



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ



*Иновационная образовательная программа
«Наукоемкие технологии и экономика инноваций»
Московского физико-технического института
(государственного университета)
на 2006–2007 годы*

Э.М. Габидулин, Н.И. Пилипчук

**ЛЕКЦИИ
ПО ТЕОРИИ ИНФОРМАЦИИ**

519

Г1

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное агентство по образованию

Государственное образовательное учреждение

высшего профессионального образования

Московский физико-технический институт

(государственный университет)

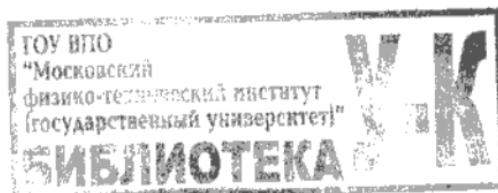
Кафедра радиотехники

Э.М. Габидулин, Н.И. Пилипчук

ЛЕКЦИИ ПО ТЕОРИИ ИНФОРМАЦИИ

Рекомендовано

Учебно-методическим объединением
высших учебных заведений Российской Федерации
по образованию в области прикладных математики и физики
в качестве учебного пособия



МОСКВА 2007



100086500148

-148

УДК 62-50
ББК 22.18я73
Г 12

Р е ц е н з е н т ы:

Кафедра теории вероятностей
механико-математического факультета МГУ им. М.В. Ломоносова
доктор физико-математических наук, профессор *А.Г. Дьячков*

Кандидат технических наук,
старший сотрудник ИППИ РАН *В.Б. Афанасьев*

Габидулин Э.М., Пилипчук Н.И.

Г 12 Лекции по теории информации: Учебное пособие. – М.:
МФТИ, 2007. – 214 с.

ISBN 5-7417-0197-3

Данное учебное пособие предназначено для студентов технических вузов, начинающих изучать теорию информации. Оно также может быть полезно инженерам, интересующимся шенноновской теорией информации. Здесь на простых моделях представлены с доказательством основные теоремы Шеннона. Особое место занимает материал, посвященный доказательству границ оптимального декодирования, связанных с именами Бхаттачария и Галлагера. В нем продемонстрированы современные методы исследования, используемые в этой области науки. Для чтения необходимо знание основ теории вероятностей. Некоторые сведения, облегчающие понимание излагаемых вопросов, приведены в Приложении.

УДК 62-50

ISBN 5-7417-0197-3

© Московский физико-технический институт
(государственный университет), 2007
© Габидулин Э.М., Пилипчук Н.И., 2007

Оглавление

Предисловие	7
Введение	10
Глава 1. Передача информации	14
1.1. Модель системы связи	14
1.2. Шенноновские меры информации	16
1.2.1. Собственная информация и энтропия	16
1.2.2. Взаимная информация	23
Глава 2. Источники информации	28
2.1. Статистические источники	28
2.2. Энтропия стационарного источника	30
2.3. Предельные условная и средняя энтропии	33
2.4. Источники без памяти и источники с памятью	34
Глава 3. Кодирование источника	38
3.1. Классификация методов кодирования	38
3.2. Разделимые и неразделимые коды	39
3.3. Префиксный код	43
3.4. Теоремы Шеннона для источника	49
3.5. Коды Шеннона и Фано	53
3.6. Код Хаффмена	55

Глава 4. Кодирование блоков	65
4.1. Кодирование источника без памяти	65
4.2. Кодирование источников с памятью	66
4.3. Код Танстолла	66
4.4. Универсальное кодирование	70
4.4.1. Алгоритм Лемпела–Зива–Велча	70
4.4.2. Алгоритм Лемпела–Зива	74
Глава 5. Типические последовательности	78
5.1. Эпсилон-типические последовательности	78
5.2. Вероятность ε -типической последовательности	79
5.3. Вероятность множества ε -типических последовательностей	80
5.4. Число ε -типических последовательностей	82
5.5. Вероятность множества $M_\varepsilon(P_X)$ при других распределениях	83
5.6. Двумерные типические последовательности	84
Глава 6. Кодирование с потерями	89
6.1. Неоднозначность декодирования	89
6.2. Прямая теорема Шеннона	91
6.3. Обратная теорема Шеннона	91
Глава 7. Кодирование для канала	94
7.1. Лемма об обработке информации	94
7.2. Лемма Фано	97
7.3. Классификация дискретных каналов	99
7.3.1. Общий случай	99
7.3.2. Каналы с нулевой пропускной способностью	104
7.3.3. Каналы, симметричные по входу	105
7.3.4. Каналы, симметричные по выходу	108
7.3.5. Каналы, симметричные по входу и выходу	111

Глава 8. Передача по каналу и декодирование	113
8.1. Основные характеристики передачи	113
8.2. Теоремы Шеннона для канала	121
8.2.1. Обратная теорема Шеннона	122
8.2.2. Прямая теорема Шеннона	126
8.2.3. Первое доказательство	126
8.2.4. Второе доказательство	130
Глава 9. Граница Бхаттачария	132
9.1. Код из двух векторов	132
9.2. Случайное кодирование	136
9.3. Код из $M = 2^{RL}$ векторов	138
Глава 10. Граница Галлагера	141
10.1. Вывод основного неравенства	141
10.2. Функция Галлагера и экспонента вероятности ошибки	144
Глава 11. Непрерывные источники и непрерывные каналы	151
11.1. Непрерывные сообщения и их информационные характеристики	151
11.2. Гауссовские случайные величины и их информационные характеристики	153
11.3. Непрерывный канал без памяти с дискретным временем	155
11.4. Канал с аддитивным гауссовым шумом	156
11.5. Система параллельных каналов с аддитивным гауссовым шумом	160
11.6. Каналы с непрерывным временем и аддитивным белым гауссовским шумом	165
11.6.1. Теорема Котельникова	166
11.6.2. Непрерывные каналы с аддитивным белым гауссовским шумом	168
Приложение А. Сведения из теории вероятностей	174

Приложение Б. Некоторые неравенства	179
Б.1. Неравенство логарифма	179
Б.2. Неравенства Гёльдера–Иенсена	179
Приложение В. Задачи и упражнения	184
В.1. Упражнения к главе 2	184
В.2. Задачи и упражнения к главе 3	189
В.3. Задачи и упражнения к главе 4	191
В.4. Задачи и упражнения к главе 5	194
В.5. Задачи и упражнения к главе 6	196
В.6. Задачи и упражнения к главе 8	200
В.7. Задачи к главе 9	203
В.8. Задачи к главе 10	205
В.9. Задачи и упражнения к главе 11	207
Список литературы	213

Предисловие

Данное учебное пособие предназначено для студентов технических вузов, изучающих теорию информации, и может быть полезно инженерам связи, интересующимся шенноновской теорией информации. Здесь представлены с доказательством все основные теоремы Шенна.

Особое место занимает глава, посвященная доказательству верхних границ оптимального декодирования, связанных с именами Бхаттачария и Галлагера. В ней показаны современные методы, используемые в этой области науки. Для чтения данного пособия необходимо знание основ теории вероятностей. Некоторые сведения приведены в приложении к основному материалу.

Пособие включает вопросы, рассмотренные в течение одного семестра в курсе теории информации для студентов Московского физико-технического института радиотехнической специальности. Краткость курса по времени ограничила и определила выбор представленных здесь тем. Основные из них – информационные меры по Шенону, кодирование источника, кодирование канала, теорема Котельникова и смежные вопросы. Пособие построено следующим образом.

В главе "Передача информации" приведена типовая обобщенная модель системы связи и рассмотрены основные информационные меры.

В главе "Источники информации" определены статистические источники, приведены основные теоремы для стационарного источника и рассмотрены источники без памяти и источники с памятью.

В главе "Кодирование источника" дана классификация методов кодирования, введены понятия разделимых и неразделимых кодов и связанные с этими понятиями критерии разделимости и теоремы. Особое внимание уделено префиксным кодам.

В главе "Кодирование блоков" рассмотрены вопросы кодирования блоков сообщений для стационарных источников. Приведен код Танстолла, алгоритмы универсального кодирования для источников с неизвестной статистикой.

В главе "Типические последовательности" объяснено понятие эпсилон-типических последовательностей и приведены их характеристики.

В главе "Кодирование с потерями" рассмотрена задача специального вида неоднозначного кодирования, доказаны теоремы Шеннона для этого типа кодирования.

Главы "Кодирование для канала" и "Передача по каналу и декодирование" посвящены основным проблемам помехоустойчивого кодирования для дискретных каналов без памяти. Доказана Лемма об обработке информации, устанавливающая невозможность увеличения количества информации при обработке данных, и Лемма Фано, позволяющая оценить вероятность ошибки через условную энтропию передаваемого и принятого сообщения. Приведена классификация каналов, рассмотрена задача оптимального выбора областей декодирования.

В главах "Граница Бхаттачария" и "Граница Галлагера" приведен вывод верхней границы для экспоненты оптимального декодирования.

В главе "Непрерывные источники и каналы" представлена теорема Котельникова, описаны информационные меры непрерывных источников, даны с доказательством прямая и обратная теоремы Шеннона, описаны характеристики параллельных каналов, выведена формула для пропускной способности канала с белым гауссовым шумом и непрерывным временем.

В Приложении А представлены некоторые сведения из теории информации, которые использованы в ходе изложения основного материала. В Приложении Б даны с доказательствами используемые неравенства. В Приложении В приведены примеры и задачи по всему курсу с указанием глав, к которым они относятся.

Для дополнительного изучения предмета рекомендуем учебник В.Д. Колесника и Г.Ш. Полтырева "Курс теории информа-

ции" [6], а также в переводах с английского книги известных ученых Р. Галлагера "Теория информации и надежная связь" [7] и Р. Фано "Передача информации. Статистическая теория связи" [8]. Эти книги написаны на высоком научном уровне и хорошо и ясно излагают рассматриваемые темы. К сожалению, в последние годы они не переиздавались, так что исследования этих лет в них не отражены.

Относительно недавно, в 2004 году вышел перевод с немецкого учебника для вузов М. Вернера "Основы кодирования" [9]. Рекомендуем эту книгу для желающих продолжить изучение предмета "Теория информации".

Введение

Предмет "Теория информации" посвящен рассмотрению проблем экономного представления сообщений источника информации и передачи этих сообщений по каналам с шумом при заданном уровне достоверности. Используемый аппарат – теория вероятности, математическая статистика, математический анализ. Решаемые задачи – построение статистических моделей источников информации и каналов связи, нахождение потенциальных характеристик кодирования источников с целью сжатия данных и избыточного кодирования каналов с целью защиты информации от помех.

Теория информации возникла на основе результатов статистической теории связи. Годом ее возникновения считается 1948, когда были опубликованы основополагающие работы американского ученого, математика и инженера, Клода Шеннона "Математическая теория связи" и "Связь при наличии шума"[1]. История появления этих работ такова. Во время второй мировой войны в Америке проводились исследования, в частности, по автоматическому управлению огнем и по шифрованию. Первое направление возглавлял известный ученый Норберт Винер – создатель новой отрасли науки – кибернетики. Второе направление вел тогда еще молодой ученый Клод Шеннон. На основании исследований второго направления написаны работы К. Шеннона. В этих работах понятие информации связано только со случайностью события и не зависит ни от каких других качеств типа ценности и прочего.

Великий математик академик А.Н. Колмогоров высоко оценил эти работы, считая, что они содержат "необычайное богатство идей". В 50-е годы А.Н. Колмогоров использовал в различных областях математики понятия теории информации, введенные Шенноном.

Однако еще до 1948 года были опубликованы работы других ученых, которые внесли свой вклад в становление этой области науки. В частности, один из основоположников статистической теории связи Л. Хартли ввел единицу измерения информации в виде десятичного логарифма от отношения апостериорной вероятности события к его априорной вероятности.

За полтора десятилетия до публикации первых работ Шеннона советский ученый В.А. Котельников сформулировал и доказал теорему о потенциальных характеристиках дискретного представления непрерывного сообщения. Он представил данный результат в 1933 году на Всесоюзном съезде по вопросу реконструкции дела связи, что отражено в Материалах съезда. В 1946 году теорема вошла в докторскую диссертацию Котельникова, а в 1956 году в его монографию "Теория потенциальной помехоустойчивости"^[2].

А.Н. Колмогоров писал об этой работе следующее: "Еще в 1933 была сформулирована фундаментальная идея спектральной теории передачи информации при помощи непрерывных сигналов". С момента опубликования теорема Котельникова стала активно использоваться инженерами-связистами при расчете характеристик систем представления сообщений и передачи. Шенноном эта теорема была доказана независимо в его работах 1948 года. В иностранной технической литературе она получила название *теоремы отсчетов*.

С 1948 года начинается бурное развитие теории информации. Оно идет широким фронтом с привлечением большого числа исследователей. Многие исследователи развивали идеи, заложенные К. Шенноном в его первых работах. Некоторые из теорем Шеннона были строго доказаны математиками при достаточно общих предположениях. Весь круг проблем, связанных с основополагающими идеями Шеннона, получил название *шенноновской теории информации*.

В 60-е годы А.Н. Колмогоров [3] применил другой подход к теории информации и создал новую область математики – *алгоритмическую теорию информации*. Здесь центральным понятием является сложность конечного объекта при алгоритмическом

способе его описания. Сложность понимается как минимальный объем описания. Он ввел понятие *эпсилон-энтропии*, получил результаты по вычислению и оценке эпсилон-энтропии и скорости создания сообщений некоторых типов источников, пропускной способности каналов связи.

В область науки, называемую шенноновской теорией информации, большой вклад внесли ученые нашей и других стран, среди них Добрушин Р.Л., Галлагер(R. Gallager), Зигангиров К.Ш., Зяблов В.А, Левенштейн В.И., Мэсси (J. Massey), Пинскер М.С., Сифоров В.И., Хинчин А.Я., Файнстейн А. (A. Feinstein), Фано Р. (R. Fano), Цыбаков Б.С. и многие другие.

На практике теория информации находит применение в следующих областях: сжатие данных; передача по каналам связи и запись информации; теоретическая криптография; стеганография, цифровые водяные знаки; квантовые вычисления.

Понятие "сжатие данных" обычно относится к дискретным сообщениям или непрерывным сообщениям после дискретизации. Это означает экономное представление сообщений с точки зрения компактности занимаемой ими записи в дискретных устройствах хранения или для передачи по дискретному каналу. Особенно это важно для космических аппаратов, когда устройство получения информации из Космоса накапливает данные в течение определенного времени и затем в короткие моменты связи с Землей передает эту информацию. Устройство накопления информации имеет ограниченную емкость, аналогично и канал связи характеризуется ограниченной пропускной способностью. Сжатие данных применяется также в архиваторах, при кодировании изображений, например, в форматах JPEG и во многих других системах, где необходимо сокращать объем данных.

Передача и запись информации широко используются в различных системах связи. Организация канала связи является самой дорогой в материальном отношении частью системы связи. Стоимость канала примерно пропорциональна его пропускной способности. Поэтому задачи экономного представления информации (кодирования источника информации) и экономного введения избыточности для защиты от шумов (помехоустойчивого

кодирования) с учетом свойств данного канала являются весьма актуальными. Их инженерные решения используют достижения современной теории информации. Экономная запись информации требуется в компакт-дисках, DVD и т.п.

Теоретическая криптография представляет собой область научных исследований, посвященных шифрованию сообщений. Ее задача – сделать сообщения доступными толькосанкционированному пользователю и недоступными всем остальным. Первые открытые результаты по теоретической криптографии были получены К. Шенноном и опубликованы в работе "Теория связи в секретных системах" [1]. Основные понятия теории информации были применены для построения "идеальных" секретных систем и исследования их свойств. В последние годы криптографические методы защиты информации стали использоваться очень широко.

Одним из методов криптозащиты является стеганография, в частности, использование цифровых водяных знаков (Digital Watermarking). Небольшая часть общей записи информации выделяется для скрытой, замаскированной информации. Метки этой скрытой информации могут быть "размазаны" по всей записи или сосредоточены в определенных местах. Они так представлены, что не влияют на качество записи. Спектр полученного сигнала близок к спектру широкополосного шума, поэтому в целом получается шумоподобный сигнал. Легальный пользователь знает свой код. Применяя свой алгоритм декодирования, он получает информацию. Одно из применений – борьба с пиратскими дисками.

Квантовые вычисления основаны на теоретических и экспериментальных исследованиях в области физики квантовых явлений с применением современного математического аппарата. Введены основные понятия квантовой теории информации, такие как энтропия, взаимная информация и другие, связанные с квантовыми состояниями.

Г л а в а 1.

Передача информации

1.1. Модель системы связи

На рис. 1.1 приведена упрощенная модель системы связи.

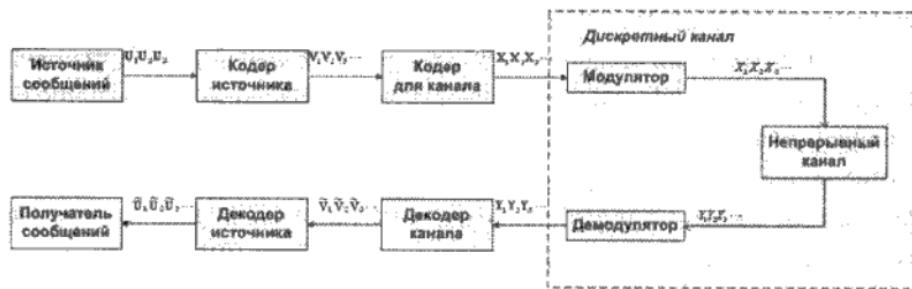


Рис. 1.1. Упрощенная модель системы связи

Сообщения U_1, U_2, U_3, \dots , порождаемые источником сообщений, должны быть переданы по каналу связи получателю сообщений.

Алфавит $\mathcal{U} = \{u_1, u_2, \dots, u_K\}$, где u_i – буквы алфавита, используется для записи сообщений. Количество сообщений, передаваемый по каналу в единицу времени, ограничено. Поэтому надо представить так сообщения источника, чтобы длина сообщения в среднем была по возможности наименьшей.

Кодер источника преобразует эти сообщения, уменьшая избыточность. Последовательности сообщений источника ставится в соответствие последовательность V_1, V_2, V_3, \dots , которая может быть записана в другом алфавите $\mathcal{V} = \{v_1, v_2, \dots, v_D\}$.

Далее буквы-символы этой последовательности поступают на вход кодера для канала, который образует с их помощью

кодовые последовательности $X_1 X_2 X_3 \dots$, способные исправлять некоторые ошибки, возникающие при передаче. Для этого с учетом свойств канала определенным образом вводится избыточность, по возможности, минимальная.

Для передачи по реальным каналам кодовые символы или их группы превращаются с помощью *модулятора* в физические сигналы x_1, x_2, x_3, \dots , которые передаются по *непрерывному каналу связи*. На выходе непрерывного канала получаем сигналы y_1, y_2, y_3, \dots . На приемной стороне системы связи операции проводятся в обратном порядке.

Демодулятор преобразует принятые сигналы y_1, y_2, y_3, \dots в последовательность дискретных символов $Y_1 Y_2 Y_3 \dots$.

Декодер канала восстанавливает входную последовательность $\tilde{V}_1 \tilde{V}_2 \tilde{V}_3 \dots$.

Декодер источника восстанавливает исходное сообщение $\tilde{U}_1, \tilde{U}_2, \tilde{U}_3, \dots$, которое и передается *получателю сообщений*.

Так как при выполнении операций декодерами возможны ошибки, то здесь используется знак "тильда" для обозначения того, что принятые последовательности могут отличаться от переданных последовательностей.

Для теоретического анализа имеет смысл заменить часть системы от входа *модулятора* до выхода *демодулятора* блоком, называемым *дискретным каналом*, входом и выходом которого являются дискретные символы.

Понятие информации в различных областях применения неодинаково. Например,

- в средствах массовой информации – это сведения о событиях и фактах;
- в информатике – это все, что можно записать на каком-либо носителе;
- в шенноновском смысле понятие информации связано с понятием случайности события.

Шенноновская теория информации основана на теории вероятностей. Основные положения шенноновской теории представле-

ны в последующих разделах этой книги.

1.2. Шенноновские меры информации

Перейдем к основным понятиям теории информации. Рассматривается дискретная случайная величина X , которая принимает конечное число значений $\{x_1, x_2, \dots, x_L\}$. Распределение случайной величины задают в виде

$$P_X(x_i) = p_i, \quad p_i \geq 0, \quad i = 1, 2, \dots, L,$$

$$\sum_{i=1}^L p_i = 1.$$

1.2.1. Собственная информация и энтропия

Собственной информацией называется случайная величина

$$I(X) = -\log P_X(X).$$

Единицы измерения информации зависят от основания логарифма: если основание логарифма 2, то бит; если используется натуральный логарифм, то нат; если используется десятичный логарифм, то хартли. Наиболее употребляемая единица измерения информации – бит.

Энтропией, или *неопределенностью* случайной величины X , называется среднее значение (математическое ожидание) собственной информации:

$$\begin{aligned} H(X) &= E(-\log P_X(X)) = -\sum_{i=1}^L P_X(x_i) \log P_X(x_i) = \\ &= -\sum_{i=1}^L p_i \log p_i. \end{aligned} \tag{1.1}$$

Здесь обозначение E означает математическое усреднение.

Пример 1.1. Пусть X принимает только два значения x_1 и x_2 с вероятностями $P_X(x_1) = p$ и $P_X(x_2) = 1 - p$. Тогда неопределенность X равна

$$H(X) = h(p) = -p \log_2 p - (1-p) \log_2(1-p).$$

График этой функции, называемой двоичной энтропией, приведен на рис. 1.2. Функция достигает максимума, равного 1 биту,

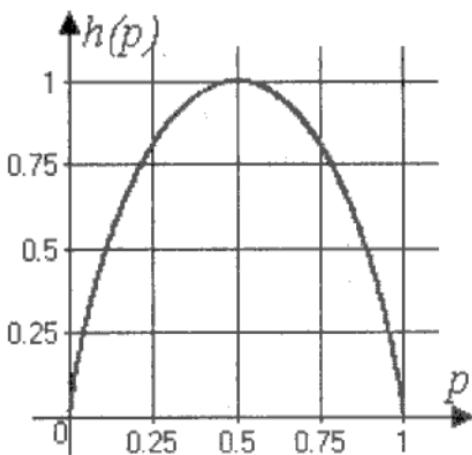


Рис. 1.2. Двоичная энтропия

при равномерном распределении двоичной случайной величины $p = 1 - p = 0.5$.

Теорема 1.1. Энтропия случайной величины X , принимающей L значений, удовлетворяет следующим неравенствам:

$$0 \leq H(X) \leq \log L. \quad (1.2)$$

Левое равенство достигается тогда и только тогда, когда X является детерминированной случайной величиной, т.е., когда $P_X(x_i) = 1$ для некоторого x_i и $P_X(x_j) = 0$ для $j \neq i$. Правое равенство достигается тогда и только тогда, когда X является случайной величиной с равномерным распределением, т.е., когда $P_X(x_i) = 1/L$ для всех i .

Доказательство. В определении энтропии (1.1) все слагаемые $-p_i \log p_i$ неотрицательны, так что левая часть неравенства очевидна. Равенство достигается только в случае, когда каждое слагаемое равно нулю, то есть, когда

$$p_i \log p_i = 0 \quad (1.3)$$

для всех $i = 1, 2, \dots, k$. Это происходит, если $p_i = 1$ для $i = i_0$ и $p_i = 0$ для всех других значений $i \neq i_0$. Для $x = 0$ функция $x \log x$ не определена. Мы доопределяем ее с помощью предельного перехода $\lim_{x \rightarrow +0} x \log x = 0$. Другими словами, энтропия равна нулю для *детерминированной* случайной величины.

Докажем правую часть неравенства (1.2):

$$\begin{aligned} H(X) - \log_2 L &= \sum_{i=1}^L p_i \log_2 \frac{1}{p_i} - \sum_{i=1}^L p_i \log_2 L = \\ &= \sum_{i=1}^L p_i \log_2 \frac{1}{L p_i} \leq \sum_{i=1}^L p_i \left(\frac{1}{L p_i} - 1 \right) \log_2 e = 0. \end{aligned} \quad (1.4)$$

Здесь использовано неравенство логарифма (см. раздел Б.1). Знак равенства в правой части соотношения (1.4), а значит и в (1.2), достигается для случайной величины с равномерным распределением, то есть, когда $p_i = \frac{1}{L}$.

Условная энтропия (или *условная неопределенность*) случайной величины X при условии, что задано значение случайной величины $Y = y_j$, определяется следующим образом:

$$\begin{aligned} H(X | Y = y_j) &= E_{X|Y=y_j}[-\log P_{X|Y=y_j}(X | Y = y_j)] = \\ &= - \sum_{i=1}^L P_{X|Y=y_j}(x_i | y_j) \log P_{X|Y=y_j}(x_i | y_j) = - \sum_{i=1}^L p_{i|j} \log p_{i|j}. \end{aligned}$$

Теорема 1.2. *Верхняя и нижняя границы для условной энтропии дискретной случайной величины X при условии, что задано значение $Y = y_0$ другой случайной величины, определяются следующими неравенствами:*

$$0 \leq H(X | Y = y_0) \leq \log_2 L.$$

Доказательство. Пусть дискретная случайная величина X принимает значения $X = x_i, i = \overline{1, L}$. Условную вероятность X при условии, что задано значение другой случайной величины $Y = y_0$, обозначим $p(X = x_i | Y = y_0)$. Запишем формулу для условной энтропии дискретной случайной величины X при условии, что задано значение $Y = y_0$ другой случайной величины Y :

$$H(X | Y = y_0) = - \sum_{i=1}^L p(X = x_i | Y = y_0) \log_2 p(X = x_i | Y = y_0).$$

Так как все члены суммы неотрицательные величины, то эта величина больше или равна нулю. Равенство нулю достигается, если при некотором $i = i_0$ условная вероятность равна единице, т. е., $p(X = x_{i_0} | Y = y_0) = 1$. Для других $i \neq i_0$ условная вероятность равна нулю, т. е., $p(X = x_i | Y = y_0) = 0$. Таким образом, нижняя граница такова: $H(X | Y = y_0) \geq 0$. Максимум функции $-\sum_{i=1}^L p(X = x_i | Y = y_0) \log_2 p(X = x_i | Y = y_0)$, если все слагаемые равны, то есть, если $p(X = x_i | Y = y_0) = 1/L$. В этом случае значение функции равно $\log_2 L$. Таким образом, для условной энтропии $H(X | Y = y_0)$ имеем следующие границы:

$$0 \leq H(X | Y = y_0) \leq \log_2 L.$$

Условная энтропия (или *условная неопределенность*) случайной величины X при условии, что задана *случайная величина* Y , определяется в виде

$$\begin{aligned} H(X | Y) &= E_{XY}[-\log P_{X|Y}(X | Y)] = \\ &= - \sum_{i=1}^L \sum_{j=1}^M P_{XY}(x_i, y_j) \log P_{X|Y}(x_i | y_j) = \\ &= - \sum_{i=1}^L \sum_{j=1}^M p_{ij} \log p_{i|j} = \\ &= \sum_{j=1}^M q_j H(X | Y = y_j). \end{aligned}$$

Пусть две случайные величины X и \tilde{X} имеют одно и то же множество значений и описываются распределениями $P_X(x)$ и $P_{\tilde{X}}(x)$.

Информационная дивергенция между распределениями P_X и $P_{\tilde{X}}$ (или *расстояние Кульбака–Лейблера* между распределениями P_X и $P_{\tilde{X}}$) определяется как математическое усреднение по распределению P_X от логарифма отношения этих распределений

$$D(P_X \| P_{\tilde{X}}) = E \left[\log \frac{P_X(X)}{P_{\tilde{X}}(X)} \right] = \sum_{i=1}^L P_X(x_i) \log \frac{P_X(x_i)}{P_{\tilde{X}}(x_i)}. \quad (1.5)$$

Теорема 1.3. *Информационная дивергенция не отрицательна:*

$$D(P_X \| P_{\tilde{X}}) \geq 0,$$

причем равенство нулю достигается тогда и только тогда, когда $P_X(x) = P_{\tilde{X}}(x)$ для всех x .

Доказательство. В соответствии с определением используем формулу для информационной дивергенции и неравенство логарифма:

$$\begin{aligned} -D(P_X \| P_{\tilde{X}}) &= - \sum_{i=1}^L P_X(x_i) \log \frac{P_X(x_i)}{P_{\tilde{X}}(x_i)} = \\ &= \sum_{i=1}^L P_X(x_i) \log \frac{P_{\tilde{X}}(x_i)}{P_X(x_i)} \leq \sum_{i=1}^L P_X(x_i) \left(\frac{P_{\tilde{X}}(x_i)}{P_X(x_i)} - 1 \right) \log_2 e = 0. \end{aligned} \quad (1.6)$$

Если распределения совпадают, т. е., $P_X(x) = P_{\tilde{X}}(x)$ для всех x , то в формуле для дивергенции под знаком логарифма стоит единица, следовательно, дивергенция равна нулю.

Теорема 1.4. *Для любых двух дискретных случайных величин X и Y условная энтропия не превосходит безусловной:*

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

причем равенство достигается тогда и только тогда, когда случайные величины X и Y независимы.

Доказательство. Рассмотрим разность безусловной и условной энтропии:

$$\begin{aligned} H(X) - H(X | Y) &= - \sum_{i=1}^k \sum_{j=1}^M p_{ij} \log_2 p_i + \\ &+ \sum_{i=1}^k \sum_{j=1}^M p_{ij} \log_2 p_{i|j} = \sum_{i=1}^k \sum_{j=1}^M p_{ij} \log_2 p_{i|j}/p_i. \end{aligned}$$

Умножим числитель и знаменатель под знаком логарифма на q_j : $p_{i|j}/p_i = q_j p_{i|j}/p_i q_j$. Получаем формулу для разности безусловной и условной энтропии в виде

$$H(X) - H(X | Y) = \sum_{i=1}^k \sum_{j=1}^M p_{ij} \log_2 \frac{p_{ij}}{p_i q_j}.$$

Правая часть этого соотношения определяет расстояние Кульбака–Лейблера (информационную дивергенцию) между двумерными распределениями p_{ij} и $p_i q_j$ и по доказанному выше является неотрицательной. Равенство нулю достигается, если распределения совпадают. В данном случае это означает независимость величин X и Y .

Следствие 1.1. Для трех дискретных случайных величин X , Y и Z верно неравенство

$$H(X | Y, Z = z) \leq H(X | Z = z),$$

причем равенство достигается тогда и только тогда, когда

$$P_{XY|Z}(x, y | z) = P_{X|Z}(x | z) P_{Y|Z}(y | z) \text{ для всех } x \text{ и } y.$$

Следствие 1.2.

$$H(X | YZ) \leq H(X | Z),$$

причем равенство достигается тогда и только тогда, когда X и Y условно независимы при известной величине Z .

Следствие 1.3.

Верхняя и нижняя границы для условной энтропии дискретной случайной величины X при условии, что задана другая случайная величина Y , определяются следующими неравенствами:

$$0 \leq H(X | Y) \leq H(X) \leq \log_2 L.$$

Следствие 1.4.

$$0 \leq H(X | Y) \leq \log L,$$

причем левое равенство достигается тогда и только тогда, когда для каждого y_j существует x_i такое, что $P_{X|Y}(x_i | y_j) = 1$. Другими словами: случайная величина X есть детерминированная функция от случайной величины Y . Правое равенство достигается тогда и только тогда, когда для каждого y_j условная вероятность равна $P_{X|y_j}(x | y_j) = 1/L$ для всех x .

Условная энтропия (или условная неопределенность) случайной величины X при условии, что заданы случайная величина Y и значение случайной величины $Z = z$, определяется как

$$H(X | Y, Z = z) = E_{XYZ=z}[-\log P_{X|YZ=z}(X | YZ = z)] =$$

$$= - \sum_{i=1}^L \sum_{j=1}^M P_{XY|Z=z}(x_i, y_j | z) \log P_{X|YZ=z}(x_i | y_j z).$$

Пусть случайная величина Z принимает N значений. Условная энтропия (или условная неопределенность) случайной величины X при условии, что заданы случайная величина Y и случайная величина Z , определяется формулой

$$H(X \mid Y, Z) = \sum_{k=1}^N P_Z(z_k) H(X \mid Y, z_k).$$

Многомерная энтропия $H(X_1 X_2 \dots X_n)$ определена в виде статистического среднего по многомерному распределению $P_{X_1 \dots X_n}(x_1 \dots x_n)$ от функции $\log_2 \frac{1}{P_{X_1 X_2 \dots X_n}(x_1 x_2 \dots x_n)}$:

$$H(X_1 X_2 \dots X_n) = E_{X_1 X_2 \dots X_n} \log_2 \frac{1}{P_{X_1 X_2 \dots X_n}(x_1 x_2 \dots x_n)}. \quad (1.7)$$

Теорема 1.5. *Многомерная энтропия может быть представлена в виде цепного равенства*

$$H(X_1 X_2 \dots X_n) = H(X_1) + H(X_2 \mid X_1) + \dots + H(X_n \mid X_1 \dots X_{n-1}). \quad (1.8)$$

Доказательство. В формуле для многомерной энтропии (1.7) выразим многомерное распределение через произведение распределений:

$$P_{X_1 X_2 \dots X_n}(x_1 x_2 \dots x_n) = \\ = P_{X_1}(x_1) P_{X_2 \mid X_1}(x_2 \mid x_1) \dots P_{X_n \mid X_1 X_2 \dots X_{n-1}}(x_n \mid x_1 x_2 \dots x_{n-1}) \quad (1.9)$$

и учтем представление логарифма от произведений в виде суммы логарифмов сомножителей. Получим правую часть цепного равенства.

