) = 0$ можно записать с помощью двух ограничений — неравенств, а именно $h(\mathbf{x}) \geq 0$, $-h(\mathbf{x}) \geq 0$. Поэтому множество $\tilde{D}(\mathbf{x})$ (см. (2.8.11)) можно определить следующим образом:

$$\begin{aligned} \tilde{D}(\mathbf{x}) = \{&\mathbf{d}: \nabla g_j(\mathbf{x}) \mathbf{d}^\tau \geq 0, \quad j \in J(\mathbf{x}), \quad \nabla h_k(\mathbf{x}) \mathbf{d}^\tau \geq 0, \\ &-\nabla h_k(\mathbf{x}) \mathbf{d}^\tau \geq 0, \quad k = 1, \dots, t\}. \end{aligned}$$

Теперь из (2.8.13) следует, что для векторов

$$-\nabla f(\mathbf{x}^*), \quad \nabla g_j(\mathbf{x}^*), \quad j \in J(\mathbf{x}^*), \quad \nabla h_k(\mathbf{x}^*), \quad -\nabla h_k(\mathbf{x}^*), \quad k = 1, \dots, t,$$

выполнены условия леммы Фаркаша и поэтому существуют такие неотрицательные числа λ_j , $j \in J(\mathbf{x}^*)$, μ'_k , μ''_k , $k = 1, \dots, t$, что

$$0 = \nabla f(\mathbf{x}^*) + \sum_{j \in J(\mathbf{x}^*)} \lambda_j \nabla g_j(\mathbf{x}^*) + \sum_{k=1}^t \mu'_k \nabla h_k(\mathbf{x}^*) - \sum_{k=1}^t \mu''_k \nabla h_k(\mathbf{x}^*).$$

Если обозначить $\mu_k = \mu'_k - \mu''_k$ и положить $\lambda_j = 0$ для всех $j \notin J(\mathbf{x}^*)$, то последнее соотношение примет вид (2.8.14). Отсюда также следует, что для всех $j = 1, \dots, s$ имеет место равенство $\lambda_j g_j(\mathbf{x}^*) = 0$. Теорема доказана.

2.8.3. Достаточность условий Куна—Таккера для выпуклых функций. Пусть $f(\mathbf{x})$ — выпуклая вверх функция, определенная в некоторой выпуклой области пространства E_m , заданной с помощью ограничений $g_j(\mathbf{x}) \geq 0$, $j = 1, \dots, s$, где $g_j(\mathbf{x})$ — также выпуклые вверх функции. Тогда функция $f(\mathbf{x})$ обладает тем свойством, что локальный ее максимум совпадает с глобальным. Мы покажем, что в этом случае необходимые условия максимума являются также достаточными.

Задача поиска максимума выпуклой вверх функции $f(\mathbf{x})$ в выпуклой области пространства E_m , задаваемой s выпуклыми вверх функциями, может быть сформулирована следующим образом:

$$\text{максимизировать } f(\mathbf{x}) \quad (2.8.15)$$

при ограничениях

$$g_j(\mathbf{x}) \geq 0, \quad j = 1, \dots, s, \quad (2.8.16)$$

где $f(\mathbf{x})$ и $g_j(\mathbf{x})$, $j = 1, \dots, s$, — выпуклые вверх функции.

Теорема 2.8.2. Предположим, что $\mathbf{x}^* \in E_m$ удовлетворяет условиям Куна—Таккера для задачи (2.8.15) при ограничениях (2.8.16). Тогда $f(\mathbf{x}^*) \geq f(\mathbf{x}_0)$ для любой точки $\mathbf{x}_0 \in E_m$, удовлетворяющей ограничениям (2.8.16).

Доказательство. Так как ограничения — выпуклые вверх функции, то множество допустимых точек выпукло. Следовательно, все точки прямой, соединяющей \mathbf{x}^* и \mathbf{x}_0 , допустимы, а вектор $\mathbf{d}_0 = \mathbf{x}_0 - \mathbf{x}^*$ является допустимым направлением. По лемме 2.8.1 $\nabla g_j(\mathbf{x}^*) \mathbf{d}_0^T \geq 0$, $j \in J(\mathbf{x}^*)$. Согласно теореме Куна—Таккера существуют числа λ_j , $j = 1, \dots, s$, такие, что $\lambda_j \geq 0$, $\lambda_j g_j(\mathbf{x}) = 0$ и

$$\nabla f(\mathbf{x}^*) + \sum_{j=1}^s \lambda_j \nabla g_j(\mathbf{x}^*) = 0.$$

Поэтому

$$-\nabla f(\mathbf{x}^*) \mathbf{d}_0^T = \sum_{j=1}^s \lambda_j \nabla g_j(\mathbf{x}^*) \mathbf{d}_0^T \geq 0$$

и для направления \mathbf{d}_0 имеет место неравенство $\nabla f(\mathbf{x}^*) \mathbf{d}_0^T \leq 0$. В силу выпуклости функции $f(\mathbf{x})$ из (2.8.6) имеем

$$\begin{aligned} 0 &\geq \nabla f(\mathbf{x}^*) \mathbf{d}_0^T = \\ &= \lim_{\theta \rightarrow 0} \frac{f(\mathbf{x}^* + \theta \mathbf{d}_0) - f(\mathbf{x}^*)}{\theta} = \lim_{\theta \rightarrow 0} \frac{f[\theta(\mathbf{x}^* + \mathbf{d}_0) + (1 - \theta)\mathbf{x}^*] - f(\mathbf{x}^*)}{\theta} \geq \\ &\geq \lim_{\theta \rightarrow 0} \frac{\theta f(\mathbf{x}^* + \mathbf{d}_0) + (1 - \theta)f(\mathbf{x}^*) - f(\mathbf{x}^*)}{\theta} = \\ &= f(\mathbf{x}^* + \mathbf{d}_0) - f(\mathbf{x}^*) = f(\mathbf{x}_0) - f(\mathbf{x}^*). \end{aligned}$$

Следовательно, $f(\mathbf{x}^*) \geq f(\mathbf{x}_0)$ для любой допустимой точки $\mathbf{x}_0 \in E_m$. Теорема доказана.

2.8.4. Поиск экстремумов на множестве вероятностных векторов. Ниже будет рассмотрен важный частный случай, когда множество векторов, на котором ищется экстремум, является выпуклым множеством вероятностных векторов $\mathbf{x} = (x_1, \dots, x_m)$, определяемым ограничениями

$$\begin{cases} g_j(\mathbf{x}) = x_j \geq 0, & j = 1, \dots, m, \\ h(\mathbf{x}) = \sum_{j=1}^m x_j - 1 = 0. \end{cases} \quad (2.8.17)$$

В задаче 2.8.2 показано, что в этом случае условие регулярности выполнено и теорема Куна—Таккера о необходимых условиях максимума может быть сформулирована следующим образом.

Следствие 2.8.1 (теорема Куна—Таккера для вероятностных ограничений). Пусть R — множество вероятностных векторов, определяемое условиями (2.8.17), и $f(\mathbf{x})$ — функция, определенная на R и имеющая частные производные f'_{x_i} , $i = 1, \dots, m$. Пусть $\mathbf{x}^* \in R$ максимизирует $f(\mathbf{x})$ на множестве R . Тогда существует число μ такое, что

$$f'_{x_i}|_{\mathbf{x}=\mathbf{x}^*} \leq \mu, \quad i = 1, \dots, m, \quad (2.8.18)$$

причем (2.8.18) выполняется со знаком равенства для всех таких i , для которых $x_i^* > 0$.

Доказательство. Так как $h'_{x_i} = 1$ и

$$\frac{\partial g_j(\mathbf{x})}{\partial x_i} = \begin{cases} 1, & \text{если } i = j, \\ 0, & \text{если } i \neq j, \end{cases}$$

то из (2.8.14) следует, что

$$f'_{x_i}|_{\mathbf{x}=\mathbf{x}^*} = -\lambda_i + \mu \leq \mu, \quad i = 1, \dots, m,$$

так как λ_i — неотрицательные величины и $\lambda_i = 0$, если $x_i^* > 0$. Следствие доказано.

Поскольку область R вероятностных векторов выпукла, то из теоремы 2.8.2 вытекает, что в случае, когда $f(\mathbf{x})$ — выпуклая вверх функция, необходимые условия максимума являются также достаточными. Таким образом, верно следующее утверждение.

Следствие 2.8.2. Пусть R — множество вероятностных векторов, определяемое условиями (2.8.17), и $f(\mathbf{x})$ — выпуклая вверх функция, определенная на R и имеющая частные производные f'_{x_i} , $i = 1, \dots, m$. Если существует точка \mathbf{x}^* и число μ такое, что

$$f'_{x_i}|_{\mathbf{x}=\mathbf{x}^*} \leq \mu, \quad i = 1, \dots, m,$$

причем знак равенства имеет место для всех таких i , для которых $x_i^* > 0$, то \mathbf{x}^* есть точка, в которой $f(\mathbf{x})$ принимает максимальное значение.

З а м е ч а н и я. 1. Для того чтобы минимизировать выпуклую вниз функцию $f(\mathbf{x})$ на множестве R , достаточно максимизировать выпуклую вверх функцию $-f(\mathbf{x})$ на том же множестве. Необходимые и достаточные условия того, что \mathbf{x}^* минимизирует $f(\mathbf{x})$, состоят в том, что существует такое μ , что

$$f'_{x_i}|_{\mathbf{x}=\mathbf{x}^*} \leq \mu, \quad i = 1, \dots, m,$$

причем знак равенства имеет место для всех таких i , для которых $x_i^* > 0$.

2. Если функция $h(\mathbf{x})$ в (2.8.17) имеет вид $\sum_{i=1}^m x_i - c$, где c — положительная константа, то с помощью нормировки (деления всех переменных на c) множество, на котором ищется экстремум, приводится к множеству вероятностных векторов.

Полезно прояснить необходимые условия максимума (2.8.18) с помощью простых рассуждений. Рассмотрим задачу максимизации выпуклой вверх функции $f(\mathbf{x})$ в области вероятностных векторов. Так как для каждого вероятностного вектора (x_1, \dots, x_m) должно выполняться равенство $\sum_{i=1}^m x_i = 1$, то, не обращая пока

внимания на требование неотрицательности величин x_i , $i = 1, \dots, m$, можно воспользоваться методом неопределенных множителей Лагранжа и попытаться найти максимум функции Лагранжа $L(\mathbf{x}) \triangleq f(\mathbf{x}) - \mu \left(\sum_{i=1}^m x_i - 1 \right)$. Если максимум существует внутри области вероятностных векторов, т. е. при $x_i^* > 0$ для всех i , то производная функции Лагранжа по каждой из переменных x_i , $i = 1, \dots, m$, будет равна нулю в точке максимума и условия (2.8.18) будут выполняться со знаком равенства. Если же максимум функции Лагранжа существует вне области вероятностных векторов, то хотя бы для одного индекса i будет выполняться неравенство $x_i < 0$. Мы предположили, что $f(\mathbf{x})$ — выпуклая вверх функция, поэтому выпуклой вверх является и функция $L(\mathbf{x})$. Отсюда следует, что максимум функции $f(\mathbf{x})$ в области вероятностных векторов достигается на границе, т. е. при условии, что $x_i^* = 0$ для некоторых индексов i . Так как выпуклая функция может иметь только один максимум, то производные функции $L(\mathbf{x})$ по тем переменным, для которых $x_i^* = 0$, не положительны

$$L'_{x_i}|_{\mathbf{x}=\mathbf{x}^*} \leq 0, \quad i \in J(\mathbf{x}^*).$$

Условия Куна—Таккера сами по себе не дают метода отыскания экстремальных точек. В общем случае нахождение таких точек, исходя из необходимых или необходимых и достаточных условий, приведенных выше, является довольно сложной задачей. Однако в ряде случаев эта задача может иметь простое решение. В следующих примерах иллюстрируется применение условий Куна—Таккера.

Пример 2.8.1. Найдем распределение вероятностей на дискретном ансамбле, которое максимизирует энтропию $H(X) = -\sum_{i=1}^m p_i \log p_i$. Хотя ответ известен и получается просто с помощью применения неравенства для логарифма, можно формально использовать теорему Куна—Таккера для отыскания необходимых условий экстремума. В этой задаче $f(\mathbf{p}) = -\sum_{i=1}^m p_i \log p_i$ и $\sum_{i=1}^m p_i = 1$, $p_i \geq 0$, $i = 1, \dots, m$. Согласно следствию 2.8.1 в точке максимума должны выполняться соотношения

$$\frac{\partial f(\mathbf{p}^*)}{\partial p_i} = -\log e [\ln p_i^* + 1] \leq \mu, \quad i = 1, \dots, m, \quad (2.8.19)$$

для некоторого вероятностного вектора $\mathbf{p}^* = (p_1^*, \dots, p_m^*)$ и некоторого числа μ , причем для всех $p_i^* > 0$ должно выполняться равенство. Из (2.8.19) получим, что для оптимального распределения \mathbf{p}

$$p_i^* \geq \frac{2^{-\mu}}{e}, \quad i = 1, \dots, m,$$

со знаком равенства для всех $p_i^* > 0$. Другими словами, необходимое условие того, чтобы \mathbf{p}^* было распределением, максимизирующим энтропию, состоит в том, что $p_i^* = \text{const}$, $i = 1, \dots, m$. Поскольку имеется только одно такое распределение, а именно равномерное, то это условие является и достаточным.

Пример 2.8.2. В задаче 2.3.5 предлагается получить следующее выражение для средней взаимной информации между двумя гауссовскими векторами \mathbf{X} и \mathbf{Y} , где $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$, причем \mathbf{X} и \mathbf{Z} статистически независимы и имеют независимые компоненты:

$$I(\mathbf{X}; \mathbf{Y}) = \frac{1}{2} \sum_{i=1}^n \log \left(1 + \frac{\lambda_i}{v_i} \right) = f(\boldsymbol{\lambda}).$$

В этой формуле λ_i, v_i , $i = 1, \dots, n$, дисперсии компонент векторов \mathbf{X} и \mathbf{Z} соответственно. Найдем максимальное значение $I(\mathbf{X}; \mathbf{Y})$ по всем векторам $\boldsymbol{\lambda}$, для которых $\sum_{i=1}^n \lambda_i = P$. Таким образом, задача заключается в максимизации функции $f(\boldsymbol{\lambda})$ при ограничениях $\sum_{i=1}^n \lambda_i = P$, $\lambda_i \geq 0$, $i = 1, \dots, n$. Нетрудно убедиться

в том, что $f(\boldsymbol{\lambda})$ является выпуклой вверх функцией, определенной на множестве таких векторов $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)$, что $\lambda_i \geq 0$ и $\sum_{i=1}^n (\lambda_i/P) = 1$. Необходимые и

достаточные условия максимума определяются теоремой Куна—Таккера. Из следствия 2.8.2 получим, что для оптимального вектора λ^* выполняются неравенства

$$v_i + \lambda_i^* \geq \frac{\log e}{2\mu} \triangleq B$$

для некоторого μ , причем равенство имеет место для всех $\lambda_i^* > 0$. Таким образом, оптимальный выбор дисперсий компонент вектора λ при ограничении на суммарную дисперсию задается следующими уравнениями:

$$\lambda_i^* = \begin{cases} B - v_i, & \text{если } v_i \leq B, \\ 0, & \text{если } v_i > B, \end{cases}$$

а постоянную B можно найти из условия на суммарную дисперсию

$$\sum_{i: v_i \leq B} (B - v_i) = P.$$

Задачи, упражнения и дополнения

2.1.1. Предположим, что XYZ — дискретный ансамбль и распределение вероятностей на тройках $(x, y, z) \in XYZ$ таково, что для всех (x, y, z) имеет место равенство $p(x, y, z) = p_1(x) \cdot p(y, z)$. Покажите, что

$$I(y; z|x) = I(y; z), \quad I(x; z|y) = I(x; z) = 0$$

и, следовательно, $I((x, y); z) = I(x; z) + I(y; z)$. Найдите другие случаи, когда выполняется последнее равенство.

2.1.2. Пусть задан ансамбль $\{XYUV, p(x, y, u, v)\}$, элементами которого являются всевозможные четверки (x, y, u, v) , и распределение вероятностей таково, что $p(x, y, u, v) = p_1(x, y) p_2(u, v)$ для всех четверок. Это означает, что ансамбли XY и UV статистически независимы.

а) Пользуясь аддитивностью и симметричностью взаимной информации, покажите, что

$$I((x, u); (y, v)) = I(x; y) + I(x; v|y) + I(u; y|x) + I(u; v|x, y).$$

б) Пользуясь теперь независимостью пар (x, y) и (u, v) , покажите, что $I((x, u); (y, v)) = I(x; y) + I(u; v)$. Таким образом, взаимная информация между парами сообщений, содержащими независимые компоненты, равна сумме взаимных информаций, где слагаемые соответствуют независимым парам. Этому случаю соответствует, например, следующая физическая модель. Пусть имеются два независимых канала связи со входными сигналами X и U и выходными сигналами Y и V соответственно. Тогда количество информации между сообщениями на входе и выходе таких каналов (их называют параллельными) равно сумме количеств информации для каждого канала в отдельности.

2.1.3. Покажите, что все следующие средние взаимные информации $I(X; Y|z)$, $I(X; Y|Z)$, $I(XY; Z)$ неотрицательны. Тем самым получает некоторое полезное обобщение теоремы 2.1.1.

2.1.4. Пусть $Z = \varphi(X)$ — отображение множества X в множество Z , описанное в теореме 2.1.2. Как было указано, этому отображению соответствует такое распределение вероятностей на тройках, что $p(z|x, y) = p(z|x)$ для всех $(x, y, z) \in XYZ$. Покажите, что это условие эквивалентно следующему: $p(y|x, z) = p(y|x)$ для всех $(x, y, z) \in XYZ$, т. е. при данном x сообщения y и z независимы.

2.2.1. Дискретные случайные величины обычной ф. п. в. не имеют. Вводя в рассмотрение обобщенные функции и, в частности, дельта-функцию Дирака (см. 2.2.18), можно расширить понятие ф. п. в. и определить ф. п. в. для дискретных случайных величин.

а) Пусть $X = \{x_1, \dots, x_M\}$ — произвольное числовое множество и $\{p_1, \dots, p_M\}$ — распределение вероятностей на X . Очевидно, что функция распределения $F(x) = \sum p_i$, где суммирование выполняется по всем таким i , что $x_i \leq x$. Покажите, что

$$f(x) = \sum_{i=1}^n p_i \delta(x - x_i)$$

есть формальное выражение для ф. п. в. на X . Для этого проверьте, что для любого x

$$\int_{-\infty}^x f(x) dx = F(x).$$

Покажите, что k -й начальный момент с. в. X , т. е. величина $\sum_i x_i^k p_i$, может быть записан как

$$\int x^k f(x) dx.$$

Объясните, почему энтропию $H(X)$ ансамбля X нельзя выразить через обобщенную ф. п. в.

б) Пусть $X = \{x_1, \dots, x_M\}$ и $Y = \{y_1, \dots, y_N\}$ — два дискретных множества и $p(x_i, y_j)$ — распределение вероятностей на XY . Пусть $f(x)$, $f(y)$ и $f(x, y)$ — обобщенные ф. п. в. на X , Y и XY соответственно (см. (2.2.21), (2.2.23)). Покажите, что информация $I(x_i; y_j)$ может быть записана в виде

$$I(x_i; y_j) = \log \frac{f(x, y)}{f(x) f(y)} \quad \text{при } x = x_i, \quad y = y_j.$$

Покажите, что средние взаимные информации

$$I(X; y_j) = \int_X f(x) \log \frac{f(x|y_j)}{f(x)} dx \quad \text{для } y = y_j,$$

$$I(X; Y) = \int_{X, Y} f(x, y) \log \frac{f(x|y)}{f(x)} dx dy = \int_{X, Y} f(x, y) \log \frac{f(y|x)}{f(y)} dx dy.$$

в) Пусть $d(x, y)$ — некоторая функция на XY . Покажите, что

$$M_y = d(x, y) \triangleq \sum_{x_i} p(x_i|y) d(x_i, y) = \int_X d(x, y) f(x|y) dx$$

для $y = y_1, \dots, y_N$.

2.2.2. Пусть $\{X, f(x)\}$ — непрерывный ансамбль с равномерным на отрезке $(0, a)$ распределением вероятностей, т. е. $f(x) = c$ для $x \in (0, a)$ и $f(x) = 0$ для остальных x . Всякая функция $\varphi(x)$, определенная на X , является по определению случайнай величиной на ансамбле X .

а) Определите константу c .

б) Найдите функцию $\varphi(x)$, дающую дискретную с. в. из примера 2.2.1.

в) Пусть $a = 1$ и $\Phi(y) \triangleq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y \exp\left(-\frac{x^2}{2}\right) dx$ — так называемый интеграл вероятностей. Очевидно, функция распределения $F(y)$ гауссовой с. в. с нулевым средним и дисперсией 1 совпадает с $\Phi(y)$. Покажите, что функция

$\varphi_1(x) \triangleq \Phi^{-1}(x)$, где $\Phi^{-1}(x)$ — функция, обратная к $\Phi(y)$, т. е. $\Phi(\Phi^{-1}(x)) = x$, задает гауссовскую с. в. с нулевым средним и дисперсией 1. Покажите, что $\varphi_2(x) = \varphi_1(x) + m$ задает гауссовскую с. в. с математическим ожиданием m и дисперсией σ^2 . Указано: воспользуйтесь тем, что $\Phi(y)$ — монотонная функция, и тем, что $\Pr(\varphi(x) \leq y) = \Pr(x \leq \varphi^{-1}(y))$ для монотонной функции $\varphi(x)$.

2.2.3. Покажите, что необходимым и достаточным условием независимости двух ансамблей с гауссовским совместным распределением вероятностей является равенство нулю коэффициента корреляции.

2.2.4. Пусть $f(x) = (\sigma\sqrt{2\pi})^{-1} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]$ — ф. п. в. гауссовского

распределения со средним значением m и дисперсией σ^2 . Пусть $\varphi(x) \triangleq \exp(\omega x^2)$, где $\omega < 1/\sigma^2$ — некоторая константа.

Покажите, что

$$\mathbf{M}\varphi(x) = (1 - 2\omega\sigma^2)^{-1/2} \exp[\omega m^2/(1 - 2\omega\sigma^2)].$$

Указано: воспользуйтесь тем, что при всех m , σ интеграл от $f(x)$ по всей оси X равен 1.

2.2.5. Пусть $f(x, y)$ определено, как в примере 2.2.4. Покажите, что

$$\begin{aligned} f(x|y) &= \\ &= \frac{1}{\sigma_X \sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{1}{2\sigma_X^2(1-\rho^2)} \left(x - m_X - \rho \frac{\sigma_X}{\sigma_Y} (y - m_Y)\right)^2\right]. \end{aligned}$$

Средняя взаимная информация между непрерывным ансамблем X и фиксированным сообщением y определяется соотношением

$$I(X; y) \triangleq \mathbf{M}_y I(x; y) = \int_X f(x|y) \log \frac{f(x|y)}{f(x)} dx.$$

Покажите, что

$$I(X; y) = -\frac{1}{2} \log(1 - \rho^2) + \frac{\rho^2}{2} \left(1 - \left(\frac{y - m_Y}{\sigma_Y}\right)^2\right) \log e$$

и, следовательно,

$$I(X; Y) = \mathbf{MI}(X; y) = -\frac{1}{2} \log(1 - \rho^2),$$

как это было найдено в примере 2.2.4.

2.2.6. Пусть X и Y — две с. в., имеющие совместное гауссовское распределение $f(x, y)$ (см. пример 2.2.4). Цель этой задачи — показать, что сумма $Z = X + Y$ двух гауссовых с. в. снова является гауссовой с. в. Для этого обозначим через $f_Y(y)$ и $f_X(x|y)$ безусловную и условную ф. п. в. для с. в. Y и X соответственно, кроме того, обозначим через $f_Z(z)$ ф. п. в. для с. в. Z .

а) Покажите, что $f_Z(z) = \int f_Y(y) f_X(z - y|y) dy$.

б) Пользуясь задачей 2.2.5, покажите, что ф. п. в. $f_Z(z)$ может быть представлена в форме

$$f_Z(z) = \frac{1}{\sigma_Z \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_Z^2} (z - m_Z)^2\right]$$

и, следовательно, Z является гауссовой с. в. Найдите параметры m_Z и σ_Z^2 .

2.2.7. Покажите, что для непрерывных с. в. с нулевым математическим ожиданием и конечной дисперсией имеет место неравенство Чебышева

$$\Pr(|X| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}, \quad \varepsilon > 0.$$

Покажите также, что для всякого $k \geq 0$ имеет место следующее неравенство (обобщенное неравенство Чебышева):

$$\Pr(|X| \geq \varepsilon) \leq \mu_k \varepsilon^k, \quad \varepsilon > 0,$$

где

$$\mu_k \triangleq \int |x|^k f(x) dx < \infty.$$

2.2.8. Покажите, что для независимых одинаково распределенных непрерывных с. в. $\{X_j, f(x_j)\}$, $j = 1, 2, \dots, n$, с ограниченными математическим ожиданием m и дисперсией σ^2 имеет место закон больших чисел

$$\Pr\left(\left|\frac{1}{n} \sum_{j=1}^n X_j - m\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}, \quad \varepsilon > 0.$$

2.2.9. Пусть в предыдущей задаче X_j — гауссовские с. в. с нулевым средним и дисперсией 1.

а) Найдите функцию плотности вероятностей с. в. $\frac{1}{n} \sum_{j=1}^n X_j - m$. Указано: воспользуйтесь результатом задачи 2.2.6.

б) Известно [11], что функция $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{x^2}{2}} dx$ может быть определена следующим образом:

$$\frac{1}{\sqrt{2\pi}} \left(\frac{1}{x} - \frac{1}{x^3}\right) e^{-\frac{x^2}{2}} < 1 - \Phi(x) < \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{x} e^{-\frac{x^2}{2}}, \quad x > 0.$$

Уточните правую часть неравенства в предыдущей задаче.

2.3.1. Докажите, что $H_0(Y|X) \leq H_0(Y)$. Опишите случай равенства.

2.3.2. Пусть $f_1(x)$ и $f_2(x)$ — функции плотности вероятностей двух случайных величин, отличающихся только математическим ожиданием. Пусть $H_0(X_1)$ и $H_0(X_2)$ — относительные энтропии этих с. в. Покажите, что $H_0(X_1) = H_0(X_2)$.

2.3.3. Пусть X — с. в. с плотностью вероятностей $f(x)$ и $Y = kX + m$, где k и m — фиксированные неслучайные величины. Покажите, что $H_0(Y) = H_0(X) + \log k$.

2.3.4. Пусть $\Phi_{a,b}$ — класс функций плотности вероятностей такой, что для любой функции $f(x) \in \Phi_{a,b}$, $f(x) = 0$, если x не принадлежит интервалу (a, b) . Покажите, что

$$H_0(X) = - \int f(x) \log f(x) dx \leq \log(b-a)$$

для любой функции $f(x) \in \Phi_{a,b}$, причем равенство имеет место в том и только том случае, когда $f(x) = 1/(b-a)$ для всех $x \in (a, b)$.

2.3.5. Предположим, что $Z = (Z_1, \dots, Z_n)$ и $X = (X_1, \dots, X_n)$ — два случайных вектора, $Y = X + Z$, причем X и Z статистически независимы и

$$K_Z = \begin{bmatrix} \lambda_1 & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}.$$

а) Покажите, что средняя взаимная информация

$$I(X; Y) \leq \frac{1}{2} \log \frac{\det(K_X + K_Z)}{\lambda_1 \dots \lambda_n},$$

где равенство достигается в случае, когда X — гауссовский случайный вектор с корреляционной матрицей K_X .

б) Пусть $A = [a_{ij}]$ — корреляционная матрица некоторой системы с. в., тогда $\det A \leq \prod_{i=1}^n a_{ii}$ (см., например, [1]). Используя это неравенство, покажите, что

$$I(X; Y) \leq \frac{1}{2} \sum_{i=1}^n \log \left(1 + \frac{\sigma_{X_i}^2}{\lambda_i} \right),$$

где равенство достигается в том случае, когда X_i , $i = 1, \dots, n$, — независимые гауссовские с. в. с дисперсиями $\sigma_{X_i}^2$, $i = 1, 2, \dots, n$.

в) Пусть $n = 2$, $\lambda_1 = \lambda_2 = 1$, $X = (X_1, X_2)$ — гауссовский случайный вектор и

$$K_X = \begin{bmatrix} 1,5 & 0,5 \\ 0,5 & 1,5 \end{bmatrix}.$$

Найдите величину $I(X; Y)$. Найдите матрицу K_X , обеспечивающую максимум $I(X; Y)$ при условии, что дисперсии с. в. X_1 и X_2 сохраняют свои значения.

2.3.6. Пусть $X = (X_1, \dots, X_n)$ — вырожденная система с. в., т. е. либо некоторые из этих с. в. принимают неслучайные значения, либо некоторые из них являются линейными комбинациями остальных. Последнее означает, что существует неслучайный вектор a такой, что $aX^T = 0$.

а) Пусть $X_1 = (X_2 + X_3)/3$. Укажите вектор a такой, что $a(X_1, X_2, X_3)^T = 0$.

б) Известно, что определитель матрицы K равен нулю, если некоторые строки или столбцы матрицы линейно зависимы. В матрице K строки линейно зависимы, если существует ненулевой вектор a такой, что $aK = 0$ (где 0 — это нулевая строка).

Покажите, что в случае, когда $X_1 = b$, где b — неслучайная величина и K — корреляционная матрица системы с. в. X , имеет место равенство $\det K = 0$.

в) Предположим, что некоторые из с. в. в указанной выше системе являются линейными комбинациями остальных. Покажите, что в этом случае $\det K = 0$.

2.3.7. В каждой системе вырожденных с. в. имеется подсистема, образующая невырожденную систему. Пусть X_1, \dots, X_m — максимальная невырожденная подсистема системы с. в. X_1, \dots, X_n , т. е. такая, что корреляционная матрица K_m этой подсистемы не вырождена, а каждая из с. в. X_{m+1}, \dots, X_n либо принимает фиксированное значение, либо линейно зависит от с. в. X_1, \dots, X_m . Покажите, что в этом случае

$$I(X_1, \dots, X_n; Y) = I(X_1, \dots, X_m; Y).$$

Указание: покажите вначале, что $I(X_{m+1}, \dots, X_n; Y | X_1, \dots, X_m) = 0$.

2.4.1. Ниже мы приведем некоторые дополнительные сведения, относящиеся к многомерным гауссовским распределениям. В § 2.3 отмечалось, что матрица

$$K = [\mathbf{M}(X_i - m_i)(X_j - m_j)], \quad i, j = 1, \dots, n,$$

— корреляционная матрица системы случайных величин X_1, \dots, X_n — является симметрической, т. е. $K = K^T$, где « T » — символ транспозиции (замены местами строк и столбцов). Всякой невырожденной корреляционной матрице

соответствует n -мерное гауссовское распределение, задаваемое формулой (2.3.25). Вопрос, который представляет интерес, заключается в том, какая симметрическая матрица является корреляционной матрицей некоторой системы с. в.

Всякая корреляционная матрица является неотрицательно определенной матрицей. Поэтому необходимым условием того, чтобы K была корреляционной матрицей, является неотрицательная определенность K . Можно показать, что симметричность и неотрицательная определенность являются также достаточными условиями.

а) Пусть K неотрицательно определена и A — произвольная ненулевая квадратная $n \times n$ -матрица. Покажите, что матрица AKA^T также неотрицательно определена.

б) Пусть Q — ортогональная матрица, т. е. такая, что

$$QQ^T = Q^TQ = I,$$

где I — единичная матрица. Согласно предыдущему пункту матрица QKQ^T неотрицательно определена.

Известно [1], что для любой симметрической матрицы K существует такая ортогональная матрица Q , что

$$QKQ^T = \Lambda,$$

где $\Lambda = [\lambda_1, \dots, \lambda_n]$ — диагональная матрица с элементами $\lambda_1, \dots, \lambda_n$ на главной диагонали. Покажите, что

$$\det K = \det QKQ^T = \prod_{i=1}^n \lambda_i.$$

Покажите, что условие $\det K \neq 0$ является необходимым и достаточным условием положительной определенности матрицы K . Покажите, что отсюда следует положительность собственных чисел.

в) Пусть K_X — положительно определенная матрица и

$$K_Z = \begin{bmatrix} \varepsilon_0 & & 0 \\ & \ddots & \\ 0 & & \varepsilon_0 \end{bmatrix}$$

— диагональная матрица с неотрицательными элементами $\varepsilon_0, \dots, \varepsilon_0$ на главной диагонали. Покажите, что матрица $K_Y \triangleq K_X - K_Z$ положительно определена, если $\varepsilon_0 < \lambda_i$, где $\lambda_1, \dots, \lambda_n$ — собственные числа матрицы K_X . Покажите также, что

$$\det K_Y = \prod_{i=1}^n (\lambda_i - \varepsilon_0).$$

Указание: воспользуйтесь тем, что для любой ортогональной матрицы Q выполняется равенство $QK_ZQ^T = K_Z$.

2.4.2. Пусть $n \times n$, $m \times m$ -матрицы K_Y и K_Z положительно определены. Покажите, что существует $n+m$ -мерное невырожденное гауссовское распределение вероятностей такое, что системы с. в. Y и Z , для которых K_Y и K_Z являются корреляционными матрицами, являются независимыми.

Указание: воспользуйтесь тем, что из некоррелированности гауссовых с. в. вытекает их независимость.

2.4.3. Пусть симметрические матрицы A и B положительно определены. Покажите, что для любых $\alpha > 0$, $\beta \geq 0$ матрица $\alpha A + \beta B$ — симметрическая и $\det(\alpha A + \beta B) > 0$. Покажите, что то же утверждение справедливо и в случае, когда матрица B неотрицательно определена.

2.4.4. Пусть Y — ортогональное преобразование случайного вектора X , т. е. $Y = XQ$, где Q — ортогональная матрица. Покажите, что

$$H_0(Y) = H_0(X).$$

Таким образом, относительная энтропия n -мерного распределения вероятностей не зависит от выбора системы координат в n -мерном пространстве. Указание: рассмотрите якобиан преобразования $y = xQ$ и воспользуйтесь правилом замены переменных в кратных интегралах.

2.5.1. Предположим, что для некоторой функции $\varphi(v)$, для любых неотрицательных λ_1, λ_2 таких, что $\lambda_1 + \lambda_2 = 1$, и любых v_1, v_2 из выпуклой области R выполняется неравенство

$$\lambda_1\varphi(v_1) + \lambda_2\varphi(v_2) \leq \varphi(\lambda_1v_1 + \lambda_2v_2). \quad (*)$$

Покажите, что тогда для любых неотрицательных $\alpha_1, \alpha_2, \alpha_3$ таких, что $\alpha_1 + \alpha_2 + \alpha_3 = 1$, и любых v'_1, v'_2, v'_3 из R выполняется неравенство

$$\alpha_1\varphi(v'_1) + \alpha_2\varphi(v'_2) + \alpha_3\varphi(v'_3) \leq \varphi(\alpha_1v'_1 + \alpha_2v'_2 + \alpha_3v'_3). \quad (**)$$

Если это так, то в определении выпуклой функции (2.4.1) достаточно положить $k = 2$. Указание: положите $v'_1 = v_1, v'_2 = v_2, v'_3$ — произвольный элемент из R и $\alpha_1 = (1 - \alpha_3)\lambda_1, \alpha_2 = (1 - \alpha_3)\lambda_2$ и два раза примените неравенство (*) к левой части неравенства (**).

2.5.2. Пусть $\varphi_1(v), \varphi_2(v)$ — выпуклые вверх (вниз) функции в некоторой выпуклой области R . Покажите, что при любых $c_1 > 0, c_2 > 0$ функция $\varphi(v) \triangleq c_1\varphi_1(v) + c_2\varphi_2(v)$ является выпуклой вверх (вниз) в той же области R . Покажите, что для любого $c > 0$ функции $c\varphi(v)$ является выпуклой вверх (вниз). Покажите, что для любого $c < 0$ функции $c\varphi(v)$ является выпуклой вниз (вверх).

2.5.3. Пусть $\varphi(v)$ — непрерывная на числовой оси R функция, которая имеет первую и вторую производные для всех $v \in R$. Известно, что необходимым и достаточным условием выпуклости вверх функции $\varphi(v)$ на R является условие $\varphi''(v) \leq 0$ для всех $v \in R$. Покажите, что функция $-v \log v$ — выпуклая вверх на $R = [0, \infty)$. Покажите, что функция $\exp(\omega v)$ — выпуклая вниз на $R = (-\infty, \infty)$ при всех ω .

2.5.4. Покажите, что энтропия $H(X) \triangleq -\sum_{i=1}^M p_i \log p_i$ является выпуклой вверх функцией в области R всех вероятностных векторов $p = (p_1, \dots, p_M)$. Указание: введите в рассмотрение множество $XY, Y = \{y_1, \dots, y_k\}$ и задайте на нем распределение с помощью соотношения $p(x, y) = p(x|y)p(y)$, $p(y_i) = \lambda_i, \lambda_i \geq 0, \sum \lambda_i = 1$. Покажите, что для доказательства выпуклости достаточно доказать неравенство $H(X|Y) \leq H(X)$.

2.5.5. Аналогичным образом покажите, что относительная энтропия $H_0(X) \triangleq -\int f(x) \log f(x) dx$ является выпуклой вверх функцией в области R всех функций плотности вероятностей.

2.5.6. Покажите, что условная энтропия дискретного ансамбля $H(X|y)$ при каждом фиксированном сообщении $y \in Y$ является выпуклой вверх функцией в области R всех условных распределений, т. е. вероятностных векторов $p_y = (p(x_1|y), \dots, p(x_M|y))$. Покажите, что отсюда вытекает выпуклость вверх условной энтропии $H(X|Y)$ в области R при фиксированном распределении вероятностей $p(y), y \in Y$.

2.5.7. Пусть $\varphi(v)$ — выпуклая вверх функция в области $R \subseteq X, X$ — числовая ось. Пусть $X^* = \{x_1, \dots, x_M\}$ — некоторое множество из M элементов из R , на котором задано распределение вероятностей p_1, \dots, p_M . Обозначим через $\varphi(X^*)$ математическое ожидание случайной величины $\varphi(X^*)$ на множе-

стве X^* , обозначим также через $\bar{X}^* \triangleq \sum p_i x_i$ среднее значение элемента в этом множестве. Тогда

$$\overline{\varphi(X^*)} \leq \varphi(\bar{X}^*).$$

Выведите это неравенство (так называемое неравенство Йенсена) из определения выпуклой вверх функции. Покажите, что для выпуклой вниз функции знак неравенства заменяется на противоположный. Пользуясь неравенством Йенсена, покажите, что для любого множества X^* и любого распределения вероятностей p_1, \dots, p_M имеют место неравенства $\bar{X}^a \geq (\bar{X})^a$ при $a \gg 1$ и обратное неравенство при $a \ll 1$.

2.5.8. Пусть X — гауссовская с. в. с математическим ожиданием m и дисперсией D .

а) Покажите, что $M(X - m)^4 \leq D^2, MX^4 \leq (D + m^2)^2$.

б) Покажите, что для любого нечетного числа k

$$M(X - m)^k = 0.$$

в) Покажите, что для любого целого $s > 0$ и $k = 2^s$

$$M(X - m)^k \leq D^{2^{s-1}}, \quad MX^k \leq (D + m^2)^{2^{s-1}}.$$

Указание: воспользуйтесь предыдущей задачей.

2.5.9. Пусть $\{XY, p(x, y)\}$ — пара совместно заданных дискретных ансамблей, $p(x, y) = p(x)p(y|x)$ и $\tilde{p}(x|y)$ — произвольное семейство условных распределений вероятностей на X , т. е. $\tilde{p}(x|y) \geq 0$ для всех $x \in X, y \in Y$ и $\sum_X \tilde{p}(x|y) = 1$ для всех $y \in Y$. Пусть p — вектор вероятностей, определяемый распределением вероятностей $p(x), x \in X$. Рассмотрим функцию

$$f(p) \triangleq \sum_X \sum_Y p(x) p(y|x) \log \frac{\tilde{p}(x|y)}{p(x)},$$

определенную на множестве всех вероятностных векторов. Покажите, что на этом множестве $f(p)$ — выпуклая вверх функция. Указание: воспользуйтесь выпуклостью вверх средней взаимной информации $I(X; Y)$ как функции от p .

2.6.1. Рассмотрим на интервале $[-T/2, T/2]$ гармоническую систему ортогональных функций $\left\{ \exp\left(-j2\pi \frac{k}{T} t\right) \right\}, k = 0, \pm 1, \pm 2, \dots$. Покажите, что для произвольного стационарного случайного процесса $X(t)$ с корреляционной функцией $K(t_1, t_2)$ при любом действительном f

$$\lim_{\substack{T \rightarrow \infty \\ k \rightarrow \infty \\ k/T \rightarrow f}} \int_{-T/2}^{T/2} K(t_1, t_2) \exp\left(-j2\pi \frac{k}{T} t_1\right) dt_1 = N(f) \exp(-j2\pi f t_2),$$

где $N(f)$ — спектральная плотность мощности процесса $X(t)$. Таким образом, в этой задаче устанавливается, что гармоническая система функций асимптотически является системой собственных функций для корреляционного ядра произвольного стационарного случайного процесса.

2.6.2. В этой задаче последовательно шаг за шагом будет построено доказательство леммы 2.6.1, в которой даются необходимые и достаточные условия существования предела в среднеквадратическом смысле последовательности случайных величин. Оно полезно не только само по себе, но и потому, что использует

важный математический метод представления случайных величин как элементов линейного нормированного пространства, сходимость по норме которого соответствует с. в. сходимости последовательности с. в.

Пусть H — множество с. в. с конечным вторым моментом: $\mathbf{M}X^2 < \infty$ для любой с. в. $X \in H$. Для любых двух с. в. X и Y имеет место неравенство Коши—Буняковского

$$\mathbf{M}|XY| \leq (\mathbf{MX}^2 \cdot \mathbf{MY}^2)^{1/2}$$

(см. также неравенство (2.6.6)).

а) Покажите, что множество всех с. в. с конечным вторым моментом является линейным пространством. Указание: воспользуйтесь неравенством Коши—Буняковского с тем, чтобы получить неравенство $\mathbf{M}(X+Y)^2 \leq (\sqrt{\mathbf{MX}^2} + \sqrt{\mathbf{MY}^2})^2$.

б) В линейном пространстве H можно ввести скалярное произведение $\mathbf{M}XY$ и норму \mathbf{MX}^2 , в результате чего оно становится линейным нормированным пространством. Проверьте, что $\mathbf{M}XY$ удовлетворяет определению скалярного произведения.

в) Рассмотрим теперь свойство непрерывности введенной выше нормы \mathbf{MX}^2 в пространстве H . Пусть последовательность с. в. $X_1, X_2, \dots \in H$ с. к. сходится к с. в. $X \in H$. Покажите, что в этом случае $\lim_{m \rightarrow \infty} \mathbf{MX}_m^2 = \mathbf{MX}^2$. Указание:

покажите с помощью неравенства треугольника, что $\mathbf{MX}^2 - \mathbf{M}(X_m - X)^2 \leq \mathbf{MX}_m^2 \leq \mathbf{MX}^2 + \mathbf{M}(X_m - X)^2$.

г) Из непрерывности нормы выводится непрерывность скалярного произведения. Пусть последовательность с. в. $X_1, X_2, \dots \in H$ с. к. сходится к с. в. $X \in H$ и последовательность $Y_1, Y_2, \dots \in H$ с. к. сходится к с. в. $Y \in H$. Покажите, что в этом случае

$$\lim_{m \rightarrow \infty} \mathbf{MX}_m Y = \mathbf{M}XY \quad \text{и} \quad \lim_{m, n \rightarrow \infty} \mathbf{MX}_m Y_n = \mathbf{M}XY.$$

Указание: покажите вначале, что $\lim_{m \rightarrow \infty} \mathbf{M}(X_m - Y)^2 = \mathbf{M}(X - Y)^2$.

д) Теперь легко доказать лемму 2.6.1. Необходимость следует из непрерывности скалярного произведения. Достаточность — из того, что $\lim_{m \rightarrow \infty} \mathbf{MX}_m^2 = \lim_{m, n \rightarrow \infty} \mathbf{MX}_m X_n$, поэтому $\mathbf{M}(X_m - X_n)^2 \rightarrow 0$. Воспользуйтесь этими указаниями, чтобы самостоятельно закончить доказательство.

2.6.3. Ниже будет рассмотрено доказательство первой половины теоремы 2.6.1, а именно — дано обоснование необходимых и достаточных условий существования интеграла в среднеквадратическом смысле от случайного процесса.

Пусть $\varphi(t)$ — произвольная функция на интервале $[T_1, T_2]$, для которой существует и ограничен интеграл

$$\int_{T_1}^{T_2} \int_{T_1}^{T_2} K(t_1, t_2) \varphi(t_1) \varphi(t_2) dt_1 dt_2, \quad (*)$$

где $K(t_1, t_2)$ — функция корреляции случайного процесса $X(t)$, $\mathbf{MX}^2 < \infty$, для каждого $t \in [T_1, T_2]$. Пусть $T_1 = t_{0n}, t_{1n}, \dots, t_{nn} = T_2$ — некоторое разбиение интервала $[T_1, T_2]$ и ξ_n — с. в., определяемая соотношением

$$\xi_n = \sum_{k=0}^n X_{tk} \varphi(t_k) \Delta t_{nk}; \quad \Delta t_{nk} = t_{n, k+1} - t_{n, k}.$$

а) Покажите, что $\mathbf{M}\xi_n^2 < \infty$. Указание: воспользуйтесь предыдущей задачей.

б) Покажите, что необходимым и достаточным условием существования с. к. предела последовательности с. в. ξ_n при $n \rightarrow \infty$ и $\max_k \Delta t_{nk} \rightarrow 0$ является существование обычного предела последовательности интегральных сумм

$$\sum_{k=1}^m \sum_{l=1}^n K(t_{mk}, t_{nl}) \varphi(t_{mk}) \varphi(t_{nl}) \Delta t_{mk} \Delta t_{nl},$$

т. е. существование и ограниченность интеграла (*). Указание: воспользуйтесь леммой 2.6.1.

2.6.4. Пусть $X(t)$ — случайный процесс с функцией корреляции $K(t_1, t_2)$. Предположим, что наблюдаются n с. в. X_{t_1}, \dots, X_{t_n} , соответствующие моментам времени t_1, \dots, t_n , и требуется построить наилучшую в смысле минимума среднеквадратической ошибки линейную оценку значения процесса в момент времени t . Другими словами, требуется подобрать неслучайные величины c_1, \dots, c_n так, чтобы величина $\mathbf{M}(X_t - \hat{X}_t)^2$ была минимальной, где

$$\hat{X}_t \triangleq \sum_{i=1}^n c_i X_{t_i}. \quad (**)$$

Эту задачу можно представить на языке линейного пространства H случайных величин со скалярным произведением $\mathbf{M}XY$.

Случайные величины X_{t_1}, \dots, X_{t_n} и все их линейные комбинации, т. е. суммы вида (**), образуют подпространство H_n пространства H . С. в. X_t является элементом H и в общем случае может не принадлежать подпространству H_n .

Так как $\mathbf{M}(X_t - \hat{X}_t)^2$ представляет собой квадрат расстояния между элементами X_t, \hat{X}_t в метрике пространства H , то наилучшая линейная оценка есть такая с. в. $\hat{X}_t \in H_n$, которая находится на ближайшем расстоянии от с. в. X_t . Для того, чтобы найти точку в H_n , ближайшую к точке $X_t \in H$, необходимо опустить перпендикуляр из X_t на H_n . Таким образом, для наилучшей линейной оценки \hat{X}_t с. в. $X_t - \hat{X}_t$ должна быть ортогональна всем с. в. из H_n .

Покажите, что коэффициенты c_1, \dots, c_n , определяющие оптимальную линейную оценку с. в. X_t , получаются в результате решения следующей системы линейных уравнений

$$K(t, t_i) = \sum_{j=1}^n c_j K(t_i, t_j), \quad i = 1, \dots, n.$$

2.6.5. В этой задаче приведены основные моменты доказательства теоремы 2.6.3 о разложении случайных процессов в ряд Карунена—Лозва. Доказательство этой теоремы опирается на следующее утверждение (теорему Мерсера). Пусть $K(t, s)$, $0 < t, s < T$, — непрерывное ограниченное неотрицательно определенное ядро и $\{\lambda_i\}$, $\{\varphi_i(t)\}$ — система собственных чисел и собственных функций этого ядра. Тогда $\sum_{i=1}^{\infty} \lambda_i \varphi_i(t) \varphi_i(s)$ представляет собой ряд, который сходится к $K(t, s)$ абсолютно и непрерывно для всех $t, s \in [0, T]$.

Пусть $X(t)$ — случайный процесс с нулевым математическим ожиданием и корреляционной функцией $K(t, s)$, которая удовлетворяет условиям теоремы Мерсера. Тогда с. в. $X \triangleq \int X(f) \varphi(f) dt$, где $\varphi(f)$ — произвольная функция из $L_2[0, T]$, существует в среднеквадратическом смысле (см. теорему 2.6.1).

Рассмотрим с. в. $\xi_n \triangleq \sum_{i=1}^n X_i \varphi_i(t)$, где $\{\varphi_i(t)\}$ — собственные функции $K(t, s)$ и $X_i = \int X(t) \varphi_i(t) dt$, $t \in [0, T]$.

а) Покажите, что $\mathbf{M}X_i = 0$, $i = 1, 2, \dots$, и

$$\mathbf{M}X_i X_j = \begin{cases} \lambda_i, & \text{если } i = j, \\ 0, & \text{если } i \neq j. \end{cases}$$

б) Покажите, что для всех $t \in [0, T]$

$$\mathbf{M}X_i X(t) = \lambda_i \varphi_i(t), \quad i = 1, 2, \dots$$

в) Пусть $\varepsilon_n(t) \triangleq \mathbf{M}[X(t) - \xi_n(t)]^2$. Покажите, что для всех $t \in [0, T]$

$$\lim_{n \rightarrow \infty} \varepsilon_n(t) = 0.$$

Указание: воспользуйтесь тем, что согласно теореме Мерсера $\lim_{n \rightarrow \infty} \sum_{i=1}^n \varphi_i^2(t) \lambda_i = K(t, t) = \mathbf{M}X^2(t)$ и ряд сходится абсолютно и равномерно для всех $t \in [0, T]$.

2.7.1. Пусть $X(t)$ и $Z(t)$ — два статистически независимых гауссовских стационарных случайных процесса и $Y(t) = X(t) + Z(t)$. Пусть $\{\hat{X}_k\}$, $\{Z_k\}$ и $\{Y_k\}$ — коэффициенты разложения на интервале $[-T/2, T/2]$ случайных процессов $X(t)$, $Z(t)$ и $Y(t)$ по гармонической системе ортонормированных функций $\left\{ \frac{1}{\sqrt{T}} \exp\left(-i2\pi \frac{k}{T} t\right) \right\}$, $k = 0, \pm 1, \pm 2, \dots$, $\sigma_{X_k}^2$, $\sigma_{Y_k}^2$, $\sigma_{Z_k}^2$ — дисперсии коэффициентов X_k , Y_k , Z_k соответственно.

а) Является ли выражение

$$\frac{1}{2} \sum_{k=-\infty}^{\infty} \log \frac{\sigma_{X_k}^2 + \sigma_{Z_k}^2}{\sigma_{Y_k}^2}$$

средней взаимной информацией между случайными процессами $X(t)$ и $Z(t)$, рассматриваемыми на интервале $[-T/2, T/2]$?

б) Покажите, что

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \sum_{k=-\infty}^{\infty} \log \left(1 + \frac{\sigma_{X_k}^2}{\sigma_{Z_k}^2} \right) = \int_{-\infty}^{\infty} \frac{1}{2} \log \left(1 + \frac{N_X(f)}{N_Z(f)} \right) df,$$

где $N_X(f)$ и $N_Z(f)$ — спектральные плотности мощности процессов $X(t)$ и $Z(t)$ соответственно. Указание: воспользуйтесь результатом задачи 2.6.1.

2.8.1. Рассмотрим следующие ограничения: $(1 - x_1 - x_2)^2 \geq 0$, $x_1 \geq 0$, $x_2 \geq 0$. Докажите, что в точке $\mathbf{x}^* = (1/2, 1/2)$ $D_\sigma(\mathbf{x}^*) \neq \tilde{D}(\mathbf{x}^*)$ ни для какого $\sigma > 0$. Докажите также, что ограничения $x_1 + x_2 \leq 1$, $x_1 \geq 0$, $x_2 \geq 0$ определяют то же самое допустимое множество векторов.

2.8.2. Ограничения $x_i \geq 0$, $i = \overline{1, m}$, и $\sum_{i=1}^m x_i = 1$ определяют множество вероятностных векторов. В этом случае $\psi_g(\mathbf{x})$ есть вектор, который содержит 1 на j -й позиции и нули на остальных, а $\psi_h(\mathbf{x})$ есть вектор из всех единиц. По-

кажите, что в этом случае существует $\sigma > 0$ такое, что $D_\sigma(\mathbf{x}) = \tilde{D}(\mathbf{x})$ для любого вектора \mathbf{x} , удовлетворяющего указанным выше ограничениям.

2.8.3. Пусть $\{p_i\}$, $\{q_i\}$, $i = \overline{1, m}$, — два распределения вероятностей на множестве целых чисел $\{1, \dots, m\}$. Величина

$$\mathcal{H}(p, q) \triangleq \sum_{i=1}^m p_i \log \frac{p_i}{q_i}$$

называется *энтропией распределения p относительно распределения q* . Предположим, что распределение q зафиксировано, а требуется найти распределение p , максимизирующее $\mathcal{H}(p, q)$.

а) Покажите, что $\mathcal{H}(p, q) \geq 0$, причем равенство имеет место, только когда $p_i = q_i$ для всех $i = \overline{1, m}$.

б) Найдите необходимые условия максимума. Обратите внимание на то, что в этом случае необходимые условия не являются достаточными.

в) Покажите, что максимум достигается на вырожденном распределении $p_i^* = 1$ для некоторого i . Укажите, для какого i .

Количество информации между сообщениями и ансамблями было введено К. Шенноном [15]. Им же были установлены основные свойства средней взаимной информации и относительной энтропии. Средняя взаимная информация между случайными векторами и случайными процессами была найдена И. М. Гельфандом и А. М. Яглом [5]. А. Н. Колмогоров [9] предложил общий теоретико-вероятностный подход к введению количества информации и наметил программу строгого математического обоснования. Им [10, 11] также были предложены новые логические основания теории информации, основанные на понятии сложности последовательности, которые позволяют ввести понятие «информация» и «энтропия» без ссылок на вероятности. Разработке общего теоретико-вероятностного подхода к обоснованию теории информации посвящены работы Р. Л. Добрушиной [7], М. С. Пинскера [13].

Многие результаты, относящиеся к исследованию свойств средней взаимной информации, можно найти в книге Галлагера [4]. С основами нелинейного программирования и теоремой Куна—Таккера можно познакомиться по книге У. И. Зангвила [8].

- Б е л л м а н (Bellman R.). *Introduction to Matrix Analysis* — New York: McGraw-Hill, 1960. [Русский перевод: Б е л л м а н Р. Введение в теорию матриц. — М.: Наука, 1969.]
- В е н т ц е л ь Е. С. Теория вероятностей. — М.: Наука, 1964.
- В у л и х Б. З. Введение в функциональный анализ. — М.: Наука, 1967.
- Г а л л а г е р (Gallager R.) *Information Theory and Reliable Communication*. New York: Wiley, 1968. [Русский перевод: Г а л л а г е р Р. Теория информации и надежная связь. — М.: Советское радио, 1974.]
- Г е л ь ф а н д И. М., Я г л о м А. М. О вычислении количества информации о случайной функции, содержащейся в другой такой же функции. — Успехи матем. наук, 1957, т. 12, № 1.
- Д а в е н п о р т, Р у т (Davenport V. B., Root W. L.) — *Random Signals and Noise* — New York: McGraw-Hill, 1958. [Русский перевод: Д а в е н п о р т В. Б., Р у т В. Л. Введение в теорию случайных сигналов и шумов. М.: ИЛ, 1960.]
- Д о б р у ш и н Р. Л. Математические вопросы шенноновской теории оптимального кодирования информации. — Проблемы передачи информации, 1961, т. 10, 63—107.

8. Зангвилл (Zangwill W. I.) Nonlinear Programming, a Unified Approach. [Русский перевод: Зангвилл У. И. Нелинейное программирование. Единый подход. — М.: Советское радио, 1973.]
9. Колмогоров А. Н. Теория передачи информации. — В сб.: Сессия АН СССР по научн. пробл. автоматизации произв. Пленарные заседания. — М.: Изд. АН СССР, 1956.
10. Колмогоров А. Н. Три подхода к определению понятия количества информации. — Проблемы передачи информации, 1965, т. 1, № 1.
11. Колмогоров А. Н. К логическим основам теории информации и теории вероятностей. — Проблемы передачи информации, 1969, т. 5, № 3.
12. Крамер, Лидбеттер (Cramer H., Leadbetter M. R.) Stationary and Related Stochastic Processes, 1967. [Русский перевод: Крамер Г., Лидбеттер М. Стационарные случайные процессы. — М.: Мир, 1969.]
13. Пинскер М. С. Информация и информационная устойчивость случайных величин и процессов. — Проблемы передачи информации, 1960, т. 7.
14. Феллер (Feller W.) An Introduction to Probability Theory and its Applications. — New York: Wiley, 1966. [Русский перевод: Феллер В., Введение в теорию вероятностей и ее приложение, т. I. — М.: Мир, 1967.]
15. Шеннон (Shannon C. E.), 1948, A Mathematical Theory of Communications. Bell. Syst. Tech. J., 1948, 27, 379—423 (Part I), 623—656 (Part II). [Русский перевод: Шеннон К. Математическая теория связи. В сб.: Работы по теории информации и кибернетике. — М.: ИЛ, 1963.]

Глава 3

КОДИРОВАНИЕ В ДИСКРЕТНЫХ КАНАЛАХ

В этой и следующей главах будет рассматриваться кодирование при передаче сообщений по каналам связи, подверженным воздействию шума. В любом реальном канале связи, который используется для передачи сообщений, всегда в той или иной степени существует шум. В результате воздействия шума на приемной стороне никогда не может быть вынесено абсолютно достоверное решение о том, какое сообщение передавалось по каналу. Наличие такой неопределенности приводит к существованию ненулевой вероятности ошибочной передачи сообщения. Если не принимать соответствующих мер для защиты передаваемых сообщений, то эта вероятность может оказаться весьма большой.

Защита сообщений от влияния шума канала реализуется с помощью специальных методов кодирования. Легко понять, что одним из возможных методов защиты является увеличение энергии, затрачиваемой на передачу каждого сообщения, например, с помощью многократного повторения передачи одного и того же сообщения. Однако при таком методе защиты время передачи одного сообщения становится весьма большим и, следовательно, скорость передачи — весьма низкой. Возможен и другой метод увеличения энергии, основанный на увеличении мощности передатчика. Однако зачастую на практике мощность передатчика не может быть увеличена в силу различных технических ограничений.

Повторение сообщений является тривиальным методом кодирования. Оказывается, что имеются нетривиальные методы кодирования, которые позволяют осуществлять передачу сообщений со сколь угодно высокой достоверностью и относительно высокой скоростью. Основной задачей, которую мы будем решать в этой и следующей главе, является определение по заданной статистической модели канала величины наибольшей скорости, при которой возможна передача сообщений с произвольно малой вероятностью ошибки.

§ 3.1. Классификация каналов связи

На рис. 3.1.1 приведена структурная схема системы связи. Всякая система связи использует некоторый канал связи. Физически канал представляет собой среду, в которой распространяются

сигналы, соответствующие передаваемым сообщениям. Например, это меняющиеся во времени значения напряжения или тока, если канал образован парой проводов, или меняющаяся во времени напряженность электромагнитного поля в случае радиоканалов. Однако часто в канал включают не только физическую среду, но и некоторые устройства, сопряженные с входом и выходом физического канала. Например, это могут быть антенные устройства, выходные цепи передатчиков и входные цепи приемников. В зависимости от этого получаются различные модели реальных каналов связи.

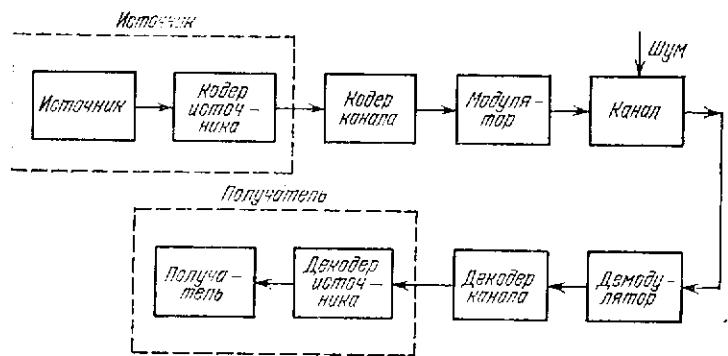


Рис. 3.1.1. Структурная схема системы связи.

Шум, действующий в канале, имеет такую же физическую природу, что и сигналы, и, как обычно предполагается в статистической теории связи, никогда не известен точно наблюдателю, находящемуся на приемной стороне системы связи. Поэтому наблюдатель всегда находится перед проблемой определения того, что же было передано по каналу.

Назначение кодера и декодера источника (см. рис. 3.1.1), а также их основные характеристики были подробно рассмотрены в первой главе. Здесь мы будем предполагать, что кодер источника выбран достаточно хорошо, поэтому можно считать, что символы, появляющиеся на его выходе, независимы и равновероятны. Таким образом, пару «источник — кодер источника» можно рассматривать как новый источник дискретных сообщений. Аналогично пару «декодер источника — получатель» мы будем рассматривать как получателя сообщений в системе передачи. Роль остальных блоков системы связи сводится к тому, что обеспечить максимально надежную передачу последовательности независимых равновероятных символов. Основную роль в решении этой задачи играет пара «кодер — декодер» канала.

Роль пары «модулятор — демодулятор» можно пояснить следующим образом. Предположим, что задано множество символов

на выходе кодера канала. Эти символы могут передаваться с помощью различных сигналов. Например, если символов всего два, то они могут быть переданы либо двумя значениями амплитуды несущих колебаний, либо двумя значениями частоты, либо двумя значениями длительности колебаний при фиксированной частоте и амплитуде и т. д.

Форма сигналов влияет на результатирующее действие шумов. Например, при одинаковом источнике шума частота ошибок при использовании второго метода передачи может быть меньшей, чем при использовании первого.

Устройство, сопоставляющее каждому символу или группе символов на выходе кодера соответствующий входной сигнал канала, называется *модулятором*. Устройство, выполняющее обратное преобразование, называется *демодулятором*. Задачей конструктора системы связи является построение таких пар «кодер — декодер», «модулятор — демодулятор», которые наиболее эффективно уменьшают влияние шумов.

Для осуществления этой задачи необходимо совместное проектирование указанных пар. Однако довольно часто встречается ситуация, когда проектировщик системы не имеет возможности выбирать способ модуляции и демодуляции. Такая ситуация имеет место, например, когда пользователю предоставляется канал вместе с модулятором и демодулятором, а возможно выбирать только метод кодирования и декодирования. В этом случае каналом для проектировщика системы связи является та часть на рис. 3.1.1, которая находится между выходом кодера канала и входом декодера канала. Такой канал называют *дискретным*. Действие шума проявляется в том, что символ на выходе кодера может не совпадать с соответствующим ему символом на входе декодера.

Общее проектирование кодера и модулятора, а также декодера и демодулятора является весьма сложной задачей. В этой главе мы всегда будем предполагать, что модулятор — демодулятор и, следовательно, система используемых сигналов выбрана, а задача состоит в выборе пары кодер — декодер. Несколько более общее рассмотрение будет проведено в следующей главе.

Для теории информации физическая природа сигналов и шумов является несущественной. Так же, как и при кодировании источников, мы будем рассматривать сигналы на входе и выходе канала как элементы некоторых абстрактных множеств (алфавитов). В предыдущем изложении мы различали дискретные и непрерывные источники в зависимости от выбора множества сообщений. Аналогичная классификация имеет место и для каналов.

Определение 3.1.1. Канал называют *дискретным по входу (выходу)*, если множество входных (выходных) сигналов конечно.

З а м е ч а н и е. Иногда дискретным называют такой канал, в котором эти множества, или одно из них, счетно. В настоящем изложении каналы со счетными бесконечными алфавитами не встречаются.

О п р е д е л е н и е 3.1.2. Канал называют *непрерывным по входу (выходу)*, если множество входных (выходных) сигналов несчетно.

О п р е д е л е н и е 3.1.3. Канал называют каналом, *дискретным по входу и непрерывным по выходу (полунепрерывным)*, если множество входных сигналов конечно, а множество выходных сигналов несчетно.

Обычно множество входных сигналов канала будет обозначаться через X , а некоторый элемент этого множества — через x . Аналогичные обозначения Y и y используются для обозначения выходных сигналов канала.

Говоря о непрерывных или полунепрерывных каналах, мы в дальнейшем будем предполагать, что их вероятностное описание может быть дано в терминах функций плотностей вероятностей. Такое ограничение не является существенным, оно принято только в целях упрощения изложения.

О п р е д е л е н и е 3.1.4. Канал называют *каналом с дискретным временем*, если сигналы на его входе и выходе представляют собой конечные или бесконечные последовательности с элементами из алфавитов X и Y соответственно. Дискретный по входу и выходу канал с дискретным временем мы будем называть *дискретным каналом*.

О п р е д е л е н и е 3.1.5. Канал называют *каналом с непрерывным временем*, если сигналы на его входе и выходе представляют собой действительные функции времени. Непрерывный по входу и выходу канал с непрерывным временем мы будем называть *непрерывным каналом*.

В этой главе рассматриваются только дискретные каналы. Непрерывные каналы (с дискретным и непрерывным временем) будут рассмотрены в следующей главе.

Для полного задания канала необходимо задать статистическое описание процесса передачи. Как уже отмечалось, наличие шума может привести к тому, что один и тот же входной сигнал канала может перейти в различные выходные сигналы. С математической точки зрения такие переходы могут описываться с помощью распределений вероятностей. В случае дискретного канала переходы входных сигналов в выходные задаются условными вероятностями $p(y|x)$, $x \in X$, $y \in Y$, получения на выходе сигнала y , если на входе был сигнал x .

В дальнейшем X и Y будут рассматриваться как множества сигналов на входе и выходе канала, которые появляются в некоторый фиксированный момент времени. Поэтому условные вероят-

ности $\{p(y|x)\}$ будут описывать только процесс однократной передачи (передачи одного сигнала) в этот фиксированный момент времени. Однако по каналу никогда не передается один-единственный сигнал, а передается, как правило, достаточно длинная последовательность сигналов. Поэтому задание только одномерных условных вероятностей или условных плотностей вероятностей в общем случае не описывает процесс передачи полностью.

Мы будем говорить, что дискретный канал задан, если для любых целых чисел n и j и любых последовательностей

$$(x^{(j)}, x^{(j+1)}, \dots, x^{(n+j-1)}), (y^{(j)}, y^{(j+1)}, \dots, y^{(n+j-1)})$$

с элементами из дискретных множеств X и Y соответственно заданы условные (или переходные) вероятности $p(y^{(j)}, y^{(j+1)}, \dots, y^{(n+j-1)} | x^{(j)}, x^{(j+1)}, \dots, x^{(n+j-1)})$ получения на выходе канала последовательности $(y^{(j)}, y^{(j+1)}, \dots, y^{(n+j-1)})$, если на входе канала была последовательность $(x^{(j)}, x^{(j+1)}, \dots, x^{(n+j-1)})$.

О п р е д е л е н и е 3.1.6. Дискретный канал называют *каналом без памяти*, если для любых n и j , а также для любых последовательностей $(x^{(j)}, \dots, x^{(n+j-1)})$ и $(y^{(j)}, \dots, y^{(n+j-1)})$ имеют место равенства

$$p(y^{(j)}, \dots, y^{(n+j-1)} | x^{(j)}, \dots, x^{(n+j-1)}) = \prod_{i=j}^{n+j-1} p_i(y^{(i)} | x^{(i)}), \quad (3.1.1)$$

где $p_i(y|x)$ — вероятность для момента времени i получения на выходе канала сигнала y , если на входе был сигнал x .

Название «без памяти» подчеркивает тот факт, что если выполняются соотношения (3.1.1), то при очередной передаче канал как бы не помнит результатов предыдущих передач.

О п р е д е л е н и е 3.1.7. Будем говорить, что дискретный канал без памяти удовлетворяет *условию стационарности*, если для любых i , j , $x \in X$, $y \in Y$

$$p_j(y|x) = p_i(y|x). \quad (3.1.2)$$

Другими словами, статистические характеристики процесса передачи последовательностей сигналов по стационарному каналу без памяти не зависят от момента начала передачи и сохраняются постоянными на протяжении всего времени передачи.

Из определения 3.1.7 следует, что для задания дискретного канала без памяти, удовлетворяющего условию стационарности, достаточно задать лишь одномерные переходные вероятности. В дальнейшем мы всегда будем предполагать, если это не оговорено особо, что дискретные каналы без памяти удовлетворяют условию стационарности. При этом дискретный канал без памяти мы иногда будем обозначать как $\{XY, p(y|x)\}$, где X , Y — входной и выходной алфавиты и $p(y|x)$, $x \in X$, $y \in Y$, — переходные вероятности канала.

Ниже мы будем всюду предполагать, что зафиксирован момент времени j , в который начинается передача по каналу связи. Для простоты будем полагать $j = 1$. Из приведенных определений следует, что в общем случае для задания дискретного канала, по которому сообщения передаются, начиная с момента $j = 1$, необходимо задать переходные вероятности $p(y|x)$ для всех $n = 1, 2, \dots$ и всех последовательностей $\mathbf{x} \triangleq (x^{(1)}, \dots, x^{(n)}) \in X^n$ и $\mathbf{y} \triangleq (y^{(1)}, \dots, y^{(n)}) \in Y^n$. Мы будем предполагать, что переходные вероятности удовлетворяют следующим условиям согласованности:

$$\sum_{y^{n-k+1} \dots y^n} p(y^{(1)}, \dots, y^{(n)} | x^{(1)}, \dots, x^{(n)}) = p(y^{(k)} | x^{(1)}, \dots, x^{(k)}),$$

$$n = 1, 2, \dots, k = 1, \dots, n - 1.$$

Каналы, удовлетворяющие условиям согласованности, называются каналами без предвосхищения. В таких каналах вероятность появления выходного сигнала в некоторый момент времени не зависит от сигналов, которые появляются на входе канала в последующие моменты времени.

Отметим, что распределение вероятностей на входе канала не входит в описание канала, поскольку входное распределение определяется устройствами на входе канала (источником, кодером источника и кодером канала), но не самим каналом. Однако, если некоторое входное распределение, скажем $p(\mathbf{x})$, задано, то оно вместе с условными вероятностями $p(y|x)$ задает совместное распределение вероятностей на парах $(\mathbf{x}, \mathbf{y}) \in X^n Y^n$

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y}|\mathbf{x}) \cdot p(\mathbf{x}) \quad (3.1.3)$$

и распределение вероятностей на выходных последовательностях канала

$$p(\mathbf{y}) = \sum_{X^n} p(\mathbf{x}) p(\mathbf{y}|\mathbf{x}). \quad (3.1.4)$$

§ 3.2. Постановка задачи кодирования в дискретном канале

Как отмечалось в предыдущем параграфе, назначение кодера и декодера канала состоит в том, чтобы уменьшить влияние шумов в канале и обеспечить надежную связь между источником и получателем сообщений. В следующем примере рассматривается один из методов повышения надежности связи.

Пример 3.2.1. Пусть множества входных и выходных сигналов дискретного канала без памяти состоят из двух элементов $\{0, 1\}$ и пусть $p(0|1) = p(1|0) = p$. Такой канал называется двоичным симметричным и полностью определяется

заданием величины p . Действительно, если \mathbf{x}, \mathbf{y} — последовательности длины n из нулей и единиц на входе и выходе канала, то

$$p(\mathbf{y}|\mathbf{x}) = p^t (1-p)^{n-t},$$

где t — количество позиций, на которых последовательности \mathbf{x} и \mathbf{y} различаются, другими словами, t — количество ошибок при передаче \mathbf{x} и получении \mathbf{y} .

Предположим, что $p < 0,5$ и требуется передать одно из двух сообщений z_1 или z_2 . Можно было бы закодировать эти сообщения так: $z_1 \rightarrow 0$, $z_2 \rightarrow 1$. Однако при этом вероятность неправильного приема сообщения равнялась бы p .

Рассмотрим другой метод кодирования (передачу с помощью повторений): если надо передать z_1 , то по каналу передается последовательность из n нулей, если же надо передать z_2 , то по каналу передается последовательность из n единиц. Приемник работает по следующему правилу: если в принятой последовательности количество нулей больше количества единиц, считается, что передавалось z_1 , в противном случае считается, что передавалось z_2 .

Очевидно, что ошибка декодирования возникает всякий раз, когда при передаче последовательности длины n число ошибок t превосходит или равно $n/2$. Так как в рассматриваемом канале вероятность ошибочного приема сигнала равна p и не зависит от того, какой сигнал, 0 или 1, передавался, то вероятность λ неправильного приема сообщения можно определить следующим образом:

$$\lambda \triangleq \Pr \left(t \geq \frac{n}{2} \right) = \sum_{t \geq n/2} C_n^t p^t (1-p)^{n-t}. \quad (3.2.1)$$

Так как математическое ожидание числа ошибок в последовательности длины n равно $np < n/2$, то в силу закона больших чисел λ стремится к нулю при возрастании n .

Таким образом, мы видим, что вероятность неправильной передачи сообщений по каналу может быть сделана сколь угодно малой, если это сообщение передается посредством достаточно большого числа повторений одного и того же входного сигнала. Время передачи при таком методе кодирования пропорционально числу повторений. Поэтому, чтобы вероятность неправильного приема была достаточно малой, необходимо иметь достаточно большое время передачи. При этом скорость передачи, т. е. количество информации, передаваемое в единицу времени, будет стремиться к нулю, так как за все время передачи будет передано одно из двух сообщений или не более 1 бита информации.

В этой главе мы хотим показать, что произвольно малая вероятность ошибки может быть также достигнута и при скоростях передачи, отличных от нуля, за счет усложнения методов кодирования и соответственно декодирования.

Ниже будет описана общая ситуация, имеющая место при передаче сообщений по дискретному каналу связи.

Определение 3.2.1. Кодом с длиной n и объемом M для канала называется множество из M пар $\{\mathbf{u}_1, A_1; \mathbf{u}_2, A_2; \dots; \mathbf{u}_M, A_M\}$, где $\mathbf{u}_i \in X^n$, $i = 1, \dots, M$, — последовательности длины n , образованные входными сигналами канала и называемые кодовыми словами ($\mathbf{u}_i \neq \mathbf{u}_j$ при $i \neq j$), и $A_i \subseteq Y^n$, $i = 1, 2, \dots, M$, — решающие области, образованные выходными последо-

вательностями канала, причем при $i \neq j$ множества A_i и A_j не пересекаются.

Если задан код, то тем самым задано как множество кодовых слов, так и правило, по которому приемник принимает решение о переданном кодовом слове: если на выходе канала появляется последовательность y и $y \in A_i$, то приемник принимает решение о том, что передавалось слово u_i .

Определение 3.2.2. Скоростью кода (или скоростью передачи) называется величина

$$R \triangleq \frac{1}{n} \log M, \quad (3.2.2)$$

где M — объем и n — длина кода.

Из этого определения следует, что скорость кода представляет собой максимальное количество информации, которое может быть передано с помощью одного сигнала (или символа), так как $\log M$ есть максимальное количество информации, которое может быть передано с помощью одного кодового слова. Это количество информации действительно передается, когда кодовые слова имеют одинаковые вероятности появления. Скорость измеряется в битах на символ. Если скорость кода равна R бит/символ, то с помощью такого кода можно передавать nR двоичных единиц информации за время передачи одного кодового слова (за n единиц времени).

Очевидно, что число кодовых слов не может превышать общего числа последовательностей длины n , образованных символами входного алфавита (входными сигналами), канала. Для дискретных каналов это число равно L^n , где L — число элементов множества входных сигналов. Следовательно, в случае дискретных каналов $R \leq \log L$.

Следует отметить разницу в определениях скорости кода канала и скорости кода источника. В случае кода источника скорость определяется как отношение логарифма числа кодовых слов к длине отрезков кодируемых сообщений. В случае кода канала скорость определяется как отношение того же числа к длине кодовых слов (к длине кодирующих последовательностей).

Очевидно, что код длины n , имеющий скорость R , имеет объем $M = 2^{nR}$. Такой код в дальнейшем будем обозначать символом $G(n, R)$.

Пример 3.2.2. Предположим, что двоичный источник без памяти имеет энтропию $H(X) < 1$. Как было показано в гл. 1, при кодировании сообщений такого источника можно достичь скорости, близкой к $H(X)$. Это означает, что при появлении на входе кодера источника n двоичных символов, где n достаточно велико, на выходе кодера появляется примерно $nH(X)$ двоичных символов, что меньше, чем n . Если теперь рассматривать последовательности длины $nH(X)$ как входные сообщения для кодера двоичного канала, осуществляющего кодирование со скоростью $R < 1$, то длина кодовых слов будет равна $n \cdot \frac{H(X)}{R}$,

что больше, чем $nH(X)$. Таким образом, кодирование источника понижает длину последовательностей сообщений, а кодирование в канале ее увеличивает. В связи с этим кодирование источника иногда называют устранением избыточности, а кодирование в канале — введением избыточности. Последовательное применение этих двух операций в большинстве случаев увеличивает эффективность передачи по сравнению с непосредственной передачей сообщений источника без какого-либо кодирования.

Если заданы некоторый канал и код, то мы можем определить вероятность ошибки декодирования данного кода при передаче по данному каналу. Пусть передается слово u_i некоторого фиксированного кода $G(n, R)$, тогда ошибка декодирования возникает в случае, когда последовательность на выходе канала не принадлежит решающей области A_i . Обозначая через λ_i вероятность ошибки декодирования при условии передачи слова u_i , получим

$$\lambda_i = \sum_{y \in \bar{A}_i} p(y|u_i), \quad (3.2.3)$$

где \bar{A}_i — дополнение к множеству A_i .

В качестве количественной меры надежности передачи с помощью кода $G(n, R)$ мы будем использовать две величины. Первая — максимальная вероятность ошибки

$$\Lambda \triangleq \max \{\lambda_1, \dots, \lambda_M\}. \quad (3.2.4)$$

Вторая — средняя вероятность ошибки

$$\lambda \triangleq \sum_{i=1}^M \lambda_i p(u_i), \quad (3.2.5)$$

где $p(u_i)$ — вероятность передачи i -го кодового слова.

Так как распределение вероятностей $p(u_i)$ характеризует источник сообщений и никак не связано ни с каналом, ни с кодом, то средняя вероятность ошибки декодирования часто определяется следующим образом:

$$\lambda \triangleq \frac{1}{M} \sum_{i=1}^M \lambda_i. \quad (3.2.6)$$

Выражение (3.2.6) совпадает с (3.2.5) в случае оптимального кодирования источника, когда $p(u_i) = 1/M$, $i = 1, \dots, M$.

Определение 3.2.3. Пропускной способностью канала с дискретным временем называется максимальное число C такое, что для любого сколь угодно малого δ , $\delta > 0$, и для любого R , $R < C$, существует код $G(n, R)$ такой, что максимальная вероятность ошибки удовлетворяет неравенству

$$\Lambda < \delta. \quad (3.2.7)$$

C — это верхняя грань скоростей кодов, для которых выполняется (3.2.7), поэтому передача с произвольно малой

вероятностью ошибки при скоростях $R > C$ невозможна и, следовательно, для любого $R > C$ существует положительное число δ' такое, что $\Lambda \geq \delta'$ для любого n и любого кода $G(n, R)$.

В последующей части этой главы мы будем заниматься вычислением пропускной способности различных дискретных каналов. Для того чтобы доказать, что некоторое число C является пропускной способностью канала, необходимо доказать два утверждения:

1) при любом $R < C$ и любом положительном δ существует код длины n , скорость которого равна R и максимальная вероятность ошибки удовлетворяет неравенству $\Lambda < \delta$ (прямая теорема кодирования);

2) для всякого $R > C$ найдется положительное число δ' такое, что $\Lambda \geq \delta'$ для любого n и любого кода $G(n, R)$ (обратная теорема кодирования).

Пропускная способность канала была определена относительно максимальной вероятности ошибки Λ . Очевидно, что, если для некоторого кода максимальная вероятность ошибки не превосходит δ , то и средняя вероятность ошибки для этого кода также не превосходит δ . Следующая лемма устанавливает, что в определенном смысле верно и обратное утверждение. Поэтому пропускную способность можно определять как максимальную скорость, при которой средняя вероятность ошибки не превосходит δ .

Л е м м а 3.2.1. Пусть существует код объема M с вероятностью ошибки λ , определенной соотношением (3.2.6), тогда существует код объема $M/2$, максимальная вероятность ошибки которого удовлетворяет неравенству $\Lambda \leq 2\lambda$.

Д о к а з а т е л ь с т в о. Предположим, что в коде объема M кодовые слова u_1, \dots, u_M упорядочены по невозрастанию вероятности ошибки, т. е. $\lambda_i \geq \lambda_j$ при $i < j$. Имеет место следующая цепочка неравенств:

$$\lambda = \frac{1}{M} \sum_{i=1}^M \lambda_i \geq \frac{1}{M} \sum_{i=1}^{M/2} \lambda_i \geq \frac{1}{2} \lambda_{M/2}. \quad (3.2.8)$$

В силу упорядоченности вероятностей ошибок из (3.2.8) вытекает, что

$$\lambda_j \leq \lambda_{M/2} \leq 2\lambda \quad (3.2.9)$$

для всех $j > M/2$. Тогда код объема $M/2$, образованный словами $u_{M/2+1}, \dots, u_M$, имеет максимальную вероятность ошибки, не превышающую 2λ . Лемма доказана.

Из леммы 3.2.1 следует, что, если при любом $R' < C$ существует код $G(n, R')$ такой, что

$$\lambda < \delta/2, \quad (3.2.10)$$

где δ — произвольное положительное число, то при любом $R < C$ существует код $G(n, R)$, для которого $\Lambda < \delta$.

Действительно, скорость кода, построенного в доказательстве леммы 3.2.1, равна $R = R' - 1/n$. Поэтому для любого $R < C$ найдется такое, быть может большое, значение n , что $R' < C$ и, следовательно, существует код, для средней вероятности ошибки которого выполняется неравенство (3.2.10). По лемме 3.2.1 подкод объема $M/2$ этого кода имеет максимальную вероятность ошибки $\Lambda < \delta$ и скорость R .

§ 3.3. Неравенство Фано

В этом параграфе будет рассмотрено основное неравенство, с помощью которого доказываются обратные теоремы кодирования для различных каналов.

Пусть задан дискретный ансамбль $\{UW, p(u, w)\}$, где $U = \{u_1, u_2, \dots, u_M\}$, $W = \{w_1, w_2, \dots, w_L\}$. Обозначим через E событие, состоящее в появлении пары (u_i, w_j) , $i \neq j$. Это событие будем называть «ошибкой». Положим

$$\lambda_j \triangleq \Pr(E | w_j) = \sum_{i: i \neq j} p(u_i | w_j) = 1 - p(u_j | w_j), \quad (3.3.1)$$

$$\lambda \triangleq \Pr(E) = \sum_{i, j: i \neq j} p(u_i, w_j) = \sum_{j=1}^L \lambda_j p(w_j), \quad (3.3.2)$$

где

$$p(w_j) = \sum_{i=1}^M p(u_i, w_j), \quad p(u_i | w_j) = p(u_i, w_j)/p(w_j).$$

Величину λ_j будем называть условной вероятностью ошибки при фиксированном $w_j \in W$, а величину λ — средней вероятностью ошибки.

Рассмотрим множество $E = \{E, \bar{E}\}$, состоящее из двух событий E и \bar{E} , где \bar{E} — событие, дополнительное к E , оно наступает при появлении любой пары (u_i, w_j) , для которой $i = j$. На множестве E для каждого $w_j \in W$ определено условное распределение вероятностей $\{\lambda_j, 1 - \lambda_j\}$. Это распределение совместно с безусловным распределением $p(w_j)$, $j = 1, \dots, L$, задает ансамбль EW , для которого

$$H(E | w_j) \triangleq -\lambda_j \log \lambda_j - (1 - \lambda_j) \log (1 - \lambda_j) = h(\lambda_j), \quad (3.3.3)$$

где $h(p) = -p \log p - (1 - p) \log (1 - p)$, и

$$H(E | W) := \sum_{j=1}^L H(E | w_j) p(w_j). \quad (3.3.4)$$

Безусловное распределение вероятностей на E есть $\{\lambda, 1 - \lambda\}$. При этом

$$H(E) = h(\lambda). \quad (3.3.5)$$

В следующей теореме устанавливается связь между условной энтропией $H(U|W)$ и вероятностью ошибки λ .

Теорема 3.3.1 (неравенство Фано). Для любого дискретного ансамбля $\{UW, p(u, w)\}, |U| = M$, справедливо неравенство

$$H(U|W) \leq h(\lambda) + \lambda \log M. \quad (3.3.6)$$

Доказательство. Рассмотрим условную энтропию $H(U|w_j)$. При $j < M$ имеем

$$\begin{aligned} H(U|w_j) &\triangleq - \sum_{i=1}^M p(u_i|w_j) \log p(u_i|w_j) = \\ &= -p(u_j|w_j) \log p(u_j|w_j) - [1 - p(u_j|w_j)] \log [1 - p(u_j|w_j)] - \\ &- \sum_{i:i \neq j} p(u_i|w_j) \log p(u_i|w_j) + [1 - p(u_j|w_j)] \log [1 - p(u_j|w_j)] = \\ &= H(E|w_j) - \lambda_j \sum_{i:i \neq j} \frac{p(u_i|w_j)}{\lambda_j} \log \frac{p(u_i|w_j)}{\lambda_j}, \end{aligned} \quad (3.3.7)$$

где последнее равенство следует из (3.3.1) и (3.3.3). Из соотношения (3.3.1) следует также, что

$$\sum_{i:i \neq j} \frac{p(u_i|w_j)}{\lambda_j} = 1. \quad (3.3.8)$$

Поэтому второе слагаемое в последнем выражении в (3.3.7) представляет умноженную на λ_j энтропию ансамбля, состоящего из $(M - 1)$ сообщений, вероятности которых указаны как слагаемые в сумме (3.3.8). Если эту энтропию оценить сверху величиной $\log M$, то

$$H(U|w_j) \leq H(E|w_j) + \lambda_j \log M = h(\lambda_j) + \lambda_j \log M. \quad (3.3.9)$$

При $j > M$ (такое j найдется, если $L > M$) имеем

$$H(U|w_j) = - \sum_{i=1}^M p(u_i|w_j) \log p(u_i|w_j) \leq \log M. \quad (3.3.10)$$

Так как при $j > M$ всегда происходит ошибка, то при таких значениях j $\lambda_j = 1$ и $H(E|w_j) = h(\lambda_j) = 0$. Следовательно, из (3.3.10) вытекает, что неравенство (3.3.9) имеет место при всех $j = 1, \dots, L$.

Усредним обе части неравенства (3.3.9) по ансамблю W . Для этого умножим правую и левую части неравенства на $p(w_j)$ и просуммируем по всем j . В результате получим, что

$$H(U|W) \leq H(E|W) + \lambda \log M. \quad (3.3.11)$$

Поскольку условная энтропия $H(E|W)$ не превосходит безусловную $H(E) = h(\lambda)$, то из (3.3.11) следует неравенство (3.3.6). Теорема доказана.

Рассмотрим, как неравенство Фано может применяться для оценки вероятности ошибки декодирования в дискретном канале связи. Пусть задан дискретный канал, т. е. заданы множества входных X и выходных Y сигналов, а также при всех $n = 1, 2, \dots$ заданы условные вероятности $p(y|x)$, $y \in Y^n$, $x \in X^n$. Предположим, что для передачи по каналу используется код $G(n, R) = \{u_1, A_1; \dots; u_M, A_M\}$ длины n и объема $M = 2^{nR}$.

Обозначим через W множество решений $\{w_1, \dots, w_M, w_{M+1}\}$, которые принимает приемник о передаваемых кодовых словах. Решение есть w_j , $j \neq M + 1$, если выходная последовательность канала принадлежит области A_j ; решение есть w_{M+1} , если выходная последовательность канала не принадлежит ни одной решашющей области A_j , $j = 1, \dots, M$. Пусть $U = \{u_1, \dots, u_M\}$ — множество кодовых слов и $p(u_i)$ — вероятность появления слова u_i на входе канала. Тем самым определен ансамбль $\{UW, p(u, w)\}$, элементами которого являются пары (u, w) — (переданное слово, решение), а распределение вероятностей

$$p(u_i, w_j) = p(u_i)p(w_j|u_i), \quad i = 1, \dots, M, \quad j = 1, \dots, M + 1,$$

где

$$p(w_j|u_i) = \sum_{y \in A_j} p(y|u_i), \quad j \neq M + 1,$$

$$p(w_{M+1}|u_i) = 1 - \sum_{j=1}^M p(w_j|u_i).$$

При этом величина λ_j (см. (3.3.1)) представляет собой условную вероятность ошибки декодирования для кода $G(n, R)$ при условии, что в результате декодирования вынесено решение $w_j \in W$, а величина λ (см. (3.3.2)) представляет собой среднюю вероятность ошибки декодирования. Эта средняя вероятность ошибки может быть вычислена также по формуле (3.2.5):

$$\lambda \triangleq \Pr(E) = \sum_{i=1}^M \Pr(E|u_i)p(u_i).$$

Энтропия $H(U|W)$ в рассматриваемом случае представляет собой среднюю условную информацию ансамбля кодовых слов при фиксированном множестве решений. Величину $H(U|W)$

иногда называют ненадежностью передачи с помощью кода $G(n, R)$. Она характеризует количество информации, потерянное при передаче из-за шума в канале. Неравенство Фано (3.3.6) устанавливает связь между ненадежностью передачи и средней вероятностью ошибки декодирования для кода $G(n, R)$.

Неравенство Фано можно интерпретировать следующим образом. Для того чтобы наблюдатель, находящийся в декодере, мог точно установить переданное сообщение, он, во-первых, должен знать, допускает или не допускает ошибку декодер. Среднее количество информации, необходимое для этого, равно $h(\lambda)$. Если наблюдатель знает, что при декодировании произошла ошибка, то ему необходимо дополнительно установить, какое из оставшихся $M - 1$ кодовых слов было действительно передано. Среднее количество информации, необходимое для этого, не превосходит $\log M$. Так как такая необходимость возникает с вероятностью λ , то среднее количество дополнительной информации не превосходит $\lambda \log M$. Неравенство (3.3.6) обосновывает тот интуитивно

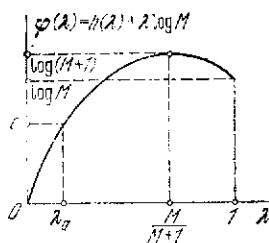


Рис. 3.3.1. График функции $\varphi(\lambda)$.

ясный факт, что потеря информации в канале из-за действия шумов, т. е. величина $H(U|W)$, не превосходит величины $h(\lambda) + \lambda \log M$, которая является верхней оценкой количества информации, необходимого для точного установления переданного сообщения.

Правая часть неравенства Фано является функцией только от λ ; обозначим ее через $\varphi(\lambda)$:

$$\varphi(\lambda) \triangleq h(\lambda) + \lambda \log M. \quad (3.3.12)$$

Заметим, что $\varphi(\lambda) \geq 0$, причем равенство имеет место только при $\lambda = 0$. Функция $\varphi(\lambda)$ является непрерывной на интервале $[0, 1]$. Беря производную по λ , можно убедиться в том, что она монотонно возрастает, если $0 < \lambda < M/M + 1$, убывает, если $M/M + 1 < \lambda < 1$, и имеет максимум в точке $M/M + 1$. График функции $\varphi(\lambda)$ изображен на рис. 3.3.1.

Пусть a — некоторое положительное число, меньшее или равное $\log M$, и λ_a — наименьшее решение уравнения $\varphi(\lambda) = a$. Нетрудно видеть, что следующие два неравенства

$$\begin{aligned} \varphi(\lambda) &\geq a, \\ \lambda &\geq \lambda_a \end{aligned} \quad (3.3.13)$$

равносильны, т. е. первое влечет второе и наоборот.

§ 3.4. Общая обратная теорема кодирования для дискретных каналов

Теперь мы используем неравенство Фано для доказательства обратной теоремы кодирования для широкого класса дискретных каналов.

Рассмотрим некоторый дискретный канал. Пусть задано распределение вероятностей $p(x)$ на входных последовательностях канала $x \in X^n$. Это распределение совместно с условными вероятностями, посредством которых задается канал, определяет ансамбль $\{X^n Y^n, p(y|x)p(x)\}$. Пусть $I(X^n; Y^n)$ — средняя взаимная информация между последовательностями длины n на входе и выходе канала

$$I(X^n; Y^n) = \sum_{x^n} \sum_{y^n} p(x)p(y|x) \log \frac{p(y|x)}{p(y)}, \quad (3.4.1)$$

где

$$p(y) = \sum_{x^n} p(x)p(y|x). \quad (3.4.2)$$

Обозначим через C^* максимальное значение средней взаимной информации в единицу времени между входом и выходом канала

$$C^* \triangleq \sup_{n, \{p(x)\}} \frac{1}{n} I(X^n; Y^n), * \quad (3.4.3)$$

где верхняя грань берется по всем n и всевозможным распределениям $p(x)$, $x \in X^n$, на входных последовательностях длины n . Мы будем называть этот максимум *информационной емкостью дискретного канала*.

Теорема 3.4.1 (обратная теорема кодирования для дискретных каналов). Пусть C^* — информационная емкость дискретного канала и $R = C^* + \varepsilon$, где ε — произвольное положительное число. Тогда существует положительное число δ , зависящее от R , такое, что для всякого кода $G(n, R)$

$$\lambda \geq \delta. \quad (3.4.4)$$

* Точная верхняя грань $\sup f(x)$, $x \in X$, где $f(x)$ — некоторая функция на X , есть наименьшее число f_0 такое, что $f_0 \geq f(x)$ для каждого $x \in X$. Если в множестве X существует такой элемент x_0 , для которого $f_0 = f(x_0)$, то говорят, что верхняя грань достигается на X , и пишут $f(x_0) = \max f(x)$, $x \in X$. Если X — конечное множество, то верхняя грань всегда достигается. В этом случае всегда $\sup f(x) = \max f(x)$. Если X — бесконечное множество, то верхняя грань может не достигаться ни на одном элементе из X . Например, если X — множество натуральных чисел и $f(x) = 1 - \frac{1}{x}$, то $\sup f(x) = 1$, но $f(x) \neq 1$ ни для одного элемента из X . Заметим также, что верхняя грань достигается, если X — замкнутое множество и функция $f(x)$ непрерывна.

Доказательство. Зафиксируем некоторое n и рассмотрим код $G(n, R)$ с $M = 2^{nR}$ кодовыми словами $\{\mathbf{u}_1, \dots, \mathbf{u}_M\}$. Зададим распределение вероятностей на X^n следующим образом. Положим

$$p(\mathbf{x}) = \begin{cases} 1/M & \text{для всех } \mathbf{x} \in \{\mathbf{u}_1, \dots, \mathbf{u}_M\}, \\ 0 & \text{для остальных } \mathbf{x} \in X^n. \end{cases} \quad (3.4.5)$$

Пусть $I(X^n; Y^n)$ — средняя взаимная информация между входом и выходом канала, вычисленная для распределения вероятностей (3.4.5). Тогда $I(X^n; Y^n) = I(U; Y^n)$, где U — ансамбль слов рассматриваемого кода, и из определения информационной емкости следует, что

$$nC^* \geq I(X^n; Y^n) = I(U; Y^n). \quad (3.4.6)$$

Пусть W — ансамбль решений. Этот ансамбль можно рассматривать как результат отображения ансамбля Y^n всех последовательностей на выходе канала в множество решений. Это отображение задается посредством набора решающих областей A_1, \dots, A_M . Каждая последовательность $\mathbf{y} \in Y^n$ однозначно определяет решение $w \in W$ по следующему правилу:

$$w = \begin{cases} w_i, & \text{если } \mathbf{y} \in A_i, i = 1, \dots, M, \\ w_{M+1}, & \text{если } \mathbf{y} \notin \bigcup A_i. \end{cases} \quad (3.4.7)$$

Поскольку информация не возрастает в результате преобразований (см. теорему 2.1.2), то

$$I(U; Y^n) \geq I(U; W). \quad (3.4.8)$$

Так как $H(U|W) = H(U) - I(U; W)$ и согласно (3.4.5) $H(U) = \log M$, то используя неравенство (3.4.6), получим, что

$$H(U|W) = \log M - I(U; W) \geq \log M - nC^*, \quad (3.4.9)$$

или

$$H(U|W) \geq n(R - C^*) = ne. \quad (3.4.10)$$

Теперь можно воспользоваться неравенством Фано, которое, как было показано выше, выполняется для любого кода и для любого распределения вероятностей $p(\mathbf{x})$ на кодовых словах и, в частности, для кода $G(n, R)$ и распределения вероятностей (3.4.5). Обозначим через λ_{0n} наименьший корень уравнения

$$h(\lambda) + \lambda \log M = ne. \quad (3.4.11)$$

Тогда из неравенства Фано и неравенства (3.4.10) следует, что средняя вероятность ошибки λ для кода $G(n, R)$ удовлетворяет неравенству $\lambda \geq \lambda_{0n}$. Легко увидеть, что λ_{0n} стремится к e/R при $n \rightarrow \infty$. Из свойств функции $\varphi(\lambda)$ (см. предыдущий параграф) следует, что при $M \geq 1$ число λ_{0n} остается положительным при

всех n и $\lambda_{0n} \geq \lambda_{01} > 0$. Полагая $\lambda_{01} = \delta$, получим, что $\lambda \geq \delta$ для любого кода $G(n, R)$. Теорема доказана.

Цель дальнейшего изложения состоит в том, чтобы показать, что для широкого класса каналов информационная емкость и, пропускная способность совпадают. Для этого нужно доказать прямую теорему кодирования, в которой утверждается существование кода со скоростью $R < C^*$, обеспечивающего сколь угодно малую заданную наперед вероятность ошибки. Путь, которому мы следуем, состоит в том, что вначале вычисляются величины информационных емкостей ряда достаточно простых каналов, а затем для каждого из рассмотренных каналов доказываются индивидуальные прямые теоремы кодирования. Доказательства для простых каналов обладают необходимой прозрачностью и позволяют наиболее выпукло показать фундаментальные идеи теории информации.

§ 3.5. Информационная емкость дискретных каналов без памяти

В этом параграфе мы покажем, что в случае дискретных каналов без памяти формула (3.4.3), по которой вычисляется информационная емкость, может быть упрощена, а именно можно опустить максимизацию по n и всегда полагать $n = 1$. Кроме того, мы выведем итерационный алгоритм вычисления информационной емкости произвольных дискретных каналов без памяти.

3.5.1. Упрощение формулы (3.4.3). Пусть $p(y|x)$, $x \in X$, $y \in Y$, — переходные вероятности, задающие дискретный канал без памяти. По определению такого канала

$$p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n p(y^{(i)}|x^{(i)}) \quad (3.5.1)$$

для любых последовательностей $\mathbf{x} \in X^n$ и $\mathbf{y} \in Y^n$, $\mathbf{x} = (x^{(1)}, \dots, x^{(n)})$, $\mathbf{y} = (y^{(1)}, \dots, y^{(n)})$.

Ансамбли X^n и Y^n последовательностей сообщений на входе и выходе канала можно представить как произведение ансамблей X_1, \dots, X_n и Y_1, \dots, Y_n соответственно, где X_i и Y_i — ансамбли входных и выходных сигналов канала в момент времени i . Конечно, множества входных и выходных сигналов в каждый момент времени — это множества X и Y соответственно.

Теорема 3.5.1. Информационная емкость C^* дискретного канала без памяти определяется соотношением

$$C^* = \max_{\{p(x)\}} I(X; Y), \quad (3.5.2)$$

где максимум разыскивается по всем распределениям вероятностей $p(x)$ на X .

Доказательство. Пусть $p(\mathbf{x})$ — произвольное распределение вероятностей на входе канала. Рассмотрим среднюю взаимную информацию $I(X^n; Y^n)$ между входным и выходным ансамблями канала, вычисленную в соответствии с распределением $p(\mathbf{x})$. Имеем

$$\begin{aligned} I(X^n; Y^n) &= H(Y^n) - H(Y^n | X^n) = \\ &= H(Y^n) + \sum_{X^n} \sum_{Y^n} p(\mathbf{x}) p(\mathbf{y} | \mathbf{x}) \log \prod_{i=1}^n p(y^{(i)} | x^{(i)}) = \\ &= H(Y^n) + \sum_{i=1}^n \sum_{X_i^n} \sum_{Y_i^n} p(\mathbf{x}) p(\mathbf{y} | \mathbf{x}) \log p(y^{(i)} | x^{(i)}) = \\ &= H(Y^n) + \sum_{i=1}^n \sum_{X_i} \sum_{Y_i} p_i(x^{(i)}) p(y^{(i)} | x^{(i)}) \log p(y^{(i)} | x^{(i)}) = \\ &= H(Y^n) - \sum_{i=1}^n H(Y_i | X_i) \leq \sum_{i=1}^n H(Y_i) - \sum_{i=1}^n H(Y_i | X_i) = \\ &= \sum_{i=1}^n I(X_i; Y_i), \quad (3.5.3) \end{aligned}$$

где $I(X_i; Y_i)$ — средняя взаимная информация между ансамблями X_i, Y_i в момент времени i , определяемая переходными вероятностями канала и распределением вероятностей $p_i(x^{(i)})$ на входе в момент времени i :

$$p_i(x^{(i)}) = \sum_{X_1} \dots \sum_{X_{i-1}} \sum_{X_{i+1}} \dots \sum_{X_n} p(\mathbf{x}), \quad i = 1, \dots, n.$$

В неравенстве (3.5.3) имеет место знак равенства, если ансамбли Y_1, \dots, Y_n статистически независимы, т. е. если

$$p(\mathbf{y}) = \prod_{i=1}^n p_i(y^{(i)}) \quad (3.5.4)$$

для всех $\mathbf{y} \in Y^n$. Нетрудно видеть, что для дискретного канала без памяти это условие выполняется, если выбрать $p(\mathbf{x}) = \prod_{i=1}^n p_i(x^{(i)})$. Действительно, в этом случае

$$\begin{aligned} p(\mathbf{y}) &= \sum_{X^n} p(\mathbf{x}) p(\mathbf{y} | \mathbf{x}) = \\ &= \sum_{X_1} \dots \sum_{X_n} \prod_{i=1}^n p_i(x^{(i)}) p(y^{(i)} | x^{(i)}) = \\ &= \prod_{i=1}^n \sum_{X_i} p_i(x^{(i)}) p(y^{(i)} | x^{(i)}) = \prod_{i=1}^n p_i(y^{(i)}). \quad (3.5.5) \end{aligned}$$

Далее, для произвольного входного распределения $p(\mathbf{x}), \mathbf{x} \in X^n$, из (3.5.3) имеем

$$I(X^n; Y^n) \leq \sum_{i=1}^n \max_{\{p_i(x^{(i)})\}} I(X_i; Y_i) = n \max_{\{p(\mathbf{x})\}} I(X; Y), \quad (3.5.6)$$

где последнее равенство имеет место в силу стационарности канала (независимости переходных вероятностей $p(y^{(i)} | x^{(i)})$ от i).