1.2.2. Взаимная информация

Взаимная информация между случайными величинами X и Y определяется как

$$I(X; Y) = H(X) - H(X \mid Y).$$

Можно дать такую эвристическую интерпретацию этого определения:

$H(X)$ – энтропия случайной величины X , характеризующая степень ее "неопределенности"; $H(X | Y)$ – степень неопределенности той же случайной величины при условии, что известна другая случайная величина Y . Разность этих энтропий и есть количество информации, содержащееся в величине Y относительно величины X .

Теорема 1.6. Взаимная информация является симметричной функцией случайных величин X и Y :

$$I(X; Y) = I(Y; X).$$

Доказательство. Исходя из определения, приведенное соотношение эквивалентно следующему равенству:

$$H(X) - H(X | Y) = H(Y) - H(Y | X).$$

Энтропия $H(X)$ и условная энтропия $H(X | Y)$ имеют следующий вид соответственно определениям:

$$H(X) = E_X \left(\log \frac{1}{P_X(X)} \right) = E_{XY} \left(\log \frac{1}{P_X(X)} \right)$$

и

$$H(X | Y) = E_{XY} \left(\log \frac{1}{P_X(X)} \right) = E_{XY} \left(\log \frac{1}{P_{X|Y}(X | Y)} \right),$$

где E_X, E_{XY} – это обозначение статистического усреднения по распределениям случайных величин, указанных в индексе.

Учитывая правые части этих соотношений, запишем разность этих энтропий и используем формулы для совместной вероятности двух случайных величин:

$$\begin{aligned} H(X) - H(X | Y) &= E_{XY} \left(\log \frac{P_{X|Y}(X | Y) P_Y(Y)}{P_X(X) P_Y(Y)} \right) = \\ &= E_{XY} \left(\log \frac{P_X(X) P_{Y|X}(X | Y)}{P_X(X) P_Y(Y)} \right) = E_{XY} \left(\log \frac{P_{Y|X}(Y | X)}{P_Y(Y)} \right) = \\ &= H(Y) - H(Y | X). \end{aligned}$$

Условная взаимная информация при условии, что задано значение случайной величины $Z = z$, есть

$$I(X;Y \mid Z = z) = H(X \mid Z = z) - H(X \mid Y, Z = z).$$

Условная взаимная информация при условии, что задана случайная величина Z , есть

$$I(X;Y|Z) = H(X \mid Z) - H(X \mid YZ).$$

Условная взаимная информация и безусловная взаимная информация являются симметричными функциями:

$$I(X;Y \mid Z = z) = I(Y;X \mid Z = z),$$

$$I(X;Y \mid Z) = I(Y;X \mid Z).$$

Доказательство. Доказательство аналогично предыдущему;

$$\begin{aligned} I(X;Y \mid Z = z) &= H(X \mid Z = z) - H(X \mid Y, Z = z) = \\ &= E_{XY|Z=z} \log \frac{P_{X|Y,Z=z}(x|y,Z=z)}{P_{X|Z=z}(x|Z=z)} \frac{P_{Y|Z=z}(y|Z=z)}{P_{Y|Z=z}(y|Z=z)} = \\ &= E_{XY|Z=z} \log \frac{P_{XY|Z=z}(xy|Z=z)}{P_{X|Z=z}(x|Z=z)P_{Y|Z=z}(y|Z=z)} = \\ &= E_{XY|Z=z} \log \frac{P_{X|Z=z}(x|Z=z)P_{Y|X,Z=z}(y|x,Z=z)}{P_{X|Z=z}(x|Z=z)P_{Y|Z=z}(y|Z=z)} = \\ &= H(Y \mid Z = z) - H(Y \mid X, Z = z) \equiv I(Y;X \mid Z = z). \end{aligned}$$

Теорема 1.7. Взаимная информация подчиняется следующим ограничениям:

$$0 \leq I(X;Y) \leq \min[H(X), H(Y)],$$

причем левое равенство достигается тогда и только тогда, когда случайные величины X и Y независимы, а правое – если одна из случайных величин является детерминированной функцией другой.

Доказательство. Используя определение и свойство симметрии взаимной информации, запишем соотношение

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X). \quad (1.10)$$

Для независимых X и Y энтропия и условная энтропия равны: $H(X) = H(X|Y)$ и $H(Y) = H(Y|X)$. В других случаях энтропия больше условной энтропии. Значит, нижняя граница в виде нуля справедлива. Из соотношения (1.10) видно, что максимальное значение взаимной информации достигается, если условная энтропия равна нулю. Это происходит, если случайная величина является детерминированной функцией другой случайной величины, указанной в условии. Тогда максимальное значение взаимной информации равно $H(X)$ или $H(Y)$. Поэтому в качестве верхней границы следует использовать минимальное из этих двух значений, что и указано в формулировке теоремы. Аналогично,

Теорема 1.8. Условная взаимная информация удовлетворяет неравенствам

$$0 \leq I(X;Y|Z=z) \leq \min[H(X|Z=z), H(Y|Z=z)],$$

$$0 \leq I(X;Y|Z) \leq \min[H(X|Z), H(Y|Z)].$$

Доказательство. Здесь приведем доказательство первого из этих соотношений. Доказательство второго соотношения аналогично. Оно приведено в Приложении 3 в качестве упражнения.

Запишем, используя определение и свойство симметрии:

$$\begin{aligned} I(X;Y|Z=z) &= H(X|Z=z) - H(X|Y,Z=z) = \\ &= H(Y|Z=z) - H(Y|X,Z=z). \end{aligned} \quad (1.11)$$

Как было показано ранее, добавление случайной величины в условие не увеличивает энтропию, но может ее уменьшить. Поэтому знак неравенства относительно нуля в приведенном соотношении поставлен правильно. Равенство нулю условной взаимной информации соответствует равенству условных энтропий:

$$\begin{aligned} H(X|Z=z) &= H(X|Y,Z=z) \text{ или} \\ H(Y|Z=z) &= H(Y|X,Z=z). \end{aligned}$$

Это выполняется, если равны следующие условные вероятности: $P_{X|Z=z} = P_{X|Y,Z=z}$ или

$$P_{Y|Z=z}(y | Z = z) = P_{Y|X,Z=z}(y | x, Z = z),$$

т. е., случайные величины X и Y при известном значении Z условно независимы. Таким свойством обладают случайные величины X, Z, Y , а также Y, Z, X , образующие марковскую цепь первого порядка. В этом случае достигается нижняя граница в виде нуля.

Максимальное значение условной взаимной информации достигается, если в соотношении (1.11) условная энтропия при заданной другой случайной величине и значению третьей случайной величины равна нулю. Это происходит, если случайная величина является детерминированной функцией этой другой случайной величины при заданном значении третьей случайной величины. Тогда максимум взаимной информации $I(X; Y | Z)$ равен $H(X | Z = z)$ или $H(Y | Z = z)$.

Поэтому в качестве верхней границы следует использовать минимальное из этих двух значений, что и указано в формулировке теоремы.

Г л а в а 2.

Источники информации

2.1. Статистические источники

Здесь рассматриваются некоторые простые источники сообщений,

Источник сообщений (см. рис. 1.1) – это устройство, порождающее полубесконечную последовательность символов (букв) $U_1 U_2 U_3 \dots$. Предполагается, что алфавит $\mathcal{U} = \{u_1, u_2, \dots, u_K\}$ источника состоит из K букв.

В теории рассматриваются *статистические* источники сообщений. Это означает, что задано совместное распределение случайных величин $U_1 U_2 U_3 \dots$, или можно задать начальное распределение $P_{U_1}(U_1 = a_1)$ первого символа и все условные распределения

$$P_{U_L|U_1 U_2 \dots U_{L-1}}(U_L = a_L | U_1 = a_1, U_2 = a_2, \dots, U_{L-1} = a_{L-1}),$$

для всех $a_j \in \mathcal{U}$ и $L \geq 2$.

Источник называется *стационарным*, если для любых целых чисел $n \geq 1, L \geq 1$ случайные векторы

$$(U_1, U_2, \dots, U_L) \text{ и } (U_{n+1}, U_{n+2}, \dots, U_{n+L})$$

имеют одинаковые распределения.

Источник называется источником *без памяти*, если случайные величины $U_1 U_2 U_3 \dots$ независимы в совокупности. Чтобы описать источник без памяти, необходимо задать все одномерные распределения $P_{U_j}(U_j = a_j), j = 1, 2, \dots$. В этом случае для любого целого L совместное распределение первых L случайных величин задается формулой

$$P_{U_1 U_2 \dots U_L}(U_1 = a_1, U_2 = a_2, \dots, U_L = a_L) = \\ = P_{U_1}(U_1 = a_1) \cdot P_{U_2}(U_2 = a_2) \cdots P_{U_L}(U_L = a_L). \quad (2.1)$$

Распределения величин U_1 и U_2 называются одинаковыми, если для всех $a \in \mathcal{U}$ $P_{U_1}(U_1 = a) = P_{U_2}(U_2 = a)$. Это определение естественным образом обобщается на многомерные распределения.

Если все одномерные распределения одинаковы, то источники без памяти называют также *бернульевыми* источниками.

Источник называется *марковским* порядка m , если для любого $L > m$ и любых $a_1, a_2, \dots, a_L \in \mathcal{U}$ имеет место равенство

$$P_{U_L|U_1 \dots U_{L-1}}(U_L = a_L | U_1 = a_1, U_2 = a_2, \dots, U_{L-1} = a_{L-1}) = \\ = P_{U_L|U_{L-m} \dots U_{L-1}}(U_L = a_L | U_{L-m} = a_{L-m}, \dots, U_{L-1} = a_{L-1}). \quad (2.2)$$

Другими словами, условное распределение величины U_L зависит не от всей предыстории, а только от предыдущих m значений.

Совместное распределение при $L > m$ задается формулой

$$P_{U_1 U_2 \dots U_L}(U_1 = a_1, U_2 = a_2, \dots, U_L = a_L) = \\ = P_{U_1 U_2 \dots U_m}(U_1 = a_1, U_2 = a_2, \dots, U_m = a_m) \times \\ \times P_{U_{m+1}|U_1 \dots U_m}(U_{m+1} = a_{m+1} | U_1 = a_1, U_2 = a_2, \dots, U_m = a_m) \times \\ \cdots \times P_{U_L|U_{L-m} \dots U_{L-1}}(U_L = a_L | U_{L-m} = a_{L-m}, \dots, U_{L-1} = a_{L-1}). \quad (2.3)$$

Вероятности

$$P_{U_L|U_{L-m} \dots U_{L-1}}(U_L = a_L | U_{L-m} = a_{L-m}, \dots, U_{L-1} = a_{L-1})$$

называются переходными вероятностями марковской цепи. Они описывают вероятности перехода блока $\{a_{L-m}, \dots, a_{L-1}\}$, порожденного источником в предыдущие моменты времени $L-m, L-m+1, \dots, L-1$, в символ a_L в момент L . Часто эти вероятности располагают в матрицу переходных вероятностей размеров $K^m \times K$. Число строк равно числу возможных предыдущих блоков символов, а число столбцов – числу возможных символов в текущий момент. Каждая строка соответствует некоторому фиксированному блоку, поэтому сумма элементов каждой строки равна 1.

В общем случае матрица переходных вероятностей меняется от одного момента к другому. Для стационарных марковских источников матрица переходных вероятностей одинакова для всех моментов времени.

Источник без памяти является частным случаем марковского источника при $m = 0$. В этом случае источник порождает последовательность независимых случайных величин.

2.2. Энтропия стационарного источника

Далее рассматриваются стационарные источники. Основной характеристикой источника является его *энтропия*. Для определения энтропии существует два подхода.

Условной L-энтропией на букву, $L = 1, 2, \dots$, называется условная энтропия случайной величины U_L при заданных случайных величинах $U_1 \dots U_{L-1}$:

$$H_C(L) = H(U_L | U_1 \dots U_{L-1}). \quad (2.4)$$

Условной энтропией источника на букву называется предел (если он существует)

$$R_1 = \lim_{L \rightarrow \infty} H_C(L) = \lim_{L \rightarrow \infty} H(U_L | U_1 \dots U_{L-1}). \quad (2.5)$$

Средней L-энтропией на букву, $L = 1, 2, \dots$, называется относенная к одному символу энтропия случайных величин $U_1 U_2 \dots U_L$

$$H_L = \frac{1}{L} H(U_1 U_2 \dots U_L). \quad (2.6)$$

Средней энтропией источника на букву называется предел (если он существует)

$$R_2 = \lim_{L \rightarrow \infty} H_L = \lim_{L \rightarrow \infty} \frac{1}{L} H(U_1 U_2 \dots U_L). \quad (2.7)$$

Ниже будет показано, что оба предела существуют и равны. Предел называется *энтропией стационарного источника* и обозначается H_∞ . Другое название – *скорость создания информации*.

Лемма 2.1. Условная L -энтропия на букву является монотонно невозрастающей функцией L , ограниченной снизу:

$$0 \leq H_C(L+1) \leq H_C(L), L = 1, 2, \dots, \quad (2.8)$$

так что существует предел

$$R_1 = \lim_{L \rightarrow \infty} H_C(L) = \lim_{L \rightarrow \infty} H(U_L | U_1 \dots U_{L-1}).$$

Доказательство. Левое неравенство в (2.8) – общее свойство любой энтропии. Следующая цепочка неравенств доказывает монотонное невозрастание условной L -энтропии:

$$\begin{aligned} H_C(L+1) &= H(U_{L+1} | U_1 U_2 \dots U_L) \leq \\ &\leq H(U_{L+1} | U_2 \dots U_L) = \\ &= H(U_L | U_1 U_2 \dots U_{L-1}) = H_C(L). \end{aligned} \quad (2.9)$$

Неравенство в (2.9) верно, так удаление случайной величины U_1 из условия может лишь увеличить энтропию. Последующее равенство верно потому, что сдвиг на единицу индексов случайных величин, фигурирующих в энтропии, не меняет ее величины в силу стационарности источника. Из доказанных неравенств следует существование предела R_1 .

Лемма 2.2. Условная L -энтропия на букву не превосходит среднюю L -энтропию на букву:

$$H_C(L) \leq H_L, L = 1, 2, \dots \quad (2.10)$$

Доказательство. Цепное равенство для набора случайных величин $U_1 U_2 \dots U_L$ имеет вид

$$H(U_1U_2\dots U_L) = H(U_1) + H(U_2 \mid U_1) + \dots + H(U_L \mid U_1U_2\dots U_{L-1}). \quad (2.11)$$

Для стационарного источника каждое слагаемое в правой части этого равенства не превосходит предыдущего. Например, всегда $H(U_2 \mid U_1) \leq H(U_2)$ по свойству условной энтропии. Но $H(U_2) = H(U_1)$, так в стационарном источнике случайные величины U_1 и U_2 имеют одинаковые распределения и энтропии. Следовательно, $H(U_2 \mid U_1) \leq H(U_1)$. Заменяя в правой части (2.11) все слагаемые на наименьшее последнее, приходим к неравенству

$$LH(U_L \mid U_1U_2\dots U_{L-1}) \leq H(U_1U_2\dots U_L),$$

или

$$H_C(L) = H(U_L \mid U_1U_2\dots U_{L-1}) \leq \frac{1}{L}H(U_1U_2\dots U_L) = H_L.$$

Лемма 2.3. Средняя L -энтропия на букву является монотонно невозрастающей функцией L , ограниченной снизу:

$$0 \leq H_{L+1} \leq H_L, L = 1, 2, \dots, \quad (2.12)$$

так что существует предел

$$R_2 = \lim_{L \rightarrow \infty} H_L = \lim_{L \rightarrow \infty} \frac{1}{L}H(U_1U_2\dots U_L).$$

Доказательство. Левое неравенство (2.12) – общее свойство любой энтропии. Далее, цепное равенство для случайных величин $U_1U_2\dots U_{L+1}$ можно записать в виде

$$H(U_1U_2\dots U_{L+1}) = H(U_1U_2\dots U_L) + H(U_{L+1} \mid U_1U_2\dots U_L). \quad (2.13)$$

Так как

$$\begin{aligned} H(U_{L+1} | U_1 U_2 \dots U_L) &\leq H(U_{L+1} | U_2 \dots U_L) = \\ &= H(U_L | U_1 U_2 \dots U_{L-1}) \leq \frac{1}{L} H(U_1 U_2 \dots U_L) \end{aligned}$$

в силу стационарности источника и леммы 2.2, то из соотношения (2.13) получаем неравенство

$$H(U_1 U_2 \dots U_{L+1}) \leq H(U_1 U_2 \dots U_L) + \frac{1}{L} H(U_1 U_2 \dots U_L),$$

или

$$H_{L+1} = \frac{1}{L+1} H(U_1 U_2 \dots U_{L+1}) \leq \frac{1}{L} H(U_1 U_2 \dots U_L) = H_L.$$

Из доказанных неравенств следует существование предела R_2 . Из лемм 2.2 и 2.3 следует неравенство между предельными условной энтропией на букву и средней энтропией на букву:

$$R_1 \leq R_2. \quad (2.14)$$

2.3. Предельные условная и средняя энтропии

В следующей теореме будет доказано, что на самом деле предельные энтропии равны.

Теорема 2.1. *Предельные условная и средняя энтропии равны:*

$$R_1 = R_2 = H_\infty. \quad (2.15)$$

Доказательство. Достаточно доказать, что $R_1 \geq R_2$. Выберем целые числа L и n , запишем определение функции H_{L+n} и применим цепное равенство:

$$\begin{aligned} H_{L+n} &= \frac{1}{L+n} H(U_1 \dots U_L U_{L+1} \dots U_{L+n}) = \\ &= \frac{1}{L+n} (H(U_1 \dots U_L) + H(U_{L+1} \dots U_{L+n} | U_1 \dots U_L)). \end{aligned} \quad (2.16)$$

В свою очередь используем цепное равенство для второго слагаемого:

$$\begin{aligned}
& H(U_{L+1} \dots U_{L+n} \mid U_1 \dots U_L) = \\
& = H(U_{L+1} \mid U_1 \dots U_L) + H(U_{L+2} \mid U_1 \dots U_{L+1}) + \dots \\
& \quad \dots + H(U_{L+n} \mid U_1 \dots U_{L+n-1}) = \\
& = H_C(L+1) + H_C(L+2) + \dots + H_C(L+n).
\end{aligned} \tag{2.17}$$

Из леммы 2.1 следует, что каждое слагаемое в правой части (2.17) больше последующего. Заменяя их наибольшим первым слагаемым, приходим к неравенству

$$H(U_{L+1} \dots U_{L+n} \mid U_1 \dots U_L) \leq nH_C(L+1). \tag{2.18}$$

Подставляя это неравенство в соотношение (2.16), получим

$$H_{L+n} \leq \frac{1}{L+n} H(U_1 \dots U_L) + \frac{n}{L+n} H_C(L+1). \tag{2.19}$$

Зафиксируем целое L и перейдем к пределу при $n \rightarrow \infty$ в (2.19). Левая часть дает в пределе R_2 , а в правой часть предел равен $H_C(L+1)$, так что $R_2 \leq H_C(L+1)$. Перейдя затем к пределу при $L \rightarrow \infty$, получаем, что $R_2 \leq R_1$. Вместе с соотношением (2.14) это дает равенство $R_2 = R_1 = H_\infty$.

2.4. Источники без памяти и источники с памятью

Для источников без памяти имеем

$$\begin{aligned}
H(U_1 U_2 \dots U_L) &= H(U_1) + H(U_2) + \dots + H(U_L) = LH(U); \\
H_L &= \frac{1}{L} H(U_1 U_2 \dots U_L) = H(U); \\
H_\infty &= H(U).
\end{aligned} \tag{2.20}$$

Таким образом, энтропия стационарного источника без памяти совпадает с энтропией одномерной случайной величины (буквы) U .

По определению, условная L -энтропия на букву равна

$$H_C(L) = E_{U_1 U_2 \dots U_L} (-\log_2 P_{U_L | U_1 \dots U_{L-1}}(U_L | U_1, U_2, \dots, U_{L-1})). \quad (2.21)$$

Для марковских источников порядка m с учетом соотношения (2.2) это выражение можно записать как

$$H_C(L) = E_{U_1 U_2 \dots U_L} (-\log_2 P_{U_L | U_1 \dots U_{L-1}}(U_L | U_1, U_2, \dots, U_{L-1})) \\ \text{для } L \leq m;$$

$$H_C(L) = E_{U_{L-m} \dots U_L} (-\log_2 P_{U_L | U_{L-m} \dots U_{L-1}}(U_L | U_{L-m} \dots U_{L-1})) \\ \text{для } L > m. \quad (2.22)$$

С учетом стационарности источника, вторая часть в (2.22) приобретает вид

$$H_C(L) = E_{U_1 U_2 \dots U_{m+1}} (-\log_2 P_{U_{m+1} | U_1 \dots U_m}(U_{m+1} | U_1, U_2, \dots, U_m)), \quad (2.23)$$

т.е. правая часть не зависит от L . Следовательно, энтропия марковского источника равна условной $(m+1)$ -энтропии:

$$H_\infty = H_C(m+1) = \\ = E_{U_1 U_2 \dots U_{m+1}} (-\log_2 P_{U_{m+1} | U_1 \dots U_m}(U_{m+1} | U_1, U_2, \dots, U_m)). \quad (2.24)$$

Для ее вычисления необходимо знать совместное распределение первых $m+1$ случайных величин $U_1 U_2 \dots U_{m+1}$, причем одномерные распределения всех величин должны быть одинаковы, распределения смежных пар случайных величин должны быть одинаковы и т.д.

Эквивалентно стационарный марковский источник полностью определяется матрицей переходных вероятностей

$$P_{U_{m+1} | U_1 \dots U_m}(U_{m+1} = a_{m+1} | U_1 = a_1, U_2 = a_2, \dots, U_m = a_m), \\ a_j \in \mathcal{U}, j = 1, 2, \dots, m+1, \quad (2.25)$$

одинаковой для всех моментов времени $L > m$, совместно с распределением m случайных величин $U_1 U_2 \dots U_m$, причем одномерные распределения всех величин $U_1 U_2 \dots U_{m+1}$ должны быть одинаковы, распределения смежных пар случайных величин должны быть одинаковы и т.д.

Пример 2.1. Рассмотрим стационарный двоичный марковский источник $U_1 U_2 U_3 \dots$ порядка $m = 1$. Он полностью определяется совместным распределением первых двух величин U_1, U_2 . Выясним требования к этому распределению. Пусть

$$\begin{aligned} P_{U_1 U_2}(U_1 = 0, U_2 = 0) &= q_0; & P_{U_1 U_2}(U_1 = 0, U_2 = 1) &= q_1; \\ P_{U_1 U_2}(U_1 = 1, U_2 = 0) &= q_2; & P_{U_1 U_2}(U_1 = 1, U_2 = 1) &= q_3; \\ q_0 + q_1 + q_2 + q_3 &= 1. \end{aligned} \tag{2.26}$$

Чтобы источник был стационарным, необходимо и достаточно, чтобы распределения случайных величин U_1 и U_2 совпадали: $P_{U_1}(U_1 = 0) = P_{U_2}(U_2 = 0)$, $P_{U_1}(U_1 = 1) = P_{U_2}(U_2 = 1)$. Следовательно,

$$\begin{aligned} P_{U_1}(U_1 = 0) &= P_{U_1 U_2}(U_1 = 0, U_2 = 0) + P_{U_1 U_2}(U_1 = 0, U_2 = 1) = q_0 + q_1; \\ P_{U_2}(U_2 = 0) &= P_{U_1 U_2}(U_1 = 0, U_2 = 0) + P_{U_1 U_2}(U_1 = 1, U_2 = 0) = \\ &= q_0 + q_2. \end{aligned} \tag{2.27}$$

Отсюда следует, что

$$q_1 = q_2, \quad P_U(U = 0) = q_0 + q_1, \quad P_U(U = 1) = q_1 + q_3.$$

Энтропия источника равна условной 2-энтропии $H(U_2 | U_1)$, которая может быть найдена из цепного равенства:

$$\begin{aligned} H_\infty &= H(U_2 | U_1) = H(U_1 U_2) - H(U_1) = \\ &= -q_0 \log_2 q_0 - 2q_1 \log_2 q_1 - q_3 \log_2 q_3 + \\ &\quad + (q_0 + q_1) \log_2 (q_0 + q_1) + (q_1 + q_3) \log_2 (q_1 + q_3). \end{aligned}$$

В частности, если $q_0 = q_3$, то $2q_0 + 2q_1 = 1$, так что

$$H_\infty = -(2q_0)p \log_2(2q_0) - (1 - 2q_0) \log_2(1 - 2q_0).$$

Пример 2.2. Рассмотрим стационарный двоичный марковский источник $U_1 U_2 U_3 \dots$ порядка $m = 1$. Он полностью определяется матрицей переходных вероятностей

$$\begin{aligned} P_{U_2|U_1}(U_2 = 0 \mid U_1 = 0) &= 1 - p_1; & P_{U_2|U_1}(U_2 = 1 \mid U_1 = 0) &= p_1; \\ P_{U_2|U_1}(U_2 = 1 \mid U_1 = 1) &= 1 - p_2; & P_{U_2|U_1}(U_2 = 0 \mid U_1 = 1) &= p_2 \end{aligned} \quad (2.28)$$

и распределением величины U_1 . Введем обозначения

$P_{U_1}(U_1 = 0) = \alpha$, $P_{U_1}(U_1 = 1) = 1 - \alpha$. Выясним требования к этому распределению. Вычислим совместное распределение величин U_1, U_2 :

$$\begin{aligned} P_{U_1 U_2}(U_1 = 0, U_2 = 0) &= \alpha(1 - p_1); & P_{U_1 U_2}(U_1 = 0, U_2 = 1) &= \alpha p_1; \\ P_{U_1 U_2}(U_1 = 1, U_2 = 0) &= (1 - \alpha)p_2; & P_{U_1 U_2}(U_1 = 1, U_2 = 1) &= \\ &= (1 - \alpha)(1 - p_2). \end{aligned} \quad (2.29)$$

Найдем распределение величины U_2 :

$$P_{U_2}(U_2 = 0) = \alpha(1 - p_1) + (1 - \alpha)p_2; \quad P_{U_2}(U_2 = 1) = \alpha p_1 + (1 - \alpha)(1 - p_2).$$

Так как оно обязано совпадать с распределением величины U_1 , то получаем

$$P_{U_2}(U_2 = 0) = \alpha(1 - p_1) + (1 - \alpha)p_2 = \alpha = P_{U_1}(U_1 = 0).$$

Следовательно,

$$P_{U_1}(U_1 = 0) = \alpha = \frac{p_2}{p_1 + p_2}; \quad P_{U_1}(U_1 = 1) = \frac{p_1}{p_1 + p_2}.$$

Отсюда следует также, что $\alpha p_1 = (1 - \alpha)p_2$.

Энтропия источника равна условной 2-энтропии $H(U_2 \mid U_1)$, которая может быть найдена из цепного равенства:

$$\begin{aligned} H_\infty &= H(U_2 \mid U_1) = H(U_1 U_2) - H(U_1) = \\ &= -\alpha(1 - p_1) \log_2 \alpha(1 - p_1) - 2\alpha p_1 \log_2 \alpha p_1 - \\ &\quad -(1 - \alpha)(1 - p_2) \log_2 (1 - \alpha)(1 - p_2) + \\ &\quad + \alpha \log_2 \alpha + (1 - \alpha) \log_2 (1 - \alpha). \end{aligned}$$

В частности, если матрица переходных вероятностей симметрична, $p_1 = p_2 = p$, то $\alpha = 0.5$, так что

$$H_\infty = -p \log_2 p - (1 - p) \log_2 (1 - p) = h(p).$$

Г л а в а 3

Кодирование источника

3.1. Классификация методов кодирования

Рассмотрим следующую модель обработки данных. Источник сообщений порождает последовательность символов (букв) $U_1 U_2 U_3 \dots$ из алфавита $\mathcal{U} = \{u_1, u_2, \dots, u_K\}$, содержащего K букв. Эта последовательность поступает на вход кодера источника, который преобразует ее в последовательность выходных символов $V_1 V_2 V_3 \dots$ из алфавита кодера $\mathcal{V} = \{v_1, v_2, \dots, v_D\}$, состоящего из D букв (рис. 3.1).

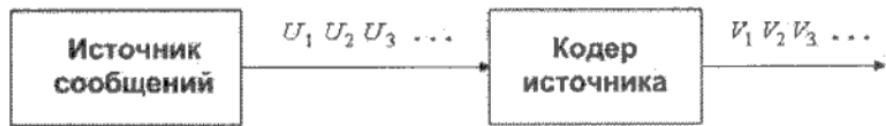


Рис. 3.1. Схема кодирования источника

Целью кодирования источника является уменьшение избыточности. Неформально это означает следующее. Пусть входной блок $U_1 U_2 U_3 \dots U_N$ из N символов преобразуется кодером в выходной блок $V_1 V_2 V_3 \dots V_M$ из M символов. Преобразование должно быть таким, чтобы по выходному блоку можно было однозначно восстановить входной блок. Для описания входного блока потребуется (с точностью до округления) $N \log_2 K$ бит. Для описания выходного блока потребуется (с точностью до округления) $M \log_2 D$ бит. Будем считать случайными величинами как символы входных блоков, так и их длину. Тогда и выходные символы, и длины выходных блоков также будут случайными величинами. Средние длины блоков обозначим \bar{N} и \bar{M} . Кодер источника будет уменьшать избыточность, если вы-

полняется соотношение

$$\overline{M} \log_2 D < \overline{N} \log_2 K.$$

Проблема сжатия данных состоит в выборе такого кодера источника, для которого левая часть этого соотношения является минимально возможной.

Различные методы кодирования можно классифицировать следующим образом.

1. Входные блоки постоянной длины N преобразуются в выходные блоки переменной длины. В этом случае, кроме входных и выходных символов, только длины выходных блоков являются случайными величинами.
2. Входные блоки переменной длины преобразуются в выходные блоки постоянной длины M . В этом случае, кроме входных и выходных символов, только длины входных блоков являются случайными величинами.
3. Входные блоки переменной длины преобразуются в выходные блоки переменной длины.
4. Входные блоки постоянной длины преобразуются в выходные блоки постоянной длины.

3.2. Разделимые и неразделимые коды

Кодирование называется взаимно однозначным, если разным входным сообщениям соответствуют различные кодовые слова.

Код называется кодом с *однозначным декодированием* (или *разделимым*), если по выходному потоку кодера можно однозначно восстановить входные сообщения.

Анализ методов кодирования источника начнем со случая, когда длина входных блоков равна 1 (побуквенное кодирование).

Входные блоки – буквы, порождаемые источником, – преобразуются кодером в кодовые слова. Буквам u_i , $i = 1, 2, \dots, K$, из входного алфавита $\mathcal{U} = \{u_1, u_2, \dots, u_K\}$ ставятся в соответствие

кодовые слова z_i в D -ичном алфавите с (различными) длинами $w_i = w(z_i)$. Последовательности букв источника

$U_1 = u_{i_1}, U_2 = u_{i_2}, \dots, U_L = u_{i_L}$ соответствует последовательность кодовых слов (без пробелов или запятых) $z_{i_1}, z_{i_2}, \dots, z_{i_L}$ с общей длиной $w_{i_1} + w_{i_2} + \dots + w_{i_L}$. Если код разделимый, то любая конечная последовательность кодовых слов однозначно разбивается на отдельные кодовые слова.

Пример 3.1. Код \mathcal{C}_1 , состоящий из слов $z_1 = 0, z_2 = 10, z_3 = 11$, является разделимым. Код \mathcal{C}_2 , состоящий из слов

$$z_1 = 0, z_2 = 10, z_3 = 11, z_4 = 01,$$

является неразделимым, так последовательность 0110110 может быть разбита на кодовые слова двумя способами:

$$0110110 = z_1 z_3 z_4 z_2 = z_4 z_2 z_3 z_1.$$

Рассмотрим вопрос о возможных длинах кодовых слов разделимого кода. Составим сумму:

$$S = \sum_{i=1}^K D^{-w_i} = D^{-w_1} + D^{-w_2} + \dots + D^{-w_K}. \quad (3.1)$$

Теорема 3.1 (Мак-Миллан). *Если существует разделимый код z_1, \dots, z_K с длинами $w_1 = w(z_1), w_2 = w(z_2), \dots, w_K = w(z_K)$, то*

$$S = \sum_{i=1}^K D^{-w_i} = D^{-w_1} + D^{-w_2} + \dots + D^{-w_K} \leq 1.$$

Доказательство. Возведем обе части в степень L , где L – положительное целое:

$$S^L = \sum_{i_1=1}^k \sum_{i_2=1}^k \dots \sum_{i_L=1}^k D^{-w_{i_1}-w_{i_2}-\dots-w_{i_L}}. \quad (3.2)$$

Сумма $w_{i_1} + w_{i_2} + \dots + w_{i_L}$ равна длине $W(\underline{Z})$ некоторой последовательности $\underline{Z} = (z_{i_1}, z_{i_2}, \dots, z_{i_L})$ из L кодовых слов. Для различных наборов из L кодовых слов длина $W(\underline{Z})$ ограничена сверху и снизу:

$$LW_{\min} \leq W(\underline{Z}) \leq LW_{\max}, \quad (3.3)$$

где W_{\min} и W_{\max} – минимальная и максимальная длина кодового слова. Переберем все возможные последовательности из L кодовых слов и упорядочим их по длинам. Обозначим через $A_i(L)$ число последовательностей из L кодовых слов, имеющих длину, равную i . Тогда равенство (3.2) можно переписать в виде

$$S^L = \sum_{i=LW_{\min}}^{LW_{\max}} A_i(L) D^{-i}. \quad (3.4)$$

Оценим число $A_i(L)$. Так как код является *разделимым*, то все последовательности из L кодовых слов с одинаковой длиной i в алфавите из D букв должны быть различными. Общее число различных последовательностей длины i равно D^i . Следовательно, $A_i(L) \leq D^i$. С учетом этого из (3.4) получаем оценку:

$$S^L = \sum_{i=LW_{\min}}^{LW_{\max}} A_i(L) D^{-i} \leq \sum_{i=LW_{\min}}^{LW_{\max}} (D^i D^{-i}) = L(W_{\max} - W_{\min} + 1), \quad (3.5)$$

или

$$S \leq \sqrt[L]{L(W_{\max} - W_{\min} + 1)}. \quad (3.6)$$

Так как это неравенство верно для любого положительного целого L , то, переходя к пределу при $L \rightarrow \infty$, получим

$$S \leq \lim_{L \rightarrow \infty} \sqrt[L]{L(W_{\max} - W_{\min} + 1)} = 1. \quad (3.7)$$

Следствие 3.1. *Если $S > 1$, то разделимый код не существует.*

Например, если $D = 2$ и $w_1 = w_2 = 1, w_3 = 2$, то значение $S = 1,25 > 1$. Разделимый код не существует.