Таким образом, мы показали, что для произвольного распределения вероятностей $p(\mathbf{x})$ на входе дискретного канала без памяти имеет место неравенство (3.5.6), равенство в котором может быть достигнуто при $p(\mathbf{x}) = \prod_{i=1}^n p_i(x^{(i)})$, т. е. когда входные сигналы статистически независимы и одинаково распределены. Отсюда следует, что

$$\sup_{\{p(\mathbf{x})\}} \frac{1}{n} I(X^n; Y^n) = \max_{\{p(\mathbf{x})\}} I(X; Y). \quad (3.5.7)$$

В правой части соотношения (3.5.7) стоит знак \max , а не \sup , так как множество всех распределений $\{p(\mathbf{x})\}$ на конечном множестве X замкнуто (см. сноску на стр. 167). Теорема доказана.

3.5.2*. Вычисление информационной емкости дискретного канала без памяти. Выше мы показали, что общая формула (3.4.3) для информационной емкости в случае дискретного канала без памяти упрощается и приводится к виду (3.5.2). Заметим, что формула (3.4.3) в общем случае не является вычислимой. Мы называем формулу вычислимой, если существует алгоритм, который позволяет вычислить задаваемое ею значение с произвольной, заданной наперед точностью, за конечное количество шагов. С точки зрения такого определения вычислимости формула (3.5.2) вычислима. Действительно, для вычисления ее значений необходимо найти максимум выпуклой функции от конечного числа аргументов. Алгоритмы для нахождения максимума выпуклой функции с заданной точностью для случая конечного числа аргументов хорошо известны.

В некоторых частных случаях удается получить явное выражение для информационной емкости. Один из таких случаев будет рассмотрен в следующем параграфе, а другой — в задаче 3.5.1. Однако получить простую формулу для информационной емкости произвольного канала без памяти не удается.

В этом пункте будет выведен итерационный алгоритм, позволяющий вычислять информационную емкость произвольного дискретного канала без памяти с произвольной, заданной наперед точностью. Этот алгоритм существенно проще стандартных алго-

ритмов. Упрощение достигается за счет использования некоторых специфических свойств взаимной информации.

Начнем с вывода необходимых и достаточных условий, которым должно удовлетворять распределение вероятностей, максимизирующее правую часть формулы (3.5.2).

Теорема 3.5.2. Пусть фиксирован дискретный канал без памяти с переходными вероятностями $p(y|x)$, $x \in X$, $y \in Y$. Необходимые и достаточные условия, которым должно удовлетворять распределение вероятностей $p(x)$, $x \in X$, на входе канала, максимизирующее среднюю взаимную информацию $I(X; Y)$, суть

$$I(X; Y) \begin{cases} = C \text{ для всех } x, \text{ для которых } p(x) > 0, \\ \leq C \text{ для всех } x, \text{ для которых } p(x) = 0, \end{cases} \quad (3.5.8)$$

где C — постоянная, определяемая из условия $\sum_x p(x) = 1$, и

$$I(X; Y) = \sum_Y p(y|x) \log \frac{p(y|x)}{\sum_x p(x)p(y|x)}.$$

Более того, число C равно информационной емкости канала.
Доказательство. Требуется найти максимум функции

$$I(X; Y) = \sum_X \sum_Y p(x)p(y|x) \log \frac{p(y|x)}{\sum_x p(x)p(y|x)} \quad (3.5.9)$$

по всем распределениям вероятностей $p(x)$, $x \in X$. Так как средняя взаимная информация $I(X; Y)$ является выпуклой вверх функцией (см. § 2.5) и максимизация осуществляется в выпуклой области всех вероятностных векторов $p = (p(x_1), \dots, p(x_L))$, $L = |X|$, то необходимые и достаточные условия, которым должно удовлетворять максимизирующее распределение $p(x)$, $x \in X$, суть условия Куна—Таккера (см. § 2.8), которые можно записать следующим образом:

$$\begin{cases} \frac{\partial I(X; Y)}{\partial p(x)} = \lambda, & \text{если } p(x) > 0, \\ \frac{\partial I(X; Y)}{\partial p(x)} \leq \lambda, & \text{если } p(x) = 0, \end{cases} \quad (3.5.10)$$

где λ — неопределенный множитель Лагранжа, определяемый из условия $\sum_x p(x) = 1$.

Зафиксируем сообщение $\hat{x} \in X$ и найдем частную производную

$$\begin{aligned} \frac{\partial I(X; Y)}{\partial p(\hat{x})} &= \sum_Y p(y|\hat{x}) \log \frac{p(y|\hat{x})}{\sum_{x' \in X} p(x')p(y|x')} - \\ &- \log e \sum_X \sum_Y p(x)p(y|x) \frac{p(y|\hat{x})}{\sum_{x' \in X} p(x')p(y|x')} = I(\hat{x}; Y) - \log e. \end{aligned} \quad (3.5.11)$$

Подставляя (3.5.11) в (3.5.10) и обозначая $\lambda + \log e$ через C , получим условия (3.5.8). Умножая обе части первого соотношения в (3.5.8) на $p(x)$ и суммируя по всем $x \in X$, получим в левой части информационную емкость канала, а в правой — число C . Теорема доказана.

В соответствии с теоремой 3.5.2 распределение вероятностей, максимизирующее среднюю взаимную информацию между входом и выходом канала, должно обладать тем свойством, что для каждого входного сообщения, появляющегося с ненулевой вероятностью, средняя взаимная информация между этим сообщением и выходом канала должна равняться одному и тому же числу. Необходимость этого условия интуитивно понятна. Действительно, если бы некоторое сообщение имело бы большую взаимную информацию с выходом канала, чем другое, то средняя взаимная информация могла быть увеличена за счет более частого использования входных сообщений с большей информацией $I(X; Y)$.

Перейдем теперь к построению итерационного алгоритма для вычисления информационной емкости. Заметим вначале, что функция $I(X; Y)$ может быть записана следующим образом:

$$I(X; Y) = \sum_X \sum_Y p(x)p(y|x) \log \frac{p(x|y)}{p(x)}, \quad (3.5.12)$$

где

$$p(x|y) = \frac{p(x)p(y|x)}{\sum_X p(x)p(y|x)}, \quad x \in X, \quad y \in Y. \quad (3.5.13)$$

Введем в рассмотрение следующую функцию:

$$I(\tilde{P}_{X|Y}; p_X) \triangleq \sum_X \sum_Y p(x)p(y|x) \log \frac{\tilde{P}(x|y)}{p(x)}, \quad (3.5.14)$$

где $\tilde{P}_{X|Y}$ — стохастическая матрица с элементами $\tilde{P}(x|y)$, $x \in X$, $y \in Y$, такими, что $\sum_X \tilde{P}(x|y) = 1$ для всех $y \in Y$, а p_X — вероятностный вектор с элементами $p(x)$, $x \in X$. Нетрудно видеть,

что $I(X; Y) = I(\tilde{P}_{X|Y}; p_X)$, где элементы стохастической матрицы $\tilde{P}_{X|Y}$ определяются соотношениями (3.5.13).

Лемма 3.5.1. Пусть $p(y|x)$, $x \in X$, $y \in Y$ — переходные вероятности канала без памяти. Тогда имеют место следующие утверждения:

1) $\max_{p_X} I(X; Y) = \max_{p_X} \max_{\tilde{P}_{X|Y}} I(\tilde{P}_{X|Y}; p_X)$, где максимум в правой

части разыскивается по всем стохастическим матрицам $\tilde{P}_{X|Y}$ и по всем вероятностным векторам p_X ;

2) при любом фиксированном вероятностном векторе p_X необходимые и достаточные условия того, чтобы матрица $\tilde{P}_{X|Y}$ максимизировала $I(\tilde{P}_{X|Y}; p_X)$, состоят в том, что

$$\tilde{p}(x|y) = \frac{p(x)p(y|x)}{\sum_X p(x)p(y|x)} \triangleq p(x|y), \quad x \in X, \quad y \in Y; \quad (3.5.15)$$

3) при любой фиксированной матрице $\tilde{P}_{X|Y}$ такой, что $\tilde{p}(x|y) > 0$, если $p(y|x) > 0$ необходимые и достаточные условия того, чтобы вероятностный вектор p_X максимизировал $I(\tilde{P}_{X|Y}; p_X)$, состоят в том, что

$$p(x) = \frac{\exp_2 \left(\sum_Y p(y|x) \log \tilde{p}(x|y) \right)}{\sum_X \exp_2 \sum_Y p(y|x) \log \tilde{p}(x|y)}, \quad (3.5.16)$$

где $\exp_2 z \triangleq 2^z$.

Доказательство. Для доказательства утверждения 1 достаточно показать, что для любого фиксированного вероятностного вектора p_X выполняется равенство

$$I(X; Y) = \max_{\tilde{P}_{X|Y}} I(\tilde{P}_{X|Y}; p_X). \quad (3.5.17)$$

Рассмотрим разность

$$\begin{aligned} I(\tilde{P}_{X|Y}; p_X) - I(X; Y) &= \sum_X \sum_Y p(x)p(y|x) \log \frac{\tilde{p}(x|y)}{p(x|y)} \\ &= \sum_X \sum_Y p(y)p(x|y) \log \frac{\tilde{p}(x|y)}{p(x|y)}, \end{aligned} \quad (3.5.18)$$

где

$$p(y) = \sum_X p(x)p(y|x).$$

Применяя неравенство $\ln x \leq x - 1$, из (3.5.18) получим

$$I(\tilde{P}_{X|Y}; p_X) - I(X; Y) \leq 0. \quad (3.5.19)$$

Равенство в неравенстве (3.5.19) имеет место в том и только том случае, когда $\tilde{p}(x|y) = p(x|y)$ при всех $x \in X$ и $y \in Y$. Отсюда следует равенство (3.5.17) и утверждения 1 и 2 леммы.

Для доказательства утверждения 3 воспользуемся теоремой Куна—Таккера. Вначале заметим, что при фиксированных матрицах $P_{X|Y}$ и $\tilde{P}_{X|Y}$ функция $I(\tilde{P}_{X|Y}; p_X)$, определенная на множестве всех вероятностных векторов, является выпуклой вверх (см. задачу 2.5.9). Следовательно, условия Куна—Таккера являются необходимыми и достаточными для того, чтобы распределение вероятностей $p(x)$, $x \in X$, максимизировало $I(\tilde{P}_{X|Y}; p_X)$. Исходя из следствия 2.8.2, эти условия могут быть записаны следующим образом:

$$\begin{aligned} \frac{\partial I(\tilde{P}_{X|Y}; p_X)}{\partial p(x)} &= \lambda \quad \text{при } p(x) > 0, \\ \frac{\partial I(\tilde{P}_{X|Y}; p_X)}{\partial p(x)} &\leq \lambda \quad \text{при } p(x) = 0, \end{aligned} \quad (3.5.20)$$

где λ — множитель Лагранжа, определяемый из условия $\sum_X p(x) = 1$. Выполняя дифференцирование, получим для $x \in X$

$$\frac{\partial I(\tilde{P}_{X|Y}; p_X)}{\partial p(x)} = \sum_Y p(y|x) \log \frac{\tilde{p}(x|y)}{p(x)} - \log e. \quad (3.5.21)$$

Подставляя затем (3.5.21) в (3.5.20) и обозначая $\lambda + \log e$ через λ_1 , будем иметь

$$\sum_Y p(y|x) \log \frac{\tilde{p}(x|y)}{p(x)} \begin{cases} = \lambda_1 & \text{при } p(x) > 0, \\ \leq \lambda_1 & \text{при } p(x) = 0. \end{cases} \quad (3.5.22)$$

Нетрудно проверить, что при

$$p(x) = \frac{\exp_2 \sum_Y p(y|x) \log \tilde{p}(x|y)}{\exp_2 \lambda_1}$$

соотношения (3.5.22) будут выполняться со знаком равенства для всех $x \in X$. Из условия $\sum_X p(x) = 1$ находим, что

$$\exp_2 \lambda_1 = \sum_X \exp_2 \sum_Y p(y|x) \log \tilde{p}(x|y).$$

Лемма доказана.

Введем следующие обозначения:

$$I(\tilde{P}_{X|Y}) \triangleq \max_{p_X} I(\tilde{P}_{X|Y}; p_X), \quad (3.5.23)$$

$$I(p_X) \triangleq \max_{\tilde{P}_{X|Y}} I(\tilde{P}_{X|Y}; p_X). \quad (3.5.24)$$

Из леммы 3.5.1 следует, что $I(p_X) = I(X; Y)$.

Пусть $p^{(0)}(x)$ — произвольное распределение вероятностей на X и $\tilde{P}_{X|Y}^{(0)}$ — соответствующий вероятностный вектор, пусть $P_{X|Y}^{(0)}$ — стохастическая матрица с элементами $p(x|y)$, $x \in X$, $y \in Y$, определяемыми из соотношения (3.5.15), в котором $p(x)$ заменено на $p^{(0)}(x)$. Обозначим через $p_X^{(1)}$ вектор вероятностей, на котором достигается максимум $I(P_{X|Y}^{(0)})$ в формуле (3.5.23), и пусть $P_{X|Y}^{(1)}$ — стохастическая матрица, на которой достигается максимум $I(p_X^{(1)})$ в формуле (3.5.24). В общем случае $k = 1, 2, \dots$ обозначим через $p_X^{(k)}$ вероятностный вектор, на котором достигается $I(P_{X|Y}^{(k-1)})$, и через $P_{X|Y}^{(k)}$ — стохастическую матрицу, на которой достигается $I(p_X^{(k)})$. Из леммы 3.5.1 следует, что вектор $p_X^{(k)}$ и матрица $P_{X|Y}^{(k)}$ могут быть получены итеративно из исходного распределения $p_X^{(0)}$ с помощью соотношений (3.5.15), (3.5.16).

Теперь покажем, что этот итеративный процесс сходится.

Теорема 3.5.3. Для произвольного дискретного канала без памяти, задаваемого переходными вероятностями $p(y|x)$, $x \in X$, $y \in Y$, и произвольного начального распределения вероятностей на входе, задаваемого вектором $p_X^{(0)}$ со строго положительными компонентами, имеет место равенство

$$C^* = \lim_{k \rightarrow \infty} I(p_X^{(k)}), \quad (3.5.25)$$

где предел вычисляется по последовательности $p_X^{(k)}$, $k = 1, 2, \dots$, получаемой итеративно с помощью соотношений (3.5.15), (3.5.16).

Доказательство. Рассмотрим последовательность пар

$$I(p_X^{(0)}), I(P_{X|Y}^{(0)}), I(p_X^{(1)}), I(P_{X|Y}^{(1)}), \dots, I(p_X^{(k)}), I(P_{X|Y}^{(k)}), \dots$$

Так как k -я пара в этой последовательности, $k \geq 1$, получена в результате максимизации либо функции $I(P_{X|Y}^{(k-1)}; p_X)$ по p_X , либо функции $I(\tilde{P}_{X|Y}; p_X^{(k)})$ по $\tilde{P}_{X|Y}$, то элементы этой последовательности не убывают с ростом номера пары k . Таким образом, в силу ограниченности величины $I(p_X^{(k)})$, для любого k , $I(p_X^{(k)}) \leq \log |X|$, и неубывания членов последовательности $I(p_X^{(k)})$, $k = 1, 2, \dots$, предел, написанный в правой части (3.5.25), существует.

Заметим далее, что равенство

$$I(p_X^{(k)}) = I(p_X^{(k+1)}) \quad (3.5.26)$$

может выполняться в том и только том случае, когда $P_{X|Y}^{(k)}$ удовлетворяет условиям (3.5.15) при $p_X = p_X^{(k)}$ и $p_X^{(k+1)}$ удовлетворяет условиям (3.5.16) при $\tilde{P}_{X|Y} = P_{X|Y}^{(k)}$. Если при этом (3.5.15) подставить в (3.5.16), то получим, что $p_X^{(k)}$ будет удовлетворять условиям теоремы 3.5.2 и, следовательно, $I(p_X^{(k)}) = C^*$.

В общем случае, однако, равенство (3.5.26) может не выполняться ни при каких конечных k . Из существования предела в (3.5.25) и непрерывности $I(p_X)$ как функции от p_X следует, что существует предельное распределение $p_X^{(\infty)}$, которое удовлетворяет условиям теоремы 3.5.2. Поэтому предел последовательности (3.5.25) равен информационной емкости канала. Теорема доказана.

Из доказанной теоремы следует, что фиксируя произвольным образом начальное распределение, задаваемое вектором $p_X^{(0)}$ со строго положительными компонентами, для любого $\epsilon > 0$ можно найти такое k , что $C^* \geq I(p_X^{(k)}) \geq C^* - \epsilon$. При этом распределение $p_X^{(k)}$ находится итеративно с помощью формул (3.2.15), (3.2.16). Для завершения построения алгоритма следует указать правило остановки. Это правило основано на следующей лемме.

Лемма 3.5.2. Определим число $D(p_X)$ следующим соотношением:

$$D(p_X) \triangleq \max_{x \in X} \sum_y p(y|x) \log \frac{p(y|x)}{\sum_x p(x)p(y|x)}. \quad (3.5.27)$$

Для произвольного дискретного канала без памяти и произвольного распределения вероятностей p_X на его входе имеет место неравенство

$$C^* \leq D(p_X). \quad (3.5.28)$$

Доказательство. Пусть p_X — вероятностный вектор, на котором достигается информационная емкость C^* канала, и $0 < \theta < 1$. Из выпуклости средней взаимной информации относительно распределений вероятностей на входе канала следует, что

$$\theta I(p_X) + (1-\theta) I(p_X) \leq I(\theta p_X + (1-\theta) p_X)$$

или

$$I(p_X^*) \leq I(p_X) + \frac{I(\theta p_X + (1-\theta)p_X) - I(p_X)}{\theta}, \quad (3.5.29)$$

где использован тот факт, что $I(p_X) = I(X; Y)$. Пусть

$$\Delta(x) \triangleq \theta(p^*(x) - p(x)), \quad (3.5.30)$$

$$\hat{p}_X \triangleq \theta p_X^* + (1-\theta)p_X, \quad (3.5.31)$$

$$I_x(p_X) \triangleq p(x) \sum_Y p(y|x) \log \frac{p(y|x)}{\sum_x p(x) p(y|x)}. \quad (3.5.32)$$

Из этих определений можно записать

$$\frac{I(\theta p_X^* + (1-\theta)p_X) - I(p_X)}{\theta} = \sum_x \frac{I_x(\hat{p}_X) - I_x(p_X)}{\Delta(x)} (\hat{p}_X(x) - p(x)). \quad (3.5.33)$$

Из (3.5.30) и (3.5.31) следует, что $\hat{p}(x) = p(x) + \Delta(x)$ для всех $x \in X$. Подставляя (3.5.33) в (3.5.29) и переходя к пределу при $\theta \rightarrow 0$, получим, что

$$I(p_X^*) \leq I(p_X) + \sum_x \frac{\partial I_x(p_X)}{\partial p(x)} (p^*(x) - p(x)). \quad (3.5.34)$$

Найдем частные производные $\frac{\partial I_x(p_X)}{\partial p(x)}$ для всех $x \in X$:

$$\frac{\partial I_x(p_X)}{\partial p(x)} = \sum_Y p(y/x) \log \frac{p(y/x)}{\sum_x p(x) p(y/x)} - \log e. \quad (3.5.35)$$

Подстановка (3.5.35) в (3.5.34) дает

$$I(p_X^*) \leq \sum_x p^*(x) \sum_Y p(y|x) \log \frac{p(y|x)}{\sum_x p(x) p(y|x)}. \quad (3.5.36)$$

Учитывая теперь, что по условию выбора распределения вероятностей $p^*(x)$ имеет место равенство $I(p_X^*) = C^*$, из (3.5.36) получим

$$C^* \leq \max_{x \in X} \sum_Y p(y|x) \log \frac{p(y|x)}{\sum_x p(x) p(y|x)},$$

что совпадает с (3.5.28). Лемма доказана.

Ниже формулируется алгоритм вычислений, который для любого дискретного канала без памяти и для любого $\varepsilon > 0$ позволяет вычислить значение \hat{C}^* такое, что $C^* - \hat{C}^* \leq \varepsilon$.

Алгоритм. Пусть $L = |X|$, положить $p^{(0)}(x) = \frac{1}{L}$ для всех $x \in X$. Для $k = 1, 2, \dots$ выполнять:

Шаг k . По $\mathbf{p}_X^{(k-1)}$ с помощью (3.5.15) вычислить $p^{(k-1)}(x|y)$ для всех $x \in X, y \in Y$. По $\mathbf{P}_{X|Y}^{(k-1)}$ с помощью (3.5.16) вычислить $p^{(k)}(x)$ для всех $x \in X$. По $\mathbf{p}_X^{(k)}$ с помощью (3.5.27) и (3.5.9) вычислить $D(p_X^{(k)})$ и $I(X; Y) = I(p_X^{(k)})$. Если $D(p_X^{(k)}) - I(p_X^{(k)}) > \varepsilon$, то перейти к шагу $k+1$, если $D(p_X^{(k)}) - I(p_X^{(k)}) \leq \varepsilon$, то положить $\hat{C}^* = I(p_X^{(k)})$. Конец.

Нетрудно видеть, что этот алгоритм заканчивается после некоторого конечного числа шагов и что полученное в результате его работы число \hat{C}^* удовлетворяет неравенству $C^* - \hat{C}^* \leq \varepsilon$. Действительно, если при некотором k

$$D(p_X^{(k)}) - I(p_X^{(k)}) \leq \varepsilon, \quad (3.5.37)$$

то из леммы 3.5.2 и определения информационной емкости в соответствии с формулой (3.5.2) следует, что

$$C^* - \hat{C}^* \leq D(p_X^{(k)}) - I(p_X^{(k)}) \leq \varepsilon.$$

Тот факт, что неравенство (3.5.37) имеет место при некотором конечном k , следует из теоремы 3.5.3 и того, что для распределения, удовлетворяющего условиям теоремы 3.5.2, в соотношении (3.5.28) имеет место знак равенства.

§ 3.6. Симметричные дискретные каналы без памяти

Рассмотрим дискретный канал с множеством входных сигналов $X = \{x_1, \dots, x_L\}$ и множеством выходных сигналов $Y = \{y_1, \dots, y_N\}$. Обозначим через $p(y_j|x_i)$, $i = 1, 2, \dots, L$, $j = 1, 2, \dots, N$, одномерные условные вероятности, задающие канал. В случае канала без памяти распределение вероятностей для последовательностей определяется соотношением (3.5.1).

Одномерные условные вероятности иногда удобно записывать в виде $L \times N$ матрицы, которая называется *матрицей переходных вероятностей канала*:

$$\begin{bmatrix} p(y_1|x_1) & p(y_2|x_1) & \dots & p(y_N|x_1) \\ p(y_1|x_2) & p(y_2|x_2) & \dots & p(y_N|x_2) \\ \dots & \dots & \dots & \dots \\ p(y_1|x_L) & p(y_2|x_L) & \dots & p(y_N|x_L) \end{bmatrix}. \quad (3.6.1)$$

Заметим, что сумма элементов любой строки равна единице.

Определение 3.6.1. Дискретный канал называется *симметричным по входу*, если все строки матрицы переходных вероятностей образованы перестановками элементов первой строки.

Определение 3.6.2. Дискретный канал называется *симметричным по выходу*, если все столбцы матрицы переходных вероятностей образованы перестановками элементов первого столбца.

Определение 3.6.3. Дискретный канал называется *симметричным*, если он симметричен и по входу, и по выходу. Другими словами, и строки, и столбцы матрицы переходных вероятностей симметричного канала образованы перестановками одного и того же набора чисел.

Имеют место следующие два свойства рассматриваемых каналов.

Свойство 1. Если канал симметричен по входу, то условная энтропия $H(Y|X)$ не зависит от распределения вероятностей на входе и равна

$$H(Y|X) = - \sum_{j=1}^N p_j \log p_j, \quad (3.6.2)$$

где p_1, \dots, p_N — элементы первой строки матрицы (3.6.1).

Для доказательства заметим, что $H(Y|X) = M H(Y|x)$, причем для любого $x \in X$

$$H(Y|x) = - \sum_{j=1}^N p(y_j|x) \log p(y_j|x) = - \sum_{j=1}^N p_j \log p_j. \quad (3.6.3)$$

Свойство 2. Если канал симметричен по выходу и распределение вероятностей на его входных сигналах равномерное, то равномерным является распределение вероятностей на его выходных сигналах.

Действительно, пусть $p(x_i) = 1/L$, $i = 1, 2, \dots, L$, тогда

$$p(y_j) = \sum_{i=1}^L p(x_i) p(y_j|x_i) = \frac{1}{L} \sum_{i=1}^L q_i, \quad j = 1, 2, \dots, N, \quad (3.6.4)$$

где q_1, \dots, q_L — элементы первого столбца матрицы (3.6.1). Так как правая часть (3.6.4) не зависит от j , то выходные сигналы канала равновероятны.

Найдем информационную емкость C^* симметричного дискретного канала без памяти.

Как показано в предыдущем параграфе,

$$C^* = \max_{\{p(x)\}} I(X; Y) = \max_{\{p(x)\}} (H(Y) - H(Y|X)), \quad (3.6.5)$$

где максимум разыскивается по всем возможным распределениям вероятностей $p(x)$ на множестве X входных сигналов канала.

Согласно свойству 1 условная энтропия $H(Y|X)$ не зависит от распределения на входе, и для нахождения C^* нужно выбирать входное распределение, максимизирующее энтропию $H(Y)$ выходных сигналов. Известно, что энтропия дискретного ансамбля максимальна в том случае, когда элементы этого ансамбля имеют одинаковые вероятности. Согласно свойству 2 выходные сигналы канала равновероятны, если входное распределение приписывает одинаковые вероятности всем входным сигналам. Таким образом, максимизирующее распределение в (3.6.5) равномерное. В этом случае $H(Y) = \log N$ и

$$C^* = \log N + \sum_{j=1}^N p_j \log p_j. \quad (3.6.6)$$

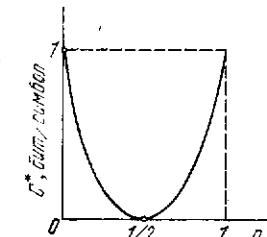


Рис. 3.6.1. Информационная емкость ДСК как функция от вероятности ошибки.

Если канал симметричен по входу, но не симметричен по выходу, может не существовать распределения на входе, при котором выходные сигналы равновероятны. В этом случае

$$C^* < \log N + \sum_{j=1}^N p_j \log p_j. \quad (3.6.7)$$

Пример 3.6.1. Найдем информационную емкость L -ичного симметричного канала, переходные вероятности которого

$$p(y_j|x_i) = \begin{cases} 1-p & \text{при } i=j, \\ \frac{p}{L-1} & \text{при } i \neq j, \quad i, j = 1, \dots, L, \end{cases} \quad (3.6.8)$$

где число p называется вероятностью ошибки. Такой канал, очевидно, является симметричным и, следовательно, применима формула (3.6.6)

$$C^* = \log L + (1-p) \log (1-p) + p \log \frac{p}{L-1} = \log L - h(p) - p \log(L-1), \quad (3.6.9)$$

где $h(p) \triangleq -p \log p - (1-p) \log (1-p)$ — энтропия ансамбля из двух сообщений, одно из которых имеет вероятность p .

При $L = 2$ получается двоичный симметричный канал (ДСК), для которого

$$C^* = 1 - h(p) \quad (\text{бит/символ}). \quad (3.6.10)$$

На рис. 2.5.2 приведен график переходов в таком канале. На рис. 3.6.1 показана зависимость информационной емкости ДСК от вероятности ошибки p . Вероятность $p = 1/2$ соответствует «ковырю» канала и нулевой пропускной способности. При $p > 1/2$ стратегия приемника заключается в том, чтобы при получении y_j решение принимать в пользу x_i , $i \neq j$. Такая стратегия соответствует каналу с вероятностью ошибки $1 - p < 1/2$.

Пример 3.6.2. Найдем информационную емкость двоичного стирающего канала. Предположим, что множество X состоит из двух сигналов, которые мы

условно обозначим через 0 и 1, а множество Y — из трех сигналов 0, 1, \emptyset . Матрица переходных вероятностей имеет вид

$$\begin{bmatrix} 1-p-q & p & q \\ p & 1-p-q & q \end{bmatrix}, \quad (3.6.11)$$

где обозначено

$$\begin{aligned} p(0|0) &= p(1|1) = 1-p-q, \\ p(0|1) &= p(1|0) = p, \quad p(\emptyset|0) = p(\emptyset|1) = q. \end{aligned} \quad (3.6.12)$$

Такой канал называется двоичным симметричным стирающим каналом (ДСтК). Он может быть получен, например, когда приемник может отказаться от принятия решения 0 или 1 и выдать в этом случае стирание « \emptyset ». Очевидно, ДСтК симметричен по входу, но не является симметричным по выходу. Поэтому нельзя воспользоваться выражением (3.6.6). Однако нетрудно убедиться с помощью непосредственной проверки, что в ДСтК энтропия $H(Y)$ максимизируется при равномерном распределении вероятностей на входе (см. также задачу 3.5.1). В этом случае

$$p(y=0) = p(y=1) = \frac{1-q}{2}, \quad p(y=\emptyset) = q \quad (3.6.13)$$

и

$$C^* = -(1-q)\log\frac{1-q}{2} + p\log p + (1-p-q)\log(1-p-q) \quad (\text{бит/символ}). \quad (3.6.14)$$

В частном случае, когда в ДСтК ошибки отсутствуют, т. е. когда $p=0$

$$C^* = 1-q \quad (\text{бит/символ}). \quad (3.6.15)$$

§ 3.7. Дискретные стационарные каналы с аддитивным по модулю L шумом

В этом параграфе будет рассмотрен класс дискретных каналов, каждый из которых обладает свойством симметрии, но не является в общем случае каналом без памяти. Без потери общности можно предположить, что как множество X сигналов на входе, так и множество Y сигналов на выходе канала представляют собой множество, состоящее из L чисел $\{0, 1, \dots, L-1\}$.

Пусть $x \in X$ и $y \in Y$. Суммой $x + y$ по модулю L называется такое число z , $0 \leq z \leq L-1$, что $x + y - z$ нацело делится на L . Это записывается так:

$$z = x + y \bmod L \quad (3.7.1)$$

(например, $3+4=2 \bmod 5$, $2+3=0 \bmod 5$).

Пусть $\mathbf{x} \in X^n$ и $\mathbf{y} \in Y^n$ — две последовательности длины n , элементы которых принадлежат множеству $\{0, 1, \dots, L-1\}$. Сумма двух последовательностей $(\mathbf{x} + \mathbf{y}) \bmod L$ определяется как

их покомпонентная сумма по модулю L . Так, если $\mathbf{x} = (x^{(1)}, \dots, x^{(n)})$, $\mathbf{y} = (y^{(1)}, \dots, y^{(n)})$, $\mathbf{z} = (z^{(1)}, \dots, z^{(n)})$, $\mathbf{z} = \mathbf{x} + \mathbf{y} \bmod L$, то

$$z^{(i)} = x^{(i)} + y^{(i)} \bmod L, \quad i = 1, 2, \dots, n. \quad (3.7.2)$$

Обозначим через Z^n множество всех последовательностей длины n над алфавитом $\{0, 1, \dots, L-1\}$. Рассмотрим дискретный стационарный источник U_Z и обозначим через $\{Z^n, q(\mathbf{z})\}$ ансамбль последовательностей длины n сообщений такого источника, $n = 1, 2, \dots$. Таким образом, для каждого n распределение вероятностей $q(\mathbf{z})$ задается источником U_Z .

Определение 3.7.1. Дискретным стационарным каналом с аддитивным по модулю L шумом (*AL-каналом*) называется канал, переходные вероятности которого определяются соотношением

$$p(\mathbf{y}|\mathbf{x}) = q(\mathbf{z}), \quad (3.7.3)$$

где \mathbf{z} — корень уравнения $\mathbf{z} + \mathbf{x} = \mathbf{y} \bmod L$ и $q(\mathbf{z})$ — распределение, задаваемое для каждого n , $n = 1, 2, \dots$, стационарным источником U_Z . В этом случае U_Z называется источником шума.

Нетрудно видеть, что при каждом n матрица, составленная из переходных вероятностей $p(\mathbf{y}|\mathbf{x})$, $\mathbf{x} \in X^n$, $\mathbf{y} \in Y^n$, *AL*-канала, удовлетворяет определению 3.6.3. Следовательно,

$$\max_{\{p(\mathbf{x})\}} I(X^n; Y^n) = n \log L + \sum_{Z^n} q(\mathbf{z}) \log q(\mathbf{z}) = n \log L - H(Z^n) \quad (3.7.4)$$

и

$$\max_{\{p(\mathbf{x})\}} \frac{1}{n} I(X^n; Y^n) = \log L - H_n(Z), \quad (3.7.5)$$

где $H_n(Z) \triangleq \frac{1}{n} H(Z^n)$.

Из свойств стационарного источника (см. § 1.5) известно, что $H_{n+1}(Z) \leq H_n(Z)$ и $\lim_{n \rightarrow \infty} H_n(Z) = H(Z|Z^\infty)$, где $H(Z|Z^\infty)$ — энтропия на сообщение стационарного источника U_Z . Поэтому правая часть (3.7.5) не убывает с ростом n и информационная емкость *AL*-канала

$$\begin{aligned} C^* &= \sup_{n, \{p(\mathbf{x})\}} \frac{1}{n} I(X^n; Y^n) = \lim_{n \rightarrow \infty} (\log L - H_n(Z)) = \\ &= \log L - H(Z|Z^\infty). \end{aligned} \quad (3.7.6)$$

Предложим теперь, что в *AL*-канале источник шума является эргодическим. В этом случае можно дать более наглядное, чем в теореме 3.4.1, доказательство обратной теоремы кодирования.

Теорема 3.7.1 (обратная теорема кодирования для AL-каналов с эргодическим шумом). Пусть C^* — информационная емкость AL-канала с эргодическим шумом. Тогда для любого $R > C^*$ и для любой последовательности кодов $\{G(n, R)\}$, $n = 1, 2, \dots$,

$$\lim_{n \rightarrow \infty} \Lambda_n = 1, \quad (3.7.7)$$

где Λ_n — максимальная вероятность ошибки для кода $G(n, R)$.

Доказательство. Положим $R = C^* + \varepsilon$, $\varepsilon > 0$ и рассмотрим произвольный код $G(n, R) = \{\mathbf{u}_1, A_1; \dots; \mathbf{u}_M, A_M\}$, $M = 2^{nR}$. Предположим, что кодовые слова упорядочены так, что решающая область A_1 имеет наименьший объем. Пусть $|A|$ есть число элементов в множестве A . Тогда

$$|A_1| \leq \frac{2^{n \log L}}{2^{nR}} = 2^{n(\log L - R)}, \quad (3.7.8)$$

где числитель $2^{n \log L}$ есть общее число выходных последовательностей канала, а знаменатель — количество решающих областей. Из (3.7.6) и (3.7.8) вытекает, что

$$|A_1| \leq 2^{n[H(z|z^\infty) - \varepsilon]}. \quad (3.7.9)$$

Определим множество B_1 следующим образом:

$$B_1 \triangleq \{\mathbf{z}: \mathbf{u}_1 + \mathbf{z} \bmod L \in A_1\}. \quad (3.7.10)$$

Если передается слово \mathbf{u}_1 и шумовая последовательность \mathbf{z} принадлежит B_1 , то происходит правильное декодирование, в противном случае происходит ошибка. Пусть λ_{1n} — вероятность ошибки при передаче \mathbf{u}_1 . Тогда

$$\Lambda \geq \lambda_{1n} = \Pr(\mathbf{z} \in \bar{B}_1 | \mathbf{u}_1) = 1 - \Pr(\mathbf{z} \in B_1), \quad (3.7.11)$$

причем последнее равенство следует из независимости источника шума и передаваемых сообщений.

Число элементов в множестве B_1 равно числу элементов в множестве A_1 и, следовательно, не превосходит правой части неравенства (3.7.9). Из обратной теоремы для случая равномерного кодирования эргодического источника (теорема 1.9.4) следует, что для любого множества B_1 с таким числом элементов вероятность события $\{\mathbf{z} \in B_1\}$ стремится к нулю, когда n стремится к бесконечности. Это означает, что вероятность ошибки λ_{1n} стремится к единице. Теорема доказана.

В основе доказанной теоремы лежат простые соображения о соотношении объема некоторой решающей области и объема высоковероятного множества эргодического источника. Для того чтобы вероятность ошибки могла быть сделана сколь угодно малой, необходимо (но не достаточно), чтобы каждая решающая область

имела объем не меньший, чем объем высоковероятного множества. Тогда общее число кодовых слов могло бы быть не меньше, чем

$$\frac{2^{n \log L}}{2^{n[H(z|z^\infty) + \varepsilon]}} = 2^{n(C^* - \varepsilon)}, \quad (3.7.12)$$

и, следовательно, скорость кода могла бы быть не меньше, чем $C^* - \varepsilon$. К сожалению, это утверждение не является доказательством прямой теоремы кодирования, поскольку дополнительно нужно доказать, что существует такой выбор решающих областей указанного объема, для которого каждая решающая область является некоторым высоковероятным множеством. Доказательство этого будет дано ниже.

§ 3.8. Неравенство Файнштейна

В этом параграфе будет рассмотрено одно из важнейших теоретико-информационных неравенств, с помощью которого могут быть доказаны прямые теоремы кодирования для различных каналов связи, а именно неравенство Файнштейна.

Рассмотрим дискретный канал, задаваемый переходными вероятностями $p(y|x)$, $x \in X^n$, $y \in Y^n$. Пусть $p(x)$, $x \in X^n$, — некоторое распределение вероятностей на входных последовательностях канала и $I(x; y)$ — информация между двумя последовательностями $x \in X^n$ и $y \in Y^n$, вычисленная по распределению $p(x)$:

$$I(x; y) = \log \frac{p(y|x)}{p(y)}, \quad (3.8.1)$$

где $p(y) = \sum_{x^n} p(y|x)p(x)$. Будем обозначать через S некоторое подмножество множества X^n , а через V_τ — множество таких пар $(x, y) \in X^n Y^n$, что взаимная информация $I(x; y)$ каждой пары больше, чем τt , где t — некоторое положительное число:

$$V_\tau \triangleq \{(x, y): I(x; y) > \tau t\}. \quad (3.8.2)$$

Теорема 3.8.1 (неравенство Файнштейна). Пусть τ — произвольное положительное число, S — произвольное подмножество множества X^n и $p(x)$ — произвольное распределение вероятностей на X^n . Тогда для каждого n , $n = 1, 2, \dots$, существует код $G(n, R)$, каждое слово которого принадлежит S , а максимальная вероят-

нность Λ_n ошибки декодирования удовлетворяет неравенству

$$\Lambda_n \leq \frac{1}{\Pr(S)} [2^{-n(\tau-R)} + 1 - \Pr(V_\tau)], \quad (3.8.3)$$

где

и

$$\Pr(S) \triangleq \sum_{x \in S} p(x) \quad (3.8.4)$$

$$\Pr(V_\tau) \triangleq \sum_{(x, y) \in V_\tau} p(y|x)p(x).$$

Доказательство. Для доказательства существования кода, указанного в теореме, мы будем пользоваться методом, который называется методом максимальных кодов. В соответствии с этим методом, содержание которого будет изложено ниже, мы покажем, что для всякого фиксированного $\alpha > 0$ можно построить код $G(n, R)$ с $\Lambda_n = \alpha$, каждое слово которого принадлежит S , а скорость R удовлетворяет неравенству

$$2^{nR} \geq 2^{n\tau} [\Pr(S)\alpha - (1 - \Pr(V_\tau))]. \quad (3.8.5)$$

Очевидно, что это неравенство эквивалентно неравенству (3.8.3) и, следовательно, доказательство существования кода, скорость которого удовлетворяет неравенству (3.8.5), эквивалентно доказательству утверждения теоремы. Далее мы будем предполагать, что $\Lambda_n \Pr(S) > 1 - \Pr(V_\tau)$. В противном случае утверждение теоремы тривиально, так как неравенство (3.8.5) выполняется при $2^{nR} = 1$, а код из одного кодового слова имеет вероятность ошибки декодирования, равную нулю.

Зафиксируем некоторое n и построим код длины n , выбирая кодовые слова из множества S . Для того чтобы определить решающие области, введем в рассмотрение для каждого $x \in X^n$ множество $B(x)$ (см. 3.8.1):

$$B(x) \triangleq \{y: y \in Y^n, I(x, y) > n\tau\}. \quad (3.8.6)$$

Тогда множество V_τ можно определить следующим образом:

$$V_\tau = \{(x, y): x \in X^n, y \in B(x)\}. \quad (3.8.7)$$

Пусть u_1, u_2, \dots, u_M — кодовые слова. Будем считать, что решающие области A_1, A_2, \dots, A_M определяются следующим способом:

$$A_1 \triangleq B(u_1), \quad A_i \triangleq B(u_i) \setminus \bigcup_{j=1}^{i-1} A_j, \quad i = 2, \dots, M, \quad (3.8.8)$$

где \setminus — знак вычитания множеств ($C \setminus D$ есть множество всех элементов, принадлежащих C , но не принадлежащих D).

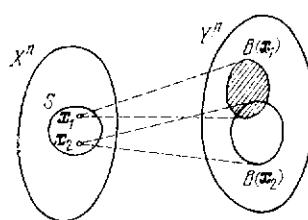


Рис. 3.8.1. К доказательству теоремы 3.8.1.

Предположим, что можно выбрать M кодовых слов, каждое из которых принадлежит множеству S , и M решающих областей, удовлетворяющих условиям (3.8.8), причем условные вероятности ошибок декодирования λ_i удовлетворяют неравенствам

$$\lambda_i \triangleq \Pr(y \notin A_i | u_i) \leq \alpha, \quad i = 1, \dots, M, \quad (3.8.9)$$

где α — произвольное заданное число из интервала $(0, 1]$. Предположим также, что M — это максимальное число, для которого можно осуществить указанный выбор.

Покажем, что $M \geq 1$, т. е. что существует по крайней мере одно кодовое слово из S , удовлетворяющее указанным выше условиям. Для этого предположим противное, а именно что для всех $x \in S$ выполняется неравенство

$$\Pr(y \in B(x) | x) \leq 1 - \alpha. \quad (3.8.10)$$

Тогда должно выполняться следующее неравенство:

$$\sum_{x \in S} \Pr(y \in B(x) | x) p(x) \leq (1 - \alpha) \Pr(S).$$

С другой стороны, из формулы (3.8.7) следует, что

$$\begin{aligned} \Pr(V_\tau) &= \sum_{X^n} \Pr(y \in B(x) | x) p(x) = \\ &= \sum_S \Pr(y \in B(x) | x) p(x) + \sum_{X^n \setminus S} \Pr(y \in B(x) | x) p(x). \end{aligned} \quad (3.8.11)$$

Так как

$$\sum_{X^n \setminus S} \Pr(y \in B(x) | x) p(x) \leq \sum_{X^n \setminus S} p(x) = 1 - \Pr(S), \quad (3.8.12)$$

то

$\Pr(V_\tau) \leq (1 - \alpha) \Pr(S) + 1 - \Pr(S) = 1 - \alpha \Pr(S) \leq 1 - \Lambda_n \Pr(S)$, что противоречит исходному предположению о том, что $\Lambda_n \Pr(S) > 1 - \Pr(V_\tau)$. Таким образом, $M \geq 1$ и хотя бы один шаг при построении кода выполнить можно.

Как было сказано, M есть объем максимального кода. Другими словами, к выбранному коду нельзя добавить ни одного слова так, чтобы оно принадлежало множеству S и не были нарушены условия (3.8.9). Следовательно, для любого $x \in S$ должно выполняться неравенство

$$\Pr(y \in B(x) \setminus \bigcup_{i=1}^M A_i | x) \leq 1 - \alpha. \quad (3.8.13)$$

В противном случае выбор кодовых слов можно было бы продолжить, положив $u_{M+1} = x$ и $A_{M+1} = B(x) \setminus \bigcup_{i=1}^M A_i$; при этом $\lambda_{M+1} \leq \alpha$ и построенный код не был бы максимальным.

Так как для произвольных событий C, D выполняется неравенство $\Pr(C \setminus D) \geq \Pr(C) - \Pr(D)$, то, используя это неравенство для оценки левой части соотношения (3.8.13), получим

$$\Pr(y \in B(x) | x) \leq 1 - \alpha + \Pr\left(y \in \bigcup_{i=1}^M A_i | x\right). \quad (3.8.14)$$

Это неравенство имеет место для всех последовательностей $x \in S$, поэтому левую и правую его части можно осреднить по S . Для этого умножим обе части (3.8.14) на $p(x)$ и просуммируем по $x \in S$. В результате получим

$$\sum_S \Pr(y \in B(x) | x) p(x) \leq (1 - \alpha) \Pr(S) + \Pr\left(x \in S, y \in \bigcup_{i=1}^M A_i\right).$$

Из соотношений (3.8.11) и (3.8.12) следует, что

$$\begin{aligned} \sum_S \Pr(y \in B(x) | x) p(x) &= \Pr(V_\tau) - \sum_{x^n \setminus S} \Pr(y \in B(x) | x) p(x) \geq \\ &\geq \Pr(V_\tau) - 1 + \Pr(S). \end{aligned}$$

Используя это и предыдущее неравенство, можно получить, что

$$\Pr\left(x \in S, y \in \bigcup_{i=1}^M A_i\right) \geq \alpha \Pr(S) - [1 - \Pr(V_\tau)]. \quad (3.8.15)$$

Теперь, чтобы завершить доказательство теоремы, необходимо связать левую часть этого неравенства с числом M — объемом кода. Для этого заметим, что

$$\Pr\left(x \in S, y \in \bigcup_{i=1}^M A_i\right) \leq \Pr\left(y \in \bigcup_{i=1}^M A_i\right) = \sum_{i=1}^M \Pr(y \in A_i). \quad (3.8.16)$$

Оценим вероятность $\Pr(y \in A_i)$ того, что выходная последовательность канала попадет в решающую область A_i . Для любой пары (x, y) , для которой $I(x; y) > n\tau$, имеет место неравенство

$$\log \frac{p(y|x)}{p(y)} > n\tau$$

и, следовательно,

$$p(y|x) > p(y) 2^{n\tau}. \quad (3.8.17)$$

Суммируя обе части последнего неравенства по всем $y \in B(x_i)$, получим следующую цепочку неравенств:

$$1 \geq \Pr(y \in A(x_i) | x_i) > \Pr(y \in B(x_i)) 2^{n\tau} \geq \Pr(y \in A_i) 2^{n\tau}, \quad (3.8.18)$$

где последнее неравенство — следствие того, что A_i — подмножество множества $B(x_i)$. Таким образом,

$$\Pr(y \in A_i) < 2^{-n\tau}. \quad (3.8.19)$$

Учитывая неравенства (3.8.15), (3.8.16), (3.8.19), а также что $M = 2^{nR}$ и $\Lambda_n < \alpha$, получим неравенство (3.8.3). Теорема доказана.

Вернемся теперь к формулировке теоремы и обсудим возможности ее применения. В теореме утверждается существование кода, слова которого принадлежат множеству $S \subseteq X^n$ и максимальная вероятность ошибки декодирования удовлетворяет неравенству (3.8.3). Значение правой части этого неравенства можно варьировать, изменяя величины слагаемых и соотношение между ними за счет выбора параметра τ и распределения вероятностей на входе. Для доказательства прямой теоремы кодирования необходимо установить, можно ли подобрать τ и входное распределение так, чтобы оба слагаемых убывали к нулю при возрастании n .

Покажем, что эта задача сводится к исследованию поведения случайной величины $\frac{1}{n} I(x; y)$ — информации на одно сообщение между входными и выходными последовательностями канала.

Предположим, что $S = X^n$, тогда $\Pr(S) = 1$ при любом распределении на входе. Положим $R = C^* - \varepsilon$, $\varepsilon > 0$, и

$$\tau = C^* - \frac{\varepsilon}{2}, \quad (3.8.20)$$

где C^* — информационная емкость дискретного по времени канала, определяемая соотношением

$$C^* = \sup \frac{1}{n} I(X^n; Y^n), \quad (3.8.21)$$

и верхняя грань разыскивается по всем n и по всем распределениям вероятностей $p(x)$ на входных последовательностях. Тогда, как это следует из (3.8.3), первое слагаемое равно $2^{-n\varepsilon/2}$ и стремится к нулю при возрастании n . Второе слагаемое определяется поведением вероятности $\Pr(\bar{V}_\tau)$, которую можно представить следующим образом:

$$\begin{aligned} \Pr(\bar{V}_\tau) &\triangleq 1 - \Pr(V_\tau) = \Pr\left(\frac{1}{n} I(x; y) < C^* - \frac{\varepsilon}{2}\right) = \\ &= \Pr\left\{\frac{1}{n} I(x; y) < \frac{1}{n} I(X^n; Y^n) - \frac{\varepsilon}{2} + \left(C^* - \frac{1}{n} I(X^n; Y^n)\right)\right\}. \end{aligned} \quad (3.8.22)$$

Предположим, что при каждом n выбирается такое входное распределение $p(x)$, при котором достигается максимум средней

взаимной информации $\frac{1}{n} I(X^n; Y^n)$. Тогда найдется n такое, что

$$C^* - \frac{1}{n} I(X^n; Y^n) < \frac{\epsilon}{4} \quad (3.8.23)$$

и

$$\Pr(\bar{V}_\tau) < \Pr\left\{\frac{1}{n} I(x; y) \geq \frac{1}{n} I(X^n; Y^n) - \frac{\epsilon}{4}\right\}. \quad (3.8.24)$$

Для моделей каналов связи, представляющих интерес, верхняя грань в (3.8.21) достигается при $n \rightarrow \infty$, поэтому неравенство (3.8.23) выполняется для всех достаточно больших n и вопрос о поведении вероятности ошибки сводится к исследованию правой части неравенства (3.8.24). Если вероятность того, что информация на сообщение $\frac{1}{n} I(x; y)$ отличается от среднего количества информации на сообщение $\frac{1}{n} I(X^n; Y^n)$ на величину, большую чем $\epsilon/4$, убывает с ростом n , тогда убывает с ростом n также и максимальная вероятность ошибки.

Ниже мы применим неравенство Файнштейна и указанное выше рассуждение для доказательства прямых теорем кодирования для некоторых каналов. Мы покажем, что в рассматриваемых каналах существуют коды с произвольно малой вероятностью ошибки при условии, что скорость кода не превосходит информационную емкость канала. Фактически вместе с обратной теоремой кодирования это является доказательством того, что информационная емкость и пропускная способность, определенная ранее в § 3.2, совпадают.

§ 3.9. Прямая теорема кодирования для дискретных каналов без памяти

В этом параграфе будет рассмотрен произвольный дискретный канал без памяти. Напомним, что его информационная емкость

$$C^* = \max_{\{p(x)\}} I(X; Y), \quad (3.9.1)$$

где максимум берется по всем распределениям вероятностей $p(x)$ на множестве X входных сигналов канала.

Для дискретного канала без памяти при любом распределении вероятностей $p(x) = \prod_{i=1}^n p(x^{(i)})$, $x \in X^n$,

$$\frac{1}{n} I(x; y) = \frac{1}{n} \sum_{i=1}^n I(x^{(i)}; y^{(i)}), \quad x \in X^n, \quad y \in Y^n, \quad (3.9.2)$$

где

$$I(x^{(i)}; y^{(i)}) = \log \frac{p(y^{(i)} | x^{(i)})}{p(y^{(i)})}, \quad i = 1, 2, \dots, n, \quad (3.9.3)$$

— независимые одинаково распределенные случайные величины, имеющие ограниченные дисперсии. Этого оказывается достаточно, чтобы доказать убывание правой части неравенства (3.8.24) к нулю при увеличении n .

Теорема 3.9.1. Пусть C^* — информационная емкость дискретного канала без памяти. При любом $R < C^*$ и любом положительном δ существует код $G(n, R)$, максимальная вероятность ошибки которого удовлетворяет неравенству $\Lambda_n < \delta$.

Доказательство. Пусть $\epsilon = C^* - R$. Выберем параметр τ в соответствии с (3.8.20), а распределение $p(x)$ таким, которое максимизирует среднюю взаимную информацию в формуле (3.9.1). По теореме 3.8.1 при $S = X^n$ существует код $G(n, R)$ максимальная вероятность ошибки декодирования которого удовлетворяет неравенству

$$\Lambda_n \leq 2^{-n(\tau-R)} + \Pr(\bar{V}_\tau) = 2^{-n\epsilon/2} + \Pr(\bar{V}_\tau), \quad (3.9.4)$$

где $\Pr(\bar{V}_\tau)$ удовлетворяет неравенству (3.8.24).

Поскольку для рассматриваемого канала выполняется соотношение (3.9.2), то

$$I(X^n; Y^n) = nI(X; Y) \quad (3.9.5)$$

и

$$\begin{aligned} \Pr(\bar{V}_\tau) &\leq \Pr\left\{\frac{1}{n} \sum_{i=1}^n I(x^{(i)}; y^{(i)}) - I(X; Y) < -\frac{\epsilon}{4}\right\} \leq \\ &< \Pr\left\{\left|\frac{1}{n} \sum_{i=1}^n I(x^{(i)}; y^{(i)}) - I(X; Y)\right| \geq \frac{\epsilon}{4}\right\}. \end{aligned} \quad (3.9.6)$$

Так как случайные величины $I(x^{(i)}; y^{(i)})$ независимы, одинаково распределены, имеют математическое ожидание $I(X; Y)$ и ограниченную дисперсию, то в силу закона больших чисел правая часть последнего неравенства стремится к нулю при увеличении n к бесконечности. Таким образом, оба слагаемых в (3.9.4) стремятся к нулю, и поэтому найдется такое n и такой код $G(n, R)$, что максимальная вероятность ошибки меньше любого заданного наперед положительного числа δ . Теорема доказана.

Этот результат совместно с обратной теоремой кодирования (теорема 3.4.1) позволяет сформулировать следующее утверждение.

Следствие 3.9.1. Пусть C — пропускная способность дискретного канала без памяти, т. е. такое число, что для каждого $R < C$ существует код $G(n, R)$ с максимальной вероятностью ошибки, меньшей чем заданное наперед произвольное положитель-

ное число, и что для любого $R > C$ не существует кода с таким свойством. Пусть C^* — информационная емкость, определяемая формулой (3.9.1). Тогда

$$C = C^*. \quad (3.9.7)$$

§ 3.10. Прямая теорема кодирования для дискретных стационарных каналов с аддитивным эргодическим шумом

Как было показано в § 3.7, информационная емкость дискретного стационарного канала с аддитивным по модулю L шумом (AL -канала) определяется соотношением

$$C^* = \log L - H(Z|Z^\infty), \quad (3.10.1)$$

где L — объем входного алфавита канала и $H(Z|Z^\infty)$ — энтропия на сообщение источника шума, причем средняя взаимная информация на сообщение $\frac{1}{n} I(X^n; Y^n)$ максимизируется при равномерном распределении вероятностей $p(x)$ на X^n при каждом значении n . Так как AL -канал является симметричным, то при равномерном распределении на входе распределение вероятностей $p(y)$ на выходе Y^n канала также равномерное:

$$p(y) = 2^{-n \log L}, \quad y \in \mathbb{Y}^n, \quad (3.10.2)$$

при этом нетрудно видеть, что

$$I(x, y) = \log \frac{p(y|x)}{p(y)} = n \log L + \log q(z). \quad (3.10.3)$$

Здесь $z = y - x \bmod L$ и $q(z)$ — распределение вероятностей для источника шума.

Теорема 3.10.1. Пусть C^* — информационная емкость дискретного стационарного канала с аддитивным по модулю L шумом. Пусть, кроме того, источник шума в таком канале — эргодический. Тогда при любом $R < C^*$ и любом положительном δ существует код $G(n, R)$, максимальная вероятность ошибки которого удовлетворяет неравенству $\Lambda_n \leq \delta$.

Доказательство. Пусть $\epsilon = C^* - R$. Выберем τ в соответствии с (3.8.20), а распределение $p(x)$ таким, которое максимизирует среднюю взаимную информацию на сообщение, т. е. $p(x)$ — равномерное на X^n распределение вероятностей. По теореме 3.8.1 при $S = X^n$ существует код, максимальная вероятность ошибки которого в рассматриваемом канале удовлетворяет неравенству

$$\Lambda_n \leq 2^{-n(\tau-R)} + \Pr(\bar{V}_\tau) = 2^{-n\epsilon/2} + \Pr(\bar{V}_\tau), \quad (3.10.4)$$

где

$$\begin{aligned} \Pr(\bar{V}_\tau) &= \Pr(I(x; y) \leq n\tau) = \Pr\left(\frac{1}{n} I(x; y) \leq C^* - \frac{\epsilon}{2}\right) = \\ &= \Pr\left(\frac{1}{n} \log q(z) \leq -H(Z|Z^\infty) - \frac{\epsilon}{2}\right) \leq \\ &\leq \Pr\left(\left|\frac{1}{n} I(z) - H(Z|Z^\infty)\right| \geq \frac{\epsilon}{2}\right). \end{aligned} \quad (3.10.5)$$

Теперь из леммы Мак-Миллана (см. § 1.9) следует, что, выбирая n достаточно большим, можно сделать вероятность $\Pr(\bar{V}_\tau)$ меньшей, чем произвольное заданное на перед положительное число. Вместе с (3.10.4) это завершает доказательство теоремы.

Теорема 3.10.1 и обратная теорема кодирования 3.4.1 дают возможность высказать следующее утверждение.

Следствие 3.10.1. Пусть C — пропускная способность AL -канала с эргодическим шумом и пусть C^* — информационная емкость, определяемая формулой (3.10.1). Тогда

$$C = C^*. \quad (3.10.6)$$

Следующий пример показывает, что предположение об эргодичности шума является существенным условием совпадения пропускной способности и информационной емкости AL -канала.

Пример 3.10.1. Рассмотрим AL -канал с неэргодическим источником шума U_Z , представляющим собой комбинацию двух эргодических источников U_1 и U_2 . Структура такого источника показана на рис. 3.10.1.

Имеется переключатель K , который подключает на выход источник U_Z с равными вероятностями либо выход источника U_1 , либо — источника U_2 . Переключение происходит в момент включения источника U_Z , и затем положение переключателя остается неизменным в течение всего времени работы источника. Предположим, что источники U_1 и U_2 имеют различные энтропии на сообщение: $H(Z_1|Z_1^\infty) > H(Z_2|Z_2^\infty)$. Тогда можно показать (см. задачу 3.10.1), что для источника U_Z

$$H(Z|Z^\infty) = \frac{1}{2} (H(Z_1|Z_1^\infty) + H(Z_2|Z_2^\infty)) \quad (3.10.7)$$

и, следовательно,

$$H(Z_2|Z_2^\infty) < H(Z|Z^\infty) < H(Z_1|Z_1^\infty). \quad (3.10.8)$$

Для AL -канала с таким источником шума передача сообщений производится либо в канале с источником шума U_1 , либо с источником шума U_2 в зависимости от положения переключателя K . Из (3.10.8) следует, что информационные емкости рассматриваемых каналов соотносятся так:

$$C_2^* > C^* > C_1^*. \quad (3.10.9)$$

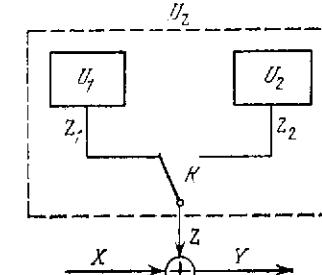


Рис. 3.10.1. AL -канал с аддитивным неэргодическим шумом.

Если $C_1^* < R < C^*$, то максимальная вероятность ошибки может быть сделана сколь угодно малой только при условии, что в канале действует источник шума U_2 . В противном случае ни для одного кода вероятность ошибки не может быть сделана произвольно малой (по теореме 3.7.1 эта вероятность близка к единице при больших n). Так как этот канал появляется с вероятностью $1/2$, то средняя вероятность ошибки для любого кода $G(n, R)$ в неэргодическом канале близка к $1/2$. Поэтому C^* не является пропускной способностью. Можно показать, что пропускная способность $C = \min(C_1^*, C_2^*)$.

Этот результат легко обобщается на случай неэргодического AL -канала, образованного более чем двумя эргодическими компонентами.

§ 3.11. Декодирование для кодов с заданным множеством кодовых слов

Рассмотрим некоторый код для канала $\{X^n Y^n, p(y|x)\}$ и обозначим через u_1, \dots, u_M , $u_i \in X^n$, его кодовые слова. Предположим, что набор кодовых слов фиксирован, а требуется указать наилучший выбор решающих областей A_1, \dots, A_M , $A_i \subseteq Y^n$, при котором средняя или максимальная вероятность ошибки минимизируются. В общем случае решающие области (3.8.8), которые участвуют в доказательстве неравенства Файнштейна, не являются оптимальными в указанном выше смысле.

Пусть $\{p(u_1), \dots, p(u_M)\}$ — распределение вероятностей на множестве кодовых слов, $\sum_{i=1}^M p(u_i) = 1$. Тогда апостериорные вероятности кодовых слов могут быть найдены для каждого $y \in Y^n$ по следующим формулам:

$$p(u_i|y) = \frac{p(y|u_i)p(u_i)}{p(y)}, \quad (3.11.1)$$

где

$$p(y) = \sum_{i=1}^M p(y|u_i)p(u_i) \quad (3.11.2)$$

— безусловное распределение вероятностей на выходе канала. Обозначим через $\lambda(y)$ вероятность ошибки при условии, что на выходе канала появилась последовательность y . Среднюю вероятность ошибки тогда можно представить следующим образом:

$$\lambda = \sum_{y^n} \lambda(y) p(y) = \sum_{i=1}^M \sum_{A_i} \lambda(y) p(y). \quad (3.11.3)$$

Для всех y из A_i ошибка декодирования происходит в том случае, когда переданное слово отличается от u_i , следовательно,

$$\lambda(y) = 1 - p(u_i|y), \quad y \in A_i. \quad (3.11.4)$$

Тогда соотношение (3.11.3) дает

$$\lambda = 1 - \sum_{i=1}^M \sum_{A_i} p(y) p(u_i|y). \quad (3.11.5)$$

3.11. ДЕКОДИРОВАНИЕ

Таким образом, при данном наборе кодовых слов средняя вероятность ошибки минимальна, если

$$A_i = \{y: p(u_i|y) \geq p(u_j|y) \text{ для всех } j = 1, 2, \dots, M\}. \quad (3.11.6)$$

Всякое разбиение множества выходных последовательностей канала на решающие области задает некоторое правило декодирования. Правило декодирования, определяемое разбиением (3.11.6), выбирает при каждом $y \in Y^n$ то кодовое слово, которое имеет наибольшую апостериорную вероятность $p(u_i|y)$. В связи с этим оно называется декодированием по *максимуму апостериорной вероятности* (МАВ-декодирование). Как выше показано, МАВ-декодирование минимизирует среднюю вероятность ошибки.

МАВ-декодирование зависит от априорного распределения вероятностей $p(u_i)$ на множестве кодовых слов. Рассмотрим декодирование по *максимуму правдоподобия* (МП-декодирование), которое по построению не зависит от априорного распределения. МП-декодированием называют такое правило декодирования, которое задается разбиением

$$A_i \triangleq \{y: p(y|u_i) \geq p(y|u_j) \text{ для всех } j = 1, 2, \dots, M\}. \quad (3.11.7)$$

Когда все кодовые слова равновероятны, т. е. $p(u_i) = 1/M$ для всех i , разбиения на решающие области (3.11.6) и (3.11.7) совпадают. Действительно, из (3.11.1) следует, что если u_i максимизирует апостериорную вероятность $p(u_i|y)$, то это же u_i максимизирует также функцию правдоподобия $p(y|u_i)$.

Пример 3.11.1. Рассмотрим МП-декодирование в двоичном симметричном канале (ДСК). Предположим, что в этом канале вероятность ошибки $p < 1/2$. Пусть $x = (x^{(1)}, \dots, x^{(n)})$ и $y = (y^{(1)}, \dots, y^{(n)})$. Тогда

$$p(y|x) = \prod_{i=1}^n p(y^{(i)}|x^{(i)}) = p^t(1-p)^{n-t}, \quad (3.11.8)$$

где t — количество позиций, в которых последовательность x отличается от последовательности y . В случае МП-декодирования y отображается в то слово используемого кода, которому соответствует минимальное значение t .

Количество позиций, в которых последовательность x отличается от последовательности y , называется *расстоянием Хемминга между x и y* . Поэтому МП-декодирование в ДСК отображает выходную последовательность канала в такое кодовое слово, которое находится на наименьшем расстоянии Хемминга от него. В этом случае принято говорить, что декодирование производится по *минимуму расстояния Хемминга*.

Пример 3.11.2. Рассмотрим МП-декодирование в двоичном стирающем канале (ДСТК). Предположим теперь, что в рассматриваемом канале вероятность ошибки равна p и вероятность стирания равна q , причем $q + 2p < 1$. Из стационарности канала следует, что

$$p(y|x) = \prod_{i=1}^n p(y^{(i)}|x^{(i)}) = p^t q^s (1-p-q)^{n-s-t}, \quad (3.11.9)$$

где s — число стираний в последовательности \mathbf{y} и t — количество нестертых позиций, в которых последовательности \mathbf{x} и \mathbf{y} отличаются. Число s определяется только выходом канала и не зависит от того, какое кодовое слово передавалось. Нетрудно проверить, что при фиксированном s и при $q + 2p < 1$ правая часть (3.11.9) монотонно убывает с ростом t . Поэтому МП-декодирование в ДСтК отображает выходную последовательность канала в такое кодовое слово, которому соответствует минимальное значение t . Другими словами, МП-декодирование осуществляется по минимуму расстояния Хемминга на нестертых позициях.

Согласно определению, взаимная информация между входной \mathbf{x} и выходной \mathbf{y} последовательностями канала

$$I(\mathbf{x}; \mathbf{y}) = \log \frac{p(\mathbf{y} | \mathbf{x})}{p(\mathbf{y})}. \quad (3.11.10)$$

Поэтому МП-декодирование эквивалентно выбору такого кодового слова, которое при данной выходной последовательности \mathbf{y} максимизирует величину взаимной информации между ним и \mathbf{y} .

При выводе неравенства Файнштейна строился код, решающие области которого образовывались, исходя из другого принципа. Последовательность \mathbf{y} относилась в решающую область A_i , если информация $I(\mathbf{u}_i; \mathbf{y})$ при данном кодовом слове \mathbf{u}_i была достаточно большой:

$$I(\mathbf{u}_i; \mathbf{y}) > n_t, \quad (3.11.11)$$

а не обязательно максимальной. При этом потребовалась некоторая коррекция решающих областей для того, чтобы сделать их непересекающимися, поскольку могли найтись два кодовых слова \mathbf{u}_i и \mathbf{u}_j , для каждого из которых выполняется условие (3.11.11).

Для такого выбора решающих областей удалось достаточно простыми средствами доказать существование кода, максимальная вероятность ошибки которого могла быть сделана сколь угодно малой при условии, что скорость кода меньше пропускной способности.

В предыдущих параграфах нас не интересовала скорость сходимости вероятности ошибки декодирования к нулю с ростом длины кода. Вместе с тем, с точки зрения практических приложений представляет существенный интерес исследование зависимости величины вероятности ошибки от длины кода. Очевидно, что вероятность ошибки декодирования зависит как от кодовых слов, так и вида решающих областей. Выше мы показали, что в случае равновероятных сообщений наилучший выбор решающих областей соответствует декодированию по максимуму правдоподобия. В следующих разделах будет выведена верхняя и нижняя оценка для вероятности ошибки МП-декодирования в канале без памяти, показывающая ее экспоненциальное убывание с ростом длины кода при всех скоростях, меньших пропускной способности канала.

§ 3.12. Верхняя граница вероятности ошибки декодирования для дискретных каналов без памяти

3.12.1. Метод случайного кодирования. Рассмотрим множество всех кодов длины n и объема $M = 2^{nR}$, где R — скорость кода. Обозначим это множество через \mathfrak{G} и будем допускать, что сюда входят даже заведомо плохие коды, например, содержащие совпадающие слова. Множество \mathfrak{G} , как нетрудно видеть, содержит L^{nM} кодов, где через L обозначен объем входного алфавита канала. Предположим, что на множестве \mathfrak{G} задано некоторое распределение вероятностей, т. е. каждому коду $G_t \in \mathfrak{G}$ приписана вероятность π_t , $\sum_t \pi_t = 1$. Распределение вероятностей на \mathfrak{G} можно

вводить многими способами. Часто удобным оказывается способ задания распределения посредством указания некоторого случайного правила выбора M кодовых слов. Ниже мы более подробно остановимся на описании ансамбля кодов $\{\mathfrak{G}, \pi\}$, а пока будем предполагать, что такой ансамбль определен.

Пусть при использовании кода G_t обеспечивается средняя вероятность ошибки λ_t . Если возможно оценить среднюю по ансамблю кодов вероятность ошибки

$$\bar{\lambda} \triangleq M\lambda_t = \sum_t \pi_t \lambda_t \leq \delta, \quad (3.12.1)$$

то можно утверждать, что в ансамбле кодов найдется хотя бы один код, для которого средняя вероятность ошибки не превосходит δ . Как мы увидим ниже, построение оценки для средней по ансамблю кодов вероятности ошибки существенно проще, чем построение оценки для отдельного кода достаточно большого объема.

3.12.2. Оценка средней по ансамблю кодов вероятности ошибки декодирования для произвольного дискретного канала. Рассмотрим некоторый код $G_t = \{\mathbf{u}_1, A_1; \dots; \mathbf{u}_M, A_M\}$, $M = 2^{nR}$, где A_1, \dots, A_M — решающие области, соответствующие декодированию по максимуму правдоподобия. Введем в рассмотрение среднюю вероятность ошибки λ_t этого кода

$$\lambda_t := \frac{1}{M} \sum_{i=1}^M \sum_{\mathbf{y} \in \bar{A}_i} p(\mathbf{y} | \mathbf{u}_i), \quad (3.12.2)$$

где \bar{A}_i — дополнение множества A_i .

Пусть

$$\Phi_t(\mathbf{y}) \triangleq \begin{cases} 1, & \text{если } \mathbf{y} \in \bar{A}_i, \\ 0, & \text{если } \mathbf{y} \in A_i, \quad i = 1, \dots, M, \end{cases} \quad (3.12.3)$$

т. е. $\varphi_i(y)$ — индикатор ошибки декодирования при передаче слова u_i и получении на выходе канала последовательности y . Функцию $\varphi_i(y)$ можно оценить сверху следующим образом:

$$\varphi_i(y) \leq \left(\frac{\sum_{j:j \neq i} p(y|u_j)^{1/1+\rho}}{p(y|u_i)^{1/1+\rho}} \right)^\rho, \quad \rho > 0. \quad (3.12.4)$$

Справедливость этого неравенства устанавливается очевидным образом. Если $\varphi_i(y) = 0$, то неравенство тривиально, так как правая часть всегда неотрицательна. Если $\varphi_i(y) = 1$, то для одного из слагаемых в числителе $p(y|u_j) \geq p(y|u_i)$, именно для того, для которого j соответствует номеру области A_j , содержащей y . Поэтому правая часть (3.12.4) не меньше единицы для любого $\rho \geq 0$. Объяснить, почему используется именно такая оценка, довольно трудно. Тем не менее она дает хорошие результаты.

Подставляя (3.12.4) в (3.12.2), получим при $\rho \geq 0$

$$\begin{aligned} \lambda_t &= \frac{1}{M} \sum_{i=1}^M \sum_{y^n} p(y|u_i) \varphi_i(y) \leq \\ &\leq \frac{1}{M} \sum_{i=1}^M \sum_{y^n} p(y|u_i)^{1/1+\rho} \left(\sum_{j:j \neq i} p(y|u_j)^{1/1+\rho} \right)^\rho. \end{aligned} \quad (3.12.5)$$

Последнее неравенство выполняется для любого кода, состоящего из кодовых слов $\{u_1, \dots, u_M\}$ и декодируемого по максимуму правдоподобия.

Теперь введем в рассмотрение вероятностный ансамбль кодов и применим метод случайного кодирования. Предположим, что $q(x)$ — некоторое распределение вероятностей на последовательностях $x \in X^n$. Будем считать, что распределение вероятностей на множестве \mathcal{G} всех кодов длины n с $M = 2^{nR}$ словами задается посредством распределения $q(x)$. Если G_t — это код, состоящий из слов $\{u_1, \dots, u_M\}$, то положим

$$\pi_t = \prod_{i=1}^M q(u_i). \quad (3.12.6)$$

Последнее означает, что кодовые слова для кода G_t выбираются из множества X^n независимо в соответствии с распределением $q(x)$.

Средняя по ансамблю кодов вероятность ошибки $\bar{\lambda}$ может быть вычислена по формуле

$$\bar{\lambda} = M\lambda_t = \sum_t \pi_t \lambda_t.$$

Оценивая λ_t с помощью неравенства (3.12.5) и обозначая операцию взятия математического ожидания по ансамблю кодов чертой сверху, получим при $\rho \geq 0$

$$\bar{\lambda} \leq \frac{1}{M} \sum_{i=1}^M \sum_{y^n} \overline{p(y|u_i)^{1/1+\rho}} \left(\sum_{j:j \neq i} \overline{p(y|u_j)^{1/1+\rho}} \right)^\rho, \quad (3.12.7)$$

где, кроме того, использована перестановочность операций суммирования и взятия математического ожидания.

Выражение под чертой представляет собой произведение двух случайных величин, первая из которых зависит только от кодового слова u_i , а вторая — от всех остальных кодовых слов. В ансамбле кодов выбор i -го и j -го кодовых слов производится независимо друг от друга (см. (3.12.6)). Следовательно, статистически независимы и указанные выше две случайные величины под знаком черты. Так как математическое ожидание произведения независимых случайных величин равно произведению математических ожиданий, то

$$\bar{\lambda} \leq \frac{1}{M} \sum_{i=1}^M \sum_{y^n} \overline{p(y|u_i)^{1/1+\rho}} \left(\sum_{j:j \neq i} \overline{p(y|u_j)^{1/1+\rho}} \right)^\rho \quad (3.12.8)$$

при $\rho \geq 0$.

Теперь воспользуемся свойствами выпуклых функций. Пусть $\psi(x)$ — выпуклая вверх функция, тогда для любого набора неотрицательных чисел $\lambda_1, \dots, \lambda_m$, $\sum \lambda_i = 1$, имеет место неравенство

$$\sum_{i=1}^m \lambda_i \psi(x_i) \leq \psi\left(\sum_{i=1}^m \lambda_i x_i\right). \quad (3.12.9)$$

Если $\lambda_1, \dots, \lambda_m$ — вероятности значений x_1, \dots, x_m случайной величины X , то левая часть (3.12.9) — это математическое ожидание $\overline{\psi(x)}$ случайной величины $\psi(x)$, а правая — это $\psi(\bar{x})$. Таким образом, для выпуклых вверх функций справедливо неравенство (см. также задачу 2.5.7)

$$\overline{\psi(x)} \leq \psi(\bar{x}), \quad \bar{x} = \sum_{i=1}^m \lambda_i x_i.$$

Легко видеть, что $\psi(x) = x^\rho$ — выпуклая вверх функция при $x \geq 0$ и $0 < \rho \leq 1$. Поэтому

$$\overline{x^\rho} \leq (\bar{x})^\rho, \quad 0 < \rho \leq 1.$$

Ограничиваая величину ρ сверху единицей, из (3.12.8) получим

$$\bar{\lambda} \leq \frac{1}{M} \sum_{i=1}^M \sum_{Y^n} \overline{p(y|u_i)^{1/\rho}} \left(\sum_{j:j \neq i} \overline{p(y|u_j)^{1/\rho}} \right)^\rho, \quad 0 < \rho \leq 1. \quad (3.12.10)$$

Заметим далее, что величина

$$\overline{p(y|u_i)^{1/\rho}} = \sum_{u_i \in X^n} q(u_i) p(y|u_i)^{1/\rho}$$

не зависит от u_i , так как u_i — переменная суммирования. Используя это замечание и обозначая переменную суммирования через x вместо u_i , из неравенства (3.12.10) получим

$$\begin{aligned} \bar{\lambda} &\leq (M-1)^\rho \sum_{Y^n} (\overline{p(y|u_i)^{1/\rho}})^{1+\rho} \leq \\ &\leq M^\rho \sum_{Y^n} \left[\sum_{X^n} q(x) p(y|x)^{1/\rho} \right]^{1+\rho}. \end{aligned} \quad (3.12.11)$$

Неравенство (3.12.11) представляет собой среднюю по ансамблю кодов оценку вероятности ошибочного декодирования, справедливую для произвольного дискретного канала. Ниже эта оценка будет упрощена для каналов без памяти.

3.12.3. Оценка средней по ансамблю кодов вероятности ошибки декодирования для дискретных каналов без памяти. Рассмотрим дискретный канал без памяти, задаваемый переходными вероятностями

$$p(y|x) = \prod_{i=1}^n p(y^{(i)}|x^{(i)}), \quad x \in X^n, \quad y \in Y^n. \quad (3.12.12)$$

Будем предполагать, что распределение вероятностей $q(x)$, $x \in X^n$, которое задает ансамбль кодов, определяется следующим образом посредством распределения вероятностей $Q(x)$, $x \in X$:

$$q(x) = \prod_{i=1}^n Q(x^{(i)}), \quad x = (x^{(1)}, \dots, x^{(n)}) \in X^n. \quad (3.12.13)$$

Другими словами, мы теперь предполагаем, что не только кодовые слова независимы в ансамбле кодов, но независимыми являются также символы кодовых слов. Каждое кодовое слово образуется с помощью независимого выбора символов $x^{(i)} \in X$, $i = 1, \dots, n$, причем вероятность того, что будет выбран символ x , есть $Q(x)$.

Подставляя (3.12.13) в (3.12.11), а также учитывая (3.12.12), получим

$$\begin{aligned} \bar{\lambda} &\leq M^\rho \sum_{Y^n} \left[\sum_{X^n} \prod_{i=1}^n Q(x^{(i)}) p(y^{(i)}|x^{(i)})^{1/(1+\rho)} \right]^{1+\rho} = \\ &= M^\rho \sum_{Y^n} \left[\sum_{x^{(1)} \in X} Q(x^{(1)}) p(y^{(1)}|x^{(1)})^{1/(1+\rho)} \dots \right. \\ &\quad \left. \dots \sum_{x^{(n)} \in X} Q(x^{(n)}) p(y^{(n)}|x^{(n)})^{1/(1+\rho)} \right]^{1+\rho} = \\ &= M^\rho \left\{ \sum_Y \left[\sum_X Q(x) p(y|x)^{1/(1+\rho)} \right]^{1+\rho} \right\}^n, \end{aligned} \quad (3.12.14)$$

где последнее равенство получается в результате того, что распределения вероятностей $Q(x^{(i)})$ и $p(y^{(i)}|x^{(i)})$ от i не зависят.

Обозначим через \mathbf{Q} вероятностный вектор $(Q(x_1), \dots, Q(x_L))$ и положим

$$E_0(\rho, \mathbf{Q}) \triangleq -\log \sum_Y \left[\sum_X Q(x) p(y|x)^{1/(1+\rho)} \right]^{1+\rho}. \quad (3.12.15)$$

Функция $E_0(\rho, \mathbf{Q})$ называется функцией Галлагера. Пусть, кроме того,

$$E(R) \triangleq \max_{\rho, \mathbf{Q}} (-\rho R + E_0(\rho, \mathbf{Q})), \quad (3.12.16)$$

где максимум разыскивается по всем ρ , $0 < \rho \leq 1$, и по всем вероятностным векторам \mathbf{Q} , т. е. по всем распределениям вероятностей $Q(x)$ на X . Учитывая, что $M = 2^{nR}$, из (3.12.14) следует, что

$$\bar{\lambda} \leq 2^{-nE(R)}. \quad (3.12.17)$$

Таким образом, средняя по ансамблю кодов вероятность ошибки при декодировании по максимуму правдоподобия в дискретном канале без памяти удовлетворяет неравенству (3.12.17). Учитывая этот результат и рассуждения, связанные с методом случайного кодирования, можно сформулировать следующую теорему.

Теорема 3.12.1. Существует код со скоростью R , который в дискретном канале без памяти обеспечивает среднюю вероятность ошибки, удовлетворяющую неравенству (3.12.17), где n — длина кода, а показатель $E(R)$ зависит только от скорости кода.

Более того, имеет место следующее утверждение.

Теорема 3.12.2. Для произвольного дискретного канала без памяти $\{XY, p(y|x)\}$ существует код со скоростью R , для которого средняя вероятность ошибки декодирования удовлетворяет неравенству (3.12.17), где n — длина кода, а функция $E(R)$ зависит только от матрицы переходных вероятностей канала и от скорости кода, причем $E(R) > 0$ при всех R , $0 \leq R \leq C^*$, где C^* — информационная емкость рассматриваемого канала.

З а м е ч а н и е. Функция $E(R)$, определяемая соотношением (3.12.16), называется *экспонентой случайного кодирования*.

Д о к а з а т е л ь с т в о. Так как средняя по ансамблю кодов вероятность ошибки удовлетворяет неравенству (3.12.17), то в ансамбле найдется по крайней мере один код, вероятность ошибки декодирования которого также удовлетворяет неравенству (3.12.17).

Покажем теперь, что $E(R) > 0$ для всех скоростей R , меньших информационной емкости C^* . Для этого рассмотрим функцию $E_0(\rho, Q)$, определяемую соотношением (3.12.15). Вычислим производную этой функции по ρ в точке $\rho = 0$. Определяя дифференцирование и полагая $\rho = 0$, получим

$$\begin{aligned} \frac{\partial E_0(\rho, Q)}{\partial \rho} \Big|_{\rho=0} &= \\ &= - \sum_Y \left[\left(\sum_X Q(x) p(y|x) \right) \ln \left(\sum_X Q(x) p(y|x) \right) - \right. \\ &\quad \left. - \frac{\sum_X Q(x) p(y|x) \ln p(y|x)}{\sum_X Q(x) p(y|x)} \right] \log e = \\ &= \sum_X \sum_Y Q(x) p(y|x) \log \frac{p(y|x)}{p(y)} = I(X; Y). \end{aligned}$$

Обозначим выражение под знаком максимума в (3.12.16) через $E(R, \rho, Q)$. Тогда с учетом последнего соотношения

$$\frac{\partial E(R, \rho, Q)}{\partial \rho} \Big|_{\rho=0} = I(X; Y) - R.$$

Таким образом, если скорость кода R удовлетворяет неравенству $R < I(X; Y)$, то производная $E(R, \rho, Q)$ по ρ в точке $\rho = 0$ положительна. Из непрерывности $E(R, \rho, Q)$ по ρ и из того, что $E(R, \rho = 0, Q) = 0$, следует, что для любого вероятностного вектора Q и любого R , меньшего, чем информация $I(X; Y)$, найдется такое ρ' , $0 < \rho' \ll 1$, при котором $E(R, \rho', Q) > 0$. Если выбрать Q таким, при котором максимизируется величина $I(X; Y)$, то предыдущее утверждение будет верным для всех скоростей R , меньших, чем информационная емкость канала C^* . Для завершения доказательства теоремы достаточно заметить, что $E(R) \geq E(R, \rho, Q)$, как бы ни были выбраны вероятностный вектор Q и параметр ρ , $0 < \rho < 1$. Теорема доказана.

Теоремы 3.12.1 и 3.12.2 в совокупности содержат утверждение, аналогичное утверждению прямой теоремы кодирования 3.9.1 для дискретного канала без памяти. Отличие этих двух

утверждений состоит в том, что в последних двух теоремах указывается не только то, что можно достичь сколь угодно малой вероятности ошибки декодирования за счет выбора большого n и удачного выбора кода, но указывается достаточно сильная, как мы покажем, в дальнейшем, оценка для средней вероятности ошибки некоторого кода из ансамбля \mathcal{Q} при конкретных значениях n и R .

3.12.4*. Свойства функции $E_0(\rho, Q)$ и построение экспоненты случайного кодирования. Ниже мы исследуем функцию $E_0(\rho, Q)$ (функцию Галлагера) и установим некоторые ее важные свойства, которые дадут возможность указать способ вычислений экспоненты случайного кодирования (3.12.16).

Предположим, что на множестве X заданы два распределения вероятностей $Q(x)$ и $Q^*(x)$, $x \in X$, или два вероятностных вектора Q и Q^* . Введем понятие энтропии одного распределения относительно другого.

О п р е д е л е н и е 3.12.1. Величина $\mathcal{H}(Q^*, Q)$, определяемая следующим выражением:

$$\mathcal{H}(Q^*, Q) \triangleq \sum_X Q^*(x) \log \frac{Q^*(x)}{Q(x)}, \quad (3.12.18)$$

называется *энтропией распределения Q^* относительно распределения Q* .

Л е м м а 3.12.1. Для произвольных распределений Q и Q^* на X функция $\mathcal{H}(Q^*, Q)$ неотрицательна и равна нулю в том и только том случае, когда $Q(x) = Q^*(x)$ для всех $x \in X$.

Д о к а з а т е л ь с т в о. Используя неравенство $\ln x \leq x - 1$, получим

$$\begin{aligned} \mathcal{H}(Q^*, Q) &= - \sum_X Q^*(x) \log \frac{Q(x)}{Q^*(x)} \geq \\ &\geq - \log e \left(\sum_X Q(x) - \sum_X Q^*(x) \right) = 0. \end{aligned}$$

Так как в неравенстве для логарифма равенство имеет место тогда и только тогда, когда аргумент логарифма равен единице, то последнее неравенство обращается в равенство тогда и только тогда, когда $Q(x) = Q^*(x)$ для всех $x \in X$. Лемма доказана.

Помимо энтропии распределения Q^* относительно распределения Q будут встречаться другие сходные понятия. Нижний индекс у вероятностного вектора будет обозначать множество, на котором рассматривается данное распределение. Так, p_Y — распределение на Y , p_{XY} — распределение на XY . Энтропии $\mathcal{H}(p_Y, p'_Y)$ и $\mathcal{H}(p'_{XY}, p_{XY})$ имеют тот же смысл, что и введенная выше энтропия $\mathcal{H}(Q^*, Q)$. Если из контекста ясно, на каком множестве

рассматривается то или другое распределение вероятностей, то подстрочный индекс иногда опускается, например: $\mathcal{H}(Q'_X, Q_X) = \mathcal{H}(Q^*, Q)$. В некоторых случаях одно из распределений или оба могут быть условными, например, $p'_X = \{p'(x)\}, x \in X$, $p''_{X|y} = \{p''(x|y)\}, x \in X$. Между такими распределениями также определена энтропия одного распределения относительно другого, которую мы будем обозначать как $\mathcal{H}(p'_X, p''_{X|y})$. Такая энтропия иногда рассматривается как случайная величина на ансамбле Y . В этом случае ее математическое ожидание обозначается через $\mathcal{H}_{p_Y}(p'_X, p''_{X|y})$:

$$\begin{aligned}\mathcal{H}_{p_Y}(p'_X, p''_{X|y}) &\triangleq \sum_Y p(y) \mathcal{H}(p'_X, p''_{X|y}) = \\ &= \sum_Y p(y) \sum_X p'(x) \log \frac{p'(x)}{p''(x|y)}.\end{aligned}$$

Во всех таких случаях лемма 3.12.1 остается справедливой.

Пусть теперь $\{XY, p(y|x)\}$ — дискретный канал без памяти с матрицей переходных вероятностей $P = \{p(y|x)\}, x \in X, y \in Y$. Введем в рассмотрение условные распределения вероятностей $v(x|y), x \in X, y \in Y$. Будем считать, что $v(x|y) = 0$ только в том случае, когда для тех же значений $x \in X$ и $y \in Y$ имеет место равенство $p(y|x) = 0$. Пусть $V = \{v(x|y)\}, x \in X, y \in Y$, — матрица, элементами которой являются вероятности $v(x|y)$, и пусть

$$E_0(\rho, Q, V) \triangleq -\log \sum_X \sum_Y p(y|x) Q^{1+\rho}(x) v^{-\rho}(x|y). \quad (3.12.19)$$

В следующей теореме устанавливается связь между функциями $E_0(\rho, Q)$ и $E_0(\rho, Q, V)$, а также устанавливаются некоторые свойства функции $E_0(\rho, Q)$.

Теорема 3.12.3. Пусть фиксированы дискретный канал без памяти $\{XY, p(y|x)\}$ и распределение вероятностей $Q(x)$ на входном алфавите этого канала. Имеют место следующие утверждения.

$$1. \quad E_0(\rho, Q) = \max_V E_0(\rho, Q, V), \quad (3.12.20)$$

где максимум разыскивается по всем таким стохастическим матрицам V , что $v(x|y) = 0$, если только $p(y|x) = 0$.

2. Пусть $\tilde{V} = \{v(x|y)\}, x \in X, y \in Y$, — стохастическая матрица, на которой достигается максимум в выражении (3.12.20). Тогда величины $\tilde{v}(x|y)$ удовлетворяют следующей системе уравнений:

$$\tilde{p}(x, y) = \tilde{p}(y) \tilde{v}(x|y), \quad x \in X, \quad y \in Y, \quad (3.12.21)$$

где

$$\tilde{p}(x, y) \triangleq \frac{p(y|x) Q^{1+\rho}(x) v^{-\rho}(x|y)}{\sum_X \sum_Y p(y|x) Q^{1+\rho}(x) v^{-\rho}(x|y)}, \quad (3.12.22)$$

$$\tilde{p}(y) \triangleq \sum_X \tilde{p}(x, y). \quad (3.12.23)$$

3. Обозначим через $\tilde{v}_{X|y}$ вероятностный вектор $\{\tilde{v}(x|y)\}, x \in X$, который задает оптимальное распределение вероятностей на X при фиксированном значении $y \in Y$; обозначим через \tilde{p}_Y и \tilde{p}_{XY} вероятностные векторы, соответствующие распределениям на Y (3.12.23) и на XY (3.12.22) соответственно. Пусть \tilde{p}_X — вероятностный вектор, соответствующий распределению $\tilde{p}(x) = \sum_Y \tilde{p}(x, y)$ на X . Тогда

$$\begin{aligned}\frac{\partial E_0(\rho, Q)}{\partial \rho} &= \mathcal{H}_{\tilde{p}_Y}(\tilde{v}_{X|y}, Q) = \\ &= \sum_Y \tilde{p}(y) \mathcal{H}(\tilde{v}_{X|y}, Q) = I_{\tilde{p}_{XY}}(X; Y) + \mathcal{H}(\tilde{p}_X, Q),\end{aligned} \quad (3.12.24)$$

где $I_{\tilde{p}_{XY}}(X; Y)$ — средняя взаимная информация, вычисленная в соответствии с распределением $\tilde{p}(x, y), x \in X, y \in Y$.

4. В обозначениях п. 3 имеет место равенство

$$\frac{\partial^2 E_0(\rho, Q)}{\partial \rho^2} = -\frac{1}{\log e} \sum_X \sum_Y \tilde{p}(x, y) \left[\log \frac{\tilde{v}(x, y)}{Q(x)} - \mathcal{H}_{\tilde{p}_Y}(\tilde{v}_{X|y}, Q) \right]^2. \quad (3.12.25)$$

Доказательство. Найдем частную производную $E_0(\rho, Q, V)$ по $v(x|y)$:

$$\frac{\partial E_0(\rho, Q, V)}{\partial v(x|y)} = (\log e) \frac{\rho p(y|x) Q^{1+\rho}(x) v^{-(1+\rho)}(x|y)}{\sum_X \sum_Y p(y|x) Q^{1+\rho}(x) v^{-\rho}(x|y)}.$$

В соответствии с теоремой Куна—Таккера (см. параграф 2.8) необходимые условия достижения максимума в (3.12.20) состоят в том, чтобы при каждом $y \in Y$

$$\frac{\partial E_0(\rho, Q, V)}{\partial v(x|y)} \leq c_y, \quad x \in X, \quad (3.12.26)$$

где c_y — множитель Лагранжа, обусловленный ограничением $\sum_X v(x|y) = 1$, и строгое неравенство в (3.12.26) возможно только при условии $v(x|y) = 0$ при данном y . Условия (3.12.26) эквивалентны условиям

$$\tilde{p}(x, y) - \frac{c_y}{\rho \log e} \tilde{v}(x|y) \leq 0, \quad x \in X, \quad (3.12.27)$$

для максимизирующих значений $\tilde{v}(x|y)$, $x \in X$, $y \in Y$. Суммируя обе части последнего неравенства по $x \in X$, получим, что при каждом $y \in Y$

$$\tilde{p}(y) < \frac{c_y}{\rho \log e}.$$

Если $p(y|x) > 0$, то $\tilde{p}(x, y) > 0$ и условия (3.12.27) могут быть выполнены только в случае $\tilde{v}(x|y) > 0$. Следовательно, $c_y/\rho \log e = \tilde{p}(y)$, откуда с учетом (3.12.27) получаются соотношения (3.12.21). Утверждение 2 доказано.

Заметим далее, что выражение под знаком логарифма в правой части (3.12.19) при $\rho \geq 0$ является выпуклой вниз функцией $v(x|y)$, $x \in X$, $y \in Y$. Действительно, $v^{-\rho}$ — выпуклая вниз функция v , а вся сумма является некоторой комбинацией с неотрицательными коэффициентами выпуклых вниз функций (см. задачу 2.5.2). Поэтому в силу монотонности логарифма максимум (3.12.20) единствен и, следовательно, уравнения (3.12.21) имеют единственное решение.

Покажем, что решением является система величин

$$\tilde{v}(x|y) = \frac{Q(x)p^{1/1+\rho}(y|x)}{\sum_x Q(x)p^{1/1+\rho}(y|x)}, \quad x \in X, \quad y \in Y, \quad (3.12.28)$$

и что при этом имеет место соотношение (3.12.20). Действительно, подстановка (3.12.28) в (3.12.19) дает

$$E_0(\rho, Q, \tilde{V}) = -\log \sum_Y \left(\sum_X Q(x)p^{1/1+\rho}(y|x) \right)^{1+\rho} \triangleq E_0(\rho, Q).$$

Подставляя теперь (3.12.28) в (3.12.22) и (3.12.23), получим

$$\tilde{p}(x, y) = \frac{Q(x)p^{1/1+\rho}(y|x) \left(\sum_X Q(x)p^{1/1+\rho}(y|x) \right)^\rho}{\sum_Y \left(\sum_X Q(x)p^{1/1+\rho}(y|x) \right)^{1+\rho}},$$

$$\tilde{p}(y) = \frac{\left(\sum_X Q(x)p^{1/1+\rho}(y|x) \right)^{1+\rho}}{\sum_Y \left(\sum_X Q(x)p^{1/1+\rho}(y|x) \right)^{1+\rho}}.$$

Таким образом, величины (3.12.28) удовлетворяют системе уравнений (3.12.21) и, следовательно, максимизируют правую часть соотношения (3.12.20), что доказывает утверждение 1.

Для доказательства утверждения 3 найдем производную $E_0(\rho, Q, V)$ по ρ при фиксированной функции $v(x|y)$, $x \in X$, $y \in Y$:

$$\begin{aligned} \frac{\partial E_0(\rho, Q, V)}{\partial \rho} &= -\log e \frac{\sum_X \sum_Y p(y|x) Q^{1+\rho}(x) v^{-\rho}(x|y) \ln \frac{Q(x)}{v(x|y)}}{\sum_X \sum_Y p(y|x) Q^{1+\rho}(x) v^{-\rho}(x|y)} = \\ &= \sum_X \sum_Y p^*(x, y) \log \frac{v(x|y)}{Q(x)}, \end{aligned}$$

где $p^*(x, y)$ вычисляется по формуле (3.12.22), в которой $\tilde{v}(x|y)$ заменено на $v(x|y)$. Если теперь положить $v(x|y) = \tilde{v}(x|y)$, где $\tilde{v}(x|y)$ дается соотношениями (3.12.28), то с учетом (3.12.20) и (3.12.21) получим

$$\begin{aligned} \frac{\partial E_0(\rho, Q)}{\partial \rho} &= \frac{\partial E_0(\rho, Q, \tilde{V})}{\partial \rho} = \sum_X \sum_Y \tilde{p}(x, y) \log \frac{\tilde{v}(x|y)}{Q(x)} = \\ &= \sum_Y \tilde{p}(y) \sum_X \tilde{v}(x|y) \log \frac{\tilde{v}(x|y)}{Q(x)} \triangleq \mathcal{H}_{\tilde{p}_Y}(\tilde{v}_{X|y}, Q). \end{aligned}$$

Кроме того,

$$\begin{aligned} \mathcal{H}_{\tilde{p}_Y}(\tilde{v}_{X|y}, Q) &= \sum_X \sum_Y \tilde{p}(x, y) \log \frac{\tilde{v}(x|y) \tilde{p}(x)}{Q(x) \tilde{p}(x)} = \\ &= \sum_X \sum_Y \tilde{p}(x, y) \log \frac{\tilde{v}(x|y)}{\tilde{p}(x)} + \sum_X \tilde{p}(x) \log \frac{\tilde{p}(x)}{Q(x)} = \\ &= I_{\tilde{p}_{XY}}(X; Y) + \mathcal{H}_{\tilde{p}_X}(Q), \end{aligned}$$

что и завершает доказательство утверждения 3.

Для доказательства утверждения 4 найдем вторую производную $E_0(\rho, Q, V)$ по ρ при фиксированной функции $v(x|y)$, $x \in X$, $y \in Y$:

$$\begin{aligned} \frac{\partial^2 E_0(\rho, Q, V)}{\partial \rho^2} &= \\ &= -\log e \left[\sum_X \sum_Y \frac{p(y|x) Q^{1+\rho}(x) v^{-\rho}(x|y) \ln^2 \frac{Q(x)}{v(x|y)}}{\sum_X \sum_Y p(y|x) Q^{1+\rho}(x) v^{-\rho}(x|y)} - \right. \\ &\quad \left. - \left(\frac{\sum_X \sum_Y p(y|x) Q^{1+\rho}(x) v^{-\rho}(x|y) \ln \frac{Q(x)}{v(x|y)}}{\sum_X \sum_Y p(y|x) Q^{1+\rho}(x) v^{-\rho}(x|y)} \right)^2 \right] = \end{aligned}$$

$$\begin{aligned}
 &= -\frac{1}{\log e} \left[\sum_x \sum_y p^*(x, y) \log^2 \frac{v(x, y)}{Q(x)} - \right. \\
 &\quad \left. - \left(\sum_x \sum_y p^*(x, y) \log \frac{v(x|y)^2}{Q(x)} \right)^2 \right] = \\
 &= -\frac{1}{\log e} \sum_x \sum_y p^*(x, y) \left(\log \frac{v(x|y)}{Q(x)} - \right. \\
 &\quad \left. - \sum_x \sum_y p^*(x, y) \log \frac{v(x|y)}{Q(x)} \right)^2.
 \end{aligned}$$

Если теперь положить $v(x|y) = \tilde{v}(x|y)$, то получим утверждение 4. Теорема доказана.

Следствие 3.12.1. Пусть распределение вероятностей $Q(x)$, $x \in X$, и дискретный канал без памяти таковы, что $I(X; Y) > 0$. Тогда функция $E_0(\rho, Q)$, определяемая соотношением (3.12.15), имеет следующие свойства при $\rho \geq 0$:

1. $E_0(\rho, Q) \geq 0$, причем равенство имеет место только при $\rho = 0$.

2. $I(X; Y) \geq \frac{\partial E_0(\rho, Q)}{\partial \rho} \geq 0$, причем равенство в первом неравенстве имеет место только при $\rho = 0$.

3. $\frac{\partial^2 E_0(\rho, Q)}{\partial \rho^2} \leq 0$.

Доказательство. При $\rho = 0$ выражение под знаком логарифма в (3.12.15) равно единице, поэтому $E_0(\rho, Q) = 0$ при $\rho = 0$. Из (3.12.22) следует, что $\tilde{p}(x, y) = Q(x)p(y|x)$ при $\rho = 0$, поэтому $\tilde{p}_X = Q$ и из утверждения 3 теоремы и леммы 3.12.1 вытекает, что

$$\frac{\partial E_0(\rho, Q)}{\partial \rho} \Big|_{\rho=0} = I(X; Y).$$

Так как согласно (3.12.24) производная $\partial E_0/\partial \rho$ неотрицательна при всех $\rho \geq 0$, то отсюда следует первое доказываемое свойство. Третье свойство непосредственно следует из утверждения 4 теоремы. Это свойство показывает, что $\partial E_0/\partial \rho$ не возрастает по ρ и поэтому выполняется свойство 2. Следствие доказано.