Если $S \leq 1$, то код все же может оказаться неразделимым. Вопрос о разделимости (однозначной декодируемости) кода решается с помощью критерия Сардинаса–Паттерсона.

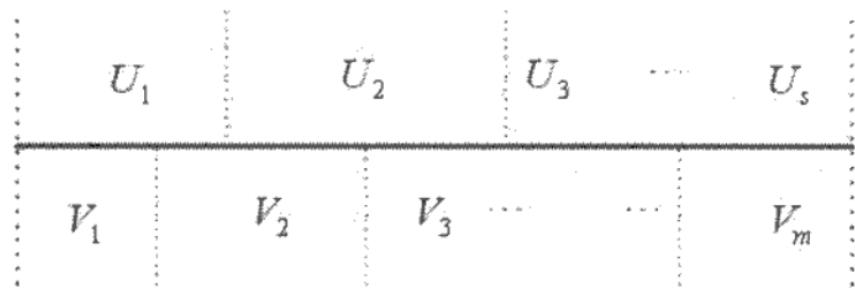


Рис. 3.2. Различные разбиения на кодовые слова для кода с неоднозначным декодированием

Предположим, что код не является разделимым. Тогда существует конечная последовательность кодовых слов, которая может быть разбита на кодовые слова двумя различными способами, скажем, как $(U_1 U_2 U_3 \dots U_s)$ и как $(V_1 V_2 V_3 \dots V_m)$ (см. рис. 3.2). Кодовое слово V_1 является префиксом кодового слова U_1 . Оставшийся суффикс называется допустимым суффиксом. Он в свою очередь должен быть или префиксом какого-нибудь кодового слова, или иметь в качестве префикса какое-нибудь кодовое слово, давая новый допустимый суффикс. Последний из возникающих суффиксов обязан быть кодовым словом. Критерий Сардинаса–Паттерсона состоит в следующем.

- По совокупности кодовых слов вычислить множество \mathcal{S} допустимых суффиксов.
- Код является разделимым (однозначно декодируемым) тогда и только тогда, когда множество \mathcal{S} допустимых суффиксов не содержит кодовых слов.

Пример 3.2. Рассмотрим код Z_1 со словами

$$Z_1 = \{z_1 = 0, z_2 = 01, z_3 = 10\}. \quad (3.8)$$

Кодовое слово z_1 является префиксом кодового слова z_2 . Оставшийся суффикс равен $s_1 = 1$. Этот суффикс является префиксом кодового слова z_3 . Новый оставшийся суффикс равен $s_2 = 0$. Так как он совпадает с кодовым словом $z_1 = 0$, то данный код является кодом с неоднозначным декодированием. Действительно, последовательность 010 разбивается на кодовые слова двумя способами: $010 = z_1 z_3 = z_2 z_1$.

Пример 3.3. Рассмотрим код Z_2 со словами

$$Z_2 = \{z_1 = 0, z_2 = 01, z_3 = 11\}. \quad (3.9)$$

Кодовое слово z_1 является префиксом кодового слова z_2 . Оставшийся суффикс равен $s_1 = 1$. В свою очередь этот суффикс является префиксом кодового слова z_3 . Новый оставшийся суффикс равен $s_2 = 1$. Таким образом, множество возможных суффиксов S состоит только из одного суффикса $S = \{s_1 = s_2 = 1\}$, который не совпадает ни с одним из кодовых слов. Этот код является разделимым (с однозначным декодированием).

Пример 3.4. Рассмотрим код Z_3 со словами

$$Z_3 = \{z_1 = 0, z_2 = 10, z_3 = 11\}. \quad (3.10)$$

В этом примере ни одно из кодовых слов не является префиксом другого кодового слова, поэтому множество оставшихся суффиксов пусто: $S = \emptyset$. Код является разделимым (с однозначным декодированием).

3.3. Префиксный код

Определение 3.1. Код, в котором ни одно из кодовых слов не является префиксом другого кодового слова, называется префиксным.

Префиксный код, очевидно, является разделимым.

Для анализа удобно представлять коды графами специального вида.

Граф $\{V, E\}$ задается множеством вершин V и множеством ребер E . Ребро соединяет пару вершин.

На рис. 3.3 приведены примеры графов с циклами и без циклов. Граф без циклов называется деревом.

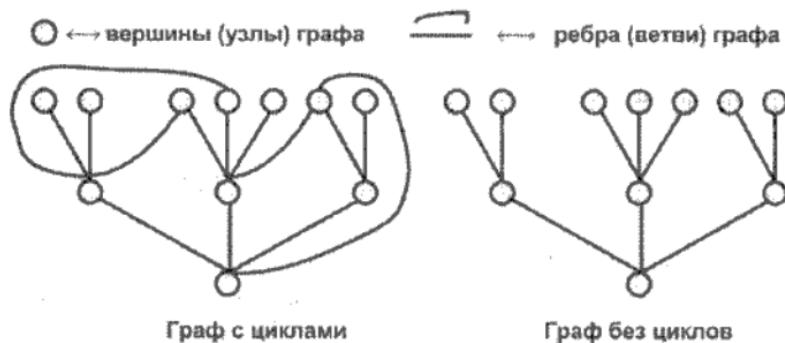


Рис. 3.3. Примеры графов

Особую роль в теории кодирования играют полные D -ичные кодовые деревья с L ярусами. Они строятся следующим образом. Ярус 0-го уровня состоит из одного узла, называемого корнем дерева. Из этого узла выходят D нумерованных ветвей, заканчивающихся D узлами 1-го яруса. Сами ветви нумеруются числами $0, 1, 2, \dots, D - 1$ в фиксированном порядке (например, слева направо). В свою очередь из каждого узла 1-го яруса выходит D нумерованных ветвей, заканчивающихся D узлами. Эти узлы являются узлами 2-го яруса. Их число равно D^2 . Число узлов на w -м промежуточном ярусе равно D^w . Построение продолжается до тех пор, пока не будет достигнут L -й ярус. Из узлов этого яруса не выходят ветви. Такие узлы называются листьями. Число листьев в полном дереве с L ярусами равно D^L . В полном дереве в каждый узел, кроме корня, входит точно одна ветвь. Точно D ветвей, кроме листьев, выходит из каждого узла.

Существует взаимно однозначное соответствие между множест-

жеством листьев полного дерева с L ярусами и множеством D -ичных последовательностей длины L . К каждому листу существует единственный путь из корня дерева. Ему можно поставить в соответствие последовательность, составленную из номеров ветвей; по которым этот путь проходит.

На рис. 3.4 приведен пример полного дерева для $D = 3$ и $L = 3$. В этом дереве число узлов соответственно на 1-м и 2-м ярусах равно $3^1 = 3$ и $3^2 = 9$. Число листьев, т.е. узлов на последнем, 3-м ярусе, равно $3^3 = 27$. Крайнему левому листу соответствует троичная последовательность (000). Черному листу соответствует троичная последовательность (102), а серому – троичная последовательность (200).

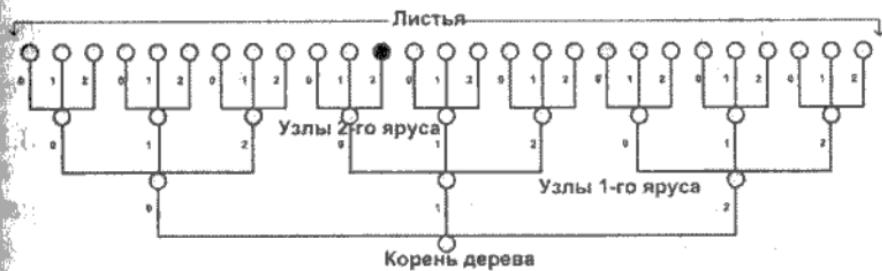


Рис. 3.4. Полное троичное дерево с 3 ярусами

Произвольный D -ичный код, заданный в виде набора последовательностей различной длины, можно представить в виде дерева. Если максимальная длина кодовой последовательности равна L , то это дерево получается из полного D -ичного дерева с L ярусами, если в последнем сохранить только пути, соответствующие кодовым последовательностям. Деревья для рассмотренных выше кодов Z_1, Z_2, Z_3 показаны на рис. 3.5. Они построены на основе двоичного полного дерева с двумя ярусами.

Черные узлы соответствуют кодовым словам. Для кодов Z_1, Z_2 слову $z_1 = 0$ соответствует промежуточный узел. Из него выходит ветвь к другому кодовому слову. Остальным словам соответствуют листья. Для префиксного кода Z_3 всем кодовым словам соответствуют листья. Это верно для любых префиксных кодов: код является префиксным тогда и только тогда, когда

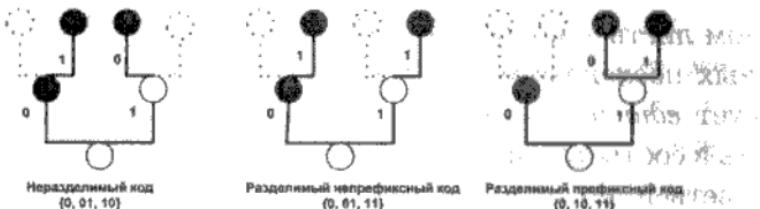


Рис. 3.5. Примеры представления кодов деревьями

где путь, соответствующий каждому кодовому слову, кончается листом.

Кодовые деревья являются полезным инструментом при доказательстве существования разделимых кодов.

Теорема 3.2 (Крафт). *Пусть положительные целые числа w_1, w_2, \dots, w_K удовлетворяют условию Мак-Миллана*

$$S = D^{-w_1} + D^{-w_2} + \dots + D^{-w_K} \leq 1. \quad (3.11)$$

Тогда существует D -ичный префиксный (следовательно, разделимый) код с этими длинами.

Доказательство. Для доказательства будет предъявлена конструкция префиксного кода. Без ограничения общности будем считать, что длины кодовых слов, удовлетворяющие неравенству Мак-Миллана, упорядочены по возрастанию:

$$w_1 \leq w_2 \leq \dots \leq w_K = L.$$

Построим D -ичное полное дерево с L ярусами. Первоначально на последнем ярусе присутствует D^L листьев.

1. На ярусе w_1 выбираем любой узел. Удаляем все пути, выходящие из этого узла. Тем самым этот узел становится листом. Путь из корня дерева к этому листу, соответствует первому кодовому слову длины w_1 . Число удаленных листьев на последнем ярусе равно D^{L-w_1} , так как все пути, исходящие из выбранного узла, образуют D -ичное полное дерево с $L - w_1$ ярусами. На последнем ярусе осталось

$D^L - D^{L-w_1} > 1$ листьев. Следовательно, не все узлы на ярусе w_2 были удалены при формировании первого кодового слова.

2. На ярусе w_2 выбираем любой из сохранившихся узлов. Удаляем все пути, выходящие из выбранного узла, превращая его в лист. Путь из корня дерева к этому листу соответствует второму кодовому слову длины w_2 . Число листьев, удаленных на последнем ярусе при формировании второго кодового слова, равно D^{L-w_2} , так как все пути, исходящие из выбранного узла, образуют D -ичное полное дерево с $L - w_2$ ярусами. На последнем ярусе осталось $D^L - D^{L-w_1} - D^{L-w_2} = D^L(1 - D^{-w_1} - D^{-w_2})$ листьев. Это целое число строго положительно, так как из условия (3.11) следует, что выражение в скобках строго положительно. Следовательно, не все узлы на ярусе w_3 были удалены при формировании первого и второго кодовых слов.
3. Продолжаем процедуру построения кодовых слов.
4. На ярусе w_{K-1} выбираем любой из сохранившихся узлов. Удаляем все пути, выходящие из выбранного узла, превращая его в лист. Путь из корня дерева к этому листу, соответствует $(K-1)$ -му кодовому слову длины w_{K-1} . Число листьев, удаленных на последнем ярусе при формировании $(K-1)$ -го кодового слова, равно $D^{L-w_{K-1}}$, так как все пути, исходящие из выбранного узла, образуют D -ичное полное дерево с $L - w_{K-1}$ ярусами. На последнем ярусе осталось

$$D^L - D^{L-w_1} - \dots - D^{-w_{K-1}} = D^L(1 - D^{-w_1} - \dots - D^{-w_{K-1}})$$

листьев. Это целое число строго положительно, так как из условия (3.11) все еще следует, что выражение в скобках строго положительно. Следовательно, не все листья на последнем ярусе $w_K = L$ были удалены при формировании первых $K-1$ кодовых слов.

5. На ярусе $w_K = L$ выбираем любой из сохранившихся листьев. Удаляем все остальные листья вместе с входящими

в них ветвями. Путь из корня дерева к этому листу, соответствует последнему K -му кодовому слову длины w_K .

Таким образом, построен код из K слов, каждому из которых на кодовом дереве соответствует путь, заканчивающийся листом. Другими словами, построен префиксный код с заданными длинами.

Пример 3.5. Построим двоичный префиксный код с длинами $w_1 = w_2 = 2, w_3 = 3, w_4 = 4$. Так как $S = 2^{-2} + 2^{-2} + 2^{-3} + 2^{-4} = \frac{11}{16} < 1$, то такой код существует. На рис. 3.6 показано построение этого кода. Сначала строится полное двоичное дерево с $L = w_4 = 4$ ярусами. Затем на ярусе 2 выбирается узел, соответствующий слову z_1 с длиной $w_1 = 2$. Он превращается в лист (черный кружок) путем отбрасывания всех путей, исходящих из него. Один из оставшихся на ярусе 2 узлов выбирается для слова z_2 с длиной $w_2 = 2$. Этот узел также превращается в лист. Для слова z_3 с длиной $w_3 = 3$ выбирается один из узлов, сохранившихся на ярусе 3, с последующим превращением его в лист. Наконец, для слова z_4 с длиной $w_4 = 4$ выбирается один из сохранившихся на ярусе 4 листьев.

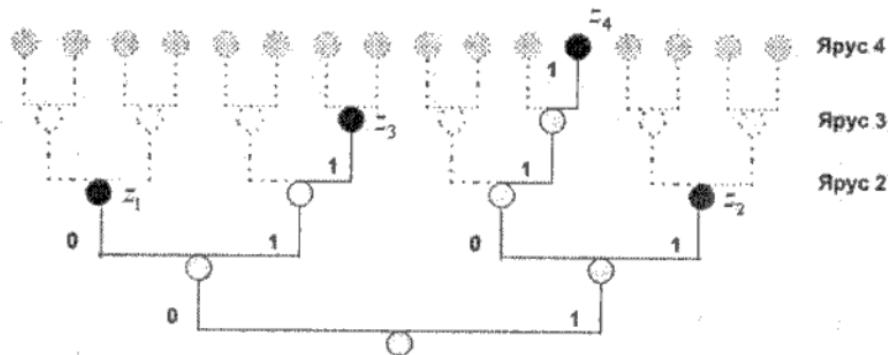
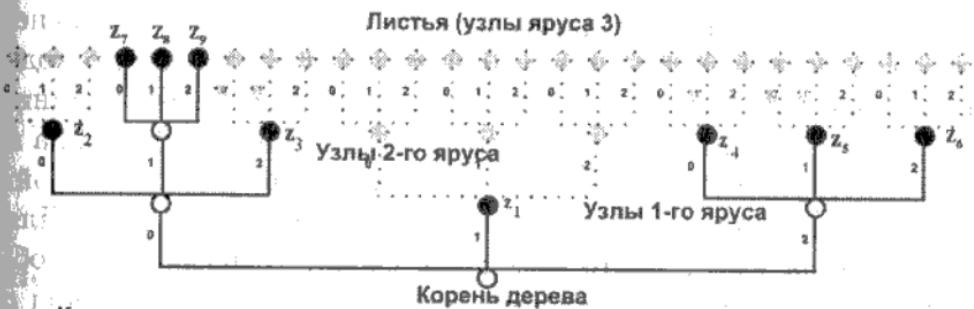


Рис. 3.6. Построение двоичного префиксного кода с длинами $w=2, 2, 3, 4$.

Кодовые слова, соответствующие листьям, образуют префиксный код $\{z_1 = 00, z_2 = 11, z_3 = 011, z_4 = 1011\}$.

Пример 3.6. Построим троичный префиксный код с длинами $w_1 = 1, w_2 = w_3 = w_4 = w_5 = w_6 = 2, w_7 = w_8 = w_9 = 3$. Так как $S = 3^{-1} + 5 \cdot 3^{-2} + 3 \cdot 3^{-3} = 1$, то такой код существует. На рис. 3.7 показано построение этого кода. Сначала строится полное троичное дерево с $L = w_9 = 3$ ярусами. Затем на ярусе 1 выбирается узел, соответствующий слову z_1 с длиной $w_1 = 1$. Он превращается в лист (черный кружок) путем отбрасывания всех путей, исходящих из него. На ярусе 2 из оставшихся узлов выбираются 5 узлов для слов z_2, z_3, z_4, z_5, z_6 с длиной 2. Эти узлы также превращаются в листья путем отбрасывания всех путей, исходящих из них. На последнем ярусе 3 сохранились три листа, которые и выбираются для слов z_7, z_8, z_9 с длиной 3.



Кодовые слова: $z_1 = 1, z_2 = 00, z_3 = 02, z_4 = 20, z_5 = 21, z_6 = 22, z_7 = 010, z_8 = 011, z_9 = 012$

Рис. 3.7. Построение троичного префиксного кода с длинами $w = \{1, 2, 2, 2, 2, 2, 3, 3, 3\}$

Кодовые слова, соответствующие листьям, образуют префиксный код $z_1 = 1, z_2 = 00, z_3 = 02, z_4 = 20, z_5 = 21, z_6 = 22, z_7 = 010, z_8 = 011, z_9 = 012$.

3.4. Теоремы Шеннона для источника

Рассмотрим источник сообщений U , порождающий буквы u_1, u_2, \dots, u_K с вероятностями p_1, p_2, \dots, p_K , $\sum_{i=1}^K p_i = 1$.

Для побуквенного кодирования можно использовать любой разделимый D -ичный код со словами z_1, z_2, \dots, z_K и длинами

w_1, w_2, \dots, w_K . Качество используемого кода оценивается *средней длиной* кодового слова в D -ичных символах:

$$E_{Uw}(U) = \sum_{i=1}^K p_i w_i; \quad (3.12)$$

или в битах:

$$(E_{Uw}(U))_{bit} = \log_2 D \cdot (Ew)_D = \log_2 D \cdot \sum_{i=1}^K p_i w_i. \quad (3.13)$$

Чем меньше средняя длина, тем лучше код.

Основной целью кодирования источника является уменьшение избыточности, т.е. выбор для источника разделимого кода с наименьшей средней длиной. Возникает вопрос, какова минимально достижимая средняя длина для заданного источника и как конструктивно выбрать код с наименьшей средней длиной.

В теории информации, по традиции, утверждения типа “Для любого кода имеет место некоторое свойство” называются обратными теоремами, а утверждения типа “Существует код с заданным свойством” — прямыми теоремами.

К. Шенноном были сформулированы и доказаны прямые и обратные теоремы для кодирования источников общего вида.

Применительно к побуквенному кодированию, обратная теорема Шеннона утверждает следующее.

Теорема 3.3. Для любого разделимого кода с длинами w_1, w_2, \dots, w_K средняя длина сообщений больше или равна энтропии источника, нормированной на двоичный логарифм от числа букв в алфавите кодера. Иными словами, всегда выполняется неравенство

$$\frac{H(U)}{\log_2 D} \leq E_{Uw}(U). \quad (3.14)$$

Доказательство. Надо доказать, что

$$H(U) - \log_2 D \cdot E_U w(U) \leq 0. \quad (3.15)$$

Расшифруем эту запись, используя определения входящих в нее величин:

$$\begin{aligned} & \sum_{i=1}^K p_i \log_2 \frac{1}{p_i} - \sum_{i=1}^K p_i w_i \log_2 D = \\ & = \sum_{i=1}^K p_i \log_2 \frac{1}{p_i} + \sum_{i=1}^K p_i \log_2 D^{-w_i} = \sum_{i=1}^K p_i \log_2 \frac{D^{-w_i}}{p_i}. \end{aligned} \quad (3.16)$$

Применим неравенство логарифма к правой части соотношения (3.16):

$$\begin{aligned} & \sum_{i=1}^K p_i \log_2 \frac{D^{-w_i}}{p_i} \leq \sum_{i=1}^K p_i \left(\frac{D^{-w_i}}{p_i} - 1 \right) \log_2 e = \\ & = \left(\sum_{i=1}^K D^{-w_i} - \sum_{i=1}^K p_i \right) \log_2 e \leq 0. \end{aligned} \quad (3.17)$$

Неравенство в правой части (3.17) верно, так как

$$\sum_{i=1}^K D^{-w_i} \leq 1 \quad (3.18)$$

$$\sum_{i=1}^K p_i = 1. \quad (3.19)$$

Теперь докажем прямую теорему Шеннона.

Теорема 3.4. *Существует префиксный, т.е. разделимый код, для которого средняя длина сообщений отличается от нормированной энтропии не более, чем на единицу:*

$$E_U w(U) < \frac{H(U)}{\log_2 D} + 1. \quad (3.20)$$

Доказательство. Для доказательства будет предъявлен конкретный префиксный код, построенный по методу Шеннона–Фано.

Функция $\lceil x \rceil$ для вещественного аргумента x определяется как ближайшее к x сверху целое число. Эта функция удовлетворяет очевидным неравенствам

$$x \leq \lceil x \rceil < x + 1. \quad (3.21)$$

Выбор длин кодовых слов в коде Шеннона–Фано производится по правилу

$$w_i = \lceil \log_D \frac{1}{p_i} \rceil, \quad i = 1, 2, \dots, K. \quad (3.22)$$

Проверяем, что неравенство Мак–Миллана–Крафта выполняется при таком выборе длин. Используя левое неравенство в (3.21), имеем

$$\begin{aligned} \log_D \frac{1}{p_i} &\leq w_i, \\ D^{-w_i} &\leq p_i, \\ \sum_{i=1}^K D^{-w_i} &\leq \sum_{i=1}^K p_i = 1. \end{aligned}$$

Оценим сверху среднюю длину кодового слова, используя правое неравенство в (3.21):

$$\begin{aligned} E_U w(U) &= \sum_{i=1}^K p_i w_i = \sum_{i=1}^K p_i \lceil \log_D \frac{1}{p_i} \rceil \leq \\ &\leq \sum_{i=1}^K p_i (\log_D \frac{1}{p_i} + 1) = \frac{\sum_{i=1}^K p_i \log_2 \frac{1}{p_i}}{\log_2 D} + 1 = \\ &= \frac{H(U)}{\log_2 D} + 1. \end{aligned} \quad (3.23)$$

3.5. Коды Шеннона и Фано

Фактически коды Шеннона и Фано с длинами (3.22) строятся по-разному. Код Шеннона строится следующим образом.

1. Символы нумеруются в порядке убывания вероятностей:
 $p_1 \geq p_2 \geq \dots \geq p_K$.
2. Вычисляются длины кодовых слов в соответствии с (3.22):
 $w_1 \leq w_2 \leq \dots \leq w_K$.
3. Вычисляется таблица кумулятивных сумм:
 $0, D^{-w_1}, D^{-w_1} + D^{-w_2}, \dots, D^{-w_1} + D^{-w_2} + \dots + D^{-w_{K-1}}$.
4. Вычисляется представление кумулятивных сумм в виде D -ичных дробей.
5. В качестве i -го кодового слова выбираются первые w_i цифр D -ичной дроби, соответствующей i -й кумулятивной сумме.

Код Фано строится следующим образом.

1. Символы нумеруются в порядке убывания вероятностей:
 $p_1 \geq p_2 \geq \dots \geq p_K$.
2. Вычисляется таблица кумулятивных вероятностей: $0, p_1, p_1 + p_2, p_1 + p_2 + p_3, \dots, p_1 + p_2 + \dots + p_{K-1}$.
3. Вычисляется представление кумулятивных вероятностей в виде D -ичных дробей.
4. В качестве i -го кодового слова выбираются первые w_i цифр D -ичной дроби, соответствующей i -й кумулятивной вероятности.

Пример 3.7. Пусть $p_1 = 0.4; p_2 = 0.3; p_3 = 0.2; p_4 = 0.1; D = 2; K = 4$. Найдем длины кодовых слов в соответствии с алгоритмом Шеннона–Фано:

$$w_1 = \lceil \log_2 \frac{1}{0.4} \rceil = 2; \quad w_2 = \lceil \log_2 \frac{1}{0.3} \rceil = 2;$$
$$w_3 = \lceil \log_2 \frac{1}{0.2} \rceil = 3; \quad w_4 = \lceil \log_2 \frac{1}{0.1} \rceil = 4.$$

Проверим выполнение неравенства Крафта:

$$S = 2^{-2} + 2^{-2} + 2^{-3} + 2^{-4} = \frac{11}{16} < 1,$$

т.е. неравенство выполняется.

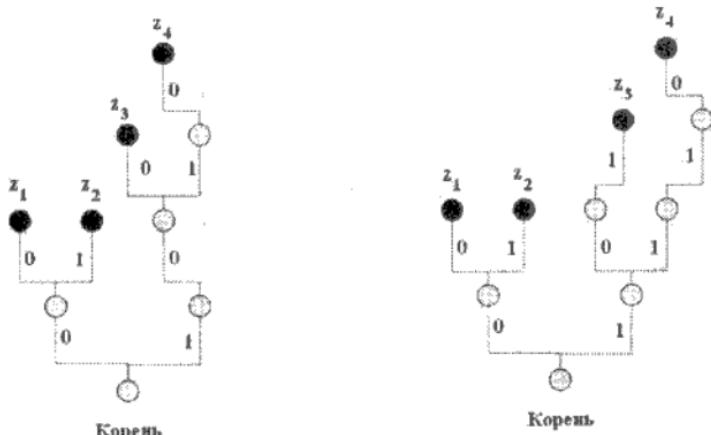
Построим этот код по методу Шеннона:

Символ i	p_i	w_i	$\sum D^{-w_i}$	Дробь	Код
1	0.4	2	0	0.0000...	$z_1 = 00$
2	0.3	2	2^{-2}	0.0100...	$z_2 = 01$
3	0.2	3	$2^{-2} + 2^{-2}$	0.1000...	$z_3 = 100$
4	0.1	4	$2^{-2} + 2^{-2} + 2^{-3}$	0.1010...	$z_4 = 1010$

Построим этот код по методу Фано:

Символ i	p_i	w_i	$\sum p_i$	Дробь	Код
1	0.4	2	0	0.0000...	$z_1 = 00$
2	0.3	2	0.4	0.0110...	$z_2 = 01$
3	0.2	3	0.7	0.1010...	$z_3 = 101$
4	0.1	4	0.9	0.1110...	$z_4 = 1110$

Средняя длина равна 2.4, энтропия источника равна $H = 1,894$.



Код Шеннона:
 $z_1 = 00; z_2 = 01; z_3 = 100; z_4 = 1010.$

Код Фано:
 $z_1 = 00; z_2 = 01; z_3 = 101; z_4 = 1110.$

Рис. 3.8. Кодовые деревья кодов Шеннона и Фано

На рис. 3.8 приведены кодовые деревья кодов Шеннона и Фано. Из рисунка видно, что конструкции кодов не оптимальны. Среднюю длину можно уменьшить, если два последних слова заменить на $\tilde{z}_3 = 10, \tilde{z}_4 = 11$. В этом случае средняя длина равна 2, что меньше 2.4.

3.6. Код Хаффмена

Оптимальная конструкция префиксного кода с минимальной средней длиной предложена Хаффменом. В кодовом дереве буквам соответствуют листья, а соответствующие им кодовые слова отождествляются с путем к этому листу из корня дерева.

В оптимальном коде свободные узлы, не соответствующие кодовым словам, могут существовать только на последнем ярусе. В противном случае можно было бы кодовый лист последнего яруса переместить на свободный узел предыдущих ярусов, уменьшив тем самым среднюю длину кода. Как будет показано, свободные узлы на последнем ярусе возможны только для $D \geq 3$.

Сначала рассмотрим двоичный случай, $D = 2$. Будем нумеровать символы источника в порядке убывания вероятностей: $p_1 \geq p_2 \geq \dots \geq p_K$.

Лемма 3.1. Наименее вероятной букве соответствует самое длинное кодовое слово.

Доказательство. Рассмотрим в сумме $Ew = \sum_{i=1}^K p_i w_i$ слагаемые $p_K w_K$ и $p_j w_j$, где $p_j > p_K$. Предположим, что $w_K < w_j$. Покажем, что такой код не оптимален и что среднюю длину можно уменьшить. Поменяем в кодовом дереве местами листья, соответствующие буквам u_K и u_j , оставив остальные листья (кодовые слова) на месте. Тогда средняя длина уменьшится, так как разность средних длин

$$p_j w_j + p_K w_K - p_K w_j - p_j w_K = (p_j - p_K)(w_j - w_K)$$

положительна.

Лемма 3.2. Двум наименее вероятным буквам соответствуют самые длинные кодовые слова одинаковой длины, причем они отличаются только в последнем символе.

Доказательство

В оптимальном двоичном коде не может существовать свободных узлов на последнем ярусе. Действительно, как уже показано, наименее вероятной букве соответствует лист в последнем ярусе оптимального дерева. Однако структура последнего яруса не может иметь вид, показанный на рис. 3.9A, так как в этом случае кодовый лист можно переместить в предыдущий ярус.



Рис. 3.9. Структура последнего яруса оптимального двоичного кода Хаффмена и создание виртуальной буквы

дущий ярус, не меняя остальных кодовых листьев. Средняя длина при этом уменьшится. В оптимальном коде (рис. 3.9B) из промежуточного узла предыдущего яруса должно возникнуть два кодовых листа на последнем ярусе. Один из них – наименее вероятная буква согласно лемме 3.1, другой кодовый лист соответствует следующей наименее вероятной букве. Пути к ним отличаются только в последних ветвях. На рис. 3.9C показано, что две наименее вероятных буквы можно объединить в виртуальную букву $\hat{u}_{K-1} = [u_{K-1}, u_K]$, присвоить ей вероятность $\hat{p}_{K-1} = p_{K-1} + p_K$, равную сумме вероятностей двух наименее вероятных букв, и поместить ее в узел предыдущего яруса. Остальные кодовые листья оставим без изменений. Полученное

кодовое дерево будет соответствовать оптимальному префиксному коду для источника с числом букв на 1 меньшему, чем в исходном источнике. Это наблюдение позволяет сформулировать алгоритм Хаффмена для рекуррентного построения оптимального префиксного кода.

Алгоритм Хаффмена построения оптимального двоичного префиксного кода состоит в следующем.

Шаг 1. Упорядочить буквы источника $u_1, u_2, \dots, u_{K-1}, u_K$ в порядке убывания их вероятностей:

$$p(u_1) \geq p(u_2) \geq \dots \geq p(u_{K-1}) \geq p(u_K).$$

Приписать буквы и их вероятности листьям строящегося дерева.

Шаг 2. Создать из двух наименее вероятных букв виртуальную букву $v_1 = [u_{K-1}, u_K]$. Приписать ей вероятность $p(v_1) = p(u_{K-1}) + p(u_K)$. Создать новый текущий виртуальный источник, удалив из исходного две наименее вероятных буквы с их вероятностями и добавив созданную виртуальную букву с ее вероятностью. Повторить для текущего виртуального источника шаг 1. На каждом шаге новая виртуальная буква состоит из объединения двух наименее вероятных виртуальных букв.

Шаг 3. Продолжать, пока в текущем источнике не останется $D = 2$ виртуальных буквы с их вероятностями.

Шаг 4. Построить дерево из корня и первого яруса с двумя ветвями, кончающимися листьями. Листьям приписать виртуальные буквы текущего источника после выполнения шага 3. Если одна из виртуальных букв состоит только из *одной* буквы исходного источника, то этот лист является окончательным кодовым листом, соответствующим этой букве. Если же виртуальная буква состоит из *двух* букв, то этот лист превращается в узел, из которого строится ярус

следующего уровня с двумя листьями, которым приписываются компоненты виртуальной буквы. Продолжать, пока всем возникающим листам не будут приписаны буквы исходного источника.

Пример 3.8. Рассмотрим источник \mathcal{U} , порождающий 7 букв с вероятностями

Буква	u_1	u_2	u_3	u_4	u_5	u_6	u_7
Вероятность	0.3	0.25	0.15	0.1	0.08	0.07	0.05

Из источника удаляются две буквы u_6, u_7 . Вместо них создается новая виртуальная буква $v_1 = [u_6, u_7]$ с вероятностью

$$p(v_1) = p(u_6) + p(u_7) = 0.12.$$

Новый виртуальный источник \mathcal{U}_1 упорядочивается по убыванию вероятностей:

Буква	u_1	u_2	u_3	v_1	u_4	u_5
Вероятность	0.3	0.25	0.15	0.12	0.1	0.08

Из виртуального источника удаляются две буквы u_4, u_5 . Вместо них создается новая виртуальная буква $v_2 = [u_4, u_5]$ с вероятностью $p(v_2) = p(u_4) + p(u_5) = 0.18$. Новый виртуальный источник \mathcal{U}_2 упорядочивается по убыванию вероятностей:

Буква	u_1	u_2	v_2	u_3	v_1
Вероятность	0.3	0.25	0.18	0.15	0.12

Из виртуального источника удаляются две буквы u_3, v_1 . Вместо них создается новая виртуальная буква $v_3 = [u_3, v_1]$ с вероятностью $p(v_3) = p(u_3) + p(v_1) = 0.27$. Новый виртуальный источник \mathcal{U}_3 упорядочивается по убыванию вероятностей:

Буква	u_1	v_3	u_2	v_2
Вероятность	0.3	0.27	0.25	0.18

Из виртуального источника удаляются две буквы u_2, v_2 . Вместо них создается новая виртуальная буква $v_4 = [u_2, v_2]$ с вероятностью $p(v_4) = p(u_2) + p(v_2) = 0.43$. Новый виртуальный источник \mathcal{U}_4 упорядочивается по убыванию вероятностей:

Буква	v_4	u_1	v_3
Вероятность	0.43	0.3	0.27

Из виртуального источника удаляются две буквы u_1, v_3 . Вместо них создается новая виртуальная буква $v_5 = [u_1, v_3]$ с вероятностью $p(v_5) = p(u_1) + p(v_3) = 0.57$. Новый виртуальный источник \mathcal{U}_5 упорядочивается по убыванию вероятностей:

Буква	v_5	v_4
Вероятность	0.57	0.43

Он состоит из $D = 2$ букв, и можно приступить к построению оптимального дерева. Оптимальное дерево для источника \mathcal{U}_5 показано на рис. 3.10А. Листьям этого кода соответствуют вирту-

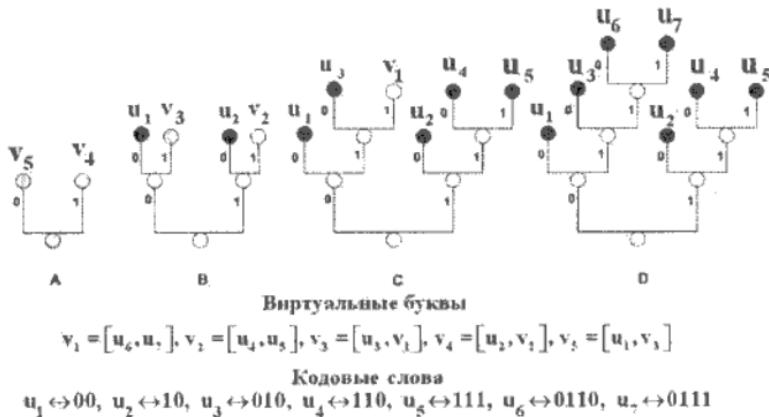


Рис. 3.10. Построение оптимального двоичного кода Хаффмена

альные буквы v_5 и v_4 , поэтому из них выпускаются ветви для построения оптимальных кодов виртуальных источников \mathcal{U}_4 (две ветви из левого листа) и \mathcal{U}_3 (две ветви из правого листа). На рис. 3.10В изображено оптимальное дерево для виртуального источника \mathcal{U}_3 . В построенном дереве появились листья, соответствующие буквам u_1 и u_2 исходного источника \mathcal{U} . Они в дальнейшем не изменяются. Остальные листья соответствуют виртуальным буквам $v_3 = [u_3, v_1]$ и $v_2 = [u_4, u_5]$. Из них выпускаются ветви, которые приводят к оптимальному дереву для источника \mathcal{U}_1 .

(рис. 3.10С). В этом дереве только один лист соответствует виртуальной букве $v_1 = [u_6, u_7]$. Остальные листья соответствуют буквам u_1, u_2, u_3, u_4, u_5 источника U и в дальнейшем не меняются. Наконец, выпущенные из виртуального листа $v_1 = [u_6, u_7]$ две ветви заканчиваются листьями для букв u_6 и u_7 . Построение завершено. Кодовые слова имеют вид

Буква	u_1	u_2	u_3	u_4	u_5	u_6	u_7
Код буквы	00	10	010	110	111	0110	0111

Теперь перейдем к недвоичным кодам, $D > 2$. Если $D > K$, то сжатие невозможно. Обычно выполняется противоположное неравенство $K > D$. При $D > 2$ в оптимальном кодовом дереве не может быть свободных листьев на всех ярусах, кроме последнего, но на последнем ярусе могут быть свободные листья, которым не соответствует никакая буква источника. Обозначим их число через r . Возможные структуры последнего яруса кодового дерева показаны на рис. 3.11. Для оптимального кода число r не может равняться D (рис. 3.11А) или $D - 1$ (рис. 3.11В), так как в обоих случаях кодовый лист последнего яруса может быть



Рис. 3.11. Структура последнего яруса неоптимальных недвоичных кодов и оптимального недвоичного кода Хаффмена

перемещен в узел предыдущего яруса, что приведет к уменьшению средней длины. Основная проблема: как найти число r . Следующая лемма показывает, как это можно сделать.

Лемма 3.3. Поделим целое (отрицательное) число $D - K$ на целое число $D - 1$ с остатком:

$$D - K = q(D - 1) + r, \quad \text{где } 0 \leq r < D - 1. \quad (3.24)$$

В оптимальном кодовом дереве число свободных листьев на последнем ярусе равно остатку r .

Доказательство. Сначала докажем, что в любом D -ичном дереве число листьев N_{leaves} равно

$$N_{\text{leaves}} = D + s(D - 1), \quad (3.25)$$

где s – число промежуточных узлов дерева, не считая корня. Действительно, эта формула верна при $s = 0$ (см. рис. 3.11А), так как в этом случае промежуточных узлов, кроме корня, нет, а из корня выходит D ветвей, кончающихся листьями. Чтобы построить более сложное дерево, необходимо один лист превратить в промежуточный узел и выпустить из него D ветвей, кончающихся листьями, так что общее число листьев станет равным $D + (D - 1)$. Так как при каждом добавлении нового промежуточного узла число листьев возрастает на $D - 1$, то по индукции приходим к формуле (3.25).

В оптимальном кодовом дереве число листьев также определяется формулой (3.25) для некоторого s . Из них K листьев будут кодовыми, а оставшиеся $r = N_{\text{leaves}} - K$ – свободными. Но эти листья могут быть только на последнем ярусе, так что $0 \leq r < D - 1$. Итак, имеем

$$r = N_{\text{leaves}} - K = D + s(D - 1) - K,$$

или

$$D - K = -s(D - 1) + r, \quad \text{где } 0 \leq r < D - 1.$$

Но это и есть соотношение (3.24), если положить $q = -s$. Алгоритм Хаффмена построения оптимального недвоичного префиксного кода состоит в следующем.

Шаг 0. Найти число r свободных листьев на последнем ярусе по формуле (3.24).

Шаг 1. Упорядочить буквы источника $u_1, u_2, \dots, u_{K-1}, u_K$ в порядке убывания их вероятностей:

$$p(u_1) \geq p(u_2) \geq \cdots \geq p(u_{K-1}) \geq p(u_K).$$

Приписать буквы и их вероятности листьям строящегося дерева.

Шаг 2. Создать из $D - r$ наименее вероятных букв виртуальную букву

$$[u_{K-r+1}, \dots, u_{K-1}, u_K].$$

Приписать ей вероятность

$$p([u_{K-r+1}, \dots, u_{K-1}, u_K]) = p(u_{K-r+1}) + \cdots + p(u_K).$$

Создать новый текущий виртуальный источник, удалив из исходного $D - r$ наименее вероятные буквы с их вероятностями и добавив созданную виртуальную букву с ее вероятностью. Новый источник будет содержать

$K - D + r + 1 = s(D - 1) + 1$ букв. Упорядочить буквы источника в порядке убывания их вероятностей.

Шаг 3. Создать из D наименее вероятных букв виртуальную букву, определяемую как их объединение. Приписать ей вероятность, равную сумме вероятностей этих букв. Создать новый текущий виртуальный источник, удалив из предыдущего источника D наименее вероятные буквы с их вероятностями и добавив созданную виртуальную букву с ее вероятностью. Новый источник будет содержать $(s - 1)(D - 1) + 1$ букв. Упорядочить буквы источника в порядке убывания их вероятностей. Повторить для него шаг 3. На каждом шаге новая виртуальная буква состоит из объединения D наименее вероятных виртуальных букв.

Шаг 4. Продолжать до тех пор, пока в текущем виртуальном источнике не останется точно D виртуальных букв с их вероятностями.

Шаг 5. Построить дерево из корня и первого яруса с D листьями. Листьям приписать виртуальные буквы последнего текущего источника после выполнения шага 4. Если какая-

нибудь из виртуальных букв состоит только из *одной* буквы исходного источника, то этот узел объявляется окончательным кодовым листом, соответствующим этой букве. Если виртуальная буква состоит из D букв, то из этого узла строится ярус следующего уровня с D листьями, которым приписываются компоненты виртуальной буквы. Продолжать, пока всем возникающим листьям не будут приписаны буквы исходного источника. На последнем этапе только $D - r$ листьям будут приписаны буквы исходного источника. Оставшиеся r листьев останутся свободными.

Пример 3.9. Пусть источник \mathcal{U} , порождающий $K = 8$ букв, кодируется в алфавите из $D = 4$ символов. Так как

$D - K = -4$, $D - 1 = 3$, то в соответствии с (3.24) получаем: $-4 = -2 \cdot 3 + 2$. Следовательно, на последнем ярусе оптимального дерева будет $r = 2$ свободных листа. Пусть вероятности букв источника \mathcal{U} упорядочены:

Буква	u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8
Вероятность	0.3	0.25	0.15	0.1	0.07	0.06	0.04	0.03

Из источника удаляются $D - r = 2$ буквы u_7, u_8 . Вместо них создается новая виртуальная буква $v_1 = [u_7, u_8]$ с вероятностью $p([u_7, u_8]) = p(u_7) + p(u_8) = 0.07$. Новый виртуальный источник \mathcal{U}_1 упорядочивается по убыванию вероятностей:

Буква	u_1	u_2	u_3	u_4	u_5	$v_1 = [u_7, u_8]$	u_6
Вероятность	0.3	0.25	0.15	0.1	0.07	0.07	0.06

На этом шаге из виртуального источника удаляются $D = 4$ наименее вероятных буквы $u_4, u_5, v_1 = [u_7, u_8], u_6$. Вместо них создается новая виртуальная буква $v_2 = [u_4, u_5, v_1, u_6]$, вероятность которой равна

$$p([u_4, u_5, [u_7, u_8], u_6]) = p(u_4) + p(u_5) + p([u_7, u_8]) + p(u_6) = 0.3.$$

Новый виртуальный источник \mathcal{U}_2 упорядочивается по убыванию вероятностей:

Буква	$v_2 = [u_4, u_5, [u_7, u_8], u_6]$	u_1	u_2	u_3
Вероятность	0.3	0.3	0.25	0.15

Этот источник порождает точно $D = 4$ буквы, поэтому можно приступать к построению оптимального дерева. Оптимальное дерево для источника \mathcal{U}_2 показано на рис. 3.12А. Листьям этого

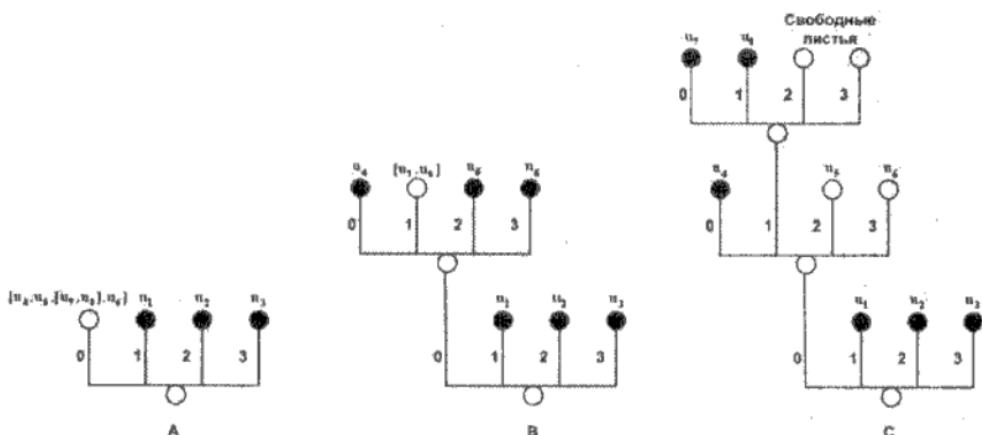


Рис. 3.12. Построение оптимального двоичного кода Хаффмена

кода соответствуют виртуальная буква $v_2 = [u_4, u_5, [u_7, u_8], u_6]$ и буквы u_1, u_2, u_3 . Три последних листа являются кодовыми и в дальнейшем не меняются. Из первого листа выпускаются $D = 4$ ветви для построения оптимального кода виртуального источника \mathcal{U}_1 , как показано на рис. 3.12В. В построенном дереве появились листья, соответствующие буквам u_4, u_5 и u_6 исходного источника \mathcal{U} . Они в дальнейшем не изменяются. Оставшийся лист соответствует виртуальной букве $v_1 = [u_7, u_8]$. Из него выпускаются $D = 4$ ветви, которые приводят к оптимальному дереву для исходного источника \mathcal{U} (рис. 3.12С). В этом дереве $D - r = 2$ листа соответствуют буквам u_7 и u_8 , а остальные $r = 2$ листа остаются свободными. Построение завершено. Кодовые слова имеют вид

Буква	u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8
Код буквы	1	2	3	00	02	03	010	011

Глава 4

Кодирование блоков

Теперь перейдем к кодированию блоков. Блоки постоянной длины кодируются блоками переменной длины. Всю полу бесконечную последовательность сообщений источника U_1, U_2, \dots разделяем на блоки из n букв. Обозначим

$$U_j^n = (U_j, U_{j+1}, \dots, U_{j+n-1}).$$

Тогда можно считать, что источник порождает последовательность блоков U_1^n, U_2^n, \dots . Блоки можно рассматривать как буквы нового алфавита. Число букв равно K^n . Распределение последовательности блоков можно найти, зная распределение последовательности букв исходного источника. В частности, можно найти вероятности порождения каждой новой буквы-блока.

4.1. Кодирование источника без памяти

Рассмотрим побуквенное кодирование преобразованного источника, рассматривая блок как одну новую букву. Тогда обратная и прямая теоремы Шеннона приводят к соотношениям:

$$\frac{1}{n} \left(\frac{H(U^n)}{\log_2 D} \right) \leq \frac{1}{n} (E_{U^n} w(U^n)) < \frac{1}{n} \left(\frac{H(U^n)}{\log_2 D} + 1 \right). \quad (4.1)$$

Здесь $E_{U^n} w(U^n)$ означает среднюю длину кодового слова на *входной блок* длины n . Для стационарного источника без памяти цепное равенство имеет вид

$$\begin{aligned} H(U^n) &= H(U_1 U_2 \dots U_n) = \\ &= H(U_1) + H(U_2) + \dots + H(U_n) = nH(U). \end{aligned} \quad (4.2)$$

Перепишем неравенства (4.1) в следующем виде:

$$\frac{H(U)}{\log_2 D} \leq \frac{E_{U^n} w(U^n)}{n} < \frac{H(U)}{\log_2 D} + \frac{1}{n}. \quad (4.3)$$

Как видно из соотношений (4.3), кодирование блоками более эффективно, так как средняя длина на *входной символ* меньше, чем при побуквенном кодировании исходного источника. При больших длинах блоков нижняя и верхняя границы средней длины сближаются и при $n \rightarrow \infty$ становятся равными нормированной энтропии $\frac{H(U)}{\log_2 D}$.

4.2. Кодирование источников с памятью

Для общего стационарного источника по-прежнему верны соотношения (4.1), которые можно переписать в виде

$$\frac{H_n}{\log_2 D} \leq \frac{1}{n} (E_{U^n} w(U^n)) < \frac{H_n}{\log_2 D} + \frac{1}{n}. \quad (4.4)$$

Так как функция H_n монотонно не возрастает по n и имеет предел $H_\infty \leq H_n$, то для любого $\varepsilon > 0$ найдется такое n_0 , что $H^n < H_\infty + \varepsilon$ при $n > n_0$. Это означает, что

$$\frac{H_\infty}{\log_2 D} \leq \frac{1}{n} (E_{U^n} w(U^n)) < \frac{H_\infty + \varepsilon}{\log_2 D} + \frac{1}{n}, \quad (4.5)$$

где левое неравенство верно для всех n , а правое – для $n > n_0$. Таким образом, средняя длина кодового слова $\frac{1}{n} (E_{U^n} w(U^n))$ на *входной символ* может быть сколь угодно близкой к энтропии источника при достаточной длине n .

4.3. Код Танстолла

Здесь рассматривается источник без памяти и новый способ кодирования, когда входным блокам переменной длины ставятся в соответствие выходные блоки одинаковой длины.

Пусть $\mathcal{U} = \{u_1, u_2, \dots, u_K\}$ – входной алфавит, N – длина выходного блока, $\mathcal{V} = \{v_1, v_2, \dots, v_D\}$ – выходной алфавит.

Если $K \leq D^N$, то проблемы кодирования нет. Можно было бы каждой букве u_i , $i = 1, 2, \dots, K$, приписать некоторый выходной блок z_i длины N , однако уменьшения избыточности в этом случае не происходит.

Если $K < D^N$, то уменьшить избыточность можно при подходящем кодировании. Идея кодирования состоит в следующем. Любая последовательность входных символов должна быть разбита на последовательность входных блоков переменной длины, причем количество M различных блоков не должно превосходить числа D^N различных выходных блоков. Чтобы обеспечить однозначность декодирования, различные входные блоки должны образовывать *префиксный* код. В кодовом дереве входного префиксного кода число листьев равно $M = K + q(K - 1)$, где q – число промежуточных узлов, не считая корня (ср. (3.25)). Число q – это наибольшее целое, для которого выполняется соотношение

$$M = K + q(K - 1) \leq D^N. \quad (4.6)$$

Критерий выбора данного префиксного кода должен быть обратным по отношению к рассмотренному ранее префиксному кодированию, т.е. средняя длина входного префиксного кода при заданной длине выходного блока должна быть *максимальна*.

Метод построения оптимального входного префиксного кода Танстолла состоит в следующем.

1. Определяем из (4.6) число промежуточных узлов q .
2. Строим первый ярус кодового дерева, выпуская из корня K ветвей с их концевыми узлами. Каждую ветвь снабжаем меткой, состоящей из буквы u_i и вероятности p_i появления этой буквы. Каждому узлу приписываем это значение вероятности.
3. В качестве первого промежуточного узла выбираем узел с *наибольшей* вероятностью. Выпускаем из него K ветвей с их концевыми узлами. Пересчитываем вероятности всех свободных узлов. Например, если узел второго яруса соответствует пути $u_1 u_3$, то его вероятность равна $p_1 p_3$.

4. В качестве второго промежуточного узла выбираем узел с *наибольшей* вероятностью. Выпускаем из него K ветвей с их концевыми узлами. Пересчитываем вероятности всех свободных узлов.
5. Продолжаем процедуру, пока все q промежуточных узлов не будут исчерпаны. После этого все оставшиеся свободные вершины обзываются листьями, соответствующими кодовым словам.

После построения дерева Танстолла кодирование заключается в приписывании каждому листу дерева выходного блока длины N . Порядок использования выходных блоков не имеет значения. Для определенности можно, например, упорядочить листья по возрастанию их вероятностей, а выходные блоки – в лексикографическом порядке. "Лишние" выходные блоки (если таковые окажутся) не используются.

Метод Танстолла приводит к префиксному коду с *максимальной* средней длиной \bar{L} . Критерием сжатия служит отношение \bar{L}/N . Можно показать, что для этой величины верны теоремы Шеннона в следующей формулировке:

$$\frac{\log_2 D}{H(U)} - \frac{\log_2(2/p_{min})}{NH(U)} \leq \frac{\bar{L}}{N} \leq \frac{\log_2 D}{H(U)}. \quad (4.7)$$

Приведем конкретный пример построения кода Танстолла для $U = \{u_1, u_2, u_3, u_4\}$, $p_1 = 0.1$, $p_2 = 0.1$, $p_3 = 0.3$, $p_4 = 0.6$, $D = 2$, $N = 4$. Число выходных блоков равно $2^N = 2^4 = 16$. Вычисляем $M = K + q(K-1) = 4 + 3q$ и находим целое значение $q = 4$, $M = 16$. Построим префиксный код.

Из корня дерева выходят $K = 4$ ветви. Далее из наиболее вероятного узла ($p_4 = 0.5$) делаем первое расширение. Затем из узла второго яруса u_4u_4 , имеющего наибольшую текущую вероятность $p_4p_4 = 0.25$, делаем второе расширение. Из узла первого яруса u_3 , имеющего наибольшую текущую вероятность $p_3 = 0.3$, делаем третье расширение. Наконец, на втором ярусе из узла u_3u_4 , имеющего наибольшую текущую вероятность

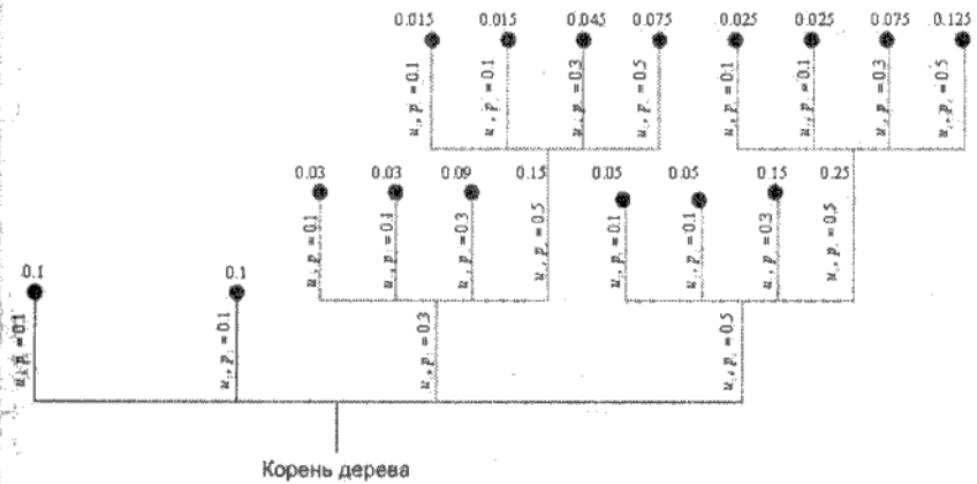


Рис. 4.1. Код Танстолла

$p_3p_4 = 0.15$, делаем четвертое расширение. Построение дерева Танстолла завершено.

Входной блок	Его вероятность	Выходной блок
$u_3u_4u_1$	0.015	0000
$u_3u_4u_2$	0.015	0001
$u_4u_4u_1$	0.025	0010
$u_4u_4u_2$	0.025	0011
u_3u_1	0.030	0100
u_3u_2	0.030	0101
$u_3u_4u_3$	0.045	0110
u_4u_1	0.050	0111
u_4u_2	0.050	1000
$u_4u_4u_3$	0.075	1001
$u_3u_4u_4$	0.075	1101
u_3u_3	0.090	1010
u_1	0.100	1011
u_2	0.100	1100
$u_4u_4u_4$	0.125	1110
u_4u_3	0.150	1111

Выше приведены входные блоки префиксного кода Танстолл-

ла, их вероятности и соответствующие выходные кодовые блоки.

В данном примере средняя длина входной последовательности равна $\bar{L} = 2.2$. Энтропия рассматриваемого источника равна $H(U) = 1.685$ бит. Выполняется неравенство Шеннона: $\frac{\bar{L}}{N} = 0.550 < \frac{\log_2 2}{H(U)} = 0.593$.

4.4. Универсальное кодирование

Пусть имеется источник информации с неизвестными вероятностями создания сообщений, т.е. источник с *неизвестной статистикой*. Большинство встречающихся на практике источников являются таковыми.

Сообщения источника обозначены $U_1 U_2 \dots U_n \dots$. Блокам сообщений источника, вообще говоря, различной длины $n_1, n_2 \dots$ поставим в соответствие кодовые слова $z_1, z_2 \dots$ одинаковой длины N : $|z_1| = N, |z_2| = N \dots$ Метод универсального кодирования разработан для целого класса источников. Основное его свойство таково: при увеличении длины блока отношение числа бит на выходе к числу бит на входе стремится к нормированной энтропии для любого источника этого класса, т.е.

$$\text{При } N \rightarrow \infty \left(\frac{\text{число бит на выходе}}{\text{число букв на входе}} \right) \rightarrow \frac{H(U)}{\log_2(D)} \quad (4.8)$$

Конкретные методы универсального кодирования: алгоритмы Лемпела–Зива, Лемпела–Зива–Велча, арифметическое кодирование (Рябко и др.).

4.4.1. Алгоритм Лемпела–Зива–Велча

Рассмотрим алгоритм Лемпела–Зива–Велча, для которого используем обозначение LZW . Предположим, что имеем исходный словарь из 256 слов. Пусть эти слова перенумерованы цифрами от 0 до 255. Каждому слову будет соответствовать один из байтов

$00000000 = 0, 00000001 = 1, \dots, 11111111 = 255.$

Обычно это таблица ASCII кодов. Предполагается, что этот словарь известен заранее как при сжатии (компрессии) исходного файла, так и при декомпрессии сжатого файла.

В процессе компрессии сжимаемый файл, состоящий из байт, преобразуется в другой файл, состоящий из символов большей длины. В стандартных программах длина выходных символов равна 12 битам. При этом символы-байты исходного словаря удлиняются до 12 добавлением 4-х нулей в старших разрядах. В процессе сжатия к исходному словарю добавляются новые слова, формируя динамический словарь. Первое добавленное слово получает номер 256, второе – 257, и т.д. Максимальное число слов в динамическом словаре с 12-битными символами равно 4096. Процедура сжатия продолжается до тех пор, пока либо файл не закончится, либо динамический словарь достигнет размера 4096. В последнем случае остаток входного файла начинает сжиматься заново, так что сжатый файл может состоять из нескольких независимых частей. Как именно производится компрессия в алгоритме Лемпела–Зива–Велча поясняется в следующей программе, записанной в псевдокодах. Длины входных и выходных символов в битах являются параметрами программы. В стандартных реализациях они равны 8 и 12.

Алгоритм сжатия LZW:

$w := \emptyset;$

while (there is input) {

$K :=$ next symbol from input;

 if (wK exists in the dictionary) {

$w := wK;$

 } else

 add wK to the dictionary;

 output (code for w);

$w := K;$

}

Переменная w имеет смысл текущего символа. Перед нача-

лом этой переменной присваивается значение "пустого" слова \emptyset . В процессе сжатия при извлечении из входного файла очередного символа K образуется конкатенация двух символов wK . Если такая комбинация символов уже есть в динамическом словаре, то переменной w присваивается значение wK , выходной символ не формируется, из входной последовательности извлекается следующий символ. В противном случае слово wK добавляется в динамический словарь, формируется выходной символ в виде номера текущего символа w , текущему символу w присваивается значение K , из входной последовательности извлекается следующий символ.

Каждая последовательность заканчивается словом EOF (end of file), что означает "конец файла извлечение которого означает конец процесса компрессии.

Для примера проведем сжатие последовательности $bbbaaacc$ для алгоритма с длиной входных символов 8, а выходных – 9. Все входные символы байты, которые содержатся в исходном словаре. Их удлиненные 9-битные версии обозначаются заключением в квадратные скобки ([b] и т.п.).

Составим таблицу

n	w	K	$output$	$index$	$word$
1	\emptyset	b	–	–	
2	b	b	[b]	256	[bb] = 256
3	b	b	–	–	–
4	[bb]	a	256	257	[bba] = 257
5	a	a	[a]	258	[aa] = 258
6	a	a	–	–	–
7	[aa]	c	258	–	[aac] = 259
8	c	c	[c]	–	[cc] = 260
9	c	c	–	–	–
10	[cc]	EOF	260	–	–

Последовательность выходных символов равна

[b] 256 [a] 258 [c] 260.

Так как входные символы – 8-битные слова, а выходные – 9-битные слова, то длина последовательностей в битах на входе равна $8 \times 9 = 72$, на выходе – $9 \times 6 = 54$. Коэффициент сжатия $54/72 = 0.75$.

В практических приложениях осуществляют сжатие файлов большой длины – больше 100 бит.

Осуществим декомпрессию для рассмотренного выше примера. При декомпрессии снова формируется динамический словарь, но из выходных символов, уже преобразованных в формат входных символов. Сжатый файл имеет вид $[b] 256 [a] 258 [c] 260$. При декомпрессии первый символ такого файла (в нашем случае $[b]$) всегда является удлинением символа из исходного словаря. Укорачивая его, получаем

<i>n</i>	<i>w</i>	<i>K</i>	<i>output</i>	<i>index</i>	<i>word</i>
1	\emptyset	<i>b</i>	–	–	

Следующим обрабатываемым символом является символ 256. Очевидно, он является конкатенацией входного символа *b* и одного из входных символов *K*. Но этим символом может быть только символ *b*. В противном случае для любого другого входного символа алгоритм сжатия выдал бы в качестве второго выходного символа символ $[b]$. Итак, имеем

<i>n</i>	<i>w</i>	<i>K</i>	<i>output</i>	<i>index</i>	<i>word</i>
1	\emptyset	<i>b</i>	–	–	
2	<i>b</i>	<i>b</i>	$[b]$	256	$[bb] = 256$
3	<i>b</i>	<i>b</i>	–	–	–

Третий выходной символ равен $[a]$. Укорачивая его, имеем

<i>n</i>	<i>w</i>	<i>K</i>	<i>output</i>	<i>index</i>	<i>word</i>
1	\emptyset	<i>b</i>	–	–	
2	<i>b</i>	<i>b</i>	$[b]$	256	$[bb] = 256$
3	<i>b</i>	<i>b</i>	–	–	–
4	$[bb]$	<i>a</i>	256	257	$[bba] = 257$
5	<i>a</i>	?	?	?	?

Четвертый выходной символ равен 258. Этот символ является конкатенацией двух символов, первый из которых должен быть a . Как и ранее, устанавливаем, что вторым символом может быть только a . Имеем

n	w	K	$output$	$index$	$word$
1	\emptyset	b	—	—	
2	b	b	[b]	256	[bb] = 256
3	b	b	—	—	—
4	[bb]	a	256	257	[bba] = 257
5	a	a	[a]	258	[aa] = 258
6	a	a	—	—	—
7	[aa]	c	258	—	[aac] = 259
8	c	?	?	—	?

Последним обрабатываемым символом является символ 260. Как и ранее, устанавливаем, что $260 = [cc]$. Окончательно имеем

n	w	K	$output$	$index$	$word$
1	\emptyset	b	—	—	
2	b	b	[b]	256	[bb] = 256
3	b	b	—	—	—
4	[bb]	a	256	257	[bba] = 257
5	a	a	[a]	258	[aa] = 258
6	a	a	—	—	—
7	[aa]	c	258	—	[aac] = 259
8	c	c	[c]	—	[cc] = 260
9	c	c	—	—	—
10	[cc]	EOF	260	—	—

Третий столбец содержит символы исходного файла.

Для сжатия изображений применяется программа GIF, в модемах программы V.42 и Compress (Unix).

4.4.2. Алгоритм Лемпела–Зива

В алгоритме Лемпела–Зива (LZ77) используется идея "скользящего окна". Скользящее окно – это буфер заранее заданной

длины, разделенный на две части. На рис. 4.2 показан пример скользящего окна. В первой части буфера, называемого поисковым буфером (Search buffer), на длине L_s размещены ранее обработанные символы входной (сжимаемой) последовательности.

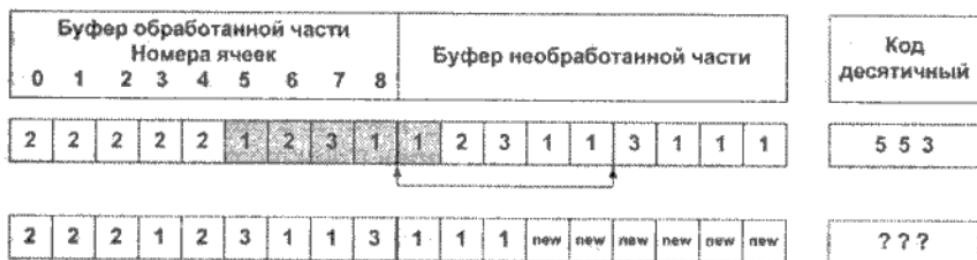


Рис. 4.2. Скользящее окно в алгоритме LZ77

ковым буфером (Search buffer), на длине L_s размещены ранее обработанные символы входной (сжимаемой) последовательности.

В нашем примере там размещены десятичные символы

2, 2, 2, 2, 1, 2, 3, 1. Ячейки буфера обработанной части нумеруются целыми числами от 0 до $L_s - 1$. В другой части буфера, называемого буфером необработанной части (Lookahead buffer), на длине L_l размещены очередные необработанные символы входной последовательности.

В нашем примере – это символы 1, 2, 3, 1, 1, 3, 1, 1, 1.