С помощью следствия 3.12.1 можно легко максимизировать правую часть соотношения (3.12.16) по ρ при заданном вероятностном векторе Q . Пусть

$$E(R, Q) \triangleq \max_{0 \leq \rho \leq 1} (E_0(\rho, Q) - \rho R). \quad (3.12.29)$$

Дифференцируя по ρ выражение, стоящее под знаком максимума,

и приравнивая производную к нулю, получим следующее уравнение для максимизирующего значения ρ :

$$\frac{\partial E_0(\rho, Q)}{\partial \rho} = R. \quad (3.12.30)$$

Используя свойство 3 следствия, получим, что любое неотрицательное решение этого уравнения максимизирует (3.12.29). Так как $\partial E_0/\partial \rho$ является непрерывной и убывающей функцией от ρ , то решение уравнения (3.12.30), лежащее в интервале $[0, 1]$, существует, если

$$\begin{aligned}
 R_{\text{кр}} &\triangleq \frac{\partial E_0(\rho, Q)}{\partial \rho} \Big|_{\rho=1} \leq R \leq \\
 &\leq \frac{\partial E_0(\rho, Q)}{\partial \rho} \Big|_{\rho=0}.
 \end{aligned}$$

Значение $R_{\text{кр}}$ называется *критической скоростью* для данного распределения Q .

Уравнения

$$R = \frac{\partial E_0(\rho, Q)}{\partial \rho}, \quad 0 \leq \rho \leq 1, \quad (3.12.31)$$

$$E(R, Q) = E_0(\rho, Q) - \rho \cdot \frac{\partial E_0(\rho, Q)}{\partial \rho}$$

задают значения $E(R, Q)$ при $R_{\text{кр}} \leq R \leq I(X; Y)$ для фиксированного распределения вероятностей на входе $Q(x)$, $x \in X$.

Метод построения функции $E_0(R, Q)$ можно иллюстрировать с помощью рис. 3.12.1. Согласно следствию 3.12.1 функция $E_0(\rho, Q)$, изображенная на рисунке, монотонно возрастает от нуля, оставаясь выпуклой вверх. Если провести касательную к этой функции с наклоном R , то абсцисса точки касания будет равна корню уравнения (3.12.30). Касательная отсекает на оси ординат отрезок, равный $E(R, Q)$. Этот отрезок остается положительным для всех значений скорости R , меньших информации $I(X; Y)$, вычисленной относительно распределения вероятностей $Q(x)$ на входе канала.

Если выбрать распределение вероятностей $Q(x)$ так, чтобы информация $I(X; Y)$ была равна своему максимальному значению, т. е. была равной пропускной способности C рассматриваемого канала, то показатель экспоненты $E(R)$ вероятности ошибки (см. (3.12.17)) будет положительной величиной для всех скоростей, меньших пропускной способности C .

Теперь можно дать общий метод построения функции $E(R, Q)$. Из (3.12.31), дифференцируя по ρ , получим

$$\frac{\partial R}{\partial \rho} = \frac{\partial^2 E_0(\rho, Q)}{\partial \rho^2}, \quad \frac{\partial E(R, Q)}{\partial \rho} = -\rho \frac{\partial^2 E_0(\rho, Q)}{\partial \rho^2}.$$

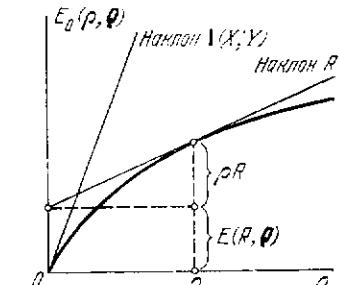


Рис. 3.12.1. Функция $E_0(\rho, Q)$.

Следовательно, при изменении ρ от 0 до 1 значение R убывает монотонно от $I(X; Y)$ до $R_{\text{кр}}$, а $E(R, Q)$ монотонно возрастает от нуля до $E_0(1, Q) = R_{\text{кр}}$. Взяв отношение производных, получим

$$\frac{\partial E(R, Q)}{\partial R} = -\rho, \quad 0 < \rho < 1.$$

Таким образом, параметр $-\rho$ представляет собой наклон касательной к функции $E(R, Q)$ в точке R .

При $R < R_{\text{кр}}$ значение $E_0(\rho, Q) - \rho R$ достигает максимума при $\rho = 1$ при условии, что $0 < \rho < 1$. Это можно увидеть, например, из рис. 3.12.1: отрезок, отсекаемый касательной на оси ординат, монотонно увеличивается с ростом ρ . Поэтому

$$E(R, Q) = E_0(1, Q) - R, \quad R < R_{\text{кр}}.$$

Для каждого значения R из интервала $[R_{\text{кр}}, I(X; Y)]$ касательная к кривой $E(R, Q)$ в точке с абсциссой R имеет наклон $-\rho$. Эта касательная отсекает на оси ординат отрезок, равный

$$E(R, Q) + \rho R = E_0(\rho, Q).$$

Для всех R из интервала $[0, R_{\text{кр}}]$ график функции $E(R, Q)$ представляет собой прямую с наклоном -1 , отсекающую на оси ординат отрезок $E_0(1, Q)$. Поэтому график функции $E(R, Q)$ может быть построен следующим образом (см. рис. 3.12.2). Для каждого значения ρ из интервала $[0, 1]$ может быть вычислено $E_0(\rho, Q)$ и проведена прямая с наклоном $-\rho$ из точки $E_0(\rho, Q)$ на оси ординат. Семейство таких прямых является семейством касательных, верхняя огибающая которых и представляет собой функцию $E(R, Q)$.

Показатель экспоненты случайного кодирования, $E(R)$, можно теперь представить следующим образом:

$$E(R) = \max_Q E(R, Q) = \max_{0 < \rho < 1} (E_0(\rho) - \rho R), \quad (3.12.32)$$

где

$$E_0(\rho) \triangleq \max_Q E_0(\rho, Q). \quad (3.12.33)$$

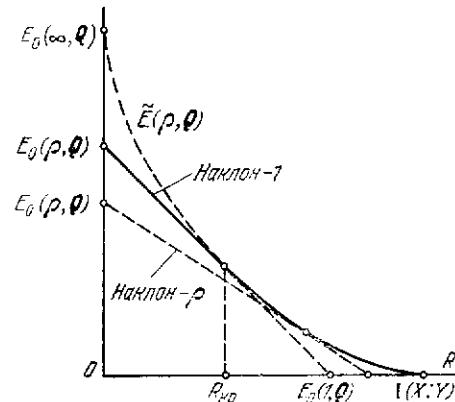


Рис. 3.12.2. Функция $E(R, Q)$ как огибающая семейства касательных.

$E(R, Q)$ в точке с абсциссой R имеет наклон $-\rho$. Эта касательная отсекает на оси ординат отрезок, равный

Отыскание распределения вероятностей $Q(x)$, максимизирующего функцию $E(R, Q)$ при каждом R или функцию $E_0(\rho, Q)$ при каждом ρ , в общем случае представляет собой довольно сложную задачу, во многом аналогичную задаче вычисления информационной емкости канала. Однако в случае симметричных каналов эта задача имеет простое решение. Ниже мы покажем, что в этом случае при всех R оптимальным распределением на X является равномерное: $Q(x) = L^{-1}$ для всех $x \in X$, где L — число элементов в X .

Если оптимизирующее распределение $Q(x)$ найдено и величина $E_0(\rho)$ может быть вычислена для каждого ρ , то для построения графика функции $E(R)$ могут быть использованы те же рассуждения, которые были использованы при построении графика функции $E(R, Q)$ (см. рис. 3.12.2) с заменой $E_0(\rho, Q)$ на $E_0(\rho)$.

3.12.5*. Показатель экспоненты случайного кодирования для симметричных каналов без памяти. Докажем вначале лемму, в которой устанавливаются необходимые и достаточные условия, которым должно удовлетворять распределение вероятностей $Q(x)$, максимизирующее $E_0(\rho, Q)$ при каждом фиксированном значении ρ в случае произвольных дискретных каналов без памяти.

Лемма 3.12.2. Для того чтобы распределение вероятностей $Q(x)$ максимизировало $E_0(\rho, Q)$ при фиксированном ρ , необходимо и достаточно выполнения следующих условий:

$$\frac{\sum_y p^{1/1+\rho}(y|x) \left(\sum_x Q(x) p^{1/1+\rho}(y|x) \right)^\rho}{\sum_y \left(\sum_x Q(x) p^{1/1+\rho}(y|x) \right)^{1+\rho}} \geq 1 \quad (3.12.34)$$

при всех $x \in X$, причем знак неравенства может иметь место только для тех $x \in X$, для которых $Q(x) = 0$.

Доказательство. Используя теорему Куна—Таккера, получаем, что необходимые условия максимума функции $E_0(\rho, Q)$ (3.12.15) суть

$$\frac{\sum_y p^{1/1+\rho}(y|x) \left(\sum_x Q(x) p^{1/1+\rho}(y|x) \right)^\rho}{\sum_y \left(\sum_x Q(x) p^{1/1+\rho}(y|x) \right)^{1+\rho}} \geq \frac{c}{(1+\rho) \log e}, \quad x \in X, \quad (3.12.35)$$

причем если $Q(x) > 0$, то имеет место знак равенства. Домножая обе части неравенства (3.12.35) на $Q(x)$ и суммируя по всем $x \in X$, для которых $Q(x) > 0$, получим, что $c/(1+\rho) \log e = 1$. Заметим далее, что выражение под знаком логарифма в (3.12.15) выпукло вниз по $Q(x)$, так как $1+\rho \geq 1$. Поэтому в силу монотонности логарифма максимум $E_0(\rho, Q)$ по Q единствен и, следо-

вательно, необходимые условия (3.12.35) являются также достаточными. Лемма доказана.

Покажем теперь, что в случае симметричных каналов без памяти оптимальным является равномерное распределение вероятностей $\mathbf{Q} = (1/L, \dots, 1/L)$. Напомним, что в случае симметричных каналов матрица переходных вероятностей обладает симметрией по строкам и столбцам: все строки и все столбцы образованы перестановками одного и того же набора чисел $\{p_1, p_2, \dots, p_L\}$. Поэтому

$$\sum_X p^{1/1+\rho}(y|x) = \sum_Y p^{1/1+\rho}(y|x) = \sum_{i=1}^L p_i^{1/1+\rho} = k, \quad (3.12.36)$$

где k — некоторая константа.

Следствие 3.12.2. При любом $\rho \geq 0$ функция $E_0(\rho, \mathbf{Q})$ в случае симметричных каналов принимает максимальное значение на равномерном распределении вероятностей.

Доказательство. Подставляя в левую часть (3.12.34) $\mathbf{Q} = \{1/L, \dots, 1/L\}$ и используя соотношения (3.12.36), получим

$$\begin{aligned} & \frac{\sum_Y p^{1/1+\rho}(y|x) \left(\sum_X Q(x) p^{1/1+\rho}(y|x) \right)^\rho}{\sum_Y \left(\sum_X Q(x) p^{1/1+\rho}(y|x) \right)^{1+\rho}} = \\ & = \frac{L^{-\rho} \sum_Y p^{1/1+\rho}(y|x) \left(\sum_X p^{1/1+\rho}(y|x) \right)^\rho}{L^{-1-\rho} \sum_Y \left(\sum_X p^{1/1+\rho}(y|x) \right)^{1+\rho}} = \frac{Lk^{1+\rho}}{\sum_Y k^{1+\rho}} = 1. \end{aligned} \quad (3.12.37)$$

Отсюда, применяя лемму 3.12.2, получим доказываемое утверждение.

Теперь легко выписать функцию $E_0(\rho) = \max_{\mathbf{Q}} E_0(\rho, \mathbf{Q})$. Подставляя оптимизирующее распределение вероятностей в формулу (3.12.15), получим, что

$$E_0(\rho) = \rho \log L - (1 + \rho) \log \sum_{i=1}^L p_i^{1/1+\rho}. \quad (3.12.38)$$

Выше был описан способ построения графика функции $E(R)$ с помощью построения семейства касательных, верхней огибающей которых и является функция $E(R)$. Этот метод геометрически нагляден, но неудобен тогда, когда нужно иметь аналитическую запись для $E(R)$. Ниже мы покажем, как можно найти решение системы уравнений (3.12.31) в случае симметричных каналов. В рассматриваемом случае после оптимизации по распре-

делению $Q(x)$ система уравнений, параметрически задающая функцию $E(R)$, имеет вид:

$$\begin{aligned} R &= \frac{\partial E_0(\rho)}{\partial \rho}, \quad 0 < \rho < 1, \\ E(R) &= E_0(\rho) - \rho \frac{\partial E_0(\rho)}{\partial \rho}. \end{aligned} \quad (3.12.39)$$

Мы будем пользоваться теоремой 3.12.3, поэтому вначале подставим $Q(x) = 1/L$ в формулу (3.12.22) и найдем распределения $\tilde{p}(x, y)$, $\tilde{p}(x)$, $\tilde{p}(y|x)$, $x \in X$, $y \in Y$:

$$\begin{aligned} \tilde{p}(x, y) &= \frac{L^{-1} p^{1/1+\rho}(y|x) L^{-\rho} k^\rho}{L^{-\rho} k^{1+\rho}} = \frac{1}{L} \frac{p^{1/1+\rho}(y|x)}{k}, \\ \tilde{p}(x) &= \sum_Y \tilde{p}(x, y) = \frac{1}{L} = Q(x), \\ \tilde{p}(y|x) &= \frac{\tilde{p}(x, y)}{\tilde{p}(x)} = \frac{p^{1/1+\rho}(y|x)}{k}, \end{aligned} \quad (3.12.40)$$

где k — константа, определяемая соотношением (3.12.36).

Таким образом, из утверждения 3 теоремы 3.12.3 с учетом того, что $\mathcal{H}(\tilde{\mathbf{p}}_X, \mathbf{Q}_X) = 0$, получим, что первое соотношение в (3.12.39) эквивалентно следующему:

$$R = I_{\tilde{\mathbf{p}}_{XY}}(X; Y), \quad (3.12.41)$$

где средняя взаимная информация вычислена в соответствии с распределением $\tilde{p}(x, y)$, $x \in X$, $y \in Y$.

Заметим далее, что $\tilde{p}(y|x)$, $x \in X$, $y \in Y$, представляют собой переходные вероятности некоторого симметричного канала, если исходный канал является симметричным. Матрица переходных вероятностей этого нового канала образована элементами

$$\tilde{p}_i = \frac{p_i^{1/1+\rho}}{\sum_{i=1}^L p_i^{1/1+\rho}}. \quad (3.12.42)$$

Так как распределение вероятностей $\tilde{p}(x)$ — равномерное, то средняя взаимная информация $I_{\tilde{\mathbf{p}}_{XY}}(X; Y)$ равна пропускной способности $C(\tilde{\mathbf{p}})$ этого канала. В результате соотношение (3.12.41) можно записать следующим образом:

$$R = C(\tilde{\mathbf{p}}) = \log L + \sum_{i=1}^L \tilde{p}_i \log \tilde{p}_i. \quad (3.12.43)$$

Для вычисления $E(R)$ как функции от R можно воспользоваться следующим приемом. Будем менять параметр ρ в интервале от 0 до 1. Для каждого значения ρ , используя соотношения

(3.12.42) и (3.12.43), можно найти соответствующее значение скорости R , а используя соотношение (3.12.38) — значение $E_0(p)$. Значение $E(R)$ можно тогда найти по формуле $E(R) = E_0(p) - pR$. При этом скорость меняется в интервале от C до R_{kp} , где R_{kp} определяется соотношениями (3.12.42), (3.12.43) при $p = 1$. Для $R < R_{kp}$ функция $E(R)$ определяется соотношением $E(R) = E_0(1) - R$.

Пример 3.12.1. Рассмотрим построение экспоненты случайного кодирования для двоичного симметричного канала с вероятностью ошибки p . Для этого канала

$$E_0(p) = p - (1 + p) \log(p^{1/1+p} + (1 - p)^{1/1+p}).$$

Положим

$$\delta = \frac{p^{1/1+p}}{p^{1/1+p} + (1 - p)^{1/1+p}}.$$

Тогда для скоростей в диапазоне от R_{kp} до $C = 1 - h(p)$

$$E(R) = \delta \log \frac{\delta}{p} + (1 - \delta) \log \frac{1 - \delta}{1 - p} \triangleq h(\delta, p), \quad (3.12.44)$$

$$R = 1 - h(\delta). \quad (3.12.45)$$

Для скоростей в диапазоне от 0 до R_{kp}

$$E(R) = 1 - 2 \log(V\sqrt{p} + V\sqrt{1-p}) - R.$$

Значение критической скорости

$$R_{kp} = 1 - h\left(\frac{\sqrt{p}}{\sqrt{p} + \sqrt{1-p}}\right).$$

В табл. 3.12.1 приведены численные значения экспоненты случайного кодирования для ДСК с двумя значениями вероятности ошибки $p = 0,1$ и $p = 0,01$.

Таблица 3.12.1

p	$p = 0,1$		$p = 0,01$		Примечание
	R	$E(R)$	R	$E(R)$	
0	0,531	0	0,979	0	$R = C$
0,1	0,472	$2,88 \cdot 10^{-3}$	0,887	$1,64 \cdot 10^{-3}$	
0,2	0,421	$1,05 \cdot 10^{-2}$	0,851	$6,99 \cdot 10^{-3}$	
0,3	0,376	$2,17 \cdot 10^{-2}$	0,814	$1,63 \cdot 10^{-2}$	
0,4	0,337	$3,52 \cdot 10^{-2}$	0,775	$2,99 \cdot 10^{-2}$	
0,5	0,303	$5,04 \cdot 10^{-2}$	0,737	$4,73 \cdot 10^{-2}$	
0,6	0,274	$6,65 \cdot 10^{-2}$	0,699	$6,82 \cdot 10^{-2}$	
0,7	0,248	$8,32 \cdot 10^{-2}$	0,662	$9,23 \cdot 10^{-2}$	
0,8	0,226	$1,00 \cdot 10^{-1}$	0,626	$1,19 \cdot 10^{-1}$	
0,9	0,206	$1,17 \cdot 10^{-1}$	0,592	$1,48 \cdot 10^{-1}$	
1,0	0,189	$1,33 \cdot 10^{-1}$	0,559	$1,79 \cdot 10^{-1}$	
—	0	$3,22 \cdot 10^{-1}$	0	$7,38 \cdot 10^{-1}$	$R = R_{kp}$ $E(R) = E_0(1)$

§ 3.13*. Нижняя граница вероятности ошибки декодирования для дискретных каналов без памяти (граница сферической упаковки)

В предыдущем параграфе мы показали, что для любого дискретного канала без памяти средняя по ансамблю кодов \mathcal{G} вероятность ошибки декодирования удовлетворяет неравенству

$$\bar{\lambda} < 2^{-nE(R)}, \quad (3.13.1)$$

где показатель экспоненты $E(R)$ зависит только от скорости и переходных вероятностей канала и положителен для всех скоростей, меньших пропускной способности. Это фактически означает, что в ансамбле \mathcal{G} существует код, средняя вероятность ошибки которого удовлетворяет неравенству (3.13.1). Каждый отдельно взятый код может иметь вероятность ошибки, отличную от вероятности, полученной осреднением по всем кодам. В частности, в ансамбле могут быть коды, вероятность ошибки которых может быть существенно большей, чем правая часть (3.13.1), например, коды, имеющие большое количество совпадающих кодовых слов. С другой стороны, может оказаться, что в ансамбле всех кодов существует очень хороший код, вероятность ошибки которого существенно меньше правой части (3.13.1).

Пусть λ^* — наименьшая по всем кодам $G(n, R)$ средняя вероятность ошибки декодирования. Основная цель этого раздела состоит в том, чтобы показать, что

$$\lambda^* \geq 2^{-n(E_{sp}(R)+\epsilon)}, \quad (3.13.2)$$

где $E_{sp}(R)$ — некоторая функция, зависящая от скорости и переходных вероятностей канала, а ϵ — некоторая положительная величина, которая может быть выбрана сколь угодно малой при достаточно больших n . Более того, мы покажем, что для достаточно широкого диапазона скоростей, а именно для всех $R \geq R_{kp}$, функции $E(R)$ и $E_{sp}(R)$ совпадают. Это означает, что при $R \geq R_{kp}$ и при достаточно больших n показатели экспонент в оценках (3.13.1) и (3.13.2) сколь угодно близки друг к другу и, следовательно, в ансамбле не существует кодов, вероятность ошибки декодирования которых в дискретном канале без памяти была бы существенно меньшей, чем правая часть неравенства (3.13.1). С другой стороны, это означает, что при указанных выше условиях почти все коды в ансамбле имеют вероятность ошибки декодирования, равную правой части неравенства (3.13.1). Таким образом, при $R \geq R_{kp}$ верхняя и нижняя оценки вероятности ошибки в дискретном канале без памяти асимптотически совпадают (или экспоненциально точны).

3.13.1. Нижняя граница вероятности ошибки для ДСК. Прежде чем приступить к выводу нижней границы для вероятности

ошибки декодирования в произвольном дискретном канале без памяти, мы построим эту границу для ДСК. Приводимый ниже вывод обладает достаточной простотой и наглядностью, которых, к сожалению, лишен вывод для произвольного дискретного канала без памяти.

Введем вначале несколько вспомогательных понятий. Пусть $X = \{0, 1\}$ — двоичный алфавит и $\mathbf{x}_1, \mathbf{x}_2$ — две двоичные последовательности длины n . Расстоянием Хемминга между последовательностями \mathbf{x}_1 и \mathbf{x}_2 называется число $d(\mathbf{x}_1, \mathbf{x}_2)$ позиций, в которых эти последовательности различаются. Например, пусть $\mathbf{x}_1 = (1010011)$, $\mathbf{x}_2 = (1100010)$, тогда $d(\mathbf{x}_1, \mathbf{x}_2) = 3$. Пусть фиксирована последовательность (точка) $\mathbf{x}_0 \in X^n$. Сферой Хемминга с центром в точке \mathbf{x}_0 и радиусом t называется множество всех таких двоичных последовательностей $\mathbf{x} \in X^n$, для которых $d(\mathbf{x}_0, \mathbf{x}) \leq t$. Например, сфера Хемминга с центром в точке (011) и радиусом 2 содержит следующие 7 последовательностей из X^3 : (011), (111), (001), (010), (101), (000), (110); точка (100) находится на расстоянии 3 от центра и в данную сферу не входит.

Пусть вероятность ошибки в ДСК равна p , $0 < p < 1/2$. Рассмотрим некоторый код $G(n, R) = \{\mathbf{u}_1, A_1; \dots; \mathbf{u}_M, A_M\}$ для этого канала. Заметим, что в любом коде для ДСК слова представляют собой двоичные последовательности длины n , а решаютые области — подмножества множества всех 2^n двоичных последовательностей. Вероятность $p(\mathbf{y}|\mathbf{x})$ получения двоичной последовательности \mathbf{y} , если передавалась двоичная последовательность \mathbf{x} , определяется соотношением

$$p(\mathbf{y}|\mathbf{x}) = p^t(1-p)^{n-t},$$

где $i = d(\mathbf{x}, \mathbf{y})$. Учитывая это соотношение, можно записать следующее выражение для условной вероятности ошибки λ_j при передаче слова \mathbf{u}_j из кода $G(n, R)$:

$$\lambda_j = 1 - \sum_{\mathbf{y} \in A_j} p(\mathbf{y}|\mathbf{u}_j) = 1 - \sum_{i=0}^n k_i p^i (1-p)^{n-i},$$

где k_i — число таких последовательностей из решающей области A_j , которые отличаются от \mathbf{u}_j в i позициях. Так как $p < 1/2$, то $p^{i'}(1-p)^{n-i'} > p^{i''}(1-p)^{n-i''}$ при $i' < i''$.

Следовательно,

$$\lambda_j \geq 1 - \sum_{i=0}^t C_n^i p^i (1-p)^{n-i}, \quad (3.13.3)$$

где t — наибольшее целое, удовлетворяющее неравенству

$$|A_j| \geq \sum_{i=0}^t C_n^i. \quad (3.13.4)$$

Неравенства (3.13.3) и (3.13.4) отражают тот факт, что при фиксированном объеме решающей области A_j , наименьшую вероятность ошибки дает область, которая является сферой Хемминга с центром в кодовом слове \mathbf{u}_j .

Так как общее число выходных последовательностей канала равно 2^n , то найдется j_0 такое, что

$$|A_{j_0}| \leq \frac{2^n}{M} = 2^{n(1-R)}. \quad (3.13.5)$$

Обозначим через t_0 наибольшее целое число, удовлетворяющее неравенству (3.13.4) при $j = j_0$. Тогда из (3.13.4) и (3.13.5) имеем

$$2^{n(1-R)} \geq \sum_{i=0}^{t_0} C_n^i \geq C_n^{t_0}. \quad (3.13.6)$$

Обозначим t_0/n через τ и воспользуемся формулой Стирлинга для факториалов, чтобы оценить величину C_n^t . Учитывая задачу 1.6.3, можно записать

$$2^{n(1-R)} \geq C_n^{t_0} \geq \frac{1}{\sqrt{2\pi n\tau(1-\tau)}} 2^{nh(\tau)}$$

или

$$1 - h(\tau) \geq R - \epsilon_n, \quad (3.13.7)$$

где

$$\epsilon_n = \frac{1}{2n} \log 2\pi n\tau(1-\tau) \rightarrow 0, \quad n \rightarrow \infty. \quad (3.13.8)$$

Оценим теперь величину λ_{j_0} . Из (3.13.3), снова учитывая задачу 1.6.3, получим

$$\begin{aligned} \lambda_{j_0} &\geq \sum_{i=t_0+1}^n C_n^i p^i (1-p)^{n-i} \geq C_n^{t_0+1} p^{t_0+1} (1-p)^{n-t_0-1} \geq \\ &\geq \frac{1}{\sqrt{2\pi n\tau'(1-\tau')}} 2^{-n[-h(\tau')-\tau' \log p-(1-\tau') \log(1-p)]} = \\ &= [2\pi n\tau'(1-\tau')]^{-1/2} 2^{-nh(\tau', p)} = 2^{-n[h(\tau', p)+\epsilon_n]}, \end{aligned} \quad (3.13.9)$$

где

$$\tau' \triangleq (t_0 + 1)/n = \tau + 1/n \quad (3.13.10)$$

и

$$\begin{aligned} h(\tau, p) &\triangleq -h(\tau) - \tau \log p - (1-\tau) \log(1-p) = \\ &= \tau \log \tau + (1-\tau) \log(1-\tau) - \tau \log p - (1-\tau) \log(1-p) = \\ &= \tau \log \frac{\tau}{p} + (1-\tau) \log \frac{1-\tau}{1-p}. \end{aligned} \quad (3.13.11)$$

Так как $h(\tau, p)$ — непрерывная функция τ , то из (3.13.9) и (3.13.10) следует, что

$$\lambda_{j_0} \geq 2^{-n[h(\tau, p)+\epsilon'_n]}, \quad (3.13.12)$$

где ϵ'_n — положительная величина, которая может быть выбрана сколь угодно малой при больших n .

Обозначим через $E_{sp}(R)$ функцию, определяемую следующим образом:

$$E_{sp}(R) \triangleq h(\tau, p), \quad (3.13.13)$$

где τ — корень уравнения

$$1 - h(\tau) = R. \quad (3.13.14)$$

Легко видеть, что $E_{sp}(R)$ — непрерывная функция аргумента R . Поэтому из (3.13.7), (3.13.12), (3.13.13) и (3.13.14) следует, что Λ — максимальная вероятность ошибки декодирования любого кода $G(n, R)$ в ДСК удовлетворяет неравенству

$$\Lambda \triangleq \max_j \lambda_j \geq \lambda_{j_0} \geq 2^{-n[E_{sp}(R)+\epsilon''_n]}, \quad (3.13.15)$$

где ϵ''_n — положительная величина, которая может быть выбрана сколь угодно малой при больших n .

С помощью леммы 3.2.1 можно сделать аналогичное утверждение для средней вероятности ошибки декодирования. Из этой леммы следует, что если не существует кода $G(n, R)$ с максимальной вероятностью ошибки Λ , то не существует и кода $G(n, R')$ со средней вероятностью ошибки, меньшей чем $\Lambda/2$, где $R' = R + \frac{1}{n}$. Учитывая это замечание, из (3.13.15) получим, что для любого кода $G(n, R)$ средняя вероятность ошибки декодирования в ДСК удовлетворяет неравенству

$$\lambda \geq \frac{1}{2} \cdot 2^{-n[E_{sp}(R-\frac{1}{n})+\epsilon]} = 2^{-n[E_{sp}(R)+\epsilon]},$$

где ϵ — положительная величина, стремящаяся к нулю с ростом n .

Сравнивая (3.13.13) и (3.13.14) с (3.12.44) и (3.12.45), можно убедиться в том, что в случае ДСК функции $E_{sp}(R)$ и $E(R)$ совпадают при всех $R \geq R_{kp}$ и, следовательно, верхняя и нижняя оценки для вероятности ошибки декодирования в ДСК экспоненциально точны.

Заметим, что приведенный выше вывод нижней границы вероятности ошибки декодирования в ДСК основан на следующих двух утверждениях: 1. При фиксированном объеме решающей области наименьшую вероятность ошибки дает область, представляющая собой сферу Хемминга с центром в соответствующем кодовом слове. 2. Наибольшее число непересекающихся сфер радиуса t в множестве всех двоичных последовательностей длины n

не превосходит величины $2^n / \sum_{i=0}^t C_n^i$ и равно ей, когда достигается полное заполнение (упаковка) всего множества сферами радиуса t . В соответствии с этими утверждениями наименьшей вероятностью ошибки в ДСК обладает код, для которого все решающие области являются сферами Хемминга радиуса t , где t определяется из условия $2^{nR} \sum_{i=0}^t C_n^i = 2^n$. Такой код называется «сферически плотно упакованным» или просто «плотно упакованным». Использованный метод построения нижней границы заключается в построении оценки для гипотетического плотно упакованного кода, скорость которого равна скорости исходного кода. При данных n и R плотно упакованного кода почти во всех случаях не существует. Однако асимптотическая точность полученных оценок свидетельствует о том, что при больших n существуют почти плотно упакованные коды, для которых решающие области почти сферичны.

3.13.2. Коды с фиксированной композицией. Пусть X — конечное множество, X^n — множество последовательностей длины n с элементами из X и $\mathbf{x} = (x^{(1)}, \dots, x^{(n)})$ — некоторая последовательность из X^n . Обозначим через $k(x)$ количество элементов в последовательности \mathbf{x} , которые равны данному $x \in X$. Очевидно, $\sum_{x \in X} k(x) = n$ для любой последовательности $\mathbf{x} \in X^n$.

Определение 3.13.1. Вероятностный вектор $\tau = \{\tau(x)\}$, $\tau(x) = \frac{k(x)}{n}$, $x \in X$, называется *композицией последовательности \mathbf{x}* . Множество всех последовательностей из X^n , композиции которых совпадают и равны τ , называется *множеством последовательностей с фиксированной композицией τ* . Такое множество будет обозначаться символом X_τ^n .

Лемма 3.13.1. Пусть T — количество различных композиций τ для последовательностей $\mathbf{x} \in X^n$. Имеет место неравенство

$$T \leq (n+1)^{L-1}, \quad (3.13.16)$$

где L — число элементов в множестве X .

Доказательство. Пусть $X = \{x_1, x_2, \dots, x_L\}$. При фиксированном n каждая композиция τ взаимно однозначно определяется набором целых чисел $k_i = k(x_i)$, $i = \overline{1, L}$, таких, что $\sum_{i=1}^L k_i = n$. Таким образом, для того чтобы указать некоторую композицию, достаточно задать $L-1$ целых чисел k_1, k_2, \dots, k_{L-1} (k_L определяется как $k_L = n - \sum_{i=1}^{L-1} k_i$). Каждое из чисел k_i может

принимать одно из $(n+1)$ значений, от 0 до n . Поэтому количество различных наборов чисел k_1, k_2, \dots, k_{L-1} , определяющих различные композиции, удовлетворяет неравенству (3.13.16). Знак неравенства в (3.13.16) имеет место вследствие того, что среди различных наборов чисел k_1, k_2, \dots, k_{L-1} имеются и такие, что $\sum_{i=1}^{L-1} k_i > n$, т. е. которые не определяют никакой композиции.

Лемма доказана.

Определение 3.13.2. Код $G(n, R) = \{\mathbf{u}_1, A_1; \dots; \mathbf{u}_M, A_M\}$, $\mathbf{u}_i \in X^n$, $A_i \subseteq Y^n$, $i = 1, M$, $M = 2^{nR}$, называется кодом с фиксированной композицией τ , если все его кодовые слова имеют композицию τ : $\mathbf{u}_i \in X_{\tau}^n$, $i = 1, M$. Код с фиксированной композицией τ будет обозначаться через $G(n, R, \tau)$.

Лемма 3.13.2. Пусть для дискретного канала без памяти существует код $G(n, R)$ со средней вероятностью ошибки λ . Тогда существует композиция τ и код $G(n, R', \tau)$ со средней вероятностью ошибки λ' такой, что

$$R - (L-1) \frac{\log(n+1)}{n} \leq R' \leq R, \quad (3.13.17)$$

$$\lambda' \leq (n+1)^{L-1} \lambda, \quad (3.13.18)$$

где L — число элементов во входном алфавите канала.

Доказательство. Код $G(n, R) = \{\mathbf{u}_1, A_1; \mathbf{u}_2, A_2; \dots; \mathbf{u}_M, A_M\}$ можно рассматривать как объединение N кодов с фиксированными композициями:

$$G(n, R) = \bigcup_{j=1}^N G(n, R_j, \tau_j), \quad G(n, R_j, \tau_j) = \\ = \{\mathbf{u}_{1j}, A_{1j}; \dots; \mathbf{u}_{M_j j}, A_{M_j j}\}, \\ \mathbf{u}_{ij} \in X_{\tau_j}^n, \quad M_j = 2^{nR_j},$$

где $M = \sum_{j=1}^N M_j$ и

$$\{\mathbf{u}_1, \dots, \mathbf{u}_M\} = \bigcup_{j=1}^N \{\mathbf{u}_{1j}, \dots, \mathbf{u}_{M_j j}\}.$$

Из леммы 3.13.1 следует, что такое представление всегда возможно при некотором N , $1 \leq N \leq (n+1)^{L-1}$. Среди кодов $G(n, R_j, \tau_j)$ найдется по крайней мере один код, объем которого

$$M' = M_j \geq (n+1)^{L-1} M$$

и, следовательно, скорость которого удовлетворяет неравенствам (3.13.17). Обозначим скорость, вероятность ошибки и композицию этого кода через R' , λ' и τ' соответственно.

Пусть $\lambda^{(j)}$ — вероятность ошибки для кода $G(n, R_j, \tau_j)$, тогда вероятность ошибки λ для кода $G(n, R)$ может быть представлена следующим образом:

$$\lambda \triangleq \frac{1}{M} \sum_{i=1}^M \sum_{y \notin A_i} p(y | \mathbf{u}_i) = \frac{1}{M} \sum_{j=1}^N \sum_{i=1}^{M_j} \sum_{y \notin A_{ij}} p(y | \mathbf{u}_{ij}) = \\ = \sum_{j=1}^N \frac{M_j}{M} \frac{1}{M_j} \sum_{i=1}^{M_j} \sum_{y \notin A_{ij}} p(y | \mathbf{u}_{ij}) = \sum_{j=1}^N \frac{M_j}{M} \lambda^{(j)} \geq \\ \geq \frac{M'}{M} \lambda' \geq (n+1)^{-L+1} \lambda',$$

что совпадает с неравенством (3.13.18). Лемма доказана.

Лемма 3.13.2 позволяет строить нижнюю оценку для вероятности ошибки декодирования произвольных кодов, исходя из нижней оценки вероятности ошибки для кодов с фиксированной композицией. В п. 3.13.3 мы построим нижнюю оценку вероятности ошибки для таких кодов.

3.13.3. Нижняя граница для вероятности ошибки декодирования кода с фиксированной композицией. Пусть дан дискретный канал без памяти с матрицей переходных вероятностей $\mathbf{P} = \{p(y|x)\}$, $x \in X$, $y \in Y$. Введем в рассмотрение некоторую новую матрицу переходных вероятностей $\mathbf{P}^* = \{p^*(y|x)\}$, $x \in X$, $y \in Y$. Будем говорить, что матрица \mathbf{P}^* согласована с каналом, если из равенства $p(y|x) = 0$ следует, что $p^*(y|x) = 0$ для тех же сообщений $x \in X$, $y \in Y$. Пусть $\tau = \{\tau(x)\}$ — некоторая фиксированная композиция и $p_{XY}^* = \{p^*(x, y)\}$ — распределение вероятностей на XY , задаваемое следующим образом: $p^*(x, y) = \tau(x) p^*(y|x)$, $x \in X$, $y \in Y$. Введем также в рассмотрение среднюю взаимную информацию, вычисленную относительно распределения p_{XY}^* :

$$I_{p_{XY}^*}(X; Y) \triangleq \sum_x \sum_y \tau(x) p^*(y|x) \log \frac{p^*(y|x)}{\sum_x p^*(x, y)}, \quad (3.13.19)$$

и среднюю энтропию распределения \mathbf{P}^* относительно распределения \mathbf{P}

$$\mathcal{H}_{\tau}(\mathbf{P}^*, \mathbf{P}) \triangleq \mathcal{H}_{\tau}(p_{Y|X}^*, p_{Y|X}) = \sum_X \tau(x) \sum_Y p^*(y|x) \log \frac{p^*(y|x)}{p(y|x)}.$$

Пусть, кроме того, для фиксированного $\epsilon > 0$ и для каждой

последовательности $\mathbf{x} \in X_{\tau}^n$ определено множество

$$B(\mathbf{x}) \triangleq \left\{ \mathbf{y}: \left| \frac{1}{n} \log \frac{p^*(\mathbf{y}|\mathbf{x})}{p(\mathbf{y}|\mathbf{x})} - \mathcal{H}_{\tau}(P^*, P) \right| \leq \varepsilon, \right. \\ \left. \left| \frac{1}{n} \log \frac{p^*(\mathbf{y}|\mathbf{x})}{p^*(\mathbf{y})} - I_{P_{XY}^*}(X;Y) \right| \leq \frac{\varepsilon}{2} \right\}, \quad (3.13.20)$$

где

$$p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n p(y^{(i)}|x^{(i)}), \quad p^*(\mathbf{y}|\mathbf{x}) \triangleq \prod_{i=1}^n p^*(y^{(i)}|x^{(i)}), \quad (3.13.21)$$

$$p^*(\mathbf{y}) = \prod_{i=1}^n p^*(y^{(i)}), \quad p^*(y) \triangleq \sum_{\mathbf{x}} p^*(\mathbf{x}, y).$$

Имеет место следующее утверждение, показывающее, что для каждого $\mathbf{x} \in X_{\tau}^n$ множество $B(\mathbf{x})$ является высоковероятным множеством выходных последовательностей канала с переходными вероятностями $p^*(y|x)$.

Лемма 3.13.3. Для любого $\varepsilon > 0$ и для любого $\mathbf{x} \in X_{\tau}^n$ найдется n_0 такое, что при $n > n_0$

$$\Pr(B(\mathbf{x})) \triangleq \sum_{\mathbf{y} \in B(\mathbf{x})} p^*(\mathbf{y}|\mathbf{x}) \geq 1 - \varepsilon. \quad (3.13.22)$$

Доказательство. Рассмотрим функцию

$$l_y(x) \triangleq \log \frac{p^*(y|x)}{p(y|x)},$$

которая при каждом фиксированном $x \in X$ отображает множество Y в числовую ось. Эту функцию можно рассматривать как случайную величину на ансамбле $\{Y, p^*(y|x)\}$, ее математическое ожидание равно

$$Ml_y(x) = \sum_Y p^*(y|x) \log \frac{p^*(y|x)}{p(y|x)} = \mathcal{H}(P_{Y|x}^*, P_{Y|x}).$$

Из (3.13.21) следует, что

$$l_y(x) \triangleq \log \frac{p^*(y|x)}{p(y|x)} = \sum_{i=1}^n \log \frac{p^*(y^{(i)}|x^{(i)})}{p(y^{(i)}|x^{(i)})} \quad (3.13.23)$$

и при каждом фиксированном $\mathbf{x} \in X_{\tau}^n$ представляет собой сумму независимых случайных величин $l_{y^{(i)}}(x^{(i)})$, $i = \overline{1, n}$. При этом

$$Ml_y(x) = \sum_{i=1}^n Ml_{y^{(i)}}(x^{(i)}) = \sum_{i=1}^n \mathcal{H}(P_{Y|x^{(i)}}^*, P_{Y|x^{(i)}}).$$

Если $\mathbf{x} \in X_{\tau}^n$, то для каждого $x \in X$ в последней сумме имеется $n t(x)$ слагаемых, равных $\mathcal{H}(P_{Y|x}^*, P_{Y|x})$, поэтому

$$Ml_y(x) = n \sum_{x \in X} t(x) \mathcal{H}(P_{Y|x}^*, P_{Y|x}) = n \mathcal{H}_{\tau}(P^*, P).$$

В силу того, что значение $l_y(x)$ конечно для всех таких x, y , для которых $p^*(y|x) > 0$, случайная величина, определяемая функцией $l_y(x)$ на ансамбле $\{Y, p^*(y|x)\}$, имеет конечную дисперсию, которую мы будем обозначать как D_x , $D_x < \infty$. В сумме (3.13.23) $n t(x)$ величин имеют дисперсию D_x , $x \in X$. Обозначая через D_x дисперсию случайной величины $l_y(x)$, получим

$$D_x = \sum_{i=1}^n D_{x(i)} = n \sum_{x \in X} t(x) D_x \leq n \max_{x \in X} D_x = n D < \infty,$$

где $D \triangleq \max_{x \in X} D_x$.

Применим теперь неравенство Чебышева. В результате получим, что

$$\Pr(|l_y(x) - Ml_y(x)| \geq n\varepsilon) = \\ = \Pr\left(\left| \frac{1}{n} \log \frac{p^*(y|x)}{p(y|x)} - \mathcal{H}_{\tau}(P^*, P) \right| \geq n\varepsilon\right) \leq \frac{D}{n\varepsilon^2}. \quad (3.13.24)$$

Аналогичным образом можно получить неравенство

$$\Pr\left(\left| \frac{1}{n} \log \frac{p^*(y|x)}{p^*(y)} - I_{P_{XY}^*}(X;Y) \right| \geq \frac{\varepsilon}{2}\right) \leq \frac{4D_I}{n\varepsilon^2}, \quad (3.13.25)$$

где D_I — максимальная по x , $x \in X$, дисперсия случайной величины, определяемой функцией $I(x; y) \triangleq \log \frac{p^*(x|y)}{p^*(y)}$ на ансамбле $\{Y, p^*(y|x)\}$.

Пусть $B'(\mathbf{x})$ — множество всех $\mathbf{y} \in Y^n$, которые удовлетворяют первому неравенству в (3.13.20), а $B''(\mathbf{x})$ — множество всех таких $\mathbf{y} \in Y^n$, которые удовлетворяют второму неравенству в (3.13.20). Очевидно, что $B(\mathbf{x}) = B'(\mathbf{x}) \cap B''(\mathbf{x})$. Обозначая через $\bar{B}(\mathbf{x})$ дополнение к множеству $B(\mathbf{x})$, получим с учетом (3.13.24) и (3.13.25), что

$$\Pr(B(\mathbf{x})) = 1 - \Pr(\bar{B}(\mathbf{x})) = 1 - \Pr(\bar{B}'(\mathbf{x}) \cup \bar{B}''(\mathbf{x})) \geq \\ \geq 1 - \Pr(\bar{B}'(\mathbf{x})) - \Pr(\bar{B}''(\mathbf{x})) \geq 1 - \frac{D}{n\varepsilon^2} - \frac{4D_I}{n\varepsilon^2}. \quad (3.13.26)$$

Из (3.13.26) следует, что для любого $\varepsilon > 0$ и любого $\mathbf{x} \in X_{\tau}^n$ найдется n_0 такое, что при $n > n_0$ выполняется (3.13.22). Лемма доказана.

Докажем теперь теорему о нижней границе вероятности ошибки для кодов с фиксированной композицией.

Теорема 3.13.1. Пусть фиксированы дискретный канал без памяти с матрицей переходных вероятностей $\mathbf{P} = \{p(y|x)\}$, $x \in X$, $y \in Y$, код $G(n, R, \tau)$ для этого канала с композицией τ и некоторая вспомогательная матрица переходных вероятностей $\mathbf{P}^* = \{p^*(y|x)\}$, согласованная с каналом (такая, что $p^*(y|x) = 0$ при $p(y|x) = 0$). Тогда для любого $\epsilon > 0$, любого R такого, что

$$R \geq I_{p_{XY}^*}(X; Y) + \epsilon, \quad (3.13.27)$$

где $I_{p_{XY}^*}(X; Y)$ определяется соотношением (3.13.19), и любого $\epsilon' > 0$ найдется n_0 такое, что при $n > n_0$ выполняется следующее неравенство для средней вероятности ошибки декодирования λ_τ :

$$\lambda_\tau \geq 2^{-n[\mathcal{H}_\tau(\mathbf{P}^*, \mathbf{P}) + \epsilon']}. \quad (3.13.28)$$

Доказательство. Пусть $G(n, R, \tau) = \{\mathbf{u}_1, A_1; \dots, \mathbf{u}_M, A_M\}$, $\mathbf{u}_i \in X_\tau^n$, $A_i \subseteq Y^n$, $i = 1, M$, $M = 2^{nR}$. Тогда среднюю вероятность ошибки λ_τ этого кода можно оценить следующим образом:

$$\lambda_\tau \triangleq \frac{1}{M} \sum_{i=1}^M \sum_{y \in \bar{A}_i} p(y|\mathbf{u}_i) \geq \frac{1}{M} \sum_{i=1}^M \sum_{y \in \bar{A}_i \cap B(\mathbf{u}_i)} p(y|\mathbf{u}_i), \quad (3.13.29)$$

где $B(\mathbf{u}_i)$ — множество, определяемое для данного ϵ соотношением (3.13.20). Из (3.13.20) (первое неравенство) имеем для каждого $y \in B(\mathbf{u}_i)$

$$p(y|\mathbf{u}_i) \geq p^*(y|\mathbf{u}_i) 2^{-n[\mathcal{H}_\tau(\mathbf{P}^*, \mathbf{P}) + \epsilon]}. \quad (3.13.30)$$

Подставляя (3.13.30) в (3.13.29), получим

$$\lambda_\tau \geq 2^{-n[\mathcal{H}_\tau(\mathbf{P}^*, \mathbf{P}) + \epsilon']} \frac{1}{M} \sum_{i=1}^M \sum_{y \in \bar{A}_i \cap B(\mathbf{u}_i)} p^*(y|\mathbf{u}_i). \quad (3.13.31)$$

Используем соотношение $\bar{A} \cap B = B \setminus (A \cap B)$. Тогда получим, что

$$\begin{aligned} \sum_{y \in \bar{A}_i \cap B(\mathbf{u}_i)} p^*(y|\mathbf{u}_i) &= \sum_{y \in B(\mathbf{u}_i)} p^*(y|\mathbf{u}_i) - \sum_{y \in A_i \cap B(\mathbf{u}_i)} p^*(y|\mathbf{u}_i) = \\ &= \Pr(B(\mathbf{u}_i)) - \sum_{y \in A_i \cap B(\mathbf{u}_i)} p^*(y|\mathbf{u}_i). \end{aligned} \quad (3.13.32)$$

Из (3.13.20) (второе неравенство) следует, что для всех $y \in B(\mathbf{u}_i)$

$$p^*(y|\mathbf{u}_i) \leq p^*(y) 2^{-n[\mathcal{H}_{\mathbf{P}^*}(X; Y) + \frac{\epsilon}{2}]}. \quad (3.13.33)$$

Из (3.13.27) и (3.13.33) получим теперь

$$\begin{aligned} \frac{1}{M} \sum_{i=1}^M \sum_{y \in \bar{A}_i \cap B(\mathbf{u}_i)} p^*(y|\mathbf{u}_i) &\leq \\ &\leq 2^{-n(R - \mathcal{H}_{\mathbf{P}^*}(X; Y) - \frac{\epsilon}{2})} \sum_{i=1}^M \sum_{y \in \bar{A}_i \cap B(\mathbf{u}_i)} p^*(y) \leq \\ &\leq 2^{-n\frac{\epsilon}{2}} \sum_{i=1}^M \sum_{y \in \bar{A}_i} p^*(y) = 2^{-n\frac{\epsilon}{2}} \sum_{y \in \bigcup_{i=1}^M \bar{A}_i} p^*(y) \leq 2^{-n\frac{\epsilon}{2}}, \end{aligned} \quad (3.13.34)$$

где последнее равенство в (3.13.34) следует из того, что решающие множества $\{\bar{A}_i\}$ не пересекаются, а последнее неравенство — из того, что

$$\sum_{y \in \bigcup_{i=1}^M \bar{A}_i} p^*(y) \leq 1.$$

Из (3.13.31), (3.13.32), (3.13.34) и леммы 3.13.3 следует окончательно, что

$$\lambda_\tau \geq 2^{-n[\mathcal{H}_\tau(\mathbf{P}^*, \mathbf{P}) + \epsilon]} \left(1 - \epsilon - 2^{-n\frac{\epsilon}{2}}\right).$$

Для любого $\epsilon' > 0$ можно так выбрать ϵ и n_0 , что при $n > n_0$ будет выполняться неравенство (3.13.28). Теорема доказана.

Доказанная теорема вместе с леммой 3.13.2 дает возможность построить нижнюю границу вероятности ошибки декодирования произвольного кода в дискретном канале без памяти. Заметим сначала, что неравенство (3.13.28) справедливо для любой стохастической матрицы \mathbf{P}^* , для которой выполняется условие (3.13.27). Следовательно, вероятность ошибки декодирования любого кода $G(n, R, \tau)$ с фиксированной композицией τ удовлетворяет неравенству

$$\lambda_\tau \geq \max_{\{\mathbf{P}^*\}} 2^{-n[\mathcal{H}_\tau(\mathbf{P}^*, \mathbf{P}) + \epsilon]}, \quad (3.13.35)$$

где максимум разыскивается по всем стохастическим матрицам \mathbf{P}^* , согласованным с каналом и удовлетворяющим условию

$$R \geq I_{p_{XY}^*}(X; Y) + \epsilon'', \quad (3.13.36)$$

где $p^*(x, y) = p^*(y|x) \tau(x)$, причем ϵ' , ϵ'' — положительные величины, которые могут быть выбраны сколь угодно малыми

при достаточно больших n . Другими словами, не существует кода с композицией τ , скорости которого удовлетворяла бы условию (3.13.36), а вероятность ошибки была бы меньше правой части соотношения (3.13.35). Если рассматривать все коды $G(n, R, \tau)$ со всем возможными фиксированными композициями, то не существует кода, вероятность ошибки которого в дискретном канале без памяти $\{XY, p(y|x)\}$ была бы меньше, чем

$$\tilde{\lambda} \triangleq \min_{\{\tau\}} \max_{\{P^*\}} 2^{-n} [\mathcal{H}_\tau(P^*, P) + \varepsilon], \quad (3.13.37)$$

где минимум разыскивается по всем композициям τ , а максимум — по всем стохастическим матрицам P^* , согласованным с каналом, для которых удовлетворяется условие (3.13.36) для данной композиции τ .

Заметим, что множество всех композиций зависит от n , так как компонентами вектора τ являются дроби со знаменателем n .

Для того чтобы избавиться от этой зависимости в дальнейшем, мы будем рассматривать в (3.13.37) вместо минимума по $\{\tau\}$ минимум по всем вероятностным векторам, задающим распределение вероятностей на X . При такой замене неравенство ослабляется, однако при больших n ослабление не является существенным, так как в этом случае множество всех композиций достаточно хорошо аппроксимирует множество всех вероятностных векторов.

Положим

$$E_{sp}(R, Q) \triangleq \min_{P^*: \mid P_{XY}^* \mid \leq R} \mathcal{H}_Q(P^*, P) \quad (3.13.38)$$

и

$$\begin{aligned} E_{sp}(R) &\triangleq \max_{\{Q\}} E_{sp}(R, Q) = \\ &= \max_{\{Q\}} \min_{P^*: \mid P_{XY}^* \mid \leq R} \mathcal{H}_Q(P^*, P), \end{aligned} \quad (3.13.39)$$

где минимум разыскивается по всем вспомогательным матрицам переходных вероятностей $P^* = \{p^*(y|x)\}$, согласованным с каналом, и $p^*(x, y) = Q(x)p^*(y|x)$, $x \in X$, $y \in Y$. Имеет место следующее утверждение.

Теорема 3.13.2. Для любого $\varepsilon > 0$ найдется n_0 такое, что при $n > n_0$ средняя вероятность ошибки декодирования λ любого кода $G(n, R)$ удовлетворяет неравенству

$$\lambda \geq 2^{-n} [E_{sp}(R) + \varepsilon]. \quad (3.13.40)$$

Доказательство. Из леммы 3.13.2 следует, что для любого кода $G(n, R)$ найдется композиция τ такая, что

$$\lambda \geq (n+1)^{-L+1} \lambda'_\tau, \quad (3.13.41)$$

где λ'_τ — средняя вероятность ошибки для кода $G(n, R', \tau)$, имеющего композицию τ и скорость $R' \geq R - n^{-1}(L-1)\log \times (n+1)$. Из неравенств (3.13.35), (3.13.36), (3.13.41) с учетом обозначения (3.13.38) следует, что для любых положительных $\varepsilon', \varepsilon''$ при достаточно больших n справедливо следующее неравенство:

$$\begin{aligned} \lambda &\geq (n+1)^{-L+1} 2^{-n} [E_{sp}((R-\varepsilon''-\varepsilon_n), \tau) + \varepsilon'] = \\ &= 2^{-n} [E_{sp}((R-\varepsilon''-\varepsilon_n), \tau) + \varepsilon' + \varepsilon_n] = 2^{-n} [E_{sp}(R, \tau) + \varepsilon], \end{aligned} \quad (3.13.42)$$

где $\varepsilon_n \triangleq n^{-1}(L-1)\log(n+1)$, $\varepsilon \geq \varepsilon' + \varepsilon_n$ и последнее равенство получено в силу непрерывности и невозрастания $E_{sp}(R, \tau)$ как функции от R . Величина ε положительна, но может быть сделана произвольно малой за счет выбора достаточно большого n . Заметим далее, что неравенство (3.13.42) справедливо лишь при определенной композиции, зависящей от кода $G(n, R)$. Поэтому для получения нижней границы вероятности ошибки, справедливой для любого кода, следует взять минимум правой части (3.13.42) по всем композициям τ . Тогда, учитывая обозначение (3.13.39), получим, что для любого $\varepsilon > 0$ найдется n_0 такое, что при $n > n_0$ средняя вероятность ошибки декодирования любого кода $G(n, R)$ удовлетворяет неравенству

$$\lambda \geq \min_{\{\tau\}} 2^{-n} [E_{sp}(R, \tau) + \varepsilon] \geq \min_{\{Q\}} 2^{-n} [E_{sp}(R, Q) + \varepsilon] = 2^{-n} [E_{sp}(R) + \varepsilon],$$

что совпадает с (3.13.40). Теорема доказана. ■

Из сравнения (3.13.13) и (3.13.14) с (3.13.39) можно заключить, что показатель экспоненты в нижней границе для ДСК, полученный в п. 3.13.1, также может быть представлен в форме (3.13.39). При этом для случая ДСК максимум по $\{Q\}$ достигается при выборе равномерного распределения вероятностей на X : $Q = (1/2, 1/2)$, а минимум по $p^*(y|x)$ достигается при выборе следующих переходных вероятностей: $p^*(y|x) = \tau = t_0/n$, если $x \neq y$, и $p^*(y|x) = 1 - \tau$, если $x = y$. Функцию $E_{sp}(R)$ (см. (3.13.39)), частным случаем которой является показатель экспоненты сферической упаковки для ДСК, также принято называть экспонентой сферической упаковки (*sp* — сокращение от английского «sphere packing» — «сферическая упаковка»). В следующем пункте мы покажем, что, как и в случае ДСК, экспоненты случайного кодирования $E(R)$ и сферической упаковки $E_{sp}(R)$ совпадают при всех скоростях $R \geq R_{kp}$ для произвольного дискретного канала без памяти.

3.13.4. Совместное рассмотрение экспонент случайного кодирования и сферической упаковки. Наша цель заключается в том, чтобы найти общую форму, в которой могут быть представлены

обе экспоненты — случайного кодирования и сферической упаковки. Затем, исходя из этого общего представления, можно будет найти условия, при которых экспоненты совпадают.

Будем отправляться от определений (3.13.38) и (3.13.39). Напомним, что в этих определениях $\mathbf{P} = \{p(y|x)\}$, $x \in X$, $y \in Y$ — матрица переходных вероятностей канала, $\mathbf{P}^* = \{p^*(y|x)\}$, $x \in X$, $y \in Y$ — некоторая вспомогательная согласованная с каналом матрица переходных вероятностей и $\mathbf{P}_{XY}^* = \{p^*(x,y) = Q(x)p^*(y|x)\}$, $x \in X$, $y \in Y$.

Введем в рассмотрение еще одну функцию

$$\tilde{E}(R, \mathbf{Q}) \triangleq \max_{\rho \geq 0} (E_0(\rho, \mathbf{Q}) - \rho R). \quad (3.13.43)$$

Заметим, что определения функций $E(R, \mathbf{Q})$ (см. (3.12.29)) и $\tilde{E}(R, \mathbf{Q})$ (см. (3.13.43)) различаются только областью изменения параметра ρ : в первом случае ρ находится в интервале $[0, 1]$, во втором — $\rho \geq 0$. Очевидно, что $\tilde{E}(R, \mathbf{Q}) = E(R, \mathbf{Q})$ при $R \geq R_{kp}$ и $\tilde{E}(R, \mathbf{Q}) \geq E(R, \mathbf{Q})$ при $R < R_{kp}$. Как было показано ранее, при $R < R_{kp}$ функция $E(R, \mathbf{Q})$ представляет собой прямую с наклоном -1 , касательную к $\tilde{E}(R, \mathbf{Q})$ в точке $R = R_{kp}$. Значение ρ , максимизирующее правую часть соотношения (3.13.43), как и раньше, определяется как корень уравнения (3.12.30). Поэтому график функции $\tilde{E}(R, \mathbf{Q})$ может быть построен таким же образом, как и график функции $E(R, \mathbf{Q})$ (см. § 3.12). На рис. 3.12.2 показаны обе функции $E(R, \mathbf{Q})$ и $\tilde{E}(R, \mathbf{Q})$.

З а м е ч а н и е. Пусть $R(\rho)$ — значение скорости при данном ρ , которое удовлетворяет уравнению (3.12.30). Пусть $R_\infty \triangleq \lim_{\rho \rightarrow \infty} R(\rho)$. Если $R_\infty > 0$, то $\tilde{E}(R, \mathbf{Q}) = \infty$ при $R \leq R_\infty$. Отметим, что это возможно только для таких каналов, для которых $p(y|x) = 1$ при некоторых $x \in X$, $y \in Y$. Детальное обсуждение свойств таких каналов выходит за рамки данной книги.

Наша ближайшая задача состоит в установлении связи между функциями $\tilde{E}(R, \mathbf{Q})$ и $E_{sp}(R, \mathbf{Q})$, а также между функциями $\tilde{E}(R) \triangleq \max_{\{\mathbf{Q}\}} \tilde{E}(R, \mathbf{Q})$ и $E_{sp}(R)$.

Л е м м а 3.13.4. Пусть \mathbf{p}_{XY}^* — произвольное распределение вероятностей на XY и $\mathbf{p}_X^*, \mathbf{p}_Y^*$ — распределения вероятностей на X, Y , которые порождаются распределением \mathbf{p}_{XY}^* , т. е. $p^*(x) = \sum_Y p^*(x, y)$, $p^*(y) = \sum_X p^*(x, y)$. Пусть $v^*(x|y) = p^*(x, y) / p^*(y)$ и $\mathbf{v}_{X|y}^*$ — соответствующий вероятностный вектор. Пусть, кроме того, $p(y|x)$ — распределения вероятностей, за-

дающие канал, $\mathbf{Q} = \{Q(x)\}$ — произвольное распределение вероятностей на X и $\mathbf{p}_{XY} = \{p(y|x) = p(y|x)Q(x)\}$. Тогда

$$\tilde{E}(R, \mathbf{Q}) = \min_{\mathbf{p}_{XY}^*: \mathcal{H}_{\mathbf{p}_Y^*}(\mathbf{v}_{X|y}^*, \mathbf{Q}) \leq R} \mathcal{H}(\mathbf{p}_{XY}^*, \mathbf{p}_{XY}). \quad (3.13.44)$$

Если $R > R_\infty$, то минимум в (3.13.44) достигается при $\mathbf{p}_{XY}^* = \mathbf{p}_{XY}$, где распределение \mathbf{p}_{XY} определяется соотношениями (3.12.22) при таком значении ρ , что $\partial E_0(\rho, \mathbf{Q}) / \partial \rho = R$.

Д о к а з а т е л ь с т в о. Для доказательства леммы достаточно рассматривать только такие значения R , что $R_\infty \leq R \leq I(X; Y)$, где средняя взаимная информация вычислена относительно распределения \mathbf{p}_{XY} . Действительно, при $R > I(X; Y)$ можно положить $\mathbf{p}_{XY}^* = \mathbf{p}_{XY}$, так как в этом случае $\mathcal{H}_{\mathbf{p}_Y^*}(\mathbf{V}_{X|y}^*, \mathbf{Q}) = I(X; Y)$. Тогда $\mathcal{H}(\mathbf{p}_{XY}^*, \mathbf{p}_{XY}) = \tilde{E}(R, \mathbf{Q}) = 0$. Если же $R \leq R_\infty$ и $R_\infty > 0$, то в соответствии с замечанием, предшествующим формулировке теоремы, левая часть соотношения (3.13.44) обращается в бесконечность. Правая часть также обращается в бесконечность при условии, что обе части (3.13.44) совпадают при всех $R \geq R_\infty$, поскольку правая часть является монотонно невозрастающей функцией R , а левая — непрерывная функция R .

Пусть зафиксированы $R \in [R_\infty, I(X; Y)]$ и распределение \mathbf{p}_{XY}^* такое, что $\mathcal{H}_{\mathbf{p}_Y^*}(\mathbf{V}_{X|y}^*, \mathbf{Q}) \leq R$, и пусть параметр ρ удовлетворяет уравнению (3.12.30). Пусть $\tilde{p}(x, y)$ и $\tilde{v}(x|y)$, $x \in X$, $y \in Y$ — распределения вероятностей, определенные соотношениями (3.12.21)–(3.12.23). Тогда

$$\begin{aligned} E_0(\rho, \mathbf{Q}, \tilde{V}) &\triangleq -\log \sum_X \sum_Y p(y|x) Q^{1+\rho} (x) \tilde{v}^{-\rho} (x|y) = \\ &= -\sum_X \sum_Y p^*(x, y) \log \sum_X \sum_Y p(y|x) Q^{1+\rho} (x) \tilde{v}^{-\rho} (x|y) = \\ &= \sum_X \sum_Y p^*(x, y) \log \frac{p(y|x) Q^{1+\rho} (x) \tilde{v}^{-\rho} (x|y)}{\sum_X \sum_Y p(y|x) Q^{1+\rho} (x) \tilde{v}^{-\rho} (x|y)} = \\ &\quad -\sum_X \sum_Y p^*(x, y) \log p(y|x) Q^{1+\rho} (x) \tilde{v}^{-\rho} (x|y) = \\ &= \sum_X \sum_Y p^*(x, y) \log \tilde{p}(y|x) - \sum_X \sum_Y p^*(x, y) \log Q(x) p(y|x) + \\ &\quad + \rho \sum_X \sum_Y p^*(x, y) \log \frac{\tilde{v}(x|y)}{Q(x)}. \end{aligned}$$

Так как для функции $\tilde{v}(x|y)$ имеет место равенство $E_0(\rho, \mathbf{Q}, \tilde{\mathbf{V}}) = E_0(\rho, \mathbf{Q})$, то, добавляя и вычитая в правой части последнего выражения величину

$$\sum_x \sum_y p^*(x, y) \log p^*(x, y) + \rho \sum_x \sum_y p^*(x, y) \log v^*(x|y),$$

получим

$$\begin{aligned} E_0(\rho, \mathbf{Q}) &= \mathcal{H}(p_{XY}^*, p_{XY}) + \rho \mathcal{H}_{p_Y^*}(v_{X|Y}^*, \mathbf{Q}) - \\ &\quad - \mathcal{H}(p_{XY}^*, \tilde{p}_{XY}) - \rho \mathcal{H}_{p_Y^*}(V^*, \tilde{\mathbf{V}}), \end{aligned} \quad (3.13.45)$$

где $V^* = \{v^*(x|y)\}$, $\tilde{V} = \{\tilde{v}(x|y)\}$, $x \in X$, $y \in Y$, и

$$\mathcal{H}_{p_Y^*}(V^*, \tilde{V}) \triangleq \sum_x \sum_y p^*(x, y) \log \frac{v^*(x|y)}{\tilde{v}(x|y)}.$$

Из условий выбора p_{XY}^* и ρ , а также из неотрицательности энтропии одного распределения относительно другого, получим, что

$$\begin{aligned} \tilde{E}(R, \mathbf{Q}) &= E_0(\rho, \mathbf{Q}) - \rho R \leq \mathcal{H}(p_{XY}^*, p_{XY}) + \\ &\quad + \rho \mathcal{H}_{p_Y^*}(v_{X|Y}^*, \mathbf{Q}) - \rho R \leq \mathcal{H}(p_{XY}^*, p_{XY}). \end{aligned} \quad (3.13.46)$$

Заметим, что оба неравенства в последнем соотношении переходят в точные равенства, если $p_{XY}^* = \tilde{p}_{XY}$. Действительно, при этом $v_{X|Y}^* = \tilde{v}_{X|Y}$, кроме того, из (3.12.30) и (3.12.24) следует, что $\mathcal{H}_{p_Y^*}(\tilde{v}_{X|Y}, \mathbf{Q}) = R$. Такой выбор распределения p_{XY}^* удовлетворяет условию, при котором разыскивается минимум в (3.13.44). Следовательно, из (3.13.46) вытекает утверждение леммы. Лемма доказана.

Л е м м а 3.13.5. Функция $E_{sp}(R, \mathbf{Q})$, определяемая соотношением (3.13.38), представима в виде

$$E_{sp}(R, \mathbf{Q}) = \max_{\rho \geq 0} (E_0^*(\rho, \mathbf{Q}) - \rho R), \quad (3.13.47)$$

где

$$E_0^*(\rho, \mathbf{Q}) \triangleq \max_{(\mathbf{Q}')} (E_0(\rho, \mathbf{Q}') - (1 + \rho) \mathcal{H}(\mathbf{Q}, \mathbf{Q}')), \quad (3.13.48)$$

причем максимум отыскивается по всем таким распределениям вероятностей \mathbf{Q}' , что $Q'(x) = 0$, если $Q(x) = 0$ для того же $x \in X$.

Д о к а з а т е л ь с т в о. Положим

$$\begin{aligned} E_0^*(\rho, \mathbf{Q}, \mathbf{Q}') &\triangleq E_0(\rho, \mathbf{Q}') - (1 + \rho) \mathcal{H}(\mathbf{Q}, \mathbf{Q}') = \\ &= -\log \sum_Y \left(\sum_X Q'(x) p^{1/\rho+1}(y|x) \right)^{1+\rho} - (1 + \rho) \sum_X Q(x) \log \frac{Q(x)}{Q'(x)}. \end{aligned}$$

Далее будем предполагать, что $Q(x) > 0$ для всех $x \in X$. Это предположение не нарушает общности, так как всегда можно заменить множество X , по которому производится суммирование, на такое его подмножество X' , что $Q(x) > 0$ для всех $x \in X'$ и $\sum_{X'} Q(x) = 1$. Найдем частную производную $E_0^*(\rho, \mathbf{Q}, \mathbf{Q}')$ по $Q'(x)$:

$$\begin{aligned} \frac{\partial E_0^*(\rho, \mathbf{Q}, \mathbf{Q}')}{\partial Q'(x)} &= \\ &= -(1 + \rho) \log e \left(\frac{\sum_Y p^{1/\rho+1}(y|x) \left(\sum_X Q'(x) p^{1/\rho+1}(y|x) \right)^\rho}{\sum_Y \left(\sum_X Q'(x) p^{1/\rho+1}(y|x) \right)^{1+\rho}} - \frac{Q(x)}{Q'(x)} \right). \end{aligned} \quad (3.13.49)$$

Обозначим через $Q^*(x)$, $x \in X$, распределение вероятностей, на котором достигается максимум в выражении (3.13.48). Из теоремы Куна—Таккера и из (3.13.49) следует, что оптимизирующее распределение $Q^*(x)$ должно удовлетворять следующей системе соотношений:

$$\begin{aligned} \frac{\sum_Y p^{1/\rho+1}(y|x) \left[\sum_X Q^*(x) p^{1/\rho+1}(y|x) \right]^\rho}{\sum_Y \left[\sum_X Q^*(x) p^{1/\rho+1}(y|x) \right]^{1+\rho}} - \frac{Q(x)}{Q^*(x)} &\geq \\ &\geq \frac{c}{(1 + \rho) \log e}, \quad x \in X, \end{aligned} \quad (3.13.50)$$

где c — неопределенный множитель Лагранжа и знак неравенства может иметь место только для тех $x \in X$, для которых $Q^*(x) = 0$. Заметим, однако, что $\mathcal{H}(\mathbf{Q}, \mathbf{Q}^*) = \infty$, если для некоторого $x \in X$, $Q(x) > 0$, но $Q^*(x) = 0$. Поэтому соотношения (3.13.50) выполняются со знаком равенства для всех $x \in X$. Домножая левую и правую части неравенства (3.13.50) на $Q^*(x)$ и суммируя по X , получим, что $c/(1 + \rho) \log e = 0$. Отсюда следует, что необходимые условия максимума в (3.13.48) задаются следующей системой уравнений:

$$\frac{\sum_Y Q^*(x) p^{1/\rho+1}(y|x) \left[\sum_X Q^*(x) p^{1/\rho+1}(y|x) \right]^\rho}{\sum_Y \left[\sum_X Q^*(x) p^{1/\rho+1}(y|x) \right]^{1+\rho}} = Q(x), \quad x \in X. \quad (3.13.51)$$

Легко видеть, что, как и в теореме 3.13.1, эти условия являются и достаточными.

Обозначим теперь через \tilde{p}_{XY} распределение вероятностей на XY , определяемое соотношением (3.12.22) при условии, что $Q(x)$ заменено на $Q^*(x)$. Тогда из (3.13.51) получим, что

$$\tilde{p}^*(x) \triangleq \sum_Y \tilde{p}^*(x, y) = Q(x). \quad (3.13.52)$$

Далее из (3.12.24), (3.12.25) и (3.13.48) следует, что

$$\begin{aligned} \frac{\partial E_0^*(\rho, Q)}{\partial \rho} &= I_{\tilde{p}_{XY}}(X; Y) \geq 0, \\ \frac{\partial^2 E_0^*(\rho, Q)}{\partial \rho^2} &= \frac{\partial^2 E_0(\rho, Q^*)}{\partial \rho^2} \leq 0. \end{aligned}$$

Поэтому значение параметра ρ , на котором достигается максимум в (3.13.47), является корнем уравнения

$$I_{\tilde{p}_{XY}^*}(X; Y) = R. \quad (3.13.53)$$

Введем следующие обозначения. Пусть $\mathbf{P}^* = \{p^*(y|x)\}$, $x \in X, y \in Y$, — матрица переходных вероятностей, на которой достигается минимум в (3.13.38). Пусть p_{XY}^* и p_{XY} — распределения вероятностей на XY такие, что $p^*(x, y) = Q(x)p^*(y|x)$ и $p(x, y) = Q^*(x)p(y|x)$, $x \in X, y \in Y$, где $p(y|x)$ — переходные вероятности канала. Пусть p_Y^* — распределение вероятностей на Y такое, что $p^*(y) = \sum_x p^*(x, y)$, и $V^* = \{v^*(x|y)\}$, $x \in X, y \in Y$, — матрица переходных вероятностей, элементы которой определяются из условия $v^*(x|y) = p^*(x, y)/p^*(y)$. Пусть, наконец, $\tilde{V}^* = \{\tilde{v}^*(x|y)\}$, $x \in X, y \in Y$, — еще одна матрица переходных вероятностей, элементы которой определяются из условия $\tilde{v}^*(x|y) = \tilde{p}^*(x, y)/\tilde{p}^*(y)$. Тогда из (3.13.45) имеем

$$\begin{aligned} E_0(\rho, Q^*) &= \mathcal{H}_Q(\mathbf{P}^*, \mathbf{P}) + \mathcal{H}(Q, Q^*) + \rho [I_{p_{XY}^*}(X; Y) + \mathcal{H}(Q, Q^*)] - \\ &\quad - \mathcal{H}(p_{XY}^*, \tilde{p}_{XY}^*) - \rho \mathcal{H}_{p_Y^*}(V^*, \tilde{V}^*). \end{aligned} \quad (3.13.54)$$

Учитывая, что распределение Q^* доставляет максимум правой части (3.13.48), из (3.13.54) и (3.13.48) получим, что

$$\begin{aligned} E_0^*(\rho, Q) &= \mathcal{H}_Q(\mathbf{P}^*, \mathbf{P}) + \rho I_{p_{XY}^*}(X; Y) - \mathcal{H}(p_{XY}^*, \tilde{p}_{XY}^*) - \\ &\quad - \rho \mathcal{H}_{p_Y^*}(V^*, \tilde{V}^*). \end{aligned} \quad (3.13.55)$$

Пусть параметр ρ удовлетворяет уравнению (3.13.53). Тогда из (3.13.55) следует, что

$$E_0^*(\rho, Q) - \rho R \leq \mathcal{H}_Q(\mathbf{P}^*, \mathbf{P}) + \rho I_{p_{XY}^*}(X; Y) - \rho R \leq \mathcal{H}_Q(\mathbf{P}^*, \mathbf{P}). \quad (3.13.56)$$

Оба неравенства в (3.13.56) превращаются в равенства при $p_{XY}^* = \tilde{p}_{XY}^*$ или, как это следует из (3.13.52), при $\mathbf{P}^* = \tilde{\mathbf{P}}^*$. Из (3.13.52) и (3.13.53) вытекает, что матрица $\tilde{\mathbf{P}}^*$ доставляет минимум в (3.13.38). Поэтому из (3.13.56) следует утверждение леммы. Лемма доказана.

Положим

$$\tilde{E}(R) \triangleq \max_{\{Q\}} \tilde{E}(R, Q). \quad (3.13.57)$$

Имеет место следующее утверждение.

Теорема 3.13.3. Для произвольного дискретного канала без памяти

$$\tilde{E}(R, Q) \leq E_{sp}(R, Q) \quad (3.13.58)$$

для любого распределения вероятностей Q на X ; кроме того,

$$\tilde{E}(R) = E_{sp}(R). \quad (3.13.59)$$

Доказательство. Пусть фиксирована скорость R и пусть \mathbf{P}^* — матрица переходных вероятностей, на которой достигается минимум в (3.13.38). Положим $p^*(x, y) = Q(x)p^*(y|x)$, $x \in X, y \in Y$. Определенное таким образом распределение p_{XY}^* удовлетворяет условию, при котором отыскивается минимум в (3.13.44). Действительно, в этом случае

$$\mathcal{H}_{p_Y^*}(V_{X|Y}^*, Q) = I_{p_{XY}^*}(X; Y) + \mathcal{H}(Q, Q^*) = I_{p_{XY}^*}(X; Y) \leq R. \quad (3.13.60)$$

Кроме того,

$$\begin{aligned} \mathcal{H}(p_{XY}^*, p_{XY}) &\triangleq \sum_X \sum_Y Q(x) p^*(y|x) \log \frac{Q(x) p^*(y|x)}{Q(x) p(y|x)} = \\ &= \mathcal{H}_Q(\mathbf{P}^*, \mathbf{P}) = E_{sp}(R, Q), \end{aligned} \quad (3.13.61)$$

где последнее равенство обусловлено выбором матрицы \mathbf{P}^* . Из (3.13.60), (3.13.61) и леммы 3.13.4 следует (3.13.58). Докажем теперь справедливость равенства (3.13.59). Из леммы 3.13.5 и (3.13.39) имеем

$$\begin{aligned} E_{sp}(R) &= \max_{\{Q\}} \max_{\{\rho \geq 0\}} (E_0^*(\rho, Q) - \rho R) = \\ &= \max_{\{\rho \geq 0\}} \max_{\{Q\}} (E_0(\rho, Q') - (1 + \rho) \mathcal{H}(Q, Q') - \rho R) = \\ &= \max_{\{Q'\}} \max_{\{\rho \geq 0\}} (E_0(\rho, Q') - (1 + \rho) \mathcal{H}(Q, Q') - \rho R). \end{aligned} \quad (3.13.62)$$

Учитывая, что при фиксированных \mathbf{Q}' и ρ в выражении под знаком \max в (3.13.62) от распределения \mathbf{Q} зависит только $\mathcal{H}(\mathbf{Q}, \mathbf{Q}')$, а также учитывая лемму 3.12.1, из (3.13.62) получим

$$E_{sp}(R) = \max_{\{\mathbf{Q}'\}} \max_{\rho \geq 0} (E_0(\rho, \mathbf{Q}') - \rho R) \triangleq \max_{\{\mathbf{Q}'\}} \tilde{E}(R, \mathbf{Q}') \triangleq \tilde{E}(R).$$

Теорема доказана.

Таким образом, из теоремы 3.13.3 следует, что экспоненты случайного кодирования $E(R)$ и сферической упаковки $E_{sp}(R)$ совпадают при всех $R \geq R_{kp}$. Поэтому построенные ранее верхняя и нижняя границы вероятности ошибки для дискретного канала без памяти экспоненциально точны при всех скоростях $R \geq R_{kp}$.

Задачи, упражнения и дополнения

3.1.1. Что значит, что задан некоторый дискретный канал? Выберите правильный ответ и объясните, почему остальные ответы неправильны: а) заданы множества X и Y сигналов на входе и выходе канала и заданы условные распределения вероятностей $p(y|x)$ появления на выходе сигнала $y \in Y$ при условии, что на входе имеется сигнал $x \in X$; б) заданы множества X и Y (см. выше) и для любого n заданы условные распределения $p(y|x)$, $x \in X^n$, $y \in Y^n$; в) то же, что и в п. б), кроме того, задано распределение вероятностей $p(x)$ на множестве входных последовательностей X^n ; г) для всех n заданы вероятности $p(x, y)$ пар $x \in X^n$, $y \in Y^n$.

3.1.2. Рассмотрим канал, в котором сигналы на входе и выходе принимают только два значения, скажем 0 и 1.

а) Покажите, что канал, в котором $p(y|x) = 2^{-n}$ для всех n , всех $x \in X^n$ и всех $y \in Y^n$, является каналом без памяти и удовлетворяет условию стационарности;

б) Покажите, что канал, в котором $p(y|x) = 2^{-n+1}$ для всех n , всех $x \in X^n$ и всех $y \in Y^n$ таких, что $y^{(1)} = 0$, является каналом без памяти, но не удовлетворяет условию стационарности.

3.2.1. Предположим, что в двоичном канале без памяти ошибки (т. е. события, состоящие в том, что выходной сигнал канала не совпадает со входным) случаются независимо с вероятностью p . Допустим, что для увеличения надежности передачи применяются повторения. Сколько раз достаточно повторить передачу одного и того же сигнала, чтобы при $p = 0,1$ вероятность ошибки после декодирования была меньше, чем 0,03, чем 0,01? Чему равна скорость кода в каждом из указанных случаев?

3.2.2. Предположим, что в условиях предыдущей задачи используется не повторение, а более сложное кодирование и соответственно декодирование такое, что при скорости R и при длине кода n исправляются все сочетания ошибок, если только их число на длине n не превышает $t = tn$ (такие сочетания ошибок называются ошибками кратности t). Величина t является корнем уравнения $h(2t) = 1 - R$, где $h(x) = -x \log x - (1-x) \log(1-x)$ — двоичная энтропия. Коды с указанными характеристиками (n, R, t) существуют для любых n и называются *корректирующими кодами*, лежащими на границе *Варшамова—Гилberta* (ВГ). Укажите диапазон скоростей, в котором коды, лежащие на ВГ-границе, позволяют достичь сколь угодно малой вероятности ошибки.

3.2.3. В этой задаче будет дана граница для вероятности больших уклонений сумм независимых случайных величин и ее применение для оценки вероят-

ности ошибки при передаче сообщений с помощью корректирующих кодов. Эта граница иногда называется границей Чернова.

Пусть Z_1, \dots, Z_n — независимые случайные величины. Обозначим через $G_i(s) \triangleq M \exp(sZ_i)$ функцию, называемую *производящей функцией моментов* случайной величины Z_i .

а) Покажите, что производящая функция моментов суммы $Z = Z_1 + \dots + Z_n$ равна $G(s) = \prod_{i=1}^n G_i(s)$.

б) Покажите, что $\Pr(Z > t) \leq e^{-st}G(s)$ при $s > 0$. Указание: рассмотрите $G(s) = \sum_z e^{sz} p(z)$ и ограничьте область суммирования теми z , которые больше или равны t .

в) Используя предыдущий результат, для случая суммы n одинаково распределенных слагаемых покажите, что $\Pr(Z > t) \leq \exp(-n(st - g(s)))$, где $t = tn/n$, $g(s) = \ln G_i(s)$, $s \geq 0$, и что оптимальное значение s_0 параметра s , минимизирующее правую часть неравенства, определяется из уравнения $dg/ds_0 = -t$. Используйте тот факт, что $Q(z) \triangleq e^{sz} p(z)/G(s)$ можно рассматривать как распределение вероятностей, для которого $G'(s)/G(s) = M_Q Z$ и $G'(s)/G(s) = M_Q Z^2$, где M_Q — математическое ожидание по распределению $Q(z)$, для доказательства того, что функция $st - g(s)$ выпукла вверх и, следовательно, s_0 , которое является решением уравнения $dg/ds = t$, минимизирует оценку вероятности $\Pr(Z > t)$.

г) Пусть Z_i принимает два значения 0 и 1, причем $p(Z_i = 0) = q$ и $p(Z_i = 1) = p$, $i = 1, 2, \dots, n$. Покажите, что $g(s) = \ln(q + e^s p)$, $dg/ds = pe^s/(q + pe^s)$ и $s_0 = \ln(t(1-p)/(1-t)p)$. Подставляя это значение s_0 в оценку, покажите, что при $t > p$

$$\Pr(Z_1 + \dots + Z_n > tn) \leq 2^{-nh(t, p)},$$

где

$$h(t, p) = t \log(t/p) + (1-t) \log(1-t)/(1-p).$$

д) Покажите, что при использовании кода $G(n, R)$, исправляющего ошибки кратности t и меньше, в ДСК с вероятностью ошибки p (т. е. в канале, описанном в задаче 3.2.1), вероятность ошибки декодирования имеет следующую оценку:

$$\lambda \leq 2^{-nh(t, p)}, \quad t > p,$$

где $t = \frac{t+1}{n}$.

3.2.4. В условиях задачи 3.2.2, используя результат задачи 3.2.3, оцените длину кода, который позволяет получить вероятности ошибки декодирования, меньшие, чем 10^{-3} , 10^{-6} при $R = 0,1$; $R = 0,2$, $R = 0,3$.

3.2.5. Предположим, что используемый в канале код имеет скорость R (в двоичных единицах). Сколько символов должно быть передано по каналу, чтобы при этом могло быть передано одно из 128 сообщений? Как нужно изменить скорость кода, чтобы при таком же числе символов канала можно было бы передать одно из 256 сообщений.

3.2.6. Рассмотрим двоичный канал из задачи 3.2.1. Для этого канала $p(y|x) = p^t(1-p)^{n-t}$, где x, y — последовательности длины n на входе и выходе, t — количество позиций, в которых последовательности x и y отличаются, а p — вероятность ошибки в канале. Предположим, что $p = 0,1$, $n = 3$ и используется код, состоящий из равновероятных кодовых слов $\mathbf{u}_1 = (010)$ и $\mathbf{u}_2 = (101)$.

а) Пусть решающие области выбраны следующим образом: $A_1 = \{010, 000, 110, 011\}$ и $A_2 = \{101, 001, 111, 100\}$. Найдите вероятности ошибок λ_1 , λ_2 , λ , Λ .

6) Пусть решашные области выбраны следующим образом: $A_1 = \{000, 100, 010, 001\}$ и $A_2 = \{110, 101, 011, 111\}$. Найдите вероятности ошибок $\lambda_1, \lambda_2, \lambda, \Lambda$. Какое из двух приведенных разбиений на решающие области лучше? Объясните причину. Как должны быть выбраны кодовые слова, чтобы худшее разбиение стало лучшим? Зависит ли выбор лучшего разбиения от вероятности p ?

3.2.7. Покажите, что средняя вероятность ошибки декодирования может быть рассчитана по любой из следующих формул:

$$\lambda := \sum_i \lambda_i p(u_i), \quad \lambda := \sum_i \Pr(e|w_i) p(w_i),$$

где $p(u)$ — распределение вероятностей на множестве кодовых слов, λ_i — вероятность ошибки при передаче слова u_i , $\Pr(e|w_i)$ — вероятность ошибки при принятии решения w_i , $p(w)$ — распределение вероятностей на множестве решений, определяемое каналом, декодером и распределением $p(u)$. Вычислите $\Pr(e|w_1), \Pr(e|w_2)$ для предыдущих двух задач.

3.3.1. Используя неравенство Фано, докажите обратную теорему кодирования дискретного источника без памяти, т. е. покажите, что при скорости кодирования $R < H(X)$, где $H(X)$ — энтропия ансамбля сообщений источника, найдется $\delta > 0$ такое, что вероятность ошибки больше, чем δ , для любого равномерного кода источника. Указание: положите $U = X^n$ и обозначьте через W ансамбль кодовых слов. Затем воспользуйтесь неравенствами $H(U; W) \leq H(W) \leq nR$ и неравенством (3.3.6).

3.3.2. Вычислите $H(U|W)$ и $H(E|W)$ для задачи 3.2.6. Проверьте выполнение неравенства Фано.

3.3.3. Предположим, что при использовании кода G ($n = 100, R = 1/2$) оказалось, что $H(U/W) = 10$. Оцените среднюю вероятность ошибки снизу.

3.3.4. Укажите достаточные условия для того, чтобы $H(U|W) = H(U|Y^n)$. Каковы необходимые условия того, чтобы $H(U|W) > H(U|Y^n)$? Вычислите $H(U|Y^n)$ для задачи 3.2.6.

3.4.1. Пусть C^* — информационная емкость канала и скорость кода, используемого в этом канале, равна $R = C^* + \varepsilon$, где ε — некоторое положительное число. Покажите, что с ростом n наименьший корень — λ_{\min} уравнения $H(E) + \lambda \log M = n\varepsilon, M = 2^{nR} \geq 1$, стремится к ε/R возрастаю. Для этого покажите вначале, что $\lambda_{\min} < \varepsilon/R$ при всех n . Затем, рассматривая n как непрерывную переменную, покажите, что производная $d\lambda_{\min}/dn$ положительна для всех n и для всех $\lambda < M/(M+1)$.

3.4.2. Покажите, что в условии задачи 3.3.3 и при равномерном распределении вероятностей на множестве кодовых слов информационная емкость имеет следующую оценку: $C^* \geq 0.4$. Указание: воспользоваться обратной теоремой кодирования.

3.5.1. Будем говорить, что дискретный канал является *строго симметричным по входу*, если для матрицы $P = \{p(y|x)\}, x \in X, y \in Y$, образованной переходными вероятностями канала, выполнены два условия: (а) все строки обозначены перестановками элементов первой строки, (б) множество Y выходных сигналов канала может быть разбито на такие непересекающиеся подмножества Y_1, \dots, Y_k , $\bigcup_{i=1}^k Y_i = Y$, что каждая подматрица P_i , соответствующая подмножеству Y_i , имеет все строки и все столбцы, обозначенные перестановками одного и того же множества чисел (это подмножество чисел может быть своим для каждой подматрицы). Например, канал со следующей матрицей переходных вероятностей:

$$P = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.1 & 0.2 & 0.7 \end{bmatrix}$$

(столбцы соответствуют выходным сигналам, строки — входным) строго симметричен по входу. Для этого канала $Y = \{y_1, y_2, y_3\}$, $Y_1 = \{y_1, y_3\}$, $Y_2 = \{y_2\}$

$$P_1 = \begin{bmatrix} 0.7 & 0.1 \\ 0.1 & 0.7 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 0.2 \\ 0.2 \end{bmatrix}.$$

Покажите, что информационная емкость строго симметричного дискретного канала без памяти достигается при равномерном распределении вероятностей на входе канала. Указание: воспользуйтесь теоремой 3.5.2.

3.5.2 *). Мы говорили (см. § 3.1), что канал задан, если заданы множества входных X и выходных Y сигналов, а также для всех $n, x \in X^n$ и $y \in Y^n$ определены (заданы) переходные вероятности $p(y|x)$. Конечно, такой метод задания каналов удобен только в случае каналов без памяти, когда достаточно задавать только одномерные ($n = 1$) переходные вероятности. Для того чтобы описывать более сложные каналы, которые имеют память, обычно используют специальные математические модели. Эти модели выбираются, исходя из свойств реального канала, а также исходя из простоты математического анализа. Ниже будет приведена одна из широко распространенных моделей дискретного канала с памятью.

Предположим, что канал может находиться в одном из T состояний и $S = \{s_0, s_1, \dots, s_{T-1}\}$ есть множество состояний. Для каждого состояния заданы переходные вероятности $p(y^{(i)}|x^{(i)}, s^{(i)})$, определяющие канал в i -й момент времени. В этой модели предполагается, что начальное состояние $s^{(1)}$ фиксировано и при известной последовательности состояний $s \in S^n$ канал не имеет памяти, т. е. $p(y|x, s) = \prod_{i=1}^n p(y^{(i)}|x^{(i)}, s^{(i)})$. Это, конечно, не означает, что канал не имеет памяти. Просто память канала обусловлена памятью «переключателя» состояний.

Для описания процесса, образованного последовательными состояниями, используется марковская цепь, в которой будущее состояние определяется только одним или несколькими прошлыми состояниями и не зависит от прошлых входных или выходных сигналов канала. В связи с независимостью состояний от сигналов такие каналы называют иногда *каналами без межсимвольной интерференции*.

На рисунке показана так называемая *модель Гильберта*. В соответствии с этой моделью канал может находиться в одном из двух состояний s_0, s_1 ; последовательность состояний образует простую марковскую цепь, переходные вероятности которой показаны в верхней части рисунка. В обоих состояниях s_0 (хорошем) и s_1 (плохом) канал является двоичным симметричным, причем он имеет вероятность ошибки $p_0 = 0$ в состоянии s_0 и вероятность ошибки $p_1 = 1/2$ (обрыв канала) в состоянии s_1 . В более общем случае (в обобщенной модели Гильберта) вероятность p_0 может быть больше нуля, а вероятность p_1 меньше $1/2$.

*). В этой задаче предполагается знакомство читателя с элементами теории цепей Маркова.

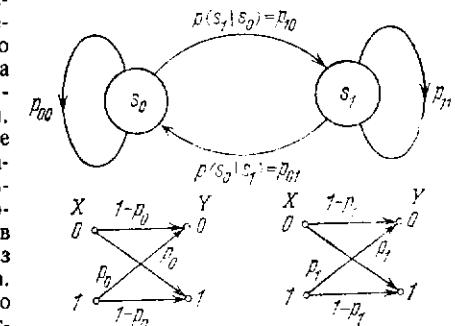


Рис. к задаче 3.5.2.

а) Покажите, что безусловные (стационарные или финальные) вероятности состояний могут быть найдены из соотношений $p(s_0) = p_{01}/(p_{01} + p_{10})$, $p(s_1) = p_{10}/(p_{01} + p_{10})$, где p_{ij} — вероятность появления состояния i при условии, что в предыдущий момент времени имелось состояние j . В таком канале хорошие и плохие состояния имеют тенденцию объединяться в серии (пакеты). Поэтому рассматриваемая модель хорошо описывает каналы, в которых ошибки группируются в пакеты. Покажите, что средняя длина серии плохих состояний равна $p(s_1)/p_{10}$. Найдите среднюю длину серии хороших состояний.

б) Найдите среднюю вероятность ошибки в обобщенной модели Гилберта ($p_0 \neq 0$, $p_1 \neq 1/2$), определяемую как вероятность события $x \neq y$. Покажите, что условные вероятности ошибок $\Pr(\varepsilon|x_1)$ и $\Pr(\varepsilon|x_2)$ совпадают с этой величиной, где x_1, x_2 — входные сигналы канала.

в) Покажите, что в случае каналов без межсимвольной интерференции $I(X^n; Y^n) = I(X^n; Y^n | S^n) = I(X^n; S^n | Y^n)$ для любого n , где $\{S^n, p(s)\}$ — ансамбль состояний. Указание: воспользуйтесь тем, что распределение вероятностей $p(x, y, s)$ на $X^n Y^n S^n$ может быть записано как $p(y|x, s)p(x)p(s)$; рассмотрите информацию $I(X^n; Y^n | S^n)$.

г) Покажите, что для модели Гилберта $\max_{\{p(x)\}} \frac{1}{n} I(X^n; Y^n | S^n) = p(s_0)$

и максимум достигается при равномерном распределении на входе $p(x) = 2^{-n}$, $x \in X^n$. Покажите, что отсюда вытекает следующая граница для информационной емкости рассматриваемого канала:

$$C^* \leq p_{01}/(p_{01} + p_{10}).$$

д) Покажите, что в случае $p_0 \neq 0$ и $p_1 \neq 1/2$ максимальное значение $\frac{1}{n} I(X^n; Y^n | S^n)$ достигается при равномерном распределении на входе и

$$\max_{\{p(x)\}} \frac{1}{n} I(X^n; Y^n | S^n) = \frac{p_{10}}{p_{10} + p_{01}} C_0^* + \frac{p_{01}}{p_{10} + p_{01}} C_1^*,$$

где C_0^* и C_1^* — информационные емкости двоичных симметрических каналов, которые соответствуют состояниям s_0 и s_1 соответственно. Покажите, что отсюда вытекает следующая граница для информационной емкости:

$$C^* \leq 1 - \frac{p_{10}}{p_{10} + p_{01}} h(p_0) - \frac{p_{01}}{p_{10} + p_{01}} h(p_1),$$

где $h(p) = -p \log p - (1-p) \log(1-p)$.

3.5.3. В более сложных моделях каналов с памятью последующие состояния могут определяться не только предыдущими состояниями, но и предыдущими сигналами на входе или выходе канала. В этом случае говорят, что модель описывает канал с межсимвольной интерференцией. Наиболее простой пример канала с межсимвольной интерференцией получается, когда состояние в каждый момент времени определяется только входным сигналом канала в предыдущий момент времени (см. рисунок). Как и в задаче 3.5.2, канал может находиться в одном из двух состояний s_0 или s_1 в каждом состоянии описывается моделью ДСК:

$$p(y^{(i)} | x^{(i)}, s^{(i)} = s_k) = \begin{cases} 1 - p_k, & y^{(i)} = x^{(i)}, \\ p_k, & y^{(i)} \neq x^{(i)}, \end{cases} \quad k = 0, 1.$$

Отличие состоит в том, что состояние равно s_k , если входной символ канала в предыдущий момент равен k ($k = 0, 1$). Таким образом, в этом канале $p(y^{(i)} | x^{(i)} x^{(i-1)} \dots) = p(y^{(i)} | x^{(i)}, s^{(i)})$. При известной последовательности

состояний, как и раньше, канал по определению является каналом без памяти, т. е.

$$p(y | x, s) = \prod_{i=1}^n p(y^{(i)} | x^{(i)}, s^{(i)}).$$

Покажите, что для канала с межсимвольной интерференцией, приведенного на рисунке, вероятности $p(y|x)$ определяются посредством задания распределения вероятностей на множестве состояний в первый момент времени.

Предположим, что $p(s^{(1)} = s_k) = p_k$, $k = 0, 1$. Покажите, что $p(10101/00000) = p_0^3 (1 - p_0)^2 (p_0 p_0 + p_1 p_1)$. Найдите вероятности $p(10101/10101)$ и $p(10101/11100)$. Укажите общую формулу для расчета $p(y|x)$.

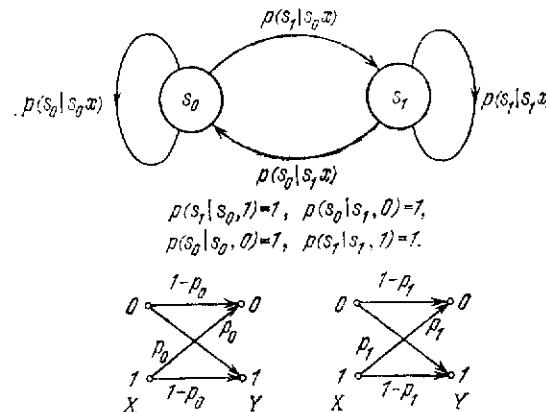


Рис. к задаче 3.5.3.

3.5.4. Теорема 3.5.2 может быть использована для отыскания информационной емкости C^* некоторых каналов без памяти. В этой задаче дается иллюстрация применения теоремы 3.5.2.

Пусть $X = \{x_1, \dots, x_L\}$, $Y = \{y_1, \dots, y_M\}$ — множества входных и выходных сигналов канала. Обозначим через p_k вероятность появления на входе сигнала x_k .

а) Покажите, что при $p_k \neq 0$ и невырожденной матрице переходных вероятностей вероятностный вектор $p = \{p_1, p_2, \dots, p_L\}$, максимизирующий информацию, можно получить из системы уравнений

$$p(y_j) = \sum_{k=1}^L p_k p(y_j | x_k), \quad j = 1, 2, \dots, M,$$

где $p(y_j)$ — решения следующей системы уравнений:

$$\sum_{j=1}^M p(y_j | x_k) [C^* + \log p(y_j)] = \sum_{j=1}^M p(y_j | x_k) \log p(y_j | x_k), \quad k = 1, \dots, L.$$

б) Если величины β_j , $j = 1, 2, \dots, M$, суть решения предыдущей системы уравнений относительно неизвестных $[C^* + \log p(y_j)]$, то $C^* = \log \sum_{j=1}^M 2^{\beta_j}$.

Покажите это.

3.5.5. Рассмотрим двоичный канал без памяти с матрицей переходных вероятностей $P = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$. Пусть $b = (b_1, b_2)^T$ (τ — знак транспонирования) — решение системы уравнений $Pb = a$, где $a = [-H(Y|0), -H(Y|1)]^T$ и $H(Y|x_k) = -\sum_y p(y_j|x_k) \log p(y_j|x_k)$, $x_k = 0, 1$.

а) Найдите b (ответ: $b = (-0.43 - 0.79)^T$).

б) Вычислите C^* (ответ: $C^* = 0.4$). Указание: воспользуйтесь результатами задачи 3.5.4.

3.5.6. Используйте итеративный алгоритм для нахождения информационной емкости канала, указанного в предыдущей задаче. Для этого положите $p_X^{(0)} = (p(0), p(1)) = (\frac{1}{2}, \frac{1}{2})$ и вычислите последовательность величин $I(p_X^{(k)})$, $k = 0, 1, \dots$. Покажите, что для первых трех шагов

$$p_X^{(0)} = (1/2, 1/2), \quad I(p_X^{(0)}) = 0.3973,$$

$$p_X^{(1)} = (0.5087, 0.4913), \quad I(p_X^{(1)}) = 0.3976,$$

$$p_X^{(2)} = (0.5131, 0.4869), \quad I(p_X^{(2)}) = 0.3977.$$

3.6.1. Предположим, что m дискретных каналов соединены последовательно. Говорят, что каналы соединены последовательно, если выходные сигналы одного канала являются входными сигналами другого. Пусть в рассматриваемых каналах множества X и Y совпадают и для каждого из каналов P есть матрица переходных вероятностей. Обозначим через $p_X^{(k)}$ и $p_Y^{(k)}$ вероятностные векторы, соответствующие распределениям на входе и на выходе k -го канала, $k = 1, 2, \dots, m$.

а) Покажите, что матрица переходных вероятностей канала, являющегося последовательным соединением m указанных каналов, равна P^m и $p_Y^{(m)} = p_X^{(1)} P^m$.

б) Покажите, что при известном сообщении на выходе k -го канала, $1 \ll i \ll k$, и на выходе j -го канала, $k < j \ll m$, статистически независимы.

в) Покажите, что последовательное соединение симметричных каналов является симметричным каналом. Предположим, что последовательно соединяются ДСК с вероятностью ошибки p . Покажите, что информационная емкость результирующего канала равна $1 - h(p_m)$, где $p_m = p * \dots * p$ (m раз) и $p * q \triangleq (1 - p)q + (1 - q)p$. Найдите предел информационной емкости при $m \rightarrow \infty$.

3.6.2. Рассмотрим вычисление информационной емкости ДСК при ограничениях на входе. Будем рассматривать ограничение, согласно которому на входе канала не могут появляться последовательности, содержащие слишком много единиц. Пусть p — вероятность ошибки ДСК и $\delta < 0.5$ — некоторое заданное число.

а) Покажите, что максимальное количество средней взаимной информации $I(X; Y)$ между входом и выходом канала, вычисленное по всем входным распределениям вероятностей, для которых $p(1) \leq \delta$, определяется соотношением $C_\delta^* = h(\delta * p) - h(p)$, где $h(\cdot)$ — двоичная энтропия и $\delta * p \triangleq \delta(1 - p) + (1 - \delta)p$.

б) Предположим, что на входных последовательностях ДСК задано распределение вероятностей следующего вида: $p(x) = (C_n^d)^{-1}$ для всех $x \in X^n$, для которых число единиц в x фиксировано и равно $d < n/2$, и $p(x) = 0$ для остальных x . Покажите, что $p(x^{(i)} = 1) = d/n$ для любого $i = 1, 2, \dots, n$. Покажите, что $1/n I(X^n; Y^n) \leq C_\delta^*$ при $d = \delta n$. Проверьте, достигается ли точное равенство в последнем неравенстве.

в) Исследуйте зависимость C_δ^* от вероятности ошибки p . Нарисуйте график и сравните с графиком информационной емкости ДСК без ограничений на входе.