Далее в обработанной части отыскивается отрезок наибольшей длины, совпадающий с префиксом необработанной части (разрешается, чтобы "хвост" этого отрезка находился в буфере необработанной части) и формируется кодовое слово, соответствующее этому отрезку.

В нашем примере таким отрезком является затемненный сегмент 1, 2, 3, 1, 1. Соответствующий ему префикс в необработанной части отмечен стрелками снизу.

Кодовое слово состоит из трех частей: 1) номер ячейки буфера обработанной части, в которой начинается найденный отрезок; 2) длина найденного отрезка; 3) символ в необработанной части, следующий непосредственно за префиксом. Если символы представлены в D -ичном алфавите, а ячейки буфера также хранят D -ичные символы, то длина кодового символа равна

$$N = \log_D L_s + \log_D L_l + 1.$$

В нашем примере найденный отрезок начинается в ячейке 5 (нумерация с 0), его длина равна 5, а символом, следующим за префиксом, является символ 3. Следовательно, кодовое слово имеет вид $C = 5\ 5\ 3$.

После формирования кодового слова производится сдвиг символов в сторону буфера обработанной части на число позиций, равное длине обработанного префикса плюс 1. В буфере необработанной части освободившиеся ячейки заполняются очередными входными символами.

В нашем примере производится сдвиг на 6 позиций. Словом "new" обозначены очередные символы входной последовательности, заполнившие освободившиеся места. Для формирования очередного кодового слова в буфере обработанной части хранятся символы 2, 2, 2, 1, 2, 3, 1, 1, 3, а в буфере необработанной части – 1, 1, 1, new, new, new, new, new.

Возможна ситуация, когда в обработанной части нет отрезка, совпадающего с любым префиксом в необработанной части. Другими словами, длина такой части равна 0. В этом случае формируется кодовое слово

(0 0, *Первый символ необработанной части*).

После этого производится сдвиг в буфере на 1 позицию.

Перед началом работы буфер обработанных символов заполняется нулями, а в буфер необработанных символов вводятся первые символы сжимаемой последовательности.

Ниже представлен упрощенный алгоритм компрессии (сжатия) LZ77.

```
Ll := length(LookAheadBuffer)
while (LookAheadBuffer not empty) {
    get a reference (position, length) to longest match;
    if (Ll > length > 0) {
        output (position, length, next symbol);
        shift the window length+1 positions along;
    if (length = Ll) {
```

```

length :=  $L_l - 1$ ; output (position, length, next symbol);
shift the window length+1 positions along;
}
} else {
    output (0, 0, first symbol in the lookahead buffer);
    shift the window 1 character along;
}

```

Пример 4.1. Рассмотрим пример: пусть последовательность для сжатия представлена в троичном алфавите:

$$S = 001010210210212021021200.$$

На рис. 4.3 показано, как производится компрессия этой последовательности.

Длина буфера обрабатываемой части (Lookahead buffer) = 9
Длина буфера обработанной части (Search buffer) = 9
Вход S = 001010210210212021021200 ...
Выход T = 2202121102202202120220 ...

	Search buffer Positions	Lookahead buffer	Код десктичный	Код троичный
1.	0 1 2 3 4 5 6 7 8	0 0 0 0 0 0 0 0 1 0 1 0 2 1 0 2 1 0 2 1 1 ...	821	22 02 1
2.	0 0 0 0 0 0 0 0 1 0 1 0 2 1 0 2 1 0 2 1 0 2 1 ...	1 0 1 0 2 1 0 2 1 0 2 1 0 2 1 0 2 1 0 2 1 ...	732	21 10 2
3.	0 0 0 0 1 0 1 0 2 1 0 2 1 0 2 1 0 2 1 2 0 2 1 ...	0 1 0 2 1 0 2 1 0 2 1 0 2 1 2 0 2 1 2 0 2 1 ...	672	20 21 2
4.	2 1 0 2 1 0 2 1 2 0 2 1 0 2 1 0 2 1 2 0 0 ...	1 0 2 1 0 2 1 0 2 1 0 2 1 2 0 0 ...	280	02 22 0

Рис. 4.3. Пример компрессии в алгоритме LZ77

Существуют усовершенствования этого алгоритма. Имеются алгоритмы Лемпела–Зива–Хаффмена (LZH), zip, gzip, stacker.

Г л а в а 5

Типические последовательности

5.1. Эпсилон-тиปические последовательности

Одним из важных понятий теории информации является понятие "типической" последовательности. Это понятие может вводиться различными способами, но во всех случаях с его помощью удается описывать удобные с теоретической точки зрения множества "высоко вероятных асимптотически равномерно распределенных" последовательностей. Такие множества полезны в теории кодирования с потерями и в теории кодирования для каналов с шумами.

Пусть $\mathbf{X} = (X_1, X_2, \dots, X_L)$ – случайный вектор длины L , компонентами которого являются независимые одинаково распределенные случайные величины с распределением $P_X(x)$, принимающие значения из алфавита $\mathcal{U} = \{u_1, u_2, \dots, u_K\}$. Этот вектор можно интерпретировать как последовательность случайных величин, порожденную стационарным источником без памяти. Распределение случайного вектора имеет вид

$$P_{\mathbf{X}}(\mathbf{x}) = P_{\mathbf{X}}(x_1, x_2, \dots, x_L) = P_X(x_1)P_X(x_2)\dots P_X(x_L). \quad (5.1)$$

Введем обозначение $p_s = P_X(u_s)$, $s = 1, 2, \dots, K$.

Пусть $\mathbf{x} = (x_1, x_2, \dots, x_L)$ – значение случайного вектора \mathbf{X} . Обозначим через $n_s(\mathbf{x})$, $s = 1, 2, \dots, K$, число координат вектора \mathbf{x} , равных u_s . Ясно, что $\sum_{s=1}^K n_s(\mathbf{x}) = L$. Выберем число $\varepsilon > 0$.

Последовательность (x_1, x_2, \dots, x_L) (эквивалентно, вектор \mathbf{x}) называется ε -типической для распределения $P_X(x)$, если одновременно выполняются соотношения

$$p_s(1 - \varepsilon) < \frac{n_s(\mathbf{x})}{L} < p_s(1 + \varepsilon), \quad s = 1, 2, \dots, K. \quad (5.2)$$

5.2. Вероятность ε -типической последовательности

Вероятность вектора \mathbf{x} равна $P_{\mathbf{X}}(\mathbf{x})$. Оценим вероятность появления ε -типической последовательности (вектора). Пусть $H(X) = -\sum_{s=1}^K p_s \log_2 p_s$ – энтропия случайной величины X .

Лемма 5.1. Пусть $\mathbf{x} = (x_1, x_2, \dots, x_L)$ – ε -типическая последовательность (вектор) для распределения $P_X(x)$. Тогда вероятность $P_{\mathbf{X}}(\mathbf{x} | \varepsilon\text{-тип})$ появления такой последовательности удовлетворяет неравенствам

$$2^{-L(1+\varepsilon)H(X)} \leq P_{\mathbf{X}}(\mathbf{x} | \varepsilon\text{-тип}) \leq 2^{-L(1-\varepsilon)H(X)}. \quad (5.3)$$

Доказательство. Из (5.1) имеем

$$\begin{aligned} P_{\mathbf{X}}(\mathbf{x} | \varepsilon\text{-тип}) &= p_1^{n_1(\mathbf{x})} p_2^{n_2(\mathbf{x})} \cdots p_K^{n_K(\mathbf{x})} = \\ &= 2^{n_1(\mathbf{x}) \log_2 p_1 + n_2(\mathbf{x}) \log_2 p_2 + \cdots + n_K(\mathbf{x}) \log_2 p_K}. \end{aligned} \quad (5.4)$$

Для оценки этой вероятности сверху подставим вместо $n_s(\mathbf{x})$ их оценки $L(1 - \varepsilon)p_s \leq n_s(\mathbf{x})$ из левых неравенств (5.2):

$$\begin{aligned} 2^{n_1(\mathbf{x}) \log_2 p_1 + n_2(\mathbf{x}) \log_2 p_2 + \cdots + n_K(\mathbf{x}) \log_2 p_K} &\leq \\ &\leq 2^{L(1-\varepsilon) \sum_{s=1}^K p_s \log_2 p_s} = 2^{-L(1-\varepsilon)H(X)}. \end{aligned} \quad (5.5)$$

Аналогично, для оценки снизу подставим вместо $n_s(\mathbf{x})$ их оценки $L(1 + \varepsilon)p_s \geq n_s(\mathbf{x})$ из правых неравенств (5.2):

$$\begin{aligned} 2^{n_1(\mathbf{x}) \log_2 p_1 + n_2(\mathbf{x}) \log_2 p_2 + \cdots + n_K(\mathbf{x}) \log_2 p_K} &\geq \\ &\geq 2^{L(1+\varepsilon) \sum_{s=1}^K p_s \log_2 p_s} = 2^{-L(1+\varepsilon)H(X)}. \end{aligned} \quad (5.6)$$

Объединяя эти оценки, получим (5.3).

Соотношения (5.3) показывают, что при больших L и малых ε вероятности ε -типических последовательностей одинаковы и приближенно равны $2^{-LH(X)}$.

5.3. Вероятность множества ε -типических последовательностей

Обозначим через

$$\mathcal{M}_\varepsilon(P_X) = \{\mathbf{x} : \text{вектор } \mathbf{x} \text{ } \varepsilon\text{-типический для } P_X(x)\} \quad (5.7)$$

множество векторов \mathbf{x} , ε -типических для распределения $P_X(x)$, а через $M_\varepsilon(P_X) = |\mathcal{M}_\varepsilon(P_X)|$ – их число.

Обозначим через

$$\mathcal{F}_\varepsilon = \{\mathbf{x} \in \mathcal{M}_\varepsilon(P_X)\} \quad (5.8)$$

событие, состоящее в том, что источник породит ε -типический для распределения $P_X(x)$ вектор \mathbf{x} .

Пусть $\overline{\mathcal{F}}_\varepsilon$ – противоположное событие.

Оценим вероятности $\Pr(\overline{\mathcal{F}}_\varepsilon)$ и $\Pr(\mathcal{F}_\varepsilon) = 1 - \Pr(\overline{\mathcal{F}}_\varepsilon)$ этих событий.

Лемма 5.2. Для любого $\varepsilon > 0$ при $L \rightarrow \infty$ имеем

$$\Pr(\overline{\mathcal{F}}_\varepsilon) \rightarrow 0, \quad \Pr(\mathcal{F}_\varepsilon) \rightarrow 1.$$

Доказательство. Для $i = 1, 2, \dots, K$ определим событие \mathcal{B}_i как

$$\mathcal{B}_i = \left\{ \mathbf{x} : \left| \frac{n_i(\mathbf{x})}{L} - p_i \right| \geq \varepsilon p_i \right\}. \quad (5.9)$$

Тогда событие $\overline{\mathcal{F}}_\varepsilon$ равно

$$\overline{\mathcal{F}}_\varepsilon = \bigcup_{i=1}^K \mathcal{B}_i.$$

Действительно, вектор \mathbf{x} не будет ε -типическим для распределения $P_X(x)$ тогда и только тогда, когда реализуется хотя бы одно из событий \mathcal{B}_i , так как в этом случае хотя бы для одного i не выполняется условие ε -типичности. Вероятность этого события равна

$$\Pr(\overline{\mathcal{F}_\varepsilon}) = \Pr\left(\bigcup_{i=1}^K \mathcal{B}_i\right) \leq \sum_{i=1}^K \Pr(\mathcal{B}_i). \quad (5.10)$$

Здесь при получении верхней оценки использовано общее неравенство, гласящее, что вероятность объединения событий не превосходит суммы вероятностей этих событий. Для оценки вероятности события \mathcal{B}_i применим неравенство Чебышева:

$$\begin{aligned} \Pr(\mathcal{B}_i) &= \Pr\left(\left|\frac{n_i(\mathbf{X})}{L} - p_i\right| \geq \varepsilon p_i\right) \leq \frac{E_{\mathbf{X}}\left(\frac{n_i(\mathbf{X})}{L} - p_i\right)^2}{\varepsilon^2 p_i^2} = \\ &= \frac{p_i(1-p_i)}{L\varepsilon^2 p_i^2} = \frac{(1-p_i)}{L\varepsilon^2 p_i} < \frac{1}{L\varepsilon^2 p_{\min}}, \end{aligned} \quad (5.11)$$

где $p_{\min} = \min_i p_i$. Здесь

$$E_{\mathbf{X}}\left(\frac{n_i(\mathbf{X})}{L} - p_i\right)^2 = \frac{p_i(1-p_i)}{L}$$

— дисперсия случайной величины $n_i(\mathbf{X})/L$, нормированного числа появлений символа u_i среди координат вектора \mathbf{X} .

Более точная оценка следует из теории больших уклонений. Обозначим через $\theta(p)$ выражение

$$\theta(p) = 2^{-[h(p) - h(p+p\varepsilon) + p\varepsilon \log_2(\frac{1-p}{p})]}, \quad (5.12)$$

где $h(p) = -p \log_2 p - (1-p) \log_2(1-p)$ — функция двоичной энтропии. Можно показать, что при $0 < p < 1/(1+\varepsilon)$ выполняется неравенство $0 < \theta(p) < 1$. Тогда

$$P\left(\left|\frac{n_i(\mathbf{X})}{L} - p_i\right| \geq \varepsilon p_i\right) \leq \frac{2}{\sqrt{L}} \theta(p_i)^L \leq \frac{2}{\sqrt{L}} \theta_{\max}^L, \quad (5.13)$$

где $\theta_{\max} = \theta(p_{\min})$.

Подставляя (5.11) и (5.12) в (5.10), получим

$$\begin{aligned}\Pr(\bar{\mathcal{F}}_\varepsilon) &\leq \begin{cases} \frac{K}{L\varepsilon^2 p_{\min}}, \\ \frac{2K}{\sqrt{L}} \theta_{\max}^L, \end{cases} \rightarrow 0, \quad L \rightarrow \infty \\ \Pr(\mathcal{F}_\varepsilon) &= 1 - \Pr(\bar{\mathcal{F}}_\varepsilon) \rightarrow 1, \quad L \rightarrow \infty.\end{aligned}\tag{5.14}$$

5.4. Число ε -типовических последовательностей

Точное число таких последовательностей определяется суммой полиномиальных коэффициентов

$$M_\varepsilon = \sum_{n_1 + \dots + n_K = L, \text{ } \varepsilon\text{-тип}} \frac{L!}{n_1! n_2! \dots n_K!}, \tag{5.15}$$

где сумма берется по всем значениям n_s ($s = 1, 2, \dots, K$), удовлетворяющим определению ε -типовичности в соотношении (5.2). Однако точное вычисление этой суммы затруднительно, поэтому перейдем к получению верхних и нижних асимптотических оценок.

Лемма 5.3. Число ε -типовических для распределения $P_X(x)$ последовательностей удовлетворяет неравенствам

$$(1 - \Pr(\bar{\mathcal{F}}_\varepsilon)) 2^{(1-\varepsilon)LH(U)} \leq M_\varepsilon(P_X) \leq 2^{(1+\varepsilon)LH(U)}. \tag{5.16}$$

Это означает, что число эпсилон-типовических для распределения $P_X(x)$ векторов при больших L и малых ε равно примерно $2^{LH(U)}$.

Доказательство. Для оценки числа последовательностей сверху заметим, что

$$\Pr(\mathcal{F}_\varepsilon) = \sum_{\mathbf{x} | \varepsilon\text{-тип}} P_X(\mathbf{x} | \varepsilon\text{-тип}) \geq M_\varepsilon(P_X) 2^{-L(1+\varepsilon)H(X)}.$$

Здесь для оценки снизу каждого слагаемого использовано левое неравенство из соотношения (5.3). Отсюда следует, что

$$M_\varepsilon(P_X) \leq 2^{L(1+\varepsilon)H(X)} \Pr(\mathcal{F}_\varepsilon) \leq 2^{L(1+\varepsilon)H(X)}.$$

Для оценки числа последовательностей снизу заметим, что

$$\Pr(\mathcal{F}_\varepsilon) = \sum_{\mathbf{x}|\varepsilon\text{-тип}} P_X(\mathbf{x} | \varepsilon\text{-тип}) \leq M_\varepsilon(P_X) 2^{-L(1-\varepsilon)H(X)}.$$

Здесь для оценки сверху каждого слагаемого использовано *правое* неравенство из соотношения (5.3). Отсюда следует, что

$$M_\varepsilon(P_X) \geq 2^{L(1-\varepsilon)H(X)} \Pr(\mathcal{F}_\varepsilon) = (1 - \Pr(\bar{\mathcal{F}}_\varepsilon)) 2^{L(1-\varepsilon)H(X)}.$$

Для оценки $\Pr(\bar{\mathcal{F}}_\varepsilon)$ можно использовать одну из оценок в соотношении (5.14).

5.5. Вероятность множества $\mathcal{M}_\varepsilon(P_X)$ при других распределениях

Предположим, что для распределения $P_X(x)$ найдено множество ε -типовических векторов $\mathcal{M}_\varepsilon(P_X)$. Рассмотрим теперь другое распределение $Q_X(x)$, с вероятностями $Q_X(u_s) = r_s$, (где $s = 1, 2, \dots, K$), отличающееся от $P_X(x)$. Оценим сверху вероятность того, что случайный вектор $\mathbf{X} = (X_1, X_2, \dots, X_L)$, компонентами которого являются независимые одинаково распределенные случайные величины с распределением $Q_X(x)$, примет значение, попадающее в множество $\mathcal{M}_\varepsilon(P_X)$. Обозначим эту вероятность через $\Pr(\mathbf{x} \in \mathcal{M}_\varepsilon(P_X) | Q_X)$. По определению, эта вероятность равна

$$\Pr(\mathbf{x} \in \mathcal{M}_\varepsilon(P_X) | Q_X) = \sum_{\mathbf{x} \in \mathcal{M}_\varepsilon(P_X)} r_1^{n_1(\mathbf{x})} r_2^{n_2(\mathbf{x})} \dots r_K^{n_K(\mathbf{x})}. \quad (5.17)$$

Информационная дивергенция Кульбака–Лейблера $D(P_X \| Q_X)$ между распределениями $P_X(x)$ и $Q_X(x)$ равна

$$D(P_X | Q_X) = \sum_{i=1}^K p_i \log_2 \frac{p_i}{q_i}. \quad (5.18)$$

Как известно, эта величина строго положительна при несовпадающих $P_X(x)$ и $Q_X(x)$.

Лемма 5.4. Для достаточно больших L и малых $\varepsilon > 0$

$$\Pr(\mathbf{x} \in \mathcal{M}_\varepsilon(P_X | Q_X)) \leq 2^{-L(1-\varepsilon') D(P_X \| Q_X)}, \quad (5.19)$$

где $\varepsilon' = \varepsilon(1 + \frac{2D(P_X | Q_X)}{H(X)})$.

Доказательство. С учетом ограничений (5.2) и неравенств (5.16) получаем из (5.17) оценку сверху:

$$\begin{aligned} \Pr(\mathbf{x} \in \mathcal{M}_\varepsilon(P_X) | Q_X(x)) &\leq M_\varepsilon(P_X) 2^{L(1-\varepsilon) \sum_{i=1}^K p_i \log_2 r_i} \leq \\ &\leq 2^{L(1+\varepsilon)H(X)+L(1-\varepsilon) \sum_{i=1}^K p_i \log_2 r_i} = \\ &= 2^{-L(1-\varepsilon) \sum_{i=1}^K p_i \log_2 \frac{p_i}{r_i}} 2^{2L\varepsilon H(X)} = \\ &= 2^{-L(1-\varepsilon) D(P_X \| Q_X)} 2^{2L\varepsilon H(X)} = 2^{-L(1-\varepsilon') D(P_X \| Q_X)}. \end{aligned}$$

Отсюда следует, что для достаточно малых ε эта вероятность стремится к нулю при $L \rightarrow \infty$. Таким образом, при "неправильном" распределении $Q_X(x)$ вероятность породить последовательность из множества $\mathcal{M}_\varepsilon(P_X)$ пренебрежимо мала, тогда как для "правильного" распределения $P_X(x)$ аналогичная вероятность близка к 1 (см. лемма 5.2, (5.14)).

5.6. Двумерные типические последовательности

До сих пор рассматривались последовательности одномерных случайных величин. Однако все результаты легко обобщаются и на случай многомерных случайных величин. Приведем

такое обобщение для двумерных случайных величин. Пусть совместное распределение пары случайных величин $\{X, Y\}$ описывается функцией $P_{\{X,Y\}}(\{x, y\})$. Величина X принимает значения из алфавита $\mathcal{U} = \{u_1, u_2, \dots, u_K\}$, величина Y принимает значения из алфавита $\mathcal{V} = \{v_1, v_2, \dots, v_D\}$, так что возможные значения, которые может принимать пара, равны

$$\{x, y\} = \{u_i, v_j\}, i = 1, 2, \dots, K; j = 1, 2, \dots, D.$$

Представим пару случайных векторов $\mathbf{X} = (X_1, X_2, \dots, X_L)$ и $\mathbf{Y} = (Y_1, Y_2, \dots, Y_L)$ длины L как случайную последовательность из L пар $(\{\mathbf{X}, \mathbf{Y}\}) = (\{X_1, Y_1\}, \{X_2, Y_2\}, \dots, \{X_L, Y_L\})$ с совместным распределением

$$P_{\{\mathbf{X}, \mathbf{Y}\}}(\{\mathbf{x}, \mathbf{y}\}) = P_{\{X, Y\}}(\{x_1, y_1\}) \dots P_{\{X, Y\}}(\{x_L, y_L\}). \quad (5.20)$$

Это означает, что при $s \neq m$ пары случайных величин $\{x_m, y_m\}$ и $\{x_s, y_s\}$ независимы. Случайные величины внутри пары в общем случае зависимы. Для конкретной пары векторов $(\{\mathbf{x}, \mathbf{y}\})$ обозначим через $n_{ij}(\{\mathbf{x}, \mathbf{y}\})$ число пар вида $\{u_i, v_j\}$,

$(i = 1, 2, \dots, K; j = 1, 2, \dots, D)$ в этой последовательности.

Очевидно, что $\sum_{j=1}^D \sum_{i=1}^K n_{ij}(\{\mathbf{x}, \mathbf{y}\}) = L$.

Последовательность пар $(\{\mathbf{x}, \mathbf{y}\}) = (\{x_1, y_1\}, \dots, \{x_L, y_L\})$ называется ε -типической для распределения $P_{\{X, Y\}}(\{x, y\})$, если одновременно выполняются следующие соотношения:

$$p_{ij}(1-\varepsilon) < \frac{n_{ij}(\{\mathbf{x}, \mathbf{y}\})}{L} < p_{ij}(1+\varepsilon), \quad i = 1, 2, \dots, K; j = 1, 2, \dots, D, \quad (5.21)$$

где для краткости обозначено $p_{ij} = P_{\{X, Y\}}(\{u_i, v_j\})$.

Соответствующая пара векторов $(\{\mathbf{x}, \mathbf{y}\})$ также называется совместно ε -типической для распределения $P_{\{X, Y\}}(\{x, y\})$.

Маргинальные (одномерные) распределения вероятностей $p_i = P_X(u_i)$ и $q_j = P_Y(v_j)$ получаются из совместного двумерного распределения p_{ij} : $p_i = \sum_{j=1}^D p_{ij}$, $q_j = \sum_{i=1}^K p_{ij}$.

Число символов u_i в последовательности \mathbf{x} равно

$\sum_{j=1}^D n_{ij}(\{\mathbf{x}, \mathbf{y}\}) = n_i(\mathbf{x})$. Число символов v_j в последовательности \mathbf{y} равно $\sum_{i=1}^K n_{ij}(\{\mathbf{x}, \mathbf{y}\}) = m_j(\mathbf{y})$.

Суммирование в соотношениях (5.21) по i или по j дает следующие неравенства:

$$\begin{aligned} p_i(1 - \varepsilon) &< \frac{n_i(\mathbf{x})}{L} < p_i(1 + \varepsilon), \\ q_j(1 - \varepsilon) &< \frac{m_j(\mathbf{y})}{L} < q_j(1 + \varepsilon). \end{aligned} \quad (5.22)$$

Соотношения (5.22) показывают, что если векторы $(\{\mathbf{x}, \mathbf{y}\})$ являются совместно ε -типическими для распределения $P_{XY}(\{x, y\})$, то оба вектора \mathbf{x} и \mathbf{y} являются ε -типическими для своих маргинальных распределений $P_X(x)$ и $P_Y(y)$. Обратное утверждение не верно.

Пусть $H(XY) = -\sum_{i=1}^K \sum_{j=1}^D p_{ij} \log_2 p_{ij}$ – энтропия случайной пары величины $\{X, Y\}$.

Обозначим через $P_{\{\mathbf{X}, \mathbf{Y}\}}(\{\mathbf{x}, \mathbf{y}\} | \varepsilon\text{-тип})$ вероятность совместно ε -типической для распределения вероятностей $P_{\{X, Y\}}(\{x, y\})$ пары векторов $\{\mathbf{x}, \mathbf{y}\}$.

Обозначим через $\mathcal{M}_\varepsilon(P_{\{X, Y\}})$ множество пар векторов $(\{\mathbf{x}, \mathbf{y}\})$ длины L , совместно ε -типических для распределения вероятностей $P_{\{X, Y\}}(x, y)$, а через $M_\varepsilon(P_{\{X, Y\}}) = |\mathcal{M}_\varepsilon(P_{\{X, Y\}})|$ – их число.

Лемма 5.5.

$$2^{-L(1+\varepsilon)H(XY)} \leq P_{\{\mathbf{X}, \mathbf{Y}\}}(\{\mathbf{x}, \mathbf{y}\} | \varepsilon\text{-тип}) \leq 2^{-L(1-\varepsilon)H(XY)} \quad (5.23)$$

Доказательство. Дословное повторение доказательства Леммы 5.1 с заменой вероятностей p_i на p_{ij} .

Соотношения (5.23) показывают, что при больших L и малых ε вероятности ε -типических последовательностей равны, грубо говоря, $2^{-LH(XY)}$.

Обозначим через

$$\mathcal{F}_{2D, \varepsilon} = \{ \{\mathbf{x}, \mathbf{y}\} \in \mathcal{M}_\varepsilon(P_{\{X, Y\}}) \} \quad (5.24)$$

событие, состоящее в том, что случайный вектор $\{\mathbf{x}, \mathbf{y}\}$ с распределением $P_{\{\mathbf{X}, \mathbf{Y}\}}(\{\mathbf{x}, \mathbf{y}\})$ примет значение в множестве $\mathcal{M}_\varepsilon(P_{\{X, Y\}})$.

Лемма 5.6. Для любого $\varepsilon > 0$ при $L \rightarrow \infty$ имеем

$$\Pr(\bar{\mathcal{F}}_{2D, \varepsilon}) \leq \begin{cases} \frac{K}{L\varepsilon^2 p_{\min}}, \\ \frac{2K}{\sqrt{L}} \theta_{\max}^L, \end{cases} \rightarrow 0, \quad L \rightarrow \infty \quad (5.25)$$

$$\Pr(\mathcal{F}_{2D, \varepsilon}) = 1 - \Pr(\bar{\mathcal{F}}_{2D, \varepsilon}) \rightarrow 1, \quad L \rightarrow \infty,$$

где $p_{\min} = \min_{i,j} p_{ij}$, $\theta_{\max} = \theta(p_{\min})$, а $\theta(p)$ определяется соотношением (5.12).

Доказательство. Дословное повторение доказательства леммы 5.2 с заменой вероятностей p_i на p_{ij} .

Лемма 5.7. Число ε -типовых последовательностей пар для распределения $P_{\{X,Y\}}(\{x,y\})$ удовлетворяет неравенствам

$$(1 - \Pr(\bar{\mathcal{F}}_{2D, \varepsilon}))2^{(1-\varepsilon)LH(XY)} \leq M_\varepsilon(P_X) \leq 2^{(1+\varepsilon)LH(XY)}. \quad (5.26)$$

Доказательство. Дословное повторение доказательства леммы 5.3 с заменой вероятностей p_i на p_{ij} и $\Pr(\bar{\mathcal{F}}_\varepsilon)$ на $\Pr(\bar{\mathcal{F}}_{2D, \varepsilon})$.

Пусть для распределения $P_{\{X,Y\}}(\{x,y\})$ найдено множество ε -типовых пар векторов $\mathcal{M}_\varepsilon(P_{\{X,Y\}})$. Рассмотрим теперь другое распределение $Q_{\{X,Y\}}(\{x,y\})$, отличающееся от $P_X(x)$:

$Q_{\{X,Y\}}(\{u_i, v_j\}) = r_{ij}$, $i = 1, 2, \dots, K$; $j = 1, 2, \dots, D$. Оценим сверху вероятность того, что случайный вектор пар

$$\{\mathbf{X}, \mathbf{Y}\} = (\{X_1, Y_1\}, \{X_2, Y_2\}, \dots, \{X_L, Y_L\}),$$

компонентами которого являются независимые одинаково распределенные случайные пары с распределением $Q_{\{X,Y\}}(\{x,y\})$, примет значение, попадающее в множество $\mathcal{M}_\varepsilon(P_{\{X,Y\}})$. Обозначим $\Pr(\{\mathbf{x}, \mathbf{y}\} \in \mathcal{M}_\varepsilon(P_{\{X,Y\}}) \mid Q_{\{X,Y\}}(\{x,y\}))$ эту вероятность. По определению,

$$\Pr(\{\mathbf{x}, \mathbf{y}\} \in \mathcal{M}_\varepsilon(P_{\{X,Y\}}) \mid Q_{\{X,Y\}}) = \\ = \sum_{\{\mathbf{x}, \mathbf{y}\} \in \mathcal{M}_\varepsilon(P_{\{X,Y\}})} \prod_{i=1}^K \prod_{j=1}^D r_{ij}^{n_{ij}(\{\mathbf{x}, \mathbf{y}\})}. \quad (5.27)$$

Информационная дивергенция $D(P_{\{XY\}} \| Q_{\{X,Y\}})$ между распределениями $P_{\{XY\}}(\{x, y\})$ и $Q_{\{XY\}}(\{x, y\})$ равна

$$D(P_{\{XY\}} \| Q_{\{X,Y\}}) = \sum_{i=1}^K \sum_{j=1}^D p_{ij} \log_2 \frac{p_{ij}}{q_{ij}} \quad (5.28)$$

и что эта величина строго положительна при несовпадающих $P_{\{X,Y\}}(\{x, y\})$ и $Q_{\{X,Y\}}(\{x, y\})$.

Лемма 5.8. Для достаточно больших L и малых $\varepsilon > 0$

$$\Pr(\{\mathbf{x}, \mathbf{y}\} \in \mathcal{M}_\varepsilon(P_{\{X,Y\}}) \mid Q_{\{X,Y\}}) \leq 2^{-L(1-\varepsilon') D(P_{\{XY\}} \| Q_{\{X,Y\}})}, \quad (5.29)$$

где $\varepsilon' = \varepsilon(1 + 2D(P_{\{XY\}} \| Q_{\{X,Y\}} \| Q_X)/H(XY))$.

Доказательство. Дословное повторение доказательства леммы 5.4 с заменой вероятностей p_i на p_{ij} и вероятностей r_i на r_{ij} .

Таким образом, при "неправильном" распределении $Q_{\{X,Y\}}$ последовательность из множества $\mathcal{M}_\varepsilon(P_{\{X,Y\}})$ крайне мало вероятна, тогда как для "правильного" распределения $P_X(x)$ вероятность близка к 1 (см. лемму 5.6, (5.25)).

Интересен случай, когда распределения $P_{\{XY\}}$ и $Q_{\{XY\}}$ имеют одинаковые маргинальные распределения $P_X(x)$ и $P_Y(y)$, причем для распределения $Q_{\{XY\}}$ случайные величины внутри пары независимы. Тогда дивергенция равна средней взаимной информации между X и Y :

$$D(P_{\{XY\}} \| Q_{\{X,Y\}}) = I(Y; X) = H(X) + H(Y) - H(XY) = \\ = H(Y) - H(Y | X). \quad (5.30)$$

Соотношение (5.29) приобретает вид:

$$\Pr(\{\mathbf{x}, \mathbf{y}\} \in \mathcal{M}_\varepsilon(P_{\{X,Y\}}) \mid Q_{\{X,Y\}}) \leq 2^{-L(1-\varepsilon')I(Y;X)}. \quad (5.31)$$

Г л а в а 6

Кодирование с потерями

6.1. Неоднозначность декодирования

Рассмотрим кодирование входных блоков постоянной длины L в выходные блоки постоянной длины N . Пусть алфавит источника состоит из букв $\mathcal{U} = \{u_1, u_2, \dots, u_K\}$, а алфавит кодера — из букв $\mathcal{V} = \{v_1, v_2, \dots, v_D\}$. Число возможных входных блоков длины L равно K^L . Число возможных выходных блоков длины N равно D^N . Если основное требование — однозначность декодирования, то условие однозначности следующее:

$$D^{N-1} < K^L \leq D^N,$$

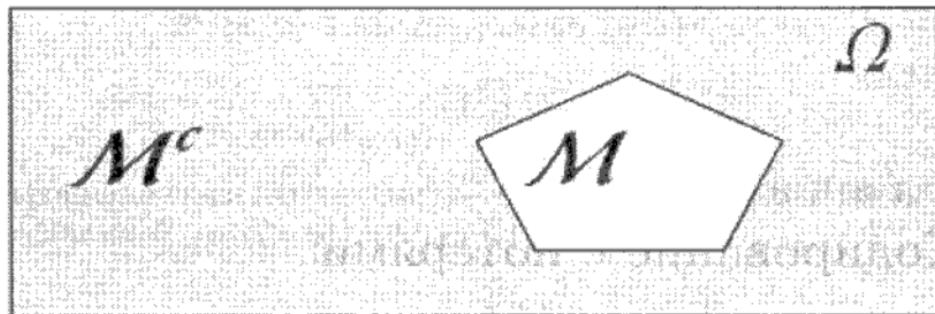
или

$$(N-1) \log_2 D < L \log_2 K \leq N \log_2 D.$$

Предложенный способ является просто перекодированием источника с использованием другого алфавита. Соотношение длин входных и выходных блоков не зависит от статистических свойств источника. Длины блоков, выраженные в битах, примерно одинаковы. Сжатие данных невозможно.