3.6.3. Пусть код $G(n, R)$ для ДСК удовлетворяет ограничению на максимальное количество единиц в кодовом слове: количество единиц в каждом кодовом слове не превосходит $n\delta$.

а) Покажите, что при распределении на входе $p(x)$ таком, что $p(x) = 2^{-nR}$, если x совпадает с некоторым кодовым словом, и $p(x) = 0$ в противном случае,

$$\frac{1}{n} \sum_{i=1}^n p(x^{(i)} = 1) \leq \delta.$$

б) Покажите, что

$$\frac{1}{n} I(X^n; Y^n) \leq C_\delta^*.$$

Для этого воспользуйтесь результатом задачи 3.6.2 (а) и выпуклостью информации.

в) Докажите обратную теорему кодирования для ДСК и кодов, удовлетворяющих ограничению на число единиц в кодовом слове.

3.7.1. Является ли канал, описываемый моделью Гилберта (см. задачу 3.5.2), каналом с аддитивным шумом? Если да, то опишите источник U_Z шума в этом канале.

3.7.2. Пусть U_Z — простой марковский источник с переходными вероятностями $p(z^{(i)} = 0 | z^{(i-1)} = 1) = 0, 1$ и $p(z^{(i)} = 1 | z^{(i-1)} = 0) = 0, 2$. Найдите информационную емкость канала с аддитивным по модулю 2 шумом, в котором U_Z является источником шума. (Ответ: $C^* = 0,45$.)

3.8.1. Предположим, что на множестве кодовых слов задано равномерное распределение вероятностей. Покажите, что $B(u_i)$ — множество, определенное в теореме 3.8.1, есть множество таких $y \in Y^n$, для которых априорная вероятность $p(u_i | y)$ больше некоторой фиксированной величины. Найдите эту величину.

3.8.2. Предположим, что $p(x) = \prod_{i=1}^n p(x^{(i)})$, $x \in X^n$. Покажите, что в слу-

чае каналов без памяти вероятность $\Pr(y \in B(x) | x)$ стремится к единице при стремлении n к бесконечности, если $\tau = I(x; Y^n) - \varepsilon$ и ε — любое положительное число, где $B(x)$ и τ определены в теореме 3.8.1.

3.8.3. Предположим, что канал представляет собой ДСК с вероятностью ошибки p_0 , $S = X^n$ и $p(x) = 2^{-n}$ для любого $x \in X^n$. Покажите, что пара (x, y) , $x \in X^n$, $y \in Y^n$, принадлежит множеству V_τ , если расстояние Хемминга $d(x, y)$, т. е. количество позиций, в которых последовательности x и y отличаются, удовлетворяет условию

$$\frac{1}{n} d(x, y) \leq (1 - \tau + \log(1 - p_0)) \log^{-1}\left(\frac{1 - p_0}{p_0}\right).$$

Покажите, что выбирая $\tau = 1 - h(p_0) - \varepsilon$, где ε — произвольное положительное число, можно добиться того, чтобы при достаточно больших n величина $\Pr(V_\tau)$ была сколь угодно малой.

3.8.4. Метод максимальных кодов (метод исчерпывания) является одним из важных методов теории информации. Используя этот метод, получите границу Варшамова—Гилберта, приведенную в задаче 3.3.2. Для этого выполните следующие шаги.

а) Минимальным расстоянием d кода называется минимальное из попарных расстояний Хемминга между различными кодовыми словами. Покажите, что

необходимым и достаточным условием того, чтобы двоичный код исправлял все ошибки кратности t или меньше, является выполнение равенства $d = 2t + 1$.

б) Пусть $\{u_1, \dots, u_M\}$ максимальный двоичный код длины n с минимальным расстоянием d . Максимальность кода означает, что к этому коду нельзя добавить ни одной двоичной последовательности длины n такой, чтобы код $\{u_1, \dots, u_M, u_{M+1}\}$ имел минимальное расстояние, равное d . Шаром радиуса d с центром в точке u_i назовем множество всех двоичных последовательностей, находящихся от u_i на расстоянии d или меньше. Покажите, что для максимального двоичного кода длины n с минимальным расстоянием d любая двоичная последовательность $x \in X^n$ принадлежит по крайней мере одному из шаров радиуса $d - 1$ с центрами в u_i , $i = 1, \dots, M$.

в) Используя результат пп. а) и б), оцените снизу скорость двоичного кода длины n , исправляющего все ошибки кратности t или меньше.

3.9.1. Покажите, что в ДСК вероятность $\Pr(\bar{V}_t)$, определяемая формулой (3.9.6), имеет экспоненциальный характер убывания: $\Pr(\bar{V}_t) \leq 2^{-nE}$. Укажите значение коэффициента E в зависимости от вероятности ошибки в канале. Указание: воспользуйтесь границей задачи 3.2.3; для этого оцените требуемую вероятность с помощью производящей функции моментов.

3.10.1. Эргодичность шума является достаточным условием того, чтобы C^* было пропускной способностью AL -канала. В параграфе 3.10 в примере 3.10.1 отмечалось, что в случае неэргодического источника шума, состоящего из двух эргодических компонент с различными энтропиями $H(Z|Z^\infty, \alpha_1)$, $H(Z|Z^\infty, \alpha_2)$, где параметр α нумерует источник, выполняются неравенства $C^*(\alpha_1) < C^* < C^*(\alpha_2)$. Ниже это утверждение будет обосновано.

Пусть $q_1 = p(\alpha_1)$, $q_2 = p(\alpha_2) = 1 - q_1$ — вероятности компонент и $H(Z|Z^\infty A) \triangleq q_1 H(Z|Z^\infty \alpha_1) + q_2 H(Z|Z^\infty \alpha_2)$. Если $H(Z|Z^\infty)$ — энтропия на сообщение, то информационная емкость $C^* = 1 - H(Z|Z^\infty)$.

а) Покажите, что при любом выборе q_1 , q_2 имеет место неравенство $H(Z|Z^\infty A) < H(Z|Z^\infty)$. Указание: рассмотрите последовательность $n^{-1}H(Z^n|A)$, $n = 1, 2, \dots$

б) Покажите, что из предыдущего неравенства следует, что энтропия $H(Z|Z^\infty)$ — выпуклая вверх функция, определенная на множестве всех распределений вероятностей на множестве всех последовательностей.

в) Покажите, что $\frac{1}{n} I(Z^n; A)$ стремится к нулю с ростом n . Выведите отсюда, что $H(Z|Z^\infty) = H(Z|Z^\infty, A)$.

г) Пользуясь предыдущим результатом и определением энтропии $H(Z|Z^\infty, A)$, покажите, что $C^*(\alpha_1) < C^* < C^*(\alpha_2)$.

д) Покажите, что источник шума, рассматриваемый в этой задаче, является источником с памятью, даже тогда, когда каждая эргодическая компонента памяти не имеет. Указание: достаточно показать, что $H(Z|Z^\infty) < H(Z)$. Сделайте это, показав вначале, что $H(Z|Z^\infty) < H(Z|A)$.

е) Рассмотрим троичный канал с аддитивным по модулю 3 шумом, для которого $\{p(z|\alpha_1)\} = \{2/3, 1/6, 1/6\}$ и $\{p(z|\alpha_2)\} = \{1/3, 2/3, 1/6\}$ — два различных распределения вероятностей на компонентах источника шума, причем $q_1 = p_1 = q_2 = 1/2$, и компоненты памяти не имеют. Найдите $H(Z)$, $H(Z|Z^\infty)$, $C^*(\alpha_1)$, $C^*(\alpha_2)$, C^* .

3.11.1. Пусть для передачи четырех сообщений по ДСК с вероятностью ошибки 0,1 используется код, состоящий из следующих слов: $u_1 = (00000)$, $u_2 = (01111)$, $u_3 = (10101)$, $u_4 = (11010)$, причем $p(u_1) = 0,8$, $p(u_2) = 0,1$, $p(u_3) = 0,05$. Укажите разбиения на решающие области, выпишите элементы этих областей при декодировании по МАВ и при декодировании по МП.

3.11.2. В условии предыдущей задачи вычислите вероятности ошибок.

3.11.3. Решите предыдущие две задачи для случая несимметричного канала с переходными вероятностями $p(0|1) = 0,01$ и $p(1|0) = 0,1$.

3.12.1. Используя границу для вероятности ошибки декодирования в ДСК с $p = 0,1$ (см. (3.12.44), (3.12.45)), укажите длину кода, позволяющего достигнуть вероятности ошибки 10^{-6} при $R = 0,1, 0,3, 0,5$.

3.12.2. Покажите, что для кода из двух слов $\{u_1, u_2\}$ вероятность ошибки при декодировании по МП в произвольном дискретном канале не зависит от передаваемого слова и оценивается сверху величиной $\sum_y \sqrt{p(y|u_1)p(y|u_2)}$.

3.12.3. Покажите, что в случае кода со словами $\{u_1, \dots, u_M\}$ вероятность ошибки при передаче слова u_i и МП-декодировании удовлетворяет неравенству $\Pr(e|u_i) \leq \sum_{j=2}^M \lambda(u_i, u_j)$, где $\lambda(u_i, u_j)$ — вероятность ошибки для кода из двух слов u_i , u_j .

3.12.4. Используя результат задачи 3.12.2, покажите, что для дискретного канала без памяти средняя по ансамблю кодов из двух слов вероятность ошибки удовлетворяет неравенству

$$\overline{\lambda(u_1, u_2)} \leq \left[\sum_Y \left(\sum_X Q(x) \sqrt{p(y|x)} \right)^2 \right]^n.$$

3.12.5. Используя задачи 3.12.3 и 3.12.4, покажите, что средняя по ансамблю кодов с $M = 2^{nR}$ словами вероятность ошибки МП-декодирования в дискретном канале без памяти имеет следующую оценку:

$$\lambda \leq 2^{-n} [R_0(Q) - R],$$

$$R_0(Q) \triangleq \left[-\log \sum_Y \left(\sum_X Q(x) \sqrt{p(y|x)} \right)^2 \right].$$

Эта оценка называется *аддитивной границей* вероятности ошибки, а число $R_0 = \max_Q R_0(Q)$, зависящее только от переходных вероятностей канала, — *вычислительной скоростью* канала.

3.12.6. Покажите, что необходимые и достаточные условия максимума функции $E_0(p, Q, V)$ (см. (3.12.19)) по распределению Q при фиксированной матрице $V \triangleq (v(x|y))$, $x \in X$, $y \in Y$, есть

$$Q(x) = \frac{\left[\sum_Y p(y|x) v^{-p}(x|y) \right]^{1/p}}{\sum_X \left[\sum_Y p(y|x) v^{-p}(x|y) \right]^{1/p}}, \quad x \in X.$$

3.12.7. Различные функции, участвующие в построении экспоненты случайного кодирования для произвольного дискретного канала без памяти, содержат по определению максимизацию или минимизацию некоторого выражения по распределению вероятностей. Поэтому их вычисление представляет собой достаточно сложный процесс, особенно если количество входных и выходных сигналов канала велико. В этой задаче будет указан итеративный алгоритм вычисления одной из таких функций, подобный итеративному алгоритму вычисления информационной емкости.

Пусть $E_0(p) \triangleq \max_Q E_0(p, Q)$, где $E_0(p, Q)$ определено соотношением (3.12.15) (см. также теорему 3.12.3), и Q — распределение вероятностей на X .

Постройте итеративный алгоритм вычисления $E_0(\rho)$ при фиксированном ρ . Указание: построение алгоритма производите по аналогии с построением итеративного алгоритма для вычисления информационной емкости дискретного канала без памяти (п. 3.5.2), используя при этом соотношения (3.12.20), (3.12.28) и результат предыдущей задачи.

3.12.8. Покажите, что для строго симметричных по входу каналов (см. задачу 3.5.1) максимум функции $E_0(\rho, Q)$ по Q достигается на равномерном распределении $Q = \{1/L, \dots, 1/L\}$, $L = |X|$, при всех $\rho \geq 0$. Указание: воспользуйтесь леммой 3.12.3.

3.13.1. Пусть X_τ^n — множество всех последовательностей длины n с композицией τ .

а) Покажите, что $|X_\tau^n| \leq 2^{nH_\tau(X)}$, где $H_\tau(X) \triangleq -\sum_X \tau(x) \log \tau(x)$. Указание: покажите, что для распределения вероятностей $q_\tau(x) \triangleq \prod_{i=1}^n \tau(x^{(i)})$ и любой последовательности $x \in X_\tau^n$ имеет место равенство

$$q_\tau(x) = 2^{-nH_\tau(X)}.$$

б) Покажите, что

$$|X_\tau^n| \geq (n+1)^{-L+12-nH_\tau(X)}, \quad L = |X|.$$

Указание: покажите сначала, что для любой композиции τ^* выполняется неравенство $\Pr_{\tau^*}(X_\tau^n) \geq \Pr_{\tau^*}(X_{\tau^*}^n)$, где $\Pr_{\tau^*}(A)$ — вероятность множества $A \subseteq X^n$, вычисленная в соответствии с распределением $q_{\tau^*}(x)$; для доказательства рассмотрите отношение $\Pr_{\tau^*}(X_{\tau^*}^n)/\Pr_{\tau^*}(X_\tau^n)$ и воспользуйтесь тем, что

$$|X_\tau^n| = \frac{n!}{[\tau(x_1)n]! \dots [\tau(x_L)n]!}.$$

в) Пусть $g_i(x)$, $i = \overline{1, n}$, — произвольные неотрицательные функции, заданные на X , и пусть Q — некоторое распределение на X . Покажите, что

$$\sum_{X_\tau^n} |X_\tau^n|^{-1} \prod_{i=1}^n g_i(x^{(i)}) \leq (n+1)^{L-1} 2^{n\mathcal{H}(\tau, Q)} \prod_{i=1}^n \sum_X Q(x) g_i(x).$$

Указание: используйте результат п. б) и следующие равенства:

$$\sum_{X^n} Q(x) \prod_{i=1}^n g_i(x^{(i)}) = \sum_{\tau \in T} \sum_{x \in X_\tau^n} Q(x) \prod_{i=1}^n g_i(x^{(i)}),$$

$$Q(x) = 2^{-n[\mathcal{H}(\tau, Q) - H_\tau(X)]},$$

где $Q(x) \triangleq \prod_{i=1}^n Q(x^{(i)})$, T — множество всех композиций и $x \in X_\tau^n$.

3.13.2. В этой задаче с помощью метода случайного кодирования строится верхняя оценка для вероятности ошибки декодирования кода с фиксированной композицией в дискретном канале без памяти.

Пусть τ — некоторая композиция. Рассмотрим ансамбль \mathfrak{G}_τ кодов с фиксированной композицией τ , распределение вероятностей на котором дается соотношением (3.12.6) при

$$q(x) = \begin{cases} |X_\tau^n|^{-1} & \text{при } x \in X_\tau^n, \\ 0 & \text{для остальных } x \in X^n. \end{cases}$$

а) Покажите, что средняя по ансамблю кодов \mathfrak{G}_τ вероятность ошибки МП-декодирования удовлетворяет неравенству

$$\bar{\lambda} \leq M^\rho \sum_{Y^n} \left(\sum_{X_\tau^n} |X_\tau^n|^{-1} \prod_{i=1}^n [p(y^{(i)} | x^{(i)})]^{1/1+\rho} \right)^{1+\rho}, \quad 0 \leq \rho \leq 1,$$

где $M = 2^{nR}$ — объем кода.

б) Используя результат задачи 3.13.1 (в), покажите, что

$$\begin{aligned} \sum_{X_\tau^n} |X_\tau^n|^{-1} \prod_{i=1}^n p^{1/1+\rho}(y^{(i)} | x^{(i)}) &\leq \\ &\leq (n+1)^{L-1} 2^{nH(\tau, Q)} \prod_{i=1}^n \sum_X Q(x) p(y^{(i)} | x). \end{aligned}$$

в) Используя результаты задач 3.13.2 (а) и 3.13.2 (б), покажите, что существует код $G(n, R, \tau)$, средняя вероятность ошибки $\bar{\lambda}$ которого удовлетворяет неравенству

$$\bar{\lambda} \leq \begin{cases} 2^{-n[F_{sp}(R, \tau) - \epsilon_n]} & \text{при } R \geq R_{kp}(\tau), \\ 2^{-n[E_0^*(\rho, \tau) - R - \epsilon_n]} & \text{при } R < R_{kp}(\tau), \end{cases}$$

где ϵ_n — положительное число, сколь угодно малое при больших n , $R_{kp}(\tau) \triangleq \partial E_0^*(\rho, \tau)/\partial \rho$ и функция $E_0^*(\rho, \tau)$ определена в лемме 3.13.5.

3.13.3. В этой задаче даются указания к построению итеративного алгоритма для вычисления функции $E_{sp}(R, Q)$.

а) Пусть

$$E_0^*(\rho, Q, Q', V) \triangleq E_0(\rho, Q', V) - (1+\rho)\mathcal{H}(Q, Q').$$

Покажите, что необходимые и достаточные условия максимума функции $E_0^*(\rho, Q, Q', V)$ по распределению Q' при фиксированных ρ , Q и V есть

$$Q'(x) = \frac{\left[\sum_Y Q^{-1}(x) p(y|x) v^{-\rho} (x|y) \right]^{-1/1+\rho}}{\sum_X \left[\sum_Y Q^{-1}(x) p(y|x) v^{-\rho} (x|y) \right]^{-1/1+\rho}}, \quad x \in X.$$

б) Постройте итеративный алгоритм, аналогичный алгоритму вычисления информационной емкости (п. 3.5.2), используя соотношения (3.12.20), (3.12.28), лемму 3.13.5 и результат п. а).

КРАТКИЙ ИСТОРИЧЕСКИЙ КОММЕНТАРИЙ И ЛИТЕРАТУРА

Задача кодирования в канале с шумом впервые была рассмотрена К. Шеноном [13]. Первое доказательство прямой теоремы кодирования для дискретных каналов без памяти было получено А. Файнштейном [9]. Позднее эта же теорема была доказана К. Шенном методом случайного кодирования [14].

Большое число работ посвящено теоремам кодирования для различных моделей дискретных каналов. К их числу в первую очередь относятся работы К. Шенна [14], А. Я. Хинчина [12], Р. Л. Добрушина [7], А. Файнштейна [10] и Дж. Вольфовича [4]. Изложенный в § 3.5 итеративный алгоритм вычисления пропускной способности был построен независимо С. Аrimoto [1] и Р. Блэйхутом [3]. Экспоненциально точные в области высоких скоростей границы для вероятности ошибки декодирования в ДСК были впервые получены П. Элайесом [16]. Р. Л. Добрушин [8] получил аналогичные границы для произвольных симметричных каналов без памяти. Р. Фано [11] построил экспоненциально точные в области высоких скоростей границы для вероятности ошибки декодирования в произвольном дискретном канале без памяти. Позднее Р. Галлагер [5] предложил более простой, чем способ Фано, метод построения верхних границ. Параграф 3.12 основывается на методе Галлагера. В той же работе Галлагер улучшил верхнюю границу в области малых скоростей. К. Шенон, Р. Галлагер и Е. Берлекэмп [15] построили нижнюю границу, более точную при малых скоростях, чем граница Фано. Метод построения границы сферической упаковки, использованный в п. 3.13.2, принадлежит Е. А. Арутюняну [2]. Наиболее подробно вопросы кодирования в дискретных каналах рассмотрены в книгах Вольфовича [4] и Галлагера [6].

1. А р имото (Arimoto S.) An Algorithm for Computing the Capacity of Arbitrary Discrete Memoryless Channels. — IEEE Trans. on Inf. Theory, 1972, v. IT-18, № 1.
2. А р у тюн я н Е. А. Оценки экспоненты вероятности ошибки для полу-непрерывного канала без памяти. — Проблемы передачи информации, 1968, т. 4, № 4.
3. Б л э й х у т (Blahut R. E.) Computation of Channel Capacity and Rate Distortion Functions, IEEE Trans. on Inf. Theory, 1972, v. IT-18, № 4.
4. В оль фов и ц (Wolfowitz J.) Coding Theorems of Information Theory. Englewood Cliffs: New York, Springer-Verlag and Prentice-Hall, 1961. [Русский перевод: В оль фов и ц Дж. Теоремы кодирования теории информации. — М.: Мир, 1967.]
5. Г а л л а г е р (Gallager R. G.) A Simple Derivation of the Coding Theorem and Some Applications. — IEEE, Trans. Inform. Theory, 1965, v. IT-11, № 1. [Русский перевод: Г а л л а г е р Р. Г. Простой вывод теоремы кодирования и некоторые применения, Кибернетический сб., 1966, новая серия, вып. 3. — М.: Мир.]
6. Г а л л а г е р (Gallager R. G.) Information Theory and Reliable Communication. New York MIT, Wiley, 1968. [Русский перевод: Г а л л а г е р Р. Г. Теория информации и надежная связь. — М.: Советское радио, 1974.]
7. Д о б р у ш и н Р. Л. Общая формулировка основной теоремы Шенона в теории информации. — Успехи матем. наук, 1959, т. 14, № 6.
8. Д о б р у ш и н Р. Л. Асимптотика вероятностей ошибок при передаче информации по каналу без памяти с симметрической матрицей вероятностей перехода. — Докл. Акад. наук СССР, 1960, т. 133, № 2.
9. Ф айн стейн (Feinstein A.) A New Basic Theorem of Information Theory. — IRE Trans. Inf. Theory, 1954, PGIT-4, 2—22.
10. Ф айн стейн (Feinstein A.) Foundations of Information Theory. New York: McGraw-Hill, 1958. [Русский перевод: Ф айн стейн А. Основы теории информации. — М.: ИЛ, 1960.]

11. Ф а н о (Fano R. M.) Transmission of Information. A statistical theory of communication. — New York: Wiley, 1961. [Русский перевод: Ф а н о Р. Передача информации. Статистическая теория связи. — М.: Мир, 1965.]
12. Х и н ч и н А. Я. Об основных теоремах теории информации. — Успехи матем. наук, 1956, т. 11, № 1.
13. Ш ен н о н (Shannon C. E.) A Mathematical Theory of Communication. Bell. Syst. Tech. J., 1948, 27, 379—423, 623—656. [Русский перевод: Ш ен н о н К., Математическая теория связи. — В сб.: Работы по теории информации и кибернетике. — М.: ИЛ, 1963.]
14. Ш ен н о н (Shannon C. E.) Certain Results in Coding Theory for Noisy Channels. — Inf. Contr., 1957, 6—25. [Русский перевод: Ш ен н о н К. Некоторые результаты теории кодирования для каналов с шумами. — В сб.: Работы по теории информации и кибернетике. — М.: ИЛ, 1963.]
15. Ш ен н о н (Shannon C. E.), Г а л л а г е р (Gallager R. G.), Б е р л е к э м п (Berlekamp E. R.) Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels. — Inf. Contr., 1967, 10, 65—103 (Part I), 522—552 (part II). (Русский перевод: Ш ен н о н К., Г а л л а г е р Р., Б е р л е к э м п Е., Нижние границы вероятности ошибки для кодирования в дискретном канале без запоминания. — Зарубежная радиоэлектроника, 1968, вып. 2 и 6, 52—81, 41—64.)
16. Э л а й е с (Elias P.) Coding for Two Noisy Channels. — Inform. Theory (Cherry C. (Ed.)) London: Butterworth, 1956. [Русский перевод: Э л а й е с П. Ко-дирование для двух каналов с шумами. — В сб.: Теория передачи сооб-щений. — М.: ИЛ, 1957.]

Глава 4

КОДИРОВАНИЕ В НЕПРЕРЫВНЫХ КАНАЛАХ

В этой главе мы продолжим изучение задачи кодирования при передаче сообщений по каналам с шумом. Объектом рассмотрения теперь будут каналы, непрерывные по входу и выходу как с дискретным, так и с непрерывным временем. В предыдущей главе при изучении кодирования в дискретных каналах основной нашей задачей являлось определение максимальной скорости, с которой возможна передача при произвольно малой вероятности ошибки. Аналогичная задача будет решаться и в этой главе по отношению к непрерывным каналам.

Вначале мы рассмотрим кодирование в непрерывных каналах с дискретным временем. Оказывается, что все результаты, полученные в предыдущей главе для дискретных каналов, легко переносятся на непрерывные каналы. Такое перенесение достигается заменой переходных вероятностей, задающих дискретный канал, на условные функции плотности вероятностей, задающие непрерывный канал с дискретным временем, а также сумм на интегралы. Однако при кодировании в непрерывных каналах с дискретным временем возникает некоторая специфика, связанная с необходимостью учитывать ограниченность мощности передатчика. По существу именно этим специфическим вопросам кодирования в непрерывных каналах с дискретным временем и посвящены соответствующие параграфы этой главы.

Затем мы рассмотрим кодирование в непрерывных каналах с непрерывным временем (в непрерывных каналах). Здесь будет рассмотрена только одна, наиболее простая модель непрерывного канала, а именно модель частотно-ограниченного канала с аддитивным белым гауссовским шумом. Вопросы, относящиеся к каналам с небелым шумом и линейными фильтрами на входе и выходе, не рассматриваются. Ограничение рассмотрения только простейшей моделью обусловлено, с одной стороны, тем, что результаты для нее могут быть получены достаточно простыми средствами, а с другой стороны — тем, что изучение именно этой простой модели позволяет наиболее наглядно пояснить постановку задачи и основные результаты, относящиеся к кодированию в непрерывных каналах.

§ 4.1. Непрерывные каналы с дискретным временем. Обратная теорема кодирования

Пусть множества X и Y сигналов на входе и выходе соответственно непрерывного канала с дискретным временем — числовые оси. Передача одного сигнала $x \in X$ (однократная передача) в некоторой фиксированый момент времени задается с помощью условной (или переходной) функции плотности вероятностей (ф. п. в.) $f(y|x)$, $y \in Y$. Передача последовательностей задается с помощью условных (или переходных) многомерных ф. п. в. в точности так же, как задается передача последовательностей по дискретному каналу. Мы будем считать, что передача сообщений начинается в момент времени $i = 1$, поэтому непрерывный канал с дискретным временем считается заданным, если для любого $n = 1, 2, \dots$ и любых последовательностей $\mathbf{x} = (x^{(1)}, \dots, x^{(n)}) \in X^n$, $\mathbf{y} = (y^{(1)}, \dots, y^{(n)}) \in Y^n$ заданы n -мерные ф. п. в. $f(\mathbf{y}|\mathbf{x})$, описывающие передачу последовательностей длины n в таком канале.

Непрерывный канал с дискретным временем называется каналом без памяти, если для всех $n = 1, 2, \dots$, $\mathbf{x} \in X^n$, $\mathbf{y} \in Y^n$,

$$f(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n f_i(y^{(i)}|x^{(i)}). \quad (4.1.1)$$

Если, кроме того, ф. п. в., задающие передачу в моменты времени i и j , одинаковы для всех i и j , т. е. если

$$f_i(y|x) = f_j(y|x), \quad x \in X, \quad y \in Y, \quad (4.1.2)$$

то говорят, что канал с дискретным временем удовлетворяет условию стационарности. В дальнейшем мы всегда будем предполагать, если противное не оговорено особо, что условие стационарности всегда выполняется.

Как нетрудно догадаться, мы будем следовать схеме изложения предыдущей главы, где основная роль принадлежит средней взаимной информации между входом и выходом канала. В случае дискретного канала максимум средней взаимной информации на сообщение по всем распределениям вероятностей на входе давал пропускную способность канала, т. е. наибольшую скорость передачи, при которой вероятность ошибки могла быть сделана сколь угодно малой. В случае непрерывного канала это не так. Средняя взаимная информация может быть сделана какой угодно большой соответствующим выбором распределения вероятностей на входе. В этом основное отличие дискретных и непрерывных каналов. Следующий пример поясняет суть этого явления; он также подсказывает правильную постановку задачи кодирования.

Пример 4.1.1. Пусть X, Y, Z — гауссовские случайные величины (с. в.), связанные следующим соотношением: $Y = X + Z$, причем X и Z статистически независимы. Можно рассматривать X как входной и Y как выходной сигналы канала в некоторый момент времени. Величина Z называется шумом, а соответствующий канал — канал с аддитивным гауссовским шумом. Мы рассматриваем случай, когда распределение на входе канала является гауссовским. В этом случае

$$f(y|x) = \frac{1}{\sigma_Z \sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma_Z^2} (y-x)^2 \right\}, \quad (4.1.3)$$

$$f_1(x) = \frac{1}{\sigma_X \sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma_X^2} x^2 \right\},$$

где σ_X^2 и σ_Z^2 — дисперсии (мощности) сигнала на входе канала и шума. Очевидно, что $\sigma_Y^2 = \sigma_X^2 + \sigma_Z^2$.

Средняя взаимная информация $I(X; Y)$ может быть представлена как разность относительных энтропий

$$I(X; Y) = H_0(Y) - H_0(Y|X). \quad (4.1.4)$$

Из свойств относительной энтропии следует, что

$$H_0(Y) = \frac{1}{2} \log 2\pi e (\sigma_X^2 + \sigma_Z^2), \quad (4.1.5)$$

$$H_0(Y|X) = M H_0(Y|x), \quad (4.1.6)$$

где

$$H_0(Y|x) = \frac{1}{2} \log 2\pi e \sigma_Z^2. \quad (4.1.7)$$

и, следовательно,

$$H_0(Y|X) = H_0(Z) = \frac{1}{2} \log 2\pi e \sigma_Z^2. \quad (4.1.8)$$

Используя (4.1.4), (4.1.5) и (4.1.8), получим, что

$$I(X; Y) = \frac{1}{2} \log \left(1 + \frac{\sigma_X^2}{\sigma_Z^2} \right). \quad (4.1.9)$$

Из этого выражения видно, что средняя взаимная информация между входными и выходными сигналами канала в некоторый момент времени может быть сделана сколь угодно большой за счет выбора достаточно большого отношения σ_X^2/σ_Z^2 . Если канал задан, то задана и мощность шума σ_Z^2 , поэтому, выбирая мощность входных сигналов σ_X^2 достаточно большой, можно получить сколь угодно большое значение $I(X; Y)$.

На практике входным сигналам нельзя придавать сколь угодно большую мощность, так как мощность передатчика ограничена. Поэтому входные сигналы непрерывных каналов, а, следовательно, и распределения вероятностей на входе канала должны подчиняться так называемым мощностным ограничениям. В принципе возможны и другие ограничения, связанные с условиями передачи, но мы в дальнейшем будем рассматривать только мощностные ограничения.

Определение 4.1.1. Пусть $\mathbf{u}_1, \dots, \mathbf{u}_M$ — последовательности длины n (кодовые слова), образованные входными сигналами канала ($\mathbf{u}_i = (u_i^{(1)}, \dots, u_i^{(n)}) \in X^n$, $i = 1, \dots, M$), и A_1, \dots, A_M — непересекающиеся подмножества (решающие области), образованные выходными сигналами канала, $A_i \subseteq Y^n$, $i = 1, \dots, M$ *). Кодом для непрерывного канала с дискретным временем, удовлетворяющим ограничению P на среднюю мощность, будем называть множество пар $\{\mathbf{u}_1, A_1; \dots; \mathbf{u}_M, A_M\}$, такое, что

$$\frac{1}{n} \sum_{j=1}^n (u_i^{(j)})^2 \leq P, \quad i = 1, \dots, M, \quad (4.1.10)$$

для каждого кодового слова \mathbf{u}_i . Число

$$R = \frac{\log M}{n} \quad (4.1.11)$$

называется *скоростью*, а число n — *длиной* кода. Так же как в дискретном случае, код для непрерывного канала будет обозначаться символом $G(n, R)$.

Набор решающих областей задает правило декодирования: если выходная последовательность \mathbf{y} канала принадлежит множеству A_i , то принимается решение о том, что передавалось кодовое слово \mathbf{u}_i . Если при передаче \mathbf{u}_i последовательность \mathbf{y} не принадлежит A_i , то происходит ошибка. Вероятность этого события определяется соотношением

$$\lambda_i \triangleq 1 - \int_{A_i} f(\mathbf{y}|\mathbf{u}_i) d\mathbf{y}. \quad (4.1.12)$$

Как и в случае дискретных каналов, для каждого кода определены максимальная Λ и средняя $\bar{\lambda}$ вероятности ошибок.

Определение 4.1.2. Пропускной способностью непрерывного канала с дискретным временем при ограничении P на среднюю мощность входных сигналов называется максимальное число C такое, что для любого сколь угодно малого положительного δ и любого $R < C$ существует код $G(n, R)$, все слова которого удовлетворяют ограничению (4.1.10) и максимальная вероятность ошибки которого удовлетворяет неравенству $\Lambda \ll \delta$.

Введем теперь понятие информационной емкости непрерывного канала с дискретным временем при ограничении на среднюю мощность входных сигналов. Для этого рассмотрим всевозможные

*). Здесь и далее мы всегда предполагаем, что для множеств A_i , $i = 1, \dots, M$, выполнены условия измеримости относительно всех условных распределений вероятностей на выходе канала, соответствующих используемым кодовым словам. Грубо говоря, измеримость множеств A_i предполагает существование интегралов в определении вероятности ошибки (см. ниже).

ф. п. в. на X^n . Обозначим через $\Phi_n(P)$ множество всех ф. п. в. на X^n таких, что

$$\frac{1}{n} \sum_{i=1}^n M X_i^2 \leq P, \quad (4.1.13)$$

где

$$M X_i^2 \triangleq \int_{X^n} (x^{(i)})^2 f(\mathbf{x}) d\mathbf{x} = \int_X (x^{(i)})^2 f_i(x^{(i)}) dx^{(i)} \quad (4.1.14)$$

и

$$f_i(x^{(i)}) = \int \cdots \int f(\mathbf{x}) dx^{(1)} \dots dx^{(i-1)} dx^{(i+1)} \dots dx^{(n)}. \quad (4.1.15)$$

Определение 4.1.3. Информационной емкостью непрерывного канала с дискретным временем при ограничении P на среднюю мощность входных сигналов называется число C^* , определяемое следующим соотношением:

$$C^* = \sup_{n, \Phi_n(P)} \frac{1}{n} I(X^n; Y^n), \quad (4.1.16)$$

где верхняя грань разыскивается по всем n и по всем ф. п. в. $f(\mathbf{x}) \in \Phi_n(P)$, а $I(X^n; Y^n)$ — средняя взаимная информация, вычисленная для данного канала и для ф. п. в. $f(\mathbf{x}) \in \Phi_n(P)$.

Понятие информационной емкости позволяет сформулировать и доказать обратную теорему кодирования. В основе доказательства обратной теоремы лежит неравенство Фано (см. § 3.3), которое справедливо для произвольного канала. Заметим, что в случае непрерывных каналов с дискретным временем условные вероятности $p(w_j | \mathbf{u}_i)$ приятия решения w_j при условии, что передано кодовое слово \mathbf{u}_i , $j = 1, \dots, M$, вычисляются следующим образом:

$$p(w_j | \mathbf{u}_i) = \int_A f(\mathbf{y} | \mathbf{u}_i) dy. \quad (4.1.17)$$

Теорема 4.1.1 (обратная теорема кодирования для непрерывных каналов с дискретным временем при ограничении на среднюю мощность сигналов на входе). Пусть C^* — информационная емкость указанного выше канала при ограничении P на среднюю мощность сигналов на входе. Пусть ϵ — произвольное положительное число и $R = C^* + \epsilon$. Тогда найдется такое положительное число δ , зависящее от R , что для всякого кода $G(n, R)$, удовлетворяющего ограничению P на среднюю мощность, средняя вероятность ошибки $\lambda \geq \delta$.

Доказательство. Зафиксируем n и рассмотрим некоторый код $G(n, R)$ при $R = C^* + \epsilon$, $\epsilon > 0$, все слова которого удовлетворяют условию (4.1.10). Обозначим через $\mathbf{u}_1, \dots, \mathbf{u}_M$ слова этого кода и положим

$$f(\mathbf{x}) \triangleq \frac{1}{M} \sum_{i=1}^M \delta(\mathbf{x} - \mathbf{u}_i), \quad \mathbf{x} \in X^n, \quad (4.1.18)$$

где $\delta(\mathbf{x})$ — дельта-функция Дирака *). Функция $f(\mathbf{x})$ является обобщенной ф. п. в. дискретного распределения на X^n , приписывающей одинаковые вероятности $1/M$ всем кодовым словам и нулевые вероятности всем остальным последовательностям из X^n . Легко проверить, что функция $f(\mathbf{x})$ принадлежит $\Phi_n(P)$. Для того чтобы в этом убедиться, заметим, что неравенство (4.1.10) можно записать в следующей векторной форме:

$$\frac{1}{n} (\mathbf{u}_i \mathbf{u}_i^T) \leq P, \quad (4.1.19)$$

где « T » — символ транспонирования вектора \mathbf{u}_i . Рассмотрим левую часть неравенства (4.1.13). Очевидно,

$$\begin{aligned} \frac{1}{n} \int_{X^n} (\mathbf{x} \mathbf{x}^T) f(\mathbf{x}) d\mathbf{x} &= \frac{1}{n} \cdot \frac{1}{M} \sum_{i=1}^M \int_{X^n} (\mathbf{x} \mathbf{x}^T) \delta(\mathbf{x} - \mathbf{u}_i) d\mathbf{x} = \\ &= \frac{1}{n} \cdot \frac{1}{M} \sum_{i=1}^M (\mathbf{u}_i \mathbf{u}_i^T) \leq P, \end{aligned} \quad (4.1.20)$$

т. е. для ф. п. в. (4.1.18) неравенство (4.1.13) выполняется.

Из определения информационной емкости (4.1.16) следует, что для ф. п. в. (4.1.18) выполняется следующая цепочка неравенств:

$$n C^* \geq I(X^n; Y^n) = I(U; Y^n) \geq I(U; W), \quad (4.1.21)$$

где U — ансамбль кодовых слов с равномерным распределением вероятностей, W — ансамбль решений, и последнее неравенство есть следствие невозрастания средней взаимной информации при преобразованиях. Доказательство теоремы завершается применением неравенства Фано и рассуждений, приведенных при доказательстве обратной теоремы кодирования для дискретных каналов. Теорема доказана.

*) Если $\mathbf{x} = (x^{(1)}, \dots, x^{(n)})$, то согласно определениям § 2.2

$$\delta(\mathbf{x}) = \prod_{i=1}^n \delta(x^{(i)}).$$

§ 4.2. Непрерывные каналы без памяти с дискретным временем

В этом параграфе будет рассмотрена прямая теорема кодирования. Хотя в дальнейшем будут рассматриваться только непрерывные каналы без памяти с дискретным временем, мы начнем с общего результата, который позволяет получать прямые теоремы кодирования не только в случае каналов без памяти. Этим результатом является неравенство Файнштейна, вывод которого для непрерывных каналов с дискретным временем в точности повторяет соответствующий вывод для дискретных каналов и поэтому здесь опускается. Ниже будет приведена только формулировка теоремы, устанавливающей это неравенство.

Теорема 4.2.1 (неравенство Файнштейна). Пусть задан произвольный непрерывный канал с дискретным временем. Пусть τ — произвольное положительное число, S — произвольное подмножество множества X^n сигналов на входе канала, $f(\mathbf{x})$ — произвольная ф. п. в. на X^n . Тогда для любых значений n и R существует код $G(n, R)$, каждое слово которого принадлежит подмножеству S , а максимальная вероятность ошибки удовлетворяет неравенству

$$\Lambda_n \leq \frac{1}{\Pr(S)} [2^{-n(\tau-R)} + \Pr(\bar{V}_\tau)], \quad (4.2.1)$$

где

$$V_\tau \triangleq \left\{ (\mathbf{x}, \mathbf{y}) : I(\mathbf{x}; \mathbf{y}) = \log \frac{f(\mathbf{y}|\mathbf{x})}{f(\mathbf{y})} > n\tau \right\}, \quad (4.2.2)$$

$$\Pr(\bar{V}_\tau) = 1 - \int_{V_\tau} f(\mathbf{x}) f(\mathbf{y}|\mathbf{x}) d\mathbf{x} d\mathbf{y}, \quad (4.2.3)$$

$$\Pr(S) = \int_S f(\mathbf{x}) d\mathbf{x}. \quad (4.2.4)$$

Читателю нетрудно догадаться, что при использовании неравенства Файнштейна для доказательства прямых теорем кодирования для непрерывных каналов с дискретным временем при ограничении P на среднюю мощность сигналов на входе в качестве множества S будет выбрано множество всех последовательностей $\mathbf{x} = (x^{(1)}, \dots, x^{(n)}) \in X^n$ таких, что

$$\frac{1}{n} \sum_{i=1}^n (x^{(i)})^2 \leq P.$$

Теперь мы рассмотрим непрерывные каналы без памяти с дискретным временем и покажем, что в этом случае имеет место теорема, аналогичная теореме 3.5.1 для дискретных каналов без памяти.

Из нее вытекает, что информационная емкость каналов без памяти может быть вычислена более простым образом, чем в определении 4.1.3, а именно верхнюю грань в (4.1.16) можно искать, полагая $n = 1$.

Пусть $f(y|x)$, $x \in X$, $y \in Y$, — условные ф. п. в., задающие канал без памяти:

$$f(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n f(y^{(i)}|x^{(i)}). \quad (4.2.5)$$

Определим множество $\Phi(P)$ ф. п. в. на X следующим образом:

$$\Phi(P) \triangleq \left\{ f(x) : \int_{-\infty}^{\infty} x^2 f(x) dx \leq P \right\}. \quad (4.2.6)$$

Для каждой функции $f(x) \in \Phi(P)$ определена средняя взаимная информация $I(X; Y)$ между ансамблями X и Y .

Теорема 4.2.2. Информационная емкость непрерывного канала без памяти с ограничением P на среднюю мощность входных сигналов определяется соотношением

$$C^* = \max_{\Phi(P)} I(X; Y), \quad (4.2.7)$$

где максимум разыскивается по всем ф. п. в. $f(x)$ из множества $\Phi(P)$.

Доказательство. Пусть $f(\mathbf{x})$ — произвольная ф. п. в. на X^n , которая принадлежит множеству $\Phi_n(P)$, т. е. которая удовлетворяет ограничению (4.1.13). Тогда из (4.2.5), как и при доказательстве теоремы 3.5.1, следует неравенство

$$I(X^n; Y^n) \leq \sum_{i=1}^n I_i(X; Y), \quad (4.2.8)$$

где

$$I_i(X; Y) = \int_X \int_Y f_i(x^{(i)}) f(y^{(i)}|x^{(i)}) \log \frac{f(y^{(i)}|x^{(i)})}{f_i(y^{(i)})} dx^{(i)} dy^{(i)}, \quad (4.2.9)$$

и

$$f_i(y^{(i)}) = \int_X f_i(x^{(i)}) f(y^{(i)}|x^{(i)}) dx^{(i)}. \quad (4.2.10)$$

В неравенстве (4.2.8) равенство достигается в том случае, когда выходные сигналы канала в различные моменты времени статистически независимы (см. доказательство теоремы 3.5.1). Заметим, что в случае каналов без памяти независимость выходных сигналов обеспечивается выбором статистически независимых сигналов на входе, т. е. таким выбором, что

$$f(\mathbf{x}) = \prod_{i=1}^n f_i(x^{(i)}). \quad (4.2.11)$$

Средняя взаимная информация является выпуклой вверх функцией относительно входных распределений (см. § 2.5). Поэтому

$$\frac{1}{n} \sum_{i=1}^n I_i(X; Y) \leq I_0(X; Y), \quad (4.2.12)$$

где

$$I_0(X; Y) = \int_X \int_Y f_0(x) f(y|x) \log \frac{f(y|x)}{f_0(y)} dx dy, \quad (4.2.13)$$

$$f_0(y) = \int_X f_0(x) f(y|x) dx \quad (4.2.14)$$

и

$$f_0(x) \triangleq \frac{1}{n} \sum_{i=1}^n f_i(x), \quad (4.2.15)$$

причем $f_i(x)$ — ф. п. в. случайной величины X_i на входе канала в момент времени i , $i = 1, \dots, n$, определяемая соотношением (4.1.15). Легко увидеть, что ф. п. в. $f_0(x)$ принадлежит множеству $\Phi(P)$. Действительно, для любой ф. п. в. $f(\mathbf{x}) \in \Phi_n(P)$ из (4.2.15) следует, что

$$P \geq \frac{1}{n} \sum_{i=1}^n M X_i^2 = \frac{1}{n} \sum_{i=1}^n \int_X f_i(x) x^2 dx = \int_X x^2 f_0(x) dx, \quad (4.2.16)$$

т. е. $f_0(x) \in \Phi(P)$.

В неравенстве (4.2.12) равенство достигается в том случае, когда $f_i(x) = f_0(x)$ для всех $i = 1, \dots, n$. Таким образом,

$$\begin{aligned} \max_{f(\mathbf{x}) \in \Phi_n(P)} \frac{1}{n} I(X^n; Y^n) &\leq \max_{f(\mathbf{x}) \in \Phi_n(P)} \frac{1}{n} \sum_{i=1}^n I_i(X; Y) \leq \\ &\leq \max_{f(\mathbf{x}) \in \Phi(P)} I(X; Y). \end{aligned} \quad (4.2.17)$$

Легко указать функцию $f(\mathbf{x}) \in \Phi_n(P)$, на которой достигается максимум. Пусть $f_0(x)$ — ф. п. в. из $\Phi(P)$, на которой достигается максимум в последнем выражении в (4.2.17). Положим

$$f(\mathbf{x}) = \prod_{i=1}^n f_0(x^{(i)}). \quad (4.2.18)$$

Очевидно, что $f(\mathbf{x})$ принадлежит множеству $\Phi_n(P)$. При таком выборе входного распределения выходные сигналы канала статистически независимы. Следовательно, имеет место первое равенство в (4.2.17), кроме того, $f_i(x) = f_0(x)$, $i = 1, \dots, n$, и, следовательно, имеет место второе равенство в (4.2.17).

Тем самым, мы показали, что

$$C^* = \sup_{n, f(\mathbf{x}) \in \Phi_n(P)} \frac{1}{n} I(X^n; Y^n) = \max_{\Phi(P)} I(X; Y). \quad (4.2.19)$$

Теорема доказана.

Теперь мы еще более сузим класс рассматриваемых непрерывных каналов, а именно, рассмотрим канал без памяти с дискретным временем и аддитивным гауссовским шумом.

Предположим, что для любой случайной последовательности входных сигналов $\mathbf{X} = (X_1, \dots, X_n)$ выходная последовательность $\mathbf{Y} = (Y_1, \dots, Y_n)$ может быть записана в виде

$$\mathbf{Y} = \mathbf{X} + \mathbf{Z} = (X_1 + Z_1, \dots, X_n + Z_n), \quad (4.2.20)$$

причем случайные последовательности \mathbf{X} и \mathbf{Z} статистически независимы и Z_i , $i = 1, \dots, n$, — независимые гауссовские с. в. Если обозначить через $f_Z(\mathbf{z})$ ф. п. в. случного вектора \mathbf{Z} , то

$$f_Z(\mathbf{z}) = \prod_{i=1}^n f_Z(z^{(i)}), \quad (4.2.21)$$

где

$$f_Z(z^{(i)}) = \frac{1}{\sigma_Z \sqrt{2\pi}} \exp \left[-\frac{1}{2\sigma_Z^2} (z^{(i)})^2 \right]. \quad (4.2.22)$$

Число σ_Z^2 называется при этом мощностью шума.

Из (4.2.20) и статистической независимости \mathbf{X} и \mathbf{Z} следует, что

$$f(\mathbf{y}|\mathbf{x}) = f_Z(\mathbf{y} - \mathbf{x}) = \prod_{i=1}^n f_Z(y^{(i)} - x^{(i)}), \quad (4.2.23)$$

т. е. рассматриваемый канал является непрерывным каналом без памяти, в котором действует аддитивный гауссовский шум.

Теорема 4.2.3. Информационная емкость непрерывного канала без памяти с дискретным временем, с аддитивным гауссовским шумом мощности σ_Z^2 и ограничением P на среднюю мощность входных сигналов определяется соотношением

$$C^* = \frac{1}{2} \log \left(1 + \frac{P}{\sigma_Z^2} \right). \quad (4.2.24)$$

Доказательство. Для канала с аддитивным гауссовским шумом

$$\begin{aligned} I(X; Y) &= H_0(Y) - H_0(Y|X) = H_0(Y) - H_0(Z) \leq \\ &\leq \frac{1}{2} \log 2\pi e (P + \sigma_Z^2) - \frac{1}{2} \log 2\pi e \sigma_Z^2 = \frac{1}{2} \log \left(1 + \frac{P}{\sigma_Z^2} \right), \end{aligned} \quad (4.2.25)$$

где неравенство $H_0(Y) \leq \frac{1}{2} \log 2\pi e (P + \sigma_Z^2)$ выполняется в силу того, что с. в. Y имеет ограниченный средний квадрат:

$$\mathbf{M}Y^2 = \mathbf{M}(X+Z)^2 = \mathbf{M}X^2 + \mathbf{M}Z^2 \leq P + \sigma_Z^2. \quad (4.2.26)$$

Равенство в (4.2.25) достигается в том случае, когда с. в. Y является гауссовой и имеет дисперсию $P + \sigma_Z^2$. Это условие выполняется, когда с. в. X является гауссовой, имеет нулевое математическое ожидание и дисперсию P .

Таким образом,

$$C^* = \max_{\Phi(P)} I(X; Y) = \frac{1}{2} \log \left(1 + \frac{P}{\sigma_Z^2} \right), \quad (4.2.27)$$

где максимум достигается выбором гауссовой ф. п. в. $f(x)$ с параметрами $\mathbf{M}X = 0$, $\mathbf{M}X^2 = P$. Теорема доказана.

Перейдем теперь к формулировке и доказательству прямой теоремы кодирования для непрерывных каналов без памяти с аддитивным гауссовским шумом.

Теорема 4.2.4 (прямая теорема кодирования). Пусть C^* — информационная емкость непрерывного канала без памяти с дискретным временем и аддитивным гауссовским шумом мощности σ_Z^2 при ограничении P на среднюю мощность входных сигналов, определяемая формулой (4.2.24). При любом $R < C^*$ и любом положительном δ существует код $G(n, R)$, удовлетворяющий ограничению P на среднюю мощность кодовых слов, максимальная вероятность ошибки декодирования которого удовлетворяет неравенству $\Lambda_n \leq \delta$.

Доказательство. Воспользуемся неравенством Файнштейна (теорема 4.2.1), в котором множество S будем рассматривать как множество всех последовательностей $\mathbf{x} = (x^{(1)}, \dots, x^{(n)}) \in X^n$, удовлетворяющих ограничению P на среднюю мощность:

$$S \triangleq \left\{ \mathbf{x}: \frac{1}{n} \sum_{i=1}^n (x^{(i)})^2 \leq P \right\}. \quad (4.2.28)$$

Согласно теореме 4.2.1 для любой ф. п. в. $f(\mathbf{x})$ на X^n и любого положительного числа τ существует код $G(n, R)$, каждое слово которого принадлежит множеству S , т. е. каждое слово которого удовлетворяет ограничению P на среднюю мощность, и максимальная вероятность ошибки Λ_n удовлетворяет неравенству

$$\Lambda_n \leq \frac{1}{\Pr(S)} [2^{-n(\tau-R)} + \Pr(\bar{V}_\tau)], \quad (4.2.29)$$

где

$$\Pr(S) \triangleq \int_S f(\mathbf{x}) d\mathbf{x}, \quad (4.2.30)$$

$$V_\tau \triangleq \{(\mathbf{x}, \mathbf{y}): I(\mathbf{x}; \mathbf{y}) > n\tau\}. \quad (4.2.31)$$

Пусть $C^* - R = \varepsilon > 0$. Выберем ф. п. в. $f(\mathbf{x})$ следующим образом. Положим

$$f(\mathbf{x}) = \prod_{i=1}^n f(x^{(i)}), \quad (4.2.32)$$

где

$$f(x) = \frac{1}{\sigma_1 \sqrt{2\pi}} \exp \left[-\frac{1}{2\sigma_1^2} x^2 \right]. \quad (4.2.33)$$

Рассмотрим среднюю взаимную информацию $I(X; Y)$, вычисленную относительно распределения на XY с ф. п. в. $f(x)f(y|x)$, где функция $f(y|x)$ определяется каналом, а $f(x)$ — равенством (4.2.33). Пусть

$$\sigma_1^2 \triangleq P - \delta_1, \quad (4.2.34)$$

где δ_1 — корень уравнения

$$C^* - I(X; Y) = \frac{\varepsilon}{4} \quad (4.2.35)$$

или, что то же самое, уравнения

$$\frac{1}{2} \log \left(1 + \frac{P}{\sigma_Z^2} \right) - \frac{1}{2} \log \left(1 + \frac{P - \delta_1}{\sigma_Z^2} \right) = \frac{\varepsilon}{4}. \quad (4.2.36)$$

Из последнего соотношения видно, что δ_1 — непрерывная функция ε , причем $\delta_1 = 0$ при $\varepsilon = 0$ и $\delta_1 > 0$ при $\varepsilon > 0$.

Покажем, что при достаточно больших n вероятность $\Pr(S)$ достаточно близка к единице. Действительно,

$$\Pr(S) = \Pr \left(\frac{1}{n} \sum_{i=1}^n X_i^2 \leq P \right) \geq \Pr \left(\left| \frac{1}{n} \sum_{i=1}^n X_i^2 - \sigma_1^2 \right| \leq \delta_1 \right). \quad (4.2.37)$$

В соответствии с (4.2.32) с. в. X_1^2, \dots, X_n^2 статистически независимы, одинаково распределены и имеют математическое ожидание σ_1^2 . К этим с. в. применим закон больших чисел, согласно которому при больших n вероятность, записанная в правой части соотношения (4.2.37), близка к единице. Точнее, существует такое N_1 , что для всех $n > N_1$

$$\Pr(S) \geq \Pr \left(\left| \frac{1}{n} \sum_{i=1}^n X_i^2 - \sigma_1^2 \right| \leq \delta_1 \right) \geq 1 - \delta_2 \quad (4.2.38)$$

для любого положительного числа δ_2 .

Выберем параметр τ следующим образом:

$$\tau = C^* - \frac{\varepsilon}{2}. \quad (4.2.39)$$

Тогда при $n > N_1$, согласно неравенству Файнстейна найдется код $G(n, R)$, $R = C^* - \varepsilon$, такой, что

$$\Lambda_n \leq \frac{1}{1-\delta_2} [2^{-n\varepsilon/2} + \Pr(\bar{V}_\tau)]. \quad (4.2.40)$$

Рассмотрим вероятность $\Pr(\bar{V}_\tau)$. Так как для наступления события $(x, y) \in \bar{V}_\tau$ необходимо наступление события $I(x; y) \leq n\tau$, то

$$\Pr(\bar{V}_\tau) = \Pr(I(x; y) \leq n(C^* - \frac{\varepsilon}{2})). \quad (4.2.41)$$

Правую часть последнего соотношения можно оценить следующим образом:

$$\begin{aligned} \Pr(I(x; y) \leq n(C^* - \frac{\varepsilon}{2})) &= \Pr\left(\frac{1}{n} I(x; y) - I(X; Y) \leq -\frac{\varepsilon}{4}\right) \leq \\ &\leq \Pr\left(\left|\frac{1}{n} \sum_{i=1}^n I(x^{(i)}; y^{(i)}) - I(X; Y)\right| \geq \frac{\varepsilon}{4}\right). \end{aligned} \quad (4.2.42)$$

В силу выбора ф. п. в. $f(x)$, указанного в (4.2.32), с. в. $I(x^{(i)}; y^{(i)})$, $i = 1, \dots, n$, статистически независимы, одинаково распределены и все имеют математическое ожидание $I(X; Y)$. К этим с. в. применим закон больших чисел, согласно которому для любых ε и δ_2 найдется такое число N_2 , что для всех $n > N_2$

$$\Pr\left(\left|\frac{1}{n} \sum_{i=1}^n I(x^{(i)}, y^{(i)}) - I(X; Y)\right| \geq \frac{\varepsilon}{4}\right) \leq \delta_2. \quad (4.2.43)$$

Наконец, найдется такое число N_3 , что для всех $n > N_3$

$$2^{-n\varepsilon/2} \leq \delta_2. \quad (4.2.44)$$

Таким образом, если δ_2 выбрать из условия

$$\delta = \frac{2\delta_2}{1-\delta_2}, \quad (4.2.45)$$

то при $n > \max\{N_1, N_2, N_3\}$ существует код $G(n, R)$, для которого максимальная вероятность ошибки

$$\Lambda_n \leq \frac{1}{1-\delta_2} (\delta_2 + \delta_2) = \delta. \quad (4.2.46)$$

Теорема доказана.

Обратная и прямая теоремы кодирования позволяют теперь сформулировать следующее утверждение.

Следствие 4.2.1. Пусть C^* — информационная емкость непрерывного канала без памяти с дискретным временем, аддитивным гауссовским шумом и ограничением на среднюю мощность входных сигналов. Пусть C — пропускная способность этого канала, т. е. такое максимальное число, что для любого положительного δ и любого $R < C$ существует код $G(n, R)$, удовлетворяющий ограничению на среднюю мощность кодовых слов, максимальная вероятность ошибки которого не превосходит δ . Тогда $C = C^*$.

§ 4.3*. Каналы с непрерывным временем. Обратная теорема кодирования

До сих пор мы рассматривали лишь такие модели каналов, для которых предполагалось, что изменения значений входных и выходных сигналов происходят в дискретные моменты времени. Однако реально передаваемый и принимаемый в физическом канале сигналы могут представлять собой непрерывно меняющиеся во времени функции. Предположение о том, что время в канале дискретно, фактически означает, что реальные сигналы непрерывного канала подвергаются некоторой промежуточной обработке, дискретизации. Теперь мы хотим рассмотреть более общую модель канала связи, не делая предположений о наличии дискретизирующих устройств. Вероятностное описание таких непрерывных каналов, входные и выходные сигналы которых могут быть произвольными функциями времени, опирается на понятие случайного процесса непрерывного времени (см. § 2.6).

Определение 4.3.1. Непрерывным каналом с непрерывным временем (или просто непрерывным каналом) называется канал, входные и выходные сигналы которого могут быть произвольными функциями времени. Если на входе канала фиксирована некоторая функция $x(t)$ (некоторая реализация из множества входных сигналов канала), то выход канала является случайным процессом $Y_x(t)$, статистические характеристики которого зависят от фиксированной функции $x(t)$.

Вообще говоря, множество возможных входных сигналов канала бесконечно и несчетно. Поэтому задание непрерывного канала в общем случае требует задания несчетного множества случайных процессов $Y_x(t)$. В дальнейшем мы будем рассматривать, главным образом, так называемые каналы с аддитивным шумом, которые имеют существенно более простое описание.

Определение 4.3.2. Непрерывным каналом с аддитивным шумом называется такой непрерывный канал, процесс $Y_x(t)$ на

выходе которого при любой фиксированной функции $x(t)$ на его входе определяется соотношением

$$Y_x(t) = x(t) + Z(t), \quad (4.3.1)$$

где $Z(t)$ — случайный процесс, не зависящий от $x(t)$. Этот процесс называется *шумовым процессом* или просто *шумом*.

Таким образом, непрерывный канал с аддитивным шумом полностью определяется только одним случайнм процессом, а именно шумом. При любой фиксированной функции $x(t)$, принадлежащей множеству входных сигналов канала, выходной сигнал отличается от шумового процесса только математическим ожиданием. Очевидно, что случайный процесс $Y_x(t)$ имеет математическое ожидание, равное $x(t)$, если шум $Z(t)$ имеет нулевое математическое ожидание. В дальнейшем всегда предполагается, что шум имеет нулевое математическое ожидание. В противном случае, если $m_Z(t) \triangleq MZ(t) \neq 0$, то можно полагать, что входным сигналом канала является функция $x(t) + m_Z(t)$, и тем самым перейти к случаю, когда шум в канале имеет нулевое среднее.

Введем теперь понятие кода для непрерывного канала непрерывного времени.

Определение 4.3.3. Пусть $u_i = u_i(t)$, $i = 1, \dots, M$, $0 \leq t \leq T$, — произвольные интегрируемые с квадратом функции, заданные на интервале $[0, T]$ (кодовые слова). Пусть Y^t — множество всех сигналов на выходе канала, образованное функциями $y(t)$, заданными на том же интервале $[0, T]$, и A_1, \dots, A_M * — непересекающиеся подмножества множества Y^t (решающие области). Кодом для непрерывного канала будем называть множество пар $\{u_i, A_i\}$. Если для каждого кодового слова $u_i(t)$ имеет место неравенство

$$\frac{1}{T} \int_0^T u_i^2(t) dt \leq P, \quad i = 1, \dots, M, \quad (4.3.2)$$

то будем говорить, что код удовлетворяет ограничению P на среднюю мощность кодовых слов. Число

$$R = \frac{\log M}{T} \quad (4.3.3)$$

называется *скоростью*, а число T — *длиной* кода. Код длины T со скоростью R будет обозначаться символом $G(T, R)$.

*). Здесь, как и в случае каналов с дискретным временем, предполагается, что множества A_i , $i = 1, \dots, M$, удовлетворяют условиям измеримости; выполнение этих условий обеспечивает существование вероятностей ошибок (см. ф-лу (4.3.4) ниже).

Данное определение, по существу, не отличается от аналогичных определений в случае каналов дискретного времени. Как и раньше, кодовые слова представляют собой сигналы, с помощью которых передаются сообщения, а множества A_1, \dots, A_M задают правило декодирования: если сигнал $y(t)$ на выходе канала принадлежит множеству A_i , то принимается решение о том, что передавалось кодовое слово u_i .

Для каждого кода определена вероятность ошибки при передаче слова u_i :

$$\lambda_i \triangleq \Pr(Y(t) \notin A_i | u_i(t)). \quad (4.3.4)$$

Кроме того, определены средняя λ и максимальная Λ вероятности ошибок. Эти вероятности определяются в точности так же, как и в случае дискретных и непрерывных каналов с дискретным временем.

Пропускная способность С непрерывного канала при ограничении на среднюю мощность сигналов на входе определяется аналогично пропускной способности непрерывных каналов с дискретным временем (см. определение 4.1.2).

Информационная емкость C^* непрерывных каналов при ограничении на среднюю мощность сигналов на входе, вообще говоря, определяется таким же образом, как и в случае каналов с дискретным временем. Мы приведем это определение, не вдаваясь в детали. Более подробное обсуждение будет дано в случае каналов с аддитивным белым гауссовским шумом с ограничением на полосу частот.

Определение 4.3.4. Пусть $X_T(t)$ — случайный процесс, заданный на интервале $[0, T]$. Пусть для случайного процесса $X_T(t)$ на входе определена величина средней взаимной информации $\frac{1}{T} I(X_T(t); Y_T(t))$ в единицу времени между случайными процессами на входе $X_T(t)$ и на выходе $Y_T(t)$, $0 \leq t \leq T$, канала, соответственно. Число

$$C^* = \sup_{T, X_T(t)} \frac{1}{T} I(X_T(t); Y_T(t)), \quad (4.3.5)$$

где верхняя грань разыскивается по всем T и по всем случайным процессам $X_T(t)$, обладающим ограниченной средней мощностью, т. е. таким, что

$$\frac{1}{T} \int_0^T M X_T^2(t) dt \leq P, \quad (4.3.6)$$

называется *информационной емкостью непрерывного канала при ограничении Р на среднюю мощность входных сигналов*.

Докажем теперь обратную теорему кодирования. Как и раньше, основным инструментом доказательства является неравенство Фано.

Теорема 4.3.1 (обратная теорема кодирования для непрерывных каналов при ограничении на среднюю мощность сигналов на входе). Пусть C^* — информационная емкость указанного выше канала при ограничении P на среднюю мощность сигналов на входе. Пусть ε — произвольное положительное число и $R = C^* + \varepsilon$. Тогда найдется положительное число δ , зависящее от R , такое, что для всякого T и всякого кода $G(T, R)$, удовлетворяющего ограничению R (см. (4.3.2)), средняя вероятность ошибки $\lambda \geq \delta$.

Доказательство. Зафиксируем T и рассмотрим некоторый код $G(T, R)$ при $R = C^* + \varepsilon$, $\varepsilon > 0$, все слова которого удовлетворяют условию (4.3.2). Рассмотрим случайный процесс $X_T(t)$ на входе канала, для которого с вероятностью единица в качестве реализаций появляются кодовые слова $u_i(t)$, $i = 1, \dots, M$, $M = 2^{RT}$, рассматриваемого кода. Будем считать, что вероятность каждой такой реализации равна 2^{-RT} . Так как каждая реализация процесса $X_T(t)$ удовлетворяет условию

$$\frac{1}{T} \int_0^T u_i^2(t) dt \leq P, \quad i = 1, \dots, M,$$

то и сам процесс $X_T(t)$ удовлетворяет условию (4.3.6). Из определения информационной емкости имеем

$$C^* \geq \frac{1}{T} I(X_T(t); Y_T(t)) \geq \frac{1}{T} I(U; W), \quad (4.3.7)$$

где U — ансамбль кодовых слов с равномерным распределением вероятностей, а W — ансамбль решений, для которого вероятности $p(w_j | u_i) \triangleq \Pr(Y_T(t) \in A_j | u_i(t))$ определены заданием канала. Второе неравенство в (4.3.7) есть следствие невозрастания средней взаимной информации при преобразованиях. Доказательство теоремы завершается применением неравенства Фано и рассуждений, приведенных при доказательстве обратной теоремы кодирования для дискретных каналов. Теорема доказана.

Наша цель по-прежнему состоит в том, чтобы показать, что по крайней мере в некоторых случаях, пропускная способность непрерывного канала равна его информационной емкости, а также в том, чтобы указать метод вычисления информационной емкости. В общем случае эта задача является весьма сложной. Для упрощения доказательств и для получения наглядных результатов мы ограничимся рассмотрением одного частного вида каналов, а именно непрерывных каналов с аддитивным белым гауссовским шумом и ограничением на полосу частот.

Говорят, что в канале действует аддитивный белый гауссовский шум, если шумовой процесс $Z(t)$ в определении 4.3.2 является процессом белого гауссовского шума. Это означает, что для любой системы ортонормированных функций $\varphi_i(t)$, $i = 1, 2, \dots$, случайные величины

$$Z_i \triangleq \int_0^T Z(t) \varphi_i(t) dt \quad (4.3.8)$$

являются совместно гауссовскими и статистически независимыми. Математическое ожидание каждой из с. в. (4.3.8) равно нулю, а дисперсия равна $N_0/2$ — интенсивности белого шума. Такой процесс можно представить себе как сумму бесконечного числа составляющих вида $Z_i \varphi_i(t)$ (гармонических составляющих, если $\varphi_i(t)$, $i = 1, 2, \dots$, — гармонические функции). Мощность каждой составляющей равна $N_0/2$ — отсюда название белый шум. Реально белый шум не может существовать, иначе процесс имел бы бесконечную мощность. Однако он является удобной математической моделью для описания реальных каналов, шум в которых имеет достаточно широкий спектр.

Предположим теперь, что некоторая функция $x(t)$, заданная на интервале $[0, T]$, представима в виде конечного ряда

$$x(t) = \sum_{i=1}^{ST} x_i \varphi_i(t), \quad 0 \leq t \leq T, \quad (4.3.9)$$

где $\varphi_i(t)$, $i = 1, 2, \dots, ST$, — некоторая фиксированная система ортонормированных на интервале $[0, T]$ функций, а S — некоторое фиксированное число. Ограничение на конечность ряда (4.3.9) можно интерпретировать как ограничение на полосу, занимаемую сигналом $x(t)$ (на полосу частот, если $\varphi_i(t)$, $i = 1, 2, \dots$ — гармонические функции и T достаточно велико).

Действительно, пусть $\{\varphi_i(t)\}$ — ортонормальные гармонические функции на интервале $[-T/2, T/2]$, т. е.

$$\varphi_i(t) = \frac{1}{\sqrt{T}} \exp\left(-j2\pi \frac{i}{T} t\right), \quad i = 0, \pm 1, \pm 2, \dots, -\frac{T}{2} \leq t \leq \frac{T}{2}.$$

Число $\omega = 2\pi \frac{i}{T}$ является круговой частотой колебания $\varphi_i(t)$. Поскольку $\varphi_i(t)$ является отрезком длины T гармонического колебания, то в его спектре имеются компоненты, частоты которых отличаются от $2\pi \frac{i}{T}$. Так как спектр функции $\varphi_i(t)$ равен

$$F(\omega) = \frac{\sin\left(2\pi \frac{i}{T} - \omega\right) \cdot \frac{T}{2}}{\left(2\pi \frac{i}{T} - \omega\right) \cdot \frac{T}{2}} \quad (4.3.10)$$

и так как функция $\frac{\sin x}{x}$ равна единице при $x = 0$ и убывает к нулю при увеличении x , то при больших значениях T спектр функции $\varphi_i(t)$ сосредоточен вблизи частоты $2\pi \frac{i}{T}$. Если некоторая функция $x(t)$ имеет разложение вида (4.3.9) в ряд не более чем с $2WT + 1$ членами, соответствующими значениям i таким, что $\left| \frac{i}{T} \right| \leq W$, то при больших T почти весь спектр функции $x(t)$ будет находиться внутри полосы частот $[-2\pi W, 2\pi W]$.

Слова «почти весь спектр» требуют, вообще говоря, уточнения, которое может быть сделано с помощью так называемых эллиптических функций вытянутого сфера. Мы отсылаем интересующегося читателя к книге Р. Галлагера [1] и к работам Г. Ландау и Г. Поллака (см. [1]). Для целей нашего изложения уточнение не требуется, поскольку мы дальше будем рассматривать только такие сигналы на входе канала, которые представимы в виде конечного ряда (4.3.9) по некоторой системе ортонормальных функций. Следуя традиции, принятой в теории информации, мы будем полагать $S = 2W$ и говорить, что сигналы, представимые в виде (4.3.9), удовлетворяют ограничению W на полосу частот.

Таким образом, мы приходим к модели непрерывного канала с аддитивным белым гауссовским шумом и ограничениями на среднюю мощность и на полосу частот входных сигналов. Код для такого канала определяется так же, как в определении 4.3.3, с тем только отличием, что каждое кодовое слово дополнительно удовлетворяет ограничению W на полосу частот, т. е. каждое кодовое слово представимо в виде конечного ряда с $2WT$ членами по некоторой фиксированной системе ортонормальных функций. Аналогичным образом определяется пропускная способность и информационная емкость. В определении информационной емкости (см. определение 4.3.4) дополнительно требуется, чтобы случайный процесс $X_T(t)$ имел представление в виде конечного ряда с $2WT$ членами относительно той же фиксированной системы ортонормальных функций:

$$X_T(t) = \sum_{i=1}^{2WT} X_i \varphi_i(t), \quad 0 < t < T. \quad (4.3.11)$$

Теорема 4.3.2. Информационная емкость C_W^* непрерывного канала с аддитивным белым гауссовским шумом при ограничении P на среднюю мощность и ограничении W на полосу частот не зависит от выбора системы из $2WT$ ортонормированных функций и определяется соотношением

$$C_W^* = W \log \left(1 + \frac{P}{WN_0} \right) \text{ бит/с}, \quad (4.3.12)$$

где $N_0/2$ — интенсивность белого шума.

Доказательство. Напомним вначале, что средняя взаимная информация между двумя случайными процессами $X_T(t)$ и $Y_T(t)$, заданными на интервале $[0, T]$, определяется как следующий предел, если он существует,

$$I(X_T(t); Y_T(t)) \triangleq \lim_{n \rightarrow \infty} I(X_1 \dots X_n; Y_1 \dots Y_n), \quad (4.3.13)$$

где

$$\begin{aligned} X_i &= \int_0^T X_T(t) \varphi_i(t) dt, \quad i = 1, 2, \dots, \\ Y_j &= \int_0^T Y_T(t) \varphi_j(t) dt, \quad j = 1, 2, \dots, \end{aligned} \quad (4.3.14)$$

и $\varphi_i(t)$, $i = 1, 2, \dots$, — произвольная полная в L_2 (в пространстве функций, интегрируемых с квадратом) система ортонормированных функций. Пусть случайный процесс $X_T(t)$ удовлетворяет условию (4.3.11). Так как случайные величины X_i при $i > 2WT$ вырождены (тождественно равны нулю), то они статистически независимы от первых $2WT$ с. в. и от всех с. в. Y_j , $j = 1, 2, \dots$. Поэтому при всех $n \geq 2WT$

$$I(X_1 \dots X_n; Y_1 \dots Y_n) = I(X_1 \dots X_{2WT}; Y_1 \dots Y_{2WT}) \quad (4.3.15)$$

и, следовательно,

$$I(X_T(t); Y_T(t)) = I(X_1 \dots X_{2WT}; Y_1 \dots Y_{2WT}). \quad (4.3.16)$$

Так как $Z_T(t)$ — белый гауссовский шум с интенсивностью $N_0/2$, то

$$Z_i = \int_0^T Z_T(t) \varphi_i(t) dt, \quad i = 1, 2, \dots, \quad (4.3.17)$$

являются в совокупности гауссовскими и статистически независимыми с. в., причем математическое ожидание каждой из этих с. в. равно нулю, а дисперсия равна $N_0/2$. В канале с аддитивным белым гауссовским шумом случайный процесс $Y_T(t)$, процесс на выходе канала, может быть представлен как

$$Y_T(t) = X_T(t) + Z_T(t), \quad 0 < t < T. \quad (4.3.18)$$

Тогда

$$Y_i = X_i + Z_i, \quad i = 1, 2, \dots, \quad (4.3.19)$$

причем с. в. X_i , Z_j , $i, j = 1, 2, \dots$, статистически независимы в силу независимости процессов $X_T(t)$ и $Z_T(t)$.

Из равенства (4.3.16) следует, что для вычисления средней взаимной информации между процессами $X_T(t)$ и $Y_T(t)$ при

условии, что $X_T(t)$ удовлетворяет ограничению W на полосу частот, достаточно рассматривать среднюю взаимную информацию между входом и выходом непрерывного канала «дискретного времени», в котором на вход подается последовательность X_1, \dots, X_n , на выходе появляется последовательность Y_1, \dots, Y_n и действует аддитивный гауссовский шум Z_1, \dots, Z_n , $n = 2WT$. Поскольку с. в. Z_1, \dots, Z_n статистически независимы и имеют одинаковое распределение вероятностей с нулевым средним и дисперсией $N_0/2$, то рассматриваемый канал является непрерывным каналом без памяти с дискретным временем и с аддитивным гауссовским шумом.

Если процесс $X_T(t)$ удовлетворяет, кроме того, ограничению P на среднюю мощность, т. е.

$$\frac{1}{T} \mathbf{M} \int_0^T X_T^2(t) dt \leq P,$$

то из (4.3.11) и из ортонормальности функций $\varphi_i(t)$, $i = 1, 2, \dots, n = 2WT$, следует, что

$$\begin{aligned} \frac{1}{T} \mathbf{M} \int_0^T X_T^2(t) dt &= \frac{1}{T} \sum_{i, j=1}^n \mathbf{M} X_i X_j \int_0^T \varphi_i(t) \varphi_j(t) dt = \\ &= 2W \frac{1}{n} \sum_{i=1}^n \mathbf{M} X_i^2 \leq P, \end{aligned} \quad (4.3.20)$$

т. е. последовательность случайных величин X_1, \dots, X_n удовлетворяет следующему ограничению на среднюю мощность:

$$\frac{1}{n} \sum_{i=1}^n \mathbf{M} X_i^2 \leq \frac{P}{2W}. \quad (4.3.21)$$

Из определения информационной емкости следует, что

$$\begin{aligned} C_W^* &= \sup_{T, X_T(t)} \frac{1}{T} I(X_T(t); Y_T(t)) = \\ &= 2W \sup_{n, (X_1, \dots, X_n)} \frac{1}{n} I(X_1 \dots X_n; Y_1 \dots Y_n), \quad n = 2WT, \end{aligned} \quad (4.3.22)$$

где в последнем выражении верхняя грань разыскивается по всем $n = 1, 2, \dots$ и по всем последовательностям с. в. X_1, \dots, X_n таким, что выполняется условие (4.3.21). Задача отыскания этой верхней грани нами рассматривалась в теоремах 4.2.2 и 4.2.3. Согласно первой из этих теорем

$$\sup_{n, (X_1, \dots, X_n)} \frac{1}{n} I(X_1 \dots X_n; Y_1 \dots Y_n) = \max_{\Phi(P/2W)} I(X; Y), \quad (4.3.23)$$

где множество $\Phi(P/2W)$ определено соотношением (4.2.6). Согласно второй из упомянутых теорем

$$\max_{\Phi(P/2W)} I(X; Y) = \frac{1}{2} \log \left(1 + \frac{P}{WN_0} \right). \quad (4.3.24)$$

Теперь из (4.3.22)–(4.3.24) следует выражение (4.3.12). Теорема доказана.

Таким образом, мы получили, что информационная емкость непрерывного канала с аддитивным белым гауссовским шумом при ограничениях на среднюю мощность и на полосу частот определяется формулой (4.3.12) и не зависит от выбора системы ортонормированных функций, устанавливающих вид возможных сигналов на входе канала, а зависит только от числа $2WT$ таких функций в системе. Легко увидеть, что C_W^* является монотонно возрастающей функцией от W . Предел $C_\infty^* \triangleq \lim_{W \rightarrow \infty} C_\infty^*$ определяется с помощью правила Лопитала

$$C_\infty^* = \lim_{W \rightarrow \infty} W \log \left(1 + \frac{P}{WN_0} \right) = \lim_{W \rightarrow \infty} \frac{\log \left(1 + \frac{P}{WN_0} \right)}{\frac{1}{W}} = \frac{P}{N_0} \log e \text{ (бит/с).} \quad (4.3.25)$$

Для непрерывного канала с аддитивным белым гауссовским шумом при ограничениях на среднюю мощность и полосу частот входных сигналов остается справедливой обратная теорема кодирования. Она устанавливает, что для всякого кода, удовлетворяющего ограничениям на среднюю мощность и полосу частот и имеющего скорость R большую, чем информационная емкость C^* указанного канала, средняя вероятность ошибки не может быть сделана произвольно малой, а остается не меньшей, чем некоторое положительное число. Доказательство такой теоремы в точности повторяет доказательство теоремы 4.3.1, и поэтому мы приведем только ее формулировку.

Теорема 4.3.3 (обратная теорема кодирования для непрерывных каналов с аддитивным белым гауссовским шумом при ограничениях на среднюю мощность и полосу частот сигналов на входе). Пусть $C_W^* = W \log (1 + P/WN_0)$ — информационная емкость указанного выше канала и e — произвольное положительное число. Тогда найдется положительное число δ , зависящее от R , такое, что для всякого T и всякого кода $G(T, R)$, $R = C_W^* + e$, удовлетворяющего ограничению P на среднюю мощность и ограничению W на полосу частот, средняя вероятность ошибки $\lambda \geq \delta$.

§ 4.4*. Прямая теорема кодирования для непрерывных каналов с аддитивным белым гауссовским шумом

В этом параграфе мы докажем прямую теорему кодирования для каналов, указанных в заголовке, при ограничениях на среднюю мощность и полосу частот сигналов на входе. Частным случаем этой теоремы является случай отсутствия ограничений на полосу частот, при котором входные сигналы могут иметь какие угодно высокие частотные составляющие.

Теорема 4.4.1. Пусть $C_W^* = W \log(1 + P/WN_0)$ — информационная емкость непрерывного канала с аддитивным белым гауссовским шумом при ограничениях P на среднюю мощность и W на полосу частот входных сигналов. Для любых положительных ε и δ существует код $G(T, R)$, $R = C_W^* - \varepsilon$, слова которого удовлетворяют ограничениям на среднюю мощность

$$\frac{1}{T} \int_0^T u_i^2(t) dt \leq P, \quad i = 1, \dots, M = 2^{TR}, \quad (4.4.1)$$

и на полосу частот

$$u_i(t) = \sum_{j=1}^{2WT} x_{ik} \varphi_k(t), \quad i = 1, \dots, M = 2^{TR}, \quad 0 \leq t \leq T, \quad (4.4.2)$$

причем максимальная вероятность ошибки $\Lambda \leq \delta$.

Доказательство. Вначале мы сведем задачу передачи сообщений по непрерывному каналу с непрерывным временем с помощью кода, удовлетворяющего ограничениям на среднюю мощность и на полосу частот, к задаче передачи сообщений по непрерывному каналу с дискретным временем с помощью кода, удовлетворяющего некоторому другому ограничению на среднюю мощность, затем воспользуемся прямой теоремой кодирования, доказанной для каналов с дискретным временем.

Пусть код $G(T, R)$ удовлетворяет ограничениям P и W , указанным выше, и используется для передачи сообщений по непрерывному каналу с аддитивным белым гауссовским шумом. Тогда случайный процесс на выходе канала можно представить в виде суммы

$$Y_T(t) = X_T(t) + Z_T(t), \quad 0 \leq t \leq T, \quad (4.4.3)$$

где $X_T(t)$ — случайный процесс на входе канала (реализациями этого случайного процесса являются слова кода $G(T, R)$), а $Z(t)$ — отрезок длины T белого гауссовского шума, статистически независимого от передаваемого сигнала $X_T(t)$. Очевидно, что процесс $X_T(t)$ представим в виде ряда

$$X_T(t) = \sum_{k=1}^n X_k \varphi_k(t), \quad (4.4.4)$$

где $n = 2WT$ и

$$X_k = \int_0^T X_T(t) \varphi_k(t) dt, \quad k = 1, \dots, n. \quad (4.4.5)$$

Пусть

$$Y_k = \int_0^T Y_T(t) \varphi_k(t) dt, \quad Z_k = \int_0^T Z_T(t) \varphi_k(t) dt, \quad (4.4.6)$$

тогда

$$Y_k = X_k + Z_k, \quad k = 1, \dots, n, \quad (4.4.7)$$

где с. в. Z_k , $k = 1, \dots, n$, статистически независимы от с. в. X_l , $l = 1, \dots, n$, и, кроме того, Z_k , $k = 1, \dots, n$, являются статистически независимыми гауссовскими с. в. Они имеют нулевое математическое ожидание и дисперсию $N_0/2$. Из ограничения P на среднюю мощность кодовых слов следует, что

$$\begin{aligned} P &\geq \frac{1}{T} \int_0^T u_i^2(t) dt = \frac{1}{T} \int_0^T \sum_{k=1}^n u_{ik} u_{ik} \varphi_k(t) \varphi_k(t) dt = \\ &= 2W \frac{1}{n} \sum_{k=1}^n u_{ik}^2, \quad i = 1, \dots, M = 2^{TR}, \quad (4.4.8) \end{aligned}$$

где u_{i1}, \dots, u_{in} — коэффициенты разложения i -го кодового слова (функции $u_i(t)$) по системе ортонормальных функций $\varphi_k(t)$, $k = 1, \dots, n$. Другими словами,

$$\frac{1}{n} \sum_{i=1}^n u_{ik}^2 \leq \frac{P}{2W} \quad (4.4.9)$$

для каждого слова $u_i(t)$ кода $G(T, R)$.

Таким образом, любому коду $G(T, R)$ объема $M = 2^{TR}$, удовлетворяющему ограничениям P на среднюю мощность и W на полосу частот для непрерывного канала с аддитивным белым гауссовским шумом с интенсивностью $N_0/2$, соответствует код $G(n, R_d)$, $n = 2WT$, того же объема, удовлетворяющий ограничению $P/2W$ на среднюю мощность, для непрерывного канала без памяти с дискретным временем и аддитивным гауссовским шумом мощности $N_0/2$.

Согласно теореме 4.2.4 для любого положительного ε' существует код $G(n, R_d')$ при $R_d' = C_d^* - \varepsilon'$, где $C_d^* = \frac{1}{2} \log(1 + P/WN_0)$ — информационная емкость соответствующего канала с дискретным временем, удовлетворяющий ограничению $P/2W$ на среднюю мощность и обеспечивающий произвольно малое значение

максимальной вероятности ошибки. Если теперь каждому кодовому слову $u_i = (u_{i1}, \dots, u_{in})$ кода $G(n, R_d)$ сопоставить функцию

$$u_i(t) = \sum_{k=1}^n u_{ik} \varphi_k(t), \quad i = 1, \dots, 2^{nR_d}, \quad (4.4.10)$$

и рассматривать множество этих функций как код $G(T, R)$ для непрерывного канала, то нетрудно убедиться, что этот код будет содержать

$$2^{nR_d} = 2^{2WT} (C_d^* - \epsilon) = 2^T (C_W^* - \epsilon' \cdot 2W) = 2^{TR} \quad (4.4.11)$$

кодовых слов, где

$$R = C_W^* - \epsilon' \cdot 2W, \quad C_W^* = 2WC_d^*, \quad (4.4.12)$$

причем каждое кодовое слово будет удовлетворять ограничениям P на среднюю мощность и W на полосу частот. Так как код $G(n, R_d)$ обеспечивает произвольно малую максимальную вероятность ошибки, то и код $G(T, R)$ также обеспечивает произвольно малую максимальную вероятность ошибки, поскольку декодирование кода $G(T, R)$ можно осуществлять, выполняя разложения (4.4.5), (4.4.6) и переходя тем самым к декодированию в канале с дискретным временем. Выбирая $\epsilon' = \epsilon/2W$, получим утверждение теоремы. Теорема доказана.

Из обратной теоремы кодирования (теорема 4.3.3) и доказанной прямой теоремы вытекает следующее утверждение.

Следствие 4.4.1. Пусть $C_W^* = W \log(1 + P/N_0 W)$ — информационная емкость непрерывного канала с аддитивным белым гауссовским шумом с интенсивностью $N_0/2$ при ограничениях P на среднюю мощность и W на полосу частот входных сигналов. Пусть C — пропускная способность этого канала, т. е. такое максимальное число, что для любого положительного δ и любого $R < C$ существует код $G(T, R)$, удовлетворяющий ограничениям P на среднюю мощность и W на полосу частот кодовых слов, максимальная вероятность ошибки которого не превосходит δ . Тогда $C = C_W^*$.

В заключение мы сделаем два замечания, в которых дается обсуждение принятой модели канала и результатов, полученных в обратной и прямой теоремах кодирования для непрерывных каналов.

1. Реальный канал с точки зрения передачи электрических сигналов представляет собой фильтр. Если шум отсутствует и на вход такого фильтра подается гармоническое колебание $A \exp\{j\omega t\}$ бесконечной длительности, то на выходе возникает гармоническое колебание $AL(\omega) \exp\{j(\omega t + \alpha)\}$, имеющее также бесконечную длительность и фазовый сдвиг α . Функция $L(\omega)$ представляет собой частотную характеристику фильтра канала. Если $L(\omega) = 0$ при некотором значении частоты ω , то колебание $A \exp\{j\omega t\}$

вообще не проходит через канал. Чаще всего частотная характеристика канала формируется с помощью специальных каналообразующих устройств с целью обеспечения возможности одновременной работы нескольких абонентов. Для этого пытаются уменьшить влияние входных сигналов одного абонента на выходные сигналы другого и подбирают частотные характеристики $L_1(\omega)$ и $L_2(\omega)$ каждой пары абонентов так, чтобы произведение $L_1(\omega) \times L_2(\omega)$ было равно или почти равно нулю.

Введенное выше ограничение на полосу частот входных сигналов можно связать со свойствами частотной характеристики канала как фильтра. Пусть

$$L(\omega) = \begin{cases} 1, & \text{если } -2\pi W < \omega \leq 2\pi W, \\ 0 & \text{в противном случае.} \end{cases} \quad (4.4.13)$$

Тогда все составляющие входного сигнала, частоты которых лежат вне полосы $[-2\pi W, 2\pi W]$, не проходят через канал, и в этом случае естественно предполагать, что на выходе канала следует использовать только те сигналы, спектр которых целиком лежит в указанной полосе. Мы показали выше, что такие сигналы можно получить, если T достаточно велико и если в качестве системы ортонормальных функций использовать гармонические функции.

Таким образом, следствие 4.4.1 дает пропускную способность непрерывного канала с аддитивным белым гауссовским шумом с сигналами ограниченной мощности и идеальным полосовым фильтром.

Строго говоря, это утверждение требует некоторого уточнения. Дело в том, что спектр любого сигнала, ограниченного по длительности, отличен от нуля в бесконечной полосе частот. Поэтому часть мощности сигналов ограниченной длительности теряется и требуется учитывать величину потерянной мощности. Заметим, что можно построить сигналы длительности T , которые имеют максимальную мощность в полосе частот $[-2\pi F, 2\pi F]$. Эти сигналы строятся с помощью упомянутых выше эллиптических функций вытянутого сфера. Однако можно показать, что часть потерянной мощности стремится к нулю при $T \rightarrow \infty$. Как нетрудно увидеть из доказательства прямой теоремы, максимальная вероятность ошибки стремится к нулю также при $T \rightarrow \infty$. Используя эти соображения, можно показать, что доказанные обратная и прямая теоремы действительно дают пропускную способность непрерывного канала с идеальным полосовым фильтром.

2. Мы показали, что пропускная способность рассматриваемого канала возрастает с увеличением допустимой полосы частот W и стремится к величине $\frac{P}{N_0} \log e$. Эта величина устанавливает наивысшую скорость передачи в битах в секунду, при которой

возможно за счет усложнения кодирующих и декодирующих устройств, а также за счет увеличения задержки T достигнуть сколь угодно малой, заданной наперед, вероятности ошибки. Существуют более сильные обратные теоремы, которые показывают, что при скоростях передачи, превышающих $\frac{P}{N_0} \log e$, вероятность ошибки не только больше некоторой положительной величины, а даже стремится к единице при увеличении T . Таким образом, надежную передачу в канале с белым гауссовским шумом можно осуществлять только, когда

$$R < \frac{P}{N_0} \log e. \quad (4.4.14)$$

Обозначим через E_b энергию, затрачиваемую на передачу одной двоичной единицы информации:

$$E_b \triangleq \frac{PT}{\log M} = \frac{P}{R} \quad (4.4.15)$$

и через E_w — энергию шума в полосе частот 1 Гц за 1 секунду:

$$E_w \triangleq \frac{P_w T}{W \cdot T} = N_0, \quad (4.4.16)$$

где $P_w = N_0 W$ — мощность шума в полосе W .

Величина

$$h_0^2 \triangleq \frac{E_b}{E_w} = \frac{P}{RN_0} \quad (4.4.17)$$

называется *удельным отношением сигнал/шум* на бит в полосе 1 Гц за 1 секунду. Очевидно, h_0^2 равно отношению энергий (или мощностей) сигнала и шума, если $W = 1$, $T = 1$ и передается 1 бит информации. Если известно удельное отношение сигнал/шум, а также объем кода M , полоса частот W и время передачи T , то можно найти отношение сигнал/шум — отношение мощности сигналов к мощности помех

$$h^2 \triangleq \frac{P}{N_0 W} = h_0^2 \frac{\log M}{WT}. \quad (4.4.18)$$

Условие (4.4.14) дает нижнюю границу для удельного отношения сигнал/шум, при котором возможно осуществить надежную передачу по каналу с аддитивным белым гауссовским шумом:

$$h_0^2 > \frac{1}{\log e} = \ln 2 = 0,6931. \quad (4.4.19)$$

Если величина h_0^2 не удовлетворяет этому неравенству, то при любом методе передачи вероятность ошибки будет оставаться

большей, чем некоторое положительное число, и стремиться к единице при увеличении T к бесконечности. Если же неравенство (4.4.18) удовлетворено, то существует метод передачи (код), для которого максимальная вероятность ошибки принимает произвольно малое значение.

Задачи, упражнения и дополнения

4.1.1. Пусть X и Y — числовые оси и $f(\mathbf{y} | \mathbf{x}) = \prod_{i=1}^n f_0(y^{(i)} - x^{(i)})$, где

$f_0(z) = (\sigma \sqrt{2\pi})^{-1} \exp \{-z^2/2\sigma^2\}$. В этом случае говорят, что в непрерывном канале с дискретным временем действует аддитивный гауссовский шум с мощностью σ^2 .

а) Пусть $n = 2$, $\sigma^2 = 1$ и для передачи используется код, состоящий из двух равновероятных кодовых слов $\mathbf{u}_1 = (1, -1)$ и $\mathbf{u}_2 = (-1, 1)$. Пусть решающие области выбраны следующим образом:

$$\begin{aligned} A_1 &= \{(y^{(1)}, y^{(2)}): y^{(1)} > 0, y^{(2)} < 0\}, \\ A_2 &= \{(y^{(1)}, y^{(2)}): y^{(1)} < 0, y^{(2)} > 0\}. \end{aligned}$$

Найдите $\lambda_1, \lambda_2, \lambda$, Λ .

б) Пусть в условиях п. а) решающие области таковы: $A_1 = \{(y^{(1)}, y^{(2)}): y^{(1)} > 0, y^{(2)} > 0\}$, $A_2 = \{(y^{(1)}, y^{(2)}): y^{(1)} < 0, y^{(2)} < 0\}$. Найдите те же вероятности ошибок. Какое из двух разбиений пространства Y^2 на решающие области лучше? Почему? Как должны быть выбраны кодовые слова для того, чтобы худшее разбиение стало лучшим? Зависит ли этот выбор от величины σ^2 ?

в) Как изменятся вероятности ошибок при фиксированных решающих областях, если выбрать $\mathbf{u}_1 = (2, -2)$, $\mathbf{u}_2 = (-2, 2)$? Тот же вопрос, если, кроме того, σ увеличилось в два раза.

Указание: при расчетах использовать таблицы интеграла вероятностей.

4.1.2. Пусть X — числовая ось. Предположим, что на X^n задано распределение вероятностей посредством ф. п. в.

$$f_n(\mathbf{x}) = \prod_{i=1}^n f(x^{(i)}),$$

где $f(x) = (\sqrt{2\pi})^{-1} \exp \{-x^2/2\}$. Пусть

$$\Phi(z) \triangleq \int_{-\infty}^z (\sqrt{2\pi})^{-1} \exp \{-x^2/2\} dx.$$

Покажите, что вероятность появления последовательности $\mathbf{x} \in X^n$, удовлетворяющей ограничению P на среднюю мощность, равна

$$\int_{-\sqrt{2P}}^{\sqrt{2P}} [\Phi(\sqrt{2P} - x^2) - \Phi(-\sqrt{2P} - x^2)] f(x) dx.$$

Указание: если $\mathbf{x} = (x_1, x_2)$ удовлетворяет ограничению P , то $x_1^2 + x_2^2 \leq 2P$.

4.1.3. Говорят, что последовательность $(x_1, \dots, x_n) \in X^n$ удовлетворяет ограничению на пиковую мощность, если $\max\{x_1^2, \dots, x_n^2\} \leq P$. Покажите, что в условиях предыдущей задачи вероятность появления такой последовательности при $n = 2$ равна $[\Phi(\sqrt{P}) - \Phi(-\sqrt{P})]^2$. Выпишите общую формулу для этой вероятности при произвольном n .

4.1.4. Пусть $S_n(0)$ — n -мерный шар радиуса $\sqrt{n}P$ с центром в точке 0: $S_n(0) \triangleq \{x : (xx^T) \leq nP\}$. Предположим, что слова u_1, \dots, u_M кода $G(n, R)$ выбираются как точки $S_n(0)$, причем расстояние $d(u_i, u_j) \triangleq \sqrt{(u_i - u_j)(u_i - u_j)^T}$ между любыми двумя точками u_i и u_j не меньше чем $\sqrt{n}\alpha$, где α — некоторое фиксированное число. Покажите, что для любого такого кода $G(n, R)$ скорость удовлетворяет неравенству $R \leq \log(1 + 2\sqrt{P}/\alpha)$. Указание: воспользуйтесь тем, что объем n -мерного шара радиуса r равен k_nr^n , где коэффициент k_n не зависит от r и от выбора центра шара, а зависит только от размерности n .

4.1.5. Предположим, что в коде $G(100, 2)$ любые два слова находятся на расстоянии 20 или больше: $d(u_i, u_j) \geq 20$ при всех $i \neq j$. Может ли для всех слов этого кода выполняться условие $(u_i u_i^T) \leq nP$ при $P = 8$, при $P = 10$?

4.2.1. Подберите ограничение на входе непрерывного канала без памяти с аддитивным гауссовским шумом так, чтобы информационная емкость была равна 0,1, 1, 10.

4.2.2. Предположим, что в непрерывном канале без памяти с аддитивным шумом распределение вероятностей шума отличается от гауссовского и имеет относительную энтропию $H_0(Z)$. Величина $\tilde{\sigma}^2$, определяемая из уравнения $H_0(Z) = \frac{1}{2} \log 2\pi eB$, называется *энтропийной мощностью шума*. Покажите, что информационная емкость рассматриваемого канала имеет следующую оценку:

$$C^* \geq \frac{1}{2} \log \left(1 + \frac{P}{\tilde{\sigma}^2} \right).$$

4.2.3. Покажите, что информационная емкость при ограничении на пиковую мощность (см. задачу 4.1.3) непрерывного канала без памяти с аддитивным гауссовским шумом не больше информационной емкости такого же канала при той же величине ограничения на среднюю мощность.

4.2.4. Непрерывным m -мерным векторным каналом без памяти с дискретным временем называется канал, входные и выходные сигналы которого представляют собой элементы действительного векторного пространства V_m размерности m . Векторный канал называют системой параллельных каналов, если

$$f(\mathbf{y}_m | \mathbf{x}_m) = \prod_{i=1}^m f_i(y_i | x_i), \quad \mathbf{x}_m, \mathbf{y}_m \in V_m,$$

где $\mathbf{x}_m = (x_1, \dots, x_m)$ — входной, $\mathbf{y}_m = (y_1, \dots, y_m)$ — выходной сигналы канала. Рассмотрим случай, когда каждый из подканалов системы m параллельных каналов является каналом с аддитивным гауссовским шумом, т. е.

$$f_i(y_i | x_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma_i^2} (y_i - x_i)^2 \right\}.$$

Говорят, что последовательность $(\mathbf{x}_m^{(1)}, \dots, \mathbf{x}_m^{(n)})$ на входе такого канала удовлетворяет ограничению P на среднюю мощность, если

$$\frac{1}{n} \sum_{j=1}^n \sum_{i=1}^m (x_i^{(j)})^2 \leq P.$$

а) Информационная емкость C^* системы параллельных каналов определяется аналогично определению 4.1.3. Покажите, что

$$C^* = \max_{P_1, \dots, P_m} \sum_{i=1}^m \frac{1}{2} \log \left(1 + \frac{P_i}{\sigma_i^2} \right),$$

где максимум отыскивается по таким значениям мощностей P_1, \dots, P_m , $P_i \geq 0$, $i = 1, \dots, m$, в подканалах, для которых $\sum_{i=1}^m P_i \leq P$.

б) Используя теорему Куна—Таккера, покажите, что распределение мощностей P_1, \dots, P_m , максимизирующее C^* , есть

$$P_i = \begin{cases} B - \sigma_i^2, & \text{если } \sigma_i^2 < B, \\ 0 & \text{в противном случае,} \end{cases}$$

где B определяется из условия

$$\sum_{i=1}^m \max \{B - \sigma_i^2, 0\} = P.$$

в) Пусть $m = 4$ и $\sigma_1^2 = 4$, $\sigma_2^2 = 1$, $\sigma_3^2 = 2$, $\sigma_4^2 = 3$. Найдите параметр B и определите, какие из подканалов должны использоваться при передаче (в каких подканалах $P_i > 0$), если средняя мощность входных сигналов равна 3,5. Указание: воспользуйтесь результатом п. б).

г) В условиях п. в) оцените допустимую среднюю мощность P на входе, если известно, что информационная емкость достигается при использовании двух подканалов. При какой мощности используются четыре подканала?

4.3.1. Пусть по каналу с непрерывным временем и аддитивным белым гауссовским шумом с интенсивностью $N_0/2$ передается один из двух сигналов $u_1(t)$ или $u_2(t)$ длительности T , причем

$$u_j(t) = \sum_{i=1}^n u_{ji} \varphi_i(t), \quad j = 1, 2,$$

где

$$u_{ji} = \int_0^T u_j(t) \varphi_i(t) dt, \quad j = 1, 2, \quad i = 1, \dots, n,$$

и $\{\varphi_i(t)\}$, $i = 1, \dots, n$, — некоторая система ортонормированных функций. Предположим, что приемник по сигналу $y(t) = u_j(t) + Z(t)$ на выходе канала вычисляет величины

$$y_i = \int_0^T y(t) \varphi_i(t) dt, \quad i = 1, \dots, n.$$

а) Покажите, что

$$y_i = \begin{cases} u_{1i} + Z_i, & \text{если передавался сигнал } u_1(t), \\ u_{2i} + Z_i, & \text{если передавался сигнал } u_2(t), \end{cases}$$

где Z_i — гауссовская случайная величина с нулевым математическим ожиданием и дисперсией $N_0/2$.

б) Покажите, что

$$f(y|u_j(t)) = (\pi N_0)^{-n/2} \exp \left\{ -\frac{1}{N_0} \sum_{i=1}^n (y_i - u_{ji})^2 \right\}, \quad i = 1, 2,$$

где $y = (y_1, \dots, y_n)$.

в) Покажите, что логарифм отношения правдоподобия

$$l(y) \triangleq \ln \frac{f(y|u_1(t))}{f(y|u_2(t))} = \frac{2}{N_0} \sum_{i=1}^n y_i (u_{1i} - u_{2i}) - \frac{1}{N_0} \sum_{i=1}^n (u_{1i}^2 - u_{2i}^2)$$

и является гауссовой с. в. с математическим ожиданием $1/N_0 \sum_{i=1}^n (u_{1i} - u_{2i})^2$

и дисперсией $2/N_0 \sum_{i=1}^n (u_{1i} - u_{2i})^2$ при условии, что передавался сигнал $u_1(t)$. Найдите математическое ожидание и дисперсию с. в. $l(y)$ при условии, что передавался сигнал $u_2(t)$.

г) Покажите, что при декодировании по максимуму правдоподобия выполняется решение о том, что передавался сигнал $u_1(t)$, если $l(y) > 0$, и противоположное решение, если $l(y) < 0$. Покажите, что вероятность ошибки при таком декодировании не зависит от передаваемого сигнала и равна

$$P_e = \Phi \left(- \left(\frac{1}{2N_0} \sum_{i=1}^n (u_{1i} - u_{2i})^2 \right)^{1/2} \right),$$

где функция $\Phi(z)$ определена в задаче 4.1.2.