Сжатие возможно, если осуществлять так называемое кодирование с потерями.

Обозначим через Ω пространство всех входных последовательностей длины L . Выделим в нем область \mathcal{M} . Обозначим дополнительную область через \mathcal{M}^c (см. рис. 6.1). Кодирование с потерями осуществляется следующим образом. Для области \mathcal{M} используется однозначное кодирование. Для кодирования всех последовательностей из дополнительного множества \mathcal{M}^c используется одно кодовое слово. Таким образом, кодирование не яв-



Разным словам из \mathcal{M} соответствуют разные кодовые слова.
Всем словам из \mathcal{M}^c соответствует одно кодовое слово.

Рис. 6.1. Разбиение пространства входных последовательностей для кодирования с потерями

ляется взаимно однозначным. Множество \mathcal{M}^c называется множеством *неоднозначного* декодирования.

Представим число $M = |\mathcal{M}|$ элементов множества \mathcal{M} в виде $M = 2^{RL}$. Число различных выходных кодовых блоков будет равно $2^{RL} + 1$. Наименьшая длина N выходных блоков, при которой это возможно, находится из условий

$$D^{N-1} \leq 2^{RL} + 1 < D^N,$$

или, что эквивалентно, из условий

$$\frac{R}{\log_2(D)} < \frac{N}{L} \leq \frac{R}{\log_2(D)} + \frac{1}{L}.$$

Основная идея кодирования с потерями состоит в выборе множества \mathcal{M} возможно меньшей мощности (малое значение R), но такого, чтобы вероятность потерь, т.е. вероятность множества неоднозначного декодирования \mathcal{M}^c , была пренебрежимо мала: $P(\mathcal{M}^c) \rightarrow 0$ при $L \rightarrow \infty$.

6.2. Прямая теорема Шеннона

Теорема 6.1. Для источника без памяти с энтропией $H(U)$ и любого $\varepsilon > 0$ существует последовательность множеств однозначного декодирования \mathcal{M}_L мощности $2^{L(1+\varepsilon)H(U)}$ такая, что вероятность множества неоднозначного декодирования стремится к нулю: $P(\mathcal{M}_L^c) \rightarrow 0$ при $L \rightarrow \infty$.

Доказательство. Для заданных $\varepsilon > 0$ и L рассмотрим множество всех ε -типовых последовательностей $\mathcal{M}_{L,\varepsilon}$. Согласно (5.16), мощность этого множества удовлетворяет неравенству $\mathcal{M}_{L,\varepsilon} \leq 2^{L(1+\varepsilon)H(U)}$. Выберем множество однозначного декодирования \mathcal{M}_L в виде

$$\mathcal{M}_L = \mathcal{M}_{L,\varepsilon} \cup \mathcal{R},$$

где \mathcal{R} – произвольное множество не ε -типовых последовательностей такое, что $|\mathcal{M}_{L,\varepsilon} \cup \mathcal{R}| = 2^{L(1+\varepsilon)H(U)}$. Оценим вероятность множества неоднозначного декодирования \mathcal{M}_L^c . Так как $\mathcal{M}_L^c = (\mathcal{M}_{L,\varepsilon} \cup \mathcal{R})^c = \mathcal{M}_{L,\varepsilon}^c \cap \mathcal{R}^c$, то

$$P(\mathcal{M}_L^c) = P(\mathcal{M}_{L,\varepsilon}^c \cap \mathcal{R}^c) \leq P(\mathcal{M}_{L,\varepsilon}^c) \leq \frac{K}{L\varepsilon^2 p_{\min}}. \quad (6.1)$$

В последнем неравенстве использована оценка (5.14). Из неравенства (6.1) следует, что $P(\mathcal{M}_L^c) \rightarrow 0$ при $L \rightarrow \infty$.

Прямая теорема Шеннона показывает, что при кодировании с потерями достигима степень сжатия $\frac{N}{L} \approx \frac{H(U)(1+\varepsilon)}{\log_2 D}$ сколь угодно близкая к энтропии источника, но все же превосходящая её. Обратная теорема показывает, что лучшего результата добиться нельзя.

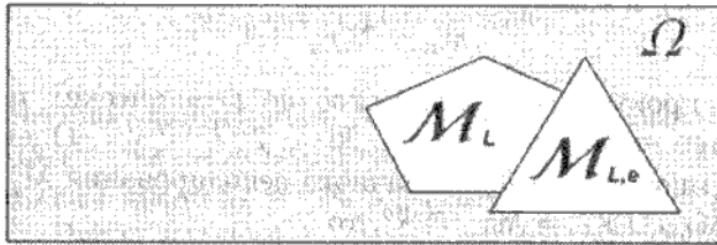
6.3. Обратная теорема Шеннона

Пусть задан источник без памяти с энтропией $H(U)$ и $\varepsilon_1 > 0$.

Теорема 6.2. Для любой последовательности множеств однозначного декодирования \mathcal{M}_L мощности $M_L = 2^{L(1-\epsilon_1)H(U)}$ вероятность множества неоднозначного декодирования стремится к 1: $P(\mathcal{M}_L^c) \rightarrow 1$ при $L \rightarrow \infty$. Другими словами, сжатие невозможно.

Доказательство. Пусть \mathcal{M}_L – область однозначного декодирования. Выберем ϵ так, чтобы $0 < \epsilon < \epsilon_1$. Выделим область $\mathcal{M}_{L,\epsilon}$ ϵ -типових последовательностей. Она может частично пересекаться с областью однозначного декодирования \mathcal{M}_L .

На рис. 6.2 схематически показаны области, используемые для кодирования с потерями.



\mathcal{M}_L – область однозначного декодирования.

$\mathcal{M}_{L,\epsilon}$ – область ϵ -типових последовательностей.

Рис. 6.2. Области, используемые при кодировании с потерями

Так как

$$\Omega = \mathcal{M}_{L,\epsilon} \cup \mathcal{M}_{L,\epsilon}^c,$$

то область однозначного декодирования \mathcal{M}_L можно представить в виде

$$\mathcal{M}_L = \mathcal{M}_L \cap \Omega = (\mathcal{M}_L \cap \mathcal{M}_{L,\epsilon}) \cup (\mathcal{M}_L \cap \mathcal{M}_{L,\epsilon}^c).$$

Поскольку вероятность объединения событий не превосходит суммы вероятностей этих событий, то можно записать оценку

$$\begin{aligned} P(\mathcal{M}_L) &= P\left((\mathcal{M}_L \cap \mathcal{M}_{L,\epsilon}) \cup (\mathcal{M}_L \cap \mathcal{M}_{L,\epsilon}^c)\right) \leq \\ &\leq P(\mathcal{M}_L \cap \mathcal{M}_{L,\epsilon}) + P(\mathcal{M}_L \cap \mathcal{M}_{L,\epsilon}^c). \end{aligned} \quad (6.2)$$

Оценим по отдельности каждое слагаемое в правой части.

$$\begin{aligned} P(\mathcal{M}_L \cap \mathcal{M}_{L,\varepsilon}) &\leq |\mathcal{M}_L| p_{\varepsilon, \max} \leq \\ &\leq 2^{(1-\varepsilon_1)LH(U)} \cdot 2^{-(1-\varepsilon)LH(U)} = 2^{-L(\varepsilon_1 - \varepsilon)H(U)}. \end{aligned}$$

Так как $(\varepsilon_1 - \varepsilon) > 0$, то правая часть этого неравенства стремится к нулю при $L \rightarrow \infty$, т.е. первое слагаемое в (6.2) стремится к нулю.

Оценим второе слагаемое в (6.2). Учтем, что вероятность пересечения событий не превосходит вероятности любого из этих событий:

$$P(\mathcal{M}_L \cap \mathcal{M}_{L,\varepsilon}^c) \leq P(\mathcal{M}_{L,\varepsilon}^c) \leq \frac{K}{L\varepsilon^2 p_{\varepsilon, \min}}.$$

Последнее неравенство следует из оценки (5.14). Правая часть этого неравенства стремится к нулю при $L \rightarrow \infty$, т.е. второе слагаемое в (6.2) стремится к нулю.

Итак, вероятность любого множества однозначного декодирования \mathcal{M}_L с числом элементов $2^{(1-\varepsilon_1)LH(U)}$ стремится к нулю при увеличении L . Это означает, что источник с вероятностью, близкой к 1, порождает последовательности из области неоднозначного декодирования. Сжатие становится невозможным.

Г л а в а 7

Кодирование для канала

7.1. Лемма об обработке информации

На рис. 7.1 приведена схема двухступенчатой обработки входного сигнала. На вход Блока 1 поступает сигнал X . Выходом блока 1 является сигнал Y , который поступает на вход Блока 2. Сигнал на выходе Блока 2 обозначен Z . Предполагается, что все сигналы являются случайными величинами. Обработка может заключаться в преобразовании, добавлении шумов и т.д. Единственное требование – это отсутствие внешних или внутренних обратных связей между входами и выходами или между блоками.



Рис. 7.1. Два последовательных блока обработки сигнала X

В этих условиях тройка случайных величин $\{X, Y, Z\}$ образует цепь Маркова $X \rightarrow Y \rightarrow Z$. Это означает, по определению, что найденное из совместного распределения $P_{XYZ}(x, y, z)$ условное распределение $P_{Z|XY}(z | x, y)$ обладает следующим свойством:

$$P_{Z|XY}(z | x, y) = P_{Z|Y}(z | y) \text{ для всех значений } x. \quad (7.1)$$

Неформально это означает, что "будущее" Z зависит только от "непосредственного прошлого" Y и не зависит от "далекого прошлого" X .

Цепь Маркова может быть определена различными способами. Ниже приводятся несколько эквивалентных определений.

Лемма 7.1. Пусть для тройки случайных величин $\{X, Y, Z\}$ выполняется одно из перечисленных ниже условий:

Свойство 1.

$$P_{Z|XY}(z | x, y) = P_{Z|Y}(z | y) \text{ для всех значений } x. \quad (7.2)$$

Свойство 2.

$$P_{XZ|Y}(x, z | y) = P_{X|Y}(x | y)P_{Z|Y}(z | y) \quad (7.3)$$

т.е. X и Z условно независимы при заданном Y .

Свойство 3.

$$H(Z | XY) = H(Z | Y). \quad (7.4)$$

Свойство 4.

$$H(X | YZ) = H(X | Y) \quad (7.5)$$

Свойство 5.

$$P_{X|YZ}(x | y, z) = P_{X|Y}(x | y) \text{ для всех значений } z. \quad (7.6)$$

Это означает, что тройка случайных величин $\{Z, Y, X\}$, взятых в обратном порядке, также образует цепь Маркова $Z \rightarrow Y \rightarrow X$, которая называется обратной цепью Маркова.

Условия 1 – 5 эквивалентны. Выполнение любого из них влечет выполнение всех остальных.

Доказательство. Здесь покажем только, что из условия 1 следует условие 3. Доказательство остальных утверждений представляется читателю в качестве упражнения.

Пусть выполняется (7.2). Запишем определение условной энтропии $H(Z | XY)$:

$$H(Z | XY) = \sum_x \sum_y \sum_z P_{XYZ}(x, y, z) \log \frac{1}{P_{Z|XY}(z | x, y)}.$$

Из условия (7.2) следует:

$$\begin{aligned} & \sum_x \sum_y \sum_z P_{XYZ}(x, y, z) \log \frac{1}{P_{Z|XY}(z | x, y)} = \\ & = \sum_x \sum_y \sum_z P_{XYZ}(x, y, z) \log \frac{1}{P_{Z|Y}(z | y)} = \\ & = \sum_y \sum_z P_{YZ}(y, z) \log \frac{1}{P_{Z|Y}(z | y)} = \\ & = H(Z | Y). \end{aligned}$$

Доказано свойство 3. Аналогично можно доказать эквивалентность других свойств.

Найдем соотношения между взаимными информационными различиями различных пар случайных величин $\{X, Y, Z\}$, связанных в цепь Маркова.

Лемма 7.2. (*Лемма об обработке сигнала*).

$$\begin{aligned} I(Z; X) & \leq I(Z; Y); \\ I(Z; X) & \leq I(Y; X). \end{aligned} \tag{7.7}$$

Эти неравенства означают, что добавление дополнительного промежуточного блока обработки может только уменьшить информацию между выходом и входом.

Доказательство. По определению,

$$I(Z; Y) = H(Z) - H(Z | Y).$$

Так как рассматриваемые величины связаны в цепь Маркова, то из (7.4) получаем $H(Z | Y) = H(Z | XY)$, т.е.

$$I(Z; Y) = H(Z) - H(Z | XY).$$

С другой стороны, для любых случайных величин верно неравенство $H(Z | XY) \leq H(Z | X)$, так что

$$I(Z; Y) \geq H(Z) - H(Z | X) = I(Z; X).$$

Это доказывает первое из неравенств (7.7).

Аналогично, по определению,

$$I(Y; X) = H(X) - H(X | Y).$$

Так как согласно (7.6) величины $\{Z, Y, X\}$ связаны в цепь Маркова, то $H(X | Y) = H(X | YZ)$. Следовательно,

$$I(Y; X) = H(X) - H(X | YZ).$$

С другой стороны, для любых случайных величин верно неравенство $H(X | YZ) \leq H(X | Z)$, так что

$$I(Y; X) \geq H(X) - H(X | Z) = I(X; Z) = I(Z; X).$$

Это доказывает второе из неравенств (7.7).

7.2. Лемма Фано

Следующий результат является весьма общим. Он позволяет оценивать ошибки в системах передачи без знания детальной структуры блоков обработки при условии, что мы можем оценить взаимную информацию между входом и выходом.

Лемма 7.3. (Лемма Фано). Пусть U и \hat{U} – случайные величины, принимающие K значений из одного и того же множества. Будем интерпретировать величину U как вход некоторой системы обработки, а величину \hat{U} – как оценку величины U на выходе. Пусть $p_e = \Pr(\hat{U} \neq U)$ – вероятность ошибочного решения. Тогда

$$H(U | \hat{U}) \leq h(p_e) + p_e \log_2(K - 1), \quad (7.8)$$

$$\text{где } h(p_e) = -p_e \log_2(p_e) - (1 - p_e) \log_2(1 - p_e).$$

Лемма показывает, что если мы умеем оценивать взаимную информацию $H(U | \hat{U})$, то можно оценивать вероятность ошибки p_e .

Доказательство. Введем новую случайную величину Z :

$$Z = \begin{cases} 0, & \text{если } U = \hat{U}, \\ 1, & \text{если } U \neq \hat{U}. \end{cases} \quad (7.9)$$

Очевидно, что $p_e = \Pr(Z = 1)$ и что $H(Z) = h(p_e)$.

Рассмотрим условную энтропию $H(UZ | \hat{U})$ и запишем для нее цепное равенство двумя способами. Сначала используем разложение

$$H(UZ | \hat{U}) = H(U | \hat{U}) + H(Z | U\hat{U}).$$

Так как Z – детерминированная функция случайных величин U, \hat{U} , то $H(Z | U\hat{U}) = 0$, так что

$$H(UZ | \hat{U}) = H(U | \hat{U}). \quad (7.10)$$

Теперь используем разложение

$$H(UZ | \hat{U}) = H(Z | \hat{U}) + H(U | \hat{U}Z).$$

Для первого слагаемого используем общую оценку

$$H(Z | \hat{U}) \leq H(Z) = h(p_e). \quad (7.11)$$

Второе слагаемое записываем в виде

$$\begin{aligned} H(U | \hat{U}Z) &= \\ &= \Pr(Z = 0)H(U | \hat{U}, Z = 0) + \Pr(Z = 1)H(U | \hat{U}, Z = 1). \end{aligned}$$

Учтем, что $H(U | \hat{U}, Z = 0) = 0$, так при $Z = 0$ величина \hat{U} однозначно определяет величину U . Кроме того, заметим, что при $Z = 1$ величина U принимает не более $K - 1$ значений, так что $H(U | \hat{U}, Z = 1) \leq \log_2(K - 1)$. Следовательно,

$$H(U | \hat{U}Z) \leq p_e \log_2(K - 1). \quad (7.12)$$

Утверждение леммы следует из (7.10), (7.11) и (7.12).

7.3. Классификация дискретных каналов

7.3.1. Общий случай

На рис. 7.2 показан дискретный канал связи без памяти. Входом канала является случайная величина X , принимающая значения из входного алфавита $\mathcal{U} = \{u_1, u_2, \dots, u_K\}$. Выходом канала является случайная величина Y , принимающая значения из выходного алфавита $\mathcal{V} = \{v_1, v_2, \dots, v_D\}$. Канал задается матрицей переходных вероятностей

$$P_{Y|X}(y | x) = \Pr(Y = y | X = x), \quad x \in \mathcal{U}, \quad y \in \mathcal{V}. \quad (7.13)$$

Элемент матрицы $P_{Y|X}(y | x)$ означает условную вероятность получить на выходе канала символ выходного алфавита y при условии, что на вход канала был подан символ входного алфавита x . Таким образом, каждая строка матрицы – это условное распределение с условием нормировки

$$\sum_{y \in \mathcal{V}} P_{Y|X}(y | x) = 1 \quad \text{для любого } x \in \mathcal{U}. \quad (7.14)$$

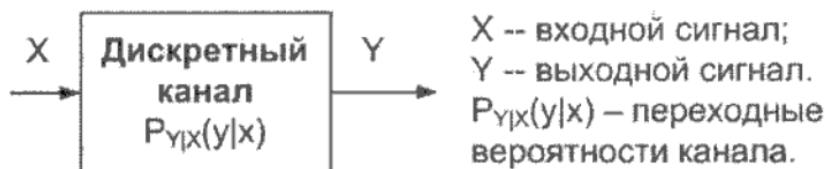


Рис. 7.2. Дискретный канал.

Для полного описания канала необходимо еще задать распределение $P_X(x)$ входной случайной величины X , т.е. вероятности появления каждого входного символа $x \in \mathcal{U}$. После этого можно найти совместное распределение

$$P_{XY}(x, y) = P_X(x)P_{Y|X}(y | x)$$

и вычислить с его помощью маргинальное распределение выходной случайной величины Y как

$$P_Y(y) = \sum_{x \in \mathcal{U}} P_{XY}(x, y) = \sum_{x \in \mathcal{U}} P_X(x)P_{Y|X}(y | x), \quad y \in \mathcal{V}.$$

Передача блоков по каналу без памяти. Пусть по дискретному каналу передается последовательность $\mathbf{x} = (x_1, x_2, \dots, x_L)$. На выходе имеем последовательность $\mathbf{y} = (y_1, y_2, \dots, y_L)$. Канал считаем стационарным без памяти. Поэтому переходные вероятности не зависят от номера символа в последовательности. Совместная вероятность входной и выходной последовательностей есть:

$$\begin{aligned} & P_{\mathbf{XY}}(x_1, x_2, \dots, x_L; y_1, y_2, \dots, y_L) = \\ & = P_{\mathbf{X}}(\mathbf{x})P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}) = P_{\mathbf{X}}(\mathbf{x}) \prod_{i=1}^L P_{Y|X}(y_i | x_i), \end{aligned}$$

так как для канала без памяти

$$\begin{aligned} & P_{\mathbf{Y}|\mathbf{X}}(y_1, y_2, \dots, y_L | x_1, x_2, \dots, x_L) = \\ & = \prod_{i=1}^L P_{Y|X}(y_i | x_i). \end{aligned}$$

В большинстве приложений случайный вектор

$$\mathbf{x} = (x_1, x_2, \dots, x_L)$$

состоит из независимых компонентов с одинаковым распределением $P_X(x)$. В этом случае

$$\begin{aligned} P_{\mathbf{X}}(\mathbf{x}) & = \prod_{i=1}^L P_X(x_i), \\ P_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) & = P_{\mathbf{X}}(\mathbf{x})P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}) = \prod_{i=1}^L P_X(x_i)P_{Y|X}(y_i | x_i). \end{aligned} \tag{7.15}$$

Пропускная способность канала без памяти. Обычно в теории информации рассматривается случай, когда матрица переходных вероятностей задана, а проектировщик может менять по своему усмотрению только распределение $P_X(x)$ входных символов. Важнейшей мерой связи между выходом канала Y и его входом X является средняя взаимная информация между этими величинами $I(Y; X)$. Напомним формулы для количества информации:

$$\begin{aligned}
 I(Y; X) &= I(X; Y) = \sum_{x \in \mathcal{U}} \sum_{y \in \mathcal{V}} P_{XY}(x, y) \log_2 \frac{P_{XY}(x, y)}{P_X(x)P_Y(y)} = \\
 &= \sum_{x \in \mathcal{U}} \sum_{y \in \mathcal{V}} P_{XY}(x, y) \log_2 \frac{P_{Y|X}(y|x)}{P_Y(y)} = \\
 &= H(Y) - H(Y | X) = \\
 &= H(X) - H(X | Y).
 \end{aligned} \tag{7.16}$$

Задача проектировщика — выбрать входное распределение таким образом, чтобы максимизировать количество информации.

Пропускной способностью канала называется следующее выражение:

$$\begin{aligned}
 C &= \sup_{P_X} I(Y; X) = \\
 &= \sup_{P_X} (H(Y) - H(Y | X)) = \\
 &= \sup_{P_X} (H(X) - H(X | Y)).
 \end{aligned} \tag{7.17}$$

Важность этой характеристики будет выяснена в последующих главах.

Пример 7.1. Приведем пример вычисления пропускной способности по общей формуле (7.17). Пусть дискретный канал без памяти с совпадающими входным и выходным алфавитами $\mathcal{U} = \mathcal{V} = \{0, 1, 2\}$ задается матрицей переходных вероятностей вида

$$\begin{bmatrix} 2/3 & 1/3 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/3 & 2/3 \end{bmatrix}.$$

Найдем оптимальное входное распределение и пропускную способность для этого канала. Введем обозначения для искомого входного распределения $P_X(x)$:

$$p_X(0) = \alpha, p_X(1) = \beta, p_X(2) = \gamma, \alpha + \beta + \gamma = 1.$$

Запишем выходное распределение $P_Y(y)$:

$$p_Y(0) = \frac{2\alpha}{3} + \frac{\beta}{3}, p_Y(1) = \frac{1}{3}, p_Y(2) = \frac{\beta}{3} + \frac{2\gamma}{3}.$$

Представим условную энтропию $H(Y | X)$ формулой

$$\begin{aligned} H(Y | X) &= \alpha h(1/3) + \beta \log_2 3 + \gamma h(1/3) = \\ &= (\alpha + \gamma)h(1/3) + \beta \log_2 3 = h(1/3) + \beta(\log_2 3 - h(1/3)), \end{aligned}$$

где

$$h(1/3) = -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3}$$

— энтропия первой и третьей строки, а $\log_2 3$ — энтропия второй строки.

Как видно из этой формулы, условная энтропия $H(Y | X)$ зависит только от β . Так как $\log_2 3 - h(p) > 0$, то она минимальна при $\beta = 0$: $H_{\min}(Y | X) = h(1/3)$.

Безусловная энтропия $H(Y)$ равна

$$H(Y) = -p_Y(0) \log_2 p_Y(0) - p_Y(2) \log_2 p_Y(2) - p_Y(1) \log_2 p_Y(1) \quad (7.18)$$

Заметим, что вероятность $p_Y(1) = 1/3$ не зависит от β , как и сумма вероятностей $(p_Y(0) + p_Y(2)) = 2/3$. Следовательно, сумма первых двух слагаемых в (7.18) по свойству энтропии будет максимальна, если $p_Y(0) = p_Y(2) = 1/3$. В этом случае энтропия $H(Y) = \log_2 3$ максимальна и не зависит от β .

В указанных условиях средняя взаимная информация $I(Y; X)$ зависит только от β и может быть записана в виде:

$$I(Y; X) = H(Y) - H(Y | X) = \log_2 3 - h(1/3) - \beta(\log_2 3 - h(1/3)) \quad (7.19)$$

Так как $\log_2 3 - h(1/3) > 0$, то информация достигает максимума при $\beta = 0$.

Таким образом, оптимальное входное распределение равно

$$p_X(0) = \frac{1}{2}, \quad p_X(1) = 0, \quad p_X(2) = \frac{1}{2},$$

а пропускная способность

$$C = \log_2 3 - h(1/3) = \frac{2}{3}.$$

Вычисление пропускной способности в общем случае является трудной задачей. В некоторых случаях задача оптимизации облегчается, если матрица переходных вероятностей имеет специальную структуру. Обозначим $p_{ij} = P_{Y|X}(v_j | x_i)$ и запишем матрицу переходных вероятностей в виде

$$P_{Y|X} = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1D} \\ p_{21} & p_{22} & \dots & p_{2D} \\ \vdots & \vdots & \dots & \vdots \\ p_{K1} & p_{K2} & \dots & p_{KD} \end{pmatrix}. \quad (7.20)$$

Условная энтропия $H(Y | X = u_i)$ случайной величины Y при заданном значении $X = u_i, i = 1, 2, \dots, K$, совпадает с энтропией i -ой строки:

$$H(Y | X = u_i) = - \sum_{j=1}^D p_{ij} \log_2 p_{ij}, \quad i = 1, 2, \dots, K. \quad (7.21)$$

Соответственно, условная энтропия $H(Y | X)$ при заданной случайной величине X равна

$$H(Y | X) = - \sum_{i=1}^K P_X(u_i) \sum_{j=1}^D p_{ij} \log_2 p_{ij}. \quad (7.22)$$

Частичная классификация каналов проводится по структурным свойствам симметрии матрицы (7.20).

7.3.2. Каналы с нулевой пропускной способностью

Лемма 7.4. Пропускная способность канала равна 0 тогда и только тогда, когда все строки матрицы переходных вероятностей одинаковы, т.е.

$$P_{Y|X} = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1D} \\ p_{11} & p_{12} & \dots & p_{1D} \\ \vdots & \vdots & \dots & \vdots \\ p_{11} & p_{12} & \dots & p_{1D} \end{pmatrix}. \quad (7.23)$$

Доказательство. Пусть матрица переходных вероятностей имеет вид (7.23). Тогда все условные энтропии $H(Y | X = u_i)$ одинаковы:

$$H(Y | X = u_i) = - \sum_{j=1}^D p_{1j} \log_2 p_{1j}, \quad i = 1, 2, \dots, K,$$

и равны условной энтропии $H(Y | X)$ при любом входном распределении $P_X(x)$:

$$\begin{aligned} H(Y | X) &= - \sum_{i=1}^K P_X(u_i) \sum_{j=1}^D p_{1j} \log_2 p_{1j} \\ &= \left(\sum_{i=1}^K P_X(u_i) \right) \left(- \sum_{j=1}^D p_{1j} \log_2 p_{1j} \right) = \\ &= - \sum_{j=1}^D p_{1j} \log_2 p_{1j}. \end{aligned}$$

С другой стороны, при любом входном распределении $P_X(x)$ маргинальное выходное распределение $P_Y(y)$ будет одним и тем же, а именно:

$$P_Y(v_j) = \sum_{i=1}^K P_X(u_i) p_{ij} = \left(\sum_{i=1}^K P_X(u_i) \right) p_{1j} = p_{1j}.$$

Следовательно,

$$H(Y) = - \sum_{j=1}^D P_Y(v_j) \log_2 P_Y(v_j) = - \sum_{j=1}^D p_{1j} \log_2 p_{1j} = H(Y | X).$$

Таким образом, при любом входном распределении верно соотношение $H(Y) - H(Y | X) = 0$, т.е. $C = 0$.

Обратно, пусть

$$C = \sup_{P_X} \{H(Y) - H(Y | X) = 0\} = \sup_{P_X} I(Y; X) = 0.$$

Это означает, что для любых распределений $P_X(x)$ средняя взаимная информация $I(Y; X) = 0$. Из определения (7.16) следует, что это возможно лишь при условии, что

$$P_Y(v_j) = P_{Y|x}(v_j | u_i) = p_{ij}, \quad i = 1, 2, \dots, K; \quad j = 1, 2, \dots, D.$$

Так как левая часть не зависит от i , то $p_{1j} = p_{2j} = \dots = p_{Dj}$, т.е. матрица переходных вероятностей имеет вид (7.23).

7.3.3. Каналы, симметричные по входу

Канал называется *симметричным по входу*, если в его матрице переходных вероятностей каждая строка получена из первой строки перестановкой ее элементов.

Пусть задана первая строка матрицы переходных вероятностей, а остальные строки являются перестановкой элементов первой строки. Тогда все условные энтропии $H(Y | X = u_i)$ одинаковы и равны энтропии первой строки:

$$H(Y | X = u_i) = - \sum_{j=1}^D p_{1j} \log_2 p_{1j}, \quad i = 1, 2, \dots, K.$$

Следовательно, при любом входном распределении $P_X(x)$ условная энтропия $H(Y | X)$ равна:

$$\begin{aligned}
 H(Y | X) &= - \sum_{i=1}^K P_X(u_i) \sum_{j=1}^D p_{1j} \log_2 p_{1j} = \\
 &= \left(\sum_{i=1}^K P_X(u_i) \right) \left(- \sum_{j=1}^D p_{1j} \log_2 p_{1j} \right) = \\
 &= - \sum_{j=1}^D p_{1j} \log_2 p_{1j}.
 \end{aligned}$$

Пропускная способность в этом случае равна

$$C = \sup_{p_X} H(Y) - H(Y | X). \quad (7.24)$$

Пример 7.2. Пусть дискретный канал без памяти задается следующей матрицей переходных вероятностей:

$$\begin{pmatrix} \frac{1}{6} & \frac{1}{6} & \frac{2}{6} & \frac{2}{6} \\ \frac{1}{6} & \frac{2}{6} & \frac{1}{6} & \frac{2}{6} \\ \frac{2}{6} & \frac{1}{6} & \frac{1}{6} & \frac{2}{6} \end{pmatrix}.$$

Найдем оптимальное входное распределение и пропускную способность этого канала. Так как канал симметричен по входу, то условная энтропия $H(Y | X)$ не зависит от входного распределения $P_X(x)$ и равна

$$H(Y | X) = -2 \frac{1}{2} \log_2 \frac{1}{6} - 2 \frac{2}{6} \log_2 \frac{2}{6} = 1.918.$$

Учтем также, что вероятность $P_Y(4)$ не зависит от входного распределения $P_X(x)$:

$$P_Y(4) = (2/6)(P_X(x_1) + P_X(x_2) + P_X(x_3) + P_X(x_4)) = 2/6.$$

Следовательно, при любом распределении $P_X(x)$ энтропия выхода равна

$$\begin{aligned}
 H(Y) &= -p_Y(1) \log_2 p_Y(1) - p_Y(2) \log_2 p_Y(2) - \\
 &\quad - p_Y(3) \log_2 p_Y(3) - \frac{2}{6} \log_2 \frac{2}{6},
 \end{aligned} \quad (7.25)$$

где $p_Y(1) + p_Y(2) + p_Y(3) = 1 - 2/6 = 2/3$.

Максимум этой величины достигается, если вероятности равны $p_Y(1) = p_Y(2) = p_Y(3) = 2/9$. Тогда максимум энтропии равен

$$\sup H(Y) = -3 \frac{2}{9} \log_2 \frac{2}{9} - \frac{2}{6} \log_2 \frac{2}{6} = 1.975.$$

В свою очередь, это условие выполняется, если входное распределение *равномерное*: $P_X(x) = \frac{1}{3}$, для любого x . Итак, пропускная способность равна

$$C = 1.975 - 1.918 = 0.057.$$

Пример 7.3. Приведем другой пример канала, симметричного по входу. Пусть матрица переходных вероятностей имеет следующий вид:

$$\begin{pmatrix} 1-r & 0 & r \\ 0 & 1-r & r \end{pmatrix}. \quad (7.26)$$

Этот канал называется двоичным симметричным каналом со стираниями. Входной алфавит состоит из двух символов 0 и 1, а выходной – из трех символов 0, 1 и θ , где θ означает символ "стирания". С вероятностью $1-r$ этот канал передает символы 0 и 1 правильно, а с вероятностью r они принимаются как символ "стирания". Граф этого канала приведен на рис. 7.3.

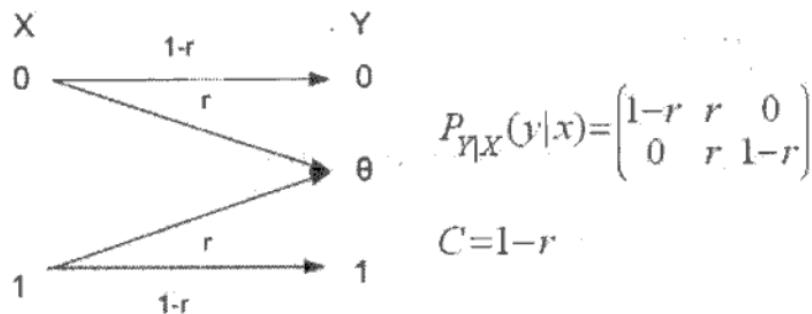


Рис. 7.3. Двоичный симметричный канал со стираниями

Найдем пропускную способность. Так как канал является симметричным по входу, то условная энтропия $H(Y | X)$ определяется элементами любой строки матрицы (7.26):

$$H(Y | X) = -r \log_2 r - (1-r) \log_2 (1-r) = h(r). \quad (7.27)$$

Теперь надо найти энтропию $H(Y)$ и ее максимум по входному распределению $P_X(x)$. Сначала найдем распределение $P_Y(y)$ символов на выходе канала:

$$\begin{aligned} P_Y(0) &= P_X(0)(1-r), \\ P_Y(1) &= P_X(1)(1-r) = (1-P_X(0))(1-r), \\ P_Y(\theta) &= P_X(0)r + P_X(1)r = r, \end{aligned}$$

где $P_X(1) = 1 - P_X(0)$.

Используя эти формулы и проведя элементарные преобразования, запишем энтропию $H(Y)$ в виде

$$H(Y) = h(r) + (1-r)h(P_X(0)). \quad (7.28)$$

Эта величина достигает максимума при $P_X(0) = P_X(1) = \frac{1}{2}$.

В этом случае $h(P_X(0)) = 1$ и максимальное значение энтропии равно $H(Y) = h(r) + (1-r)$. В результате получаем пропускную способность этого симметричного по входу канала со стиранием в виде:

$$C = \sup_{P_X} H(Y) - H(Y | X) = h(r) + (1-r) - h(r) = (1-r). \quad (7.29)$$

7.3.4. Каналы, симметричные по выходу

Канал называется *симметричным по выходу*, если в его матрице переходных вероятностей каждый столбец получен из первого столбца перестановкой его элементов.

Пример 7.4. Пусть дискретный канал без памяти с входным алфавитом $\mathcal{U} = \{0, 1, 2\}$ и выходным алфавитом $\mathcal{V} = \{0, 1, 2, 3\}$ задается следующей матрицей переходных вероятностей:

$$\begin{bmatrix} a & a & b & b \\ b & c & a & c \\ c & b & c & a \end{bmatrix}.$$

Канал симметричен по выходу, так как все столбцы являются перестановками первого столбца.

Найдем оптимальное входное распределение и пропускную способность этого канала. В данном случае симметрия по выходу не облегчает вычислений.

Обозначим входное распределение как $P_X = \{q_0, q_1, q_2\}$, где $q_0 + q_1 + q_2 = 1$.

Обозначим выходное распределение как $P_Y = \{p_0, p_1, p_2, p_3\}$, где $p_0 + p_1 + p_2 + p_3 = 1$.

Переходные вероятности таковы:

$$P_{Y|X}(0|0) = P_{Y|X}(1|0) = P_{Y|X}(2|1) = P_{Y|X}(3|2) = a;$$

$$P_{Y|X}(2|0) = P_{Y|X}(3|0) = P_{Y|X}(0|1) = P_{Y|X}(1|2) = b;$$

$$P_{Y|X}(1|1) = P_{Y|X}(3|1) = P_{Y|X}(0|2) = P_{Y|X}(2|2) = c.$$

Учтем соотношения между переходными вероятностями, исходя из вида матрицы. Вторая и третья строки матрицы состоят из одних и тех же элементов. Их перестановка не меняет пропускной способности. Поэтому входные вероятности, соответствующие этим строкам, должны быть равными, т. е. $q_1 = q_2 = \frac{1-q_0}{2}$. Так как сумма элементов любой строки равна 1, то

$$2a + 2b = 1$$

$$2c + a + b = 1$$

Отсюда следует $c = 0.25$. Выразим вероятности символов на выходе канала через вероятности символов на входе и переходные вероятности:

$$p_0 = q_0a + \frac{(1-q_0)(b+1/4)}{2}$$

$$p_1 = q_0a + \frac{(1-q_0)(b+1/4)}{2}$$

$$p_2 = q_0b + \frac{(1-q_0)(a+1/4)}{2}$$

$$p_3 = q_0b + \frac{(1-q_0)(a+1/4)}{2}$$

Из этих соотношений следует: $p_0 = p_1, p_2 = p_3, p_2 = 0.5 - p_0$.

Используя эти соотношения, запишем формулы для энтропии $H(Y)$ и условной энтропии $H(Y/X)$:

$$H(Y) = -2p_0 \log_2 p_0 - 2(0.5 - p_0) \log_2 (0.5 - p_0);$$

$$H(Y/X) = q_0(-2a \log_2 a - 2b \log_2 b) +$$

$$+(1 - q_0)(-\frac{1}{2} \log_2 \frac{1}{4} - a \log_2 a - b \log_2 b).$$

Чтобы найти оптимальное распределение на входе, запишем разность $H(Y) - H(Y/X)$, используя правые части этих соотношений, возьмем производную по q_0 и приравняем к нулю:

$$\frac{dH(y)}{dp_0} \frac{dp_0}{dq_0} - \frac{dH(Y/X)}{dq_0} = 0. \quad (7.30)$$

Решение этого уравнения относительно q_0 зависит от a как от параметра.

Возьмем, по возможности, большое значение $a = 0.4999$, соответственно малое значение $b = 0.5 - a = 0.0001$. Решаем уравнения оптимальности (7.30). В этом случае вероятности символов 0, 1, 2 равны соответственно $q_0 = 0.4846; q_1 = q_2 = 0.2577$. Этим значениям соответствуют следующие значения вероятностей на выходе $p_0 = p_1 = 0.3067, p_2 = p_3 = 0.1933$. Энтропия и условная энтропия равны соответственно: $H(Y) = 1.9625$ и $H(Y | X) = 1.2579$. Пропускная способность канала $C = 0.70458$.

Теперь выберем $a = 0.255, b = 0.245$ (близкие значения). Снова решаем уравнения оптимальности (7.30). Вероятности входных символов равны $q_0 = 0.4445; q_1 = q_2 = 0.27775$. Им соответствуют следующие значения вероятностей на выходе канала $p_0 = p_1 = 0.2492, p_2 = p_3 = 0.2508$. Энтропия и условная энтропия равны соответственно: $H(Y) = 1.9999$ и $H(Y | X) = 1.9979$. Пропускная способность канала равна $C = 0.00020$, что много меньше значения, полученного выше при большом отличии параметров a и b .

Для каналов, симметричных по выходу, справедливо следующее утверждение: если входное распределение равномерно, то и выходное распределение тоже равномерно. Равномерное входное распределение приводит к равномерному распределению на

выходе и максимизирует энтропию $H(Y)$, но не всегда минимизирует условную энтропию $H(Y/X)$.

В приведенном примере оптимальное входное распределение не является равномерным: и при большом и при малом значении параметра a вероятность символа 0 больше $\frac{1}{3}$ и близко к $\frac{1}{2}$, а вероятности двух других символов одинаковы и меньше $\frac{1}{3}$. Проанализируем отличие значений пропускной способности для оптимального входного распределения и для равномерного входного распределения. Пусть входное распределение равномерное: $q_0 = q_1 = q_2 = \frac{1}{3}$. Получим $p_0 = p_1 = p_2 = p_3 = 0.25$. При $a = 0.4999$ энтропия и условная энтропия равны соответственно: $H(Y) = 2.0$ и $H(Y/X) = 1.3336$. Пропускная способность $C = 0.66664$. При $a = 0.255$ также получаем равномерное входное распределение и то же значение энтропии $H(Y) = 2.0$. Условная энтропия $H(Y/X) = 1.99981$. Пропускная способность равна $C = 0.00019$. Анализ этого примера показывает, что в данном случае при большом отличии параметров a и b разность оптимального и неоптимального значения пропускной способности больше, чем при близких значениях a и b .

7.3.5. Каналы, симметричные по входу и выходу

Канал называется *симметричным по входу и выходу*, если в его матрице переходных вероятностей каждая строка является перестановкой элементов первой строки, а каждый столбец является перестановкой элементов первого столбца.

В этом случае пропускная способность вычисляется с учетом этих свойств следующим образом:

$$C = \sup_{p_X} H(Y) - H(Y | X), \quad (7.31)$$

где

$$\sup_{p_X} H(Y) = \log_2 D. \quad (7.32)$$

Пример 7.5. Пусть матрица переходных вероятностей имеет вид

$$\begin{bmatrix} b & a & d & c \\ a & d & c & b \\ d & c & b & a \\ c & b & a & d \end{bmatrix}, \quad a \geq 0, b \geq 0, c \geq 0, d \geq 0; a + b + c + d = 1.$$

Найдем оптимальное входное распределение и пропускную способность этого канала.

Так как все строки матрицы — перестановки первой строки и все столбцы — перестановки первого столбца, то канал симметричен по входу и по выходу. Поэтому оптимальное входное распределение — равномерное, а пропускная способность равна

$$\begin{aligned} C &= \log_2 4 + a \log_2 a + b \log_2 b + c \log_2 c + d \log_2 d = \\ &= 2 + a \log_2 a + b \log_2 b + c \log_2 c + d \log_2 d. \end{aligned}$$

7.6. Для двоичного симметричного канала без памяти матрица переходных вероятностей имеет следующий вид:

$$P_{Y|X} = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}. \quad (7.33)$$

Удобно также описывать канал с помощью графа (рис. 7.4):

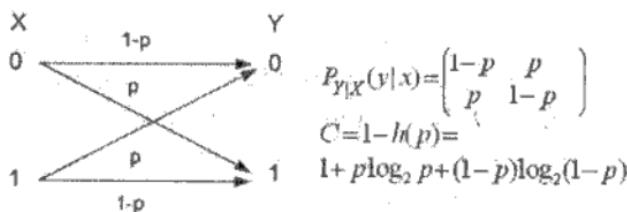


Рис. 7.4. Двоичный симметричный канал

Пропускная способность двоичного симметричного канала равна

$$C = \log_2 2 - h(p) = 1 + p \log_2 p + (1-p) \log_2 (1-p). \quad (7.34)$$

Г л а в а 8

Передача по каналу и декодирование

8.1. Основные характеристики передачи

На рис. 8.1 приведена модель передачи сообщений по дискретному каналу без памяти. Введем необходимые понятия и определения.



Рис. 8.1. Модель передачи по дискретному каналу без памяти

Пусть источник сообщений порождает поток двоичных символов (бит) U_1, U_2, \dots . Рассматриваем блоковую передачу. Поток разбивается на блоки $\mathbf{U} = (U_1, U_2, \dots, U_K)$ длины K . Число различных блоков равно $M = 2^K$. Каждый блок двоичных символов $\mathbf{U} = (U_1, U_2, \dots, U_K)$ кодируется, передается по каналу, декодируется и передается получателю в виде блока двоичных символов $\tilde{\mathbf{U}} = (\tilde{U}_1, \tilde{U}_2, \dots, \tilde{U}_K)$. Если $\tilde{U}_i \neq U_i$, то говорят, что произошла *ошибка в бите*. Вероятность ошибки обозначим

$$P_{B,i} = \Pr(\tilde{U}_i \neq U_i). \quad (8.1)$$

Если полученный блок $\tilde{\mathbf{U}}$ не совпадает с исходным \mathbf{U} , то считают, что произошла *ошибка в блоке*, вероятность этой ошибки обозначим

$$P_B = \Pr(\tilde{\mathbf{U}} \neq \mathbf{U}). \quad (8.2)$$

Так как

$$\Pr(\tilde{\mathbf{U}} \neq \mathbf{U}) = \Pr(\text{хотя бы для одного } s = 1, \dots, K, \tilde{U}_s \neq U_s),$$

то вероятность ошибки в блоке удовлетворяет неравенству

$$P_B \leq P_{B,1} + P_{B,2} + \dots + P_{B,K} = K P_{B,av}, \quad (8.3)$$

где

$$P_{B,av} = \frac{1}{K} \sum_{s=1}^K P_{B,s}$$

означает среднюю вероятность ошибки в бите.

Различные блоки \mathbf{U} можно перенумеровать целыми числами $s = 1, 2, \dots, M$ и считать, что передаваемым сообщением является число s .

Кодер канала ставит в соответствие каждому сообщению s кодовый вектор длины L :

$$\mathbf{x}_s = (x_{s,1}, x_{s,2}, \dots, x_{s,L}), \quad s = 1, 2, \dots, M.$$

Координаты $x_{s,j}$ кодовых векторов выбираются из входного алфавита канала $\mathcal{U} = \{u_1, u_2, \dots, u_K\}$. Набор кодовых векторов

$$\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$$

называется кодом. Код известен как на передающей, так и на приемной стороне.

Число передаваемых сообщений M и длина кодовых векторов L связаны между собой с помощью параметра R , называемого *скоростью передачи*. Он определяется следующим образом:

$$M = 2^{RL}, \text{ или } R = \frac{\log_2 M}{L}.$$

В рассматриваемой модели $R = K/L$. Неформально, при передаче вектора длины L канал используется L раз. Скорость передачи равна числу бит, передаваемых за одно использование канала.

При передаче по каналу вектора \mathbf{x} на выходе принимается вектор $\mathbf{y} = (y_1, y_2, \dots, y_L)$ с координатами из выходного алфавита $\mathcal{V} = \{v_1, v_2, \dots, v_D\}$.

Канал описывается матрицей переходных вероятностей: если по каналу передается вектор \mathbf{x} , то вероятность получить на выходе вектор \mathbf{y} равна $P_{Y|X}(\mathbf{y} | \mathbf{x})$. Переходные вероятности должны быть заданы для всех допустимых пар \mathbf{x}, \mathbf{y} с очевидным условием нормировки. Переходная вероятность, рассматриваемая как функция аргумента \mathbf{x} при фиксированном \mathbf{y} , называется *функцией правдоподобия*. Говорят, что вектор \mathbf{x}_1 более правдоподобен при заданном \mathbf{y} , чем вектор \mathbf{x}_2 , если

$$P_{Y|X}(\mathbf{y} | \mathbf{x}_1) > P_{Y|X}(\mathbf{y} | \mathbf{x}_2).$$

Канал без памяти описывается вероятностью $P_{Y|X}(y | x)$ получения на выходе буквы выходного алфавита y при условии, что передавалась буква входного алфавита x . При передаче блоков длины L переходные вероятности записываются следующим образом: если входной вектор равен $\mathbf{x} = (x_1, x_2, \dots, x_L)$, а выходной блок $\mathbf{y} = (y_1, y_2, \dots, y_L)$, то

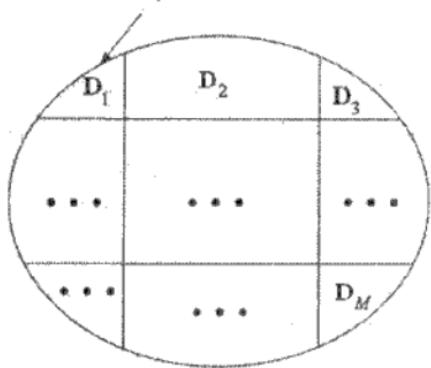
$$P_{Y|X}(\mathbf{y} | \mathbf{x}) = P_{Y|X}(y_1 | x_1) \cdot P_{Y|X}(y_2 | x_2) \cdots P_{Y|X}(y_L | x_L). \quad (8.4)$$

Декодер канала должен по принятому вектору \mathbf{y} вынести решение о том, какое кодовое слово \mathbf{x}_s передавалось и выдать получателю сообщение s или блок (U_1, U_2, \dots, U_K) . С этой целью пространство выходных векторов \mathbf{Y} разбивается на *области декодирования*. Возможны две стратегии разбиения – без отказов от декодирования и с отказами от декодирования (см. рис. 8.2).

В стратегии "без отказов от декодирования" пространство выходных векторов \mathbf{Y} разбивается на M областей декодирования D_i , $i = 1, 2, \dots, M$. Процедура передачи состоит в следующем.

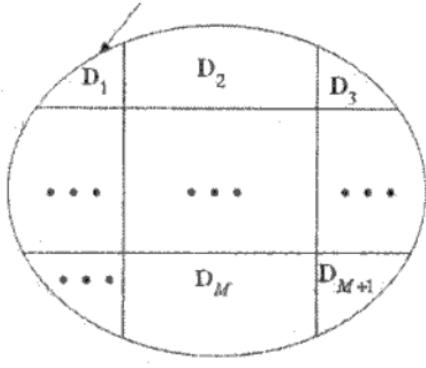
1. Если необходимо передать сообщение s , то в канал передается кодовый вектор \mathbf{x}_s .

Пространство выходных векторов \mathbf{Y}



Без отказов от декодирования

Пространство выходных векторов \mathbf{Y}



С отказами от декодирования

Рис. 8.2. Разбиение пространства выходных векторов на области декодирования

- Пусть принят вектор \mathbf{y} . Декодирование осуществляется по правилу: если вектор \mathbf{y} лежит в области D_j , т.е. $\mathbf{y} \in D_j$, то выносится решение, что передавался кодовый вектор \mathbf{x}_j , соответствующий сообщению j .

В стратегии "с отказами от декодирования" пространство выходных векторов \mathbf{Y} разбивается на $M + 1$ областей декодирования D_i , $i = 1, 2, \dots, M + 1$. Процедура передачи состоит в следующем.

- Если необходимо передать сообщение s , то в канал передается кодовый вектор \mathbf{x}_s .
- Пусть принят вектор \mathbf{y} . Декодирование осуществляется по правилу: если вектор \mathbf{y} лежит в области D_j , где j одно из чисел $\{1, 2, \dots, M\}$, то выносится решение, что передавался кодовый вектор \mathbf{x}_j , соответствующий сообщению j . Если же вектор \mathbf{y} лежит в области D_{M+1} , то выносится решение "отказ от декодирования".

Предполагаем, что все сообщения для передачи выбираются равновероятно: $Pr(\mathbf{x} = \mathbf{x}_i) = 1/M$ для всех $i = 1, 2, \dots, M$.

Если передавался вектор \mathbf{x}_s , то решение при декодировании будет вынесено правильно, если принятый вектор лежит в области D_s , т.е. $\mathbf{y} \in D_s$. Вероятность этого события равна

$$P_{\text{cor},s} = \sum_{\mathbf{y} \in D_s} P_{Y|X}(\mathbf{y} | \mathbf{x}_s).$$

Ошибка произойдет, если принятый вектор \mathbf{y} лежит в дополнительной к D_s области, которая обозначается D_s^c и равна

$$D_s^c = \mathbf{Y} \setminus D_s = \bigcup_{i=1}^M D_i \setminus D_s.$$

Вероятность ошибки при передаче кодового слова \mathbf{x}_s равна

$$P_{\text{er},s} = 1 - P_{\text{cor},s} = \sum_{\mathbf{y} \in D_s^c} P_{Y|X}(\mathbf{y} | \mathbf{x}_s). \quad (8.5)$$

Важной характеристикой является *средняя* вероятность ошибки

$$P_{\text{er}} = \frac{1}{M} \sum_{i=1}^M P_{\text{er},i}. \quad (8.6)$$

В стратегии "с отказами от декодирования" вероятность отказа включается в вероятность ошибки. Иногда событие "отказ от декодирования" анализируется отдельно. В этом случае средняя вероятность отказа определяется как

$$P_{\text{fail}} = \frac{1}{M} \sum_{i=1}^M \sum_{\mathbf{y} \in D_{M+1}^c} P_{Y|X}(\mathbf{y} | \mathbf{x}_i).$$

Задача проектировщика – выбрать области декодирования D_s оптимальным образом, чтобы минимизировать среднюю вероятность ошибки декодирования.

Сначала рассмотрим случай $M = 2$. В этом случае имеется два сообщения $s = 1, 2$ и два кодовых слова $\mathbf{x}_1, \mathbf{x}_2$. Разбиение пространства выходных векторов на области декодирования проводится следующим образом:

- если вектор y таков, что $P_{Y|X}(y | x_1) > P_{Y|X}(y | x_2)$, то его относят в область D_1 ;
- если вектор y таков, что $P_{Y|X}(y | x_2) > P_{Y|X}(y | x_1)$, то его относят в область D_2 ;
- если вектор y таков, что $P_{Y|X}(y | x_1) = P_{Y|X}(y | x_2)$, то его относят произвольно в область D_1 или D_2 , например, случайнным образом.

Такое правило разбиения на области декодирования называется разбиением по методу *максимального правдоподобия*. Решение при декодировании выносится в пользу *наиболее правдоподобного* вектора. При заранее заданных кодовых векторах такое разбиение является оптимальным в том смысле, что никакое другое разбиение не приведет к уменьшению средней вероятности ошибки. Докажем это.

В левой части рис. 8.3 показано разбиение на произвольные области декодирования.



Рис. 8.3. Преобразование неоптимальных областей декодирования в оптимальные для $M = 2$

Область D_1 в общем случае можно представить как объединение двух подобластей D_{11} и D_{12} :

$$\begin{aligned} D_1 &= D_{11} \cup D_{12}, \\ D_{11} &= \{y : P_{Y|X}(y | x_1) \geq P_{Y|X}(y | x_2)\}, \\ D_{12} &= \{y : P_{Y|X}(y | x_1) < P_{Y|X}(y | x_2)\}. \end{aligned}$$

При попадании вектора y в область D_{11} решение выносится в пользу наиболее вероятного вектора (в данном случае x_1), однако при попадании вектора в область D_{12} решение будет вынесено в пользу менее вероятного вектора (в данном случае опять-таки x_1).

Аналогично, область D_2 в общем случае можно представить как

$$\begin{aligned} D_2 &= D_{21} \cup D_{22}, \\ D_{21} &= \{y : P_{Y|X}(y | x_2) < P_{Y|X}(y | x_1)\}, \\ D_{22} &= \{y : P_{Y|X}(y | x_2) \geq P_{Y|X}(y | x_1)\}. \end{aligned}$$

Для этих областей имеем: $D_1^c = D_2$, $D_2^c = D_1$.

Средняя вероятность ошибки для этого разбиения равна

$$\begin{aligned} P_{\text{er}} &= \frac{1}{2}(P_{\text{er},1} + P_{\text{er},2}) = \\ &= \frac{1}{2} \left(\sum_{y \in D_1^c} P_{Y|X}(y | x_1) + \sum_{y \in D_2^c} P_{Y|X}(y | x_2) \right) = \\ &= \frac{1}{2} \left(\sum_{y \in D_{21}} P_{Y|X}(y | x_1) + \sum_{y \in D_{22}} P_{Y|X}(y | x_1) + \right. \\ &\quad \left. + \sum_{y \in D_{11}} P_{Y|X}(y | x_2) + \sum_{y \in D_{12}} P_{Y|X}(y | x_2) \right). \end{aligned} \tag{8.7}$$

Преобразуем теперь исходные области декодирования в области декодирования по *максимальному правдоподобию* (правая часть рис. 8.3). Для новой области $D_{1,\text{opt}}$ сохраним подобласть D_{11} , добавим подобласть D_{21} , удалим подобласть D_{12} , так что

$$\begin{aligned} D_{1,\text{opt}} &= D_{11} \cup D_{21} = \{y : P_{Y|X}(y | x_1) \geq P_{Y|X}(y | x_2)\}, \\ D_{1,\text{opt}}^c &= D_{12} \cup D_{22}. \end{aligned}$$

Аналогично,

$$D_{2,\text{opt}} = D_{12} \cup D_{22} = \{y : P_{Y|X}(y | x_2) \geq P_{Y|X}(y | x_1)\},$$

$$D_{2,\text{opt}}^c = D_{11} \cup D_{21}.$$

Средняя вероятность ошибки для этого разбиения равна

$$\begin{aligned} P_{\text{er,} \text{opt}} &= \frac{1}{2}(P_{\text{er,} 1, \text{opt}} + P_{\text{er,} 2, \text{opt}}) = \\ &= \frac{1}{2} \left(\sum_{y \in D_{1,\text{opt}}^c} P_{Y|X}(y | x_1) + \sum_{y \in D_{2,\text{opt}}^c} P_{Y|X}(y | x_2) \right) = \\ &= \frac{1}{2} \left(\sum_{y \in D_{12}} P_{Y|X}(y | x_1) + \sum_{y \in D_{22}} P_{Y|X}(y | x_1) + \right. \\ &\quad \left. + \sum_{y \in D_{11}} P_{Y|X}(y | x_2) + \sum_{y \in D_{21}} P_{Y|X}(y | x_2) \right). \end{aligned} \quad (8.8)$$

Заметим, что

$$\begin{aligned} \sum_{y \in D_{12}} P_{Y|X}(y | x_1) &< \sum_{y \in D_{12}} P_{Y|X}(y | x_2), \\ \sum_{y \in D_{21}} P_{Y|X}(y | x_2) &< \sum_{y \in D_{21}} P_{Y|X}(y | x_1), \end{aligned}$$

так как каждое слагаемое в левой сумме строго меньше соответствующего слагаемого в правой сумме. Следовательно,

$$P_{\text{er,} \text{opt}} < P_{\text{er}}.$$

Для произвольного значения M разбиение пространства выходных векторов на оптимальные по методу максимального правдоподобия области декодирования проводится следующим образом:

$$D_s = \left\{ y : P_{Y|X}(y | x_s) \geq \max_{j \neq s, j=1, \dots, M} P_{Y|X}(y | x_j) \right\}, \quad (8.9)$$

$$s = 1, 2, \dots, M.$$

Таким образом,

- если вектор y таков, что максимально правдоподобным оказывается вектор x_s , т.е.

$$P_{Y|X}(y | x_s) > \max_{j \neq s, j=1, \dots, M} P_{Y|X}(y | x_j),$$

то его относят в область D_s ;

- если вектор y таков, что

$$P_{Y|X}(y | x_i) = \max_{j \neq i} P_{Y|X}(y | x_j),$$

то его относят произвольно либо в область D_s , либо в область D_j , для которой $P_{Y|X}(y | x_s) = P_{Y|X}(y | x_j)$.

Как и в случае кода из двух слов, разбиение на области декодирования по методу максимального правдоподобия минимизирует среднюю вероятность ошибки декодирования при заданном коде x_1, x_2, \dots, x_M .

В стратегии "с отказами от декодирования" оптимальные области декодирования определяются условиями

$$D_s = \left\{ y : P_{Y|X}(y | x_s) > \max_{j \neq s, j=1, \dots, M} P_{Y|X}(y | x_j) \right\},$$

$$s = 1, 2, \dots, M.$$

Все остальные векторы y образуют область D_{M+1} (отказ от декодирования). Такое разбиение гарантирует вынесение решения в пользу кодового вектора x_j только тогда, когда он является единственным наиболее вероятным вектором при полученным векторе y . Отказ от декодирования возникает в том случае, если при принятом векторе y переходные вероятности для двух или большего числа наиболее вероятных кодовых векторов одинаковы.

8.2. Теоремы Шеннона для канала

При создании системы связи основная задача проектировщика – обеспечить малую вероятность ошибки декодирования. Если дискретный канал уже выбран, то в распоряжении проектировщика только выбор числа M передаваемых сообщений,

выбор длины кодовых векторов L и выбор оптимального кода $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$. Заслугой К. Шеннона является доказательство существования методов передачи, для которых средняя вероятность ошибочного декодирования стремится к нулю, $P_{er} \rightarrow 0$, при увеличении длины блока, $L \rightarrow \infty$, если только скорость передачи меньше пропускной способности канала, $R < C$. Это так называемая прямая теорема Шеннона для передачи по каналу. Обратная теорема Шеннона утверждает, что при $R > C$ вероятность ошибочного декодирования не стремится к нулю при любых способах передачи. Начнем с более простого доказательства обратной теоремы.

8.2.1. Обратная теорема Шеннона

Предположим, что имеется дискретный стационарный канал без памяти с переходными вероятностями $P_{Y|X}(y | x)$, у которого пропускная способность этого канала равна C .

Теорема 8.1. (*Обратная теорема Шеннона.*) *Если скорость передачи больше пропускной способности, т.е. $R > C$, то не существует таких способов передачи, при которых вероятность ошибки стремится к нулю ($P_{er} \rightarrow 0$) при увеличении длины передаваемого блока, ($L \rightarrow \infty$).*

Доказательство. В модели системы связи (рис. 8.1) будем рассматривать $\mathbf{U}, \mathbf{X}, \mathbf{Y}, \tilde{\mathbf{U}}$ как случайные векторы.

Далее предполагается, что блок \mathbf{U} составлен из K независимых равномерно распределенных двоичных случайных величин, так что \mathbf{U} принимает каждое возможное значение с вероятностью 2^{-K} . Следовательно,

$$H(\mathbf{U}) = K. \quad (8.10)$$

Вектор \mathbf{X} зависит только от текущего вектора \mathbf{U} , но не зависит от предыдущих или последующих блоков. Вектор \mathbf{Y} зависит только от текущего входного вектора \mathbf{X} , но не зависит от предыдущих или последующих входных векторов, так как

рассматривается канал без памяти. Вектор $\tilde{\mathbf{U}}$ зависит только от текущего выходного вектора \mathbf{Y} , но не зависит от предыдущих или последующих выходных векторов. Это означает, что величины $\mathbf{U}, \mathbf{X}, \mathbf{Y}, \tilde{\mathbf{U}}$ образуют простую цепь Маркова:

$$\mathbf{U} \rightarrow \mathbf{X} \rightarrow \mathbf{Y} \rightarrow \tilde{\mathbf{U}}. \quad (8.11)$$

Отсюда следует, что величины $\mathbf{U}, \mathbf{X}, \tilde{\mathbf{U}}$ и $\mathbf{X}, \mathbf{Y}, \tilde{\mathbf{U}}$ также связаны в простую цепь Маркова:

$$\mathbf{U} \rightarrow \mathbf{X} \rightarrow \tilde{\mathbf{U}}, \quad (8.12)$$

$$\mathbf{X} \rightarrow \mathbf{Y} \rightarrow \tilde{\mathbf{U}}. \quad (8.13)$$

Применим лемму 7.2 об обработке сигнала к цепи Маркова (8.12):

$$I(\tilde{\mathbf{U}}; \mathbf{U}) \leq I(\tilde{\mathbf{U}}; \mathbf{X}).$$

Применим лемму 7.2 к цепи Маркова (8.13):

$$I(\tilde{\mathbf{U}}; \mathbf{X}) \leq I(\mathbf{Y}; \mathbf{X}).$$

Объединяя эти неравенства, получаем

$$I(\tilde{\mathbf{U}}; \mathbf{U}) \leq I(\mathbf{Y}; \mathbf{X}). \quad (8.14)$$

Оценим сверху среднюю взаимную информацию между \mathbf{Y} и \mathbf{X} . По определению, $I(\mathbf{Y}; \mathbf{X}) = H(\mathbf{Y}) - H(\mathbf{Y} | \mathbf{X})$. Из цепного равенства для энтропии получаем

$$\begin{aligned} H(\mathbf{Y}) &= H(Y_1, Y_2, \dots, Y_L) \leq \\ &\leq H(Y_1) + H(Y_2) + \dots + H(Y_L) = \sum_{s=1}^L H(Y_s). \end{aligned}$$

Так как переходные вероятности для канала без памяти описываются соотношением (8.4), то для условной энтропии получаем

$$H(\mathbf{Y} \mid \mathbf{X}) = H(Y_1, Y_2, \dots, Y_L \mid X_1, X_2, \dots, X_L) = \sum_{s=1}^L H(Y_s \mid X_s).$$

Учитывая, что $I(Y_s; X_s) = H(Y_s) - H(Y_s \mid X_s) \leq C$, окончательно получаем

$$\begin{aligned} I(\mathbf{Y}; \mathbf{X}) &\leq \sum_s^L (H(Y_s) - H(Y_s \mid X_s)) = \\ &= \sum_s^L I(Y_s; X_s) \leq LC. \end{aligned} \tag{8.15}$$

Используя соотношение (8.10), найдем среднюю взаимную информацию между $\tilde{\mathbf{U}}$ и \mathbf{U} :

$$I(\tilde{\mathbf{U}}; \mathbf{U}) = H(\mathbf{U}) - H(\mathbf{U} \mid \tilde{\mathbf{U}}) = K - H(U_1 U_2 \dots U_K \mid \tilde{U}_1 \tilde{U}_2 \dots \tilde{U}_K). \tag{8.16}$$

Далее из цепного равенства получаем оценку

$$\begin{aligned} H(U_1 U_2 \dots U_K \mid \tilde{U}_1 \tilde{U}_2 \dots \tilde{U}_K) &= \\ &= H(U_1 \mid \tilde{U}_1 \tilde{U}_2 \dots \tilde{U}_K) + H(U_2 \dots U_K \mid \tilde{U}_1 \tilde{U}_2 \dots \tilde{U}_K, U_1) \leq \\ &\leq H(U_1 \mid \tilde{U}_1) + H(U_2 \dots U_K \mid \tilde{U}_2 \dots \tilde{U}_K) \leq \\ &\dots \\ &\leq H(U_1 \mid \tilde{U}_1) + H(U_2 \mid \tilde{U}_2) + \dots + H(U_L \mid \tilde{U}_K) \leq \\ &\leq h(P_{B,1}) + h(P_{B,2}) + \dots + h(P_{B,K}). \end{aligned} \tag{8.17}$$

Последнее неравенство в (8.17) следует из леммы 7.3 (леммы Фано):

$$H(U_s \mid \tilde{U}_s) \leq h(P_{B,s}),$$

где $P_{B,s} = \Pr(\tilde{U}_s \neq U_s)$ – вероятность ошибки в бите.

Комбинируя соотношения (8.14) – (8.17), получим

$$1 - \frac{C}{R} \leq \frac{1}{K} (h(P_{B,1}) + h(P_{B,2}) + \dots + h(P_{B,K})), \tag{8.18}$$

где $R = \frac{K}{L}$ – скорость передачи.