д) Пусть энергии сигналов $u_1(t)$ и $u_2(t)$ одинаковы и равны E , т. е.

$$\int_0^T u_1^2(t) dt = \int_0^T u_2^2(t) dt = E.$$

Обозначим через μ коэффициент корреляции между сигналами

$$\mu \triangleq \frac{1}{E} \int_0^T u_1(t) u_2(t) dt.$$

Покажите, что

$$P_e = \Phi \left(- \sqrt{\frac{(1-\mu)E}{N_0}} \right).$$

Покажите также, что

$$l(y) = \frac{2}{N_0} \int_0^T y(t) (u_1(t) - u_2(t)) dt.$$

Следовательно, декодирование по максимуму правдоподобия можно представить как сравнение с. в. $l_1 \triangleq \int_0^T y(t) u_1(t) dt$ и $l_2 \triangleq \int_0^T y(t) u_2(t) dt$. Если первая из них больше, то выносится решение в пользу $u_1(t)$, в противном случае — в пользу $u_2(t)$. Покажите, что это представление не зависит от того, чему равно

n , и справедливо также, когда ряды, задающие сигналы $u_1(t)$ и $u_2(t)$, — бесконечные.

Покажите, что для любого значения отношения сигнал/шум $h^2 \triangleq 2E/N_0$ вероятность ошибки минимизируется при $u_1(t) = -u_2(t)$. Такие сигналы называются противоположными. Найдите вероятность ошибки для двух ортогональных сигналов, т. е. таких, для которых $\int_0^T u_1(t) u_2(t) dt = 0$. Сопоставьте противоположные и ортогональные сигналы. Пусть $P_e(\text{пр})$ и $P_e(\text{орт})$ — вероятности ошибок соответственно для противоположных и ортогональных сигналов одинаковой энергии. Покажите, что при большом значении h^2

$$P_e(\text{пр}) \approx P_e^2(\text{орт}).$$

Указание: воспользуйтесь оценками функции $\Phi(z)$, приведенными в задаче 2.2.9.

4.4.1. Пусть по каналу с белым гауссовским шумом с интенсивностью $N_0/2$ надо передать последовательность равновероятных и независимых двоичных символов, поступающих на вход передающего устройства со скоростью R бит/с. Непосредственный метод передачи, или метод передачи без кодирования, состоит в следующем. Каждому двоичному символу сопоставляется один из двух сигналов $u_1(t)$ или $u_2(t)$ длительности $T_0 = 1/R$ и энергии E . При этом достижение малой вероятности ошибки возможно лишь за счет увеличения отношения сигнал/шум $h^2 = 2E/N_0$ (см. задачу 4.3.1 (д)). При фиксированной скорости R увеличение h^2 возможно лишь за счет увеличения энергии сигнала E , следовательно, за счет увеличения мощности передатчика $P = E/T_0 = ER$.

Естественно поставить вопрос: возможна ли передача с произвольно малой вероятностью ошибки при фиксированной, а не сколь угодно большой мощности передатчика? Положительный ответ на этот вопрос дает теорема кодирования для канала с белым гауссовским шумом. Рассмотрим следующий метод передачи (передача с кодированием). Двоичная последовательность, подлежащая передаче, разбивается на блоки по k символов в каждом. Каждый блок кодируется кодом $G(T, R)$ для канала с белым гауссовским шумом, где $T = kR$. Из прямой теоремы кодирования для канала с белым гауссовским шумом следует, что если $R < C$, где C — пропускная способность канала, то при достаточно большом T существует код $G(T, R)$, вероятность ошибки декодирования которого не больше заданной величины. Пусть $h_0^2 = \frac{P}{N_0 R}$, тогда h_0^2 есть удельное отношение сигнал/шум на бит, так как величина P/R представляет собой энергию, которая расходуется на передачу одного двоичного символа.

В § 4.4 было показано, что для канала с белым гауссовским шумом без ограничения на полосу частот входных сигналов минимальное значение h_0^2 , при котором возможна передача с произвольно малой вероятностью ошибки, не зависит от R и равно $\ln 2$.

Постройте график минимального удельного отношения сигнал/шум h_0^2 в зависимости от допустимой полосы частот W , для которого возможна передача с произвольно малой вероятностью ошибки при $R = 10^2$ бит/с, $R = 10^4$ бит/с.

4.4.2. Рассмотрим передачу сообщений по каналу с аддитивным белым гауссовским шумом с помощью ортогональных сигналов. Пусть $\{u_1(t), \dots, u_M(t)\}$ — система ортогональных на $(0, T)$ сигналов:

$$\int_0^T u_i(t) u_j(t) dt = \begin{cases} E & \text{при } i = j, \\ 0 & \text{при } i \neq j. \end{cases}$$

Предположим, что для каждого сигнала $u_j(t)$ вычисляется величина

$$l_j \triangleq \int_0^T Y(t) u_j(t) dt,$$

где $Y(t)$ — сигнал на выходе канала, который представляет собой сумму передаваемого сигнала и шума: $Y(t) = u(t) + Z(t)$. Приемник принимает решение в пользу сигнала $u_j(t)$ (в пользу сообщения u_j), если $l_j = \max\{l_1, \dots, l_M\}$.

а) Покажите, что с. в. l_1, \dots, l_M являются гауссовскими и статистически независимыми, причем

$$M_{u_k}(l_j) = \begin{cases} E & \text{при } k=j, \\ 0 & \text{при } k \neq j \end{cases}$$

и

$$M_{u_k}(l_i - M_{u_k} l_i)(l_j - M_{u_k} l_j) = \begin{cases} \frac{N_0 E}{2} & \text{при } i=j, \\ 0 & \text{при } i \neq j. \end{cases}$$

б) Пусть Q_1 — вероятность правильной передачи сообщения u_1 . Покажите, что

$$Q_1 = \Pr(l_2 < l_1, \dots, l_M < l_1 | u_1) =$$

$$\begin{aligned} &= \int_{-\infty}^{\infty} \Pr(l_2 < x | u_1) \dots \Pr(l_M < x | u_1) (\pi N_0 E)^{-1/2} \exp\left\{-\frac{(x-E)^2}{N_0 E}\right\} dx = \\ &= \frac{1}{\sqrt{\pi N_0 E}} \int_{-\infty}^{\infty} \Phi^{M-1}\left(x \sqrt{\frac{2}{N_0 E}}\right) \exp\left\{-\frac{(x-E)^2}{N_0 E}\right\} dx = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \Phi^{M-1}\left(t + \sqrt{\frac{2E}{N_0}}\right) \exp\left\{-\frac{t^2}{2}\right\} dt, \end{aligned}$$

где функция $\Phi(x)$ определена в задаче 4.1.2. Покажите, что вероятность Q_1 правильной передачи сообщения u_i не зависит от i и равна Q_1 .

в) Покажите, что аддитивная граница вероятности ошибки при передаче сообщения u_1 имеет вид

$$p_1 = 1 - Q_1 < (M-1) \Phi\left(\sqrt{\frac{E}{N_0}}\right).$$

4.4.3. Здесь будет исследовано поведение вероятности ошибки $p_1 = 1 - Q_1$, определенной в предыдущей задаче, при $M \rightarrow \infty$ и при выполнении следующих условий: 1) мощность передатчика фиксирована и равна P ; 2) скорость передачи, определяемая соотношением $R = (\log M)/T$, фиксирована и равна R .

При указанных условиях

$$\frac{2E}{N_0} = \frac{2P \log M}{RN_0}$$

и, следовательно, в формуле для вероятности Q_1 , полученной в предыдущей задаче, имеется неопределенность при $M \rightarrow \infty$. Положим

$$\varphi(M) \triangleq \ln \Phi^{M-1}\left(t + \sqrt{\frac{2E}{N_0}}\right).$$

а) Покажите, что при всех ограниченных значениях t

$$\lim_{M \rightarrow \infty} \varphi(M) = \lim_{M \rightarrow \infty} \frac{\frac{d}{dM} \ln \Phi\left(t + \sqrt{\frac{2E}{N_0}}\right)}{\frac{d}{dM} \left(\frac{1}{M-1}\right)} =$$

$$= \begin{cases} -\infty, & \text{если } R > (P \log e)/N_0, \\ 0, & \text{если } R < (P \log e)/N_0. \end{cases}$$

Следовательно, $\lim Q_1 = 0$ при $R > \frac{P \log e}{N_0} = C_\infty^*$ и $\lim Q_1 = 1$ в противном случае. Фактически доказана прямая теорема кодирования для канала с белым гауссовским шумом и неограниченной полосой частот. Более того, доказано, что произвольно малая вероятность ошибки в таком канале достигается при использовании достаточно большого числа ортогональных сигналов.

б) Используя аддитивную границу для вероятности ошибки, полученную в предыдущей задаче (п. в.), покажите, что вероятность ошибки для системы ортогональных сигналов удовлетворяет следующей (так называемой аддитивной) границе

$$p_1 < 2^{-T} \left(\frac{C_\infty^*}{2} - R \right).$$

Указание: воспользуйтесь неравенством для функции $\Phi(x)$, приведенным в задаче 2.2.9.

КРАТКИЙ ИСТОРИЧЕСКИЙ КОММЕНТАРИЙ И ЛИТЕРАТУРА

Впервые формулы для пропускных способностей непрерывного канала без памяти с дискретным временем с аддитивным гауссовским шумом и канала непрерывного времени с аддитивным белым гауссовским шумом при ограничении на среднюю мощность входных сигналов были получены К. Шенноном [3, 4]. Пропускная способность каналов с аддитивным небелым гауссовским шумом была найдена К. Шенноном [4], строгое обоснование ее вывода было позднее дано М. С. Пинскером [2]. Более полное, чем в настоящей книге, исследование кодирования в канале с гауссовским шумом (в частности, кодирование в каналах с фильтрами) содержится в книге Р. Галлагера [1]. В этой книге рассмотрено кодирование и для более общих моделей каналов.

1. Галлагер (Gallager R. G.) *Information Theory and Reliable Communication*. New York-Wiley, 1968. [Русский перевод: Галлагер Р. Теория информации и надежная связь. — М.: Советское радио, 1974.]
2. Пинскер М. С. Вычисление скорости создания сообщений стационарным случайным процессом и пропускной способности стационарного канала, Докл. Акад. наук СССР, 1956, 111, 4.
3. Шеннон (Shannon C. E.) *A Mathematical Theory of Communications*, Bell Syst. Tech. J., 1948, 27, 379–423 (Part I), 623–656 (Part II). [Русский перевод: Шеннон К. Математическая теория связи. — В сб.: Работы по теории информации и кибернетике. — М.: ИЛ, 1963.]
4. Шеннон (Shannon C. E.) *Communication in the Presence of Noise*. — Proc. IRE, 1949, 37, 10–21. [Русский перевод: Шеннон К. Связь при наличии шума. — В сб.: Работы по теории информации и кибернетике. — М.: ИЛ, 1963.]

КОДИРОВАНИЕ ИСТОЧНИКОВ С ЗАДАННЫМ КРИТЕРИЕМ КАЧЕСТВА

В этой главе мы вернемся к кодированию источников, которое рассматривалось в первой главе, с некоторым, однако, отличием в постановке задачи. Ранее требовалось определить наименьшее количество кодовых символов на сообщение источника, при котором сообщения возможно восстановить точно или со сколь угодно малой вероятностью ошибки по выходной последовательности кодера. Теперь не будем требовать точного или сколь угодно точного восстановления. Мы введем понятие критерия качества и ошибки восстановления, связанной с этим критерием качества, и потребуем, чтобы восстановление осуществлялось с ошибкой, не превосходящей заданное значение. Вопрос, который при этом будет нас интересовать, по-прежнему заключается в определении наименьшего количества кодовых символов на сообщение источника или наименьшей достижимой скорости кодирования, при которой такое восстановление можно осуществить.

Задача кодирования при заданном критерии качества естественно возникает при кодировании непрерывных источников. Действительно, всякий прибор, измеряющий сигналы на выходе непрерывных источников, обязательно вносит ошибки измерений. С другой стороны, если такой источник сопрягается с вычислительной машиной или цифровым вычислительным устройством, то его сообщения могут быть обработаны только с некоторой ошибкой, связанной с дискретностью устройства. Очевидно, что ошибка может быть сделана тем меньшей, чем больше символов некоторого алфавита, например, чем больше десятичных символов используется для представления одного сообщения непрерывного источника. Задача заключается в том, чтобы установить минимальное количество символов на сообщение, при котором величина ошибки не превосходит заданное значение.

При кодировании дискретных источников также можно представить себе ситуацию, в которой возникает задача кодирования с заданным критерием качества. Предположим, что с точки зрения некоторого получателя сообщения источника избыточны и ему достаточно иметь только какую-то часть из них. Пусть, например, дискретный датчик давления в кабине космического корабля измеряет давление с точностью 5 мм рт. ст. и каждое измерение имеет форму семиразрядного двоичного числа. Пусть также

имеется контролирующая система, которая должна срабатывать, когда давление выйдет из допустимых пределов. С точки зрения получателя, которым является эта система, измеренные значения давления избыточны. Можно ввести критерий качества, адекватный задаче контроля, и задаться вопросом о том, какое количество двоичных символов в этом случае необходимо. Понятно, что это количество будет существенно меньшим, чем семь двоичных символов.

Другой важный случай, когда возникает задача кодирования с критерием качества, связан с передачей сообщений по каналам связи. Предположим, что сообщения некоторого источника передаются по каналу связи. Если в канале есть шум, то, как было показано в предыдущих главах, возможно так закодировать передаваемые сообщения, чтобы при декодировании ошибки появлялись бы со сколь угодно малой вероятностью, т. е. так, чтобы шум в канале практически не влиял бы на передачу сообщений. Единственное требование, которое должно быть при этом удовлетворено, состоит в том, чтобы количество информации, подаваемое в единицу времени на вход канала, не было слишком большим, точнее, чтобы оно не превосходило константы C , определяемой каналом и называемой пропускной способностью канала. Пусть H есть количество информации в единицу времени, порождаемое источником. Если $H < C$, то возможно передать сообщения источника через канал так, чтобы они восстанавливались со сколь угодно малой вероятностью ошибки. Но если $H > C$, то этого сделать нельзя и при декодировании почти всегда будут возникать ошибки.

Если теперь ввести критерий качества и в соответствии с ним определять численное значение ошибки, то можно поставить вопрос о том, какова наименьшая достижимая ошибка, возникающая при передаче сообщений данного источника по данному каналу. Нетрудно понять, что ошибка связана с количеством информации H , порождаемым источником. Если обозначить через $H(\epsilon)$ количество информации в единицу времени при ошибке ϵ , то наименьшая достижимая ошибка будет равна корню уравнения $H(\epsilon) = C$.

§ 5.1. Критерии качества. Постановка задачи кодирования с заданным критерием качества

Пусть источник U_X в каждый момент времени выбирает сообщение из множества X , которое может быть как дискретным, так и непрерывным, и X^n — множество последовательностей сообщений $\mathbf{x} = (x^{(1)}, \dots, x^{(n)})$, которые порождает этот источник за n последовательных моментов времени. Предположим, что последовательность сообщений $(x^{(1)}, \dots, x^{(n)})$ должна быть представлена с помощью аппроксимирующей последовательности

$\mathbf{y} = (y^{(1)}, \dots, y^{(n)})$ из элементов, вообще говоря, некоторого другого множества Y . Нас будет интересовать только такой случай, когда все, быть может бесконечное, множество X^n последовательностей сообщений источника представляется с помощью конечного множества, содержащего M различных аппроксимирующих последовательностей $\mathbf{y}_1, \dots, \mathbf{y}_M \in Y^n$.

Введем в рассмотрение функцию $d_n(\mathbf{x}, \mathbf{y})$, каждое значение которой будем считать величиной ошибки, возникающей при аппроксимации последовательности \mathbf{x} с помощью последовательности \mathbf{y} . Мы будем считать также, что

$$d_n(\mathbf{x}, \mathbf{y}) = \frac{1}{n} \sum_{t=1}^n d(x^{(t)}, y^{(t)}), \quad (5.1.1)$$

где $d(x, y)$, $x \in X$, $y \in Y$, — некоторая неотрицательная функция, задающая величину ошибки и называемая *критерием качества*. Критерий качества задает величину ошибки при аппроксимации буквы x в последовательности \mathbf{x} соответствующей буквой y в последовательности \mathbf{y} и поэтому иногда называется побуквенным критерием качества. Для побуквенного критерия качества $d_n(\mathbf{x}, \mathbf{y})$ представляет собой среднюю величину из ошибок представления n компонент последовательности \mathbf{x} . Мы будем предполагать, что $d(x, y) \geq 0$ для всех $x \in X$ и $y \in Y$.

В дальнейшем нас будут интересовать различные способы сопоставления последовательностей сообщений $\mathbf{x} \in X^n$ и аппроксимирующих последовательностей $\mathbf{y}_1, \dots, \mathbf{y}_M$. Один из возможных способов — детерминированный. При появлении на выходе источника последовательностей \mathbf{x} для аппроксимации используется вполне определенная последовательность \mathbf{y} , которую можно обозначить через $\mathbf{y}(\mathbf{x})$. Имеется, вообще говоря, много (в непрерывном случае — бесконечно много) последовательностей $\mathbf{x} \in X^n$, которые аппроксимируются с помощью одной и той же последовательности \mathbf{y} , так что стображение $\mathbf{y}(\mathbf{x})$ не взаимно однозначно. В некоторых случаях удобно рассматривать случайный способ сопоставления, предполагая, что имеется некоторый вероятностный механизм аппроксимации, который описывается с помощью условных вероятностей $p(y|\mathbf{x})$ использования последовательности \mathbf{y} для аппроксимации последовательности \mathbf{x} . Детерминированный способ аппроксимации также может быть описан с помощью условных вероятностей $p(y|\mathbf{x})$, которые в этом случае задают вырожденное распределение, т. е.

$$p(y|\mathbf{x}) = \begin{cases} 1, & \text{если } \mathbf{y} = \mathbf{y}(\mathbf{x}), \\ 0 & \text{в остальных случаях.} \end{cases} \quad (5.1.2)$$

Таким образом, вероятностная аппроксимация включает в себя как частный случай детерминированную. В дальнейшем детерминированная аппроксимация будет называться кодированием (более точное определение кодирования будет дано ниже).

При рассмотрении непрерывных источников аппроксимирующее множество Y , как правило, также непрерывно (представляет собой числовую ось). Поэтому случайную аппроксимацию задают с помощью функции плотности вероятностей $f(y|\mathbf{x})$. Детерминированной аппроксимации при этом соответствует обобщенная ф. п. в

$$f(y|\mathbf{x}) = \delta(y - \mathbf{y}(\mathbf{x})), \quad \mathbf{x} \in X^n, \quad y \in Y^n. \quad (5.1.3)$$

Пусть задан источник U_X . Это означает, что для любого n задано распределение вероятностей $p(\mathbf{x})$, $\mathbf{x} \in X^n$ (в дискретном случае), или ф. п. в. $f(\mathbf{x})$, $\mathbf{x} \in X^n$ (в непрерывном случае). Тогда для каждого условного распределения вероятностей $p(y|\mathbf{x})$, $\mathbf{x} \in X^n$, $y \in Y^n$, или условной ф. п. в. $f(y|\mathbf{x})$, $\mathbf{x} \in X^n$, $y \in Y^n$, задан ансамбль $\{X^n Y^n, p(\mathbf{x}, y) = p(\mathbf{x}) p(y|\mathbf{x})\}$ (в дискретном случае) или ансамбль $\{X^n Y^n, f(\mathbf{x}, y) = f(\mathbf{x}) f(y|\mathbf{x})\}$ (в непрерывном случае). Очевидно, что функция $d_n(\mathbf{x}, \mathbf{y})$ представляет собой случайную величину на ансамбле $X^n Y^n$. Ее математическое ожидание \bar{d}_n называется средней ошибкой

$$\begin{aligned} \bar{d}_n &\triangleq \sum_{X^n Y^n} p(\mathbf{x}, y) d_n(\mathbf{x}, y) && \text{в дискретном случае,} \\ \bar{d}_n &\triangleq \int_{X^n Y^n} f(\mathbf{x}, y) d_n(\mathbf{x}, y) dx dy && \text{в непрерывном случае.} \end{aligned} \quad (5.1.4)$$

Пусть $X_i Y_i$ — ансамбль пар сообщений $(x^{(i)}, y^{(i)})$, соответствующих моменту времени i , $i = 1, 2, \dots, n$. Очевидно, что функция $d(x^{(i)}, y^{(i)})$ является случайной величиной на ансамбле $X_i Y_i$, ее математическое ожидание $\bar{d}^{(i)}$ называется средней ошибкой аппроксимации i -й буквы. Так как математическое ожидание суммы равно сумме математических ожиданий, то из (5.1.5) следует, что

$$\bar{d}_n = \frac{1}{n} \sum_{i=1}^n \bar{d}^{(i)}. \quad (5.1.5)$$

Пример 5.1.1. Пусть $X = Y = \{a_1, \dots, a_L\}$ — два дискретных множества, состоящие из одного и того же набора элементов. Положим

$$d(x, y) \triangleq \begin{cases} 1, & \text{если } x \neq y, \\ 0, & \text{если } x = y. \end{cases} \quad (5.1.6)$$

Математическое ожидание с. в. $d(x, y)$ на ансамбле XY

$$Md(x, y) = \sum_{XY} d(x, y) p(x, y) = \sum_{x \neq y} p(x, y) = \Pr(x \neq y) \quad (5.1.7)$$

и

$$\overline{d_n} = \frac{1}{n} \sum_{i=1}^n \Pr(x^{(i)} \neq y^{(i)}).$$

В этом случае средняя ошибка просто равна средней вероятности того, что символы в последовательностях x и y не будут совпадать. Такой критерий качества называется *вероятностным*.

Предположим, что $L = 2$, $n = 3$ и на множестве $X = \{0, 1\}$ задано равномерное распределение вероятностей. Предположим также, что аппроксимирующее множество состоит из двух последовательностей $y_1 = (000)$ и $y_2 = (111)$. В соответствии с (5.1.1) и (5.1.6) величина $d_n(x, y)$ равна относительному числу символов, в которых последовательности x и y не совпадают. Например,

$$d_n((010), (111)) = 2/3.$$

Рассмотрим неслучайное сопоставление элементов множества X^3 и последовательностей y_1 , y_2 , при котором $y(x)$ есть такая последовательность, которая минимизирует величину $d_n(x, y)$. Это сопоставление приведено в следующей таблице.

Таблица 5.1.1

(Последовательность источника x)	Аппрокси-мирующая последовательность $y(x)$	Ошибка $d_n(x, y)$	(Последовательность источника x)	Аппрокси-мирующая последовательность $y(x)$	Ошибка $d_n(x, y)$
0 0 0	0 0 0	0	1 0 0	0 0 0	1/3
0 0 1	0 0 0	1/3	1 0 1	1 1 1	1/3
0 1 0	0 0 0	1/3	1 1 0	1 1 1	1/3
0 1 1	1 1 1	1/3	1 1 1	1 1 1	0

Нетрудно найти, что средняя ошибка $d_s = 1/4$. Каждая из аппроксимирующих последовательностей появляется с вероятностью $1/2$, и, следовательно, энтропия аппроксимирующих последовательностей равна 1. Это значит, что существует однозначно декодируемый двоичный код из двух слов, для которого требуется один двоичный символ на аппроксимирующую последовательность или один двоичный символ на три сообщения источника. При этом сообщения источника будут восстанавливаться со средней ошибкой $1/4$.

Приведенный расчет показывает, что введение критерия качества и восстановление сообщений источника не точно, а с заданным уровнем ошибок, понижает по крайней мере в этом примере скорость кодирования. Заметим, что при точном восстановлении скорость равна 1 бит на сообщение, тогда как в рассматриваемом примере она равна $1/3$ бит на сообщение.

Пример 5.1.2. Предположим, что множество X выбрано, как в предыдущем примере, а множество Y совпадает с X , но содержит дополнительный $(L+1)$ -й элемент; обозначим его через a_0 . Пусть

$$d(x, y) \triangleq \begin{cases} 1, & \text{если } x \neq y \text{ и } y \neq a_0, \\ 0, & \text{если } x = y, \\ a, & \text{если } y = a_0. \end{cases}$$

В этом случае сообщение a_0 можно трактовать как стирание. Если α существенно меньше 1, то эта функция качества соответствует случаю, когда ошибки существенно более нежелательны, чем стирания.

Пример 5.1.3. Пусть X и Y — числовые оси и сообщения источника представляют собой действительные случайные величины. Положим

$$d(x, y) \triangleq (x - y)^2. \quad (5.1.8)$$

Такой критерий качества приемлем, если при аппроксимации особенно нежелательны большие ошибки. Математическое ожидание

$$Md(x, y) = M(X - Y)^2 \quad (5.1.9)$$

есть средний квадрат ошибки или дисперсия ошибки, если $M(X - Y) = 0$. Поэтому такой критерий качества называется *квадратическим*, а соответствующая средняя ошибка $\overline{d_n}$ — среднеквадратической.

Пример 5.1.4. Ниже мы покажем, что критерий качества можно подобрать так, чтобы выделить интересующие получателя свойства сообщений. Вернемся к примеру, который был приведен во введении к этой главе. Пусть имеется цифровой датчик давления и каждое измерение имеет форму семиразрядного двоич-

ного числа $x = (x^{(0)}, x^{(1)}, \dots, x^{(6)})$, причем величина давления $F \triangleq \alpha \sum_{i=0}^6 x^{(i)} 2^i$,

где α — некоторый нормирующий коэффициент. Будем для простоты считать, что давление представляет собой случайную величину, которая выходит из допустимых пределов (F_H , F_B) с вероятностью p и остается в допустимых пределах с вероятностью $q = 1 - p$.

Введем критерий качества и закодируем сообщения в соответствии с этим критерием так, чтобы по выходной последовательности кодера можно было бы однозначно определить, выходила или нет измеряемая величина из допустимых пределов. Пусть каждое сообщение x аппроксимируется семиразрядным двоичным числом $y = (y^{(0)}, y^{(1)}, \dots, y^{(6)})$. Положим $\tilde{F} \triangleq \alpha \sum y^{(i)} 2^i$ и

$$d(x, y) \triangleq \begin{cases} 1, & \text{если } \tilde{F} \in (F_H, F_B), \tilde{F} \notin (F_H, F_B) \text{ или } \tilde{F} \notin (F_H, F_B), \tilde{F} \in (F_H, F_B), \\ 0 & \text{в остальных случаях,} \end{cases} \quad (5.1.10)$$

и выберем аппроксимирующее множество из двух чисел y_1 , y_2 таких, что y_1 — любое число интервала $(F_H/\alpha, F_B/\alpha)$, а y_2 — любое число вне этого интервала. Если в качестве аппроксимирующего выбирается число, минимизирующее $d(x, y)$ при заданном x , то, очевидно, средняя ошибка d будет равна нулю. При этом интересующее получателя событие — выход давления из допустимых пределов — определяется по y однозначно. Очевидно, сообщение y_1 имеет вероятность q , а сообщение y_2 — вероятность p , поэтому энтропия аппроксимирующего ансамбля равна $h(p)$. Из результатов первой главы следует, что $h(p)$ есть минимальное количество двоичных символов на сообщение, которое позволяет кодировать заданный источник относительно критерия качества (5.1.10) с нулевой ошибкой.

Теперь дадим основные определения, относящиеся к задаче кодирования источников с заданным критерием качества.

Пусть X — множество сообщений источника и Y — множество аппроксимирующих символов. Пусть n — некоторое натуральное число — длина кодируемым сообщений, тогда X^n — множество всех последовательностей сообщений длины n на выходе источника, а Y^n — множество всех возможных аппроксимирующих последовательностей. Критерий качества $d(x, y)$, $x \in X$, $y \in Y$, определяет величину ошибки $d_n(x, y)$ (см. (5.1.1)) аппроксимации последовательности $x \in X^n$ последовательностью $y \in Y^n$.

Определение 5.1.1. Кодом для кодирования с заданным критерием качества последовательностей сообщений длины n называется произвольное множество $T_n = \{u_1, \dots, u_M\} \subseteq Y^n$ аппроксимирующих последовательностей. Кодированием для кода T_n называется произвольное отображение $u(x)$ множества X^n на множество кодовых слов T_n . Число

$$\bar{d}_n \triangleq M d_n(x, u(x)) \quad (5.1.11)$$

называется средней ошибкой кодирования относительно критерия качества $d(x, y)$. Число

$$R \triangleq \frac{\log M}{n} \quad (5.1.12)$$

называется скоростью кода T_n .

Очевидно, что средняя ошибка минимизируется при таком кодировании, при котором для каждой последовательности $x \in X^n$ выбирается кодовое слово $u(x) \in T_n$, минимизирующее величину ошибки $d_n(x, u_i)$, другими словами,

$$d_n(x, u(x)) \leq d_n(x, u_i), \quad i = 1, \dots, M. \quad (5.1.13)$$

Кодирование, определяемое соотношением (5.1.13), будем называть оптимальным для данного кода T_n и данного критерия качества $d(x, y)$.

Из приведенного выше определения следует, что каждый код характеризуется двумя величинами: средней ошибкой \bar{d}_n и скоростью кодирования R . Скорость кодирования представляет собой количество двоичных символов на сообщение источника, при котором возможно аппроксимировать сообщения источника со средней ошибкой \bar{d}_n . [36]

На рис. 5.1.1 показана структура кодера источника, на выходе которого появляется в среднем R двоичных символов на сообщение. На выходе источника U_X последовательно появляются сообщения из множества X . Эти сообщения разбиваются на блоки длины n , и каждый блок подвергается кодированию независимо от остальных блоков. Устройство, обозначенное на рисунке буквой A (аппроксиматор), каждому блоку $x \in X^n$, поданному на

его вход, сопоставляет аппроксимирующую последовательность из множества T_n в соответствии с правилом кодирования (5.1.13). Все кодовые слова множества T_n занумерованы; каждому слову сопоставляется последовательность двоичных символов, обозначающая его номер. Устройство, обозначенное на рисунке буквой B , определяет номер слова $u(x)$ в множестве T_n и представляет этот номер в двоичной форме $(\alpha^{(1)}, \alpha^{(2)}, \dots, \alpha^{(nR)})$.

Заметим, что количество двоичных символов на выходе кодера в точности равно R только в том случае, когда nR есть целое число (или M есть некоторая целая степень двойки). Если это не так, то количество двоичных символов на сообщение равно $R' = \lceil nR \rceil / n$, где $|x|$ — наименьшее целое, большее или равное x . При больших n величины R и R' близки (отличаются друг от друга не более чем на $1/n$).

Всюду ниже мы будем обозначать символом (R, d) код источника со скоростью R и средней ошибкой d .

Определение 5.1.2. Пусть X и Y — фиксированные множества и $d(x, y)$ — функция, определенная на множестве XY . Пусть U_X — источник, выбирающий сообщения из множества X . ε -скоростью создания информации источником U_X относительно критерия качества $d(x, y)$ называется наименьшее число H_ε такое, что для любого $R > H_\varepsilon$ найдется (R, d) код относительно того же критерия качества при $d \leq \varepsilon$.

В последующей части этой главы мы будем заниматься отысканием ε -скорости создания информации некоторыми источниками. Для того чтобы установить, что некоторое число H_ε является ε -скоростью создания информации относительно критерия качества $d(x, y)$, следует доказать два утверждения:

1. Для любого $R > H_\varepsilon$ найдется n и код со скоростью R , кодирующий последовательности сообщений длины n , для которого средняя ошибка относительно критерия качества $d(x, y)$ не превосходит ε (прямая теорема кодирования).

2. Для любого $R < H_\varepsilon$ для всех n и для всех кодов со скоростью R средняя ошибка относительно критерия качества $d(x, y)$ превышает ε (обратная теорема кодирования).

Далее мы будем рассматривать только непрерывные ансамбли, задаваемые функциями плотности вероятностей (ф. п. в.), и непрерывные источники сообщений, а также изучать кодирование с заданным критерием качества только для непрерывных источников с дискретным временем.

§ 5.2. Эпсилон-энтропия и ее свойства

Ключевая роль в задаче определения ε -скорости создания информации принадлежит специальной функции, эпсилон-энтропии, которую мы формально введем и изучим ее свойства. Полезность этой функции проявится позже. Дело обстоит в точности так же, как в задаче определения скорости создания информации дискретным источником (гл. 1). Мы вначале ввели и изучили энтропию, а затем показали, что скорость создания информации равна энтропии.

Пусть U_X — непрерывный стационарный источник, выбиравший сообщения из множества X . Предположим, что Y — аппроксимирующее множество и $d(x, y)$ — критерий качества. Для каждого целого n , $n = 1, 2, \dots$, определен ансамбль $\{X^n, f(\mathbf{x})\}$ последовательностей сообщений длины n , распределение вероятностей на котором задается посредством ф. п. в. $f(\mathbf{x})$, $\mathbf{x} = (x^{(1)}, \dots, x^{(n)}) \in X^n$. Удобно полагать, что $X^n = X_1 \dots \dots X_n$, где X_i — ансамбль сообщений в момент времени i .

Пусть Y^n — множество аппроксимирующих последовательностей $\mathbf{y} = (y^{(1)}, \dots, y^{(n)})$, $y^{(i)} \in Y$, и $f(\mathbf{y}|\mathbf{x})$, $\mathbf{x} \in X^n$, — произвольное семейство условных ф. п. в. на Y^n . Если Y^n — дискретное множество или дискретное подмножество непрерывного множества, то ф. п. в. $f(\mathbf{y}|\mathbf{x})$ будем рассматривать как обобщенные ф. п. в. (см. § 2.2). Эти ф. п. в. совместно с ф. п. в. $f(\mathbf{x})$ задают ансамбль $\{X^n Y^n, f(\mathbf{x}, \mathbf{y})\}$, где $f(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) f(\mathbf{y}|\mathbf{x})$, и ансамбль $\{Y^n, f(\mathbf{y})\}$, где

$$f(\mathbf{y}) = \int_{X^n} f(\mathbf{x}) f(\mathbf{y}|\mathbf{x}) d\mathbf{x}. \quad (5.2.1)$$

Удобно полагать, что $Y^n = Y_1 \dots Y_n$, где Y_i — ансамбль аппроксимирующих сообщений в момент времени i . Заметим, что в общем случае ф. п. в. $f_i(x^{(i)}, y^{(i)})$, задающая распределение вероятностей на ансамбле $X_i Y_i$, зависит от номера i .

Для ансамбля $\{X^n Y^n, f(\mathbf{x}, \mathbf{y})\}$ определены две величины. Одна из них — средняя взаимная информация $I(X^n; Y^n)$ между ансамблями X^n и Y^n :

$$I(X^n; Y^n) = \iint_{X^n Y^n} f(\mathbf{x}) f(\mathbf{y}|\mathbf{x}) \log \frac{f(\mathbf{y}|\mathbf{x})}{f(\mathbf{y})} d\mathbf{x} d\mathbf{y}. \quad (5.2.2)$$

Другая — средняя ошибка аппроксимации:

$$\overline{d}_n = \frac{1}{n} \sum_{i=1}^n \overline{d}^{(i)}, \quad (5.2.3)$$

где

$$\overline{d}^{(i)} = \int_{X_i Y_i} d(x^{(i)}, y^{(i)}) f(x^{(i)}) f_i(y^{(i)}|x^{(i)}) dx^{(i)} dy^{(i)}. \quad (5.2.4)$$

Определение 5.2.1. Пусть $\Phi_n(\varepsilon)$ есть класс всех ф. п. в. $f(\mathbf{y}|\mathbf{x})$ такой, что для каждой функции из этого класса средняя ошибка не превосходит ε :

$$\Phi_n(\varepsilon) \triangleq \{f(\mathbf{y}|\mathbf{x}) : \overline{d}_n \leq \varepsilon\}. \quad (5.2.5)$$

Пусть

$$H_n(\varepsilon) \triangleq \min_{\Phi_n(\varepsilon)} \frac{1}{n} I(X^n; Y^n), \quad (5.2.6)$$

где минимум разыскивается по всем функциям $f(\mathbf{y}|\mathbf{x})$ из $\Phi_n(\varepsilon)$. Тогда функция от ε

$$H(\varepsilon) \triangleq \inf_n H_n(\varepsilon), \quad (5.2.7)$$

где точная нижняя грань *) берется по всем n , $n = 1, 2, \dots$, называется эпсилон-энтропией непрерывного стационарного источника U_X относительно критерия качества $d(x, y)$.

Рассмотрим свойства функции $H(\varepsilon)$.

Первое очевидное свойство этой функции состоит в том, что она не отрицательна и определена только для неотрицательных значений ε . Неотрицательность $H(\varepsilon)$ следует из неотрицательности средней взаимной информации.

Второе свойство, также достаточно очевидное, состоит в том, что $H(\varepsilon)$ — невозрастающая функция. Действительно, если $\varepsilon_1 > \varepsilon_2$, то при всех n множество $\Phi_n(\varepsilon_1)$ содержит множество $\Phi_n(\varepsilon_2)$ и, следовательно, минимум по некоторому множеству не может быть больше, чем минимум по части этого множества. Таким образом, $H(\varepsilon_1) \leq H(\varepsilon_2)$.

Следующее свойство сформулируем в виде теоремы.

Теорема 5.2.1. Пусть U_X — источник без памяти, т. е. такой, что для любых n и любых $\mathbf{x} \in X^n$, $\mathbf{x} = (x^{(1)}, \dots, x^{(n)})$,

$$f(\mathbf{x}) = \prod_{i=1}^n f(x^{(i)}), \quad (5.2.8)$$

*) Точная нижняя грань $\inf f(x)$, $x \in X$, где $f(x)$ — некоторая функция на X , есть наибольшее число f_0 такое, что $f_0 \leq f(x)$ для каждого $x \in X$. Если в множестве X существует такой элемент x_0 , для которого $f_0 = f(x_0)$, то говорят, что нижняя грань достигается на X , и пишут $f(x_0) = \inf f(x)$, $x \in X$. Если X — конечное множество, то нижняя грань всегда достигается. В этом случае всегда $\inf f(x) = \min f(x)$. Если X — бесконечное множество, то нижняя грань может не достигаться ни на одном элементе из X . Например, если X — множество натуральных чисел и $f(x) = 1 + 1/x$, то $\inf f(x) = 1$, но $f(x) \neq 1$ ни для одного элемента из X . Заметим, что нижняя грань всегда достигается, если X — замкнутое множество и $f(x)$ — непрерывная функция.

где все сомножители в правой части (5.2.8) образованы с помощью одной и той же функции $f(x)$ — безусловной ф. п. в. на ансамбле X . Тогда

$$H(\varepsilon) = \min_{\Phi(\varepsilon)} I(X; Y), \quad (5.2.9)$$

где $\Phi(\varepsilon)$ — множество всех одномерных условных ф. п. в. таких, что для любой функции $f(y|x) \in \Phi(\varepsilon)$

$$\overline{d} \triangleq \iint_{X \times Y} d(x, y) f(x) f(y|x) dx dy \leq \varepsilon. \quad (5.2.10)$$

Доказательство. Имеем

$$\begin{aligned} I(X^n; Y^n) &= H_0(X^n) - H_0(X^n | Y^n) = \\ &= \sum_{i=1}^n H_0(X_i) - \sum_{i=1}^n H_0(X_i | Y^n X_1 \dots X_{i-1}), \end{aligned} \quad (5.2.11)$$

где второе равенство является следствием независимости сообщений источника $(H_0(X^n) = \sum_{i=1}^n H_0(X_i))$ и свойства аддитивности относительной энтропии. Далее, так как $Y^n = Y_1 \dots Y_n$ и относительная энтропия не уменьшается при исключении части условий, то для любого i , $1 \leq i \leq n$,

$$H_0(X_i | Y^n X_1 \dots X_{i-1}) \leq H_0(X_i | Y_i). \quad (5.2.12)$$

Поэтому

$$\frac{1}{n} I(X^n; Y^n) \geq \frac{1}{n} \sum_{i=1}^n [H_0(X_i) - H_0(X_i | Y_i)] = \frac{1}{n} \sum_{i=1}^n I_i(X; Y), \quad (5.2.13)$$

где использовано обозначение

$$I_i(X; Y) \triangleq I(X_i; Y_i)$$

и учтено то обстоятельство, что при всех i множества $X_i Y_i$ совпадают между собой и с множеством XY . Индекс i в обозначении информации $I_i(X; Y)$ означает, что она вычислена для ф. п. в. $f_i(x, y) = f(x) f_i(y|x)$, которая для произвольной ф. п. в. $f(y|x)$ может зависеть от индекса i . Если теперь положить $\lambda_1 = \dots = \lambda_n = \frac{1}{n}$, то выражение в правой части (5.2.13) в силу выпуклости вниз средней взаимной информации по условным распределениям можно оценить следующим образом:

$$\frac{1}{n} \sum_{i=1}^n I_i(X; Y) = \sum_{i=1}^n \lambda_i I_i(X; Y) \geq I_0(X; Y), \quad (5.2.14)$$

где $I_0(X; Y)$ — средняя взаимная информация, вычисленная для ф. п. в. $f_0(x, y) \triangleq f(x) f_0(y|x)$, где

$$f_0(y|x) \triangleq \frac{1}{n} \sum_{i=1}^n f_i(y|x). \quad (5.2.15)$$

Из (5.2.3) следует, что $f_0(y|x) \in \Phi(\varepsilon)$ для любой ф. п. в. $f(y|x) \in \Phi_n(\varepsilon)$.

Заметим, что равенства в (5.2.13) и (5.2.14) достигаются, если выбрать

$$f(y|x) = f_0(y|x) \triangleq \prod_{i=1}^n f_0(y^{(i)}|x^{(i)}), \quad (5.2.16)$$

т. е. если пары $(X_1 Y_1), \dots, (X_n Y_n)$ статистически независимы и одинаково распределены. Первое равенство является тогда следствием аддитивности информации, а второе — следствием того, что $I_i(X; Y) = I_0(X; Y)$ при всех $i = 1, \dots, n$.

Покажем теперь, что функция $f_0(y|x)$ принадлежит множеству $\Phi_n(\varepsilon)$, если $f_0(y|x)$ принадлежит множеству $\Phi(\varepsilon)$. Для того чтобы в этом убедиться, достаточно увидеть, что при указанном выше выборе $f_0(y|x)$ средние ошибки $\overline{d}^{(i)}$ будут одинаковы для всех $i = 1, \dots, n$ и равны \overline{d}^0 (см. (5.2.10)). Но это так, поскольку пары (X_i, Y_i) , $i = 1, \dots, n$, одинаково распределены.

Таким образом, мы показали, что

$$\min_{f(y|x) \in \Phi_n(\varepsilon)} \frac{1}{n} I(X^n; Y^n) = \min_{f(y|x) \in \Phi(\varepsilon)} I(X; Y). \quad (5.2.17)$$

Теорема доказана.

Теорема 5.2.2. Эпсилон-энтропия $H(\varepsilon)$ — выпуклая вниз функция ε .

Доказательство. Мы докажем эту теорему для случая источников без памяти и коротко обсудим общий случай. Для непрерывного источника без памяти

$$H(\varepsilon) = \min_{\Phi(\varepsilon)} I(X; Y). \quad (5.2.18)$$

Предположим, что минимум в (5.2.18) для значений ε_1 и ε_2 достигается на условных ф. п. в. $f_1(y|x)$ и $f_2(y|x)$ соответственно. Пусть λ — неотрицательное число, лежащее между нулем и единицей. Функция

$$f_0(y|x) \triangleq \lambda f_1(y|x) + (1 - \lambda) f_2(y|x) \quad (5.2.19)$$

есть ф. п. в., принадлежащая множеству $\Phi(\lambda \varepsilon_1 + (1 - \lambda) \varepsilon_2)$. Действительно, так как для функций $f_1(y|x)$ и $f_2(y|x)$ средние ошибки не превосходят величин ε_1 и ε_2 соответственно, то

$$\begin{aligned} \lambda \varepsilon_1 + (1 - \lambda) \varepsilon_2 &\geq \int_X \int_Y f(x) [f_1(y|x) \lambda + f_2(y|x) (1 - \lambda)] d(x, y) dx dy = \\ &= \int_X \int_Y f(x) f_0(y|x) d(x, y) dx dy. \end{aligned} \quad (5.2.20)$$

Поэтому имеет место неравенство

$$H(\lambda \varepsilon_1 + (1 - \lambda) \varepsilon_2) \leq I_0(X; Y), \quad (5.2.21)$$

где $I_0(X; Y)$ — средняя взаимная информация между ансамблями X и Y , вычисленная для функции $f_0(y|x)$. Теперь можно воспользоваться выпуклостью вниз средней взаимной информации по условным распределениям. Из выпуклости следует, что

$$I_0(X; Y) \leq \lambda I_1(X; Y) + (1 - \lambda) I_2(X; Y) = \lambda H(\varepsilon_1) + (1 - \lambda) H(\varepsilon_2), \quad (5.2.22)$$

где $I_i(X; Y)$ — средняя взаимная информация, вычисленная для функции $f_i(y|x)$.

Отсюда и из (5.2.21) вытекает выпуклость вниз функции $H(\varepsilon)$.

В общем случае в точности по той же схеме доказывается, что при каждом n функции $H_n(\varepsilon)$ выпуклы вниз по ε . Выпуклость $H(\varepsilon)$ следует из теоремы 5.2.3 и из того, что предел последовательности выпуклых функций также является выпуклой в ту же сторону функцией. Теорема доказана.

Теорема 5.2.3. Пусть $H_n(\varepsilon)$ ограничено при всех $n = 1, 2, \dots$. Тогда последовательность $\{H_n(\varepsilon)\}$ имеет предел и

$$\lim_{n \rightarrow \infty} H_n(\varepsilon) = H(\varepsilon). \quad (5.2.23)$$

Эта теорема приводится без доказательства. Основные моменты доказательства обсуждаются в задаче 5.2.4.

Теперь будет показано, что в некоторой области значений ε эпсилон-энтропия $H(\varepsilon)$ равна нулю. Пусть при некотором n

$$\varepsilon_0 \triangleq \min_{y \in Y^n} M_{y, d_n}(x; y)^* \quad (5.2.24)$$

и пусть $y_0 \in Y^n$ — элемент, на котором достигается этот минимум. Тогда $H_n(\varepsilon) = 0$ для всех $\varepsilon \leq \varepsilon_0$ и, следовательно, $H(\varepsilon) = 0$ при тех же значениях ε . Действительно, при выборе в качестве универсального аппроксимирующего элемента y_0 для всех $x \in X^n$ средняя ошибка будет равна

$$M_{d_n}(x, y) = M_{y, d_n}(x, y_0) = \varepsilon_0. \quad (5.2.25)$$

* Напомним, что M_y есть условное математическое ожидание по ансамблю X^n при условии, что y фиксировано.

Аппроксимации с помощью такого универсального элемента соответствует условная ф. п. в.

$$f(y|x) = \delta(y - y_0) \text{ для всех } x \in X^n. \quad (5.2.26)$$

Это означает, что ансамбли X^n и Y^n статистически независимы и, следовательно, $I(X^n; Y^n) = 0$ для ф. п. в. (5.2.26). Так как $H(\varepsilon)$ не превышает $1/n I(X^n; Y^n) = 0$ и не отрицательно, то $H(\varepsilon) = 0$ для всех $\varepsilon \geq \varepsilon_0$.

Пример 5.2.1. Рассмотрим квадратический критерий качества, введенный в примере 5.1.3. Пусть источник без памяти порождает случайные величины с нулевым математическим ожиданием $\int_{-\infty}^{\infty} xf(x) dx = 0$.

Число 0 является универсальным аппроксимирующим элементом при всех $\varepsilon \geq D$, где D — дисперсия величин на выходе источника. Действительно,

$$Md(x, y_0 = 0) = MX^2 = D. \quad (5.2.27)$$

Эпсилон-энтропия такого источника равна 0 для всех $\varepsilon \geq D$.

Типичный график эпсилон-энтропии приведен на рис. 5.2.1.

§ 5.3. Обратная теорема кодирования непрерывных источников при заданном критерии качества

В этом параграфе мы будем заниматься обратной теоремой кодирования, которая покажет, что $H(\varepsilon)$ является нижней границей скорости кода для непрерывного стационарного источника, средняя ошибка которого не превосходит ε . Вначале мы будем рассматривать источники без памяти, а затем обсудим общий случай.

Пусть U_x — непрерывный источник без памяти, т. е. такой, что для любых n и любых $x = (x^{(1)}, \dots, x^{(n)}) \in X^n$ функция плотности вероятностей $f(x) = f(x^{(1)}) \dots f(x^{(n)})$. Сообщения на выходе такого источника независимы и одинаково распределены.

Теорема 5.3.1. Для любого непрерывного источника без памяти, любого критерия качества $d(x, y)$ и любого (R, d) -кода со скоростью $R < H(\varepsilon)$, где $H(\varepsilon)$ — эпсилон-энтропия источника относительно критерия качества $d(x, y)$, имеет место неравенство $d > \varepsilon$.

Доказательство. Заметим, что для непрерывных источников без памяти и любого натурального n

$$\min_{\Phi_n(\varepsilon)} \frac{1}{n} I(X^n; Y^n) = \min_{\Phi(\varepsilon)} I(X; Y) = H(\varepsilon), \quad (5.3.1)$$

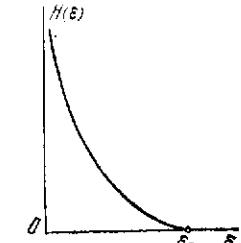


Рис. 5.2.1. Типичный график функции $H(\varepsilon)$.

где минимум в левой части равенства разыскивается по всем ф. п. в. $f(y|x) \in \Phi_n(\epsilon)$, для которых средняя ошибка \bar{d}_n не превосходит ϵ . Поэтому для любой такой ф. п. в. величина средней взаимной информации на сообщение

$$\frac{1}{n} I(X^n; Y^n) \geq H(\epsilon). \quad (5.3.2)$$

Пусть T_n — код со скоростью R , который кодирует источник U_X со средней ошибкой $d \ll \epsilon$. Условная ф. п. в., соответствующая оптимальному кодированию сообщений множества X^n с помощью кода $T_n = \{\mathbf{u}_1 \dots \mathbf{u}_M\}$, $M = 2^{nR}$, принадлежит множеству $\Phi_n(\epsilon)$ и имеет вид

$$f(y|x) = \delta(y - u(x)). \quad (5.3.3)$$

Неравенство (5.3.2) остается справедливым и для ф. п. в. (5.3.3), следовательно,

$$\frac{1}{n} I(X^n; T_n) \geq H(\epsilon). \quad (5.3.4)$$

Предположим, что при $R < H(\epsilon)$ нашелся (R, d) -код со средней ошибкой $d \ll \epsilon$. Если это так, то

$$\begin{aligned} H(\epsilon) &\leq \frac{1}{n} I(X^n; T_n) = \frac{1}{n} (H(T_n) - H(T_n|X^n)) \leq \\ &\leq \frac{1}{n} H(T_n) \leq \frac{1}{n} \log 2^{nR} = R, \end{aligned} \quad (5.3.5)$$

что противоречит предположению о скорости кода. Теорема доказана.

Рассмотрим теперь общий случай, когда источник U_X является стационарным и не обязательно не имеет памяти. Ключевым местом доказательства обратной теоремы для источника без памяти является неравенство (5.3.2), которое в рассмотренном выше случае выполняется для всех натуральных n и для всех ф. п. в. из $\Phi_n(\epsilon)$. Для того чтобы получить обратную теорему в общем случае, необходимо получить такое же неравенство для произвольного стационарного источника.

Воспользуемся теоремой 5.2.3. Из определения эпсилон-энтропии и этой теоремы вытекает, что для всех n имеет место неравенство $H_n(\epsilon) \geq H(\epsilon)$. Отсюда следует, что для произвольной ф. п. в. $f(y|x) \in \Phi_n(\epsilon)$

$$\frac{1}{n} I(X^n; Y^n) \geq H_n(\epsilon) \geq H(\epsilon). \quad (5.3.6)$$

В остальном доказательство обратной теоремы кодирования для произвольного стационарного источника сохраняется таким же, как и доказательство теоремы 5.3.1.

Таким образом, не существует кодов, кодирующих непрерывный стационарный источник с заданным критерием качества, для которых одновременно скорость $R < H(\epsilon)$ и средняя ошибка меньше или равна ϵ .

§ 5.4. ЭПСИЛОН-ЭНТРОПИЯ ГАУССОВСКОГО ИСТОЧНИКА БЕЗ ПАМЯТИ

Прежде чем переходить к рассмотрению прямых теорем кодирования, мы найдем эпсилон-энтропию одного из наиболее простых источников — непрерывного источника, на выходе которого появляются независимые гауссовские случайные величины. Как было показано в § 5.2, эпсилон-энтропия такого источника определяется только одномерными распределениями вероятностей. Поэтому можно говорить об эпсилон-энтропии гауссовой случайной величины.

Пусть $\{X, f_1(x)\}$ — непрерывный ансамбль, где $f_1(x)$ — гауссовская ф. п. в.

$$f_1(x) = \frac{1}{\sqrt{2\pi D}} \exp\left\{-\frac{x^2}{2D}\right\}. \quad (5.4.1)$$

Таким образом, рассматриваемая с. в. имеет нулевое среднее и дисперсию D .

Рассмотрим квадратический критерий качества

$$d(x, y) \triangleq (x - y)^2, \quad (5.4.2)$$

определенный для всех x, y , принимающих значения на числовой оси. Эпсилон-энтропия рассматриваемого источника (эпсилон-энтропия с. в. X), согласно теореме 5.2.1, определяется формулой

$$H(\epsilon) = \min_{\Phi(\epsilon)} I(X; Y), \quad (5.4.3)$$

где Y — аппроксимирующая с. в., задаваемая ф. п. в. $f(y|x)$, и множество $\Phi(\epsilon)$, по которому разыскивается минимум, состоит из всех ф. п. в. $f(y|x)$ таких, что среднеквадратическая ошибка $\bar{d} = M(X - Y)^2$ не превосходит ϵ .

Таким образом, для всякой ф. п. в. $f(y|x)$ из $\Phi(\epsilon)$

$$\bar{d} = M(X - Y)^2 = \iint_{X Y} (x - y)^2 f_1(x) f(y|x) dx dy \leq \epsilon. \quad (5.4.4)$$

Среднюю взаимную информацию $I(X; Y)$ можно представить как разность относительных энтропий, поэтому

$$H(\epsilon) = \min_{\Phi(\epsilon)} (H_0(X) - H_0(X|Y)) = H_0(X) - \max_{\Phi(\epsilon)} H_0(X|Y), \quad (5.4.5)$$

где второе равенство есть следствие того, что относительная энтропия $H_0(X)$ ансамбля $\{X, f_1(x)\}$ определяется только функцией $f_1(x)$:

$$H_0(X) = \frac{1}{2} \log 2\pi e D \quad (5.4.6)$$

(см. § 2.3) и не зависит от выбора функции $f(y|x)$.

Таким образом, задача определения эпсилон-энтропии свелась к нахождению максимума в формуле (5.4.5) и доказательству того, что этот максимум достигается на некоторой функции из $\Phi(\varepsilon)$. Для отыскания максимума введем в рассмотрение с. в. Z , определив ее с помощью равенства

$$X = Y + Z. \quad (5.4.7)$$

Равенство (5.4.7) означает, что значения x и y с. в. X и Y однозначно определяют значение $z = x - y$ с. в. Z . Верно также то, что при заданном значении y с. в. Y случайные величины X и Z однозначно определяют друг друга. Отсюда вытекает, что при фиксированном y две с. в. X и Z отличаются только математическим ожиданием и, следовательно, имеют одинаковые относительные энтропии:

$$H_0(Z|y) = H_0(X|y). \quad (5.4.8)$$

Усредняя обе части (5.4.8) по $y \in Y$, получим, что

$$H_0(Z|Y) = H_0(X|Y). \quad (5.4.9)$$

Если теперь воспользоваться этим равенством и тем, что

$$H_0(Z|Y) \leq H_0(Z), \quad (5.4.10)$$

то из (5.4.5) последует неравенство

$$H(\varepsilon) \geq H_0(X) - \max_{\Phi(\varepsilon)} H_0(Z). \quad (5.4.11)$$

Максимум $H_0(Z)$ в правой части (5.4.11) нетрудно найти, если заметить, что для любой функции $f(y|x)$ из $\Phi(\varepsilon)$

$$M Z^2 = M(X - Y)^2 \leq \varepsilon. \quad (5.4.12)$$

Так как для любой с. в. с ограниченным средним квадратом относительная энтропия не превосходит $\frac{1}{2} \log 2\pi e c^2$, где c^2 — ограничивающее значение для среднего квадрата (см. § 2.3), то

$$H_0(Z) \leq \frac{1}{2} \log 2\pi e \varepsilon, \quad (5.4.13)$$

причем равенство достигается в том случае, когда Z есть гауссовская с. в. с нулевым средним и дисперсией, равной ε . Используя (5.4.6), (5.4.11) и (5.4.13), получим

$$H(\varepsilon) \geq \frac{1}{2} \log 2\pi e D - \frac{1}{2} \log 2\pi e \varepsilon = \frac{1}{2} \log \frac{D}{\varepsilon}. \quad (5.4.14)$$

Вопрос, который теперь остался открытым, состоит в том, можно ли подобрать функцию $f(y|x) \in \Phi(\varepsilon)$ такую, чтобы неравенство (5.4.14) выполнялось со знаком равенства. Тогда правая часть этого выражения и была бы эпсилон-энтропией.

Чтобы в (5.4.14) имело место равенство, необходимо достичь равенства в неравенствах (5.4.10) и (5.4.13). Первое возможно только, когда с. в. Y и Z статистически независимы, второе возможно, когда Z — гауссовская с. в. с нулевым средним и дисперсией ε . Покажем, что при $D \geq \varepsilon$ оба эти требования можно удовлетворить соответствующим выбором ф. п. в. $f(y|x) \in \Phi(\varepsilon)$.

Вначале заметим, что сумма двух гауссовых с. в. есть снова гауссовская с. в. Поэтому любую гауссовскую с. в. можно представить в виде суммы двух независимых гауссовых с. в.: $X = Y + Z$. Если с. в. X имеет дисперсию D , то с. в. Y и Z можно выбрать так, чтобы они были гауссовскими, независимыми и имели дисперсии $D - \varepsilon$ и ε соответственно, причем $MZ = 0$. Очевидно, что при таком выборе с. в. Y и Z имеет место равенство $M(X - Y)^2 = \varepsilon$, т. е. ф. п. в. $f(y|x)$, задающая вместе с ф. п. в. $f(x)$ распределение вероятностей на парах (x, y) , принадлежит множеству $\Phi(\varepsilon)$. Таким образом, в неравенстве (5.4.14) достигается равенство для некоторой функции $f(y|x) \in \Phi(\varepsilon)$. Этую функцию можно выписать в явном виде, используя независимость с. в. Y и Z . Обозначим через $f_1(x)$, $f_2(y)$ и $f_3(z)$ ф. п. в. для с. в. X , Y и Z соответственно. В силу независимости с. в. Y и Z имеет место равенство $f(x|y) = f_3(x - y)$. Тогда ф. п. в., на которой достигается равенство в (5.4.14), имеет следующий вид:

$$f(y|x) = \frac{f(x|y) f_2(y)}{f_1(x)} = \frac{f_3(x-y) f_2(y)}{f_1(x)}. \quad (5.4.15)$$

Если $\varepsilon > D$, то приведенные рассуждения не справедливы и эпсилон-энтропия должна вычисляться иначе. Однако в этом случае задача оказывается еще более простой, поскольку при таких значениях среднеквадратической ошибки существует универсальный аппроксимирующий элемент, а именно $y_0 = 0$. Очевидно, что при использовании такого элемента

$$M(X - y_0)^2 = MX^2 = D, \quad (5.4.16)$$

что меньше, чем ε , для указанного выше диапазона ошибок. Поэтому

$$H(\varepsilon) = 0 \quad (5.4.17)$$

при всех $\varepsilon > D$.

Соотношения (5.4.14) и (5.4.17) можно объединить с помощью следующей формулы:

$$H(\varepsilon) = \frac{1}{2} \log \max \left\{ 1, \frac{D}{\varepsilon} \right\}, \quad (5.4.18)$$

которая и представляет собой выражение для эпсилон-энтропии гауссовской случайной величины или для эпсилон-энтропии непрерывного гауссовского источника без памяти.

§ 5.5. Прямая теорема кодирования стационарного источника независимых гауссовых сообщений при квадратическом критерии качества

Целью этого параграфа является формулировка и доказательство прямой теоремы кодирования для стационарного источника независимых гауссовых случайных величин (гауссовского источника без памяти). Для доказательства будет использован закон больших чисел для независимых непрерывных с. в. и вытекающий из него результат, называемый принципом «затвердевания сферы». Кроме этого, будет использован метод случайного кодирования и будет показано, что существует достаточно много кодов, скорость которых близка к эпсилон-энтропии при заданном значении среднеквадратической ошибки.

5.5.1. Закон больших чисел и принцип «затвердевания сферы». Пусть X_1, X_2, \dots, X_n последовательность некоррелированных одинаково распределенных с. в. с конечными математическими ожиданиями и конечными дисперсиями. Тогда для любых положительных γ и δ найдется такое N , что при всех $n > N$

$$\Pr\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - m\right| > \delta\right) \leq \gamma, \quad (5.5.1)$$

где $m \triangleq M X_i$, $i = 1, \dots, n$, — математическое ожидание каждой из с. в. X_1, \dots, X_n .

Это утверждение называется законом больших чисел. Оно ранее формулировалось и доказывалось для дискретных с. в. Доказательство этого утверждения для непрерывных с. в. производится в точности так же (см. задачу 2.2.9).

Предположим, что X_1, \dots, X_n — последовательность независимых одинаково распределенных с. в. с нулевыми математическими ожиданиями и дисперсиями, равными D . Тогда X_1^2, \dots, X_n^2 — также последовательность независимых одинаково распределенных с. в. с математическими ожиданиями, равными D , и дисперсиями, равными $M X_i^2 - D^2$. Предположим, что рассматриваемые с. в. имеют конечный четвертый момент: $M X_i^4 < \infty$ *), тогда для каждой из с. в. X_i^2 дисперсия также конечна и поэтому применим закон больших чисел. Из этого закона следует, что

*) См. задачу 2.5.8(а) для случая гауссовых с. в.

для любых положительных γ и δ при достаточно больших n выполняется следующее неравенство:

$$\Pr\left(\left|\frac{1}{n} \sum_{i=1}^n X_i^2 - D\right| > \delta\right) \leq \gamma. \quad (5.5.2)$$

Последовательность $\mathbf{X} = (X_1, \dots, X_n)$ можно рассматривать как случайный вектор в n -мерном евклидовом пространстве, тогда

$$l_n \triangleq \left(\sum_{i=1}^n X_i^2\right)^{1/2} \quad (5.5.3)$$

является длиной этого вектора. Очевидно, что l_n — случайная величина, принимающая неотрицательные значения. Мы хотим показать, что при любом положительном δ_1 и достаточно большом n событие, состоящее в том, что

$$\sqrt{D} - \delta_1 \leq \frac{1}{\sqrt{n}} l_n \leq \sqrt{D} + \delta_1, \quad (5.5.4)$$

появляется с вероятностью, близкой к единице. Другими словами, при больших n случайный вектор \mathbf{X} лежит вблизи поверхности n -мерной сферы радиуса \sqrt{Dn} с центром в начале координат.

Из (5.5.2) следует, что при достаточно больших n и произвольных положительных γ и δ событие, состоящее в том, что

$$D - \delta \leq \frac{1}{n} l_n^2 \leq D + \delta, \quad (5.5.5)$$

имеет вероятность большую или равную $1 - \gamma$. Из правого неравенства (5.5.5) следует, что $\frac{1}{n} l_n^2 \leq D + \delta + 2\sqrt{D\delta}$ и

$$\frac{1}{\sqrt{n}} l_n \leq \sqrt{D} + \sqrt{\delta}.$$

Из левого неравенства (5.5.5) следует, что $D \leq \frac{1}{n} l_n^2 + \delta + \frac{2}{n} l_n \sqrt{\delta}$ и

$$\frac{1}{\sqrt{n}} l_n \geq \sqrt{D} - \sqrt{\delta}.$$

Таким образом, событие, определяемое неравенствами (5.5.5), влечет следующее событие:

$$\sqrt{D} - \sqrt{\delta} \leq \frac{1}{\sqrt{n}} l_n \leq \sqrt{D} + \sqrt{\delta}. \quad (5.5.6)$$

Полагая $\sqrt{\delta} \triangleq \delta_1$, получим, что

$$\Pr\left(\left|\frac{1}{\sqrt{n}} l_n - \sqrt{D}\right| \leq \delta_1\right) \geq \Pr\left(\left|\frac{1}{n} l_n^2 - D\right| \leq \delta\right) \geq 1 - \gamma, \quad (5.5.7)$$

причем последнее неравенство выполняется при достаточно больших значениях n .

Отсюда следует, что при достаточно больших n случайный вектор $\mathbf{X} = (X_1, \dots, X_n)$ с высокой вероятностью попадает внутрь достаточно тонкой сферической области вблизи поверхности n -мерной сферы радиуса \sqrt{Dn} с центром в начале координат. Такая область называется «твердой» сферой (см. рис. 5.5.1).

Явление «затвердевания» n -мерной сферы является проявлением закона больших чисел.

Если ф. п. в. $f(\mathbf{x})$ случайного вектора \mathbf{X} обладает сферической симметрией, т. е. если $f(\mathbf{x}_1) = f(\mathbf{x}_2)$ для любых двух векторов \mathbf{x}_1 и \mathbf{x}_2 , имеющих одинаковые длины, то возможные значения вектора \mathbf{X} заполняют «твердую» сферу равномерно.

Другими словами, в этом случае вектор \mathbf{X} имеет равномерное распределение вероятностей на «твердой» сфере.

Отметим, что гауссовское n -мерное распределение вероятностей случайного вектора X , образованного независимыми одинаково распределенными гауссовскими с. в., обладает указанной сферической симметрией.

5.5.2. Аппроксимация векторов, лежащих на поверхности n -мерной сферы. Здесь мы хотим показать, что имеется конечное число векторов на поверхности n -мерной сферы радиуса \sqrt{Dn} , с помощью которых можно достаточно хорошо аппроксимировать любой вектор на этой поверхности. Достаточно хорошо аппроксимировать — значит обеспечить достаточно малое расстояние между каждым вектором, лежащим на сфере, и хотя бы одним из аппроксимирующих векторов.

Обозначим через $S_n(\sqrt{D})$ n -мерную сферу радиуса \sqrt{D} с центром в начале координат:

$$S_n(\sqrt{D}) \triangleq \left\{ \mathbf{x} = (x_1, \dots, x_n) : \sum_{i=1}^n x_i^2 = D \right\}. \quad (5.5.8)$$

Лемма 5.5.1. В множестве $S_n(\sqrt{Dn})$ n -мерных векторов можно выбрать подмножество $A_n = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$ такое, что для каждого вектора $\mathbf{x} = (x_1, \dots, x_n)$ из множества $S_n(\sqrt{D})$ найдется такой вектор $\mathbf{y} = (y_1, \dots, y_n)$ из подмножества A_n , что

$$\sum_{i=1}^n (x_i - y_i)^2 \leq 4D. \quad (5.5.9)$$

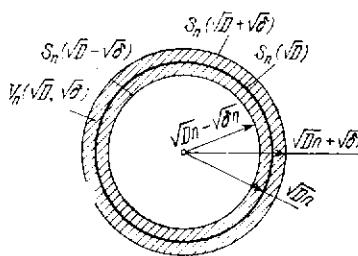


Рис. 5.5.1. К принципу «затвердевания сферы».

При этом число N элементов в множестве A_n удовлетворяет неравенству

$$N \leq (2n)^n. \quad (5.5.10)$$

Доказательство. Заметим вначале, что

$$|x_i| \leq \sqrt{Dn}, \quad i = 1, \dots, n, \quad (5.5.11)$$

для любого вектора $\mathbf{x} \in S_n(\sqrt{D})$.

Разобъем интервал $(-\sqrt{Dn}, \sqrt{Dn})$ на $2n$ непересекающихся частей: $I_1 \triangleq \left(-n\sqrt{\frac{D}{n}}, -(n-1)\sqrt{\frac{D}{n}} \right), \dots, I_{2n} \triangleq \left((n-1)\sqrt{\frac{D}{n}}, n\sqrt{\frac{D}{n}} \right)$. Каждому вектору $\mathbf{x} = (x_1, \dots, x_n) \in S_n(\sqrt{D})$ сопоставим вектор $\mathbf{z}(\mathbf{x}) = (z_1, \dots, z_n)$ по следующему правилу:

$$z_i = (k - n - 1)\sqrt{\frac{D}{n}}, \quad \text{если } x_i \in I_k. \quad (5.5.12)$$

В силу неравенства (5.5.11) такое сопоставление возможно. Каждая компонента вектора $\mathbf{z}(\mathbf{x})$ может принимать любое из $2n$ значений. Поэтому количество различных векторов, которые можно образовать из этих значений, равно

$$N_1 = (2n)^n. \quad (5.5.13)$$

Пусть $\{\mathbf{z}_1, \dots, \mathbf{z}_{N_1}\}$ — множество всех таких векторов.

Существует, вообще говоря, не один вектор, которому сопоставляется данный вектор $\mathbf{z} \in \{\mathbf{z}_1, \dots, \mathbf{z}_{N_1}\}$. Обозначим через B_j подмножество таких векторов, принадлежащих n -мерной сфере $S_n(\sqrt{D})$, которым сопоставляется один и тот же вектор \mathbf{z}_j , $j = 1, \dots, N_1$. Заметим, что некоторые из этих подмножеств будут пустыми. Например, вектору

$$\left(-n\sqrt{\frac{D}{n}}, \dots, -n\sqrt{\frac{D}{n}} \right) \quad (5.5.14)$$

соответствует пустое подмножество. Из каждого непустого подмножества B_j произвольным образом выберем один вектор. Обозначим его через \mathbf{y}_j . Пусть $A_n = \{\mathbf{y}_1, \dots, \mathbf{y}_{N_1}\}$ — полученное таким образом множество векторов. Покажем, что для этого множества утверждение леммы выполняется.

Действительно, пусть $\mathbf{x} = (x_1, \dots, x_n) \in B_j$ и $\mathbf{z}(\mathbf{x}) = (z_1, \dots, z_n)$. Пусть, кроме того, $\mathbf{y} = (y_1, \dots, y_n)$ — вектор, выбранный из подмножества B_j и включенный в множество A_n . Тогда

$$|x_i - z_i| \leq \sqrt{\frac{D}{n}}, \quad |z_i - y_i| \leq \sqrt{\frac{D}{n}}, \quad i = 1, \dots, n. \quad (5.5.15)$$

Используя неравенство треугольника, получим

$$\left(\sum_{i=1}^n (x_i - y_i)^2 \right)^{1/2} \leq \left(\sum_{i=1}^n (x_i - z_i)^2 \right)^{1/2} + \left(\sum_{i=1}^n (z_i - y_i)^2 \right)^{1/2}. \quad (5.5.16)$$

Из (5.5.15) следует, что

$$\sum_{i=1}^n (x_i - z_i)^2 \leq D, \quad \sum_{i=1}^n (z_i - y_i)^2 \leq D, \quad (5.5.17)$$

откуда вытекает неравенство (5.5.9). Очевидно, что $N \leq N_1 = (2n)^n$. Лемма доказана.

5.5.3. Аппроксимация последовательностей сообщений источника с помощью ϵ -сети на n -мерной сфере. Предположим, что возможно выбрать подмножество, состоящее из $M - 1$ n -мерных векторов, такое, что все векторы, лежащие на n -мерной сфере $S_n(\sqrt{D})$, можно аппроксимировать со среднеквадратической ошибкой ϵ . Другими словами, для каждого $\mathbf{x} = (x_1, \dots, x_n) \in S_n(\sqrt{D})$ в этом подмножестве найдется вектор $\mathbf{u}(\mathbf{x}) = (u_1, \dots, u_n)$ такой, что

$$\frac{1}{n} \sum_{i=1}^n (x_i - u_i)^2 \leq \epsilon.$$

Описанное подмножество называется *ϵ -сетью, аппроксимирующей сферу $S_n(\sqrt{D})$* . Ниже мы покажем, что ϵ -сеть, аппроксимирующая сферу $S_n(\sqrt{D})$, дополненная нулевым вектором, может служить хорошим аппроксимирующим множеством для сообщений непрерывного источника без памяти, порождающего случайные величины с нулевым средним и дисперсией D . В следующем подпараграфе при доказательстве прямой теоремы мы покажем, что ϵ -сеть, содержащая наименьшее число элементов, получается выбором точек на n -мерной сфере $S_n(\sqrt{D} - \epsilon + \delta_0)$, где δ_0 — положительное число, которое может быть выбрано сколь угодно малым.

Л е м м а 5.5.2. Пусть δ_0 — произвольное положительное число и $T_n = \{\mathbf{u}_1, \dots, \mathbf{u}_{M-1}\}$ — такое подмножество векторов, что для любого вектора $\mathbf{y} = (y_1, \dots, y_n) \in A_n$, где A_n — множество, определенное в лемме 5.5.1, найдется вектор $\mathbf{u} = (u_1, \dots, u_n) \in T_n$, для которого

$$d_n(\mathbf{y}, \mathbf{u}) \triangleq \frac{1}{n} \sum_{i=1}^n (y_i - u_i)^2 \leq \epsilon - \delta_0. \quad (5.5.18)$$

Пусть $\mathbf{u}_0 = (0, \dots, 0)$ — нулевой вектор и $T_n^* \triangleq \mathbf{u}_0 \cup T_n$. Тогда существует отображение множества X^n всех n -мерных векторов,

где X — числовая ось, на множество T_n^* , обладающее тем свойством, что при достаточно больших n

$$\bar{d}_n \triangleq M d_n(X, u(X)) \leq \epsilon, \quad (5.5.19)$$

где $u(x)$ — указанное отображение, $x \in X^n$, $u(x) \in T_n^*$, и $X = (X_1, \dots, X_n)$ — система независимых одинаково распределенных с. в., имеющих нулевое среднее и дисперсию D .

Д о к а з а т е л ь с т в о. Обозначим через $V_n(\sqrt{D}, \delta_1)$ множество векторов $\mathbf{x} \in X^n$, для которых выполняется неравенство (5.5.4). Другими словами, $V_n(\sqrt{D}, \delta_1)$ — сферическая область, заключенная между двумя n -мерными сферами $S_n(\sqrt{D} + \delta_1)$ и $S_n(\sqrt{D} - \delta_1)$. Пусть \mathbf{x} — вектор из $V_n(\sqrt{D}, \delta_1)$ и \mathbf{x}' — такой вектор из $S_n(\sqrt{D})$, что

$$d_n(\mathbf{x}, \mathbf{x}') \leq d_n(\mathbf{x}, \mathbf{x}'') \text{ для всех } \mathbf{x}'' \in S_n(\sqrt{D}). \quad (5.5.20)$$

Пусть \mathbf{y} — вектор из множества A_n , для которого

$$d_n(\mathbf{x}', \mathbf{y}) \leq d_n(\mathbf{x}'', \mathbf{y}) \text{ для всех } \mathbf{y}' \in A_n. \quad (5.5.21)$$

Наконец, пусть $\mathbf{u}(x)$ — вектор из T_n , для которого выполняется неравенство

$$d_n(\mathbf{y}, \mathbf{u}(x)) \leq \epsilon - \delta_0. \quad (5.5.22)$$

Рассмотрим отображение множества $V_n(\sqrt{D}, \delta_1)$ на множество T_n , $\mathbf{x} \rightarrow \mathbf{u}(\mathbf{x})$, задаваемое следующей цепочкой последовательных отображений:

$$\mathbf{x} \rightarrow \mathbf{x}' \rightarrow \mathbf{y} \rightarrow \mathbf{u}(\mathbf{x}).$$

Будем, кроме того, считать, что $\mathbf{x} \rightarrow \mathbf{u}_0$ для каждого вектора \mathbf{x} , не принадлежащего множеству $V_n(\sqrt{D}, \delta_1)$. Тем самым определено отображение $\mathbf{x} \rightarrow \mathbf{u}(\mathbf{x})$ множества X^n на множество T_n^* .

Оценим теперь ошибку такого отображения. Для всякого вектора $\mathbf{x} \in V_n(\sqrt{D}, \delta_1)$ имеем

$$\begin{aligned} d_n(\mathbf{x}, \mathbf{u}(\mathbf{x})) &\leq [d_n(\mathbf{x}, \mathbf{x}')^{1/2} + d_n(\mathbf{x}', \mathbf{y})^{1/2} + d_n(\mathbf{y}, \mathbf{u}(\mathbf{x}))^{1/2}]^2 \leq \\ &\leq \left[\delta_1 + 2 \sqrt{\frac{D}{n}} + \sqrt{\epsilon - \delta_0} \right]^2 \leq \epsilon - \delta_2, \end{aligned} \quad (5.5.23)$$

где δ_2 зависит от δ_0 , δ_1 и n . Обозначим эту зависимость через

$$\delta_2 \triangleq \varphi(\epsilon, \delta_1, n) \quad (5.5.24)$$

и заметим, что при достаточно больших n влияние n на δ_2 исчезает. Выбирая параметры δ_0 и δ_1 подходящим образом, можно получить любое желаемое значение параметра δ_2 . Далее, для всякого вектора $\mathbf{x} \notin V_n(\sqrt{D}, \delta_1)$ имеем

$$d_n(\mathbf{x}, \mathbf{u}(\mathbf{x})) = d_n(\mathbf{x}, \mathbf{u}_0) = \frac{1}{n} \sum_{i=1}^n x_i^2. \quad (5.5.25)$$

Ошибка $d_n(\mathbf{x}, \mathbf{u}(\mathbf{x}))$ является случайной величиной на ансамбле $\{X^n, f(\mathbf{x})\}$ всех последовательностей сообщений источника, который порождает независимые с. в. X_1, \dots, X_n , имеющие одинаковые распределения вероятностей. Оценим среднюю ошибку

$$\begin{aligned}\overline{d}_n &= \mathbf{M}d_n(\mathbf{X}, \mathbf{u}(\mathbf{X})) = \int_{X^n} d_n(\mathbf{x}, \mathbf{u}(\mathbf{x}))f(\mathbf{x})d\mathbf{x} = \\ &= \int_{V_n(\sqrt{D}, \delta_1)} d_n(\mathbf{x}, \mathbf{u}(\mathbf{x}))f(\mathbf{x})d\mathbf{x} + \int_{X^n \setminus V_n(\sqrt{D}, \delta_1)} d_n(\mathbf{x}, \mathbf{u}(\mathbf{x}))f(\mathbf{x})d\mathbf{x}.\end{aligned}\quad (5.5.26)$$

В последнем соотношении область интегрирования X^n представлена как соединение двух частей $V_n(\sqrt{D}, \delta_1)$ и $X^n \setminus V_n(\sqrt{D}, \delta_1)$ — дополнения первой части до X^n . Воспользуемся тем, что для всех векторов из $V_n(\sqrt{D}, \delta_1)$ выполняется неравенство (5.5.23). Тогда получим, что

$$\overline{d}_n \leq (\varepsilon - \delta_2) \Pr(X \in V_n(\sqrt{D}, \delta_1)) + \delta_3 \leq \varepsilon - \delta_2 + \delta_3, \quad (5.5.27)$$

где использовано то, что вероятность всякого события не превышает единицу, и обозначено

$$\delta_3 \triangleq \int_{X^n \setminus V_n(\sqrt{D}, \delta_1)} d_n(\mathbf{x}, \mathbf{u}(\mathbf{x}))f(\mathbf{x})d\mathbf{x}. \quad (5.5.28)$$

Из принципа «затвердевания» сферы следует, что для любого $\gamma > 0$ найдется N такое, что

$$\Pr(X \in V_n(\sqrt{D}, \delta_1)) \geq 1 - \gamma \quad (5.5.29)$$

для всех $n > N$. Последнее неравенство позволяет оценить величину δ_3 . Заметим, что для всех векторов $\mathbf{x} \in V_n(\sqrt{D}, \delta_1)$ имеет место неравенство

$$d_n(\mathbf{x}, \mathbf{u}_0) = \frac{1}{n} \sum_{j=1}^n x_j^2 \geq (\sqrt{D} - \delta_1)^2. \quad (5.5.30)$$

Следовательно,

$$\begin{aligned}D = \mathbf{M} \frac{1}{n} \sum_{j=1}^n X_j^2 &= \int_{V_n(\sqrt{D}, \delta_1)} d_n(\mathbf{x}, \mathbf{u}_0)f(\mathbf{x})d\mathbf{x} + \\ &+ \int_{X^n \setminus V_n(\sqrt{D}, \delta_1)} d_n(\mathbf{x}, \mathbf{u}_0)f(\mathbf{x})d\mathbf{x} \geq (\sqrt{D} - \delta_1)^2 \Pr(X \in V_n(\sqrt{D}, \delta_1)) + \\ &+ \delta_3 \geq (\sqrt{D} - \delta_1)^2(1 - \gamma) + \delta_3 \geq D - D\gamma - 2\delta_1\sqrt{D} + \delta_3, \quad (5.5.31)\end{aligned}$$

откуда следует, что для всех $n > N$

$$\delta_3 \leq 2\delta_1\sqrt{D} + D\gamma. \quad (5.5.32)$$

Подставим теперь последнее неравенство в (5.5.27). В результате получим

$$\overline{d}_n \leq \varepsilon - \delta_2 + 2\delta_1\sqrt{D} + \gamma D = \varepsilon + \psi(\delta_0, \delta_1, \gamma, n), \quad (5.5.33)$$

где

$$\psi(\delta_0, \delta_1, \gamma, n) \triangleq 2\delta_1\sqrt{D} + \gamma D - \varphi(\delta_0, \delta_1, n), \quad (5.5.34)$$

причем неравенство (5.5.33) имеет место при произвольных $\delta_0 > 0$, $\delta_1 > 0$, $\gamma > 0$ и при достаточно больших значениях n . Из неравенства (5.5.23) и соотношений (5.5.24) и (5.5.34) следует, что при фиксированном сколь угодно малом положительном δ_0 и при достаточно больших n можно подобрать $\delta_1 > 0$ и $\gamma > 0$ так, чтобы $\psi(\delta_0, \delta_1, \gamma, n) \leq 0$. Тогда средняя ошибка будет удовлетворять неравенству (5.5.19). Лемма доказана.

В лемме 5.5.2 утверждается, что при достаточно большом n всякое конечное множество T_n , которое аппроксимирует множество A_n с ошибкой $d_n \leq \varepsilon - \delta_0$ (это множество с учетом леммы 5.5.1 является $(\varepsilon - \delta_0)$ -сетью, аппроксимирующей сферу $S_n(\sqrt{D})$), будучи дополнено нулевым вектором, аппроксимирует ансамбль $\{X^n, f(\mathbf{x})\}$ сообщений на выходе непрерывного стационарного источника без памяти со средней ошибкой $\overline{d}_n = \mathbf{M}d_n \leq \varepsilon$. Величина δ_0 может быть взята сколь угодно малой за счет выбора достаточно большого n . Множество $T_n^* = T_n \cup \mathbf{u}_0$ аппроксимирует последовательности сообщений источника почти столь же точно, как множество T_n аппроксимирует множество векторов, лежащих на сфере $S_n(\sqrt{D})$. Средняя ошибка аппроксимации в первом случае близка к ошибке аппроксимации во втором, хотя источник в принципе может порождать такие последовательности сообщений, которые сильно отличаются от векторов указанной сферы.

5.5.4. Прямая теорема кодирования.

Теорема 5.5.1. Пусть источник порождает независимые гауссовские с. в. с нулевыми математическими ожиданиями и дисперсиями D . Пусть $H(\varepsilon)$ — эпсилон-энтропия этого источника относительно квадратичного критерия качества $d(x, y) = (x - y)^2$. Тогда при достаточно больших n существует (R, d) -код такой, что при любом положительном δ

$$R = \frac{\log M}{n} = H(\varepsilon) + \delta \quad (5.5.35)$$

и

$$d = \mathbf{M}d_n(\mathbf{X}, \mathbf{u}(\mathbf{X})) \leq \varepsilon. \quad (5.5.36)$$

Доказательство. Для доказательства теоремы будет указан некоторый метод построения аппроксимирующего множества кода $T_n^* = \{\mathbf{u}_0, \dots, \mathbf{u}_{M-1}\}$, число элементов которого удовлетворяет условию (5.5.35), а средняя ошибка — условию (5.5.36). Чтобы построить такое множество по лемме 5.5.2, достаточно указать аппроксимирующее множество для векторов n -мерной сферы $S_n(\sqrt{D})$. Для построения аппроксимирующего множества для $S_n(\sqrt{D})$ по лемме 5.5.1 достаточно указать аппроксимирующее множество для подмножества A_n векторов этой сферы, состоящего из $N \leq (2n)^n$ векторов.

Пусть $T_n = \{\mathbf{u}_1, \dots, \mathbf{u}_{M-1}\}$ — аппроксимирующее множество для A_n , фигурирующее в лемме 5.5.2. Будем строить это множество методом случайного кодирования, при котором векторы $\{\mathbf{u}_1, \dots, \mathbf{u}_{M-1}\}$ выбираются случайно и независимо из множества $S_n(\sqrt{D} - \varepsilon + \delta_0)$ — множества точек n -мерной сферы радиуса $\sqrt{(D - \varepsilon + \delta_0)n}$. Распределение вероятностей, в соответствии с которым происходит случайный выбор, будем считать равномерным на сфере $S_n(\sqrt{D} - \varepsilon + \delta_0)$, т. е. для любой области W_n на этой сфере положим

$$\Pr(\mathbf{u} \in W_n) = \frac{|W_n|}{|S_n(\sqrt{D} - \varepsilon + \delta_0)|}, \quad (5.5.37)$$

где $|W_n|$ и $|S_n(\sqrt{D} - \varepsilon + \delta_0)|$ — площади области W_n и сферы $S_n(\sqrt{D} - \varepsilon + \delta_0)$ соответственно.

Предположим, что $\mathbf{y} = (y_1, \dots, y_n)$ — некоторый вектор из множества A_n , лежащий на сфере $S_n(\sqrt{D})$. Рассмотрим n -мерные сферы $S_n(\sqrt{\varepsilon - \delta_0})$ с центром в точке \mathbf{y} и $S_n(\sqrt{D} - \varepsilon + \delta_0)$ с центром в начале координат:

$$S_n(\sqrt{\varepsilon - \delta_0}) \triangleq \left\{ \mathbf{x} = (x_1, \dots, x_n) : \sum_{i=1}^n (x_i - y_i)^2 \leq (\varepsilon - \delta_0)n \right\}, \quad (5.5.38)$$

$$S_n(\sqrt{D} - \varepsilon + \delta_0) \triangleq \left\{ \mathbf{x} = (x_1, \dots, x_n) : \sum_{i=1}^n x_i^2 \leq (D - \varepsilon + \delta_0)n \right\}.$$

Обозначим через $W_n(\sqrt{\varepsilon - \delta_0})$ часть поверхности сферы $S_n(\sqrt{D} - \varepsilon + \delta_0)$, лежащей внутри сферы $S_n(\sqrt{\varepsilon - \delta_0})$ (см. рис. 5.5.2). Очевидно, что вектор \mathbf{y} аппроксимируется с ошибкой меньшей или равной $\varepsilon - \delta_0$, если хотя бы один вектор $\mathbf{u} \in T_n$ содержитя в множестве $W_n(\sqrt{\varepsilon - \delta_0})$. Обозначим через Q вероятность того, что при случайном выборе $M-1$ векторов, образующих множество T_n , ни один из них не попадет в область $W_n(\sqrt{\varepsilon - \delta_0})$. Следовательно, Q есть вероятность того, что при

случайном выборе множества T_n вектор \mathbf{y} будет аппроксимироваться с ошибкой, превышающей $\varepsilon - \delta_0$. В силу независимости выбора и условия (5.5.37) эта вероятность определяется соотношением

$$Q = \left(1 - \frac{|W_n(\sqrt{\varepsilon - \delta_0})|}{|S_n(\sqrt{D} - \varepsilon + \delta_0)|} \right)^{M-1}. \quad (5.5.39)$$

Так как вероятность Q не зависит от выбора вектора \mathbf{y} , то вероятность того, что хотя бы один вектор множества A_n будет аппроксимироваться с ошибкой, превышающей $\varepsilon - \delta_0$, удовлетворяет неравенству

$$P \leq N \left(1 - \frac{|W_n(\sqrt{\varepsilon - \delta_0})|}{|S_n(\sqrt{D} - \varepsilon + \delta_0)|} \right)^{M-1}. \quad (5.5.40)$$

Число P при этом является вероятностью выбора «плохого» аппроксимирующего множества, а следовательно, и вероятностью выбора «плохого» кода, который хотя бы один вектор множества A_n аппроксимирует с ошибкой, превышающей $\varepsilon - \delta_0$. Если мы покажем, что при некоторых M и n эта вероятность меньше единицы, то это будет означать, что в множестве всех кодов, обеспечивающих кодирование источника с ошибкой $d \leq \varepsilon$, найдется хотя бы один код со скоростью $R = \frac{1}{n} \log M$, который обеспечивает кодирование с той же ошибкой. Другими словами, в множестве всех кодов, получающихся при случайном выборе множества T_n , не все коды являются «плохими». Поэтому теперь мы покажем, что $P < 1$.

Для оценки правой части неравенства (5.5.40) воспользуемся тем, что площадь поверхности n -мерной сферы радиуса r пропорциональна r^{n-1} :

$$|S_n(\sqrt{D} - \varepsilon + \delta_0)| = k_n (\sqrt{(D - \varepsilon + \delta_0)n})^{n-1}, \quad (5.5.41)$$

где коэффициент пропорциональности k_n зависит от n , но не зависит от радиуса. Оценим площадь $|W_n(\sqrt{\varepsilon - \delta_0})|$. Для этого рассмотрим рис. 5.5.2. Отрезки равны

$$OA = \sqrt{Dn},$$

$$OB = \sqrt{(D - \varepsilon + \delta_0)n}$$

$$AB = \sqrt{(\varepsilon - \delta_0)n}.$$

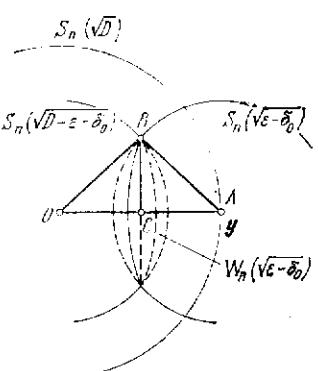


Рис. 5.5.2. К доказательству теоремы 5.5.1.

Можно легко показать (см. задачу 5.5.3), что

$$|W_n(\sqrt{\varepsilon - \delta_0})| \geq V_{n-1}(BC), \quad (5.5.43)$$

где $V_n(R)$ — объем n -мерного шара радиуса R . Так как объем n -мерного шара пропорционален R^n , причем коэффициент пропорциональности равен $V_n(1) = \frac{k_n}{n}$ (см. задачу 5.5.1), то

$$|W_n(\sqrt{\varepsilon - \delta_0})| \geq \frac{k_{n-1}}{n-1} (BC)^{n-1}. \quad (5.5.44)$$

Простые геометрические расчеты, использующие рис. 5.5.2 и формулы (5.5.42), дают

$$BC = \left[AB^2 - \left(\frac{OB^2 - OA^2 - AB^2}{2(OA)} \right)^2 \right]^{1/2} = \left[\frac{(D - \varepsilon + \delta_0)(\varepsilon - \delta_0)}{D} n \right]^{1/2}. \quad (5.5.45)$$

Таким образом,

$$\begin{aligned} P &\leq N \left[1 - \frac{k_{n-1}}{(n-1)k_n} \left(\frac{\varepsilon - \delta_0}{D} \right)^{(n-1)/2} \right]^{M-1} = \\ &= N \left[1 - 2^{-\frac{n-1}{2} \log \frac{D}{\varepsilon - \delta_0} - \log \frac{(n-1)k_n}{k_{n-1}}} \right]^{M-1}. \end{aligned} \quad (5.5.46)$$

Так как при $n \rightarrow \infty$

$$\frac{(n-1)k_n}{k_{n-1}} = \frac{(n-1)V_n(1)}{V_{n-1}(1)} \rightarrow \left(\frac{1}{2\pi n} \right)^{1/2}, \quad (5.5.47)$$

где использованы результаты задач 5.5.1—5.5.3, то при достаточно больших n будет выполняться следующее неравенство:

$$P \leq N \left[1 - 2^{\frac{n}{2} \log \frac{D}{\varepsilon - \delta_0}} \right]^{M-1}. \quad (5.5.48)$$

Положим $M = 2^{n[H(\varepsilon) + \delta]}$, тогда будет выполняться соотношение (5.5.35), и воспользуемся неравенством $N \leq (2n)^n$. Так как при $\varepsilon \leq D$ энтропия гауссовского источника без памяти равна

$$H(\varepsilon) = \frac{1}{2} \log \frac{D}{\varepsilon} \quad (5.5.49)$$

и, следовательно,

$$M = 2^{n \left(\frac{1}{2} \log \frac{D}{\varepsilon} + \delta \right)}, \quad (5.5.50)$$

то вычисляя натуральный логарифм левой и правой частей неравенства (5.5.48), а также используя указанную оценку для N , выражение (5.5.50) и неравенство $\ln x \leq x - 1$, получим

$$\begin{aligned} \ln P &\leq \ln N + M \ln \left(1 - 2^{-\frac{n}{2} \log \frac{D}{\varepsilon - \delta_0}} \right) \leq \\ &\leq n \ln (2n) - 2^{\frac{n}{2} \left(\log \frac{D}{\varepsilon} + \delta - \log \frac{D}{\varepsilon - \delta_0} \right)} + 2^{-\frac{n}{2} \log \frac{D}{\varepsilon - \delta_0}} = \\ &= n \ln (2n) - 2^{\frac{n}{2} \left(\log \frac{\varepsilon - \delta_0}{\varepsilon} + \delta \right)} + 2^{-\frac{n}{2} \log \frac{D}{\varepsilon - \delta_0}}. \end{aligned} \quad (5.5.51)$$

Если теперь задаться некоторым положительным числом α , $\alpha < \delta$, и положить

$$\delta_0 = \varepsilon (1 - 2^{2(\alpha-\delta)}), \quad (5.5.52)$$

то

$$\frac{1}{2} \log \frac{\varepsilon - \delta_0}{\varepsilon} + \delta = \alpha > 0 \quad (5.5.53)$$

и поэтому

$$\ln P \leq n \ln (2n) - 2^{\alpha n} + 2^{-\frac{n}{2} \log \frac{D}{\varepsilon - \delta_0}}. \quad (5.5.54)$$

Правая часть неравенства (5.5.54) становится отрицательной при достаточно больших значениях n . Следовательно, при $\varepsilon \leq D$ и при достаточно больших n вероятность выбрать «плохой» код меньше единицы. При $\varepsilon > D$ тривиальный код, аппроксимирующее множество которого состоит из одного нулевого вектора, удовлетворяет условиями теоремы. Теорема доказана.

5.5.5. Обсуждение. Из доказательства вспомогательных лемм и прямой теоремы кодирования следует несколько важных выводов.

1. Последовательности сообщений источника независимых гауссовых сообщений, рассматриваемые как векторы n -мерного евклидова пространства, с высокой вероятностью лежат вблизи поверхности n -мерной сферы радиуса \sqrt{Dn} с центром в начале координат. В силу симметрии гауссова n -мерного распределения на сфере нет преимущественных областей. Все сообщения распределены на «твердой» сфере равномерно.

Хотя последовательности сообщений источника и лежат вблизи поверхности n -мерной сферы радиуса \sqrt{Dn} , для их аппроксимации нужно выбирать точки, которые не лежат на этой сфере. Оказывается, что наилучший с точки зрения числа точек выбор аппроксимирующего множества состоит в выборе точек, лежащих на n -мерной сфере меньшего радиуса, который зависит от требуемой ошибки аппроксимации. При доказательстве теоремы мы сразу взяли сферу $S_n(\sqrt{D - \varepsilon + \delta_0})$, на которой выбирались аппроксимирующие векторы. Из обратной теоремы

кодирования следует, что невозможно выбрать меньше чем $2^{nH(\epsilon)}$ аппроксимирующих векторов, и следовательно, выбор сферы $S_n(\sqrt{D} - \epsilon + \delta_0)$ является наилучшим.

В теореме 5.5.1 доказано, что вероятность P выбора «плохого» кода меньше единицы. Это доказывало существование кода, удовлетворяющего условиям (5.5.35) и (5.5.36). Однако из неравенства (5.5.54) следует, что при достаточно больших n

$$P < \exp\{-2^{\alpha n-1}\},$$

т. е. вероятность выбора «плохого» кода чрезвычайно мала при больших n и почти любой случайнм образом выбираемый код, аппроксимирующее множество которого состоит примерно из $2^{nH(\epsilon)}$ векторов, будет обеспечивать среднюю ошибку, примерно равную ϵ .

2. Хотя в формулировке прямой теоремы идет речь о гауссовском источнике, однако единственное место, где использовалась гауссовость, — это формула (5.5.49). Поэтому, если источник без памяти, порождающий негауссовские сообщения, которые имеют нулевое среднее, дисперсию D и конечный четвертый момент, кодировать с помощью кода, дающего среднеквадратическую ошибку $d \leq \epsilon$ в случае гауссовского источника, то среднеквадратическая ошибка кодирования в негауссовском случае также будет меньше или равна ϵ . Отсюда следует, что величина $\frac{1}{2} \log \frac{D}{\epsilon}$ является верхней границей достижимой скорости кодирования негауссовского источника без памяти при среднеквадратической ошибке $\epsilon \leq D$.

3. Прямая теорема кодирования формулировалась для источника сообщений (случайных величин), математическое ожидание которых равно нулю. Очевидно, что все доказательство без изменений переносится на тот случай, когда сообщения имеют ненулевое математическое ожидание. Отличие состоит только в том, что вместо сферы $S_n(\sqrt{D})$ с центром в начале координат следует использовать такую же сферу с центром в точке (m, \dots, m) , где m — математическое ожидание.

4. Доказанная прямая теорема вместе с обратной теоремой кодирования позволяет сделать следующий вывод. Для источника гауссовых сообщений без памяти ϵ -скорость создания информации при среднеквадратическом критерии качества равна $H(\epsilon) = \frac{1}{2} \log \max \{1, D/\epsilon\}$ — эпсилон-энтропии гауссовой случайной величины относительно квадратического критерия качества.

Из замечания, сделанного в п. 2, следует, что источник без памяти, порождающий гауссовые сообщения, имеет наибольшую ϵ -скорость создания информации при квадратическом критерии качества среди всех источников без памяти с фиксированной дисперсией.

§ 5.6.* Эпсилон-энтропия гауссова случайного вектора

В этом параграфе мы начнем изучение эпсилон-энтропии гауссовых источников более сложных, чем источники без памяти. Хотя в дальнейшем нас будут интересовать стационарные источники, здесь мы получим выражение для эпсилон-энтропии гауссова случайного вектора, в общем случае не обязательно связанного с некоторым стационарным источником.

Пусть $\mathbf{X} = (X_1, \dots, X_n)$ — случайный вектор с ф. п. в. $f(\mathbf{x})$, $\mathbf{x} = (x^{(1)}, \dots, x^{(n)})$. Пусть $\mathbf{Y} = (Y_1, \dots, Y_n)$ — аппроксимирующий случайный вектор, причем правило аппроксимации задается условной ф. п. в. $f(\mathbf{y}|\mathbf{x})$, $\mathbf{y} = (y^{(1)}, \dots, y^{(n)})$. Если множество аппроксимирующих векторов конечно, то функцию $f(\mathbf{y}|\mathbf{x})$ будем рассматривать как обобщенную ф. п. в. Эта условная ф. п. в. вместе с функцией $f(\mathbf{x})$ задает распределение вероятностей на множестве $X^n Y^n$ всех пар (\mathbf{x}, \mathbf{y}) , где \mathbf{x} — реализация случайного вектора \mathbf{X} , а \mathbf{y} — реализация аппроксимирующего вектора \mathbf{Y} . Ф. п. в. этого распределения $f(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) f(\mathbf{y}|\mathbf{x})$. Пусть $d(\mathbf{x}, \mathbf{y})$ — критерий качества и

$$d_n(\mathbf{x}, \mathbf{y}) \triangleq \frac{1}{n} \sum_{i=1}^n d(x^{(i)}, y^{(i)})$$

— ошибка аппроксимации вектора \mathbf{x} с помощью вектора \mathbf{y} .

Определение 5.6.1. Эпсилон-энтропией случайного вектора $\mathbf{X} = (X_1, \dots, X_n)$ относительно критерия качества $d(\mathbf{x}, \mathbf{y})$ называется функция

$$H_n(\epsilon) \triangleq \min_{f(y|x) \in \Phi_n(\epsilon)} \frac{1}{n} I(X_1, \dots, X_n; Y_1, \dots, Y_n), \quad (5.6.1)$$

где $\Phi_n(\epsilon)$ — класс условных ф. п. в. $f(\mathbf{y}|\mathbf{x})$ таких, что средняя ошибка

$$\overline{d}_n \triangleq M d_n(\mathbf{x}, \mathbf{y}) \leq \epsilon. \quad (5.6.2)$$

Очевидно, $\overline{d}_n = \frac{1}{n} \sum_{i=1}^n \overline{d}^{(i)}$, где

$$\overline{d}^{(i)} \triangleq M d(x^{(i)}, y^{(i)}) \quad (5.6.3)$$

— средняя ошибка аппроксимации i -й компоненты вектора \mathbf{X} .

Ниже мы будем рассматривать только случай гауссового вектора \mathbf{X} и квадратичного критерия качества $d(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^2$. Вначале будет рассмотрен простейший случай, когда компоненты вектора \mathbf{X} представляют собой независимые гауссовые с. в., а затем будет рассмотрен общий случай.

5.6.1. Эпсилон-энтропия системы независимых гауссовских случайных величин. В случае независимых с. в. вектор \mathbf{X} имеет ф. п. в.

$$f(\mathbf{x}) = \prod_{i=1}^n f_i(x^{(i)}). \quad (5.6.4)$$

Будем предполагать, что

$$f_i(x^{(i)}) = \frac{1}{V\sqrt{2\pi D_i}} \exp\left(-\frac{(x^{(i)})^2}{2D_i}\right). \quad (5.6.5)$$

Из независимости с. в. X_1, \dots, X_n следует, что

$$\begin{aligned} I(X; Y) &= H_0(\mathbf{X}) - H_0(\mathbf{X}|Y) = \\ &= \sum_{i=1}^n H_0(X_i) - \sum_{i=1}^n H_0(X_i|Y_1, \dots, Y_{i-1}, X_{i+1}, \dots, X_{n-1}) \geqslant \\ &\geqslant \sum_{i=1}^n H_0(X_i) - \sum_{i=1}^n H_0(X_i|Y_i) = \sum_{i=1}^n I(X_i; Y_i), \end{aligned} \quad (5.6.6)$$

причем равенство достигается в том случае, когда пары с. в. $(X_1, Y_1), \dots, (X_n, Y_n)$ статистически независимы, т. е. когда

$$f(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n f_i(y^{(i)}|x^{(i)}) \quad (5.6.7)$$

для всех последовательностей \mathbf{x} и \mathbf{y} . Из (5.6.6) следует, что

$$\begin{aligned} H_n(\boldsymbol{\varepsilon}) &= \min_{\Phi_n(\boldsymbol{\varepsilon})} \frac{1}{n} \sum_{i=1}^n I(X_i; Y_i) = \\ &= \min_{\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n} \min_{\substack{f_i(y|x) \in \Phi(\boldsymbol{\varepsilon}_i), \\ i=1, \dots, n}} \frac{1}{n} \sum_{i=1}^n I(X_i; Y_i). \end{aligned} \quad (5.6.8)$$

В дальнейшем будет показано, что в множестве $\Phi_n(\boldsymbol{\varepsilon})$ существует ф. п. в., на которой достигается минимум в первом выражении в (5.6.8). Минимум по $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n$ во втором выражении в (5.6.8) ищется при условии

$$\frac{1}{n} \sum_{i=1}^n \boldsymbol{\varepsilon}_i \leq \boldsymbol{\varepsilon}, \quad (5.6.9)$$

а минимум по $f_i(y|x)$, $i = 1, \dots, n$, ищется по всем таким одномерным ф. п. в., что средняя ошибка i -й компоненты вектора \mathbf{X}

не превосходит $\boldsymbol{\varepsilon}_i$, $i = 1, \dots, n$. Так как слагаемые в (5.6.8) минимизируются независимо, то можно записать, что

$$H_n(\boldsymbol{\varepsilon}) = \min_{\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n} \frac{1}{n} \sum_{i=1}^n \min_{\Phi(\boldsymbol{\varepsilon}_i)} I(X_i; Y_i) = \min_{\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n} \frac{1}{n} \sum_{i=1}^n H(\boldsymbol{\varepsilon}_i), \quad (5.6.10)$$

где $H(\boldsymbol{\varepsilon}_i)$ — эпсилон-энтропия гауссовой с. в. относительно квадратического критерия качества. В § 5.4 было установлено, что

$$H(\boldsymbol{\varepsilon}_i) = \frac{1}{2} \log \max \left\{ 1, \frac{D_i}{\boldsymbol{\varepsilon}_i} \right\}, \quad (5.6.11)$$

поэтому

$$H_n(\boldsymbol{\varepsilon}) = \min_{\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n} \frac{1}{2n} \sum_{i=1}^n \log \max \left\{ 1, \frac{D_i}{\boldsymbol{\varepsilon}_i} \right\} \quad (5.6.12)$$

при условии, что

$$\frac{1}{n} \sum_{i=1}^n \boldsymbol{\varepsilon}_i \leq \boldsymbol{\varepsilon}. \quad (5.6.13)$$

При минимизации (5.6.12) можно полагать, что для любого i величина ошибки $\boldsymbol{\varepsilon}_i$ не превосходит D_i . Действительно, если бы $\boldsymbol{\varepsilon}_i > D_i$ для некоторого номера i , то i -е слагаемое в сумме (5.6.12) было бы равно нулю. Но тогда значение суммы могло бы быть уменьшено без увеличения общей ошибки за счет выбора $\boldsymbol{\varepsilon}_i = D_i$ и увеличения ошибки $\boldsymbol{\varepsilon}_j$ на той компоненте, для которой $\boldsymbol{\varepsilon}_j < D_j$.

Заметим, кроме того, что $H_n(\boldsymbol{\varepsilon})$ монотонно не возрастает при увеличении $\boldsymbol{\varepsilon}$, поэтому условие (5.6.13) можно заменить следующим:

$$\frac{1}{n} \sum_{i=1}^n \boldsymbol{\varepsilon}_i = \boldsymbol{\varepsilon}. \quad (5.6.14)$$

Если минимум найдется при условии (5.6.14), то этот минимум будет равен искомой величине $H_n(\boldsymbol{\varepsilon})$.

Таким образом, мы получили следующую задачу минимизации:

$$H_n(\boldsymbol{\varepsilon}) = \min_{\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n} \frac{1}{2} \sum_{i=1}^n \log \frac{D_i}{\boldsymbol{\varepsilon}_i}, \quad (5.6.15)$$

$$\frac{1}{n} \sum_{i=1}^n \boldsymbol{\varepsilon}_i = \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon}_i \leq D_i, \quad i = 1, \dots, n.$$

Нетрудно проверить, что область R допустимых векторов $(\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n)$, по которой разыскивается минимум, является выпуклой. Действительно, для любых двух векторов $(\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n)$ и $(\boldsymbol{\varepsilon}'_1, \dots, \boldsymbol{\varepsilon}'_n)$ из R и любого λ , $0 < \lambda < 1$, вектор

$$\lambda \cdot (\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n) + (1 - \lambda) \cdot (\boldsymbol{\varepsilon}'_1, \dots, \boldsymbol{\varepsilon}'_n)$$

принадлежит R . Нетрудно также проверить, что функция, стоящая под знаком минимума, является выпуклой вниз. Это следует из того, что каждое слагаемое — выпуклая вниз функция и переменные входят по одному в каждое слагаемое. Поэтому задача минимизации (5.6.15) является стандартной задачей минимизации выпуклой вниз функции на выпуклом множестве векторов. Решение ее дается теоремой Куна—Таккера (см. § 2.8, теоремы 2.8.1, 2.8.2).

Согласно этим теоремам необходимым и достаточным условием того, чтобы вектор $(\varepsilon_1, \dots, \varepsilon_n) \in R$ минимизировал функцию

$$F(\varepsilon_1, \dots, \varepsilon_n) \triangleq \frac{1}{2n} \sum_{i=1}^n \log \frac{D_i}{\varepsilon_i} \quad (5.6.16)$$

при условиях

$$\frac{1}{n} \sum_{i=1}^n \varepsilon_i = \varepsilon, \quad \varepsilon_i \leq D_i, \quad i = 1, \dots, n, \quad (5.6.17)$$

является существование такого числа μ , что

$$\frac{\partial F}{\partial \varepsilon_i} \begin{cases} = \mu, & \text{для всех } i, \text{ для которых } \varepsilon_i < D_i, \\ \leq \mu, & \text{для всех } i, \text{ для которых } \varepsilon_i = D_i. \end{cases} \quad (5.6.18)$$

Если удастся подобрать некоторое μ и вектор $(\varepsilon_1^*, \dots, \varepsilon_n^*)$, удовлетворяющий ограничениям (5.6.17), и если при этом будут выполняться условия (5.6.18), то по теореме Куна—Таккера этот вектор будет минимизировать функцию (5.6.16).

Пусть θ — корень уравнения

$$\frac{1}{n} \sum_{i=1}^n \min \{\theta, D_i\} = \varepsilon \quad (5.6.19)$$

и пусть

$$\varepsilon_i^* \triangleq \begin{cases} \theta, & \text{если } \theta < D_i, \\ D_i, & \text{если } \theta \geq D_i. \end{cases} \quad (5.6.20)$$

Очевидно, что вектор $(\varepsilon_1^*, \dots, \varepsilon_n^*)$, определяемый соотношениями (5.6.19) и (5.6.20), удовлетворяет условиям (5.6.17). Покажем, что при

$$\mu = -\frac{1}{2n} \frac{\log \varepsilon}{\theta} \quad (5.6.21)$$

условия (5.6.18) удовлетворяются.

Действительно,

$$\frac{\partial F}{\partial \varepsilon_i} = -\frac{1}{2n} \frac{\log \varepsilon}{\varepsilon_i}. \quad (5.6.22)$$

Так как $\varepsilon_i^* = \theta$ для всех i , для которых $\varepsilon_i^* < D_i$, и $\varepsilon_i^* = D_i \leq \theta$ для остальных i , то

$$\begin{aligned} \frac{\partial F}{\partial \varepsilon_i} \Big|_{\varepsilon_i=\varepsilon_i^*} &= -\frac{1}{2n} \frac{\log \varepsilon}{\theta} = \mu, \quad \text{если } \varepsilon_i^* < D_i, \\ \frac{\partial F}{\partial \varepsilon_i} \Big|_{\varepsilon_i=\varepsilon_i^*} &= -\frac{1}{2n} \frac{\log \varepsilon}{\varepsilon_i^*} \leq -\frac{1}{2n} \frac{\log \varepsilon}{\theta} = \mu, \quad \text{если } \varepsilon_i^* = D_i. \end{aligned} \quad (5.6.23)$$

Таким образом, указанный выше вектор $(\varepsilon_1^*, \dots, \varepsilon_n^*)$ минимизирует функцию $F(\varepsilon_1, \dots, \varepsilon_n)$ и

$$H_n(\varepsilon) = \frac{1}{2n} \sum_{i=1}^n \log \max \left\{ 1, \frac{D_i}{\theta} \right\} \quad (5.6.24)$$

есть эпсилон-энтропия гауссовского вектора с независимыми компонентами, дисперсии которых равны D_i , $i = 1, \dots, n$.

З а м е ч а н и е. При выводе формулы (5.6.24) мы, не оговаривая, предполагали, что все дисперсии D_1, \dots, D_n положительны. В противном случае неравенство (5.6.6) не было бы обоснованным, так как относительные энтропии с. в. с нулевыми дисперсиями равны минус бесконечности. На самом деле формула (5.6.24) остается справедливой и в случае, когда некоторые или все дисперсии равны нулю. В этом случае те компоненты вектора X , которые имеют нулевые дисперсии, принимают неслучайные значения и, следовательно, должны аппроксимироваться также неслучайными величинами. Поскольку они не случайны, то эти пары величин статистически не зависят от остальных и информация $I(X; Y)$ равна информации $I(X'; Y')$, где X' — это такая подсистема системы X , все с. в. которой имеют ненулевые дисперсии.

5.6.2. Эпсилон-энтропия системы зависимых гауссовских случайных величин. Будем теперь считать, что случайный вектор $X = (X_1, \dots, X_n)$ образован зависимыми гауссовскими с. в. и $K = [K_{ij}]$ — корреляционная матрица вектора X . Метод вычисления эпсилон-энтропии в этом случае основан на применении ортогонального преобразования (см. § 2.4), которое вектор X переводит в случайный вектор с независимыми гауссовскими компонентами. Рассмотрим следующую цепочку преобразований, где через Q обозначена ортогональная матрица, диагонализирующую корреляционную матрицу K :

$$X \xrightarrow{Q} X^* \xrightarrow{\text{аппроксимация}} Y^* \xrightarrow{Q^T} Y. \quad (5.6.25)$$

Первое преобразование

$$X^* = XQ \quad (5.6.26)$$

переводит данный случайный вектор \mathbf{X} в вектор \mathbf{X}^* с независимыми компонентами. Затем этот вектор аппроксимируется вектором \mathbf{Y}^* со среднеквадратической ошибкой $\bar{d}_n^* \ll \varepsilon$. Наконец, вектор \mathbf{Y}^* преобразуется в вектор \mathbf{Y} :

$$\mathbf{Y} = \mathbf{Y}^* \mathbf{Q}^\tau \quad (5.6.27)$$

с помощью ортогонального преобразования \mathbf{Q}^τ , обратного преобразованию \mathbf{Q} .

Ортогональные преобразования являются обратимыми и поэтому

$$I(\mathbf{X}^*; \mathbf{Y}^*) = I(\mathbf{X}; \mathbf{Y}). \quad (5.6.28)$$

Покажем, что при ортогональных преобразованиях среднеквадратическая ошибка не изменяется. Среднеквадратическая ошибка \bar{d}_n аппроксимации вектора \mathbf{X} с помощью вектора \mathbf{Y} , очевидно, может быть записана с помощью следующего матричного выражения:

$$\bar{d}_n = \mathbf{M} \frac{1}{n} (\mathbf{X} - \mathbf{Y})(\mathbf{X} - \mathbf{Y})^\tau. \quad (5.6.29)$$

Аналогичное выражение может быть записано для среднеквадратической ошибки \bar{d}_n^* аппроксимации вектора \mathbf{X}^* с помощью вектора \mathbf{Y}^* . Тогда

$$\begin{aligned} \bar{d}_n^* &= \mathbf{M} \frac{1}{n} (\mathbf{X}^* - \mathbf{Y}^*)(\mathbf{X}^* - \mathbf{Y}^*)^\tau = \\ &= \mathbf{M} \frac{1}{n} (\mathbf{X} - \mathbf{Y}) \mathbf{Q} \mathbf{Q}^\tau (\mathbf{X} - \mathbf{Y})^\tau = \mathbf{M} \frac{1}{n} (\mathbf{X} - \mathbf{Y})(\mathbf{X} - \mathbf{Y})^\tau = \bar{d}_n, \end{aligned} \quad (5.6.30)$$

где использовано то, что $\mathbf{Q} \mathbf{Q}^\tau = \mathbf{I}$ и, как следствие этого, что $\mathbf{Y}^* = \mathbf{YQ}$.

Из соотношения (5.6.30) вытекает, что при аппроксимации вектора \mathbf{X}^* с помощью вектора \mathbf{Y}^* со среднеквадратической ошибкой, меньшей или равной ε , вектор \mathbf{X} аппроксимируется вектором \mathbf{Y} со среднеквадратической ошибкой, также не превышающей ε . Если при этом аппроксимация \mathbf{X}^* с помощью \mathbf{Y}^* осуществляется так, что минимизируется величина

$$\frac{1}{n} I(\mathbf{X}^*; \mathbf{Y}^*),$$

то из (5.6.28) следует, что и аппроксимация \mathbf{X} с помощью \mathbf{Y} осуществляется так, что минимизируется величина

$$\frac{1}{n} I(\mathbf{X}; \mathbf{Y}).$$

Поэтому эпсилон-энтропия $H_n(\varepsilon)$ гауссовского вектора \mathbf{X} равна эпсилон-энтропии $H_n^*(\varepsilon)$ гауссовского вектора \mathbf{X}^* .

В п. 5.6.1 мы получили формулу, которая позволяет вычислить величину $H_n^*(\varepsilon)$. Так как дисперсии компонент вектора \mathbf{X}^* , полученного в результате ортогонального преобразования из вектора \mathbf{X} , равны собственным числам $\lambda_1, \dots, \lambda_n$ матрицы \mathbf{K} , то

$$H_n(\varepsilon) = H_n^*(\varepsilon) = \frac{1}{2n} \sum_{i=1}^n \log \max \left\{ 1, \frac{\lambda_i}{\theta} \right\}, \quad (5.6.31)$$

где θ есть корень уравнения

$$\frac{1}{n} \sum_{i=1}^n \min \{ \theta, \lambda_i \} = \varepsilon. \quad (5.6.32)$$

§ 5.7 *. ЭПСИЛОН-ЭНТРОПИЯ СТАЦИОНАРНОГО ГАУССОВСКОГО ПРОЦЕССА ДИСКРЕТНОГО ВРЕМЕНИ

В этом параграфе мы применим результаты предыдущего рассмотрения эпсилон-энтропии системы гауссовских случайных величин для вычисления эпсилон-энтропии стационарного гауссовского случайного процесса, дискретного по времени, или, что то же самое, эпсилон-энтропии стационарного дискретного по времени источника гауссовских сообщений.

Рассмотрим источник U_X , который в каждый момент времени выбирает сообщения из ансамбля $\{X, f(x)\}$ и последовательность сообщений на выходе которого представляет собой стационарный случайный процесс. Будем предполагать, что процесс является гауссовским. Это означает, что для любого $n = 1, 2, \dots$ система с. в. $\mathbf{X} = (X_1, \dots, X_n)$ имеет n -мерное гауссовское распределение вероятностей с корреляционной матрицей \mathbf{K}_n .

Напомним, что эпсилон-энтропией стационарного источника относительно критерия качества $d(x, y)$ называется следующее выражение:

$$H(\varepsilon) = \inf_n H_n(\varepsilon), \quad (5.7.1)$$

где

$$H_n(\varepsilon) = \min_{\Phi_n(\varepsilon)} \frac{1}{n} I(X_1, \dots, X_n; Y_1, \dots, Y_n) \quad (5.7.2)$$

и $\Phi_n(\varepsilon)$ — класс условных ф. п. в. $f(y|x)$ таких, что для каждой функции из этого класса средняя ошибка \bar{d}_n относительно критерия качества $d(x, y)$ не превосходит ε , т. е.

$$\Phi_n(\varepsilon) = \{f(y|x): \bar{d}_n \triangleq \mathbf{M} d_n(x, y) \leq \varepsilon\}. \quad (5.7.3)$$

В дальнейшем будет рассматриваться только квадратический критерий качества $d(x, y) = (x - y)^2$.

Для каждого конечного n мы имеем дело с гауссовским вектором X , эпсилон-энтропия которого вычислена в предыдущем разделе. Если обозначать через $\lambda_1^{(n)}, \dots, \lambda_n^{(n)}$ собственные числа матрицы K_n , то согласно (5.6.31) и (5.6.32), можно записать

$$\begin{aligned} H_n(\epsilon) &= \frac{1}{2n} \sum_{i=1}^n \log \max \left\{ 1, \frac{\lambda_i^{(n)}}{\theta} \right\}, \\ &\quad -\frac{1}{n} \sum_{i=1}^n \min \{ \theta, \lambda_i^{(n)} \} = \epsilon. \end{aligned} \quad (5.7.4)$$

Таким образом, из (5.7.1) и теоремы 5.2.3 следует, что для вычисления эпсилон-энтропии стационарного источника надо перейти к пределам в соотношениях (5.7.4) по n . Как мы увидим ниже, стационарность источника обуславливает существование этих пределов.

Введем понятие спектральной плотности мощности случайного процесса с дискретным временем. Пусть $X = \{X_1, X_2, \dots, X_n\}$ — случайный вектор, образованный сообщениями на выходе стационарного источника, и K_n — его корреляционная матрица. Из условия стационарности источника следует, что каждый элемент K_{ij} матрицы K_n зависит только от абсолютного значения разности $|i-j|$, т. е.

$$K_{ij} = K_{ji} = K_{|i-j|}. \quad (5.7.5)$$

Матрицы, элементы которых удовлетворяют свойству (5.7.5), называются *теплицевыми*. Пусть $u \triangleq i - j$, $u = -(n-1), \dots, 0, 1, \dots, n-1$. Из (5.7.5) следует, что $K_u = K_{-u}$ и что корреляционная матрица отрезка стационарного процесса полностью определяется n числами K_u , $u = 0, 1, \dots, n-1$, являющимися корреляционными моментами с. в. X_i, X_j , которые отстоят друг от друга на u единиц времени.

В дальнейшем мы будем предполагать также, что

$$\lim_{u \rightarrow \infty} K_u = 0.$$

Можно показать, что в случае гауссовских процессов это предположение равносильно предположению об эргодичности.

Рассмотрим ряд Фурье

$$\sum_{u=-\infty}^{\infty} K_u e^{-j2\pi fu} \triangleq N(f), -\frac{1}{2} \leq f \leq \frac{1}{2}. \quad (5.7.6)$$

Функция $N(f)$ в том случае, когда она существует, называется спектральной плотностью мощности процесса, порожденного стационарным источником U_X . Далее мы ограничимся рассмотрением

лишь таких источников, для которых $N(f)$ — непрерывная и ограниченная функция. При этом существует обратное преобразование, позволяющее по спектральной плотности мощности определять корреляционные моменты K_u , $u = 0, 1, 2, \dots$:

$$K_u = \int_{-1/2}^{1/2} N(f) e^{j2\pi fu} df.$$

Отметим основные свойства спектральной плотности мощности случайного процесса. Из выражений (5.7.5) и (5.7.6) следует, что $N(f)$ — действительная функция, которую можно записать в виде

$$N(f) = 2 \sum_{u=1}^{\infty} K_u \cos 2\pi fu + K_0. \quad (5.7.7)$$

Из (5.7.7) следует, что $N(f) = N(-f)$ и поэтому

$$K_u = \int_{-1/2}^{1/2} N(f) \cos 2\pi fu df. \quad (5.7.8)$$

Можно показать также, что $N(f) \geq 0$ для всех $f \in [-1/2, 1/2]$. При этом из (5.7.8) имеем

$$K_0 = \int_{-1/2}^{1/2} N(f) df. \quad (5.7.9)$$

Так как K_0 есть дисперсия (средняя мощность) случайного процесса на выходе источника, то из (5.7.9) следует, что функция $N(f)$ характеризует распределение мощности по частотам.

Вывод формулы для ϵ -энтропии гауссовского стационарного источника с дискретным временем при квадратичном критерии качества основан на следующей лемме, которую мы здесь приводим без доказательства (интересующийся доказательством читатель должен обратиться к книгам Гренандера и Сегё [2] или к книге Галлагера [1]).

Л е м м а 5.7.1. Пусть K_n — корреляционная матрица отрезка (X_1, X_2, \dots, X_n) стационарного случайного процесса с дискретным временем и $N(f)$ — спектральная плотность мощности этого процесса. Пусть $g(x)$ — неубывающая функция, определенная для всех $x \geq 0$ и такая, что выполнены два условия: 1) $g(0) = 0$; 2) найдется число B такое, что $|g(x_1) - g(x_2)| \leq B|x_1 - x_2|$ для всех $x_1 \geq 0, x_2 \geq 0$. Тогда, если функция $N(f)$ интегрируема и ограничена, то

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n g(\lambda_i^{(n)}) = \int_{-1/2}^{1/2} g(N(f)) df,$$

где $\lambda_i^{(n)}$, $i = 1, \dots, n$ — собственные числа матрицы K_n .

Используя утверждение леммы 5.7.1, легко найти пределы соотношений (5.7.4) и тем самым получить формулу для вычисления ε -энтропии гауссовского стационарного процесса.

Теорема 5.7.1. Для гауссовского стационарного случайного процесса с дискретным временем и спектральной плотностью мощности $N(f)$, которая ограничена и интегрируема, ε -энтропия $H(\varepsilon)$ при квадратичном критерии качества вычисляется по формулам

$$\begin{aligned} H(\varepsilon) &= \frac{1}{2} \int_{-1/2}^{1/2} \log \max \left\{ 1, \frac{N(f)}{\theta} \right\} df, \\ &\quad \int_{-1/2}^{1/2} \min \{ \theta, N(f) \} df = \varepsilon. \end{aligned} \quad (5.7.10)$$

Доказательство. Введем обозначения

$$A_n(\theta) \triangleq \frac{1}{2n} \sum_{i=1}^n \log \max \left\{ 1, \frac{\lambda_i^{(n)}}{\theta} \right\}, \quad (5.7.11)$$

$$B_n(\theta) \triangleq \frac{1}{2n} \sum_{i=1}^n \min \{ \theta, \lambda_i^{(n)} \}. \quad (5.7.12)$$

Рассмотрим две функции $g_1(x) \triangleq \frac{1}{2} \log \max \left\{ 1, \frac{x}{\theta} \right\}$ и $g_2(x) \triangleq \min \{ \theta, x \}$. Нетрудно видеть, что обе эти функции удовлетворяют условиям леммы 5.7.1. Применяя лемму 5.7.1 к (5.7.11) и (5.7.12), получим

$$\begin{aligned} A(\theta) &\triangleq \lim_{n \rightarrow \infty} A_n(\theta) = \frac{1}{2} \int_{-1/2}^{1/2} \log \max \left\{ 1, \frac{N(f)}{\theta} \right\} df, \\ B(\theta) &\triangleq \lim_{n \rightarrow \infty} B_n(\theta) = \int_{-1/2}^{1/2} \min \{ \theta, N(f) \} df. \end{aligned}$$

Пусть θ_ε — корень уравнения $B(\theta) = \varepsilon$ и пусть $\varepsilon_n = B_n(\theta_\varepsilon)$. Тогда для любого $\delta > 0$ найдется n_0 такое, что при $n > n_0$, $|\varepsilon_n - \varepsilon| < \delta$, $|H_n(\varepsilon_n) - A(\theta_\varepsilon)| < \delta$. В силу непрерывности ε -энтропии $H_n(\varepsilon)$ имеем $|H_n(\varepsilon_n) - H_n(\varepsilon)| < \delta_1$, где δ_1 стремится к нулю при δ , стремящемся к нулю. Объединяя полученные неравенства, получим

$$|H_n(\varepsilon) - A(\theta_\varepsilon)| < |H_n(\varepsilon_n) - A(\theta_\varepsilon)| + \delta_1 < \delta + \delta_1 = \delta_2.$$

Так как $\delta_1 \rightarrow 0$ при $\delta \rightarrow 0$, то мы показали, что для любого числа $\delta_2 > 0$ найдется достаточно большое n такое, что $|H_n(\varepsilon) - A(\theta_\varepsilon)| < \delta_2$ и, следовательно,

$$\lim_{n \rightarrow \infty} H_n(\varepsilon) = A(\theta_\varepsilon).$$

Утверждение теоремы вытекает теперь из того, что значение $A(\theta_\varepsilon)$ определяется соотношениями (5.7.10).

Второму соотношению в (5.7.10) можно придать геометрическую трактовку с помощью рассуждения о «разливании воды». Представим себе сосуд единичной длины с крышкой, форма которой задается спектральной плотностью мощности $N(f)$ (см. рис. 5.7.1). Представим себе сосуд единичной длины с крышкой, форма которой задается спектральной плотностью мощности $N(f)$ (см. рис. 5.7.1). Представим себе сосуд единичной длины с крышкой, форма которой задается спектральной плотностью мощности $N(f)$ (см. рис. 5.7.1). Тогда жидкость объема ε после опускания крышки установится на уровне θ . Для нахождения эпсилон-энтропии нужно проинтегрировать функцию $\log N(f)/\theta$ по той области частот, в которой $N(f) \geq \theta$ (на рисунке — по интервалам $(-1/2, -f_1)$, $(-f_2, f_2)$, $(f_1, 1/2)$).

Рис. 5.7.1. Интерпретация уравнений (5.7.10).

§ 5.8*. Формулировка прямой теоремы кодирования для стационарного гауссовского источника с дискретным временем

Обратная теорема кодирования (см. § 5.3) верна для произвольных непрерывных стационарных источников и произвольного критерия качества. Следовательно, и для стационарного гауссовского источника с эпсилон-энтропией $H(\varepsilon)$ относительно квадратичного критерия качества верно утверждение о том, что для любого кода (R, d) , кодирующего источник со скоростью $R < H(\varepsilon)$, средняя ошибка больше, чем ε .

Ниже мы сформулируем прямую теорему кодирования, верную для стационарного гауссовского источника, который обладает следующим свойством: корреляционный момент K_{ij} с. в. X_i и X_j на выходе источника стремится к нулю, когда $|i - j| \rightarrow \infty$. Как отмечалось выше, это требование эквивалентно эргодичности стационарного гауссовского источника. Эргодические непрерывные источники определяются в точности так же, как и дискретные (см. определение 1.9.1).

Теорема 5.8.1. Пусть источник порождает стационарный эргодический гауссовский случайный процесс дискретного времени и имеет эпсилон-энтропию $H(\varepsilon)$ относительно квадратичного критерия качества $d(x, y) = (x - y)^2$. Тогда для произволь-

ногого положительного δ при достаточно больших n найдется (R, d) -код, кодирующий отрезки сообщений источника длины n , такой, что его скорость $R = H(\varepsilon) + \delta$ и средняя ошибка $d \leq \varepsilon$.

Доказательство этой теоремы опирается на построение высоковероятного множества реализаций на выходе стационарного гауссовского источника, аналогичного множеству точек «твердой» сферы для гауссовского источника без памяти. Идея построения такого множества подсказывается задачей 5.8.2. Однако техническая реализация этой идеи довольно сложна, и поэтому доказательство теоремы 5.8.1 опускается.

Задачи, упражнения и дополнения

5.1.1. Покажите, что для примера 5.1.1 скорость кодирования равна нулю при средней ошибке, большей или равной $1/2$. Найдите наилучшее аппроксимирующее множество при скорости кодирования $2/3$. Чему равна средняя ошибка?

5.1.2. В этой задаче мы рассмотрим пример кодирования непрерывной с. в., которое называется *квантованием*. Квантование описывается следующим образом. Пусть X — числовая ось и на ней выбраны $N+1$ чисел $x_0 < x_1 < \dots < x_N$. Пусть y_1, \dots, y_N — числа, обладающие тем свойством, что $y_i \in (x_{i-1}, x_i)$, $i = 1, \dots, N$. Интервалы $(x_0, x_1), \dots, (x_{N-1}, x_N)$ называются квантами, а числа y_1, \dots, y_N называются аппроксимирующими значениями. При квантовании каждой величине x сопоставляется аппроксимирующее значение $y_i = y(x)$, если x находится в кванте $(x_{i-1}, x_i]$. Средняя ошибка квантования

$$\bar{d} \triangleq \int_{-\infty}^{\infty} d(x, y(x)) f(x) dx,$$

где $f(x)$ — ф. п. в. рассматриваемой с. в., а $d(x, y)$ — функция, задающая вид ошибок (критерий качества). Если $d(x, y) = (x - y)^2$, то средняя ошибка \bar{d} называется среднеквадратической.

Таким образом, квантование преобразует непрерывную с. в. в дискретную: каждому значению x с. в. на выходе квантователя сопоставляется число i на выходе квантователя, где i — номер кванта, содержащего значение x . Для того чтобы восстановить значение x по номеру кванта, выбирается соответствующее аппроксимирующее значение y_i . Квантование представляет собой некоторый способ кодирования непрерывных с. в. При этом скорость кодирования, очевидно, равна $\log N$ бит/сообщение.

Будем рассматривать квадратическую функцию качества. В этом случае среднюю ошибку квантования можно представить следующим образом:

$$\bar{d} = \sum_{i=1}^N \int_{x_{i-1}}^{x_i} (x - y_i)^2 f(x) dx.$$

Можно поставить задачу об оптимальном выборе границ квантов x_0, x_1, \dots, x_N и аппроксимирующих значений y_1, \dots, y_N , при которых величина среднеквадратической ошибки \bar{d} минимальна.

а) Предположим, что границы квантов x_0, x_1, \dots, x_N некоторым образом выбраны. Покажите, что аппроксимирующие значения, которые минимизируют

среднеквадратическую ошибку квантования \bar{d} при фиксированных x_0, x_1, \dots, x_N , должны удовлетворять следующей системе уравнений:

$$y_i = \frac{1}{p_i} \int_{x_{i-1}}^{x_i} x f(x) dx, \quad p_i = \int_{x_{i-1}}^{x_i} f(x) dx, \quad i = 1, \dots, N.$$

Отсюда можно заключить, что аппроксимирующее значение y_i есть координата центра тяжести фигуры, ограниченной кривой $f(x)$ на интервале $(x_{i-1}, x_i]$, а число p_i (вероятность i -го кванта) есть площадь этой фигуры.

б) Пусть $f(x)$ всюду отлична от нуля, $x_0 = -\infty$ и $x_N = \infty$, и пусть аппроксимирующие значения выбираются в соответствии с п. а). Используя дифференцирование интеграла по параметру, покажите, что необходимое условие того, чтобы x_1, \dots, x_{N-1} минимизировали среднеквадратическую ошибку \bar{d} , состоит в том, что

$$x_i = \frac{y_i + y_{i+1}}{2}.$$

Таким образом, если границы квантов находятся на одинаковом расстоянии от ближайших аппроксимирующих значений, то квантование будет минимизировать среднеквадратическую ошибку. Такое квантование называется *оптимальным неравномерным квантованием*. В общем случае квантцы имеют неодинаковые длины.

Найдите ф. п. в., для которой при оптимальном неравномерном квантовании все квантцы, имеющие отличные от нуля вероятности, имеют одинаковые длины. Вычислите среднеквадратическую ошибку как функцию от N для этого случая.

в) Предположим, что для всех границ квантов имеет место приближенное равенство $f(x_{i-1}) \approx f(x_i)$ (т. е. либо длины квантов малы, либо функция $f(x)$ слабо меняется). Покажите, что для оптимального неравномерного квантования

$$\int_{x_{i-1}}^{x_i} (x - y_i)^2 f(x) dx \approx f(x_i) \frac{(x_i - x_{i-1})^3}{12}, \quad i = 1, \dots, N,$$

и, следовательно,

$$\bar{d} \approx \sum_{i=1}^N f(x_i) \frac{(x_i - x_{i-1})^3}{12}.$$

г) Предположим, что $f(x)$ — ф. п. в. гауссовского распределения с дисперсией D . В этом случае имеется простое соотношение, связывающее число квантов N и достижимую среднеквадратическую ошибку

$$\log N = \frac{1}{2} \log \frac{D\alpha(N)}{\bar{d}}, \quad D \gg \bar{d},$$

где $\alpha(N)$ — неубывающая функция, график которой показан на рисунке. Постройте графики зависимостей скорости кодирования от ошибки при различных значениях D . Можно ли понизить скорость кодирования при данной ошибке за счет дополнительного кодирования выхода квантователя? В каком случае этого сделать нельзя?

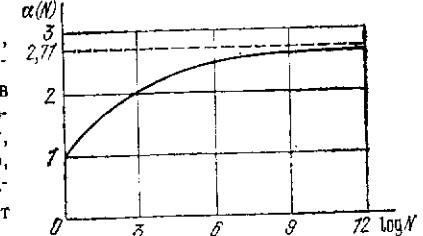


Рис. К задаче 5.1.2.

5.2.1. Пусть U_X — непрерывный источник без памяти, $d(x, y) = (x - y)^2$. и математическое ожидание с. в. X на выходе источника отлично от нуля. Покажите, что $H(\varepsilon) = 0$ для всех $\varepsilon > D$, где D — дисперсия с. в. X . Каков в этом случае универсальный аппроксимирующий элемент?

5.2.2. Пусть в условии предыдущей задачи $d(x, y) = (x - y)^{2k}$, где k — целое положительное число. Покажите, что при симметричной относительно математического ожидания ф. п. в. $f(x)$ с. в. X универсальный аппроксимирующий элемент выбирается одинаковым образом для всех $k = 1, 2, \dots$. Каким?

5.2.3. Пусть $f(x) = 1/2a$ для всех $x \in (-a, a)$ и $f(y|x) = 1/2b$ для всех $y \in (-b + x, b + x)$, $x \in (-a, a)$, $b \ll a$.

а) Покажите, что

$$f(y) = \begin{cases} 1/2a & \text{при } -a + b < y \leq a - b, \\ (a + b - y)/4ab & \text{при } a - b < y \leq a + b, \\ (a + b + y)/4ab & \text{при } -a - b < y < -a + b. \end{cases}$$

б) Покажите, что

$$H_0(Y|X) = \log 2b, \quad H_0(Y) = \log 2a + \frac{b}{2a} \log e$$

и, следовательно,

$$I(X; Y) = \log \frac{a}{b} + \frac{b}{2a} \log e.$$

в) Покажите, что средняя ошибка \bar{d} при критерии качества $d(x, y) = |x - y|$ равна $b/2$. Указание: рассмотрите сначала среднюю ошибку $\bar{d}(x)$ при фиксированном x и покажите, что $\bar{d}(x) = b/2$ для всех x .

г) Покажите, что для источника без памяти с ф. п. в. $f(x)$, определенной выше, и критерием качества $d(x, y) = |x - y|$ эпсилон-энтропия

$$H(\varepsilon) \leq \log \frac{a}{2\varepsilon} + \frac{\varepsilon}{a} \log e.$$

д) Покажите, что для всех $\varepsilon \geq a/2$ существует универсальный аппроксимирующий элемент и поэтому

$$H(\varepsilon) = \begin{cases} \leq \log \frac{a}{2\varepsilon} + \frac{\varepsilon}{a} \log e & \text{при } \varepsilon < \frac{a}{2}, \\ 0 & \text{при } \varepsilon \geq \frac{a}{2}. \end{cases}$$

е) Уточните оценку для эпсилон-энтропии источника, указанного в п. г), рассмотрев квантование случайной величины на выходе такого источника.

5.2.4. В этой задаче будет рассмотрено доказательство теоремы 5.2.3 и исследовано поведение $H_n(\varepsilon)$ по $n = 1, 2, \dots$ для стационарных источников, которое осталось вне основного текста § 5.2.

Предположим, что $H_m(\varepsilon)$ достигается на ф. п. в. $f_m(y|x)$, $x \in X^m$, $y \in Y^m$, а $H_n(\varepsilon)$ достигается на ф. п. в. $f_n(y|x)$, $x \in X^n$, $y \in Y^n$. Это означает, что для данного стационарного источника

$$\frac{1}{m} I(X^m; Y^m) = H_m(\varepsilon),$$

$$\frac{1}{n} I(X^n; Y^n) = H_n(\varepsilon).$$

Рассмотрим теперь последовательности длины $m + n$ и определим ф. п. в. $f_{m+n}(y|x)$ на таких последовательностях следующим образом:

$$f_{m+n}(y|x) \triangleq f_m(y'|x') \cdot f_n(y''|x''),$$

где $x = (x', x'') \in X^{m+n}$, $y = (y', y'') \in Y^{m+n}$. Пусть $I(X^{m+n}; Y^{m+n})$ — средняя взаимная информация, вычисленная для ф. п. в. $f_{m+n}(y|x)$. Тогда $I(X^{m+n}; Y^{m+n}) = H_0(Y^{m+n}) - H_0(Y^{m+n}|X^{m+n}) = H_0(Y^{m+n}) - H_0(Y^m|X^m) - H_0(Y^n|X^n)$.

а) Основнуйте предыдущее равенство. Покажите, что

$$I(X^{m+n}; Y^{m+n}) \leq I(X^m; Y^m) + I(X^n; Y^n).$$

б) Покажите, что

$$H_{m+n}(\varepsilon) \leq \frac{m}{m+n} H_m(\varepsilon) + \frac{n}{m+n} H_n(\varepsilon).$$

Указание: проверьте, что ф. п. в. $f_{m+n}(y|x)$ принадлежит множеству $\Phi_{m+n}(\varepsilon)$.

Покажите, что

$$H_{m+n}(\varepsilon) \leq H_n(\varepsilon)$$

для любых целых m и n .

в) Так как $H(\varepsilon) = \inf_n H_n(\varepsilon)$, то для любого $\delta > 0$ найдется такое n , что $H_n(\varepsilon) \leq H(\varepsilon) + \delta$. Покажите, используя результаты п. б), что для достаточно больших N , $N > n$, имеет место неравенство $H_N(\varepsilon) \leq H(\varepsilon) + 2\delta$. Так как δ может быть взято сколь угодно малым, то имеет место утверждение теоремы 5.2.3. Указание: воспользуйтесь представлением $N = m \cdot n + j$, где $0 \leq j < n$.

5.3.1. Пусть N — количество квантов, при которых обеспечивается квантование гауссовской с. в. со среднеквадратической ошибкой ε . Покажите, что $\log N \geq \frac{1}{2} \log \max\{1, D/\varepsilon\}$, где D — дисперсия с. в.

5.4.1. Пусть с. в. X, Y, Z связаны соотношением $X = Y + Z$. Обозначим через $f_X(y|x)$ условную ф. п. в. с. в. X при фиксированном значении y . Обозначим через $f_Z(z|y)$ такую же ф. п. в. для с. в. Z .

а) Покажите, что $f_Z(z|y) = f_X(z+y|y)$.

б) С помощью прямого вычисления относительных энтропий убедитесь в том, что $H_0(Z|y) = H_0(X|y)$.

в) Найдите $I(X; Z|y)$.

5.4.2. Пусть X, Y, Z — гауссовые с. в., причем $X = Y + Z$. Пусть $f(x, y)$ — ф. п. в. двумерного гауссова распределения вероятностей с. в. X, Y с параметрами $m_X = m_Y = 0$, $D_X = D$, $D_Y = D - \varepsilon$, $\rho_{XY} = ((D - \varepsilon)/D)^{1/2}$. Покажите, что $m_Z = 0$, $D_Z = \varepsilon$ и с. в. Y и Z некоррелированы, и следовательно независимы.

5.4.3. Покажите, что эпсилон-энтропия относительно квадратического критерия качества любой с. в. с относительной энтропией $H_0(X)$ удовлетворяет неравенству

$$H(\varepsilon) \geq H_0(X) - \frac{1}{2} \log 2\pi e \varepsilon.$$

5.5.1. n -мерным шаром радиуса R называется совокупность векторов $\mathbf{x} = (x_1, \dots, x_n)$, таких, что $\sum_{i=1}^n (x_i - y_i)^2 \leq R^2$, где $\mathbf{y} = (y_1, \dots, y_n)$ — фиксирован-

ный вектор, называемый *центром шара*. Другими словами, n -мерный шар — это такое множество n -мерных векторов, расстояние которых от вектора \mathbf{y} (центра шара) не больше чем R . Поверхность n -мерного шара, т. е. такое множество n -мерных векторов, для которых в предыдущем неравенстве имеет место строгое равенство, называется *n -мерной сферой радиуса R с центром в \mathbf{y}* .

Величина

$$V_n(R) \triangleq \int \dots \int dx_1 \dots dx_n = \int \dots \int dx_1 \dots dx_n \\ \sum_{i=1}^n (x_i - y_i)^2 \leq R^2 \quad \sum_{i=1}^n x_i^2 \leq R^2$$

называется *объемом n -мерного шара*. Последнее равенство получается в результате замены переменных. Оно показывает, что объем шара зависит только от его радиуса, но не от выбора центра. Площадь n -мерной сферы может быть определена с помощью следующего предельного соотношения:

$$S_n(R) \triangleq \lim_{\Delta R \rightarrow 0} \frac{V_n(R + \Delta R) - V_n(R)}{\Delta R} = \frac{dV_n(R)}{dR}.$$

Покажите, что

$$V_n(R) = g_n R^n, \quad S_n(R) = k_n R^{n-1},$$

где $g_n \triangleq V_n(1)$ — объем n -мерного шара радиуса единица и $k_n = nV_n(1)$.

5.5.2. В этом пункте мы рассмотрим задачу вычисления величины $V_n(1)$. Ее можно найти непосредственно, выполняя интегрирование, однако более простым является косвенный метод, рассматриваемый ниже.

Пусть $\mathbf{X} = (X_1, \dots, X_n)$ — n -мерный случайный вектор, образованный независимыми гауссовскими с. в. X_1, \dots, X_n , каждая из которых имеет нулевое среднее и единичную дисперсию. Совместная ф. п. в. этих с. в. равна

$$f(x_1, \dots, x_n) = (2\pi)^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n x_i^2 \right\}.$$

Обозначим через $w(z)$ ф. п. в. случайной величины $Z \triangleq \left(\sum_{i=1}^n X_i^2 \right)^{1/2}$. Другими словами, $w(z)$ — ф. п. в. длины случайного вектора \mathbf{X} . Вероятность того, что вектор \mathbf{X} попадает в область, заключенную между двумя концентрическими сферами с радиусами $R + \Delta R$ и R и центрами в начале координат, будет равна

$$\Pr(R \leq Z \leq R + \Delta R) = w(R) \Delta R$$

при условии, что ΔR достаточно мало. Эта же вероятность может быть найдена как произведение объема рассматриваемой области на значение функции $f(x_1, \dots, x_n)$, где (x_1, \dots, x_n) принадлежит этой области, т. е.

$$\Pr(R \leq Z \leq R + \Delta R) = (2\pi)^{-n/2} \exp \left\{ -\frac{R^2}{2} \right\} [V_n(R + \Delta R) - V_n(R)] = \\ = (2\pi)^{-n/2} \exp \left\{ -\frac{R^2}{2} \right\} S_n(R) \Delta R.$$

Учитывая, что $S_n(R) = nV_n(1)R^{n-1}$, получим

$$w(R) = V_n(1) \frac{nR^{n-1}}{(2\pi)^{n/2}} \exp \left\{ -\frac{R^2}{2} \right\}.$$

Интегрируя обе части последнего соотношения по всем R из интервала $[0, \infty]$, получим

$$1 = V_n(1) \frac{n}{(2\pi)^{n/2}} \int_0^\infty R^{n-1} \exp \left\{ -\frac{R^2}{2} \right\} dR.$$

Рассмотрим отдельно два случая: n — четное и n — нечетное число. Пусть сначала $n = 2m$ — четное число. В результате замены переменных и последовательного интегрирования по частям получим

$$\int_0^\infty R^{n-1} \exp \left\{ -\frac{R^2}{2} \right\} dR = 2^{m-1} \int_0^\infty R^{m-1} e^{-R} dR = (m-1)! 2^{m-1}.$$

Пусть теперь $n = 2m+1$ — нечетное число. Тогда величина

$$\frac{1}{\sqrt{2\pi}} \int_0^\infty R^{n-1} \exp \left\{ -\frac{R^2}{2} \right\} dR = \frac{1}{2} \mu_{n-1}$$

представляет собой половину центрального момента μ_{2m} порядка $2m$ гауссовой с. в., имеющей единичную дисперсию. Известно (это нетрудно проверить с помощью производящей функции моментов), что

$$\mu_{2m} = \frac{(2m)!}{2^m \cdot m!}.$$

Таким образом,

$$V_n(1) = \begin{cases} \frac{\pi^m}{m!} & \text{при } n = 2m, \\ \frac{\pi^m m! 2^{2m+1}}{(2m+1)!} & \text{при } n = 2m+1. \end{cases}$$

В частности, $V_1(1) = 2$, $V_2(1) = \pi$, $V_3(1) = 4\pi/3$.

Используя гамма-функцию $\Gamma(x)$, формулу для объема n -мерного единичного шара $V_n(1)$ можно представить в следующем виде, общем для четных и нечетных n :

$$V_n(1) = \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2} + 1\right)}.$$

С помощью приближения Стирлинга для факториала можно показать, что

$$\lim_{n \rightarrow \infty} \frac{V_{n-1}(1)}{V_n(1)} = \sqrt{\frac{n}{2\pi}}.$$

5.5.3. Рассмотрим пересечение n -мерного шара и плоскости. Пересечение в этом случае представляет собой $(n-1)$ -мерный шар (покажите это). Обозначим через r радиус пересечения. Пусть W_n — круговая область, вырезаемая плоскостью на поверхности n -мерного шара. Покажите, что

$$|W_n| \geq V_{n-1}(r),$$

где $|W_n|$ — площадь области W_n . Эта задача имеет ясный геометрический смысл в трехмерном ($n = 3$) случае и выражает тот очевидный факт, что площадь части поверхности шара не меньше, чем площадь круга, на который эта поверхность опирается.

5.5.4. Множество V_n векторов $\mathbf{x} = (x_1, \dots, x_n)$, лежащих на «твёрдой» сфере радиуса \sqrt{Dn} , т. е. таких, что

$$\left| \frac{1}{n} \sum_{i=1}^n x_i^2 - D \right| < \delta,$$

где δ — произвольное положительное число, образует высоковероятное множество для гауссовского источника без памяти. Покажите, что следующие два множества совпадают:

$$V_n \triangleq \left\{ \mathbf{x}: \left| \frac{1}{n} \sum_{i=1}^n x_i^2 - D \right| < \delta \right\}$$

и

$$T_n \triangleq \left\{ \mathbf{x}: \left| -\frac{1}{n} \log f(\mathbf{x}) - H_0(X) \right| < \frac{\delta}{2D} \log e \right\},$$

где $H_0(X)$ — относительная энтропия гауссовской с. в. с дисперсией D , а $f(\mathbf{x})$ — ф. п. в. гауссовского случайного вектора с независимыми одинаково распределенными компонентами.

Этот результат подсказывает способ определения высоковероятного множества для произвольного негауссовского источника.

5.5.5. Пусть эпсилон-энтропия гауссовского источника без памяти при квадратическом критерии качества и величине средней ошибки ε равна 0,5. Пусть длина n кодируемых сообщений равна 100.

а) Оцените снизу наименьшее количество кодовых слов, которые должны иметься в коде для того, чтобы средняя ошибка кодирования источника была не меньше ε .

б) Каким числом кодовых слов нужно задаться, для того чтобы половина всех кодов была хорошей в смысле обсуждения в п. 5.5.5.

5.6.1. Пусть (X_1, X_2, X_3, X_4) — гауссовский случайный вектор, компоненты которого имеют дисперсии $D_1 = 2$, $D_2 = 1$, $D_3 = D_4 = \frac{1}{2}$. Решите уравнение (5.6.19) и найдите θ для $\varepsilon = 0,6$. Убедитесь в том, что $H_4(0,6) = 0,309$.

5.6.2. Пусть $n = 2$ и X_1, X_2 — система из двух гауссовских с. в. с корреляционной матрицей

$$K = \begin{bmatrix} 1 & 0,5 \\ 0,5 & 1 \end{bmatrix}.$$

а) Покажите, что собственные числа этой матрицы равны $\lambda_1 = 1,5$, $\lambda_2 = 0,5$. Указание: воспользуйтесь тем, что λ_1, λ_2 — корни характеристического уравнения

$$\det(K - \lambda I) = 0,$$

где I — единичная матрица и λ — переменная.

Покажите, что $(1/\sqrt{2}, 1/\sqrt{2}), (1/\sqrt{2}, -1/\sqrt{2})$ — собственные векторы матрицы K .

б) Покажите, что при всех $\varepsilon \leqslant 0,5$ имеем $\theta = \varepsilon$ и

$$H_2(\varepsilon) = \frac{1}{2} \log \frac{0,865}{\varepsilon}.$$

5.8.1. Пусть $\{X_i\}$, $i = 0, \pm 1, \pm 2, \dots$, — стационарный случайный процесс дискретного времени. Если для любого $k = 1, 2, \dots$ и любой функции $\varphi(x_1, \dots, x_k)$, определенной на множестве X^k , X — числовая ось, последовательность с. в.

$$Z^{(i)} \triangleq \varphi(X^{(i+1)}, \dots, X^{(i+k)}), \quad i = 1, 2, \dots,$$

обладает тем свойством, что для любых положительных γ, δ и достаточно больших n

$$\Pr \left(\left| \frac{1}{n} \sum_{i=1}^n Z^{(i)} - MZ^{(i)} \right| \geqslant \gamma \right) < \delta,$$

то случайный процесс $\{X_i\}$ называется эргодическим.

Строго говоря, для каждого k функция $\varphi(x_1, \dots, x_k)$ должна удовлетворять некоторым дополнительным ограничениям, главное из которых состоит в том, что $M\varphi^2(X^{(i+1)}, \dots, X^{(i+k)}) < \infty$.

Очевидно, что любой стационарный процесс, в котором с. в. X_t, X_j статистически независимы при всех $i \neq j$, является эргодическим. Доказательство этого утверждения в точности повторяет доказательство теоремы 1.9.1.

Предположим, что для стационарного процесса $\{X_i\}$ выполнено следующее условие: с. в. X_t, X_{t+s} статистически независимы при всех $s \geqslant s_0$ и при всех i . Покажите, что такой процесс эргодичен. Указание: воспользуйтесь методом теоремы 1.9.1.

Коэффициент корреляции K_{XY} для независимых с. в. X и Y равен нулю. Если X и Y — гауссовские с. в., то при $K_{XY} = 0$ они независимы. Поэтому, если для стационарного гауссовского процесса $\{X_i\}$ выполнено следующее условие: коэффициент корреляции с. в. X_i и X_{i+s} равен нулю для всех $s \geqslant s_0$ и всех i , то процесс эргодичен.

Приведенные рассуждения используются для обоснования того, что гауссовский стационарный процесс, для которого $K_{ij} \rightarrow 0$ при $|i-j| \rightarrow \infty$, является эргодическим. Трудность состоит в том, чтобы доказать, что достаточно малое значение коэффициента корреляции обеспечивает в определенном смысле малую статистическую зависимость.

5.8.2. а) Покажите, что для эргодического процесса $\{X_i\}$ при $MX^4 < \infty$ выполняется следующее утверждение: для любых положительных γ, δ и достаточно большом n

$$\Pr \left(\left| \frac{1}{n} \sum_{i=1}^n X_i^2 - D \right| > \gamma \right) < \delta$$

и, следовательно, имеет место «затвердование» n -мерной сферы.

б) Пусть (R, d_n) — последовательность кодов длины n , $n = 1, 2, \dots$, со скоростью R для кодирования гауссовского источника без памяти с дисперсией D относительно квадратического критерия качества и $\lim_{n \rightarrow \infty} d_n = e$, где $e, e < D$, — корень уравнения

$$R = \frac{1}{2} \log \frac{D}{e}.$$

Пусть эти же коды применяются для кодирования некоторого эргодического гауссовского источника, дисперсия которого также равна D , и пусть d'_n — среднеквадратическая ошибка при кодировании этого источника кодом длины n . Покажите, что

$$\lim_{n \rightarrow \infty} d'_n < e.$$

КРАТКИЙ ИСТОРИЧЕСКИЙ КОММЕНТАРИЙ И ЛИТЕРАТУРА

Постановка задачи кодирования источников при заданном критерии качества принадлежит К. Шенону [7]. Он впервые вычислил величину эпсилон-энтропии для гауссовского источника без памяти. Термин эпсилон-энтропия принадлежит А. Н. Колмогорову [4], который впервые вычислил величину эпсилон-энтропии для гауссовского случайного процесса. Прямая теорема кодирования с заданным критерием качества для дискретных источников была доказана К. Шенном [8]. Общая прямая теорема кодирования для источников без памяти была доказана Р. Л. Добрушиным [3]. Обобщение этой теоремы для произвольных эргодических источников получили М. С. Пинскер [6] и К. Мартон [5].

1. Галлагер (Gallager R.). Information Theory and Reliable Communication. — New York: Wiley, 1968. [Русский перевод: Галлагер Р. Теория информации и надежная связь. — М.: Советское радио, 1974.]
2. Гренандер и Сегё (Grenander U., Szegö G.), Toeplitz Forms and their Applications. — Berkeley, Calif., 1958. [Русский перевод: Гренандер У. и Сегё Г. Тэплицевы формы и их приложения. — М.: ИЛ, 1961.]
3. Добрушин Р. Л. Общая формулировка основной теоремы Шенона в теории информации. — Успехи матем. наук, 1959, 14, 6.
4. Колмогоров А. Н. Теория передачи информации. — В сб.: Сессия АН СССР по научн. пробл. автоматизации произв. Пленарные заседания. — М.: Изд. АН СССР, 1956.
5. Мартон К. Информация и информационная устойчивость эргодических источников. Проблемы передачи информации, 1972, т. 8, № 3.
6. Пинскер М. С. Источники сообщений (а), Гауссовские источники (б). Проблемы передачи информации, 1963, т. 14, 5—20 (а), 59—100 (б).
7. Шеннон (Shannon C. E.). A Mathematical Theory of Communications, Bell Syst. Tech. J., 1948, 27, 379—423 (Part I), 623—656 (Part II). [Русский перевод: Шенон К. Математическая теория связи. — В сб.: Работы по теории информации и кибернетике. — М.: ИЛ, 1963.]
8. Шенон (Shannon C. E.). Coding Theorem for a Discrete Source with Fidelity Criterion. — IRE Nat. Conv. Rec., Part 4, 1959, 142—163. [Русский перевод: Шенон К. Теоремы кодирования для дискретного источника при заданном критерии точности. — В сб.: Работы по теории информации и кибернетике. — М.: ИЛ, 1963.]

Глава 6*

КОДИРОВАНИЕ В СИСТЕМАХ С МНОГИМИ ПОЛЬЗОВАТЕЛЯМИ

Вопросы кодирования источников сообщений и кодирования в каналах связи, рассмотренные в предыдущих главах, представляют собой вопросы традиционной или классической теории информации и соответствуют схеме передачи, изображенной на рис. 3.1.1. Отличительной особенностью такой схемы является наличие одного источника и одного получателя. Наблюдающееся в последнее время развитие сетей связи, реализующих обслуживание многих абонентов, поставило перед теорией информации новые задачи, а именно, задачи кодирования в системах с многими источниками, многими получателями и многими каналами (в системах с многими пользователями). Общая схема системы передачи с многими пользователями приведена на рис. 6.0.1. В этой системе имеются: множество источников U_1, U_2, \dots, U_m ; множество передающих устройств ПРД1, ПРД2, ..., ПРД k ; множество принимающих устройств ПРМ1, ПРМ2, ..., ПРМ l и сеть каналов, соединяющая передающие и приемные устройства. В общем случае сообщения каждого источника доступны некоторому подмножеству передающих устройств и предназначены некоторому подмножеству приемных устройств.

Как и в классическом случае, с точки зрения теории информации основной характеристикой системы передачи с большим числом пользователей является набор скоростей передачи информации. Каждый элемент этого набора R_{ij} представляет собой скорость передачи информации от источника с номером i приемнику с номером j . Количество элементов в наборе скоростей, характеризующем систему, зависит как от числа источников и приемников, так и от заданной адресации, т. е. от указания того, сообщения каких источников должны быть переданы данному получателю. Основной теоретико-информационной задачей при анализе системы передачи с многими пользователями является описание множества всех таких наборов скоростей, при которых возможна передача сообщений источникам получателям в соответствии с заданной адресацией. Очевидно, что решение этой задачи существенным образом зависит от конкретной структуры системы.

Большое многообразие возможных структур не позволяет дать единый метод определения множества всех допустимых скоростей для произвольной системы с многими пользователями.

Путь, по которому сейчас следует теория информации при решении этой задачи, состоит в выделении основных ситуаций, типичных для многих систем с большим числом пользователей, и решении задачи кодирования для каждой из таких ситуаций. В данной главе будут рассмотрены некоторые из этих задач, позволяющие, по мнению авторов, достаточно наглядно показать те новые аспекты, которые возникают при кодировании в системах с большим числом пользователей.

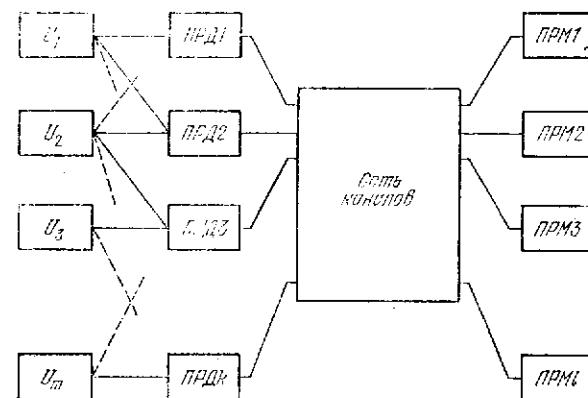


Рис. 6.0.1. Структура системы с многими пользователями.

Мы рассмотрим четыре задачи, из которых первые две относятся к кодированию многих источников, а вторые две — к кодированию в каналах с многими пользователями. Слово «многие» не следует понимать буквально. Специфика проявляется уже тогда, когда количество источников сообщений или пользователей канала связи равно двум. Обобщение на случай более двух компонент часто не является трудной задачей, хотя и может быть технически сложной.

Изложение материала этой главы значительно отличается от изложения предыдущих глав и рассчитано на читателя, хорошо овладевшего материалом предыдущих глав и выработавшего у себя некоторую теоретико-информационную интуицию. Это обусловлено тем, что задачи кодирования в системах с большим числом пользователей являются достаточно трудными и, как правило, требуют для своего решения всего арсенала теоретико-информационных средств. Вместе с тем, круг вопросов для систем с большим числом пользователей представляет значительный интерес для специалистов как в области теории, так и практики связи; эти вопросы соответствуют духу времени, они интенсивно изучаются и во многом определяют тенденции развития теории информации.

§ 6.1. Кодирование зависимых источников

6.1.1. Постановка задачи. Начнем со следующего примера. Предположим, что имеется ряд рассредоточенных в пространстве датчиков метеорологических данных, которые передают эти данные в центр обработки для составления картины погоды. Датчики можно рассматривать как источники сообщений, причем сообщения отдельных датчиков, как правило, являются зависимыми случайными величинами: сообщение на выходе одного датчика в значительной степени определяется сообщениями на выходах других, поскольку погодные условия в близких районах зависят. Из-за наличия такой зависимости сообщения всех датчиков в совокупности избыточны, и в действительности для того, чтобы воспроизвести всю картину погоды, от каждого датчика можно было бы брать только часть данных. Сокращение общего объема передаваемых данных является выгодным с многих точек зрения. Поэтому естественным является вопрос о том, какая часть всех данных достаточна.

Датчики рассредоточены в пространстве и каждый из них снабжен кодером, работающим независимо от остальных кодеров. Кодеры устраниют избыточную информацию, содержащуюся в совместных измерениях системы датчиков. Интуитивно ясно, что возможно по-разному кодировать сообщения различных датчиков, затрачивая различное количество двоичных символов на сообщение.

Например, возможно кодировать сообщения первого датчика, используя большое количество двоичных символов на сообщение, а второго — малое; возможно поступить наоборот. И в том, и в другом случае общая картина погоды может быть восстановлена. Другими словами, возможно независимое кодирование системы источников с различным набором скоростей кодирования (R_1, \dots, R_N) для N независимо работающих кодеров. С точки зрения теории информации задача заключается в определении всех таких наборов скоростей, при которых возможно восстановление сообщений всех датчиков системы.

Ниже мы рассмотрим наиболее простой случай независимого кодирования зависимых источников, а именно, будет предполагаться, что имеются два дискретных источника без памяти и требуется восстановление с произвольно малой вероятностью ошибки. Эта задача аналогична задаче кодирования дискретных источников равномерными кодами, которая была рассмотрена в первой главе.

Пусть U_x, U_y — два зависимых дискретных источника и X, Y — ансамбли их сообщений с совместным распределением вероятностей $p(x, y)$. Будем рассматривать такие источники U_x, U_y , что распределение вероятностей на парах последователь-

ностей сообщений (x, y) , $x \in X^n$, $y \in Y^n$, задается соотношением

$$p(x, y) = \prod_{i=1}^n p(x^{(i)}, y^{(i)}). \quad (6.1.1)$$

Другими словами, предполагается, что источники зависимы, но пары сообщений $(x^{(1)}, y^{(1)})$, $(x^{(2)}, y^{(2)})$, ... статистически независимы и одинаково распределены. Пару источников U_X , U_Y , для которых выполнено соотношение (6.1.1), будем называть *парой зависимых источников без памяти*.

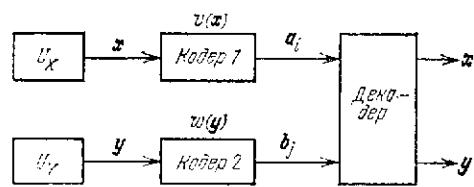


Рис. 6.1.1. Независимое кодирование зависимых источников.

множеств X^n и Y^n на множества целых чисел $V = \{1, 2, \dots, M_X\}$, $W = \{1, 2, \dots, M_Y\}$ соответственно. Каждый элемент $i \in V$ может быть взаимно однозначно представлен двоичной последовательностью a_i длины $\log M_X$ (двоичной записью номера i в множестве V). Аналогично каждый элемент $j \in W$ может быть представлен двоичной последовательностью b_j длины $\log M_Y$. Кодирование источников U_X , U_Y задается отображениями $v(x)$, $w(y)$ соответственно. Если на выходе источника U_X появляется последовательность x , а на выходе источника U_Y — последовательность y , то на выходе первого кодера появляется последовательность a_i , $i = v(x)$, а на выходе второго — b_j , $j = w(y)$ (см. рис. 6.1.1).

Восстановление пары (x, y) , появившейся на выходах источников U_X , U_Y , осуществляется декодером, работа которого задается отображением $g(i, j)$ пары чисел $i \in V$, $j \in W$ в множество $X^n Y^n$. При декодировании возможно неправильное восстановление. Вероятность ошибки λ определяется как вероятность того, что хотя бы один элемент пары будет восстановлен неверно:

$$\lambda \triangleq \Pr(g(i, j) \neq (x, y)) = \sum_{\substack{(x, y): \\ v(x) = i, w(y) = j \\ g(i, j) \neq (x, y)}} p(x, y).$$

Будем говорить, что задан код $G_n(R_X, R_Y)$ с длиной n для независимого кодирования пары дискретных источников, если кодируются последовательности сообщений длины n и заданы три

отображения $v(x)$, $w(y)$ (кодирование) и $g(i, j)$ (декодирование). Величины

$$R_X \triangleq \frac{\log M_X}{n}, \quad R_Y \triangleq \frac{\log M_Y}{n}$$

называются при этом скоростями кодирования источников U_X и U_Y соответственно.

Определение 6.1.1. Пара чисел (R_X, R_Y) , $R_X \geq 0$, $R_Y \geq 0$, называется *допустимой парой скоростей при независимом кодировании двух дискретных источников*, если для любых $\epsilon > 0$, $\delta > 0$ найдется $n_0(\epsilon, \delta)$ такое, что для любого $n > n_0$ найдется код $G_n(R_X, R_Y) = \{v, w, g\}$ со скоростями кодирования $\tilde{R}_X = R_X + \delta$, $\tilde{R}_Y = R_Y + \delta$, кодирующими отображениями $v(x)$, $w(y)$ и декодирующими отображением $g(i, j)$, для которого $\lambda \leq \epsilon$.

Очевидно, что пара скоростей $R_X = H(X)$, $R_Y = H(Y)$, где $H(X)$ и $H(Y)$ — энтропии ансамблей X и Y соответственно, допустима при независимом кодировании пары зависимых дискретных источников без памяти. При этом, как показано в гл. 1, отображения $v(x)$, $w(y)$ задают взаимно однозначное кодирование только для элементов высоковероятных множеств источников U_X , U_Y . С другой стороны, если бы каждому из кодеров 1 и 2 были доступны сообщения на выходах обоих источников, то кодирование можно было бы рассматривать как кодирование дискретного источника без памяти с ансамблем сообщений $\{XY, p(x, y)\}$. При таком зависимом кодировании была бы допустимой любая такая пара скоростей R_X, R_Y , что $R_X + R_Y = H(XY)$. Так как $H(XY) < H(X) + H(Y)$, причем строгое неравенство является следствием зависимости источников, то во втором случае можно было бы достичь меньшей суммарной скорости, чем в первом. Замечательный факт, который будет обоснован ниже, состоит в том, что независимое кодирование не приводит к увеличению минимальной допустимой суммарной скорости кодирования.

Множество всех допустимых пар скоростей (R_X, R_Y) при независимом кодировании пары источников будем обозначать через \mathfrak{J} . Основной задачей настоящего раздела является характеристизация области \mathfrak{J} .

Следующая лемма устанавливает выпукłość множества \mathfrak{J} . В этой лемме вводится один из важнейших методов кодирования для систем с многими пользователями — *метод разделения времени*.

Лемма 6.1.1. Пусть $(R_X, R_Y) \in \mathfrak{J}$, $(R'_X, R'_Y) \in \mathfrak{J}$ и α — произвольное число из интервала $[0, 1]$, тогда пара скоростей $R_X = \alpha R'_X + (1 - \alpha) R'_X$, $R_Y = \alpha R'_Y + (1 - \alpha) R'_Y$ допустима.

Доказательство. Предположим, что $\alpha = l/m$, где l и m — целые числа. Пусть фиксированы $\epsilon > 0$, $\delta > 0$ и пусть

$G'_n(\tilde{R}'_X, \tilde{R}'_Y)$, $G''_n(\tilde{R}''_X, \tilde{R}''_Y)$ — коды с длиной n и скоростями $\tilde{R}'_X = R'_X + \delta$, $\tilde{R}'_Y = R'_Y + \delta$, $\tilde{R}''_X = R''_X + \delta$, $\tilde{R}''_Y = R''_Y + \delta$, для которых вероятности ошибок удовлетворяют неравенствам: $\lambda' < \varepsilon/m$, $\lambda'' < \varepsilon/m$ соответственно. Будем кодировать последовательности сообщений длины $N = nm$ на выходах источников U_X и U_Y следующим образом. Первые l блоков длины n будем кодировать с помощью кода G'_n , а остальные $m - l$ блоков — с помощью кода G''_n . При этом скорости кодирования для источников U_X и U_Y будут равными:

$$\begin{aligned} R_X &= \alpha R'_X + (1 - \alpha) R''_X, \\ R_Y &= \alpha R'_Y + (1 - \alpha) R''_Y, \end{aligned}$$

а вероятность ошибки на блок длины N будет удовлетворять неравенствам

$$\lambda < l\lambda' + (m - l)\lambda'' < \varepsilon.$$

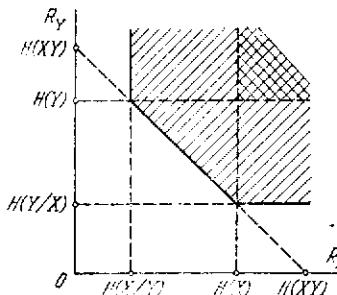
Так как ε и δ произвольны, а n , l и m могут быть выбраны сколь угодно большими, то это и доказывает утверждение леммы.

Использование части времени для одного кода, а остальной части для другого, называется *разделением времени*. Применение разделения времени дает целое семейство допустимых пар скоростей, соответствующих различному выбору параметра α . Геометрически это семейство можно представить как множество точек прямой, соединяющей в плоскости (R_X, R_Y) точки с координатами (R'_X, R'_Y) , (R''_X, R''_Y) . Поэтому для рассматриваемой задачи граница области допустимых пар скоростей представляет собой выпуклую вниз кривую, возможно, составленную из отрезков прямых.

Рис. 6.1.2. Область допустимых пар скоростей при независимом кодировании зависимых источников.

Ниже будет показано, что область допустимых пар скоростей при независимом кодировании представляет собой заштрихованную область на рис. 6.1.2. Область с двойной штриховкой представляет собой область допустимых пар скоростей при кодировании, не учитывающем зависимость источников. Область выше прямой $R_X + R_Y = H(XY)$ — область допустимых пар скоростей в случае, когда обоим кодерам доступны сообщения обоих источников (полностью зависимое кодирование).

6.1.2. Обратная теорема кодирования. В этом пункте мы докажем обратную теорему кодирования для независимого кодирования пары зависимых источников. Иными словами, мы укажем условия, при выполнении которых пара (R_X, R_Y) не является допустимой.



Прежде всего отметим, что недопустимы все пары скоростей (R_X, R_Y) , которые удовлетворяют неравенству $R_X + R_Y < H(XY)$. Это следует из обратной теоремы кодирования дискретных источников без памяти для случая источников с ансамблем сообщений $\{XY, p(x, y)\}$. Кроме этого ограничения, имеются также ограничения на минимально возможные значения R_X и R_Y . Эти ограничения могут быть пояснены с помощью следующих рассуждений.

Пусть последовательность сообщений $\mathbf{y} \in Y^n$ источника U_Y известна как кодеру 1, так и декодеру. Для типичной последовательности \mathbf{y} $H(X^n | \mathbf{y}) \approx H(X^n | Y^n) = nH(X | Y)$. При этом высоковероятное множество сообщений источника U_X будет состоять примерно из $2^{nH(X | Y)}$ последовательностей. Следовательно, скорость кодирования для источника U_X не может быть меньшей, чем $H(X | Y)$. Аналогичные рассуждения приводят к тому, что скорость кодирования для источника U_Y не может быть меньшей, чем $H(Y | X)$. Точное доказательство этих утверждений, основанное на использовании неравенства Фано, дается в следующей теореме.

Теорема 6.1.1 (Обратная теорема кодирования). Пусть U_X , U_Y — пара зависимых дискретных источников без памяти, $\{XY, p(x, y)\}$ — ансамбль сообщений для этой пары источников и $H(X, Y)$, $H(X | Y)$, $H(Y | X)$ — определяемые распределением $p(x, y)$ энтропии. Для любой пары чисел (R_X, R_Y) , удовлетворяющей хотя бы одному из следующих трех неравенств:

$$\begin{aligned} R_X + R_Y &< H(XY); \quad R_X < H(X | Y); \\ R_Y &< H(Y | X), \end{aligned} \tag{6.1.2}$$

найдется число $\varepsilon > 0$ такое, что для любого n и для любого кода $G_n(R_X, R_Y)$ вероятность ошибки удовлетворяет неравенству $\lambda \geq \varepsilon$.

Доказательство. Как уже отмечалось выше, при выполнении условия $R_X + R_Y < H(XY)$ утверждение теоремы следует из обратной теоремы кодирования для дискретного источника без памяти с ансамблем сообщений $\{XY, p(x, y)\}$. Покажем теперь справедливость теоремы для всех таких пар (R_X, R_Y) , что $R_X < H(X | Y)$.

Обозначим через \hat{X}^n множество всех последовательностей $\mathbf{x} \in X^n$, которые могут появиться на выходе декодера, $\hat{X}_n \subseteq X^n$. Заметим, что распределение вероятностей на $X^n Y^n$, кодирование $v(\mathbf{x})$ и декодирование $g(i, j)$ задают распределение вероятностей на произведении множеств $X^n Y^n V \hat{X}_n$. При этом имеют место

следующие соотношения:

$$\begin{aligned} R_X &\triangleq \frac{\log M_X}{n} \geq \frac{1}{n} H(V|Y^n) \geq \frac{1}{n} I(V; X^n|Y^n) \geq \\ &\geq \frac{1}{n} I(\hat{X}_n; X^n|Y^n) = \frac{1}{n} [H(X^n|Y^n) - H(\hat{X}_n|X^nY^n)] = \\ &= H(X|Y) - \frac{1}{n} H(\hat{X}_n|X^nY^n), \quad (6.1.3) \end{aligned}$$

где первое неравенство следует из того, что энтропия любого ансамбля не превосходит логарифма числа его элементов, второе — из того, что $H(V|X^nY^n) \geq 0$, и третье — из невозрастания информации при преобразованиях.

Предположим, что фиксирована последовательность $\mathbf{y} \in Y^n$. Рассмотрим ансамбль $\{X^n\hat{X}_n, p(\mathbf{x}, \hat{\mathbf{x}}|\mathbf{y})\}$ и обозначим через $\lambda_X(\mathbf{y})$ вероятность ошибки при декодировании последовательности $\mathbf{x} \in X^n$ при условии, что источник U_Y породил последовательность \mathbf{y} . Очевидно, $\lambda_X(\mathbf{y}) = \Pr(\mathbf{x} \neq \hat{\mathbf{x}}|\mathbf{y})$ и можно воспользоваться неравенством Фано (см. § 3.3), применив его к ансамблю $X^n\hat{X}_n$:

$$\begin{aligned} H(X^n|\hat{X}_n, \mathbf{y}) &\leq h(\lambda_X(\mathbf{y})) + \lambda_X(\mathbf{y}) \log |X^n| = \\ &= h(\lambda_X(\mathbf{y})) + n\lambda_X(\mathbf{y}) \log |X|, \end{aligned}$$

где $h(p) = -p \log p - (1-p) \log(1-p)$ и $|X|$ — число элементов в X . Усредняя левую и правую части последнего неравенства по ансамблю $\{Y^n, p(\mathbf{y})\}$ и используя выпуклость функции $h(p)$, получим

$$H(X^n|\hat{X}_nY^n) \leq h(\bar{\lambda}_X) + n\bar{\lambda}_X \log |X|, \quad (6.1.4)$$

где $\bar{\lambda}_X = \sum_{\mathbf{y}^n} \lambda_X(\mathbf{y}) p(\mathbf{y})$ — средняя вероятность ошибки декодирования сообщений источника U_X . Если $R_X = H(X|Y) - \delta$, где $\delta > 0$, то из (6.1.3) и (6.1.4) имеем

$$\frac{1}{n} h(\bar{\lambda}_X) + \bar{\lambda}_X \log |X| \geq \delta > 0,$$

откуда следует, что для некоторого $\varepsilon > 0$ $\bar{\lambda}_X \geq \varepsilon$ при любом n . Аналогично показывается, что при $R_Y < H(Y|X)$ имеет место неравенство $\bar{\lambda}_Y \geq \varepsilon > 0$. Утверждение теоремы следует из того, что $\lambda \geq \bar{\lambda}_X$ и $\lambda \geq \bar{\lambda}_Y$.

6.1.3. Прямая теорема кодирования. Теперь мы покажем, что любая пара скоростей, которая принадлежит заштрихованной

области на рис. 6.1.2, является допустимой при независимом кодировании пары зависимых дискретных источников без памяти. Заметим вначале, что для доказательства этого утверждения достаточно показать, что допустимыми являются пары скоростей $R_X = H(X|Y)$, $R_Y = H(Y)$ и $R_X = H(X)$, $R_Y = H(Y|X)$. Допустимость остальных пар, принадлежащих заштрихованной области, будет следовать из возможности разделения времени (см. лемму 6.1.1).

Рассмотрим пару скоростей $R_X = H(X|Y)$, $R_Y = H(Y)$. Доказательство допустимости этой пары будет выполнено посредством построения кода $G_n(\tilde{R}_X, \tilde{R}_Y)$, $\tilde{R}_X = H(X|Y) + \delta$, $\tilde{R}_Y = H(Y) + \delta$, который обеспечивает вероятность ошибки $\lambda < \varepsilon$ для любых положительных чисел ε и δ . Метод построения такого кода основан на следующих соображениях. Множества X^n и Y^n можно рассматривать соответственно как множества входных и выходных последовательностей дискретного канала без памяти, задаваемого переходными вероятностями $p(y|x)$, где

$$p(y|x) = \frac{p(x, y)}{\sum_y p(x, y)} \quad (6.1.5)$$

и $p(x, y)$ — распределение вероятностей, определяющее пару рассматриваемых источников. Для такого канала при любом $\varepsilon > 0$ можно построить код $G(n, R)$, содержащий $2^{nR} \approx 2^{nI(X; Y)}$ кодовых слов и обеспечивающий вероятность ошибки не большую, чем ε . Если бы на выходе источника U_X появлялись только слова этого кода, а последовательность \mathbf{y} была бы известна декодеру, то он мог бы восстановить \mathbf{x} с вероятностью ошибки, не большей чем ε . Однако на выходе источника U_X с вероятностью, близкой к единице, появляется одна из $\sim 2^{nH(X)}$ последовательностей высоковероятного множества, в то время как код $G(n, R)$ содержит самое большое $2^{nI(X; Y)} = 2^n(H(X) - H(X|Y))$ слов. Поэтому для того, чтобы исчерпать все высоковероятное множество последовательностей источника U_X , понадобится по крайней мере $2^{nH(X|Y)}$ кодов. Ниже мы покажем, что, действительно, может быть построено множество, состоящее примерно из такого числа кодов для канала без памяти с переходными вероятностями $p(y|x)$, $y \in Y$, $x \in X$, причем кодовые слова всех кодов из этого множества покрывают почти все высоковероятное множество источника U_X .

Код для кодирования источника U_X и U_Y будет строиться следующим образом. Кодирующая функция $v(\mathbf{x})$ указывает номер кода, которому принадлежит \mathbf{x} ; если последовательность \mathbf{x} не принадлежит ни одному коду, то положим $v(\mathbf{x}) = 1$. В последнем случае происходит ошибка декодирования, вероятность ко-

торой мы будем обозначать через λ_1 . Кодирующую функцию $w(y)$ указывает номер последовательности y в высоковероятном множестве источника U_Y . Если y не принадлежит этому множеству, то положим $w(y) = 1$. Вероятность ошибки декодирования, возникающей в последнем случае, будем обозначать через λ_2 . Декодер сначала восстанавливает последовательность y , а затем по номеру кода, самому коду $G(n, R)$ и y восстанавливает кодовое слово, которое и принимается за x . При этом, кроме указанных выше ошибок декодирования, с вероятностью ϵ может произойти ошибка декодирования кода $G(n, R)$. Общая вероятность ошибки декодирования λ , очевидно, не превышает $\lambda_1 + \lambda_2 + \epsilon$. Из приведенных выше соображений следует, что при $R_X > H(Y|X)$ и $R_Y > H(Y)$ величины λ_1, λ_2 и ϵ могут быть сделаны сколь угодно малыми путем выбора достаточно большого n . Аналогичные соображения могут быть приведены для обоснования допустимости пары $(R_X = H(X), R_Y = H(Y|X))$.

Перейдем теперь к точным формулировкам и доказательствам.

Лемма 6.1.2. *Пусть фиксированы дискретный канал без памяти с переходными вероятностями $\{p(y|x)\}$, $y \in Y, x \in X$, и дискретный источник без памяти U_X с ансамблем сообщений $\{X, p(x)\}$. Для любых $\epsilon > 0$ и $\delta > 0$ найдется $n_0(\epsilon, \delta)$ такое, что при $n > n_0(\epsilon, \delta)$ существует такая совокупность кодов $G_i(n, R_i)$, $i = 1, L$, для канала $\{XY, p(y|x)\}$, что*

$$1) \quad L \leq 2^{n(H(X|Y) + \delta)}, \quad (6.1.6)$$

2) вероятность ошибки каждого кода не превосходит ϵ и

$$\Pr\left(\bigcup_{i=1}^L C_i\right) \triangleq \sum_{x \in \bigcup_{i=1}^L C_i} p(x) \geq 1 - \epsilon, \quad (6.1.7)$$

где $C_i = \{x_{i1}, x_{i2}, \dots, x_{iM_i}\} \subseteq X^n$ — множество кодовых слов кода G_i , $M_i = 2^{nR_i}$.

Доказательство. Положим $\delta_1 = \delta/3$, $\epsilon_1 = \epsilon/2$. Рассмотрим высоковероятное множество источника U_X

$$T_n(\delta_1) \triangleq \left\{x: \left|\frac{1}{n} \log p(x) - H(X)\right| \leq \delta_1\right\}$$

и будем считать, что при фиксированных δ_1, ϵ_1 величина n выбрана из условия $\Pr(x \in T_n(\delta_1)) \geq 1 - \epsilon_1$. Построим код $G_1(n, R_1)$ для канала $\{XY, p(y|x)\}$ с вероятностью ошибки $\lambda \leq \epsilon$, выбирая кодовые слова в множестве $S_1 \triangleq T_n(\delta_1)$. Согласно неравенству Файнштейна (теорема 3.8.1), такой код может быть построен, причем число кодовых слов M_1 может быть выбрано не меньшим чем

$$[(1 - \epsilon_1)\lambda - \Pr(\bar{V}_t)] 2^{n\tau},$$

где $\tau > 0$ и множество V_t определено соотношением (3.8.3). Полагая $\tau = 1$ ($X; Y$) — δ_1 и учитывая (см. доказательство теоремы 3.9.1), что

$$\Pr(\bar{V}_t) \leq \Pr\left\{\left|\frac{1}{n} \sum_{i=1}^n \log \frac{p(x^{(i)}, y^{(i)})}{p(x^{(i)}) p(y^{(i)})} - I(X; Y)\right| > \delta_1\right\}_{n \rightarrow \infty} \rightarrow 0,$$

получим, что при достаточно большом n можно построить код $G_1(n, R_1)$ с числом кодовых слов

$$M_1 = 2^{nR_1} = \frac{\lambda}{2} (1 - \epsilon_1) 2^{n(I(X; Y) - \delta_1)}$$

и вероятностью ошибки λ .

Положим $S_2 \triangleq S_1 \setminus C_1$, где C_1 — множество слов кода G_1 , и построим код $G_2(n, R_2)$ для канала $\{XY, p(y|x)\}$ с вероятностью ошибки $\lambda \leq \epsilon$, выбирая кодовые слова в множестве S_2 . Аналогично предыдущему можно показать, что количество слов этого кода при достаточно большом n может быть выбрано равным

$$M_2 = 2^{nR_2} = \frac{\lambda}{2} \Pr(S_2) 2^{n(I(X; Y) - \delta_1)}.$$

Полагая $S_i \triangleq S_{i-1} \setminus C_{i-1}$, $i = 2, 3, \dots$, где C_{i-1} — множество слов кода G_{i-1} , построим код $G_i(n, R_i)$ для канала $\{XY, p(y|x)\}$ с вероятностью ошибки $\lambda \leq \epsilon$, выбирая слова из множества S_i . Количество слов кода G_i при достаточно большом n может быть выбрано равным

$$M_i = 2^{nR_i} = \frac{\lambda}{2} \Pr(S_i) 2^{n(I(X; Y) - \delta_1)}.$$

Оценим вероятность $\Pr(S_i)$. Для этого заметим, что для любой последовательности $x \in T_n(\delta_1)$

$$p(x) \geq 2^{-n(H(X) + \delta_1)}$$

и поэтому

$$\begin{aligned} \Pr(C_i) &\triangleq \sum_{x \in C_i} p(x) \geq M_i 2^{-n(H(X) + \delta_1)} = \\ &= \frac{\lambda}{2} \Pr(S_i) 2^{-n(H(X) + 2\delta_1)}. \end{aligned}$$

При этом

$$\Pr(S_2) = \Pr(S_1) - \Pr(C_1) \leq (1 - \epsilon_1) \left(1 - \frac{\lambda}{2} 2^{-n(H(X) + 2\delta_1)}\right)$$

$$\Pr(S_i) = \Pr(S_{i-1}) - \Pr(C_{i-1}) \leq$$

$$\leq (1 - \varepsilon_1) \left(1 - \frac{\lambda}{2} 2^{-n(H(X|Y) + 2\delta_1)}\right)^{i-1}.$$

Пусть $S_{L+1} \triangleq S_1 \setminus \bigcup_{i=1}^L C_i$, тогда

$$\Pr(S_{L+1}) \leq (1 - \varepsilon_1) \left(1 - \frac{\lambda}{2} 2^{-n(H(X|Y) + 2\delta_1)}\right)^L = (1 - \varepsilon_1) e^{\mu(n, L)},$$

где

$$\mu(n, L) \triangleq L \ln \left(1 - \frac{\lambda}{2} 2^{-n(H(X|Y) + 2\delta_1)}\right) \leq$$

$$\leq -L \frac{\lambda}{2} 2^{-n(H(X|Y) + 2\delta_1)}.$$

Положим $L = 2^{n(H(X|Y) + 2\delta_1)}$. Тогда

$$\Pr(S_{L+1}) \leq (1 - \varepsilon_1) \exp \left(-\frac{\lambda}{2} 2^{n\delta_1}\right).$$

Оценим теперь вероятность $\Pr \left(\bigcup_{i=1}^L C_i \right)$. Из определения множества S_{L+1} следует, что найдется $n_0(\varepsilon, \delta)$ такое, что при $n > n_0(\varepsilon, \delta)$

$$\Pr \left(\bigcup_{i=1}^L C_i \right) = \Pr(S_1) - \Pr(S_{L+1}) \geq$$

$$\geq (1 - \varepsilon_1) \left(1 - \exp \left(-\frac{\lambda}{2} 2^{n\delta_1}\right)\right) \geq 1 - \varepsilon,$$

где последнее неравенство следует из того, что при достаточно большом n величина $\exp \left(-\frac{\lambda}{2} 2^{n\delta_1}\right)$ может быть сделана как угодно малой. Таким образом, мы доказали существование не более чем $2^{n(H(X|Y) + \delta)}$ кодов с вероятностью ошибки каждого из них, не превосходящей заданного числа ε , и которые своими кодовыми словами покрывают высоковероятное множество $T_n(\delta_1)$ с точностью до подмножества, вероятность которого не превосходит ε . Заметим, что неравенство в (6.1.6) следует из того, что, начиная с некоторого $k \ll L$, вероятности $\Pr(S_k), \Pr(S_{k+1}), \dots, \Pr(S_L)$ могут все равняться нулю, поэтому количество кодов на самом деле может оказаться меньшим, чем $2^{n(H(X|Y) + \delta)}$. Лемма доказана.

Теорема 6.1.2. (Прямая теорема кодирования.) Пусть U_X, U_Y — пара зависимых дискретных источников без памяти,

$\{XY, p(x, y)\}$ — ансамбль сообщений для этой пары источников и $H(XY), H(X|Y), H(Y|X)$ — энтропии, определяемые распределением $p(x, y)$. Любая пара чисел (R_X, R_Y) такая, что $R_X \geq H(X|Y), R_Y \geq H(Y|X), R_X + R_Y \geq H(XY)$, является парой допустимых скоростей.

Доказательство. Допустимость пар скоростей $(R_X = H(X|Y), R_Y = H(Y))$ и $(R_X = H(X), R_Y = H(Y|X))$ следует из леммы и описания конструкции кода, данного непосредственно перед ней. Допустимость остальных пар скоростей, удовлетворяющих условию теоремы, вытекает из возможности использования разделения времени. Теорема доказана.

Мы рассмотрели задачу независимого кодирования двух источников. Аналогичная задача может быть поставлена и для случая любого конечного числа источников. Полученные результаты могут быть легко обобщены на этот случай (см. задачи 6.1.3—6.1.6).

§ 6.2. Кодирование источников с дополнительной информацией

6.2.1. Постановка задачи. Пусть, как и в предыдущей задаче, имеются два зависимых дискретных источника без памяти U_X и U_Y с совместным ансамблем сообщений $\{XY, p(x, y)\}$. Как и ранее, будем предполагать, что сообщения источников U_X, U_Y кодируются с помощью двух независимо работающих кодеров. Отличие рассматриваемой задачи от предыдущей состоит в том, что теперь требуется восстановление не двух последовательностей x, y на выходах источников U_X, U_Y , а только одной, например, x . При этом информация, которую доставляет декодеру второй кодер, является дополнительной. Если бы не было дополнительной информации, то для восстановления сообщений источника U_X потребовалось бы затратить примерно $H(X)$ двоичных символов на сообщение. При наличии дополнительной информации количество двоичных символов может быть уменьшено. Для восстановления сообщений источника U_X получатель может использовать выходы обоих кодеров, причем второй кодер будет при этом восполнять недостающую информацию первого.

Можно продолжить рассмотрение примера с системой метеорологических датчиков для того, чтобы показать ситуацию, в которой может возникнуть задача кодирования с дополнительной информацией. Предположим, что в центре обработки метеорологических данных требуется получить картину погоды в одном районе, скажем, районе расположения датчика U_X . Так как сообщения всех датчиков зависят, то остальные датчики системы могут рассматриваться как источники дополнительной информации. По-прежнему можно использовать различные наборы скоростей

(R_1, \dots, R_N) для N независимо работающих кодеров. Задача заключается в определении всех таких наборов скоростей, при которых возможно восстановление сообщений только одного датчика U_{X_1} .

Частичное решение этой задачи для случая двух источников было получено в предыдущем разделе. Действительно, если кодер источника U_Y передает $H(Y)$ бит на сообщение, т. е. если декодер получает полную информацию о сообщениях второго источника, то кодер источника U_X может передавать только $H(X|Y)$ бит на сообщение, $H(X|Y) < H(X)$. При этом сообщения первого источника могут быть восстановлены со сколь угодно малой вероятностью ошибки. Более того, если кодер источника U_Y передает R_Y бит на сообщение, $H(Y|X) \leq R_Y \leq H(Y)$, то кодер источника U_X может передавать $R_X \approx H(X|Y) - R_Y$ бит на сообщение. Однако при этом сообщения и первого и второго источников могут быть восстановлены сколь угодно точно, что в рассматриваемой задаче вовсе не требуется. Указанное обстоятельство позволяет надеяться на то, что, если не требовать сколь угодно точного восстановления сообщений источника U_Y , то можно понизить минимальное значение скорости R_X при фиксированной скорости R_Y . В частности, в данной задаче, как будет показано ниже, возможно кодирование со скоростью $R_X < H(X)$ при $R_Y < H(Y|X)$.

Пусть $v(x), w(y)$ — отображения множеств X^n, Y^n на множества целых чисел $V = \{1, 2, \dots, M_X\}, W = \{1, 2, \dots, M_Y\}$ соответственно. Как и раньше, будем считать, что независимое кодирование двух источников U_X, U_Y задается парой отображений $v(x), w(y)$. Декодер описывается отображением $g(i, j)$, $i \in V, j \in W$, пары чисел (i, j) в множество X^n (в отличие от предыдущей задачи, где $g(i, j)$ было отображением в $X^n Y^n$).

Вероятность ошибки λ определяется как вероятность того, что последовательность x на выходе источника U_X будет восстановлена неверно:

$$\lambda \triangleq \Pr(g(i, j) \neq x) = \sum_{\substack{x, y: \\ v(x)=i, w(y)=j, \\ g(i, j) \neq x}} p(x, y).$$

Будем говорить, что задан код $G_n(R_X, R_Y)$ с длиной n для кодирования источника U_X с дополнительной информацией, если кодируются последовательности длины n и заданы три отображения $v(x), w(y)$ (кодирование) и $g(i, j)$ (декодирование). Величины

$$R_X \triangleq \frac{\log M_X}{n}, \quad R_Y \triangleq \frac{\log M_Y}{n}$$

называются при этом скоростями кодирования источников U_X и U_Y соответственно.

Определение 6.2.1. Пара чисел (R_X, R_Y) называется допустимой парой скоростей при кодировании источника U_X с дополнительной информацией, если для любых $\epsilon > 0, \delta > 0$ найдется n и найдется код $G_n(R_X, R_Y) = \{v, w, g\}$ со скоростями кодирования $\tilde{R}_X = R_X + \delta, \tilde{R}_Y = R_Y + \delta$, кодирующими отображениями $v(x), w(y)$ и декодирующим отображением $g(i, j)$, для которого $\lambda \leq \epsilon$.

Множество всех допустимых пар скоростей при кодировании с дополнительной информацией будем обозначать через \mathcal{J}_d . Из возможности кодирования с разделением времени следует, что множество \mathcal{J}_d выпукло. Ниже будет показано, что \mathcal{J}_d имеет вид заштрихованной области на рис. 6.2.1. Двойной штриховкой на этом же рисунке показана область \mathcal{J} допустимых пар скоростей при независимом кодировании двух источников.

Прежде чем дать точное описание области допустимых пар скоростей, полезно представить себе задачу с общих позиций, чтобы попытаться найти с помощью качественных рассуждений, какими могут или не могут быть допустимые пары скоростей. Вначале заметим, что пара скоростей $R_X = H(X), R_Y = 0$, очевидно, является допустимой.

Предположим, что декодеру известен выход источника U_Y . Тогда, как было показано в предыдущем параграфе, для восстановления сообщений источника U_X требуется передать не менее $H(X|Y)$ бит на сообщение. Поэтому для любой пары допустимых скоростей (R_X, R_Y) должно выполняться неравенство $R_X \geq H(X|Y)$. Кроме того, любая пара скоростей, допустимая при независимом кодировании двух источников, допустима и при кодировании одного источника с дополнительной информацией. Таким образом, $(R_X = H(X|Y), R_Y = H(Y))$ есть пара допустимых скоростей.

Предположим теперь, что $R_Y < H(Y)$. В этом случае декодеру известна не последовательность y , появившаяся на выходе источника U_Y , а лишь подмножество, которому она принадлежит. Это следует из того, что при $R_Y < H(Y)$ кодирующее отображение $w(y)$ осуществляет разбиение высоковероятного множества источника U_Y на 2^{nR_Y} непересекающихся подмножеств таких, что все последовательности, принадлежащие одному и тому же подмножеству, отображаются в одно и то же число из W . Таким образом

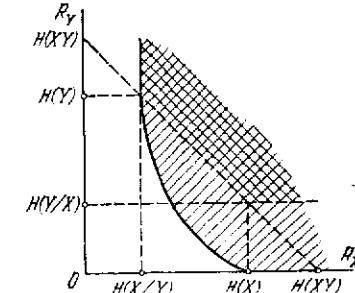


Рис. 6.2.1. Область допустимых пар скоростей при кодировании источника с дополнительной информацией.

задание кодирующего отображения $\omega(y)$ эквивалентно заданию разбиения высоковероятного множества источника U_Y на 2^{nR_Y} непересекающихся подмножеств.

В дальнейшем мы покажем, что такое разбиение может быть реализовано следующим образом. Пусть $\{Z, p(z)\}$ — некоторый вероятностный ансамбль и $\{p(y|z)\}, y \in Y, z \in Z$, — семейство переходных вероятностей из Z в Y . Будем предполагать, что эти переходные вероятности задают дискретный канал без памяти. Если распределение $p(z)$ и переходные вероятности $p(y|z)$ выбраны таким образом, что

$$p(y) = \sum_z p(z) p(y|z), \quad y \in Y,$$

то, как будет показано ниже, разбиение высоковероятного множества источника U_Y может быть реализовано с помощью совокупности решающих областей некоторого кода $G(n, R)$ для дискретного канала без памяти с переходными вероятностями $p(y|z)$. При этом число слов кода $G(n, R)$ равно $2^{nR} \approx 2^{nI(Y; Z)}$ и сами кодовые слова принадлежат высоковероятному множеству дискретного источника без памяти с ансамблем сообщений $\{Z, p(z)\}$ (источник U_Z). Такое разбиение будет задавать кодирование источника U_Y со скоростью $R_Y \approx I(Y; Z)$. При этом способе кодирования источника U_Y мы можем полагать, что декодеру известна последовательность сообщений z , появившаяся на выходе фиктивного источника U_Z . Пара источников U_X, U_Z может рассматриваться как пара зависимых дискретных источников без памяти с совместным распределением

$$p(x, z) = p(z) p(x|z) = p(z) \sum_y p(x|y) p(y|z).$$

Так как декодеру известна последовательность z и z принадлежит высоковероятному множеству источника U_Z , то в соответствии с результатами предыдущего параграфа кодирующее множество U_X отображение $v(x)$ может быть построено таким образом, что $R_X \approx H(X|Z)$. Из приведенных рассуждений следует, что пара скоростей $R_X = H(X|Z)$, $R_Y = I(Y; Z)$ должна быть допустимой для кодирования источника U_X с дополнительной информацией. Более того, как будет показано, допустимы только пары скоростей такого вида.

6.2.2. Функция $T(d)$ и ее свойства. В дальнейшем для характеристики области \mathfrak{X}_d потребуется некоторая специальная функция, которую мы введем и изучим в этом разделе.

Пусть $\{XY, p(x, y)\}$ — пара совместно заданных дискретных ансамблей. Введем в рассмотрение дискретное множество Z , $|Z| < \infty$, и зададим распределение вероятностей $p(x, y, z)$ на

тройках $(x, y, z) \in XYZ$ посредством следующих соотношений:

$$\begin{aligned} p(x, y, z) &= p(x|y) p(y|z) p(z), \\ \sum_z p(x, y, z) &= p(x, y), \end{aligned} \quad (6.2.1)$$

где $p(y, z) = p(z) p(y|z)$ — распределение вероятностей на YZ . Из (6.2.1) следует, что при данном сообщении $y \in Y$ сообщения $x \in X$ и $z \in Z$ статистически независимы. Ясно, что распределение вероятностей $p(x, y, z)$, удовлетворяющее соотношениям (6.2.1), может быть выбрано многими способами с помощью различных выборов множества Z и распределения $p(y, z)$. Каждый такой выбор определяет ансамбль $\{XYZ, p(x, y, z)\}$. Обозначим через \mathcal{A} множество различных ансамблей троек, распределения вероятностей на которых удовлетворяют условиям (6.2.1). Для каждого ансамбля XYZ из \mathcal{A} обычным образом определены энтропии $H(X|Z)$ и $H(Y|Z)$. Пусть $\mathcal{A}(d)$ — максимальное подмножество множества \mathcal{A} такое, что $H(Y|Z) \geq d$ для каждого ансамбля из $\mathcal{A}(d)$. Определим функцию $T(d)$, $d \geq 0$, следующим образом:

$$T(d) \triangleq \inf_{\mathcal{A}(d)} H(X|Z). \quad (6.2.2)$$

В определении функции $T(d)$ используется знак \inf , а не \min , так как мощность множества Z хотя и ограничена, но может быть сколь угодно большой.

Лемма 6.2.1. Функция $T(d)$ — монотонно неубывающая, выпуклая вниз функция при всех $d \geq 0$.

Доказательство. Для доказательства монотонного неубывания достаточно заметить, что $\mathcal{A}(d_1) \subseteq \mathcal{A}(d_2)$ при всех $d_1 \geq d_2 \geq 0$, и воспользоваться определением (6.2.2).

Для доказательства выпуклости достаточно показать, что для любых $d_1 \geq 0$, $d_2 \geq 0$ и любого α , $0 < \alpha < 1$, имеет место неравенство

$$T(\alpha d_1 + (1 - \alpha) d_2) \leq \alpha T(d_1) + (1 - \alpha) T(d_2). \quad (6.2.3)$$

Пусть зафиксировано $\varepsilon > 0$, тогда можно найти два ансамбля XYZ' , XYZ'' из \mathcal{A} со следующими свойствами:

$$\begin{aligned} \{XYZ', p(x|y) p_1(y|z') p_1(z')\} &\in \mathcal{A}(d_1), \\ H(Y|Z') &\geq d_1, \quad H(X|Z') \leq T(d_1) + \varepsilon, \\ \{XYZ'', p(x|y) p_2(y|z'') p_2(z'')\} &\in \mathcal{A}(d_2), \\ H(Y|Z'') &\geq d_2, \quad H(X|Z'') \leq T(d_2) + \varepsilon. \end{aligned}$$

Множества Z' , Z'' можно рассматривать как дизъюнктные. Пусть $Z \triangleq Z' \cup Z''$; для любого элемента $z \in Z$ положим

$$p(z) \triangleq \begin{cases} \alpha p_1(z) & \text{при } z \in Z', \\ (1 - \alpha) p_2(z) & \text{при } z \in Z'', \end{cases}$$

$$p(y|z) \triangleq \begin{cases} p_1(y|z) & \text{при } z \in Z', \\ p_2(y|z) & \text{при } z \in Z''. \end{cases}$$

При этом определен ансамбль $XYZ \in \mathcal{A}$, для которого

$$H(Y|Z) = \alpha H(Y|Z') + (1 - \alpha) H(Y|Z'') \geq \alpha d_1 + (1 - \alpha) d_2, \quad (6.2.4)$$

$$H(X|Z) = \alpha H(X|Z') + (1 - \alpha) H(X|Z'') \leq \alpha T(d_1) + (1 - \alpha) T(d_2) + \epsilon. \quad (6.2.5)$$

Таким образом, из (6.2.4) следует, что ансамбль XYZ принадлежит $\mathcal{A}(d)$, а из (6.2.5), произвольности ϵ и определения (6.2.2) следует неравенство (6.2.3). Лемма доказана.

Пусть U_X , U_Y — пара зависимых дискретных источников без памяти с ансамблем сообщений $\{XY, p(x, y)\}$ и $\{X^n, Y^n, p(x, y)\}$, $p(x, y) = \prod_{i=1}^n p(x^{(i)}, y^{(i)})$ — ансамбль последовательностей сообщений на выходах источников U_X , U_Y . Пусть Z , $|Z| < \infty$ — некоторое дискретное множество. Рассмотрим распределение вероятностей $p(x, y, z)$ на тройках $(x, y, z) \in X^n Y^n Z$, задаваемое следующими соотношениями:

$$\begin{aligned} p(x, y, z) &= p(x|y) p(y|z) p(z), \\ \sum_z p(x, y, z) &= p(x, y), \end{aligned} \quad (6.2.6)$$

где $p(y|z) p(z) = p(y, z)$ — распределение вероятностей на $Y^n Z$. Обозначим через \mathcal{A}_n множество различных ансамблей $X^n Y^n Z$, распределения вероятностей на которых удовлетворяют условиям (6.2.6). Пусть $\mathcal{A}_n(d)$ — подмножество множества \mathcal{A}_n такое, что $\frac{1}{n} H(Y^n|Z) \geq d$ для каждого ансамбля из $\mathcal{A}_n(d)$. Определим функцию $T_n(d)$, $d \geq 0$, следующим образом:

$$T_n(d) \triangleq \inf_{\mathcal{A}_n(d)} \frac{1}{n} H(X^n|Z). \quad (6.2.7)$$

Л е м м а 6.2.2. Для любого целого положительного n и любого $d \geq 0$

$$T_n(d) = T(d).$$

Д о к а з а т е л ь с т в о. Заметим, что при $\{XYZ, p(x, y, z)\} \in \mathcal{A}(d)$ ансамбль $\{X^n Y^n Z, p(x, y, z)\}$, где $z \in Z^n$, $p(x, y, z) =$

$= p(x, y, z) = \prod_{i=1}^n p(x^{(i)}, y^{(i)}, z^{(i)})$, принадлежит $\mathcal{A}_n(d)$. Следовательно, $T_n(d) \leq T(d)$ и для доказательства леммы достаточно показать, что $T_n(d) \geq T(d)$ или что для любого ансамбля $X^n Y^n Z \in \mathcal{A}_n(d)$ выполняется неравенство $H(X^n|Z) \geq nT(d)$.

Обозначим $X^n = X_1 \dots X_n$, $Y^n = Y_1 \dots Y_n$, где X_i , Y_i — ансамбли, соответствующие моменту времени i , $i = 1, n$. Для любого ансамбля $X^n Y^n Z$ из \mathcal{A}_n имеют место два утверждения:

1) при любых фиксированных сообщениях $(y^{(1)}, \dots, y^{(i-1)}, z) \in Y_1 \dots Y_{i-1} Z$ ансамбли X_i и $X_1 \dots X_{i-1}$ статистически независимы;

2) при любом фиксированном сообщении $y^{(i)} \in Y_i$ ансамбли X_i и $Y_1 \dots Y_{i-1} Z$ статистически независимы.

Первое утверждение проверяется с помощью следующей цепочки соотношений:

$$\begin{aligned} p(x^{(1)}, \dots, x^{(i)}, y^{(1)}, \dots, y^{(i-1)}, z) &= \\ &= \sum_{X_{i+1} \dots X_n} \sum_{Y_i \dots Y_n} p(x|y) p(y|z) p(z) = \\ &= \sum_{Y_i \dots Y_n} \prod_{j=1}^i p(x^{(j)}|y^{(j)}) p(y, z) = \\ &= \prod_{j=1}^{i-1} p(x^{(j)}|y^{(j)}) \sum_{Y_i \dots Y_n} p(x^{(i)}|y^{(1)}, \dots, y^{(i)}, z) p(y^{(1)}, \dots, y^{(i)}, z) = \\ &= \prod_{j=1}^{i-1} p(x^{(j)}|y^{(j)}) p(x^{(i)}, y^{(1)}, \dots, y^{(i-1)}, z) = \\ &= \prod_{j=1}^{i-1} p(x^{(j)}|y^{(j)}) p(x^{(i)}|y^{(1)}, \dots, y^{(i-1)}, z) p(y^{(1)}, \dots, y^{(i-1)}, z), \end{aligned} \quad (6.2.8)$$

где третье равенство следует из того, что $p(x^{(i)}|y^{(1)}, \dots, y^{(i)}, z) = p(x^{(i)}|y^{(i)})$ в соответствии с заданием пары источников как источников без памяти, а также из соотношений (6.2.6). Второе утверждение проверяется аналогично и следует из соотношения

$$\begin{aligned} p(x^{(i)}, y^{(1)}, \dots, y^{(i)}, z) &= \\ &= p(y^{(1)}, \dots, y^{(i-1)}, z) p(y^{(i)}|y^{(1)}, \dots, y^{(i-1)}, z) p(x^{(i)}|y^{(i)}). \end{aligned} \quad (6.2.9)$$

Используя эти соотношения и свойства энтропии, можно записать

$$\begin{aligned} H(X^n|Z) &= \sum_{i=1}^n H(X_i|X_1 \dots X_{i-1} Z) \geq \\ &\geq \sum_{i=1}^n H(X_i|X_1 \dots X_{i-1} Y_1 \dots Y_{i-1} Z) = \sum_{i=1}^n H(X_i|Y_1 \dots Y_{i-1} Z), \end{aligned} \quad (6.2.10)$$

где последнее равенство — следствие указанной выше независимости (соотношение (6.2.8)).

Положим $Z_i \triangleq Y_1 \dots Y_{i-1}Z$ и $d_i \triangleq H(Y_i | Y_1 \dots Y_{i-1}Z)$. Тогда из (6.2.9) следует, что $X_iY_iZ_i \in \mathcal{A}(d_i)$, откуда с учетом (6.2.10) получим

$$H(X^n | Z) \geq \sum_{i=1}^n T(d_i) = n \sum_{i=1}^n \frac{1}{n} T(d_i) \geq nT(\bar{d}),$$

где последнее неравенство следует из выпуклости функции $T(d)$ и того, что

$$\bar{d} \triangleq \frac{1}{n} H(Y^n | Z) = \frac{1}{n} \sum_{i=1}^n H(Y_i | Y_1 \dots Y_{i-1}Z) = \frac{1}{n} \sum_{i=1}^n d_i.$$

Так как $\bar{d} = \frac{1}{n} H(Y^n | Z) \geq d$, а функция $T(d)$ не убывает с ростом d , то

$$H(X^n | Z) \geq nT(\bar{d}) \geq nT(d).$$

Лемма доказана.

Неограниченность числа элементов в множестве Z , участвующем в определении функции $T(d)$, затрудняет, а возможно, делает даже нереализуемым, вычисление значений этой функции. Следующие две леммы позволяют ограничить число элементов в множестве Z и, тем самым, снять проблему вычислимости функции $T(d)$.

Лемма 6.2.3. *Пусть фиксированы множества Y, Z и распределения вероятностей $p(z), p(y|z), y \in Y, z \in Z$. Если $|Z| > |Y|$, то найдется распределение $\tilde{p}(z), z \in Z$, такое, что*

$$\sum_z \tilde{p}(z) p(y|z) = \sum_z p(z) p(y|z) \quad (6.2.11)$$

для всех $y \in Y$ и число элементов $z \in Z$, для которых $\tilde{p}(z) > 0$, не превосходит $|Y|$.

Доказательство. Пусть $|z| > |Y|$ и $p(z) > 0$ для всех $z \in Z$. Рассмотрим следующую однородную систему линейных уравнений относительно переменных $h(z), z \in Z$:

$$\sum_z h(z) p(y|z) = 0, \quad y \in Y. \quad (6.2.12)$$

Если $|Z| > |Y|$, то эта система имеет ненулевое решение $h^*(z), z \in Z$. Суммируя обе части (6.2.12) по всем $y \in Y$, получим, что $\sum_z h^*(z) = 0$, и поэтому среди элементов $\{h^*(z)\}$ имеется по крайней мере один отрицательный. Выберем $\alpha > 0$ таким образом, что для некоторого $\hat{z} \in Z$

$$p_1(\hat{z}) \triangleq p(\hat{z}) + \alpha h^*(\hat{z}) = 0 \quad \text{и} \quad p_1(z) \triangleq p(z) + \alpha h^*(z) \geq 0$$

для остальных $z \in Z$. Очевидно, что при этом $\sum_z p_1(z) = 1$ и, следовательно, $p_1(z)$ является распределением вероятностей на Z . Обозначим через Z_1 такое подмножество множества Z , что $p_1(z) > 0$ только для элементов $z \in Z_1$. Тогда $|Z_1| < |Z|$. Если $|Z_1| > |Y|$, то, заменив Z на Z_1 и проводя аналогичные рассуждения, построим множество Z_2 и распределение $p_2(z)$ такие, что $|Z_2| < |Z_1|$, $p_2(z) > 0$, если $z \in Z_2$, и $p_2(z) = 0$, если $z \notin Z_2$. Повторяя этот процесс не более чем $|Z| - |Y|$ раз, построим распределение $\tilde{p}(z), z \in Z$, удовлетворяющее условиям леммы. Лемма доказана.

Лемма 6.2.4. *Пусть $\hat{\mathcal{A}}(d)$ — такое максимальное подмножество множества $\mathcal{A}(d)$, что, если $XYZ \in \hat{\mathcal{A}}(d)$, то $|Z| \leq |Y|$. Тогда*

$$T(d) \triangleq \inf_{\hat{\mathcal{A}}(d)} H(X | Z) = \min_{\hat{\mathcal{A}}(d)} H(X | Z).$$

Доказательство. Пусть \mathcal{A}_L — такое максимальное подмножество множества \mathcal{A} , что, если $XYZ \in \mathcal{A}_L$, то $|Z| \leq L$. Для доказательства леммы достаточно показать, что для любого $L > |Y|$ выполняется равенство

$$T_L(d) \triangleq \min_{\mathcal{A}_L(d)} H(X | Z) = \min_{\hat{\mathcal{A}}(d)} H(X | Z) \triangleq \hat{T}(d), \quad (6.2.13)$$

где $\mathcal{A}_L(d) \subseteq \mathcal{A}_L$, и если $XYZ \in \mathcal{A}_L(d)$, то $H(Y | Z) \geq d$.

В соответствии с теоремой Куна—Таккера для нахождения значения функции $T_L(d)$ при каждом фиксированном значении d необходимо решить следующую задачу минимизации: найти

$$J_L(\lambda) \triangleq \min_{\mathcal{A}_L} (H(X | Z) + \lambda H(Y | Z)), \quad (6.2.14)$$

где λ — величина, определяемая условиями Куна—Таккера по заданному значению d . Тогда ансамбль $X^*Y^*Z^* \in \mathcal{A}_L$, который доставляет минимум (6.2.14), будет доставлять минимум также и (6.2.13). Если d изменяется в области своих значений, то λ также изменяется некоторым образом и поэтому $J_L(\lambda)$ можно рассматривать как функцию от λ . В общем случае $J_{L_1}(\lambda) \leq J_{L_2}(\lambda)$, если $L_1 > L_2$. Если показать, что $J_L(\lambda) = J_{L_0}(\lambda)$ для любого $L > |Y| \triangleq L_0$ и любого λ , то минимум в (6.2.14) при любых $L > L_0$ будет для каждого d достигаться на одном и том же подмножестве \mathcal{A}_L множества \mathcal{A}_L , $L > L_0$. Тем самым минимум в (6.2.13) также будет достигаться на подмножестве $\mathcal{A}_{L_0}(d)$ множества $\mathcal{A}_L(d)$, $L > L_0$. Следовательно, $T_L(d) = T_{L_0}(d)$ для всех d , и лемма будет доказана. Покажем, что $J_L(\lambda) = J_{L_0}(\lambda)$.

Пусть для некоторого фиксированного λ минимум в (6.2.14) достигается на ансамбле $XYZ \in \mathcal{A}_L$, который задается распределениями $p^*(z)$, $z \in Z$, и $p^*(y|z)$, $y \in Y$, $z \in Z$. Будем считать, что в функциях $H(X|Z)$ и $H(Y|Z)$ аргументы $p(y|z)$, $y \in Y$, $z \in Z$, фиксированы и равны $p^*(y|z)$, $y \in Y$, $z \in Z$. В этом случае

$$J_L(\lambda) = \min_{\substack{p(z): \\ \sum_z p(z) p^*(y|z) = p(y), y \in Y}} (H(X|Z) + \lambda H(Y|Z)). \quad (6.2.15)$$

Так как функции $H(X|Z)$ и $H(Y|Z)$ линейны относительно распределения $p(z)$, $z \in Z$, то выражение под знаком минимума в (6.2.1) является, тривиальным образом, выпуклой функцией. При этом из теоремы Куна—Таккера следует, что необходимые и достаточные условия, которым должно удовлетворять распределение $p(z)$, минимизирующее правую часть (6.2.15), суть

$$H(X|z) + \lambda H(Y|z) + \sum_Y \lambda_y p^*(y|z) \geq 0 \quad (6.2.16)$$

для всех $z \in Z$, где λ_y — множитель Лагранжа, обусловленный ограничением $\sum_z p(z) p^*(y|z) = p(y)$. Так как соотношение (6.2.16) не зависит от $p(z)$, $z \in Z$, то минимум в правой части (6.2.15) достигается на любом распределении $p(z)$ таком, что

$$\sum_z p(z) p^*(y|z) = p(y), \quad y \in Y. \quad (6.2.17)$$

В соответствии с леммой 6.2.3 найдется такое распределение $\tilde{p}(z)$, $z \in Z$, при котором выполняется (6.2.17) и число элементов $z \in Z$, для которых $p(z) > 0$, не превосходит $|Y|$. Отсюда очевидно следует, что $J_L(\lambda) = J_{L_0}(\lambda)$. Лемма доказана.

6.2.3. Обратная теорема кодирования. Пусть U_X , U_Y — совместно заданные дискретные источники без памяти и $\{XY, p(x, y)\}$ — совместный ансамбль сообщений. Пусть $\hat{\mathcal{A}}$ — множество всех ансамблей XYZ , распределения вероятностей на которых удовлетворяют условиям (6.2.1) и $|Z| \leq |Y|$. Определим множество \mathfrak{R}_d^* пар чисел (R_X, R_Y) следующим образом:

$$\mathfrak{R}_d^* \triangleq \{(R_X, R_Y): R_X \geq H(X|Z), R_Y \geq I(Y; Z)$$

для некоторого $XYZ \in \hat{\mathcal{A}}$.

В этом пункте будет показано, что $\mathfrak{R}_d \subseteq \mathfrak{R}_d^*$. Другими словами, если пара скоростей (R_X, R_Y) допустима при кодировании источника U_X с дополнительной информацией, то эта пара должна принадлежать множеству \mathfrak{R}_d^* .

На плоскости (R_X, R_Y) множество \mathfrak{R}_d^* имеет границу, определяемую функцией $T(d)$. Действительно, энтропия $H(Y)$ не зави-

сит от выбора распределения $p(y, z)$ и принимает одно и то же значение для всех ансамблей $XYZ \in \hat{\mathcal{A}}$. Если положить $d = H(Y) - R_Y$, то $H(Y|Z) \geq H(Y) - R_Y$ для всякого ансамбля $XYZ \in \hat{\mathcal{A}}(d)$ и, следовательно, $R_Y \geq I(Y; Z)$. С другой стороны, величина $T(d)$ равна $H(X|Z)$ и представляет собой наименьшее возможное значение скорости R_X при данном R_Y . Таким образом, для любой пары $(R_X, R_Y) \in \mathfrak{R}_d^*$

$$R_X \geq T(H(Y) - R_Y)$$

и, наоборот, любая пара (R_X, R_Y) , удовлетворяющая этому неравенству, принадлежит \mathfrak{R}_d^* .

Теорема 6.2.1 (обратная теорема кодирования). Пусть U_X , U_Y — пара зависимых дискретных источников без памяти и $(X, Y, p(x, y))$ — совместный ансамбль сообщений. Если пара скоростей (R_X, R_Y) допустима при кодировании источника U_X с дополнительной информацией, то эта пара принадлежит множеству \mathfrak{R}_d^* .

Доказательство. Пусть $G_n(R_X, R_Y) = \{v, w, g\}$ — некоторый фиксированный код для кодирования источника U_X с дополнительной информацией. Как и при доказательстве теоремы 6.1.1, имеем

$$\begin{aligned} R_X \triangleq \frac{\log M_X}{n} &\geq \frac{1}{n} H(V|W) \geq \frac{1}{n} I(V; X^n|W) \geq \\ &\geq \frac{1}{n} I(\hat{X}_n; X^n|W) = \frac{1}{n} (H(X^n|W) - H(\hat{X}_n|X^nW)), \end{aligned}$$

где сохранены обозначения теоремы 6.1.1. Далее, применяя неравенство Фано к ансамблю $(X^n \hat{X}_n, p(x, \hat{x}|j))$, $j \in W$, и следуя доказательству этой теоремы, можно получить, что при $R_X < \frac{1}{n} H(X^n|W)$ найдется $\varepsilon > 0$ такое, что для любого n и любого кода $G_n(R_X, R_Y)$ вероятность ошибки $\lambda \geq \varepsilon$. Таким образом, для любой допустимой пары скоростей (R_X, R_Y) код должен быть таким, чтобы $R_X \geq \frac{1}{n} H(X^n|W)$. Покажем теперь, что последнее условие эквивалентно требованию того, чтобы (R_X, R_Y) принадлежали \mathfrak{R}_d^* .

Очевидно, что

$$\begin{aligned} nR_Y &\geq H(W) \geq I(W; Y^n) = H(Y^n) - H(Y^n|W) = \\ &= nH(Y) - H(Y^n|W). \end{aligned}$$

Из этого соотношения следует, что

$$\frac{1}{n} H(Y^n|W) \geq H(Y) - R_Y.$$

Так как $T(d)$ — неубывающая функция d , то

$$T\left(\frac{1}{n}H(Y^n|W)\right) \geq T(H(Y) - R_Y).$$

Нетрудно заметить, что ансамбль WX^nY^n принадлежит множеству \mathcal{A}_n , и поэтому, применяя определение (6.2.7), лемму 6.2.2 и последнее неравенство, получим

$$\frac{1}{n}H(X^n|W) \geq T_n\left(\frac{1}{n}H(Y^n|W)\right) \geq T(H(Y) - R_Y).$$

Таким образом, с учетом леммы 6.2.4 для любой пары допустимых скоростей (R_X, R_Y) найдется такой ансамбль $XYZ \in \hat{\mathcal{A}}$, что $R_X \geq H(X|Z)$ и $H(Y|Z) \geq H(Y) - R_Y$ или $R_Y \geq I(Y; Z)$. Теорема доказана.

Лемма 6.2.4. Прямая теорема кодирования. Мы покажем теперь, что любая пара скоростей $(R_X, R_Y) \in \mathfrak{R}_d^*$ допустима при кодировании дискретного источника с дополнительной информацией. Совместно с доказанной выше обратной теоремой это установит, что $\mathfrak{R}_d = \mathfrak{R}_d^*$. Для произвольного ансамбля $XYZ \in \mathcal{A}$ будет указан метод построения кода со скоростью $R_X = H(X|Z) + \delta$, $R_Y = I(Y; Z) + \delta$ и вероятностью ошибки $\lambda \ll \varepsilon$, где ε и δ — произвольно малые положительные числа.

В качестве первого шага докажем утверждение, являющееся по существу непосредственным следствием из прямой теоремы кодирования для пары зависимых источников (теоремы 6.1.2).

Лемма 6.2.5. Пусть U_X, U_Y — пара зависимых дискретных источников без памяти с совместным ансамблем сообщений $\{XY, p(x, y)\}$. Пусть $w(y)$ — некоторое фиксированное отображение множества Y^n на множество $W = \{1, 2, \dots, M_Y\}$. Для любых $\varepsilon > 0$, $\delta > 0$ найдется $n_0(\varepsilon, \delta)$ такое, что при $n > n_0(\varepsilon, \delta)$ существует код $G_n(R_X, R_Y)$ для кодирования источника U_X с дополнительной информацией, для которого $R_X = \frac{1}{n}H(X^n|W) + \delta$, $R_Y = \frac{1}{n}H(W) + \delta$ и вероятность ошибки $\lambda \leq \varepsilon$.

Доказательство. Введем в рассмотрение два новых дискретных источника U_{X^n} и W . Последовательность сообщений на выходе первого представляет собой последовательность блоков длины m : $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots$, а последовательность сообщений на выходе второго — последовательность элементов из множества W : $w(y^{(1)}), w(y^{(2)}), \dots$. Из теоремы 6.1.2 следует, что для пары источников U_{X^n}, U_W существует такой код $G_l(R'_X, R'_Y)$, где l достаточно велико, что $R'_X = H(X^n|W) + \delta'$, $R'_Y = H(W) + \delta'$ и вероятность ошибки $\lambda' \leq \varepsilon'$, где ε' и δ' — произвольно малые положительные числа. Очевидно, что этот код одновременно является

и кодом $G_n(R_X, R_Y)$ с длиной $n = ml$ для кодирования источника U_X с дополнительной информацией, причем $R_X = \frac{1}{m}H(X^n|W) + \delta$, $R_Y = \frac{1}{m}H(W) + \delta$ и вероятность ошибки $\lambda \leq \varepsilon'$, где $\delta = \delta'/m$, $\varepsilon = \varepsilon'$. Лемма доказана.

Для доказательства прямой теоремы теперь достаточно показать, что для любого ансамбля $XYZ \in \mathcal{A}$ и для любого $\delta > 0$ можно построить такое отображение $w(y)$ множества Y^n на множество W , что $\frac{1}{n}H(W) \leq I(Y; Z) + \delta$ и $\frac{1}{n}H(X^n|W) \leq H(X|Z) + \delta$.

Пусть фиксирован некоторый ансамбль $\{XYZ, p(x, y, z)\}$ и $\{X^nY^nZ^n, p(x, y, z)\}$ — n -я степень ансамбля XYZ , т. е.

$$p(x, y, z) := \prod_{i=1}^n p(x^{(i)}, y^{(i)}, z^{(i)}) \quad (6.2.18)$$

для всех троек $(x, y, z) \in X^nY^nZ^n$. Следующая лемма устанавливает свойства типичного множества последовательностей троек из ансамбля $X^nY^nZ^n$.

Лемма 6.2.6. Для любых $\varepsilon > 0$, $\delta > 0$ найдется $n_0(\varepsilon, \delta)$ такое, что при $n > n_0(\varepsilon, \delta)$ существует множество $T \subseteq X^nY^nZ^n$, для которого имеют место следующие утверждения. Для любой тройки $(x, y, z) \in T$

1. $\left| -\frac{1}{n} \log p(z) - H(Z) \right| \leq \delta$;
2. $\left| -\frac{1}{n} \log p(y) - H(Y) \right| \leq \delta$;
3. $\left| -\frac{1}{n} \log p(y, z) - H(YZ) \right| \leq \delta$;
4. $\left| -\frac{1}{n} \log p(x, z) - H(XZ) \right| \leq \delta$;
5. $\left| -\frac{1}{n} \log p(x, y, z) - H(XYZ) \right| \leq \delta$;
6. $\Pr(T) \triangleq \Pr((x, y, z) \in T) \geq 1 - \varepsilon$.

Пусть, кроме того, $T_Y(z) \triangleq \{y : (x, y, z) \in T \text{ хотя бы для одной последовательности } x \in X^n\}$, $T_X(y, z) \triangleq \{x : (x, y, z) \in T\}$, $T_Z \triangleq \{z : (x, y, z) \in T \text{ хотя бы для одной пары } (x, y) \in X^nY^n\}$.

Тогда имеют место следующие неравенства:

$$7. \quad \Pr(T_Y(z)) \triangleq \sum_{y \in T_Y(z)} p(y|z) \geq 1 - \varepsilon;$$

$$8. \quad \Pr(T_X(y, z)) \triangleq \sum_{x \in T_X(y, z)} p(x|yz) \geq 1 - \varepsilon;$$

$$9. \quad \Pr(T_Z) \triangleq \sum_{z \in T_Z} p(z) \geq 1 - \varepsilon.$$

Доказательство. Обозначим через $T^{(i)}$ такое подмножество из $X^n Y^n Z^n$, для которого выполняется неравенство с номером i , $i = \overline{1, 5}$, утверждения леммы. Из (6.2.18) и из закона больших чисел следует, что для любого положительного ε_1 при достаточно большом n все пять вероятностей $\Pr(T^{(i)})$, $i = \overline{1, 5}$ не меньше чем $1 - \varepsilon_1$. Положим $\hat{T} \triangleq \bigcap_{i=1}^5 T^{(i)}$. Тогда

$$\Pr(\hat{T}) = 1 - \Pr\left(\bigcup_{i=1}^5 \bar{T}^{(i)}\right) \geq 1 - 5\varepsilon_1, \quad (6.2.19)$$

где черта означает дополнение.

Пусть

$\tilde{T}_{YZ} \triangleq \{(y, z): (x, y, z) \in \hat{T} \text{ хотя бы для одной последовательности } x \in X^n\}$,

$$\tilde{T}_X(y, z) \triangleq \{x: (x, y, z) \in \hat{T}\}.$$

Обозначим через \tilde{T}_{YZ} такое подмножество множества T_{YZ} , что для любой пары $(y, z) \in \tilde{T}_{YZ}$ при $l > 1$ выполняется неравенство $\Pr(\tilde{T}_X(y, z)) \geq 1 - 5l\varepsilon_1$. Из этих определений и (6.2.19) следует, что

$$\begin{aligned} 1 - 5\varepsilon_1 &\leq \Pr(\hat{T}) = \sum_{\tilde{T}_{YZ}} p(y, z) \sum_{\tilde{T}_X(y, z)} p(x|yz) \leq \\ &\leq \Pr(\tilde{T}_{YZ}) + (1 - \Pr(\tilde{T}_{YZ}))(1 - 5l\varepsilon_1). \end{aligned}$$

откуда получается, что

$$\Pr(\tilde{T}_{YZ}) \geq 1 - \frac{1}{l}. \quad (6.2.20)$$

Пусть

$\tilde{T}_Z \triangleq \{z: (y, z) \in \tilde{T}_{YZ} \text{ хотя бы для одной последовательности } y \in Y^n\}$,

$$\tilde{T}_Y(z) \triangleq \{y: (y, z) \in \tilde{T}_{YZ}\}$$

и пусть $T_Z^* —$ такое подмножество множества \tilde{T}_Z , что для любой последовательности $z \in T_Z^*$ выполняется неравенство $\Pr(\tilde{T}_Y(z)) \geq 1 - 1/V\bar{l}$. Из этих определений и (6.2.20) следует, что

$$\begin{aligned} 1 - \frac{1}{l} &\leq \Pr(\tilde{T}_{YZ}) = \sum_{\tilde{T}_Z} p(z) \sum_{\tilde{T}_Y(z)} p(y|z) \leq \\ &\leq \Pr(T_Z^*) + (1 - \Pr(T_Z^*))\left(1 - \frac{1}{V\bar{l}}\right), \end{aligned}$$

откуда получается, что

$$\Pr(T_Z^*) \geq 1 - \frac{1}{V\bar{l}}.$$

Положим теперь

$$T \triangleq \bigcup_{T_Z^* \subset \tilde{T}_Z} \bigcup_{\tilde{T}_Y(z)} \tilde{T}_X(y, z). \quad (6.2.21)$$

Так как $T \subseteq \hat{T}$, то для множества T неравенства 1÷5 утверждения леммы выполняются. Кроме того,

$$\begin{aligned} \Pr(T) &= \sum_{T_Z^*} p(z) \sum_{\tilde{T}_Y(z)} p(y|z) \sum_{\tilde{T}_X(y, z)} p(x|yz) \geq \\ &\geq (1 - 5l\varepsilon_1)\left(1 - \frac{1}{V\bar{l}}\right)^2 \geq 1 - 5l\varepsilon_1 - \frac{2}{V\bar{l}} - 5\varepsilon_1. \end{aligned}$$

Если выбрать $\varepsilon_1 = 1/l\sqrt{\bar{l}}$ и $V\bar{l} = 12/\varepsilon$, то легко убедиться в том, что для множества, определяемого соотношением (2.6.21), условие 6 леммы выполняется. Далее нетрудно заметить, что имеют место равенства

$$T_Z = T_Z^*, \quad T_Y(z) = \tilde{T}_Y(z), \quad T_X(y, z) = \tilde{T}_X(y, z).$$

Поэтому при указанном выше выборе величин l и ε_1 неравенства 7÷9 утверждения леммы также выполняются. Лемма доказана.

Далее нам понадобится еще одно утверждение, относящееся к типичным последовательностям ансамбля $X^n Y^n Z^n$ и свойствам кодов для канала $\{ZY, p(y|z)\}$. Мы покажем, что можно построить код, слова которого z_i , $i = \overline{1, M}$, принадлежат высоковероятному множеству ансамбля Z^n , решающие области $A_i \subseteq T_Y(z_i)$, причем число кодовых слов $M \approx 2^{nI(Y; Z)}$, а решающие области A_i почти полностью покрывают множество Y^n (т. е. вероятность попадания в непокрытую часть Y^n может быть сделана произвольно малой). Такой код мы будем использовать для задания отображения $w(y)$, определяющего кодирование источника U_Y .

Лемма 6.2.7. Пусть $T \subseteq X^n Y^n Z^n$ — подмножество, удовлетворяющее условиям леммы 6.2.6. Рассмотрим дискретный

канал без памяти с множеством входных сигналов Z , выходных сигналов Y и переходными вероятностями $p(y|z)$. Для этого канала, любых $\lambda > 0$, $\varepsilon > 0$, $\varepsilon < \lambda$, и $\delta > 0$ найдется код $G(n, R) = \{(z_1, A_1; \dots; z_M, A_M)\}$, $M = 2^{nR}$, для которого кодовые слова $z_1, \dots, z_M \in T_Z \subseteq Z^n$, решающие области $A_i \subseteq T_Y(z_i) \equiv Y^n$, $i = 1, M$, и имеют место следующие неравенства:

1. $\Pr(A_i | z_i) \triangleq p(y \in A_i | z_i) \geq 1 - \lambda$, $i = \overline{1, M}$,
2. $\Pr(A_0) \triangleq p\left(y \notin \bigcup_{i=1}^M A_i\right) \leq (1 - \lambda) \Pr(z \notin C) +$
 $\quad \quad \quad + \lambda \Pr(z \in C) + 2\varepsilon$,

где $A_0 \triangleq Y^n \setminus \bigcup_{i=1}^M A_i$, $C = \{z_1, \dots, z_M\}$ — множество кодовых слов.

Кроме того, для любого кода, для которого $z_i \in T_Z$, $A_i \subseteq T_Y(z_i)$, $i = \overline{1, M}$, и выполнены неравенства 1 и 2, число кодовых слов удовлетворяет условию

$$3. \quad M < 2^{n(1(Y; Z) + 4\delta)}.$$

Доказательство. Для фиксированного $\lambda > 0$ выберем величину n настолько большой, чтобы числа $\varepsilon > 0$ и $\delta > 0$, участвующие в формулировке леммы 6.2.6, удовлетворяли неравенствам $\delta < \varepsilon < \lambda$. Для канала $\{ZY, p(y|z)\}$ построим максимальный код с вероятностью ошибки λ (см. теорему 3.8.1), выбирая в качестве кодовых слов последовательности из множества T_Z , а в качестве решающих областей — множества $A_i = T_Y(z_i) \setminus \bigcup_{j=1}^{i-1} A_j$. При этом обеспечивается выполнение условий 1.

Для доказательства выполнения неравенства 2 учтем, что мы построили максимальный код и поэтому для любой последовательности $z \in T_Z$, не принадлежащей множеству C кодовых слов, имеет место неравенство

$$\Pr(T_Y(z)) = \Pr\left(y \in \bigcup_{j=1}^M A_j | z\right) \leq 1 - \lambda.$$

Так как $\Pr(T_Y(z)) \geq 1 - \varepsilon$, то из последнего неравенства следует, что

$$\Pr\left(y \notin \bigcup_{j=1}^M A_j | z\right) \leq 1 - \lambda + \varepsilon$$

для любой последовательности $z \in T_Z$ такой, что $z \notin C$. С другой стороны, по построению кода $\Pr(\bar{A}_i | z_i) \leq \lambda$, $i = \overline{1, M}$, и, следовательно,

$$\Pr\left(y \notin \bigcup_{j=1}^M A_j | z_i\right) \leq \lambda.$$

Учитывая теперь, что $\Pr(T_Z) \geq 1 - \delta$, получим

$$\begin{aligned} \Pr(A_0) &= \Pr\left(y \notin \bigcup_{j=1}^M A_j\right) = \sum_{z \in T_Z} p(z) \Pr\left(y \notin \bigcup_{j=1}^M A_j | z\right) = \\ &= \sum_{z \in T_Z: z \notin C} p(z) \Pr\left(y \notin \bigcup_{j=1}^M A_j | z\right) + \\ &\quad + \sum_{z \in T_Z: z \in C} p(z) \Pr\left(y \notin \bigcup_{j=1}^M A_j | z\right) + \\ &\quad + \sum_{z \in \bar{T}_Z} p(z) \Pr\left(y \notin \bigcup_{j=1}^M A_j | z\right) \leq (1 - \lambda + \varepsilon) \Pr(z \notin C) + \\ &\quad + \lambda \Pr(z \in C) + \delta \leq (1 - \lambda) \Pr(z \notin C) + \lambda \Pr(z \in C) + 2\varepsilon. \end{aligned}$$

Таким образом, неравенство 2 также имеет место.

Докажем теперь выполнение неравенства 3. Для этого нам понадобится так называемое сильное обращение теоремы кодирования в дискретном канале без памяти. В интересующем нас случае сильное обращение может быть сформулировано следующим образом. Для любой последовательности кодов $G(n, R)$, $n = 1, 2, \dots$, при $R \triangleq n^{-1} \log M = I(Y; Z) + 4\delta$, где δ — произвольное положительное число, и при условии, что слова кода $G(n, R)$ принадлежат T_Z , вероятность ошибки стремится к единице с ростом n . Из этого утверждения следует, что для любого $\lambda < 1$ при достаточно большом n количество слов построенного выше максимального кода удовлетворяет неравенству 3. Таким образом, для завершения доказательства леммы необходимо доказать справедливость сформулированного здесь сильного обращения.

Пусть $T_Y \triangleq \bigcup_{z \in T_Z} T_Y(z)$. Из условия 2 леммы 6.2.6 и теоремы о высоковероятных множествах дискретных источников без памяти следует, что

$$|T_Y| \leq 2^{n(H(Y) + \delta)}. \quad (6.2.22)$$

Кроме того, из условий 1 и 3 указанной леммы следует, что для любой пары (y, z) , $z \in T_Z$, $y \in T_Y(z)$

$$p(y|z) \leq 2^{-n(H(Y|Z) - 2\delta)}. \quad (6.2.23)$$

Предположим, что для построенного выше кода $M \geq 2^{n(1(Y; Z) + 4\delta)}$. Тогда из неравенства (6.2.22) вытекает, что по крайней мере для одного i

$$|A_i| \leq \frac{|T_Y|}{M} \leq 2^{n(H(Y|Z) - 3\delta)}. \quad (6.2.24)$$

Из (6.2.23) и (6.2.24) и того, что $A_i \subseteq T_Y(z_i)$, получим

$$1 - \lambda \leq \Pr(A_i | z_i) = \sum_{y \in A_i} p(y | z_i) \leq 2^{-n\delta},$$

т. е. $\lambda \rightarrow 1$ при $n \rightarrow \infty$. Лемма доказана.

Теорема 6.2.2 (прямая теорема кодирования). Пусть U_X, U_Y — пара зависимых источников без памяти и $\{XY, p(x, y)\}$ — совместный ансамбль сообщений. Любая пара скоростей из \mathbb{R}_+^2 допустима при кодировании источника U_X с дополнительной информацией.

Доказательство. Как уже отмечалось ранее, для доказательства теоремы достаточно показать, что для любого ансамбля $XYZ \in \mathcal{A}$ и любого $\delta' > 0$ найдется n и отображение $w(y)$ множества Y^n на множество $W = \{1, 2, \dots, M_Y\}$, для которого $\frac{1}{n} H(W) \leq I(Y; Z) + \delta'$ и $\frac{1}{n} H(X^n | W) \leq H(X | Z) + \delta'$. Для построения такого отображения воспользуемся кодом леммы 6.2.7. Положим $W = \{0, 1, 2, \dots, M\}$ и $w(y) = i$, если $y \in A_i$, $i = \overline{0, M}$, тогда из условия 3 этой леммы сразу получим

$$H(W) \leq \log(M+1) \leq n(I(Y; Z) + 5\delta) \quad (6.2.25)$$

и, следовательно, для любого δ' найдется n такое, что $\frac{1}{n} H(W) \leq I(V; Z) + \delta'$.

Оценим теперь энтропию $H(X^n | W)$. Положим

$$T_X(z_i) \triangleq \bigcup_{T_Y(z_i)} T_X(y, z_i), \quad i = 1, M,$$

где z_i — слова кода леммы 6.2.7 и множества $T_Y(z_i)$, $T_X(y, z_i)$ определены в лемме 6.2.6. Введем обозначения:

$$q \triangleq \Pr(\bar{T}_X(z_i) | w(y) = i); \quad p'(x) \triangleq p(x | w(y) = i)/q.$$

Тогда

$$\begin{aligned} H(X^n | w(y) = i) &= - \sum_{T_X(z_i)} p(x | w(y) = i) \log p(x | w(y) = i) - \\ &\quad - \sum_{\bar{T}_X(z_i)} p(x | w(y) = i) \log p(x | w(y) = i) = \\ &= - \sum_{T_X(z_i)} p(x | w(y) = i) \log p(x | w(y) = i) - q \log q - \\ &\quad - q \sum_{\bar{T}_X(z_i)} p'(x) \log p'(x) \leq \log(|T_X(z_i)| + 1) + q \log(|\bar{T}_X(z_i)|). \end{aligned} \quad (6.2.26)$$

Из определения множества $T_X(z_i)$, утверждения 8 леммы 6.2.6 и из того, что $A_i \subseteq T_Y(z_i)$ и $p(x | yz_i) = p(x | y)$, имеем

$$\begin{aligned} \Pr(T_X(z_i) | w(y) = i) &\triangleq \frac{\sum_{y \in A_i} p(y)}{\sum_{y \in A_i} p(y)} \geq \\ &\geq \frac{\sum_{y \in A_i} \Pr(T_X(y, z_i) | y) p(y)}{\sum_{y \in A_i} p(y)} \geq 1 - \delta. \end{aligned} \quad (6.2.27)$$

Из утверждений 1 и 4 леммы 6.2.6 следует, что для всех $x \in T_X(z_i)$

$$p(x | z_i) \geq 2^{-n(H(X | Z) + 2\delta)}.$$

Из последнего неравенства нетрудно вывести, что

$$|T_X(z_i)| \leq 2^{n(H(X | Z) + 2\delta)}. \quad (6.2.28)$$

Подставляя (6.2.27), (6.2.28) в (6.2.26) и учитывая, что $|\bar{T}_X(z_i)| \leq |X^n|$, получим при достаточно большом n

$$H(X^n | w(y) = i) \leq n(H(X | Z) + 3\delta) + \delta n \log |X| \quad (6.2.29)$$

для всех $i = \overline{1, M}$. В случае $i = 0$ можно использовать оценку

$$H(X^n | w(y) = 0) \leq n \log |X|. \quad (6.2.30)$$

Так как $p(w(y) = 0) = \Pr(A_0)$, то из (6.2.29) и (6.2.30) следует, что

$$\frac{1}{n} H(X^n | W) \leq H(X | Z) + 3\delta + \delta \log |X| + \Pr(A_0) \log |X|. \quad (6.2.31)$$

Для того чтобы завершить построение оценки энтропии $H(X^n | W)$, заметим, что для кода из леммы 6.2.7 имеет место неравенство

$$\Pr(z \in C) \leq M 2^{-n(H(Z) - \delta)} \leq 2^{-n(H(Y | Z) - 5\delta)}. \quad (6.2.32)$$

Другими словами, если $5\delta < H(Y | Z)$, то $\Pr(z \in C) \rightarrow 0$ при $n \rightarrow \infty$. Полагая теперь $\lambda = 1 - \delta$ для кода леммы 6.2.7 и учитывая неравенство 2 этой леммы, из (6.2.31) и (6.2.32) получим, что для любого δ' найдется n такое, что

$$\frac{1}{n} H(X^n | W) \leq H(X | Z) + \delta'.$$

Теорема доказана.

§ 6.3. Кодирование в каналах с множественным доступом

6.3.1. Постановка задачи. Предположим, что имеется несколько передающих станций, которые должны передать свои сообщения одному и тому же приемнику. Предположим также, что всем станциям выделен один канал связи и на каждой из них имеется свое кодирующее устройство. Сигналы различных станций, вообще говоря, мешают друг другу. Кроме того, в канале может действовать шум, который также искажает передаваемые сигналы. Тем не менее, приемник должен определить, какое из сообщений передавала каждая станция.

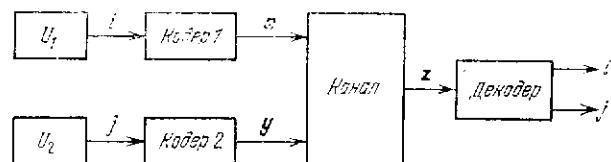


Рис. 6.3.1. Канал с множественным доступом.

Один из возможных методов передачи в такой ситуации состоит в разделении времени, когда каждой станции выделяется некоторый отрезок времени, в течение которого эта станция передает свои сообщения. Другая возможность состоит в том, чтобы разумным образом выбрать коды для каждой из станций и вести передачу всем станциям в одно и то же время. Ясно, что для каждой станции количество двоичных символов, передаваемых в единицу времени (т. е. скорость передачи), будет тем больше, чем с меньшей скоростью передают другие мешающие ей станции. Таким образом, в зависимости от канала (т. е. в зависимости от того, как взаимодействуют сигналы различных станций) и от шума в канале возможны передачи с различными наборами скоростей R_1, \dots, R_N для N одновременно работающих станций. С точки зрения теории информации основная задача анализа таких систем состоит в определении всех наборов скоростей R_1, \dots, R_N , при которых возможно передавать сообщения N станций по одному каналу со сколь угодно малой вероятностью ошибки.

Системы передачи, в которых имеется несколько передающих станций, один канал связи, доступный всем станциям, и один приемник, который должен определить, что передавала каждая станция, называются *системами передачи по каналу с множественным доступом* (КМД). На рис. 6.3.1 показана схема передачи по такому каналу в случае двух станций. Источники сообщений U_1, U_2 порождают сообщения i, j соответственно. Эти сообщения отображаются кодерами в кодовые слова \mathbf{x}, \mathbf{y} , которые и подаются на

вход канала. На выходе канала появляется последовательность \mathbf{z} , которая поступает в декодер. На выходах декодера появляются оценки i', j' передававшихся сообщений.

В дальнейшем для упрощения изложения будет рассматриваться только случай двух станций (двух пользователей). Результаты, относящиеся к обобщению на случай более чем двух пользователей, почти очевидны и поэтому отнесены в задачи.

Обозначим через X и Y множества входных сигналов КМД первого и второго пользователя соответственно. В дальнейшем рассмотрении оба множества (оба входных алфавита КМД) будут дискретными множествами с конечным, не обязательно одинаковым, количеством элементов. Обозначим через Z множество выходных сигналов канала, которое также будет предполагаться конечным.

Передача по КМД в каждый момент времени задается набором переходных вероятностей $\{p(z|xy)\}$, $(x, y, z) \in XYZ$. Процесс передачи в n моментов времени будем задавать посредством вероятностей $p(\mathbf{z}|\mathbf{xy})$ появления на выходе канала последовательности $\mathbf{z} \in Z^n$, если на первом входе была последовательность \mathbf{x} , а на втором — последовательность \mathbf{y} , причем будем полагать, что

$$p(\mathbf{z}|\mathbf{xy}) \triangleq \prod_{i=1}^n p(z^{(i)}|x^{(i)}y^{(i)}) \quad (6.3.1)$$

для всех $\mathbf{x} \in X^n$, $\mathbf{y} \in Y^n$, $\mathbf{z} \in Z^n$. Другими словами, мы будем рассматривать только дискретные КМД без памяти.

Канал, задаваемый переходными вероятностями $p(z|xy)$, можно рассматривать с общих позиций как канал $\{XYZ, p(z|xy)\}$, входным алфавитом которого является множество XY , выходным — множество Z , а переходные вероятности которого суть $\{p(z|xy)\}$. Однако специфика КМД состоит в том, что сообщения, передаваемые по этому каналу, кодируются не одним, а двумя независимо работающими кодерами, каждый из которых порождает кодовые слова в своем алфавите X и Y соответственно.

Определение 6.3.1. Кодом длины n со скоростями R_1 и R_2 для КМД называется совокупность $G_n(R_1, R_2) \triangleq \{\mathbf{x}_i, \mathbf{y}_j, A_{ij}\}$, $\mathbf{x}_i \in X^n$, $\mathbf{y}_j \in Y^n$, $A_{ij} \subseteq Z^n$, $i = 1, M_1 = 2^{nR_1}$, $j = 1, M_2 = 2^{nR_2}$, причем множества A_{ij} и $A_{i'j'}$ не пересекаются при $(i, j) \neq (i', j')$. Последовательности \mathbf{x}_i и \mathbf{y}_j , $i = 1, M_1$, $j = 1, M_2$, называются *кодовыми словами* первого и второго пользователей соответственно, а множества A_{ij} называются *решающими областями*.

Легко видеть, что код $G_n(R_1, R_2)$ для КМД является одновременно и кодом $G(n, R)$, $R = R_1 + R_2$, для обычного дискретного канала с переходными вероятностями $p(z|u) = p(z|xy)$, $u = xy$, $u \in U = XY$. Заметим, что каждое слово произвольного кода $G(n, R) = \{\mathbf{u}_k, A_k\}$, $\mathbf{u}_k \in U^n = X^nY^n$, $A_k \subseteq Z^n$, $k =$

$= \overline{1}$, $M = 2^{nR}$, представимо в виде $\mathbf{u}_k = (\mathbf{x}_k, \mathbf{y}_k)$. Между парами индексов (i, j) для слов кода $G_n(R_1, R_2)$ и между индексами слов кода $G(n, R)$ может быть установлено взаимно однозначное соответствие. Заметим однако, что кодовые слова кода $G_n(R_1, R_2)$ должны удовлетворять следующему ограничению: для двух различных пар индексов (i, j) и (i', j') , но таких, что $i = i'$, последовательности \mathbf{x}_i и $\mathbf{x}_{i'}$, входящие в кодовые слова $(\mathbf{x}_i \mathbf{y}_j)$ и $(\mathbf{x}_{i'} \mathbf{y}_{j'})$, должны совпадать (аналогично, если $j = j'$, то должны совпадать \mathbf{y}_j и $\mathbf{y}_{j'}$). В случае же произвольного кода $G(n, R)$ это ограничение отсутствует и при любых различных индексах k и k' в коде $G(n, R)$ возможно одновременное выполнение неравенств $\mathbf{x}_k \neq \mathbf{x}_{k'}$ и $\mathbf{y}_k \neq \mathbf{y}_{k'}$. Наличие указанного ограничения на множество кодовых слов кода $G_n(R_1, R_2)$ определяет специфику задачи кодирования в канале с множественным доступом.

Для каждого кода $G_n(R_1, R_2)$ определяется обычным образом средняя вероятность λ ошибки декодирования

$$\lambda \triangleq M_1^{-1} M_2^{-1} \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} \sum_{z \in A_{ij}} p(z | \mathbf{x}_i \mathbf{y}_j).$$

Определение 6.3.2. Пара скоростей (R_1, R_2) называется допустимой для КМД, если для любых $\epsilon > 0$ и $\delta > 0$ найдется $n_0(\epsilon, \delta)$ такое, что для всех $n > n_0$ существует код $G_n(\tilde{R}_1, \tilde{R}_2)$, $\tilde{R}_i = \max(0, R_i - \delta)$, $i = 1, 2$, с вероятностью ошибки декодирования $\lambda < \epsilon$. Множество \mathbb{S} , состоящее из всех пар допустимых скоростей, называется областью пропускной способности КМД.

Основной задачей настоящего параграфа является характеристизация области пропускной способности \mathbb{S} . Заметим, что как и в задачах кодирования источников, при кодировании в КМД можно использовать метод разделения времени, в соответствии с которым из допустимости пар $(R'_1, R'_2), (R''_1, R''_2)$ следует допустимость пары (R_1, R_2) , где $R_1 = \alpha R'_1 + (1 - \alpha) R''_1$, $R_2 = \alpha R'_2 + (1 - \alpha) R''_2$ для любого α из интервала $[0, 1]$. Поэтому граница области пропускной способности является выпуклой вверх кривой.

Замечание. Как и в случае кодов для обычных каналов, для кода $G_n(R_1, R_2)$ для КМД помимо средней вероятности ошибки можно определить также и максимальную вероятность ошибки. Однако в случае КМД лемма 3.2.1, связывающая среднюю вероятность ошибки кода и максимальную вероятность ошибки подкода, не справедлива. Поэтому при фиксированном отображении множества сообщений на множество кодовых слов из допустимости пары скоростей относительно средней вероятности ошибки не следует допустимость этой пары относительно максимальной вероятности ошибки.

Для пояснения особенностей кодирования в КМД и построения области пропускной способности ниже будет рассмотрен частный, но весьма показательный пример канала с множественным доступом, а именно пример двоичного суммирующего КМД.

6.3.2. Двоичный суммирующий КМД. Пусть $X = \{0, 1\}$ и $Y = \{0, 1\}$ — входные алфавиты и $Z = \{0, 1, 2\}$ — выходной алфавит КМД, показанного на рис. 6.3.2, a. Передача по этому каналу задается следующим образом. Если на первый вход канала

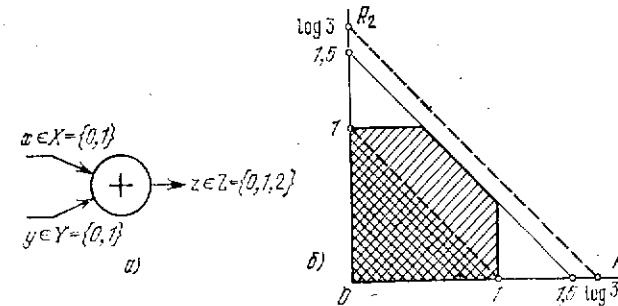


Рис. 6.3.2. Двоичный суммирующий канал с множественным доступом: логическая схема (a), область допустимых пар скоростей (б).

поступает сигнал $x \in X$, а на второй — сигнал $y \in Y$, то на выходе канала возникает сигнал $z = x + y$, принимающий значения 0, 1 или 2 в зависимости от значений x и y . Таким образом, в рассматриваемом КМД шума нет, а имеется только взаимное влияние сигналов пользователей. Такой канал с множественным доступом называется двоичным суммирующим (ДСКМД). Он имеет следующие переходные вероятности:

$$p(0 | xy) = \begin{cases} 1, & \text{если } x = y = 0 \\ 0 & \text{для остальных } x, y; \end{cases}$$

$$p(1 | xy) = \begin{cases} 1, & \text{если } x = 1, y = 0 \text{ или } x = 0, y = 1, \\ 0 & \text{для остальных } x, y; \end{cases} \quad (6.3.2)$$

$$p(2 | xy) = \begin{cases} 1, & \text{если } x = y = 1, \\ 0 & \text{для остальных } x, y. \end{cases}$$

ДСКМД можно рассматривать как обычный канал без памяти с четверичным входом $\{00, 01, 10, 11\}$, троичным выходом $\{0, 1, 2\}$ (см. рис. 6.3.3) и переходными вероятностями (6.3.2). Пропускную способность такого канала легко вычислить. Она равна

$$C = \max_{p(x, y)} I(XY; Z) = \max_{p(x, y)} H(Z) - \log 3 = 1,58, \quad (6.3.3)$$

где использовано то, что $H(Z|XY) = 0$ ввиду отсутствия шума. Максимум в (6.3.3) достигается на таком распределении вероятностей $p(x, y)$, для которого $p(00) = p(11) = p(01) + p(10) = 1/3$. Таким образом, если бы кодер располагал сообщениями обоих источников, он мог бы закодировать эти сообщения так, чтобы суммарная скорость $R = R_1 + R_2$ равнялась $\log 3$ — своему наибольшему возможному значению.

Вместе с тем для ДСКМД пары скоростей (R_1, R_2) , такая, что $R = R_1 + R_2 \leq \log 3$, не является допустимой. Это может быть

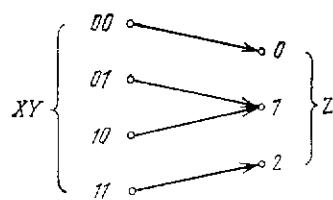


Рис. 6.3.3. ДСКМД как обычный канал с четверичным входом и троичным выходом.

канала появляется $z = 1$, то приемник не может определить, передавалась ли пара (01) или (10). С вероятностной точки зрения независимость работы кодеров в КМД соответствует (как мы покажем в следующем подразделе) тому, что при определении максимально допустимой в нем суммарной скорости передачи максимум в (6.3.3) должен отыскиваться не по всем распределениям $p(x, y)$, а лишь по таким, для которых $p(x, y) = p_1(x)p_2(y)$. Таким образом, суммарная скорость $R = R_1 + R_2$ в ДСКМД ограничена следующим образом:

$$R = R_1 + R_2 \leq \max_{p(x,y)=p_1(x)p_2(y)} I(XY; Z) = \max_{p(x,y)=p_1(x)p_2(y)} H(Z). \quad (6.3.4)$$

Нетрудно видеть, что максимум в (6.3.4) достигается при $p_1(x = 0) = p_2(y = 0) = 0,5$. При этом для любой допустимой пары (R_1, R_2) $R = R_1 + R_2 \leq 1,5 < \log 3$.

Помимо ограничения на максимальную суммарную скорость, в КМД существуют ограничения и на максимальные значения скоростей R_1 и R_2 . Для ДСКМД эти ограничения могут быть получены из следующих соображений. Так как $X = Y = \{0, 1\}$, то очевидно, что $R_1 \leq 1$ и $R_2 \leq 1$. Зафиксируем сигнал y на втором входе ДСКМД. Тогда количество информации, передаваемое по первому входу при распределении $p_1(x)$ на нем, равно $I(X; Z|y) \leq \log 2 = 1$, где равенство достигается при $p_1(x = 0) = 0,5$ для любого y . Теперь нетрудно видеть, что ограничения $R_1 \leq 1$ и

$R_2 \leq 1$ могут быть записаны в виде $R_1 \leq I(X; Z|Y)$, $R_2 \leq I(Y; Z|X)$ при $p_1(x = 0) = p_2(y = 0) = 0,5$.

Мы установили, что любая допустимая пара скоростей в ДСКМД должна удовлетворять трем неравенствам

$$R_1 + R_2 \leq I(XY; Z); \quad R_1 \leq I(X; Z|Y); \quad R_2 \leq I(Y; Z|X) \quad (6.3.5)$$

при $p(x, y) = p_1(x)p_2(y)$ и $p_1(x = 0) = p_2(y = 0) = 0,5$. Множество пар скоростей (R_1, R_2) , удовлетворяющих неравенствам (6.3.5), представляет собой часть плоскости, ограниченную прямыми $R_1 + R_2 = I(XY; Z) = 1,5$, $R_1 = I(X; Z|Y) = 1$ и $R_2 = I(Y; Z|X) = 1$ (см. рис. 6.3.2, б).

Покажем теперь, что любая пара скоростей (R_1, R_2) , удовлетворяющая условиям (6.3.5), действительно допустима при передаче по ДСКМД. Заметим вначале, что в силу возможности использования разделения времени достаточно доказать допустимость только пар $R_1 = 1$, $R_2 = 0,5$ и $R_1 = 0,5$, $R_2 = 1$.

Рассмотрим первую пару скоростей $(R_1 = 1, R_2 = 0,5)$. Пусть фиксировано произвольное положительное число δ . Построим код $G_n(\tilde{R}_1, \tilde{R}_2)$, $\tilde{R}_1 = 1$, $\tilde{R}_2 = 0,5 - \delta$ для рассматриваемого канала. Для этого в качестве кодовых слов \mathbf{x}_i , $i = \overline{1, 2^n}$, выберем все двоичные последовательности длины n , а в качестве кодовых слов \mathbf{y}_j , $j = \overline{1, M_2 = 2^{n\tilde{R}_2}}$, выберем слова кода для двоичного стирающего канала, в котором вероятность стирания равна $q = 1/2$, а вероятность ошибки равна 0. Как известно (см. § 3.6), пропускная способность такого канала равна $1 - q = 1/2$ поэтому для любых $\epsilon > 0$, $\delta > 0$ при достаточно большом n найдется код со скоростью $\tilde{R}_2 = 1/2 - \delta$ и вероятностью ошибки, не превосходящей ϵ .

Работа декодера в ДСКМД при указанном выборе кодовых слов происходит следующим образом. Выходной сигнал канала $z = 1$ рассматривается как стирание (если $z = 0$ или $z = 2$, то по выходному сигналу канала можно однозначно определить его входные сигналы, а при $z = 1$ однозначное определение невозможно). Очевидно, что вероятность стирания равна вероятности того, что $x \neq y$. Если все кодовые слова \mathbf{x}_i , $i = \overline{1, 2^n}$, используются с одинаковыми вероятностями, то $p(z = 1) = p(x \neq y) = 1/2$ и поэтому, наблюдая выходную последовательность \mathbf{z} , декодер может указать, какая из последовательностей \mathbf{y}_j была на входе канала. При этом вероятность ошибки не будет превосходить ϵ . После того как последовательность \mathbf{y}_j найдена, декодер находит разность $\mathbf{z} - \mathbf{y}_j$. Эта разность и есть декодированный вариант последовательности \mathbf{x}_i . Очевидно, что декодер сделает ошибку только в том случае, когда неправильно определит последовательность \mathbf{y}_j , т. е. вероятность ошибки для кода $G_n(\tilde{R}_1, \tilde{R}_2)$ не превосходит ϵ . Сле-

довательно, пара скоростей ($R_1 = 1, R_2 = 0,5$) допустима. Аналогичным образом доказывается допустимость пары ($R_1 = 0,5, R_2 = 1$).

Таким образом, область пропускной способности для ДСКМД — это заштрихованная область на рис. 6.3.2, б). Если бы обоим кодерам были доступны сообщения обоих источников, то сообщения можно было бы передавать с любой парой скоростей (R_1, R_2), представляющей собой точку, лежащую ниже пунктирной линии $R_1 + R_2 = \log 3$. Наконец, если просто делить время между двумя пользователями, кодеры которых используют двоичные алфавиты, то допустимыми были бы только пары (R_1, R_2), удовлетворяющие условию $R_1 + R_2 \leq 1$ (область с двойной штриховкой).

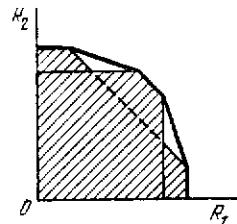


Рис. 6.3.4. Замыкание области допустимых пар скоростей.

6.3.3. Обратная теорема кодирования. Пусть $\{XY Z, p(z|xy)\}$ — дискретный канал без памяти с множественным доступом. Зафиксируем распределение вероятностей $p(x, y) = p_1(x)p_2(y)$ и рассмотрим множество $\mathfrak{R}(p_1, p_2)$ всех таких пар скоростей (R_1, R_2), которое определяется следующим образом:

$$\begin{aligned} \mathfrak{R}(p_1, p_2) &\triangleq \{(R_1, R_2): R_1 + R_2 \leq I(XY; Z), R_1 \leq I(X; Z|Y), \\ &R_2 \leq I(Y; Z|X); p(x, y, z) = p(z|xy)p_1(x)p_2(y)\}. \end{aligned} \quad (6.3.6)$$

Пусть \mathfrak{P} — множество всех распределений вероятностей на парах (x, y) , которые имеют вид $p(x, y) = p_1(x)p_2(y)$ и получаются при всевозможных выборах распределений $p_1(x)$ и $p_2(y)$. Положим

$$\mathfrak{R} \triangleq \bigcup_{\mathfrak{P}} \mathfrak{R}(p_1, p_2). \quad (6.3.7)$$

В рассмотренном выше примере ДСКМД имеется одно такое распределение вероятностей $\tilde{p}_1(x), \tilde{p}_2(y)$, именно то, на котором достигается максимум в (6.3.4), что $\mathfrak{R}(p_1, p_2) \subseteq \mathfrak{R}(\tilde{p}_1, \tilde{p}_2)$ для любого распределения $p_1(x)p_2(y)$, поэтому в этом примере $\mathfrak{R} = \mathfrak{R}(\tilde{p}_1, \tilde{p}_2)$. В общем же случае такого единственного распределения может не быть, и операция объединения в определении множества \mathfrak{R} требуется по существу. В дальнейшем мы покажем, что все пары скоростей $(R_1, R_2) \in \mathfrak{R}$ допустимы. Вместе с тем, как это видно из рис. 6.3.4, область \mathfrak{R} может не быть выпуклой (заштрихованная область на рисунке). Однако из возможности передачи с разделением времени следует, что всякая допустимая пара скоростей принадлежит выпуклому замыканию \mathfrak{R}^* области \mathfrak{R} (область, ограниченная жирной линией на рисунке).

В следующей лемме дается простая интерпретация множества \mathfrak{R}^* .

Лемма 6.3.1. Пусть множество U состоит из двух элементов $U = \{u_1, u_2\}$ и $p(u)$ — распределение вероятностей на U . Пусть распределение вероятностей на ансамбле $UXYZ$ задается следующим образом:

$$p(u, x, y, z) = p(z|xy)p_1(x|u)p_2(y|u)p(u). \quad (6.3.8)$$

Обозначим через \mathfrak{A} множество ансамблей $UXYZ$ с распределениями вероятностей (6.3.8), получающимся при различных выборах распределений $p_1(x|u), p_2(y|u)$ и $p(u)$. Пусть

$$\tilde{\mathfrak{R}} \triangleq \{(R_1, R_2): R_1 + R_2 \leq I(XY, Z|U), R_1 \leq I(X; Z|YU),$$

$$R_2 \leq I(Y; Z|XU) \text{ для некоторого } UXYZ \in \mathfrak{A}\}. \quad (6.3.9)$$

Тогда $\mathfrak{R}^* = \tilde{\mathfrak{R}}$.

Доказательство. Заметим сначала, что область \mathfrak{R}^* состоит либо из точек $(R_1, R_2) \in \mathfrak{R}$, либо из точек прямых, соединяющих пары точек из \mathfrak{R} . Поэтому любая пара скоростей $(R_1, R_2) \in \mathfrak{R}^*$ представима в виде $R_1 = \alpha R'_1 + (1 - \alpha) R''_1, R_2 = \alpha R'_2 + (1 - \alpha) R''_2$, где $(R'_1, R''_1) \in \mathfrak{R}, (R'_2, R''_2) \in \mathfrak{R}$ и $\alpha \in [0, 1]$. Пусть в этом представлении $(R'_1, R'_2) \in \mathfrak{R}(p'_1, p'_2)$ и $(R''_1, R''_2) \in \mathfrak{R}(p''_1, p''_2)$. Положим $p(u_1) = \alpha, p_1(x|u_1) = p'_1(x), p_2(y|u_1) = p''_1(y)$ и $p(u_2) = 1 - \alpha, p_1(x|u_2) = p'_2(x), p_2(y|u_2) = p''_2(y)$. Тогда для построенного ансамбля $UXYZ \in \mathfrak{A}$ пара скоростей $(R_1, R_2) \in \mathfrak{R}^*$ и удовлетворяет неравенствам

$$\begin{aligned} R_1 + R_2 &\leq I(XY; Z|U), \\ R_1 &\leq I(X; Z|YU), \quad R_2 \leq I(Y; Z|XU). \end{aligned} \quad (6.3.10)$$

Следовательно, $\mathfrak{R}^* \subseteq \tilde{\mathfrak{R}}$.

Покажем теперь, что $\mathfrak{R}^* \supseteq \tilde{\mathfrak{R}}$. Пусть $(R_1, R_2) \in \tilde{\mathfrak{R}}$, тогда найдется ансамбль $UXYZ \in \mathfrak{A}$, распределение вероятностей на котором удовлетворяет условию (6.3.8) и выполняются неравенства (6.3.10). Положим $\alpha = p(u_1), p'_1(x) = p_1(x|u_1), p''_1(y) = p_2(y|u_1), p'_1(x) = p_1(x|u_2), p''_2(y) = p_2(y|u_2)$. Введем в рассмотрение две пары чисел $(R'_1, R''_1), (R'_2, R''_2)$, определив их соотношениями

$$\begin{aligned} R'_1 + R''_1 &\leq I(XY; Z|u_1), \\ R'_1 &\leq I(X; Z|Yu_1), \quad R'_2 \leq I(Y; Z|Xu_1), \\ R'_1 + R''_2 &\leq I(XY; Z|u_2), \\ R''_1 &\leq I(X; Z|Yu_2), \quad R''_2 \leq I(Y; Z|Xu_2), \\ R_1 &= \alpha R'_1 + (1 - \alpha) R''_1, \quad R_2 = \alpha R'_2 + (1 - \alpha) R''_2. \end{aligned} \quad (6.3.11)$$

Совместность условий (6.3.11) вытекает из условий (6.3.9). Из (6.3.11) следует, что $(R'_1, R''_1) \in \mathfrak{R}(p'_1, p''_1), (R'_2, R''_2) \in \mathfrak{R}(p'_2, p''_2)$. Таким образом, $(R_1, R_2) \in \mathfrak{R}^*$. Лемма доказана.

Замечание. Так как множество \mathfrak{R}^* является выпуклым замыканием множества \mathfrak{R} , то для любого l , любых наборов чисел $\alpha_i > 0$, $i = \overline{1, l}$, $\sum_{i=1}^l \alpha_i = 1$ и пар скоростей $(R_1^{(i)}, R_2^{(i)}) \in \mathfrak{R}$, $i = \overline{1, l}$,

пара скоростей $\left(R_1 = \sum_{i=1}^l \alpha_i R_1^{(i)}, R_2 = \sum_{i=1}^l \alpha_i R_2^{(i)}\right)$ принадлежит \mathfrak{R}^* . Отсюда следует, что множество пар скоростей (6.3.9) не изменится, если в его определении использовать множество U с произвольным числом элементов.

Теорема 6.3.1 (обратная теорема кодирования). Пусть задан дискретный КМД без памяти, причем X , Y — входные алфавиты, Z — выходной алфавит и $\{p(z|x)\}$ — переходные вероятности. Пусть $\tilde{\mathfrak{R}}$ — множество пар скоростей, определенное соотношениями (6.3.9). Если в рассматриваемом КМД пара скоростей (R_1, R_2) допустима, то она принадлежит $\tilde{\mathfrak{R}}$, т. е. $\tilde{\mathfrak{R}} \ni \mathfrak{C}$.

Доказательство. Пусть $G_n(R_1, R_2)$ — некоторый произвольный код для КМД. Обозначим через W множество решений на выходе декодера, $W = \{(i, j) : i = \overline{1, M_1}, j = \overline{1, M_2}\}$. Обозначим через C_1 и C_2 множества кодовых слов для первого и второго кодов соответственно, $C_1 \triangleq \{\mathbf{x}_1, \dots, \mathbf{x}_{M_1}\}$, $C_2 \triangleq \{\mathbf{y}_1, \dots, \mathbf{y}_{M_2}\}$. Рассмотрим следующее распределение вероятностей на множестве входных последовательностей КМД

$$p(\mathbf{x}, \mathbf{y}) = p_1(\mathbf{x}) \cdot p_2(\mathbf{y}) = \begin{cases} M_1^{-1} M_2^{-1}, & \text{если } \mathbf{x} \in C_1, \mathbf{y} \in C_2, \\ 0 & \text{в противном случае.} \end{cases} \quad (6.3.12)$$

Это распределение вероятностей определяет ансамбль $X^n Y^n Z^n$. Декодирование для кода $G_n(R_1, R_2)$ определяет, кроме того, ансамбль W . Таким образом, распределение (6.3.12), КМД и код в совокупности определяют ансамбль $X^n Y^n Z^n W$. Для этого ансамбля

$$\begin{aligned} I(X^n Y^n; Z^n) &\geq I(X^n Y^n; W) = H(X^n Y^n) - H(X^n Y^n | W) = \\ &= n(R_1 + R_2) - H(X^n Y^n | W). \end{aligned} \quad (6.3.13)$$

Пусть λ — средняя вероятность ошибки для рассматриваемого кода $G_n(R_1, R_2)$. Тогда, используя неравенство Фано, из (6.3.13) получим

$$\begin{aligned} I(X^n Y^n; Z^n) &\geq n(R_1 + R_2) - h(\lambda) - \lambda \log M_1 M_2 \geq \\ &\geq n(R_1 + R_2) - h(\lambda) - n\lambda \log(|X| \cdot |Y|), \end{aligned} \quad (6.3.14)$$

где использованы неравенства $R_1 \leq \log |X|$, $R_2 \leq \log |Y|$. Из (6.3.14) следует, что

$$R_1 + R_2 \leq \frac{1}{n} I(X^n Y^n; Z^n) + \frac{h(\lambda)}{n} + \lambda \log(|X| \cdot |Y|). \quad (6.3.15)$$

Справедливы также следующие соотношения для ансамбля $X^n Y^n Z^n W$:

$$\begin{aligned} I(X^n; Z^n | \mathbf{y}) &\geq I(X^n; W | \mathbf{y}) = H(X^n | \mathbf{y}) - H(X^n | W \mathbf{y}) = \\ &= nR_1 - H(X^n | W \mathbf{y}), \end{aligned}$$

где использована независимость X^n и Y^n , а также что $H(X^n) = -nR_1$. Обозначим через $\lambda_1(\mathbf{y}_j)$ вероятность ошибки декодирования сообщения первого кодера при условии, что второй кодер передает слово \mathbf{y}_j , и через λ_1 — среднюю вероятность ошибки декодирования сообщений первого кодера: $\lambda_1 \triangleq \sum_{Y^n} \lambda_1(\mathbf{y}) p_2(\mathbf{y})$. Тогда

из неравенства Фано вытекают неравенства

$$\begin{aligned} I(X^n; Z^n | \mathbf{y}_j) &\geq nR_1 - H(X^n | W \mathbf{y}_j) \geq \\ &\geq nR_1 - h(\lambda_1(\mathbf{y}_j)) - n\lambda_1(\mathbf{y}_j) \log M_1 \geq \\ &\geq nR_1 - h(\lambda_1(\mathbf{y}_j)) - n\lambda_1(\mathbf{y}_j) \log |X|. \end{aligned}$$

Усредняя обе части последнего неравенства по ансамблю $\{Y^n, p_2(\mathbf{y})\}$ и используя выпуклость функции $h(x)$, получим

$$I(X^n; Z^n | Y^n) \geq nR_1 - h(\lambda_1) - n\lambda_1 \log |X|,$$

откуда следует, что

$$R_1 \leq \frac{1}{n} I(X^n; Z^n | Y^n) + \frac{h(\lambda_1)}{n} + \lambda_1 \log |X|. \quad (6.3.16)$$

Аналогичным образом выводится неравенство

$$R_2 \leq \frac{1}{n} I(Y^n; Z^n | X^n) + \frac{h(\lambda_2)}{n} + \lambda_2 \log |Y|, \quad (6.3.17)$$

где λ_2 — средняя вероятность ошибки декодирования сообщений второго кодера.

Оценим теперь информацию $I(X^n Y^n; Z^n)$, $I(X^n; Z^n | Y^n)$ и $I(Y^n; Z^n | X^n)$. Учитывая, что рассматриваемый КМД не имеет памяти, получим

$$\begin{aligned} I(X^n Y^n; Z^n) &= H(Z^n) - H(Z^n | X^n Y^n) = \\ &= H(Z^n) - \sum_{i=1}^n H(Z_i | X_i Y_i) \leq \sum_{i=1}^n I_i(X_i Y_i; Z_i) \end{aligned} \quad (6.3.18)$$

и аналогично

$$I(X^n; Z^n | Y^n) \leq \sum_{i=1}^n I_i(X_i; Z_i | Y_i), \quad (6.3.19)$$

$$I(Y^n; Z^n | X^n) \leq \sum_{i=1}^n I_i(Y_i; Z_i | X_i). \quad (6.3.20)$$

Заметим, что i -е слагаемое в каждой из сумм (6.3.18) — (6.3.20) вычисляется по распределению вероятностей $p_1^{(i)}(x, y)$ — распределению, определяемому из распределения $p(x, y)$ (см. (6.3.12)): $p_1^{(i)}(x, y) = p_1^{(i)}(x)p_2^{(i)}(y) =$

$$= \left(\sum_{x^{(1)}, \dots, x^{(i-1)}, x^{(i+1)}, \dots, x^{(n)}} p_1(x) \right) \left(\sum_{y^{(1)}, \dots, y^{(i-1)}, y^{(i+1)}, \dots, y^{(n)}} p_2(y) \right). \quad (6.3.21)$$

Можно ввести в рассмотрение множество $U = \{u_1, \dots, u_n\}$, состоящее из n элементов, и распределение вероятностей на нем, положив $p(u_i) = 1/n$. Можно также обозначить $p_1^{(i)}(x) \triangleq p_1(x|u_i)$, $p_2^{(i)}(y) \triangleq p_2(y|u_i)$. Тогда из (6.3.15) — (6.3.20) следует, что для любого кода $G_n(R_1, R_2)$ для КМД, обеспечивающего вероятность ошибки λ , выполняются следующие неравенства:

$$\begin{aligned} R_1 + R_2 &\leq I(XY; Z|U) + \frac{h(\lambda)}{n} + \lambda \log(|X| \cdot |Y|), \\ R_1 &\leq I(X; Z|YU) + \frac{h(\lambda_1)}{n} - \lambda_1 \log |X|, \\ R_2 &\leq I(Y; Z|XU) + \frac{h(\lambda_2)}{n} + \lambda_2 \log |Y|. \end{aligned} \quad (6.3.22)$$

Предположим, что пара скоростей (R_1, R_2) допустима. Тогда неравенства (6.3.22) должны выполняться для любых сколь угодно малых $\lambda > 0$. Следовательно, учитывая замечание, приведенное вслед за леммой 6.3.1, имеем $(R_1, R_2) \in \mathbb{R}$. Теорема доказана.

6.3.4. Прямая теорема кодирования. Докажем теперь, что любая пара скоростей $(R_1, R_2) \in \mathbb{R}(p_1, p_2)$ при произвольном выборе распределений вероятностей $p(x, y) = p_1(x)p_2(y)$ допустима. В силу возможности использования разделения времени отсюда будет следовать, что допустима любая пара скоростей из \mathbb{R}^* или, что то же самое, из \mathbb{R} .

Доказательство допустимости пары $(R_1, R_2) \in \mathbb{R}(p_1, p_2)$ мы будем проводить методом случайного кодирования, аналогичным тому, который использовался при построении верхней границы вероятности ошибки для дискретных каналов без памяти (см. § 3.12). Для этого введем в рассмотрение ансамбль $\mathfrak{G}_n(R_1, R_2)$ кодов $G_n(R_1, R_2)$, распределение вероятностей на котором зададим следующим образом.

Пусть распределения вероятностей $p_1(x)$ и $p_2(y)$ фиксираны. Положим

$$p_1(\mathbf{x}) = \prod_{i=1}^n p_1(x^{(i)}), \quad p_2(\mathbf{y}) = \prod_{i=1}^n p_2(y^{(i)}) \quad (6.3.23)$$

для всех последовательностей $\mathbf{x} \in X^n$, $\mathbf{y} \in Y^n$.

Вероятность кода $G_n(R_1, R_2) = \{\mathbf{x}_i, \mathbf{y}_j, A_{ij}\}$, $i = \overline{1, M_1}$, $j = \overline{1, M_2}$, $M_1 = 2^{nR_1}$, $M_2 = 2^{nR_2}$, в ансамбле $\mathfrak{G}_n(R_1, R_2)$ положим равной

$$p(G_n) \triangleq \prod_{i=1}^{M_1} p_1(\mathbf{x}_i) \prod_{j=1}^{M_2} p_2(\mathbf{y}_j). \quad (6.3.24)$$

Такое задание распределения вероятностей на ансамбле кодов соответствует независимому выбору кодовых слов для первого и второго кодеров.

Заметим, что в рассматриваемом ансамбле кодов не все пары кодовых слов независимы. Действительно, пусть $i = i'$, $j \neq j'$. Рассмотрим две пары кодовых слов $(\mathbf{x}_i, \mathbf{y}_j)$ и $(\mathbf{x}_{i'}, \mathbf{y}_{j'})$. Вероятность того, что при случайном выборе кодовых слов будут выбраны слова $\mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_{i'}, \mathbf{y}_{j'}$, есть

$$p(\mathbf{x}_i, \mathbf{y}_j, \mathbf{x}_{i'}, \mathbf{y}_{j'}) = p(\mathbf{x}_i)p(\mathbf{y}_j)p(\mathbf{y}_{j'}).$$

Эта вероятность не равна произведению вероятностей

$$p(\mathbf{x}_i, \mathbf{y}_j) \cdot p(\mathbf{x}_{i'}, \mathbf{y}_{j'}) = p^2(\mathbf{x}_i)p(\mathbf{y}_j)p(\mathbf{y}_{j'}).$$

Аналогичная ситуация имеет место и при $i \neq i'$, $j = j'$. Вместе с тем, если $j \neq j'$ и $j \neq j'$, то пары кодовых слов $(\mathbf{x}_i, \mathbf{y}_j)$ и $(\mathbf{x}_{i'}, \mathbf{y}_{j'})$ независимы. Наличие зависимости между некоторыми парами кодовых слов вносит дополнительные, правда легко преодолимые, трудности.

Пусть фиксирован код $G_n(R_1, R_2)$. Обозначим через λ_{ij} условную вероятность ошибки при передаче кодовых слов $(\mathbf{x}_i, \mathbf{y}_j)$. Тогда

$$\lambda_{ij} \triangleq \sum_{z \in A_{ij}} p(z|\mathbf{x}_i \mathbf{y}_j) = \sum_{z \in z^n} p(z|\mathbf{x}_i \mathbf{y}_j) \varphi_{ij}(z), \quad (6.3.25)$$

где

$$\varphi_{ij}(z) = \begin{cases} 1, & \text{если } z \in A_{ij}, \\ 0 & \text{в противном случае.} \end{cases} \quad (6.3.26)$$

В дальнейшем мы будем предполагать, что декодирование производится по максимуму правдоподобия, т. е. декодер выбирает такую пару кодовых слов, для которой вероятность $p(z|\mathbf{x}_i \mathbf{y}_j)$ максимальна. Другими словами, решающая область A_{ij} состоит из таких последовательностей z , что $p(z|\mathbf{x}_i \mathbf{y}_j) > p(z|\mathbf{x}_{i'} \mathbf{y}_{j'})$ для всех пар $(i', j') \neq (i, j)$.

Оценим теперь функцию $\varphi_{ij}(z)$ подобно тому, как это делалось в параграфе 3.12. Имеет место следующее неравенство

$$\begin{aligned} \varphi_{ij}(z) &\leq \left(\frac{\sum_{i' \neq i, j' \neq j} p^{1/(1+\rho_{12})}(z|\mathbf{x}_{i'} \mathbf{y}_{j'})}{p^{1/(1+\rho_{12})}(z|\mathbf{x}_i \mathbf{y}_j)} \right)^{\rho_{12}} + \\ &+ \left(\frac{\sum_{i' \neq i} p^{1/(1+\rho_1)}(z|\mathbf{x}_{i'} \mathbf{y}_j)}{p^{1/(1+\rho_1)}(z|\mathbf{x}_i \mathbf{y}_j)} \right)^{\rho_1} + \left(\frac{\sum_{j' \neq j} p^{1/(1+\rho_2)}(z|\mathbf{x}_i \mathbf{y}_{j'})}{p^{1/(1+\rho_2)}(z|\mathbf{x}_i \mathbf{y}_j)} \right)^{\rho_2}, \end{aligned}$$

где $\rho_1 \geq 0$, $\rho_2 \geq 0$, $\rho_{12} \geq 0$. Действительно, если $z \notin A_{ij}$, то по крайней мере в одном слагаемом числитель не меньше, чем знаменатель, и поэтому правая часть этого неравенства не меньше чем 1. Подставляя оценку функции $\varphi_{ij}(z)$ в (6.3.25), получим, что

$$\lambda_{ij} \leq \lambda^{(12)} + \lambda^{(1)} + \lambda^{(2)}, \quad (6.3.27)$$

где

$$\lambda^{(12)} \triangleq \sum_{Z^n} p^{1/(1+\rho_{12})}(z|x_i y_j) \left(\sum_{i' \neq i} \sum_{j' \neq j} p^{1/(1+\rho_{12})}(z|x_{i'} y_{j'}) \right)^{\rho_{12}},$$

$$\lambda^{(1)} \triangleq \sum_{Z^n} p^{1/(1+\rho_1)}(z|x_i y_j) \left(\sum_{i' \neq i} p^{1/(1+\rho_1)}(z|x_{i'} y_j) \right)^{\rho_1},$$

$$\lambda^{(2)} \triangleq \sum_{Z^n} p^{1/(1+\rho_2)}(z|x_i y_j) \left(\sum_{j' \neq j} p^{1/(1+\rho_2)}(z|x_i y_{j'}) \right)^{\rho_2}.$$

В дальнейшем при доказательстве теоремы кодирования мы будем усреднять обе части неравенства (6.3.27) по ансамблю $\mathfrak{G}_n(R_1, R_2)$. При этом представление правой части в виде трех слагаемых позволяет избавиться от влияния зависимости некоторых пар кодовых слов.

Теорема 6.3.2 (прямая теорема кодирования). Пусть фиксированы дискретный КМД без памяти и распределение вероятностей $p(x, y) = p_1(x)p_2(y)$. Любая пара скоростей $(R_1, R_2) \in \mathcal{R}(\rho_1, \rho_2)$ допустима.

Доказательство. Обозначим через $\bar{\lambda}^{(12)}$, $\bar{\lambda}^{(1)}$, $\bar{\lambda}^{(2)}$ средние значения величин $\lambda^{(12)}$, $\lambda^{(1)}$, $\lambda^{(2)}$ по ансамблю кодов $\mathfrak{G}_n(R_1, R_2)$.

При усреднении будем учитывать что:

а) пары кодовых слов (x_i, y_j) и $(x_{i'}, y_{j'})$ при $i \neq i'$, $j \neq j'$ статистически независимы в ансамбле $\mathfrak{G}_n(R_1, R_2)$;

б) при фиксированном слове y_j кодовые слова x_i и $x_{i'}$, $i \neq i'$, статистически независимы в ансамбле $\mathfrak{G}_n(R_1, R_2)$;

в) при фиксированном слове x_i кодовые слова y_j и $y_{j'}$, $j \neq j'$, статистически независимы в ансамбле $\mathfrak{G}_n(R_1, R_2)$;

г) для всех $\rho \in [0, 1]$ имеет место неравенство $\bar{\xi}^\rho \leq \bar{\xi}^0$, где ξ — произвольная случайная величина и черта означает усреднение;

д) средние по ансамблю кодов значения вероятностей $\lambda^{(12)}$, $\lambda^{(1)}$, $\lambda^{(2)}$ не зависят от индексов i и j .

Обозначая через $\bar{\lambda}^{(12)}$ результат усреднения $\lambda^{(12)}$ по ансамблю $\mathfrak{G}_n(R_1, R_2)$ и осуществляя преобразования, аналогичные тем,

которые производились в пунктах 3.12.2 и 3.12.3, получим

$$\bar{\lambda}^{(12)} \leq 2^{-nE_{12}(R_1, R_2)}, \quad (6.3.28)$$

где

$$E_{12}(R_1, R_2) = \max_{0 \leq \rho_{12} \leq 1} (E_0^{(12)}(\rho_{12}) - \rho_{12}(R_1 + R_2)),$$

$$E_0^{(12)}(\rho_{12}) = -\log \sum_Z \left(\sum_X \sum_Y p_1(x)p_2(y) p^{1/(1+\rho_{12})}(z|xy) \right)^{1+\rho_{12}}.$$

Для того чтобы оценить $\bar{\lambda}^{(1)}$, обозначим через $\bar{\lambda}^{(1)}(y_j)$ условное математическое ожидание величины $\lambda^{(1)}$ при фиксированном кодом слове y_j . Произведя операцию усреднения, получим

$$\bar{\lambda}^{(1)}(y_j) \leq M_1^{p_1} \sum_{Z^n} \left(\sum_{X^n} p(x)p(z|xy_j) \right)^{1+p_1}, \quad 0 < p_1 \leq 1. \quad (6.3.29)$$

Усредняя затем по Y^n и выполняя стандартные преобразования, получим

$$\bar{\lambda}^{(1)} \leq 2^{-nE_1(R_1)}, \quad (6.3.30)$$

где

$$E_1(R_1) = \max_{0 \leq \rho_1 \leq 1} (E_0^{(1)}(\rho_1) - \rho_1 R_1),$$

$$E_0^{(1)}(\rho_1) = -\log \sum_Z \sum_Y p_2(y) \left(\sum_X p_1(x) p^{1/(1+\rho_1)}(z|xy) \right)^{1+\rho_1}.$$

Аналогичным образом можно получить, что

$$\bar{\lambda}^{(2)} \leq 2^{-nE_2(R_2)}, \quad (6.3.31)$$

где

$$E_2(R_2) = \max_{0 \leq \rho_2 \leq 1} (E_0^{(2)}(\rho_2) - \rho_2 R_2),$$

$$E_0^{(2)}(\rho_2) = -\log \sum_Z \sum_X p_1(x) \left(\sum_Y p_2(y) p^{1/(1+\rho_2)}(z|xy) \right)^{1+\rho_2}.$$

Функции $E_0^{(12)}(\rho)$, $E_0^{(1)}(\rho)$, $E_0^{(2)}(\rho)$ обладают свойствами, аналогичными свойствам функции $E_0(\rho)$ для обычного дискретного канала без памяти (см. п. 3.12.4). В частности, прямыми вычислениями нетрудно найти, что

$$\frac{\partial E_0^{(12)}(\rho)}{\partial \rho} \Big|_{\rho=0} = I(XY; Z),$$

$$\frac{\partial E_0^{(1)}(\rho)}{\partial \rho} \Big|_{\rho=0} = I(X; Z|Y),$$

$$\frac{\partial E_0^{(2)}(\rho)}{\partial \rho} \Big|_{\rho=0} = I(Y; Z|X).$$

Поэтому все три функции $E_{12}(R_1, R_2)$, $E_1(R_1)$, $E_2(R_2)$ положительны для любой пары скоростей (R_1, R_2) такой, что

$$R_1 + R_2 < I(XY; Z), \quad R_1 < I(X; Z|Y), \quad R_2 < I(Y; Z|X). \quad (6.3.32)$$

Следовательно, средняя по ансамблю $\mathfrak{G}_n(R_1, R_2)$ вероятность ошибки

$$\bar{\lambda} < 2^{-nE_{12}(R_1, R_2)} + 2^{-nE_1(R_1)} + 2^{-nE_2(R_2)}$$

стремится к нулю с ростом n для любой пары (R_1, R_2) , удовлетворяющей соотношениям (6.3.32). Утверждение теоремы вытекает из того, что в ансамбле имеется хотя бы один код $G_n(R_1, R_2)$, для которого выполняются неравенства (6.3.32) и вероятность ошибки не больше, чем $\bar{\lambda}$. Теорема доказана.

§ 6.4. Кодирование в широковещательных каналах

6.4.1. Постановка задачи. В этом параграфе мы рассмотрим следующую ситуацию передачи информации по каналу (см. рис. 6.4.1). Имеется один передатчик (кодер) и два независимо работающих приемника (декодера), на входы которых поступают выходные сигналы разных каналов. На передатчик поступают сообщения от трех источников U_1, U_2 и U_0 , которые он должен передать приемникам 1 и 2 так, чтобы приемник 1 мог восстановить с произвольно малой вероятностью ошибки сообщения источников U_1 и U_0 , а приемник 2 — сообщения источников U_2 и U_0 . При этом сообщения источников U_1 и U_2 будем называть частной информацией для приемников 1 и 2 соответственно, а сообщения источника U_0 — общей информацией. Такую ситуацию передачи будем называть передачей по *широковещательному каналу* (ШК). (Мы ограничиваемся здесь рассмотрением только широковещательных каналов с двумя приемниками. Некоторые обобщения на случай большего числа приемников приведены в задачах.)

Передача сигналов по ШК определяется двумя каналами с общим входным алфавитом X , выходными алфавитами Y и Z и матрицами переходных вероятностей $\mathbf{P}_1 = \{p_1(y|x)\}, \mathbf{P}_2 = \{p_2(z|x)\}, x \in X, y \in Y, z \in Z$. В дальнейшем мы будем рассматривать только *дискретные ШК без памяти*, т. е. такие ШК, для которых алфавиты X, Y и Z конечны и для любых последовательностей $\mathbf{x} \in X^n, \mathbf{y} \in Y^n, \mathbf{z} \in Z^n$

$$p_1(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n p_1(y_i^{(i)}|x^{(i)}); \quad p_2(\mathbf{z}|\mathbf{x}) = \prod_{i=1}^n p_2(z_i^{(i)}|x^{(i)}).$$

Дискретный широковещательный канал без памяти будет обозначаться символом $\{XYZ; p_1(y|x), p_2(z|x)\}$, при этом каналы $\{XY, p_1(y|x)\}$ и $\{XZ, p_2(z|x)\}$ будут называться *составляющими широковещательного канала*.

Один из возможных методов передачи по ШК состоит в разделении времени, когда в течение некоторого отрезка времени осуществляется передача одному приемнику, а в течение другого отрезка времени — второму. Другая возможность состоит в том, чтобы вести передачу обоим приемникам в одно и то же время.

Определение 6.4.1. Кодом длины n со скоростями R_1, R_2, R_0 для ШК $\{XYZ, \mathbf{P}_1, \mathbf{P}_2\}$ называется совокупность $G_n(R_1, R_2, R_0) \triangleq \{\mathbf{x}_{ijk}, A_{ik}, B_{jk}\}, \mathbf{x}_{ijk} \in X^n, A_{ik} \subseteq Y^n, B_{jk} \subseteq Z^n$,

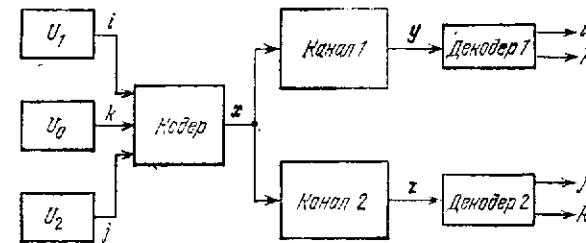


Рис. 6.4.1. Широковещательный канал.

$i = \overline{1, M_1 = 2^{nR_1}}, j = \overline{1, M_2 = 2^{nR_2}}, k = \overline{1, M_0 = 2^{nR_0}}$, причем множества A_{ik} и $A_{i'k'}$, а также множества B_{jk} и $B_{j'k''}$ не пересекаются при $(i, k) \neq (i', k')$ и $(j, k) \neq (j', k'')$ соответственно. Последовательности \mathbf{x}_{ijk} называются кодовыми словами кода для ШК, а множества A_{ik} и B_{jk} — решающими областями для приемников 1 и 2 соответственно.

Кодирование кодом $G_n(R_1, R_2, R_0)$ осуществляется следующим образом. Если на выходе источника U_1 появляется сообщение с номером i и на выходе источника U_0 — сообщение с номером k , то по ШК передается кодовое слово \mathbf{x}_{ijk} . При этом каждое из чисел M_1, M_2, M_0 представляет собой количество элементов в множествах сообщений источников U_1, U_2, U_0 соответственно. Числа $R_1 = n^{-1} \log M_1, R_2 = n^{-1} \log M_2$ представляют собой скорости передачи частной информации для приемников 1 и 2, а число $R_0 = n^{-1} \log M_0$ — скорость передачи общей информации. Работа приемников 1 и 2 (декодирование) при передаче по ШК с помощью кода $G_n(R_1, R_2, R_0)$ осуществляется следующим образом. Если $\mathbf{y} \in A_{ik}$, то приемник 1 выносит решение о том, что на выходе источника U_1 появилось сообщение с номером i , а на выходе источника U_0 — сообщение с номером k . Приемник 2 работает аналогично, используя решающие области B_{jk} .

Качество передачи с помощью кода $G_n(R_1, R_2, R_0)$ характеризуется двумя вероятностями ошибок: вероятностями λ_1 и λ_2 принятия неправильного решения приемниками 1 и 2 соответственно.

При этом

$$\lambda_1 \triangleq (M_1 \cdot M_2 \cdot M_0)^{-1} \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} \sum_{k=1}^{M_0} \sum_{y \notin A_{ijk}} p_1(y | \mathbf{x}_{ijk}), \quad (6.4.1)$$

$$\lambda_2 \triangleq (M_1 \cdot M_2 \cdot M_0)^{-1} \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} \sum_{k=1}^{M_0} \sum_{z \notin B_{ijk}} p_2(z | \mathbf{x}_{ijk}). \quad (6.4.2)$$

Определение 6.4.2. Тройка скоростей (R_1, R_2, R_0) называется допустимой для ШК, если для любых $\epsilon > 0$ и $\delta > 0$

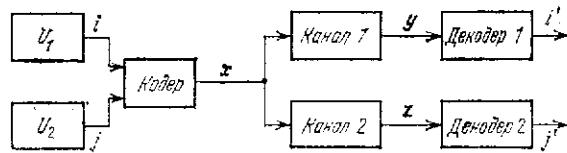


Рис. 6.4.2. Передача по ШК в ситуации КI.

найдется $n_0(\epsilon, \delta)$ такое, что для всех $n > n_0(\epsilon, \delta)$ существует код $G_n(\tilde{R}_1, \tilde{R}_2, \tilde{R}_0)$, $\tilde{R}_i = \max(0, R_i - \delta)$, $i = 0, 1, 2$, с вероятностями ошибок декодирования $\lambda_1 \ll \epsilon$ и $\lambda_2 \ll \epsilon$. Множество \mathfrak{S} , состоящее из всех допустимых троек скоростей, называется областью пропускной способности ШК.

Замечание. В определении области пропускных способностей для ШК мы использовали средние вероятности ошибок декодирования. Однако, в отличие от КМД, в случае передачи по ШК можно показать, что из допустимости тройки скоростей относительно средних вероятностей ошибок следует ее допустимость и относительно максимальных по (i, j, k) вероятностей ошибок.

В случае передачи по ШК, так же как и в ранее рассмотренных задачах, возможно использование метода разделения времени, из которого следует, что, если тройки скоростей (R'_1, R'_2, R'_0) и (R''_1, R''_2, R''_0) допустимы, то допустима и тройка скоростей (R_1, R_2, R_0) такая, что $R_i = \alpha R'_i + (1 - \alpha) R''_i$, $i = 0, 1, 2$, при любом $\alpha \in [0, 1]$. Поэтому область пропускной способности ШК является выпуклой областью неотрицательного октаната трехмерного пространства.

Основной теоретико-информационной задачей при рассмотрении кодирования в ШК является характеристизация его области пропускной способности. К сожалению, для произвольного дискретного канала без памяти эта задача в настоящее время не решена. Имеются решения лишь для некоторых частных подклассов широковещательных каналов, один из которых — подкласс так называемых ухудшающихся каналов — мы и рассмотрим в этом разделе (определение ухудшающегося канала будет дано ниже).

В общей задаче кодирования для ШК часто оказывается полезным рассматривать два частных случая. Предположим, что источник U_0 в схеме передачи по ШК отсутствует (см. рис. 6.4.2). При этом передатчик должен реализовывать передачу только частных информаций приемникам 1 и 2. Такую ситуацию передачи по ШК будем обозначать через КI.

Определение 6.4.3. Кодом длины n со скоростями R_1 и R_2 для передачи по ШК $\{XYZ, P_1(y|x), P_2(z|x)\}$ в ситуации КI называется совокупность $G_n(R_1, R_2) = \{\mathbf{x}_{ij}, A_{ij}, B_j\}$, $\mathbf{x}_{ij} \in X^n$,

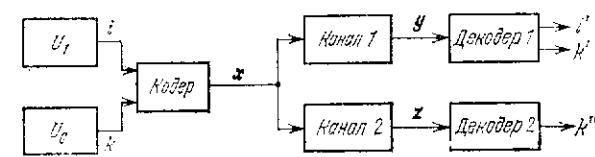


Рис. 6.4.3. Передача по ШК в ситуации КII.

$A_i \subseteq Y^n$, $B_j \subseteq Z^n$, $i = 1, M_1 = 2^{nR_1}$, $j = 1, M_2 = 2^{nR_2}$, причем множества A_i и $A_{i'}$, а также множества B_j и $B_{j'}$ не пересекаются при $i \neq i'$ и $j \neq j'$ соответственно.

Вероятности ошибок декодирования в ситуации КI определяются соотношениями (6.4.1) и (6.4.2), если в них положить $M_0 = 1$ и опустить индекс k .

Определение 6.4.4. Пара скоростей (R_1, R_2) называется допустимой для ШК в ситуации КI, если для любых $\epsilon > 0$ и $\delta > 0$ найдется $n_0(\epsilon, \delta)$ такое, что для всех $n > n_0$ существует код $G_n(\tilde{R}_1, \tilde{R}_2)$, $\tilde{R}_i = \max(0, R_i - \delta)$, $i = 1, 2$, с вероятностями ошибок декодирования $\lambda_1 \ll \epsilon$ и $\lambda_2 \ll \epsilon$. Множество \mathfrak{S}_1 , состоящее из всех допустимых пар скоростей (R_1, R_2) , называется областью пропускной способности для ШК в ситуации КI.

Второй частный случай получается, если предположить, что отсутствует источник U_2 (см. рис. 6.4.3). При этом передатчик должен реализовывать передачу частной информации приемнику 1 и общей информации приемникам 1 и 2. Такую ситуацию передачи по ШК будем обозначать через КII. (Можно, конечно, рассмотреть ситуацию, когда отсутствует источник U_1 , а не U_2 , однако нетрудно видеть, что эта ситуация аналогична ситуации КII).

Определение 6.4.5. Кодом длины n со скоростями R_1 и R_0 для передачи по ШК $\{XYZ, P_1, P_2\}$ в ситуации КII называется совокупность $G_n(R_1, R_0) = \{\mathbf{x}_{ik}, A_{ik}, B_k\}$, $\mathbf{x}_{ik} \in X^n$, $A_{ik} \subseteq Y^n$, $B_k \subseteq Z^n$, $i = 1, M_1 = 2^{nR_1}$, $k = 1, M_0 = 2^{nR_0}$,

причем множества A_{ik} и $A_{i'k'}$, а также множества B_k и $B_{k'}$ не пересекаются при $(i, k) \neq (i', k')$ и $k \neq k'$ соответственно.

Вероятности ошибок декодирования в ситуации КII определяются соотношениями (6.4.1) и (6.4.2), если в них положить $M_2 = 1$ и опустить индекс j . Область пропускной способности СII ШК в ситуации КII определяется аналогично общей ситуации и ситуации КI.

Заметим, что СI и СII являются выпуклыми подмножествами неотрицательных квадрантов плоскостей с координатами (R_1, R_2) и (R_1, R_0) соответственно. В дальнейшем общую ситуацию передачи по ШК (рис. 6.4.1) будем обозначать через КIII и соответствующую ей область пропускной способности — через СIII. Нетрудно видеть, что СI и СII являются сечениями области СIII плоскостями $R_0 = 0$ и $R_2 = 0$ соответственно.

6.4.2. Ухудшающиеся широковещательные каналы (УШК). Переходные вероятности $\{p_1(y|x)\}$ и $\{p_2(z|x)\}$ будем представлять в виде матриц P_1 и P_2 соответственно. Каждая строка таких матриц соответствует входному сигналу канала, а каждый столбец — выходному. Если два дискретных канала без памяти с матрицами переходных вероятностей P и Q соединяются последовательно, т. е. выходные сигналы первого канала являются входными сигналами второго, причем множество выходных сигналов первого совпадает со множеством входных сигналов второго, то результирующий канал является каналом без памяти с матрицей переходных вероятностей PQ .

Определение 6.4.6. Дискретный канал без памяти с матрицей переходных вероятностей $P_2 = \{p_2(z|x)\}$, $x \in X$, $z \in Z$, называется *ухудшенным вариантом* дискретного канала без памяти с матрицей переходных вероятностей $P_1 = \{p_1(y|x)\}$, $x \in X$, $y \in Y$, если существует такая стохастическая матрица $Q = \{q(z|y)\}$, $y \in Y$, $z \in Z$ ($\sum_z q(z|y) = 1$, $q(z|y) \geq 0$ для всех $y \in Y$, $z \in Z$), что $P_2 = P_1 Q$ или, что то же самое, $p_2(z|x) = \sum_y p_1(y|x) q(z|y)$ для всех $x \in X$, $z \in Z$. Широковещательный канал без памяти называется *ухудшающимся*, если одна из его составляющих является ухудшенным вариантом другой.

Из этого определения следует, что в случае УШК худшая составляющая УШК может интерпретироваться как последовательное соединение лучшей составляющей с некоторым другим каналом без памяти. На рис. 6.4.4 показан случай передачи по УШК, в котором первая составляющая, канал $\{XY, p_1(y|x)\}$, является лучшей, а вторая составляющая, канал $\{XZ, p_2(z|x)\}$, является ухудшенным вариантом первой, причем ухудшение достигается за счет дополнительного канала $\{YZ, q(z|y)\}$. Основное свойство

пары каналов, один из которых является ухудшенным вариантом другого, дается в следующей лемме.

Лемма 6.4.1. Пусть дискретный канал без памяти с матрицей переходных вероятностей $P_2 = \{p_2(z|x)\}$ является ухудшенным вариантом дискретного канала без памяти с матрицей переходных вероятностей $P_1 = \{p_1(y|x)\}$. Тогда для любого распределения вероятностей $p(x)$ на входных сигналах

$$I(Z; X) \leq I(Y; X). \quad (6.4.3)$$

Доказательство. Пусть $Q = \{q(z|y)\}$ — некоторая стохастическая матрица и $P_2 = P_1 Q$. Тогда распределение вероят-

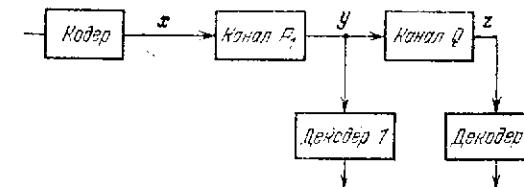


Рис. 6.4.4. Ухудшающийся ШК.

ностей для рассматриваемого ансамбля XYZ можно представить следующим образом:

$$p(x, y, z) = p(x)p_1(y|x)q(z|y) \quad (6.4.4)$$

(см. рис. 6.4.4). Из представления (6.4.4) видно, что ансамбли Z и X при фиксированном $y \in Y$ статистически независимы. Рассмотрим среднюю взаимную информацию $I(X; YZ)$. С одной стороны,

$$I(X; YZ) = I(X; Y) + I(X; Z|Y) = I(X; Y),$$

где последнее равенство следует из независимости, указанной выше. С другой стороны,

$$I(X; YZ) = I(X; Z) + I(X; Y|Z) \geq I(X; Z).$$

Сопоставляя последние два соотношения, получим неравенство (6.4.4). Лемма доказана.

Из определения УШК вытекает следующее интуитивно понятное его свойство: если передаваемые по УШК сообщения могут быть восстановлены с произвольно малой вероятностью ошибки на выходе худшей составляющей УШК, то они также могут быть восстановлены с произвольно малой вероятностью ошибки и на выходе лучшей составляющей. Строгое доказательство этого свойства будет приведено ниже при доказательстве теорем кодирования для УШК.

Далее всегда будет предполагаться, что в УШК вторая составляющая является ухудшенным вариантом первой. Тогда из ука-

занного свойства следует, что если для УШК пара скоростей (R_1, R_2) принадлежит \mathbb{C}_I , то пара скоростей (R_1, R_0) при $R_0 = R_2$ принадлежит \mathbb{C}_{II} . Так как общая информация в произвольном ШК может интерпретироваться как частная (может быть включена в частную), то из $(R_1, R_0) \in \mathbb{C}_{II}$ следует, что $(R_1, R_2) \in \mathbb{C}_I$ при $R_2 = R_0$. Таким образом, множества \mathbb{C}_I и \mathbb{C}_{II} однозначно определяют друг друга в УШК. Кроме того, из включения $(R_1, R_2, R_0) \in \mathbb{C}_{III}$ следует, что $(R_1, R_0) \in \mathbb{C}_{II}$ при $R'_0 = R_2 + R_0$. Верно и обратное, для любой пары $(R_1, R_0) \in \mathbb{C}_{II}$ всякая тройка

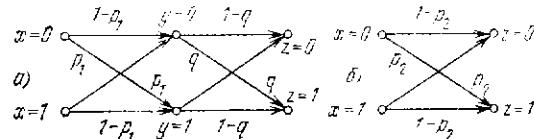


Рис. 6.4.5. Двоичный симметричный УШК.

$(R_1, R'_2, R'_0) \in \mathbb{C}_{III}$ при $R'_2 + R'_0 = R_0$. Из приведенных рассуждений вытекает, что общая задача поиска области пропускной способности для УШК сводится к поиску области пропускной способности \mathbb{C}_{II} при передаче по ШК в ситуации КII.

6.4.3. Двоичный симметричный широковещательный канал. В качестве примера УШК рассмотрим *двоичный симметричный ШК* (ДСШК), т. е. такой ШК, у которого и первая и вторая составляющие являются обычными ДСК с вероятностями ошибок p_1 и p_2 соответственно. На этом примере будет проиллюстрирована основная идея кодирования в УШК.

Покажем вначале, что ДСШК действительно является УШК. Рассмотрим последовательное соединение двух ДСК, первого с вероятностью ошибки p_1 , а второго — q (см. рис. 6.4.5). Тогда результирующий канал будет ДСК с вероятностью ошибки

$$p_2 := q(1 - p_1) + (1 - q)p_1 \triangleq p_1 * q. \quad (6.4.5)$$

Если $p_1 < p_2 < 1/2$, то последнее соотношение, рассматриваемое как уравнение относительно q , дает $q = (p_2 - p_1)/(1 - 2p_1)$ и, следовательно, рассматриваемый ШК является ухудшающимся. Мы будем называть *сверткой* операцию, обозначенную в (6.4.5) знаком $*$, таким образом, p_2 есть свертка p_1 и q . Свертка обладает следующими легко проверяемыми свойствами: 1) $\alpha * \beta = \beta * \alpha$; 2) $\alpha * \beta < 1/2$, если $\alpha < 1/2$ и $\beta < 1/2$; 3) $\alpha * \beta = 1/2$, если либо $\alpha = 1/2$, либо $\beta = 1/2$.

Рассмотрим следующий метод построения кода для ДСШК при передаче в ситуации КII. Пусть имеются три последовательно соединенных ДСК с вероятностью ошибок α , p_1 и q (см. рис. 6.4.6). Первый канал, имеющий вероятность ошибки α , будем называть

тест-каналом, а его входной алфавит обозначать через S , $S = \{0, 1\}$. Результирующий канал, как легко видеть, снова является ДСК. Его вероятность ошибки равна $p = \alpha * p_1 * q = \alpha * p_2$. На входном алфавите тест-канала зададим равномерное распределение вероятностей $p(s=0) = p(s=1) = 1/2$ и построим код $G(n, R_0)$ для результирующего канала. В соответствии с теоремой кодирования для дискретных каналов без памяти для любого $\epsilon > 0$ найдется n и найдется код со скоростью R_0

$$R_0 < I(S; Z) = H(Z) - H(Z|S) = 1 - h(p) = 1 - h(\alpha * p_2), \quad (6.4.6)$$

где $h(p) \triangleq -p \log p - (1 - p) \log(1 - p)$, для которого вероятность ошибки $\lambda \ll \epsilon$. Кодовые слова этого кода будем обозначать

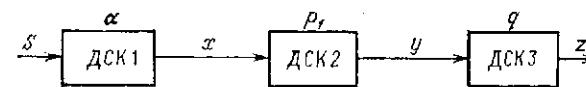


Рис. 6.4.6. К построению кода для ДСШК в ситуации КII.

через $s_k \in S^n$, $k = 1, \dots, M_0$, $M_0 = 2^{nR_0}$, и называть *центрами концентрации*.

Если на входе первого ДСК фиксирована некоторая последовательность $s \in S^n$, то в силу закона больших чисел при большом n с вероятностью, близкой к единице, появится такая последовательность $x \in X^n$, которая отличается от s примерно в αn символах. Так как фиксированной последовательности s соответствует следующее условное распределение на X^n :

$$p(x|s) = \prod_{i=1}^n p(x^{(i)}|s^{(i)}),$$

$$p(x^{(i)}|s^{(i)}) = \begin{cases} \alpha & \text{при } x^{(i)} \neq s^{(i)}, \\ 1 - \alpha & \text{при } x^{(i)} = s^{(i)}, \end{cases} \quad (6.4.7)$$

то высоковероятное множество этого распределения состоит из последовательностей $x \in X^n$, отличающихся от s примерно в αn символах. Заметим далее, что

$$I(X; Y|s) = H(Y|s) - H(Y|Xs) = H(Y|s) - H(Y|X),$$

где последнее равенство следует из независимости Y от s при фиксированном $x \in X$. В силу симметрии каналов имеем

$$H(Y|s) = H(Y|S) = h(\alpha * p_1), \quad H(Y|X) = h(p_1)$$

и, следовательно,

$$I(X; Y|s) = I(X; Y|S) = h(\alpha * p_1) - h(p_1).$$

Таким образом, если на X^n фиксировано распределение вероятностей $p(x|s)$ (см. (6.4.7)), то

$$I(X^n; Y^n | s) = I(X^n; Y^n | S^n) = nI(X; Y | S).$$

В соответствии с неравенством Файнстейна (см. теорему 3.8.1) для любого $\varepsilon > 0$ и любого $s \in S^n$ при достаточно большом n можно построить код $G_s(n, R)$ для ДСК с вероятностью ошибки p_1 , слова которого принадлежат высоковероятному множеству ансамбля $\{X^n, p(x|s)\}$ и вероятность ошибки декодирования $\lambda \ll \varepsilon$. Скорость такого кода удовлетворяет неравенству

$$R_1 < I(X; Y | S) = h(\alpha * p_1) - h(p_1).$$

Кодовые слова для ДСШК получаются так. Построим M_0 кодов $G_{s_k}(n, R_1)$ для каждого центра концентрации s_k , $k = \overline{1, M_0}$. Кодовые слова $\{\mathbf{x}_{1k}, \dots, \mathbf{x}_{M_1 k}\}$, $M_1 = 2^{nR_1}$, k -го кода принадлежат высоковероятному множеству ансамбля $\{X^n, p(x|s_k)\}$ и представляют собой последовательности, отличающиеся от своего центра концентрации примерно в αn символах, т. е. расположенные вблизи сферы радиуса αn с центром в s_k . Кодирование пары сообщений (i, k) состоит вначале в выборе k -го центра концентрации s_k , а затем в выборе i -го кодового слова \mathbf{x}_{ik} из множества кодовых слов, расположенных в сфере радиуса αn с центром в s_k .

Рассмотрим следующее правило декодирования построенного кода. Декодер на выходе второй составляющей ДСШК по принятой последовательности Z определяет центр концентрации s_k так, как если бы он передавался по ДСК с вероятностью ошибки $p = \alpha * p_2$. Так как при фиксированном k кодовые слова \mathbf{x}_{ik} , $i = \overline{1, M_1}$, принадлежат высоковероятному множеству распределения $p(x|s_k)$, а центры концентрации являются кодовыми словами кода для ДСК с вероятностью ошибки $p = \alpha * p_2$, то при $R_0 < 1 - H(\alpha * p_2)$ и достаточно большом n вероятность неправильного определения центра концентрации может быть сделана сколь угодно малой. Декодер на выходе первой составляющей ДСШК работает в два этапа. На первом этапе он определяет по принятой последовательности y центр концентрации s_k . Так как вторая составляющая является ухудшенным вариантом первой, то декодер первой составляющей также может определить s_k с произвольно малой вероятностью ошибки. На втором этапе, зная s_k , декодер осуществляет декодирование кода $G_{s_k}(n, R_1)$, в результате которого определяется переданная последовательность \mathbf{x}_{ik} . Если $R_1 < H(\alpha * p_1) - H(p_1)$, то при достаточно большом n вероятность ошибки декодирования на втором этапе при условии, что s_k определен верно, может быть сделана также сколь угодно малой.

Таким образом, мы показали, хотя и не вполне строго, что в ДСШК допустимы все такие пары скоростей (R_1, R_0) , которые при некотором α , $\alpha \in [0, 1/2]$, удовлетворяют неравенствам

$$R_1 < I(Y; X | S) = h(\alpha * p_1) - h(p_1),$$

$$R_0 < I(Z; S) = 1 - h(\alpha * p_2).$$

Область допустимых пар скоростей, задаваемая этими неравенствами, показана на рис. 6.4.7.

Рассмотренный пример наводит на мысль о том, что и для произвольного дискретного УШК без памяти допустимые пары скоростей могут быть получены аналогичным способом. А именно, пусть S — некоторое конечное множество и $p(s, x)$ — некоторое распределение вероятностей на SX . Рассмотрим ансамбль $SXYZ$, распределение вероятностей на котором имеет следующий вид:

$$p(s, x, y, z) = p(s)p(x|s)p_1(y|x)q(z|y). \quad (6.4.8)$$

Наше предположение состоит в том, что для произвольного дискретного УШК без памяти допустимой является любая пара скоростей (R_1, R_0) такая, что

$$R_1 < I(Y; X | S), \quad R_0 < I(Z; S), \quad (6.4.9)$$

где средние взаимные информации вычислены в соответствии с распределением вероятностей (6.4.8) на ансамбле $SXYZ$. Ниже мы докажем прямую и обратную теоремы кодирования, из которых будет следовать, что область пропускной способности СИИ для дискретного УШК без памяти состоит из всех пар скоростей, удовлетворяющих неравенствам (6.4.9) при некотором S и некотором распределении $p(s, x)$ на SX .

6.4.4. Обратная теорема кодирования. Для описания области пропускной способности УШК потребуется некоторая специальная функция. Вначале мы введем эту функцию и изучим ее свойства.

Пусть задан дискретный УШК без памяти с входным алфавитом X , выходными алфавитами Y, Z и переходными вероятностями $\{p_1(y|x)\}, \{p_2(z|x)\}$, $x \in X, y \in Y, z \in Z$, причем вторая составляющая канала является ухудшенным вариантом первой, т. е. $p_2(z|x) = \sum_y p_1(y|x) q(z|y)$. Пусть S — некоторое конечное множество, $|S| < \infty$. Зададим распределение вероятностей на четвертках $(s, x, y, z) \in SXYZ$ посредством соотношения (6.4.8), где $p(s)$ и $p(x|s)$ — некоторые выбранные произвольным образом распределения вероятностей на S и X соответственно,

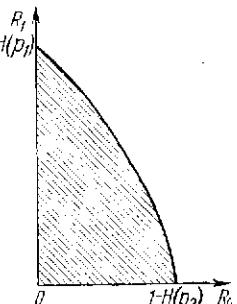


Рис. 6.4.7. Область допустимых пар скоростей ДСШК.

Обозначим через \mathcal{A} множество всех ансамблей $SXYZ$ такого вида. Пусть $\mathcal{A}(d)$ — такое максимальное подмножество множества \mathcal{A} , что $I(S; Z) \geq d$ для любого ансамбля $SXYZ \in \mathcal{A}(d)$. Определим функцию $t(d)$, $d \geq 0$, следующим образом:

$$t(d) \triangleq \sup_{\mathcal{A}(d)} I(X; Y | S). \quad (6.4.10)$$

В определении функции $t(d)$ используется знак \sup , а не \max , так как мощность множества S хотя и ограничена, но может быть сколь угодно большой. Далее мы покажем, что функция $t(d)$ задает границу каждой из областей пропускных способностей СИ и СII для дискретного УШК без памяти. Иными словами, мы покажем, что при фиксированной скорости R_2 (или R_0) максимальное значение скорости R_1 равно $t(R_2)$ (или $t(R_0)$).

Как уже отмечалось выше, область пропускных способностей выпукла. Поэтому выпуклой должна быть и ее граница. В следующей лемме устанавливается выпуклость функции $t(d)$.

Лемма 6.4.2. *Функция $t(d)$ — монотонно невозрастающая, выпуклая вверх функция при всех $d \geq 0$.*

Доказательство. Для доказательства монотонного невозрастания достаточно заметить, что $\mathcal{A}(d_1) \supseteq \mathcal{A}(d_2)$ при всех $d_2 \geq d_1 \geq 0$, и воспользоваться определением (6.4.10). Для доказательства выпуклости достаточно показать, что для любых $d_1 \geq 0$, $d_2 \geq 0$ и для любого α , $\alpha \in [0, 1]$, имеет место неравенство

$$t(\alpha d_1 + (1 - \alpha) d_2) \geq \alpha t(d_1) + (1 - \alpha) t(d_2). \quad (6.4.11)$$

Пусть зафиксировано $\varepsilon > 0$, тогда можно найти два ансамбля $S'XYZ$ и $S''XYZ$ из \mathcal{A} со следующими свойствами:

$$\begin{aligned} &\{S'XYZ, p'(s)p'(x|s)p_1(y|s)q(z|y)\} \in \mathcal{A}(d_1), \\ &I(S'; Z) \geq d_1, \quad I(X; Y | S') \geq t(d_1) - \varepsilon, \\ &\{S''XYZ, p''(s)p''(x|s)p_1(y|s)q(z|y)\} \in \mathcal{A}(d_2), \\ &I(S''; Z) \geq d_2, \quad I(X; Y | S'') \geq t(d_2) - \varepsilon. \end{aligned}$$

Множества S' и S'' можно рассматривать как дизъюнктные. Пусть $S \triangleq S' \cup S''$ и для любого элемента $s \in S$ положим

$$\begin{aligned} p(s) &= \begin{cases} \alpha p'(s) & \text{при } s \in S'; \\ (1 - \alpha)p''(s) & \text{при } s \in S''; \end{cases} \\ p(x|s) &= \begin{cases} p'(x|s) & \text{при } s \in S'; \\ p''(x|s) & \text{при } s \in S''. \end{cases} \end{aligned}$$

При этом определен ансамбль $SXYZ \in \mathcal{A}$, для которого

$$I(S; Z) \geq \alpha I(S'; Z) + (1 - \alpha) I(S''; Z) \geq \alpha d_1 + (1 - \alpha) d_2; \quad (6.4.12)$$

$$\begin{aligned} I(X; Y | S) &= \alpha I(X; Y | S') + (1 - \alpha) I(X; Y | S'') = \alpha t(d_1) + \\ &+ (1 - \alpha) t(d_2) - \varepsilon. \quad (6.4.13) \end{aligned}$$

Таким образом, из (6.4.12) следует, что ансамбль $SXYZ$ принадлежит $\mathcal{A}(d)$, а из (6.4.13), произвольности ε и определения (6.4.10) следует неравенство (6.4.11). Лемма доказана.

Введем теперь аналогичную функцию для последовательностей длины n . Пусть S — некоторое конечное множество, $|S| < \infty$, рассмотрим ансамбль $SX^nY^nZ^n$, распределение вероятностей на котором имеет следующий вид:

$$p(s, x, y, z) = p(s)p(x|s)p_1(y|x)p(z|y), \quad (6.4.14)$$

где

$$p_1(y|x) = \prod_{i=1}^n p_1(y^{(i)}|x^{(i)}), \quad q(z|y) = \prod_{i=1}^n q(z^{(i)}|y^{(i)}) \quad (6.4.15)$$

и условные распределения $p_1(y|x)$, $q(z|y)$, $x \in X$, $y \in Y$, $z \in Z$, определены заданием УШК без памяти. Обозначим через \mathcal{A}_n множество всех ансамблей такого вида. Пусть $\mathcal{A}_n(d)$ — такое максимальное подмножество множества \mathcal{A} , что $\frac{1}{n} I(S; Z^n) \geq d$ для любого ансамбля $SX^nY^nZ^n \in \mathcal{A}_n(d)$. Определим функцию $t_n(d)$, $d \geq 0$, следующим образом:

$$t_n(d) \triangleq \sup_{\mathcal{A}_n(d)} \frac{1}{n} I(X^n; Y^n | S). \quad (6.4.16)$$

Лемма 6.4.3. *Для любого положительного n и любого $d \geq 0$*

$$t_n(d) \leq t(d). \quad (6.4.17)$$

Доказательство. Для доказательства леммы достаточно показать, что для любого ансамбля $SX^nY^nZ^n \in \mathcal{A}_n(d)$ имеет место неравенство

$$\frac{1}{n} I(X^n; Y^n | S) \leq t(d). \quad (6.4.18)$$

Заметим, что из (6.4.14) и (6.4.15) следует статистическая независимость ансамблей Z_i и Z_1, \dots, Z_{i-1} , $i = \overline{1, n}$, при фиксированных $s \in S$ и $(y^{(1)}, \dots, y^{(i-1)}) \in Y_1, \dots, Y_{i-1}$, где через

Y_i и Z_i обозначены ансамбли сообщений, соответствующие моменту времени i , $Y^n = Y_1 \dots Y_n$, $Z^n = Z_1 \dots Z_n$. Действительно, $p(s, y^{(1)}, \dots, y^{(i-1)}, z^{(1)}, \dots, z^{(i-1)}z^{(i)}) =$

$$\begin{aligned} &= \sum_{X^n} \sum_{Y_i, \dots, Y_n} \sum_{Z_{i+1}, \dots, Z_n} p(s) p(\mathbf{x}|s) \prod_{j=1}^n p_1(y^{(j)}|x^{(j)}) q(z^{(j)}|y^{(j)}) = \\ &= p(s) \sum_{X^n} \sum_{Y_i} p(\mathbf{x}|s) \prod_{j=1}^i p_1(y^{(j)}|x^{(j)}) q(z^{(j)}|y^{(j)}) = \\ &= p(s) \prod_{j=1}^{i-1} q(z^{(j)}|y^{(j)}) p(y^{(1)}, \dots, y^{(i-1)}, z^{(i)}|s), \end{aligned} \quad (6.4.19)$$

причем последнее равенство вытекает из того, что $q(z^{(i)}|y^{(i)}) = q(z^{(i)}|y^{(1)}, \dots, y^{(i)}, x^{(1)}, \dots, x^{(i)})$ (канал $\{YZ, q(z|y)\}$ не имеет памяти), и из следующих соотношений:

$$\begin{aligned} &\sum_{X^n} \sum_{Y_i} p(\mathbf{x}|s) \prod_{j=1}^i p_1(y^{(j)}|x^{(j)}) q(z^{(j)}|y^{(j)}) = \\ &= \sum_{X_1, \dots, X_i} \sum_{Y_i} p(x^{(1)}, \dots, x^{(i)}|s) p(y^{(1)}, \dots, y^{(i)}|x^{(1)}, \dots, x^{(i)}) \times \\ &\quad \times q(z^{(i)}|y^{(1)}, \dots, y^{(i)}, x^{(1)}, \dots, x^{(i)}) = \\ &= \sum_{Y_i} p(y^{(1)}, \dots, y^{(i)}, z^{(i)}|s) = p(y^{(1)}, \dots, y^{(i-1)}, z^{(i)}|s), \end{aligned}$$

т. е. имеет место указанная выше независимость. Отсюда имеем $I(S; Z^n) = H(Z^n) - H(Z^n|S) \leq$

$$\begin{aligned} &\leq \sum_{i=1}^n H(Z_i) - \sum_{i=1}^n H(Z_i|SZ_1 \dots Z_{i-1}) \leq \\ &\leq \sum_{i=1}^n H(Z_i) - \sum_{i=1}^n H(Z_i|SY_1 \dots Y_{i-1}Z_1 \dots Z_{i-1}) = \\ &= \sum_{i=1}^n H(Z_i) - \sum_{i=1}^n H(Z_i|SY_1 \dots Y_{i-1}) = \sum_{i=1}^n I(Z_i; SY_1 \dots Y_{i-1}). \end{aligned} \quad (6.4.20)$$

Кроме того,

$$\begin{aligned} I(X^n; Y^n|S) &= H(Y^n|S) - H(Y^n|X^n) = \\ &= \sum_{i=1}^n H(Y_i|SY_1 \dots Y_{i-1}) - \sum_{i=1}^n H(Y_i|X_i) = \\ &= \sum_{i=1}^n I(Y_i; X_i|SY_1 \dots Y_{i-1}), \end{aligned} \quad (6.4.21)$$

где первое равенство следует из независимости ансамблей Y^n и S при каждой фиксированной последовательности $\mathbf{x} \in X^n$, второе — из того, что канал $\{XY, p(x, y)\}$ не имеет памяти, и третье — из независимости ансамблей Y_i и $SY_1 \dots Y_{i-1}$ при каждом фиксированном $x^{(i)} \in X_i$.

Положим теперь $S_1 \triangleq S$, $S_2 \triangleq SY_1 \dots S_i \triangleq SY_1 \dots Y_{i-1} \dots S_n = SY_1 \dots Y_{n-1}$. Тогда из (6.4.20) и (6.4.21) следует, что

$$\frac{1}{n} I(S; Z^n) \leq \frac{1}{n} \sum_{i=1}^n I(S_i; Z_i), \quad (6.4.22)$$

$$\frac{1}{n} I(X^n; Y^n|S) = \frac{1}{n} \sum_{i=1}^n I(X_i; Y_i|S_i). \quad (6.4.23)$$

Обозначим $I(S_i; Z_i) = d_i$, тогда $S_i X_i Y_i Z_i \in \mathcal{A}(d_i)$ и $I(X_i; Z_i|S_i) \leq t(d_i)$. Из (6.4.21) имеем

$$\frac{1}{n} I(X^n; Y^n|S) \leq \frac{1}{n} \sum_{i=1}^n t(d_i) \leq t\left(\frac{1}{n} \sum_{i=1}^n d_i\right), \quad (6.4.24)$$

где последнее неравенство следует из выпуклости функции $t(d)$. Предположим, что $S X^n Y^n Z^n \in \mathcal{A}_n(d)$, тогда $\frac{1}{n} I(S; Z^n) \geq d$

и из неравенства (6.4.22) получим, что $\frac{1}{n} \sum_{i=1}^n d_i \geq d$. Поскольку $t(d)$ — монотонно невозрастающая функция d , то неравенство (6.4.24) дает

$$\frac{1}{n} I(X^n; Y^n|S) \leq t(d),$$

что и требовалось доказать.

Рассмотрим теперь некоторый код $G_n(R_1, R_2)$ для дискретного УШК при передаче в ситуации КI. Пусть $S = \{1, 2, \dots, M_2\}$, $M_2 = 2^{nR_2}$. Выберем распределения вероятностей $\tilde{p}(s)$, $s \in S$, и $\tilde{p}(\mathbf{x}|s)$, $\mathbf{x} \in X^n$, следующим образом:

$$\tilde{p}(s) = M_2^{-1} \text{ для всех } s \in S; \quad (6.4.25)$$

$$\tilde{p}(\mathbf{x}|s=k) = \begin{cases} M_1^{-1}, & \text{если } \mathbf{x} = \mathbf{x}_{ik} \text{ при некотором } i, i = \overline{1, M_1}; \\ 0 & \text{для остальных } \mathbf{x}, \end{cases} \quad (6.4.26)$$

где $\{x_{ik}\}$, $i = \overline{1, M_1}$, $k = \overline{1, M_2}$ — множество кодовых слов кода $G_n(R_1, R_2)$. Зададим на $SX^nY^nZ^n$ следующее распределение вероятностей:

$$\tilde{p}(s, x, y, z) = \tilde{p}(s)\tilde{p}(x|s)\prod_{i=1}^n p_i(y^{(i)}|x^{(i)})q(z^{(i)}|y^{(i)}) \quad (6.4.27)$$

и обозначим через $\tilde{\Gamma}(X^n; Y^n | S)$ и $\tilde{\Gamma}(S; Z^n)$ средние взаимные информации, вычисленные в соответствии с распределениями (6.4.25)–(6.4.27). В следующей лемме делается первый шаг для доказательства обратной теоремы кодирования.

Лемма 6.4.4. *Пусть λ_1, λ_2 — вероятности ошибок при декодировании первым и вторым декодерами кода $G_n(R_1, R_2)$ для дискретного УШК без памяти. Для любого такого кода имеют место неравенства*

$$\lambda_1 + \frac{h(\lambda_1)}{nR_1} \geq R_1 - \frac{1}{n} \tilde{\Gamma}(X^n; Y^n | S),$$

$$\lambda_2 + \frac{h(\lambda_2)}{nR_2} \geq R_2 - \frac{1}{n} \tilde{\Gamma}(S; Z^n).$$

Доказательство. Обозначим через W_1, W_2 множества решений первого и второго декодеров соответственно. Правило декодирования для кода $G_n(R_1, R_2)$ и распределения (6.4.25)–(6.4.27) однозначно определяют распределение вероятностей на множестве $SX^nY^nZ^nW_1W_2$. Для этого распределения вероятностей выполняются следующие очевидные соотношения:

$$\begin{aligned} H(X^n | SW_1) &= H(X^n | S) - I(X^n; W_1 | S) = \\ &= nR_1 - I(X^n; W_1 | S) \geq nR_1 - I(X^n; Y^n | S); \end{aligned} \quad (6.4.28)$$

$$\begin{aligned} H(S | W_2) &= H(S) - I(S; W_2) = nR_2 - I(S; W_2) \geq \\ &\geq nR_2 - I(S; Z^n). \end{aligned} \quad (6.4.29)$$

Заметим, что вероятность ошибки второго декодера $\lambda_2 = p(s \neq w_2)$, $s \in S$, $w_2 \in W_2$, а условная вероятность ошибки первого декодера при условии, что передавалось одно из кодовых слов $\{x_{1k}, \dots, x_{M,k}\}$, есть $\lambda_1(k) = p(i \neq w_1 | s = k)$, $i = \overline{1, M_1}$, $w_1 \in W_1$. Утверждение леммы следует из сделанного замечания, из (6.4.28), (6.4.29) и неравенства Фано. Лемма доказана.

Из доказанной леммы вытекает, что пара скоростей (R_1, R_2) для передачи по УШК в ситуации КI не является допустимой в том случае, когда удовлетворяется хотя бы одно из неравенств

$$\begin{aligned} R_1 &> \frac{1}{n} \tilde{\Gamma}(X^n; Y^n | S), \\ R_2 &> \frac{1}{n} \tilde{\Gamma}(S; Z^n). \end{aligned} \quad (6.4.30)$$

Из определения функции $t_n(d)$ теперь легко вывести, что для любой пары скоростей (R_1, R_2) , для которой $R_1 > t_n(R_2)$, хотя бы одно из неравенств (6.4.30) имеет место. Действительно, для любой такой пары либо $R_2 > \frac{1}{n} \tilde{\Gamma}(S; Z^n)$, т. е. имеет место второе неравенство, либо $R_2 < \frac{1}{n} \tilde{\Gamma}(S; Z^n) = d'$, тогда $\frac{1}{n} \tilde{\Gamma}(X^n; Z^n | S) \leq \leq t_n(d') \leq t_n(R_2) \leq R_1$, т. е. имеет место первое неравенство. Согласно лемме 6.4.3 $t_n(d) = t(d)$, поэтому все допустимые пары скоростей (R_1, R_2) должны принадлежать области, ограниченной кривой $t(d)$.

Заметим, однако, что множество S , участвующее в определении функции $t(d)$, может иметь сколь угодно большое число элементов и поэтому $t(d)$ может оказаться невычислимой функцией. Напомним, что подобная проблема возникала в задаче кодирования источников с дополнительной информацией. К счастью, проблема вычислимости и в рассматриваемой задаче может быть решена.

Лемма 6.4.5. *Пусть $\hat{\mathcal{A}}(d)$ является максимальным подмножеством множества $\mathcal{A}(d)$ таким, что если $SXYZ \in \hat{\mathcal{A}}(d)$, то $|S| \leq |X|$. Тогда*

$$t(d) \triangleq \sup_{\hat{\mathcal{A}}(d)} I(X; Y | S) = \max_{\hat{\mathcal{A}}(d)} I(X; Y | S).$$

Доказательство. Пусть \mathcal{A}_L — такое максимальное подмножество множества \mathcal{A} , что если $SXYZ \in \mathcal{A}_L$, то $|S| = L$, и пусть $\hat{\mathcal{A}}_L(d)$ — такое подмножество множества \mathcal{A}_L , что если $SXYZ \in \hat{\mathcal{A}}_L(d)$, то $I(Z; S) \geq d$. Для доказательства утверждения леммы достаточно показать, что для любого $L > |X|$ выполняется равенство

$$t_L(d) \triangleq \max_{\mathcal{A}_L(d)} I(X; Y | S) = \max_{\hat{\mathcal{A}}_L(d)} I(X; Y | S) \triangleq \hat{t}(d). \quad (6.4.31)$$

В соответствии с методом неопределенных множителей Лагранжа и теоремой Куна—Таккера значения функции $t_L(d)$ определяются значениями функции $J_L(\lambda)$, равной (см. доказательство леммы 6.2.4)

$$J_L(\lambda) = \max_{\mathcal{A}_L} (I(Z; S) + \lambda I(Y; X | S)). \quad (6.4.32)$$

Таким образом, равенство (6.4.31), а следовательно и утверждение

леммы, будет доказано, если показать, что для любого $L > |X| \triangleq L_0$, $J_L(\lambda) = J_{L_0}(\lambda)$.

Пусть для некоторого фиксированного λ максимум в правой части (6.4.32) достигается на ансамбле $SXYZ \in \mathcal{A}_L$, задаваемом распределениями $p^*(s)$, $s \in S$, и $p^*(x|s)$, $x \in X$, $s \in S$. Будем считать, что в функциях $I(Z; S)$ и $I(Y; X|S)$ аргументы $p(x|s)$, $x \in X$, $s \in S$, фиксированы и равны $p^*(x|s)$, $x \in X$, $s \in S$. В этом случае

$$J_L(\lambda) = \max_{\{p(s)\}} (I(Z; S) + I(Y; X|S)). \quad (6.4.33)$$

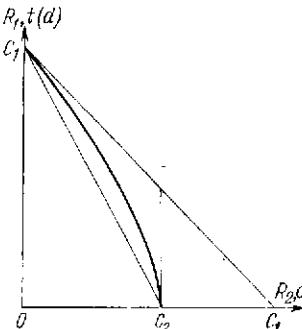


Рис. 6.4.8. Типичный вид функции $t(d)$.

в (6.4.33) по $p(s)$ и применяя теорему Куна—Таккера, получим, что необходимые и достаточные условия, которым должно удовлетворять распределение $p(s)$, $s \in S$, максимизирующее правую часть (6.4.33), суть

$$\begin{aligned} & \sum_X \sum_Z p^*(x|s) p_2(z|x) \log \frac{\sum_X p^*(x|s) p_2(z|x)}{p_2(z)} + \\ & + \lambda \sum_X \sum_Y p^*(x|s) p_1(y|x) \log \frac{p_1(y|x)}{\sum_X p^*(x|s) p_1(y|x)} \leq \lambda_1, \end{aligned} \quad (6.4.34)$$

где $p_1(y|x)$, $y \in Y$, $x \in X$, — переходные вероятности первой составляющей УШК, а $p_2(z|x)$, $z \in Z$, $x \in X$, — второй, и

$$p_2(z) = \sum_S p(s) \sum_X p^*(x|s) p_2(z|x) = \sum_X p(x) p_2(z|x). \quad (6.4.35)$$

По условию распределение $p^*(s)$ удовлетворяет соотношениям (6.4.34). Кроме того, так как правая часть (6.4.34) зависит от $p(s)$, $s \in S$, только через $p_2(z)$, $z \in Z$, то любое распределение $p(s)$, $s \in S$, порождающее распределение $p_2(z)$, $z \in Z$, равное распределению, порожденному $p^*(s)$, $s \in S$, будет также удовлет-

ворять соотношениям (6.4.34) и, следовательно, максимизировать правую часть (6.4.33). Из (6.4.35) следует, что, если для всех $x \in X$

$$\sum_S \tilde{p}(s) p^*(x|s) = \sum_S p^*(s) p^*(x|s),$$

то распределение $\tilde{p}(s)$, $s \in S$, удовлетворяет соотношениям (6.4.34). Из леммы 6.2.3 следует, что найдется такое распределение $\tilde{p}(s)$, $s \in S$, что число элементов $s \in S$, для которых $\tilde{p}(s) > 0$, не превосходит $|X|$. Отсюда следует (6.4.31), т. е. что $J_L(\lambda) = J_{L_0}(\lambda)$. Лемма доказана.

Типичный вид функции $t(d)$ показан на рис. 6.4.8. График этой функции лежит внутри трапеции, ограниченной координатными осями и прямыми $R_2 = C_2$ и $R_1 + R_2 = C_1$, где C_1 и C_2 — обычные пропускные способности первой и второй составляющих УШК соответственно. Действительно, для любого ансамбля $SXYZ \in \mathcal{A}$

$$\begin{aligned} I(S; Z) &= H(Z) - H(Z|S) \leq H(Z) - H(Z|SX) = \\ &= H(Z) - H(Z|X) = I(X; Z) \leq \max_{\{p(x)\}} I(X; Z) = C_2, \end{aligned}$$

причем первое неравенство превращается в точное равенство, когда распределение $p(x|s)$ вырождено, т. е. когда $s \in S$ однозначно определяет $x \in X$. В этом случае $I(X; Y|S) = H(X|S) - H(X|YS) = 0$. Поэтому наименьшее значение d , при котором $t(d) = 0$, есть $d = C_2$. С другой стороны, для любого ансамбля $SXYZ \in \mathcal{A}$ из леммы 6.4.1 следует, что

$$\begin{aligned} I(X; Y|S) + I(Z; S) &\leq I(X; Y|S) + I(Y; S) = \\ &= H(Y|S) - H(Y|X) + H(Y) - H(Y|S) = \\ &= I(X; Y) \leq \max_{\{p(x)\}} I(X; Y) = C_1, \end{aligned}$$

поэтому $t(d) + d \leq C_1$.

Пусть $\mathfrak{RI} \triangleq \{(R_1, R_2)\}$, $\mathfrak{RII} \triangleq \{(R_1, R_0)\}$: $R_1 \leq R_2$, $R_1 \leq t(R_2)$; $\mathfrak{RIII} \triangleq \{(R_1, R_2, R_0)\}$: $(R_1, R_2) \in \mathfrak{RI}$ при $R_2' = R_2 + R_0$. Мы показали, что любая пара скоростей $(R_1, R_2) \notin \mathfrak{RI}$ не является допустимой для передачи по УШК в ситуации КI. Так как общая информация может интерпретироваться как частная (может быть включена в частную), то тем самым показано также, что любая пара скоростей $(R_1, R_0) \notin \mathfrak{RII}$ и любая тройка скоростей $(R_1, R_2, R_0) \notin \mathfrak{RIII}$ не являются допустимыми для передачи по УШК в ситуациях КII и КIII соответственно. Таким образом, доказана следующая теорема.

Теорема 6.4.1. (обратная теорема кодирования). Для произвольного дискретного УШК без памяти

$$\mathbb{S}I \subseteq \mathbb{R}I, \quad \mathbb{S}II \subseteq \mathbb{R}II, \quad \mathbb{S}III \subseteq \mathbb{R}III,$$

где $\mathbb{S}I, \mathbb{S}II$ и $\mathbb{S}III$ — области пропускной способности для передачи по УШК в ситуациях KI, KII и $KIII$ соответственно.

6.4.5. Прямая теорема кодирования. Теперь мы покажем, что любая пара скоростей $(R_1, R_0) \in \mathbb{R}II$ допустима для дискретного УШК без памяти при передаче в ситуации KII . Для этого достаточно будет показать, что для любого распределения вероятностей на $SXYZ$, удовлетворяющего условию (6.4.8), найдется n и найдется код $G_n(\tilde{R}_1, \tilde{R}_0)$ со скоростями $\tilde{R}_1 = I(X; Y | S) - \delta$, $\tilde{R}_0 = I(S; Z) - \delta$, обеспечивающий вероятности ошибок $\lambda_1 \leq \varepsilon$, $\lambda_2 \leq \varepsilon$, где ε и δ — произвольные положительные числа. Установление этого факта завершит описание области пропускной способности для УШК.

Доказательство прямой теоремы кодирования будет проводиться методом случайного кодирования, аналогичным тому, который использовался при доказательстве прямой теоремы кодирования для каналов с множественным доступом. При построении кода для УШК будут использованы идеи, лежащие в основе построения кода для двоичного симметричного ШК (см. п. 6.4.3).

Пусть на множестве SX выбрано распределение вероятностей $p(s, x) = p(s)p(x|s)$. Введем в рассмотрение следующий ансамбль кодов $G_n(R_1, R_0)$, который будем обозначать через $\mathfrak{G}_n(R_1, R_0)$. С каждым кодом будем связывать множество $S_{R_0} = \{s_k\}$, $s_k \in S^n$, $k = \overline{1, M_0}$, $M_0 = 2^{nR_0}$, которое в дальнейшем будем называть множеством центров концентрации. Пусть $\{\mathbf{x}_{ik}\}$, $i = \overline{1, M_1}$, $M_1 = 2^{nR_1}$, $k = \overline{1, M_0}$, — множество слов кода $G_n(R_1, R_0)$ и S_{R_0} — связанное с ним множество центров концентрации. Каждому коду $G_n(R_1, R_0)$ из ансамбля $\mathfrak{G}_n(R_1, R_0)$ припишем вероятность

$$p(G_n) \triangleq \prod_{k=1}^{M_0} p(s_k) \prod_{i=1}^{M_1} p(\mathbf{x}_{ik} | s_k), \quad (6.4.36)$$

где

$$p(s) = \prod_{j=1}^n p(s^{(j)}), \quad p(\mathbf{x}|s) = \prod_{j=1}^n p(x^{(j)}|s^{(j)}). \quad (6.4.37)$$

Такое задание распределения вероятностей на $\mathfrak{G}_n(R_1, R_0)$ соответствует следующей процедуре случайного выбора кода. Вначале независимо в соответствии с распределением $p(s)$ выбираются M_0 центров концентрации, причем символы $s^{(j)} \in S$ каждой последовательности s выбираются независимо в соответ-

ствии с распределением $p(s)$. Затем для каждого фиксированного центра концентрации s_k независимо в соответствии с распределением $p(\mathbf{x}|s_k)$ выбираются M_1 кодовых слов $\{\mathbf{x}_{ik}\}$, $i = \overline{1, M_1}$, причем кодовые символы каждой последовательности \mathbf{x}_{ik} выбираются независимо в соответствии с распределением $p(x|s)$. В результате получается M_0M_1 кодовых слов $\{\mathbf{x}_{ik}\}$, $k = \overline{1, M_0}$, $i = \overline{1, M_1}$, для кода $G_n(R_1, R_0)$.

Для каждого кода $G_n(R_1, R_0)$ правило декодирования в УШК формулируется следующим образом. Декодер на выходе второй составляющей ШК по выходной последовательности $\mathbf{z} \in Z^n$ выносит решение о центре концентрации s_k , которому принадлежит передаваемое кодовое слово. Он выносит решение s_k , если

$$p^{(2)}(\mathbf{z}|s_k) < p^{(2)}(\mathbf{z}|s_{k'})$$

для всех $k' \neq k$, где

$$\begin{aligned} p^{(2)}(\mathbf{z}|s) &= \prod_{j=1}^n p^{(2)}(z^{(j)}|s^{(j)}), \\ p^{(2)}(z|s) &= \sum_X p(x|s) p_2(z|x). \end{aligned} \quad (6.4.38)$$

Декодер на выходе первой составляющей ШК работает в два этапа. На первом этапе по выходной последовательности $\mathbf{y} \in Y^n$ он выносит решение о том, какому центру концентрации принадлежит передаваемое кодовое слово. Выносится решение s_k , если

$$p^{(1)}(\mathbf{y}|s_k) > p^{(1)}(\mathbf{y}|s_{k'})$$

для всех $k' \neq k$, где

$$\begin{aligned} p^{(1)}(\mathbf{y}|s) &= \prod_{j=1}^n p^{(1)}(y^{(j)}|s^{(j)}), \\ p^{(1)}(y|s) &= \sum_X p(x|s) p_1(y|x). \end{aligned} \quad (6.4.39)$$

На втором этапе по выходной последовательности \mathbf{y} и по решению s_k , полученному на первом этапе, декодер выносит решение о том, какое кодовое слово передавалось, причем выносится решение \mathbf{x}_{ik} , если

$$p_1(\mathbf{y}|\mathbf{x}_{ik}) > p_1(\mathbf{y}|\mathbf{x}_{i'k})$$

для всех $i' \neq i$.

Обозначим через $\lambda_1(\mathbf{x}_{ik})$ и $\lambda_2(\mathbf{x}_{ik})$ условные вероятности ошибок первого и второго декодеров при условии, что передавалось

кодовое слово \mathbf{x}_{ik} . Согласно описанному выше правилу декодирования

$$\lambda_2(\mathbf{x}_{ik}) = \sum_{\mathbf{z}^n} p_2(\mathbf{z} | \mathbf{x}_{ik}) \varphi_k^{(2)}(\mathbf{z}), \quad (6.4.40)$$

$$\begin{aligned} \lambda_1(\mathbf{x}_{ik}) &\leq \sum_{\mathbf{y}^n} p_1(\mathbf{y} | \mathbf{x}_{ik}) (\varphi_k^{(11)}(\mathbf{y}) + \varphi_{ik}^{(12)}(\mathbf{y})) = \\ &= \lambda_{11}(\mathbf{x}_{ik}) + \lambda_{12}(\mathbf{x}_{ik}), \end{aligned} \quad (6.4.41)$$

где

$$\varphi_k^{(2)}(\mathbf{z}) = \begin{cases} 1, & \text{если } p^{(2)}(\mathbf{z} | s_k) \leq p^{(2)}(\mathbf{z} | s_{k'}) \text{ хотя бы} \\ & \text{для одного } k' \neq k, \\ 0, & \text{в противном случае;} \end{cases}$$

$$\varphi_k^{(11)}(\mathbf{y}) = \begin{cases} 1, & \text{если } p^{(1)}(\mathbf{y} | s_k) \leq p^{(1)}(\mathbf{y} | s_{k'}) \text{ хотя бы} \\ & \text{для одного } k' \neq k, \\ 0, & \text{в противном случае;} \end{cases}$$

$$\varphi_{ik}^{(12)}(\mathbf{y}) = \begin{cases} 1, & \text{если } p_1(\mathbf{y} | \mathbf{x}_{ik}) \leq p_1(\mathbf{y} | \mathbf{x}_{i'k}) \text{ хотя бы} \\ & \text{для одного } i' \neq i, \\ 0, & \text{в противном случае.} \end{cases}$$

Функции $\varphi_k^{(2)}(\mathbf{z})$, $\varphi_k^{(11)}(\mathbf{y})$, $\varphi_{ik}^{(12)}(\mathbf{y})$ равны единице соответственно в случаях, когда происходят ошибки декодирования на выходе второго декодера, и на первом и втором этапах на выходе первого декодера.

Теорема 6.4.2 (прямая теорема кодирования). Пусть фиксировано распределение вероятностей $p(s)p(x|s)$ на множестве SX и задан дискретный УШК без памяти $\{XYZ, P_1, P_2\}$, где $P_2 = P_1Q$. Тогда в ансамбле $\mathfrak{G}_n(R_1, R_0)$, распределение вероятностей на котором определяется соотношениями (6.4.36) и (6.4.37), существует код $G_n(R_1, R_0)$ такой, что $R_1 = \max\{0, I(X; Y|S) - \delta\}$, $R_2 = \max\{0, I(Z; S) - \delta\}$ и вероятности ошибок $\lambda_1 \leq \varepsilon$, $\lambda_2 \leq \varepsilon$, где ε, δ — произвольные положительные числа, и средние взаимные информации вычислены в соответствии с распределением $p(s)p(x|s)p_1(y|x)q(z|y)$ на $SXYZ$.

Доказательство. Мы покажем, что средние по ансамблю вероятности ошибок λ_1 и λ_2 могут быть сделаны сколь

угодно малыми. Для этого воспользуемся следующими очевидными оценками функций $\varphi_k^{(2)}(\mathbf{z})$, $\varphi_k^{(11)}(\mathbf{y})$ и $\varphi_{ik}^{(12)}(\mathbf{y})$:

$$\varphi_k^{(2)}(\mathbf{z}) = \left(\frac{\sum_{k' \neq k} p^{(2)}(\mathbf{z} | s_{k'})^{1/1+\rho}}{p^{(2)}(\mathbf{z} | s_k)^{1/1+\rho}} \right)^\rho, \quad \rho \geq 0, \quad (6.4.42)$$

$$\varphi_k^{(11)}(\mathbf{y}) = \left(\frac{\sum_{k' \neq k} p^{(1)}(\mathbf{y} | s_{k'})^{1/1+\rho}}{p^{(1)}(\mathbf{y} | s_k)^{1/1+\rho}} \right)^\rho, \quad \rho \geq 0, \quad (6.4.43)$$

$$\varphi_{ik}^{(12)}(\mathbf{y}) = \left(\frac{\sum_{i' \neq i} p_1(\mathbf{y} | \mathbf{x}_{i'k})^{1/1+\rho}}{p_1(\mathbf{y} | \mathbf{x}_{ik})^{1/1+\rho}} \right)^\rho, \quad \rho \geq 0. \quad (6.4.44)$$

Подставим неравенства (6.4.42)–(6.4.44) в (6.4.40) и (6.4.41) и усредним обе части полученных неравенств по ансамблю $\mathfrak{G}_n(R_1, R_0)$, выбирая $\rho \in [0, 1]$. При усреднении будем учитывать, что

а) центры концентрации s_k , $s_{k'}$, $k' \neq k$, статистически независимы в ансамбле кодов (см. (6.4.36));

б) при фиксированном s_k кодовые слова \mathbf{x}_{ik} , $\mathbf{x}_{i'k}$, $i' \neq i$, статистически независимы в ансамбле кодов (см. (6.4.36));

в) для всех $\rho \in [0, 1]$ имеет место неравенство $\xi^\rho \leq \bar{\xi}^\rho$, где ξ — произвольная случайная величина и черта сверху означает усреднение;

г) средние по ансамблю кодов значения вероятностей $\lambda_1(\mathbf{x}_{ik})$ и $\lambda_2(\mathbf{x}_{ik})$ не зависят от индексов i, k .

В результате операций усреднения и преобразований, подобных тем, которые производились в § 3.12, с учетом соотношений (6.4.36)–(6.4.39) получим следующие неравенства при $\rho \in [0, 1]$:

$$\begin{aligned} \bar{\lambda}_2 &\leq \sum_{\mathbf{z}^n} \sum_{s^n} p(s) \sum_{\mathbf{x}^n} p(\mathbf{x} | s_k = s) p_2(\mathbf{z} | \mathbf{x}) \times \\ &\quad \times \frac{\left(\sum_{k' \neq k} \sum_{s^n} p(s) p^{(2)}(\mathbf{z} | s_{k'})^{1/1+\rho} \right)^\rho}{p^{(2)}(\mathbf{z} | s_k = s)^{\rho/1+\rho}} = \\ &= \sum_{\mathbf{z}^n} \sum_{s^n} p(s) p^{(2)}(\mathbf{z} | s)^{1/1+\rho} \left(\sum_{k' \neq k} \sum_{s^n} p(s) p^{(2)}(\mathbf{z} | s_{k'})^{1/1+\rho} \right)^\rho \leq \\ &\leq M_0^\rho \sum_{\mathbf{z}^n} \left(\sum_{s^n} p(s) p^{(2)}(\mathbf{z} | s)^{1/1+\rho} \right)^{1+\rho} = 2^{-n(E_0^{(2)}(\rho) - \rho R_0)}; \end{aligned}$$

$$\tilde{\lambda}_{11} \leq M_0^{\rho} \sum_{S^n} \left(\sum_{y^n} p(s) p^{(1)}(y|s)^{1/1+\rho} \right)^{1+\rho} = 2^{-n(E_0^{(11)}(\rho) - \rho R_0)},$$

$$\begin{aligned} \tilde{\lambda}_{12} &\leq \sum_{Y^n} \sum_{S^n} p(s) \sum_{X^n} p(x|s_k=s) p_1(y|x) \times \\ &\quad \times \frac{\left(\sum_{i' \neq i} \sum_{X^n} p(x|s_k=s) p_1(y|x)^{1/1+\rho} \right)^{\rho}}{p_1(y|x)^{\rho/1+\rho}} = \\ &= M_1^{\rho} \sum_{S^n} p(s) \sum_{Y^n} \left(\sum_{X^n} p(x|s) p_1(y|x)^{1/1+\rho} \right)^{1+\rho} = \\ &= 2^{-n(E_0^{(12)}(\rho) - \rho R_1)}, \end{aligned}$$

где

$$E_0^{(2)}(\rho) = -\log \sum_Z \left(\sum_S p(s) p^{(2)}(z|s)^{1/1+\rho} \right)^{1+\rho},$$

$$E_0^{(11)}(\rho) = -\log \sum_Y \left(\sum_S p(s) p^{(1)}(y|s)^{1/1+\rho} \right)^{1+\rho},$$

$$E_0^{(12)}(\rho) = -\log \sum_S p(s) \sum_Y \left(\sum_X p(x|s) p_1(y|x)^{1/1+\rho} \right)^{1+\rho}.$$

Полагая

$$E_2(R_0) = \max_{0 \leq \rho \leq 1} (E_0^{(2)}(\rho) - \rho R_0),$$

$$E_{11}(R_0) = \max_{0 \leq \rho \leq 1} (E_0^{(11)}(\rho) - \rho R_0),$$

$$E_{12}(R_1) = \max_{0 \leq \rho \leq 1} (E_0^{(12)}(\rho) - \rho R_1),$$

получим, что

$$\tilde{\lambda}_2 \leq 2^{-nE_2(R_0)}, \quad \tilde{\lambda}_1 \leq 2^{-nE_{11}(R_0)} + 2^{-nE_{12}(R_1)}.$$

Непосредственной проверкой легко убедиться в том, что

$$\frac{\partial E_0^{(2)}(\rho)}{\partial \rho} \Big|_{\rho=0} = I(S; Z),$$

$$\frac{\partial E_0^{(11)}(\rho)}{\partial \rho} \Big|_{\rho=0} = I(S; Y),$$

$$\frac{\partial E_0^{(12)}(\rho)}{\partial \rho} \Big|_{\rho=0} = I(Y; X|S),$$

поэтому все три функции $E_2(R_0)$, $E_{11}(R_0)$, $E_{12}(R_1)$ одновременно положительны для любой пары скоростей (R_1, R_0) такой, что $R_1 < I(X; Y|S)$, $R_0 < \min(I(S; Z), I(S; Y))$. Согласно лемме 6.4.1 $I(S; Z) \leq I(S; Y)$ для любого дискретного УШК без памяти. Следовательно, $\tilde{\lambda}_1$ и $\tilde{\lambda}_2$ убывает к нулю с ростом n для любой пары скоростей (R_1, R_0) , $R_1 < I(X; Y|S)$, $R_0 < I(Z; S)$. Так как в ансамбле $\mathfrak{G}_n(R_1, R_0)$ существует код с вероятностями ошибок $\lambda_1 \leq \tilde{\lambda}_1$, $\lambda_2 \leq \tilde{\lambda}_2$, то теорема доказана.

Из доказанной теоремы и из определения множества $\mathfrak{H}II$ вытекает, что любая пара скоростей из $\mathfrak{H}II$ допустима для дискретного УШК без памяти и, следовательно, с учетом теоремы 6.4.1 область пропускной способности $\mathfrak{H}II$ для передачи по УШК в ситуации КИИ совпадает с $\mathfrak{H}II$. Так как множества $\mathfrak{H}I$, $\mathfrak{H}II$, $\mathfrak{H}III$ для УШК однозначно определяют друг друга, то тем самым с учетом определений множеств $\mathfrak{H}I$, $\mathfrak{H}II$, $\mathfrak{H}III$ доказано, что

$$\mathfrak{H}I = \mathfrak{H}I, \quad \mathfrak{H}II = \mathfrak{H}II, \quad \mathfrak{H}III = \mathfrak{H}III.$$

Задачи, упражнения и дополнения

6.1.1. Пусть заданы два зависимых двоичных источника без памяти U_X и U_Y , причем $X = Y = \{0, 1\}$, $p(x, y) = \alpha/2$ при $x \neq y$ и $p(x, y) = (1 - \alpha)/2$ при $x = y$. Постройте границы областей допустимых пар скоростей для всех α из множества $\{0; 0.3; 0.5; 0.7; 1\}$. Обратите внимание на роль параметра α .

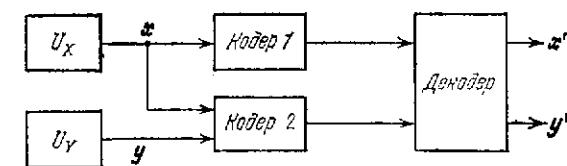


Рис. к задаче 6.1.2.

6.1.2. Пусть кодируются два зависимых источника без памяти с помощью двух независимых кодеров, причем первому доступен только выход источника U_X , а второму -- выходы обоих источников U_X и U_Y (см. рис.). Укажите область допустимых пар скоростей.

6.1.3. Рассмотрите задачу кодирования трех зависимых дискретных источников без памяти U_X , U_Y , U_Z с совместным ансамблем сообщений $\{XYZ\}$, $p(x, y, z) = \prod_{i=1}^n p(x^{(i)}, y^{(i)}, z^{(i)})$, $x \in X^n$, $y \in Y^n$, $z \in Z^n$. Покажите, что область допустимых троек скоростей (R_X, R_Y, R_Z) определяется как множество всех троек, для которых одновременно выполнены следующие условия:

$$R_X \geq H(X|YZ), \quad R_Y \geq H(Y|XZ), \quad R_Z \geq H(Z|XY),$$

$$R_X + R_Y \geq H(XY|Z), \quad R_X + R_Z \geq H(XZ|Y),$$

$$R_Y + R_Z \geq H(YZ|X), \quad R_X + R_Y + R_Z \geq H(XYZ).$$

Указание: эта и последующие задачи решаются сведением к случаю двух источников.

6.1.4. Рассмотрите модификацию предыдущей задачи, предположив, что кодеру 3 (см. рис.) доступны выходы всех источников U_X, U_Y и U_Z , а кодерам 1 и 2 выходы только своих источников U_X и U_Y соответственно. Покажите,

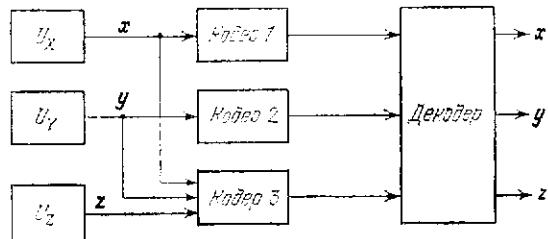


Рис. к задаче 6.1.4.

что область допустимых троек скоростей теперь определяется следующими условиями:

$$R_X \geq 0, R_Y \geq 0, R_Z \geq H(Z|XY),$$

$$R_X + R_Z \geq H(XZ|Y), \quad R_Y + R_Z \geq H(YZ|X),$$

$$R_X + R_Y + R_Z \geq H(XYZ).$$

6.1.5. Рассмотрите еще одну модификацию задачи 6.1.3, предположив теперь, что каждому кодеру доступны выходы двух источников, а именно первому

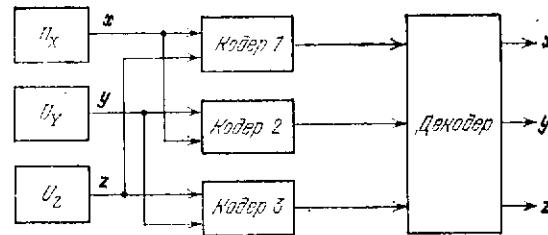


Рис. к задаче 6.1.5.

кодеру — выходы U_X, U_Z , второму — U_Y, U_X и третьему — выходы U_Z, U_Y (см. рис.). Покажите, что область допустимых троек скоростей определяется условиями:

$$R_X \geq 0, R_Y \geq 0, R_Z \geq 0,$$

$$R_X + R_Y \geq H(XY|Z), \quad R_X + R_Z \geq H(XZ|Y),$$

$$R_Y + R_Z \geq H(YZ|X), \quad R_X + R_Y + R_Z \geq H(XYZ).$$

6.1.6. Рассмотрите задачу независимого кодирования N зависимых дискретных источников без памяти $U_{X_i}, i = \overline{1, N}$, с совместным ансамблем сообщений $\{X_1, X_2, \dots, X_N\}$ и распределением вероятностей на последо-

вательностях, задаваемым по аналогии с соотношением (6.1.1). Покажите, что область \mathfrak{M} всех допустимых наборов скоростей $(R_i, i = \overline{1, N})$ определяется соотношениями

$$\sum_{i=1}^k R_{i_j} \geq H(X_{i_1}, \dots, X_{i_k} | X_{i_{k+1}}, \dots, X_{i_N}), \quad k = \overline{1, N},$$

для всех наборов индексов (i_1, i_2, \dots, i_k) , принимающих целые значения от 1 до N и таких, что $i_1 < i_2 < \dots < i_k$. Чему равно количество неравенств, определяющих область \mathfrak{M} .

6.2.1. В условиях задачи 6.1.1 постройте несколько точек границы области \mathfrak{M}_d допустимых пар скоростей при кодировании источника U_X с дополнительной информацией.

6.2.2. Рассмотрим следующее обобщение задачи кодирования с дополнительной информацией. Пусть совместно заданы три дискретных источника без памяти U_X, U_Y, U_Z с совместным ансамблем сообщений $(XYZ, p(x, y, z))$. Пусть каждый источник кодируется независимо друг от друга, каждый своим кодером.

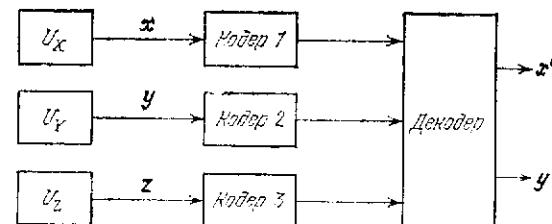


Рис. к задаче 6.2.2.

Декодер по трем кодовым словам, порожденным тремя кодерами, должен восстановить сообщения источников U_X и U_Y (см. рис.). Покажите, что при независимом кодировании источников U_X и U_Y с дополнительной информацией область $\mathfrak{M}_d = \{(R_X, R_Y, R_Z) : R_X \geq H(X|YU), R_Y \geq H(Y|XU), R_X + R_Y \geq H(XY|U), R_Z \geq I(Z; U)\}$ для некоторого ансамбля $XYZU, p(x, y, z, u) = p(xy|z)p(z|u)p(u)$ есть область допустимых троек скоростей.

6.2.3. Пусть в условиях предыдущей задачи кодирование и декодирование осуществляется по схеме, показанной на рис.: первый декодер восстанавливает

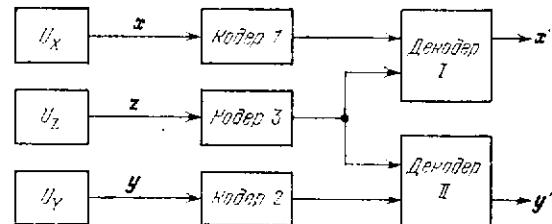


Рис. к задаче 6.2.3.

сообщения источника U_X , второй — источника U_Y , а источник U_Z является источником дополнительной информации для обоих декодеров. Покажите, что в этом случае область допустимых троек скоростей определяется соотношением

$$\mathfrak{M}_d = \{(R_X, R_Y, R_Z) : R_X \geq H(X|U), R_Y \geq H(Y|U), R_Z \geq I(Z; U)\}$$

для некоторого ансамбля $XYZU$, $p(x, y, z, u) = p(xy|z)p(z|u)p(u)$ (Вайнер [1975]).

6.3.1. Пусть задан дискретный КМД без памяти с переходными вероятностями $\{p(z|xy)\}$, $x \in X$, $y \in Y$, $z \in Z$. Рассмотрите передачу сообщений по такому КМД с помощью разделения времени. При этом для произвольного α , $0 \leq \alpha \leq 1$, α -я доли времени отводится для передачи сообщений первого источника, а $(1-\alpha)$ -я доли времени — второго. Покажите, что граница области допустимых пар скоростей (R_1, R_2) при такой передаче имеет вид прямой

$$R_1 = C_1 - \frac{C_1}{C_2} R_2,$$

где

$$\begin{aligned} C_1 &= \max_{y \in Y} \max_{p_1(x)} I(Z; X|y), \\ C_2 &= \max_{x \in X} \max_{p_2(y)} I(Z; Y|x). \end{aligned}$$

Объясните, как работают оба передатчика.

6.3.2. Пусть задан дискретный КМД без памяти с двоичными входами, двоичным выходом $X = Y = Z = \{0, 1\}$ и переходными вероятностями

$$p(z|xy) = \begin{cases} 1, & \text{если } x = y = z, \\ 1/2, & \text{если } x \neq y, \\ 0 & \text{в остальных случаях} \end{cases}$$

(см. рис.(а)).

а) Покажите, что пропускная способность этого канала, рассматриваемого как обычный канал с четверичным входом и двоичным выходом, равна 1; следовательно, в КМД максимальное значение суммы $R_1 + R_2 \leq 1$;

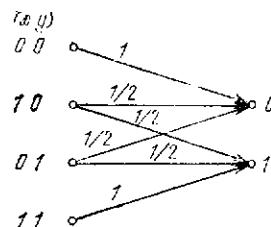


Рис. к задаче 6.3.2. (а)

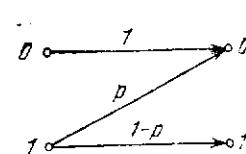


Рис. к задаче 6.3.2(б).

б) Покажите, что при разделении времени граница области допустимых скоростей задается уравнением $R_1 = C_1 - R_2$, где C_1 — пропускная способность так называемого z -канала с вероятностью ошибки $1/2$ (диаграмма переходов этого канала показана на рис. (б)). Покажите, что $C_1 = h(0.2) = 0.4 \cong 0.322$, где $h(x) = -x \log x - (1-x) \log(1-x)$. Таким образом, при использовании разделения времени в КМД $R_1 + R_2 \leq 0.322$.

в) Покажите, что в рассматриваемом КМД любая пара скоростей (R_1, R_2) , таких, что $R_1 + R_2 \leq 0.5$, $R_1 \leq h(0.25) = 0.5 \cong 0.311$, $R_2 \leq h(0.25) = 0.5 \cong 0.311$, допустима. Найдите распределения вероятностей $p_1(x)$, $p_2(y)$, при которых указанное множество пар скоростей совпадает с $\mathfrak{N}(p_1, p_2)$ и покажите, что имеется распределение $p_1(x)$ $p_2(x)$ такое, что $\mathfrak{N}(p_1, p_2) \not\subseteq \mathfrak{N}(\tilde{p}_1, \tilde{p}_2)$. Постройте область пропускной способности.

6.3.3. Рассмотрим дискретный КМД без памяти с тремя пользователями. Пусть X_1, X_2, X_3 — входные алфавиты, Z — выходной алфавит и $\{p(z|x_1, x_2, x_3)\}$ — переходные вероятности. Покажите, что область пропускных способностей описывается следующим образом. Пусть $UX_1X_2X_3Z$ — ансамбль с распределением вероятностей $p(u, x_1, x_2, x_3, z) = p(z|x_1x_2x_3)p(x_1|u)p(x_2|u)p(x_3|u)p(u)$, \mathcal{A} — множество всевозможных ансамблей такого вида. Тогда

$$\mathcal{C} = \{(R_1, R_2, R_3) : R_1 \leq I(X_1; Z|X_2X_3U),$$

$$R_2 \leq I(X_2; Z|X_1X_3U), R_3 \leq I(X_3; Z|X_1X_2U),$$

$$R_1 + R_2 \leq I(X_1X_2; Z|X_3U), R_1 + R_3 \leq I(X_1X_3; Z|X_2U),$$

$$R_2 + R_3 \leq I(X_2X_3; Z|X_1U), R_1 + R_2 + R_3 \leq I(X_1X_2X_3; Z|U)$$

для некоторого ансамбля $UX_1X_2X_3Z \in \mathcal{A}$ есть область пропускной способности.

6.3.4. В следующей задаче приводится пример КМД, для которого метод разделения времени является наилучшим методом передачи. Пусть КМД с тремя пользователями имеет двоичные входные алфавиты X_1, X_2, X_3 и двоичный выходной алфавит Z . Предположим, что выходной сигнал канала равен 1 только в том случае, когда хотя бы один из входных сигналов равен 1 (см. рис.). Канал осуществляет логическое суммирование (дизъюнкцию) входных сигналов и поэтому иногда называется дизъюнктивным КМД.

а) Покажите, что при разделении времени допустимы все тройки скоростей (R_1, R_2, R_3) такие, что $R_1 + R_2 + R_3 \leq 1$.

б) Покажите, что при независимом кодировании в рассматриваемом КМД для каждой допустимой тройки скоростей $(R_1, R_2, R_3) \in \mathfrak{N}(p_1, p_2, p_3)$, где множество $\mathfrak{N}(p_1, p_2, p_3)$ определяется аналогично множеству $\mathfrak{N}(p_1, p_2)$ (см. (6.3.7)), выполняются неравенства

$$R_1 + R_2 + R_3 \leq h(p_1(0) \cdot p_2(0) \cdot p_3(0)).$$

$$R_i + R_j = h(p_i(0) \cdot p_j(0))p_k(0), i \neq j \neq k \neq i; i, j, k = \overline{1, 3}.$$

$$R_i \leq p_i(0)p_k(0)h(p_i(0))$$

и, следовательно, при всех $p_1(x_1), p_2(x_2), p_3(x_3)$ сумма скоростей меньше 1.
в) Обобщите эту задачу на случай L пользователей.

6.3.5. Рассмотрим передачу сообщений трех источников по каналу с множественным доступом в ситуации, показанной на рис. Отличие этой ситуации от рассмотренной в § 6.3 состоит в наличии третьего источника информации U_0 , сообщения которого считаются известными обоим кодерам. Передача информации по КМД в ситуации с общим источником характеризуется тройкой скоростей

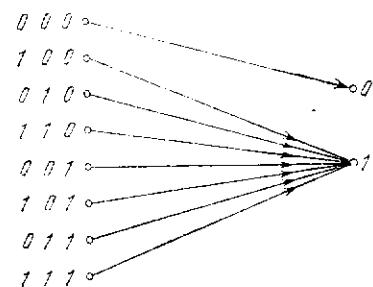


Рис. к задаче 6.3.4.

(R_1, R_2, R_0) и осуществляется с помощью кода $G_n(R_1, R_2, R_0) = \{x_{ik}, y_{jk}, A_{ijk}\}$, где $x_{ik} \in X^n$, $y_{jk} \in Y^n$, $i = \overline{1, M_1} = 2^{nR_1}$, $j = \overline{1, M_2} = 2^{nR_2}$, $k = \overline{1, M_0} = 2^{nR_0}$ — кодовые слова, а $A_{ijk} \in Z^n$ — решающие области, причем A_{ijk} и $A_{i'j'k'}$ не пересекаются для любых $(i, j, k) \neq (i', j', k')$. Покажите, что в этом случае область пропускной способности определяется следующим образом:

$$\begin{aligned} \mathcal{C} &= \{(R_1, R_2, R_0) : R_1 \leq I(X; Z | YUV), R_2 \leq I(Y; Z | XUV), \\ &\quad R_1 + R_2 \leq I(XY; Z | UV), R_1 + R_2 + R_0 \leq I(XY; Z | V) \end{aligned}$$

для некоторого ансамбля $VUXYZ \in \mathcal{A}$,

где \mathcal{A} — множество ансамблей $VUXYZ$, распределения вероятностей на которых имеют вид $p(vuxyz) = p(z|xy) p_1(x|u) p_2(y|u) p(u|v) p(v)$

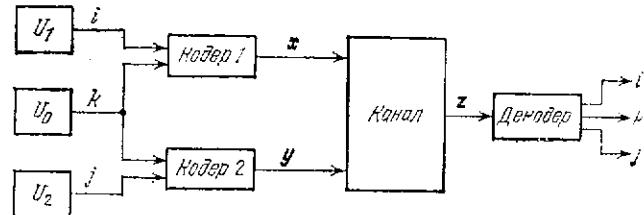


Рис. к задаче 6.3.5.

(Слепян и Вулф [1973II]). Указания: для доказательства обратной теоремы воспользуйтесь неравенством Фано. Для доказательства прямой теоремы используйте метод случайного кодирования, рассматривая ансамбль кодов $\mathfrak{G}_n(R_1, R_2, R_0)$ и используя следующее правило случайного выбора кода, определяющее распределение вероятностей на $\mathfrak{G}_n(R_1, R_2, R_0)$.

Вначале из множества U^n в соответствии с распределением $p(u) = \prod_{i=1}^n p(u^{(i)})$ независимо выбираются $M_0 = 2^{nR_0}$ последовательностей $\{u_k\}$, $k = \overline{1, M_0}$. Затем для каждой фиксированной последовательности u_k в соответствии с распределениями $p(x|u) = \prod_{i=1}^n p(x^{(i)}|u^{(i)})$ и $p(y|u) = \prod_{i=1}^n p(y^{(i)}|u^{(i)})$ независимо друг от

друга выбираются M_1 последовательностей $\{x_{ik}\}$, $i = \overline{1, M_1}$, и M_2 последовательностей $\{y_{jk}\}$, $j = \overline{1, M_2}$, из множеств X^n и Y^n соответственно. Декодирование осуществляется в два этапа. На первом этапе по принятой последовательности $z \in Z^n$ декодер определяет индекс k для переданных кодовых слов (этот индекс один и тот же для обоих кодовых слов и определяется сообщением общего источника), для чего он использует следующее правило: принимается решение о том, что второй индекс кодовых слов равен k , если $p(z|u_k) > p(z|u_{k'})$ для всех

$k' \neq k$, где $p(z|u) = \prod_{i=1}^n p(z^{(i)}|u^{(i)})$, $p(z|u) = \sum_X \sum_Y p(z|xy) p(x|u) p(y|u)$. На

втором этапе по выходной последовательности z и по определенному ранее значению индекса k декодер определяет индексы i и j передаваемых кодовых слов в соответствии со следующим правилом: декодер выносит решение о том, что значения этих индексов равны (i, j) , если $p(z|x_i y_j) > p(z|x_i y_{j'})$ для всех $(i', j') \neq (i, j)$.

6.3.6. Для двоичного суммирующего канала с множественным доступом в ситуации передачи с общим источником (см. предыдущую задачу) покажите,

что к области допустимых троек скоростей принадлежат все такие тройки (R_1, R_2, R_0) , что

- $R_1 + R_2 \leq 2/3$, $R_1 + R_2 + R_0 \leq \log 3$;
- $R_1 + R_2 \leq 1$, $R_1 + R_2 + R_0 \leq 2/3 + \frac{1}{2} \log 3$.

6.4.1. Покажите, что для дискретного ШК без памяти из допустимости тройки скоростей (R_1, R_2, R_0) относительно средних вероятностей ошибок следует, что эта тройка скоростей допустима также относительно максимальных вероятностей ошибок.

6.4.2. Рассмотрим дискретный УШК с тремя составляющими, причем вторая является ухудшенным вариантом первой, а третья — ухудшенным вариантом второй (см. рис.). Источник U_1 производит частную информацию для первого

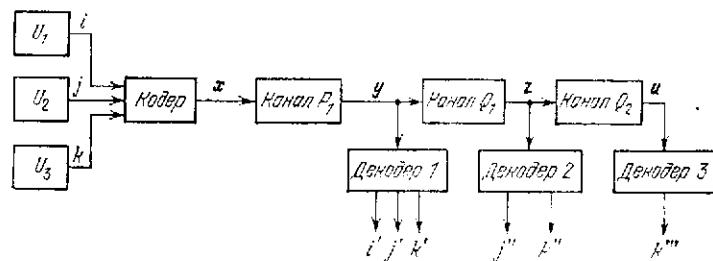


Рис. к задаче 6.4.2.

пользователя, источник U_2 производит общую информацию для первого и второго пользователей, источник U_3 производит общую информацию для всех трех пользователей. Первый канал имеет переходные вероятности $p_1(y|x)$, второй канал ухудшается с помощью канала с переходными вероятностями $q_1(z|y)$, третий канал ухудшается с помощью канала, имеющего переходные вероятности $q_2(u|z)$. Пусть \mathcal{C} — область пропускной способности в указанной ситуации передачи по УШК. Покажите, что

$$\begin{aligned} \mathcal{C} &= \{(R_1, R_2, R_3) : R_1 \leq I(X; Y | S_1 S_2), R_2 \leq I(S_1; Z | S_2), \\ &\quad R_3 \leq I(U; S_2) \text{ для некоторого ансамбля } S_1 S_2 X Y Z U \end{aligned}$$

с распределением вероятностей

$$p(s_1) p(s_2 | s_1) p(x | s_2) \cdot p_1(y|x) q_1(z|y) q_2(u|z),$$

$$s_1 \in S_1, s_2 \in S_2, x \in X, y \in Y, z \in Z, u \in U\}.$$

6.4.3. Для дискретного ШК без памяти с тремя пользователями опишите общую ситуацию передачи с семью источниками информации. На основе результата, полученного в предыдущей задаче, определите область пропускной способности в общей ситуации передачи по УШК.

6.4.4. Пусть фиксирован дискретный ШК без памяти $(XYZ, p_1(y|x), p_2(z|x))$ и пусть $\mathfrak{I} \triangleq (R_1, R_0) : R_1 \leq I(X; Y | S)$, $R_0 \leq \min\{I(S; Y), I(S; Z)\}$ для некоторого распределения вероятностей $p(sxyz) = p(s)p(x|s)p_1(y|x)p_2(z|x)$ на $SXYZ$. Покажите, что $\mathfrak{C} \ll \mathfrak{I}$ (Я. Кёрнер и К. Мартон [1977] показали, что для произвольного дискретного ШК без памяти $\mathfrak{C} \ll \mathfrak{I}$).

6.4.5. Пусть фиксирован дискретный ШК без памяти $(XYZ, p_1(y|x), p_2(z|x))$ и пусть $\mathfrak{R} \triangleq (R_1, R_2) : R_1 \leq I(S_1; Y)$, $R_2 \leq I(S_2; Z)$ для некоторого распределения вероятностей $p(s_1 s_2 xyz) = p(s_1)p(s_2)p(x|s_1 s_2)p_1(y|x)p_2(z|x)$ на

$S_1 S_2 XYZ\}.$ Покажите, что $\bar{\mathfrak{M}} \subseteq \mathbb{G}$, где $\bar{\mathfrak{M}}$ означает выпуклое замыкание множества \mathfrak{M} (Э. Ван дер Мейлен (1975), Т. Ковер (1975)).

6.4.6. Рассмотрим так называемый канал Блекуэла. Пусть ШК без памяти $\{XYZ, p_1(y|x) p_2(z|x)\}$ задается следующим образом: $X = \{0, 1, 2\}$, $Y = Z = \{0, 1\}$ и

$$p_1(y|x) = \begin{cases} 1, & \text{если } x = y \text{ или } x = 2, y = 1, \\ 0, & \text{если } x \neq y, x \neq 2; \end{cases}$$

$$p_2(z|x) = \begin{cases} 1, & \text{если } x = z \text{ или } x = 2, z = 0, \\ 0, & \text{если } x \neq z, x \neq 2 \end{cases}$$

(см. рис.). Пусть \mathfrak{M}_1 — подмножество множества \mathfrak{M} , определенного в задаче 6.4.5, которое порождается распределениями вероятностей вида $p(s_1=1)=p$, $p(s_2=0)=\alpha$, $s_1 \in S_1 = \{0, 1\}$, $s_2 \in S_2 = \{0, 1\}$ и $p(x=0|s_1=0, s_2=0) = p(x=1|s_1=0, s_2=0) = p(x=1|s_1=1, s_2=0) = p(x=1|s_1=1, s_2=1) = p(x=2|s_1=0, s_2=1) = 1$. Пусть, кроме того, $\mathfrak{M}_2 \subset \mathfrak{M}$ — аналогичное множество, порождаемое распределениями вероятностей вида: $p(s_1=1)=\alpha$, $p(s_2=0)=p$, $s_1 \in S_1 = \{0, 1\}$, $s_2 \in S_2 = \{0, 1\}$ и $p(x=0|s_1=0, s_2=0) = p(x=0|s_1=1, s_2=0) = p(x=1|s_1=1, s_2=1) = p(x=2|s_1=0, s_2=1) = 1$.

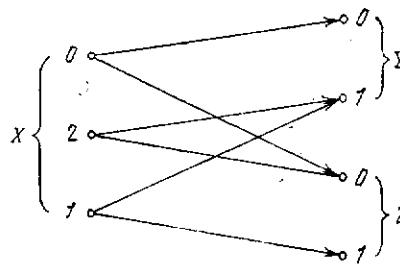


Рис. к задаче 6.4.6.

а) Покажите, что граница области \mathfrak{M}_1 задается параметрически следующим образом:

$$R_1 = h(p), \quad R_2 = C(p), \quad 0 \leq p \leq 1,$$

а граница области \mathfrak{M}_2 — следующим образом:

$$R_1 = C(p), \quad R_2 = h(p), \quad 0 \leq p \leq 1,$$

где $C(p)$ — пропускная способность двоичного z -канала (см. рис. к задаче 6.3.2(6)).

б) Вычислите $C(p)$.

в) Покажите, что для ШК Блекуэла множество $\bar{\mathfrak{M}} \triangleq \{R_1 \cup \mathfrak{M}_1 \cup \mathfrak{M}_2 | (R_1 = 1, R_2 = 0), (R_1 = 0, R_2 = 1)\}$ принадлежит области пропускной способности \mathbb{G} .

г) Постройте границу множества $\bar{\mathfrak{M}}$ и сравните ее с границей разделения времени для случая, когда α -я доля времени используется для передачи информации первому получателю, а $(1-\alpha)$ -я доля времени — второму, $\alpha \in [0, 1]$.

6.4.7. Снова рассмотрим ШК Блекуэла. Пусть $\mathfrak{M}^* \triangleq \{(R_1, R_2) : R_1 \leq H(Y), R_2 \leq H(Z), R_1 + R_2 \leq H(YZ)\}$ для некоторого распределения $p(x, y, z) = p(x)p_1(y|x)p_2(z|x)$ на XYZ . Покажите, что $\mathfrak{M}^* \subseteq \mathbb{G}$. Постройте границу множества \mathfrak{M}^* (С. И. Гельфанд (1977) показал, что для канала Блекуэла $\mathfrak{M}^* = \mathbb{G}$; М. С. Пинскер (1978) обобщил этот результат на произвольный неимущий ШК без памяти, т. е. на такой канал, у которого все переходные вероятности равны либо 0, либо 1).

КРАТКИЙ ИСТОРИЧЕСКИЙ КОММЕНТАРИЙ И ЛИТЕРАТУРА

Задача кодирования зависимых источников без памяти впервые была поставлена и решена Д. Слепяном и Дж. Вулфом [17]. Доказательства теорем кодирования, приведенные в § 6.1, принадлежат Р. Алсведе и Я. Кёрнеру [3].

Задача кодирования источников с дополнительной информацией впервые была поставлена и решена для одного частного случая двоичных источников А. Вайнером [6]. Решение этой задачи для произвольных зависимых источников без памяти было независимо получено Р. Алсведе и Я. Кёрнером [3] и А. Вайнером [17]. Обобщение задачи кодирования зависимых источников было рассмотрено в работах С. И. Гельфанда и М. С. Пинскера [10] и Я. Кёрнера и К. Мартон [13].

Постановка задачи кодирования в канале с множественным доступом впервые встречается у К. Шеннона [19]. Решение этой задачи принадлежит Р. Алсведе [1, 2]. Более общая постановка (см. задачу 6.3.5) рассмотрена в работе Д. Слепяна и Дж. Вулфа [18].

Постановка задачи кодирования в широковещательных каналах принадлежит Т. Коверу [14]. В этой же работе изложены некоторые идеи по кодированию в ШК. Область пропускной способности УШК была определена в работе Р. Галлагера [8] и Р. Алсведе и Я. Кёрнера [3].

Имеется значительное число работ, посвященных кодированию в системах с большим числом пользователей. В приведенной ниже библиографии содержатся лишь те из них, результаты которых использованы либо в основном тексте шестой главы, либо в задачах к этой главе. Интересующемуся читателю можно порекомендовать обзоры ван дер Мейлена [5] и С. И. Гельфанда и В. В. Прелова [11].

1. А л с в е д е (Ahlswede R.). Multi-Way Communication Channels. — Proc. 2nd Int. Symp. Information Theory. — Tsahkadsor, Armenian SSR: Publishing House of the Hungarian Academy of Sciences, 1973, 23—52.
2. А л с в е д е (Ahlswede R.). The Capacity Region of a Channel with Two Senders and Two Receivers. — Ann. Prob., 1974, vol. 2, 805—814.
3. А л с в е д е и К ё р н е р (Ahlswede R., Körner J.). Source Coding with Side Information and a Converse for Degraded Broadcast Channels. — IEEE Trans. on Inf. Theory, 1975, vol. IT-21, № 6.
4. В а н д е р М е й л е н (Van der Meulen E.). Random Coding Theorems for the General Discrete Memoryless Broadcast Channel. — IEEE Trans. on Inf. Theory, 1975, vol. IT-21, № 2.
5. В а н д е р М е й л е н (Van der Meulen E.). A Survey of Multi-Way Channels in Information Theory: 1961—1976. — IEEE Trans. on Inf. Theory, 1977, vol. IT-23, № 1.
6. В ай н е р (Wyner A.). A theorem on the entropy of certain binary sequences and applications, Part II. — IEEE Trans. on Inf. Theory, 1973, vol. IT-19, № 6.
7. В ай н е р (Wyner A.). On Source Coding with Side Information at the Decoder. — IEEE Trans. on Inf. Theory, 1975, vol. IT-21, № 3.
8. Г а л л а г е р Р. Пропускная способность и кодирование для ухудшающихся широковещательных каналов. — Проблемы передачи информации, 1974, т. 10, № 3.
9. Г е л ь ф а н д С. И. Пропускная способность одного широковещательного канала. — Проблемы передачи информации, 1977, т. 13, № 3.
10. Г е л ь ф а н д С. И., Пинскер М. С. Кодирование источников по наблюдениям с неполной информацией. — Проблемы передачи информации, 1979, т. 15, № 2.
11. Г е л ь ф а н д С. И., Прелов В. В. Связь с многими пользователями. — В сб.: Итоги науки и техники. Теория вероятностей, математическая статистика, техническая кибернетика, 15. — М.: ВИНИТИ, 1978.

12. Кёрнер, Мартон (Körner J., Marton K.) General Broadcast Channels with Degraded Message Sets. — IEEE Trans. on Inf. Theory, 1977, vol. 23, № 1.
13. Кёрнер, Мартон (Körner J., Marton K.) Images of a Set via Two Channels and their Role in Multi-user Communication. — IEEE Trans. on Inf. Theory, 1977, vol. 23, № 6.
14. Ковер (Cover T.). Broadcast Channels. — IEEE Trans. on Inf. Theory, 1972, vol. 18, № 1.
15. Ковер (Cover T.). An Achievable Rates Region for the Broadcast Channel. — IEEE Trans. on Inf. Theory, 1975, vol. 21, № 3.
16. Пинскер М. С. Пропускная способность широковещательных каналов без шумов. — Проблемы передачи информации, 1978, т. 14, № 2.
17. Слепян, Вульф (Slepian D., Wolf J.) Noiseless coding of Correlated Information Sources. — IEEE Trans. on Inf. Theory, 1973, vol. 19, № 4.
18. Слепян, Вульф (Slepian D., Wolf J.) A Coding Theorem for Multiple Access Channels with Correlated Sources, Bell Syst. Tech. J., 1973, vol. 52.
19. Шеннон (Shannon C. E.) Two-Way Communication Channels. — Proc. 4-th Berkeley Symp. on Math. Stat. and Prob., 1961, vol. 1. [Русский перевод: Шеннон К., Двусторонние каналы связи. — В сб. работ по теории информации и кибернетике. — М.: ИЛ, 1963.]

ПРИЛОЖЕНИЕ I

Энтропия двоичного ансамбля

p	$-\log p$	$-p \log p$	$h(p)$	$-(1-p) \times$ $\times \log(1-p)$	$-\log(1-p)$	$1-p$
0,01	6,643	0,066	0,081	0,014	0,014	0,99
0,02	5,644	0,113	0,141	0,028	0,029	0,98
0,03	5,059	0,152	0,194	0,042	0,044	0,97
0,04	4,644	0,186	0,242	0,056	0,059	0,96
0,05	4,322	0,216	0,286	0,070	0,074	0,95
0,06	4,059	0,243	0,327	0,084	0,089	0,94
0,07	3,936	0,268	0,366	0,097	0,105	0,93
0,08	3,644	0,291	0,402	0,111	0,120	0,92
0,09	3,474	0,313	0,436	0,124	0,136	0,91
0,10	3,322	0,332	0,469	0,137	0,152	0,90
0,11	3,184	0,350	0,499	0,150	0,168	0,89
0,12	3,059	0,367	0,529	0,162	0,184	0,88
0,13	2,943	0,383	0,557	0,175	0,201	0,87
0,14	2,836	0,397	0,584	0,187	0,217	0,86
0,15	2,737	0,411	0,610	0,199	0,234	0,85
0,16	2,644	0,423	0,634	0,211	0,252	0,84
0,17	2,556	0,434	0,658	0,223	0,269	0,83
0,18	2,474	0,445	0,680	0,235	0,286	0,82
0,19	2,396	0,455	0,701	0,246	0,304	0,81
0,20	2,322	0,464	0,722	0,257	0,322	0,80
0,21	2,252	0,473	0,741	0,269	0,340	0,79
0,22	2,184	0,481	0,760	0,279	0,358	0,78
0,23	2,120	0,488	0,778	0,290	0,377	0,77
0,24	2,059	0,494	0,795	0,301	0,396	0,76
0,25	2,000	0,500	0,811	0,311	0,415	0,75
0,26	1,943	0,505	0,827	0,321	0,434	0,74
0,27	1,889	0,510	0,841	0,331	0,454	0,73
0,28	1,836	0,514	0,855	0,341	0,474	0,72
0,29	1,786	0,518	0,869	0,351	0,494	0,71
0,30	1,737	0,521	0,881	0,360	0,514	0,70
0,31	1,690	0,524	0,893	0,369	0,535	0,69
0,32	1,644	0,526	0,904	0,378	0,556	0,68
0,33	1,599	0,528	0,915	0,387	0,578	0,67
0,34	1,556	0,529	0,925	0,396	0,599	0,66
0,35	1,514	0,530	0,934	0,404	0,621	0,65
0,36	1,474	0,531	0,943	0,412	0,644	0,64
0,37	1,434	0,531	0,951	0,420	0,667	0,63
0,38	1,396	0,530	0,958	0,428	0,690	0,62
0,39	1,358	0,529	0,965	0,435	0,713	0,61
0,40	1,322	0,529	0,971	0,442	0,737	0,60
0,41	1,286	0,527	0,976	0,449	0,761	0,59
0,42	1,252	0,526	0,981	0,455	0,786	0,58
0,43	1,217	0,523	0,986	0,462	0,811	0,57
0,44	1,184	0,521	0,989	0,468	0,836	0,56
0,45	1,152	0,518	0,993	0,474	0,862	0,55
0,46	1,120	0,515	0,995	0,480	0,889	0,54
0,47	1,089	0,512	0,997	0,485	0,916	0,53
0,48	1,059	0,508	0,999	0,491	0,943	0,52
0,49	1,029	0,504	0,999	0,495	0,971	0,51
0,50	1,000	0,500	1,000	0,500	1,000	0,50

Значения интеграла вероятности

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{x^2}{2}} dx \text{ для } 0,00 \leq z \leq 4,99; \Phi(-z) = 1 - \Phi(z)$$

	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7703	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8905	0,8925	0,8944	0,8962	0,8977	0,9007	0,9047
1,3	0,9032	0,9049	0,9065	0,9082	0,9098	0,9114	0,9130	0,9146	0,9162	0,9174
1,4	0,9192	0,9207	0,9220	0,9236	0,9250	0,9264	0,9278	0,9292	0,9306	0,9319
1,5	0,9331	0,9348	0,9354	0,9369	0,9382	0,9394	0,9406	0,9417	0,9429	0,9440
1,6	0,9452	0,9463	0,9473	0,9485	0,9495	0,9505	0,9515	0,9525	0,9535	0,9544
1,7	0,9554	0,9563	0,9572	0,9581	0,9590	0,9594	0,9608	0,9616	0,9624	0,9632
1,8	0,9640	0,9648	0,9652	0,9663	0,9671	0,9674	0,9685	0,9692	0,9699	0,9706
1,9	0,9719	0,9719	0,9725	0,9730	0,9738	0,9744	0,9750	0,9755	0,9761	0,9767
2,0	0,9772	0,9777	0,9781	0,9788	0,9793	0,9798	0,9803	0,9807	0,9812	0,9816
2,1	0,9821	0,9825	0,9827	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9853
2,2	0,9861	0,9861	0,9867	0,9871	0,9874	0,9877	0,9881	0,9884	0,9887	0,9889
2,3	0,9892	0,9895	0,9896	0,9898	0,9900	0,9903	0,9906	0,9908	0,9911	0,9915
2,4	0,9818	0,9824	0,9824	0,9825	0,9825	0,9826	0,9827	0,9828	0,9831	0,9836

	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
2,5	0,9379	0,9396	0,9413	0,9429	0,9446	0,9461	0,9476	0,9491	0,9506	0,9520
2,6	0,9533	0,9547	0,9564	0,9573	0,9585	0,9597	0,9609	0,9620	0,9631	0,9642
2,7	0,9653	0,9663	0,9673	0,9683	0,9692	0,9702	0,9711	0,9719	0,9726	0,9736
2,8	0,9745	0,9752	0,9759	0,9767	0,9774	0,9781	0,9788	0,9794	0,9798	0,9804
2,9	0,9813	0,9819	0,9825	0,9830	0,9835	0,9841	0,9846	0,9851	0,9855	0,9860
3,0	0,9865	0,9869	0,9873	0,9877	0,9881	0,9885	0,9889	0,9893	0,9896	0,9899
3,1	0,9832	0,9846	0,9857	0,9860	0,9865	0,9868	0,9872	0,9875	0,9878	0,9880
3,2	0,9831	0,9836	0,9839	0,9841	0,9842	0,9843	0,9844	0,9846	0,9848	0,9849
3,3	0,9851	0,9866	0,9873	0,9878	0,9883	0,9888	0,9892	0,9895	0,9898	0,9900
3,4	0,9866	0,9875	0,9882	0,9887	0,9891	0,9895	0,9898	0,9901	0,9903	0,9905
3,5	0,9877	0,9879	0,9882	0,9884	0,9886	0,9888	0,9890	0,9892	0,9894	0,9896
3,6	0,9884	0,9894	0,9896	0,9897	0,9898	0,9899	0,9900	0,9901	0,9902	0,9903
3,7	0,9889	0,9892	0,9894	0,9896	0,9898	0,9900	0,9901	0,9902	0,9903	0,9904
3,8	0,9847	0,9856	0,9862	0,9867	0,9872	0,9876	0,9880	0,9884	0,9887	0,9890
3,9	0,9851	0,9858	0,9863	0,9868	0,9873	0,9878	0,9882	0,9886	0,9890	0,9893
4,0	0,9868	0,9874	0,9879	0,9882	0,9886	0,9889	0,9892	0,9895	0,9898	0,9900
4,1	0,9840	0,9846	0,9852	0,9857	0,9863	0,9868	0,9873	0,9877	0,9881	0,9884
4,2	0,9866	0,9873	0,9879	0,9884	0,9889	0,9894	0,9898	0,9902	0,9905	0,9908
4,3	0,9847	0,9852	0,9857	0,9862	0,9867	0,9872	0,9876	0,9880	0,9884	0,9887
4,4	0,9845	0,9851	0,9858	0,9863	0,9868	0,9873	0,9878	0,9882	0,9886	0,9889
4,5	0,9860	0,9867	0,9873	0,9879	0,9885	0,9890	0,9895	0,9900	0,9904	0,9907
4,6	0,9878	0,9884	0,9889	0,9894	0,9899	0,9904	0,9909	0,9914	0,9918	0,9921
4,7	0,9886	0,9891	0,9896	0,9901	0,9906	0,9911	0,9916	0,9921	0,9925	0,9928
4,8	0,9842	0,9847	0,9852	0,9857	0,9862	0,9867	0,9872	0,9877	0,9881	0,9885
4,9	0,9850	0,9855	0,9860	0,9865	0,9870	0,9875	0,9880	0,9885	0,9889	0,9893

Пример: (3,57) = 0,938215 = 0,9998215.

Белый гауссовский шум 126
 — частотно ограниченный 126

Вектор вероятностный 112
 — гауссовский 110
 — случайный 107
 — собственный 108

Вероятностные ансамбли дискретные 12
 — непрерывные 90
 — статистически зависимые 13, 92
 — независимые 13

Вероятность ошибки декодирования 161
 — условная 14

Выпуклая область 112
 — функция 112

Выплокость взаимной информации 114, 116

Граница аддитивная для вероятности ошибки 243

— Варшамова—Гильберта 234

— сферической упаковки 227

— Чернова 235

Декодирование по максимуму апостериорной вероятности 195
 — правдоподобия 195
 — минимуму расстояния Хемминга 195

Дельта-функция 94

Демодулятор 155

Дерева порядок 58
 — ярус 58

Дерево кодовое 58

Дискретизация 93

Дискретный ансамбль 12

Дисперсия 20, 120

Допустимая пара скоростей 337, 347, 366, 381
 — точка 133
 — тройка скоростей 380

Допустимое направление 133

Закон больших чисел 23

Интеграл вероятностей 141
 — от случайного процесса 121

Интенсивность белого шума 126

Информационная емкость канала 167
 — дискретного без памяти 169
 — симметричного 180
 — с аддитивным по модулю L шумом 183
 — непрерывного канала 263
 — с аддитивным белым гауссовским шумом 266

Информация взаимная 81, 96
 — собственная 24
 — средняя 84, 98, 128
 — условная 28

Источник без памяти 19
 — дискретный 17
 — марковский 73, 75
 — непрерывный 283
 — неэргодический 49
 — постоянный 19
 — стационарный 18
 — эргодический 46

Итеративный алгоритм вычисления пропускной способности 179
 — функции $E_0(p)$ 243
 — $E_{sp}(R, Q)$ 245

Канал без памяти 157
 — предвосхищения 158
 — векторный 276
 — дискретный 155
 — непрерывный 156
 — параллельный 276
 — полуинтервальный 156
 — с аддитивным шумом 261
 — — — гауссовским 261
 — — — по модулю L 183
 — — — дискретным временем 156, 249
 — — — межсимвольной интерференцией 238
 — — — множественным доступом 364
 — — — двоичный суммирующий 367
 — — — дизьюнктивный 405
 — — — непрерывным временем 156, 261
 — — — симметричный 180
 — — — двоичный 158, 181
 — — — стирающий 181

Канал симметричный по входу 180
 — — — выходу 180
 — строго симметричный по входу 236
 — широковещательный 378
 — — Блекуэлла 408
 — — двоичный симметричный 384
 — — ухудшающийся 382

Квантование оптимальное 325

Код для канала с множественным доступом 364
 — кодирования источника с дополнительной информацией 346
 — — — критерием качества 288
 — — — пары дискретных источников 336
 — — широковещательного канала 379, 381
 — — — источника 34
 — — — канала 159
 — — — непрерывного 251, 262
 — — — максимальный 187, 242
 — — — неравномерный 34
 — — — оптимальный 61, 64
 — — — префиксный 56
 — — — равномерный 34
 — — со свойством однозначного декодирования 56
 — с фиксированной композицией 220

Кода алфавит 34
 — длина 34, 159, 251, 262
 — объем 159

Кодовые символы 34
 — слова 34, 159

Количество информации между гауссовскими ансамблями 99

Композиция 219

Корреляционный момент 21

Критерий качества 284
 — — вероятностный 286
 — — квадратический 287

Куна—Таккера условия 135

Лагранжа множитель 131
 — функция 131

Линейная оценка случайного процесса 149

Линейное нормированное пространство случайных величин 142

Мак-Миллана лемма 50

Марковская последовательность 17

Математическое ожидание 20
 — — условное 72

Матрица корреляционная 105
 — неотрицательно определенная 108
 — обратная 105

Матрица ортогональная 109
 — переходных вероятностей 179
 — положительно определенная 108
 — стохастическая 116
 — теплицева 320

Метод неопределенных множителей Лагранжа 130

Минимальное расстояние кода 241

Множество высоковероятное 41
 — конечное 11
 — последовательностей с фиксированной композицией 219

Модель Гильберта 237

Модулятор 155

Мощность средняя 103, 143

Непрерывный ансамбль 90

Неравенство для логарифма 26
 — Иенсена 147
 — Коши—Буняковского 148
 — Крафта 59
 — Файнстейна 185
 — Фано 164
 — Чебышева 22

Область допустимых скоростей при кодировании зависимых источников 337
 — — — — источника с дополнительной информацией 347
 — — — — пропускной способности канала с множественным доступом 366
 — — — — широковещательного канала 380, 381, 382
 — — — — решающая 159

Объем n -мерного шара 310, 328

Отношение сигнал—шум 273
 — — — на бит 274

Ошибка равномерного кодирования 36

Плотно упакованный код 219

Площадь n -мерной сферы 309, 328

Преобразование ортогональное 109

Производящая функция моментов 235

Пропускная способность канала 161
 — — — без памяти 191
 — — — с аддитивным гауссовским шумом 261
 — — — — эргодическим шумом 193
 — — — непрерывного канала с аддитивным белым гауссовским шумом 272

Разделения времени метод 337

Разложение Карунена—Лозва 124
 — случайного процесса в ортогональный ряд 123

- Распределение вероятностей 12
 - условное 14
 - Расстояние Хемминга 195
 - Регулярности условие 134
 - Свертка вероятностей 384
 - Свойство аддитивности взаимной информации 83
 - энтропии 29
 - Скорость вычислительная 243
 - кода 160, 251, 262
 - кодирования неравномерного 36
 - равномерного 55
 - критическая 209
 - создания информации 37, 56
 - Случайного кодирования метод 197
 - экспонента 202
 - Случайные величины дискретные 19
 - непрерывные 89
 - совместно гауссовские 106
 - статистически независимые 21
 - Случайный процесс гауссовский 125
 - непрерывного времени 120
 - стационарный 125
 - Событие элементарное 12
 - Согласованности условие 17, 158
 - Спектральная плотность мощности 121, 320
 - Стационарности условие 157, 249
 - Стирлинга формула 74
 - Сфера n -мерная 301
 - «твердая» 302
 - Хемминга 216
 - Сходимость в среднеквадратическом 121
 - Теорема Мерсера 149
 - Тест-канал 385
 - Точная верхняя грань 167
 - нижняя грань 291
 - Узел дерева 58
 - концевой 58
-
- Функции гармонические 126
 - ортогональные 123
 - ортонормированные 123
 - Функция Галлагера 201
 - корреляционная 120
 - нормированная 123
 - плотности вероятностей 89
 - — — гауссовская 101
 - распределения 89
 - Хаффмена метод 68
 - Центральный момент 20
 - Центры концентрации 385
 - Частотная характеристика фильтра канала 272
 - Числа собственные 108, 124
 - Шар n -мерный 327
 - Шеннона—Фано метод 65
 - Экспонента сферической упаковки 218
 - 227
 - Энтропия 25
 - на сообщение 32
 - одного распределения относительно другого 203
 - относительная 101
 - условная 28, 30
 - Эпсилон-сетка для аппроксимации точек n -мерной сферы 304
 - Эпсилон-скорость создания информации 289
 - Эпсилон-энтропия 291
 - гауссовой случайной величины 297
 - случайного вектора 313
 - стационарного гауссовского процесса 319
 - Эргодичность блочная 70