Заметим, что функция $h(p) = -p \log_2 p - (1-p) \log_2(1-p)$ вогнутая (см. Приложение, Б.2), поэтому в соответствии с основным неравенством Гёльдера (Б.3)

$$\begin{aligned} \frac{1}{K} (h(P_{B,1}) + h(P_{B,2}) + \cdots + h(P_{B,K})) &\leq \\ &\leq h\left(\frac{1}{K} \sum_{s=1}^K P_{B,s}\right) = h(P_{B,\text{av}}). \end{aligned}$$

Таким образом, средняя вероятность ошибки в бите $P_{B,\text{av}}$ удовлетворяет неравенству

$$h(P_{B,\text{av}}) \geq 1 - \frac{C}{R}. \quad (8.19)$$

Предположим теперь, что выполняется условие обратной теоремы Шеннона, т.е. скорость передачи R больше пропускной способности канала: $R > C$. Тогда неравенство (8.19) справедливо для значений $P_{B,\text{av}} \geq p_{e,\min}$, где $p_{e,\min}$ – решение уравнения $h(p_{e,\min}) = 1 - \frac{C}{R}$ (см. рис. 8.4). Это значит, что не существует

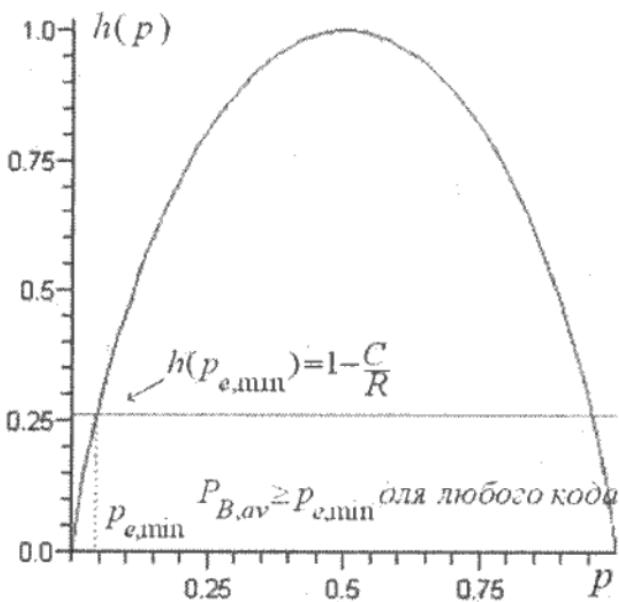


Рис. 8.4. Вероятность ошибки при превышении пропускной способности

способов передачи, при которых средняя вероятность ошибки на бит может быть сделана малой. Например, если $R = 2C$, то средняя вероятность ошибки на один бит $P_{B,\text{av}}$ не меньше

$p_{e,\min} = 0.11$ при любом выборе входных сигналов и любых методах декодирования.

Это доказывает, что при любом выборе кода $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$ средняя вероятность ошибочного декодирования не стремится к нулю ($P_{\text{er}} \not\rightarrow 0$) при увеличении длины блока ($L \rightarrow \infty$).

8.2.2. Прямая теорема Шеннона

Рассмотрим снова передачу $M = 2^{RL}$ сообщений (R – скорость передачи) по дискретному каналу без памяти с помощью кода $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$, состоящего из кодовых векторов длины L . Пусть выбраны области декодирования и правила декодирования. Пусть P_{er} – средняя вероятность ошибки декодирования, определяемая соотношением (8.6), а $P_{\text{er},\max} = \max_{1 \leq s \leq M} P_{\text{er},s}$ – максимальная ошибка декодирования.

Теорема 8.2. (Прямая теорема Шеннона.) *Если $R < C$, то существуют коды $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$ и методы декодирования такие, что средняя и максимальная вероятности ошибки декодирования стремятся к нулю, когда длина блока стремится к бесконечности, т.е. $P_{\text{er}} \rightarrow 0$, $P_{\text{er},\max} \rightarrow 0$ при $L \rightarrow \infty$.*

Будет дано 2 доказательства этой теоремы. В первом из них в качестве кодов будут выбираться случайные векторы. Области декодирования будут неоптимальными, но все же такими, что вероятности ошибок декодирования будут стремиться к нулю.

Во втором доказательстве используются оптимальные области декодирования и изучается поведение вероятностей ошибок при случайном выборе кода.

8.2.3. Первое доказательство

Задание вероятности ошибки. Задаем допустимую вероятность ошибки декодирования $\delta > 0$.

Выбор кода. Задаем распределение символов входного алфавита канала $P_X(x)$. Выбираем случайную матрицу

$X = (x_{i,j})$, $i = 1, \dots, M$; $j = 1, \dots, L$, размера $M \times L$, где все $x_{i,j}$ – случайные независимые величины с одинаковым распределением $P_X(x)$. Строки этой матрицы рассматриваем как M кодовых векторов: $\mathbf{x}_1 = (x_{1,1}, x_{1,2}, \dots, x_{1,L})$ – первый кодовый вектор и т.д. Распределение случайного вектора \mathbf{X} имеет вид

$$P_{\mathbf{X}}(\mathbf{x}) = P_{\mathbf{X}}(x_1, x_2, \dots, x_L) = P_X(x_1)P_X(x_2) \dots P_X(x_L). \quad (8.20)$$

Эпсилон-типические пары ($\{\mathbf{x}, \mathbf{y}\}$). Если X и Y – случайные вход и выход канала, то совместное распределение пары $\{X, Y\}$ равно

$$P_{\{XY\}}(x, y) = P_X(x)P_{Y|X}(y | x). \quad (8.21)$$

В этом случае говорят, что величины X и Y связаны каналом. Случайные векторы \mathbf{X} и \mathbf{Y} , связанные каналом, имеют распределение

$$P_{\{\mathbf{XY}\}}(\{\mathbf{x}, \mathbf{y}\}) = P_{\mathbf{X}}(\mathbf{x})P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}). \quad (8.22)$$

Зададим произвольное $\varepsilon > 0$.

Пара векторов ($\{\mathbf{x}, \mathbf{y}\}$) называется ε -типической для распределения $P_{\{XY\}}(\{x, y\})$, если образованная ими последовательность пар символов $\{x_s, y_s\}$, $s = 1, \dots, L$, является ε -типической для распределения $P_{\{XY\}}(\{x, y\}) = P_X(x)P_{Y|X}(y | x)$ (см. определение в соотношении (5.21)).

Из совместных распределений (8.21) и (8.22) можно найти маргинальные распределения величин Y и \mathbf{Y} :

$$\begin{aligned} P_Y(y) &= \sum_x P_X(x)P_{Y|X}(y | x), \\ P_{\mathbf{Y}}(\mathbf{y}) &= \sum_{\mathbf{x}} P_{\mathbf{X}}(\mathbf{x})P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}) = P_Y(y_1)P_Y(y_2) \dots P_Y(y_L). \end{aligned}$$

Пусть пара случайных величин ($\{X, Y\}$) и пара случайных векторов ($\{\mathbf{X}, \mathbf{Y}\}$) имеют распределения

$$\begin{aligned} Q_{XY}(x, y) &= P_X(x)P_Y(y), \\ Q_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) &= P_{\mathbf{X}}(\mathbf{x})P_{\mathbf{Y}}(\mathbf{y}). \end{aligned} \quad (8.23)$$

В этом случае величины X и Y и векторы \mathbf{X} и \mathbf{Y} не связаны каналом и независимы, но их маргинальные распределения такие же, как и у величин, связанных каналом.

Выбор областей декодирования. Пусть используется стратегия "с отказами от декодирования". Для кодового вектора \mathbf{x}_s , $s = 1, 2, \dots, M$, область декодирования определяется как

$$D_s = \left\{ \mathbf{y} : \begin{cases} \text{пара } (\{\mathbf{x}_s, \mathbf{y}\}) \text{ } \varepsilon\text{-типическая;} \\ \text{пара } (\{\mathbf{x}_j, \mathbf{y}\}) \text{ не } \varepsilon\text{-типическая, } j \neq s. \end{cases} \right\}, \quad (8.24)$$

где $s = 1, 2, \dots, M$.

Все остальные векторы \mathbf{y} , не попавшие в эти области, образуют область отказа от декодирования D_{M+1} .

Алгоритм работы декодера.

1. Пусть передавался кодовый вектор \mathbf{x}_s , а на выходе системы связи получен вектор \mathbf{y} .
2. Образуем пары векторов

$$\{\mathbf{x}_1, \mathbf{y}\}, \{\mathbf{x}_2, \mathbf{y}\}, \dots, \{\mathbf{x}_M, \mathbf{y}\}. \quad (8.25)$$

3. Проверяем пары на ε -типичность. Если для некоторого m пара $(\{\mathbf{x}_m, \mathbf{y}\})$ является ε -типической, а остальные пары не являются ε -типическими, то в этом случае выносится следующее решение: передаваемое сообщение есть \mathbf{x}_m .

Принятое решение является правильным, если названное декодером сообщение совпадает с переданным сообщением, т.е. $m = s$. Если не совпадает, то происходит ошибка декодирования.

Если не одна, а несколько пар – две или больше – являются ε -типическими, а остальные пары не являются таковыми, или все пары не являются ε -типическими, то выносится решение "отказ от декодирования".

Такой алгоритм декодирования является заведомо не оптимальным, но он дает вероятность ошибки, стремящуюся к нулю.

Вычислим вероятность правильного декодирования. Зафиксируем значение $s \neq M + 1$. Обозначим через $\mathcal{F}_{s,\varepsilon}$ событие

$\mathcal{F}_{s,\varepsilon} = \{\text{Пара } (\{\mathbf{x}_s, \mathbf{y}\}) \text{ является } \varepsilon\text{-типической}\}$

для двумерного распределения $P_X(x)P_{Y|X}(y|x)$.

Для $j \neq s$ обозначим через \mathcal{E}_j событие

$\mathcal{E}_j = \{\text{Пара } (\{\mathbf{x}_j, \mathbf{y}\}) \text{ не является } \varepsilon\text{-типической}\}$

для распределения $P_X(x)P_{Y|X}(y|x)$.

Тогда событие $\{\mathbf{y} \in D_s\}$ равно пересечению событий

$$\mathcal{F}_{s,\varepsilon} \cap \mathcal{E}_1 \cap \mathcal{E}_2 \dots \cap \mathcal{E}_{s-1} \cap \mathcal{E}_{s+1} \dots \cap \mathcal{E}_M.$$

Пусть передавался кодовый вектор \mathbf{x}_s . Вероятность правильного решения равна $P_{\text{cor},s} = \Pr(\mathbf{y} \in D_s)$, т.е.

$$P_{\text{cor},s} = \Pr(\mathcal{F}_{s,\varepsilon} \cap \mathcal{E}_1 \cap \mathcal{E}_2 \dots \cap \mathcal{E}_{s-1} \cap \mathcal{E}_{s+1} \dots \cap \mathcal{E}_M). \quad (8.26)$$

Вероятность $P_{\text{er},s}$ ошибочного декодирования или отказа при передаче вектора \mathbf{x}_s равна вероятности противоположного события и оценивается сверху таким образом

$$\begin{aligned} P_{\text{er},s} &= \Pr(\overline{\mathcal{F}_{s,\varepsilon} \cap \mathcal{E}_1 \cap \mathcal{E}_2 \dots \cap \mathcal{E}_{s-1} \cap \mathcal{E}_{s+1} \dots \cap \mathcal{E}_M}) \leq \\ &\leq \Pr(\overline{\mathcal{F}_s}) + \sum_{j \neq s} \Pr(\overline{\mathcal{E}_j}). \end{aligned} \quad (8.27)$$

Здесь использованы факты, что событие, противоположное пересечению событий, равно объединению противоположных событий, и что вероятность объединения событий не превосходит суммы вероятностей объединяемых событий.

Рассмотрим последовательность пар (8.25). В этом ряду передававшийся вектор \mathbf{x}_s и принятый вектор \mathbf{y} связаны каналом, поэтому вероятность пары $\{\mathbf{x}_s, \mathbf{y}\}$ задается соотношением (8.22), т.е.

$$P_{\{\mathbf{XY}\}}(\{\mathbf{x}_s, \mathbf{y}\}) = P_{\mathbf{X}}(\mathbf{x}_s)P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}_s).$$

Для всех других пар векторы \mathbf{x}_j и принятый вектор \mathbf{y} не связаны каналом и являются независимыми. Поэтому вероятность пары $\{\mathbf{x}_j, \mathbf{y}\}$, $j \neq s$ задается соотношением (8.23), т.е.

$$Q_{\{\mathbf{X}\mathbf{Y}\}}(\{\mathbf{x}_j, \mathbf{y}\}) = P_{\mathbf{X}}(\mathbf{x}_j)P_{\mathbf{Y}}(\mathbf{y}).$$

Согласно лемме 5.6, для любого $\delta > 0$, при достаточно больших L и достаточно малых $\varepsilon > 0$ вероятность $\Pr(\mathcal{F}_{s,\varepsilon})$ может быть сделана меньше $\delta/2$.

Согласно лемме 5.8 (см., (5.31)), каждое слагаемое $\Pr(\overline{\mathcal{E}}_j)$, где $j \neq s$, в сумме (8.27) удовлетворяют неравенству

$$\Pr(\overline{\mathcal{E}}_j) \leq 2^{-L(1-\varepsilon')I(Y;X)}.$$

Таким образом,

$$P_{\text{er},s} \leq \frac{\delta}{2} + (M-1)2^{-L(1-\varepsilon')I(Y;X)} \leq \delta/2 + 2^{-L((1-\varepsilon')I(Y;X)-R)}. \quad (8.28)$$

Если $R < I(Y;X)$, то при достаточно большом L второе слагаемое может быть сделано меньше $\delta/2$, так что

$$P_{\text{er},s} \leq \delta. \quad (8.29)$$

Эта оценка верна для любых распределений входных символов $P_X(x)$. В частности, его можно выбрать так, чтобы средняя взаимная информация $I(Y;X)$ между выходом и входом канала была сколь угодно близкой к пропускной способности канала C .

Итак, при достаточно большом L вероятность ошибочного декодирования при передаче кодового вектора может быть сделана меньше произвольно малого $\delta > 0$, если только скорость передачи меньше пропускной способности канала, т.е. $R < C$.

8.2.4. Второе доказательство

Используем теперь стратегию разбиения без отказов от декодирования (рис. 8.5) и разбиение на оптимальные области декодирования (8.9). Пространство выходных векторов канала \mathbf{Y} разбивается на области декодирования. Здесь

- $M = \lceil 2^{RL} \rceil$ – число передаваемых сообщений, равное числу кодовых слов; R – неотрицательное число, называемое скоростью передачи;

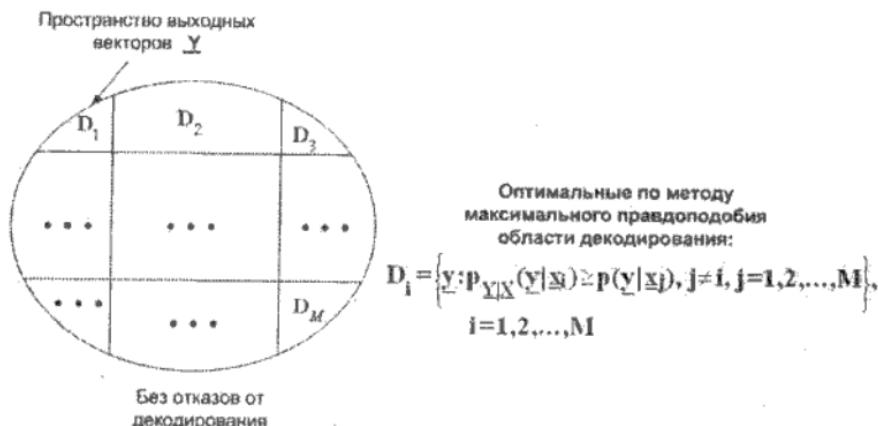


Рис. 8.5. Оптимальные области декодирования

- $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_M$ – кодовые векторы длины L ;
- $D_s, s = 1, 2, \dots, M$ – непересекающиеся области декодирования.

Вероятности ошибок задаются соотношениями (8.5) – (8.6). Их точное вычисление затруднительно из-за очень большого числа слагаемых и сложной структуры областей декодирования. Для доказательства теоремы Шеннона необходимо изучить поведение вероятностей ошибок при растущей длине кодовых блоков, $L \rightarrow \infty$. Для этого было предложено несколько методов получения простых для анализа оценок вероятностей. В теории вероятностей основными являются два класса методов. Один из них был предложен еще до создания теории информации индийским статистиком Бхаттачария (Bhattacharyya) для задач статистики. Другой метод был разработан Галлагером (R. Gallager) специально для доказательства теоремы Шеннона. Эти методы дополняют друг друга.

В Главе 9 изложен метод Бхаттачария, а в Главе 10 – метод Галлагера. Их комбинация приводит не только к доказательству теоремы Шеннона, но и к оценкам вероятностей ошибок.

Г л а в а 9

Граница Бхаттачария

9.1. Код из двух векторов

Сначала рассмотрим случай, когда имеется всего два кодовых слова ($M = 2$) и длина каждого из векторов равна L :

$$\begin{aligned} \mathbf{x}_1 &= (x_{11}, x_{12}, \dots, x_{1L}); \\ \mathbf{x}_2 &= (x_{21}, x_{22}, \dots, x_{2L}). \end{aligned} \quad (9.1)$$

Координаты кодовых векторов x_{ij} , $i = 1, 2$; $j = 1, 2, \dots, L$, принадлежат входному алфавиту $\mathcal{U} = \{u_1, u_2, \dots, u_K\}$. На выходе канала принимается вектор $\mathbf{y} = (y_1, y_2, \dots, y_L)$ с координатами из выходного алфавита $\mathcal{V} = \{v_1, v_2, \dots, v_D\}$.

Вероятности ошибок декодирования для первого и второго сообщений равны

$$\begin{aligned} \Pr_{\text{er},1} &= \sum_{\mathbf{y} \in D_1^c} P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x}_1) = \sum_{\mathbf{y} \in D_2} P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x}_1); \\ \Pr_{\text{er},2} &= \sum_{\mathbf{y} \in D_2^c} P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x}_2) = \sum_{\mathbf{y} \in D_1} P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x}_2). \end{aligned} \quad (9.2)$$

Введем характеристические функции множеств D_1^c и D_2^c :

$$\begin{aligned} \varphi_1(\mathbf{y}) &= \begin{cases} 1, & \text{если } \mathbf{y} \in D_1^c; \\ 0, & \text{если } \mathbf{y} \in D_1; \end{cases} \\ \varphi_2(\mathbf{y}) &= \begin{cases} 1, & \text{если } \mathbf{y} \in D_2^c; \\ 0, & \text{если } \mathbf{y} \in D_2. \end{cases} \end{aligned} \quad (9.3)$$

Эти функции определены на всем пространстве выходных последовательностей \mathbf{Y} . С их помощью соотношение (9.2) можно записать в виде

$$\begin{aligned} \Pr_{\text{er},1} &= \sum_{y \in Y} P_{Y|X}(y | x_1) \varphi_1(y); \\ \Pr_{\text{er},2} &= \sum_{y \in Y} P_{Y|X}(y | x_2) \varphi_2(y). \end{aligned} \quad (9.4)$$

Заметим, что суммирование в (9.4) распространено на все выходное пространство.

Пусть теперь $c_1(y)$ и $c_2(y)$ – некоторые функции, определенные на Y , причем

$$\begin{aligned} c_1(y) &\geq \varphi_1(y), \forall y \in Y; \\ c_2(y) &\geq \varphi_2(y), \forall y \in Y. \end{aligned} \quad (9.5)$$

Тогда верны оценки сверху:

$$\begin{aligned} \Pr_{\text{er},1} &= \sum_{y \in Y} P_{Y|X}(y | x_1) \varphi_1(y) \leq \sum_{y \in Y} P_{Y|X}(y | x_1) c_1(y); \\ \Pr_{\text{er},2} &= \sum_{y \in Y} P_{Y|X}(y | x_2) \varphi_2(y) \leq \sum_{y \in Y} P_{Y|X}(y | x_2) c_2(y). \end{aligned} \quad (9.6)$$

В методе Бхаттачария для оценки вероятностей выбираются функции

$$\begin{aligned} c_1(y) &= \frac{P_{Y|X}(y|x_2)^{\frac{1}{2}}}{P_{Y|X}(y|x_1)^{\frac{1}{2}}} = \sqrt{\frac{P_{Y|X}(y|x_2)}{P_{Y|X}(y|x_1)}}; \\ c_2(y) &= \frac{P_{Y|X}(y|x_1)^{\frac{1}{2}}}{P_{Y|X}(y|x_2)^{\frac{1}{2}}} = \sqrt{\frac{P_{Y|X}(y|x_1)}{P_{Y|X}(y|x_2)}}. \end{aligned} \quad (9.7)$$

Они удовлетворяют необходимым условиям (9.5). Действительно, в области D_1 имеем по определению $\varphi_1(y) = 0 \leq c_1(y)$. В области $D_1^c = D_2$ имеем по определению оптимальной области D_2

$$c_1(y) = \sqrt{\frac{P_{Y|X}(y | x_2)}{P_{Y|X}(y | x_1)}} \geq 1 = \varphi_1(y).$$

Аналогично проверяются условия для $c_2(y)$.

Подставляя (9.7) в (9.6), получим следующую оценку:

$$\begin{aligned} \Pr_{\text{er},1} &= \sum_{y \in Y} P_{Y|X}(y | x_1) \varphi_1(y) \leq \sum_{y \in Y} \sqrt{P_{Y|X}(y | x_1) P_{Y|X}(y | x_2)}; \\ \Pr_{\text{er},2} &= \sum_{y \in Y} P_{Y|X}(y | x_2) \varphi_2(y) \leq \sum_{y \in Y} \sqrt{P_{Y|X}(y | x_1) P_{Y|X}(y | x_2)}. \end{aligned} \quad (9.8)$$

Как видно из (9.8), верхние оценки для обоих сообщений совпадают.

Так как мы рассматриваем канал без памяти, то

$$P_{Y|X}(y | x_i) = P_{Y|X}(y_1 | x_{i1}) P_{Y|X}(y_2 | x_{i2}) \dots P_{Y|X}(y_L | x_{iL}),$$

где $i = 1, 2$.

Подставляя эти соотношения в (9.8), получаем

$$\begin{aligned} \Pr_{\text{er},i} &\leq \sum_{y \in Y} \sqrt{P_{Y|X}(y_1 | x_{11}) P_{Y|X}(y_1 | x_{21})} \times \\ &\quad \times \sqrt{P_{Y|X}(y_2 | x_{12}) P_{Y|X}(y_2 | x_{22})} \times \dots \\ &\quad \dots \times \sqrt{P_{Y|X}(y_L | x_{1L}) P_{Y|X}(y_L | x_{2L})}. \end{aligned} \quad (9.9)$$

В (9.9) каждое слагаемое является произведением L сомножителей, каждый из которых зависит только от одной координаты вектора y . Это позволяет перейти от суммирования по всем векторам $y \in Y$ к L -кратному суммированию по каждой координате $y_j \in \mathcal{V}$. В результате "сумма произведений" преобразуется в "произведение сумм":

$$\begin{aligned} \Pr_{\text{er},i} &\leq \left(\sum_{y \in \mathcal{V}} \sqrt{P_{Y|X}(y | x_{11}) P_{Y|X}(y | x_{21})} \right) \times \\ &\quad \times \left(\sum_{y \in \mathcal{V}} \sqrt{P_{Y|X}(y | x_{12}) P_{Y|X}(y | x_{22})} \right) \times \dots \\ &\quad \dots \times \left(\sum_{y \in \mathcal{V}} \sqrt{P_{Y|X}(y | x_{1L}) P_{Y|X}(y | x_{2L})} \right). \end{aligned} \quad (9.10)$$

Определим *расстояние Бхаттачаряя* между парами символов $x_1 \in \mathcal{U}$ и $x_2 \in \mathcal{U}$ входного алфавита с помощью формулы

$$D_{Bh}(x_1, x_2) = -\log_2 \sum_y \sqrt{P_{Y|X}(y | x_1) P_{Y|X}(y | x_2)}. \quad (9.11)$$

С помощью неравенства Коши–Буняковского убеждаемся, что $D_{Bh}(x_1, x_2) \geq 0$. Кроме того, расстояние между одинаковыми символами $x_1 = x_2$ равно $D_{Bh}(x_1, x_1) = 0$.

Расстояние Бхаттачария между двумя векторами

$$\mathbf{x}_1 = (x_{11}, x_{12}, \dots, x_{1L}) \text{ и } \mathbf{x}_2 = (x_{21}, x_{22}, \dots, x_{2L})$$

определяется как сумма расстояний между соответствующими координатами:

$$D_{Bh}(\mathbf{x}_1, \mathbf{x}_2) = \sum_{j=1}^L D_{Bh}(x_{1j}, x_{2j}). \quad (9.12)$$

С учетом этих определений можно записать

$$\sum_{\mathbf{y} \in \mathbf{Y}} \sqrt{P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}_1) P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}_2)} = 2^{-D_{Bh}(\mathbf{x}_1, \mathbf{x}_2)} \quad (9.13)$$

и оценку (9.10) переписать в виде

$$\Pr_{er,i} \leq 2^{-D_{Bh}(\mathbf{x}_1, \mathbf{x}_2)}, \quad i = 1, 2. \quad (9.14)$$

Это соотношение имеет простой физический смысл: чем больше расстояние Бхаттачария между кодовыми векторами, тем меньше вероятности ошибочного декодирования.

Для оптимального кода расстояние между векторами должно быть максимально большим. Построить такой код нетрудно. Пусть α и β – наиболее удаленные в смысле расстояния Бхаттачария символы:

$$D_{Bh}(\alpha, \beta) = D_{Bh,\max} = \max_{x_1, x_2} D_{Bh}(x_1, x_2).$$

Тогда оптимальным будет код, состоящий из векторов

$\mathbf{x}_1 = (\alpha, \alpha, \dots, \alpha)$ и $\mathbf{x}_2 = (\beta, \beta, \dots, \beta)$,

а оценка Бхаттачария примет вид

$$\Pr_{\text{er},i} \leq 2^{-LD_{Bh,\max}}, \quad i = 1, 2. \quad (9.15)$$

Таким образом, вероятность ошибочного декодирования убывает экспоненциально с ростом длины кодовых блоков L .

Пример 9.1. Пусть входной алфавит – двоичный, $\mathcal{U} = \{0, 1\}$. Тогда расстояние Бхаттачария между 0 и 1 равно

$$D_{Bh}(0, 1) = -\log_2 \sum_y \sqrt{P_{Y|X}(y | 0)P_{Y|X}(y | 1)}.$$

Для двоичного симметричного канала условные вероятности $P_{Y|X}(y = 1 | 0)$ и $P_{Y|X}(y = 0 | 1)$ равны p . В этом случае расстояние Бхаттачария равно $D_{Bh}(0, 1) = -\frac{\log_2(4p(1-p))}{2}$.

Пусть кодовые векторы длины L равны $\mathbf{x}_1 = (0, 0, \dots, 0)$ и $\mathbf{x}_2 = (1, 1, \dots, 1)$. Тогда верхняя граница для вероятности ошибки этого кода имеет вид:

$$\Pr_{\text{er},i} \leq 2^{-D_{Bh}(0,1)L}, \quad i = 1, 2.$$

9.2. Случайное кодирование

Хотя код из двух векторов всегда можно выбрать оптимальным образом, мы рассмотрим случайный выбор кода. Его характеристики будут использованы в общем случае. Пусть кодовые векторы являются случайными функциями:

$$\begin{aligned} \mathbf{X}_1 &= (X_{11}, X_{12}, \dots, X_{1L}); \\ \mathbf{X}_2 &= (X_{21}, X_{22}, \dots, X_{2L}). \end{aligned} \quad (9.16)$$

Координаты X_{ij} , $i = 1, 2$; $j = 1, 2, \dots, L$, случайных кодовых векторов являются независимыми случайными величинами с одинаковым распределением $P_X(x)$. Тогда все характеристики (области декодирования, вероятности ошибочного декодирования и др.) являются случайными величинами или функциями. В частности, обе части оценки Бхаттачария в (9.10):

$$\Pr_{\text{er},i} \leq \left(\sum_{y \in \mathcal{V}} \sqrt{P_{Y|X}(Y | X_{11})P_{Y|X}(Y | X_{21})} \right) \times \\ \times \left(\sum_{y \in \mathcal{V}} \sqrt{P_{Y|X}(Y | X_{12})P_{Y|X}(Y | X_{22})} \right) \times \dots \quad (9.17) \\ \dots \times \left(\sum_{y \in \mathcal{V}} \sqrt{P_{Y|X}(Y | X_{1L})P_{Y|X}(Y | X_{2L})} \right).$$

являются случайными величинами. Найдем их средние значения. В правой части сомножители являются независимыми случайными величинами с одинаковым распределением и одинаковыми средними значениями. Среднее значение, скажем, первого сомножителя равно

$$E \left(\sum_{y \in \mathcal{V}} \sqrt{P_{Y|X}(Y | X_{11})P_{Y|X}(Y | X_{21})} \right) = \\ = \sum_{y \in \mathcal{V}} E \left(\sqrt{P_{Y|X}(Y | X_{11})} \right) E \left(\sqrt{P_{Y|X}(Y | X_{21})} \right) = \\ = \sum_{y \in \mathcal{V}} \left(\sum_{x \in \mathcal{U}} P_X(x) \sqrt{P_{Y|X}(y | x)} \right)^2 = 2^{-R_0(P_X)},$$

где через $R_0(P_X)$ обозначена величина

$$R_0(P_X) = -\log_2 \left(\sum_{y \in \mathcal{V}} \left(\sum_{x \in \mathcal{U}} P_X(x) \sqrt{P_{Y|X}(y | x)} \right)^2 \right), \quad (9.18)$$

называемая экспонентой Бхаттачария для входного распределения $P_X(x)$. С помощью неравенства Коши–Буняковского можно показать, что $R_0(P_X) \geq 0$.

Учитывая, что средние значения всех сомножителей одинаковы, окончательно получим оценку

$$\overline{\Pr}_{\text{er},i} = E_{\mathbf{X}_1, \mathbf{X}_2} \Pr_{\text{er},i} \leq E_{\mathbf{X}_1, \mathbf{X}_2} 2^{-D_{\text{BH}}(\mathbf{X}_1, \mathbf{X}_2)} = 2^{-LR_0(P_X)}, \quad i = 1, 2. \quad (9.19)$$

Выберем входное распределение, максимизирующее экспоненту:

$$R_0 = \max_{P_X} \left(-\log_2 \left(\sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{U}} P_X(x) \sqrt{P_{Y|X}(y|x)} \right)^2 \right) \right). \quad (9.20)$$

Величина R_0 называется оптимальной экспонентой Бхаттачария.

Таким образом, для случайного кода из двух слов достижима оценка

$$\overline{Pr}_{er,i} \leq 2^{-LR_0}, \quad i = 1, 2. \quad (9.21)$$

9.3. Код из $M = 2^{RL}$ векторов

Перейдем к произвольным значениям $M = 2^{LR}$, где R – скорость передачи.

Пусть код состоит из M кодовых векторов $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M$ длины L .

Пусть выбраны оптимальные по максимальному правдоподобию области декодирования D_i , $i = 1, 2, \dots, M$.

Пусть для передачи выбран кодовый вектор \mathbf{x}_s . Тогда вероятность ошибочного декодирования этого вектора равна

$$Pr_{er,s} = \sum_{y \in D_s^c} P_{Y|X}(y|\mathbf{x}_s) = \sum_{y \in Y} P_{Y|X}(y|\mathbf{x}_s) \varphi_s(y), \quad (9.22)$$

где для распространения суммирования на все выходное пространство снова использована характеристическая функция множества D_s^c , дополнительного к D_s :

$$\varphi_s(y) = \begin{cases} 1, & \text{если } y \in D_s^c; \\ 0, & \text{если } y \in D_s. \end{cases} \quad (9.23)$$

Выберем функцию $c_s(y)$ в виде

$$c_s(y) = \frac{\sum_{j=1, j \neq s}^M P_{Y|X}(y|\mathbf{x}_j)^{\frac{1}{2}}}{P_{Y|X}(y|\mathbf{x}_s)^{\frac{1}{2}}}. \quad (9.24)$$

Эта функция удовлетворяет условию $c_s(\mathbf{y}) \geq \varphi_s(\mathbf{y}), \forall \mathbf{y}$, так как в области D_s^c в числителе (9.24) хотя бы одно слагаемое не меньше знаменателя. Действительно, если бы для некоторого \mathbf{y} каждое слагаемое в числителе было бы меньше знаменателя, то такой вектор был бы отнесен в множество D_s , а не в множество D_s^c . Поэтому из (9.22) следует оценка

$$\begin{aligned} \text{Pr}_{\text{er},s} &\leq \sum_{\mathbf{y} \in \mathbf{Y}} P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}_s) \frac{\sum_{j=1, j \neq s}^M P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}_j)^{\frac{1}{2}}}{P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}_s)^{\frac{1}{2}}} = \\ &= \sum_{j=1, j \neq s}^M \sum_{\mathbf{y} \in \mathbf{Y}} \sqrt{P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}_j) P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}_s)} = \\ &= \sum_{j=1, j \neq s}^M 2^{-D_{Bh}(\mathbf{x}_j, \mathbf{x}_s)}, \end{aligned} \quad (9.25)$$

где использовано соотношение (9.13).

Рассмотрим теперь случайный код $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M$. Координаты $X_{ij}, i = 1, 2, \dots, M; j = 1, 2, \dots, L$, случайных кодовых векторов являются независимыми случайными величинами с одинаковым распределением $P_X(x)$.

Усредняя по этому распределению оценку (9.25) и используя оценку (9.19), получим

$$\begin{aligned} E_{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M} \text{Pr}_{\text{er},s} &\leq E_{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M} \sum_{j=1, j \neq s}^M 2^{-D_{Bh}(\mathbf{x}_j, \mathbf{x}_s)} = \\ &= (M-1) 2^{-L R_0(P_X)} \leq \\ &\leq 2^{RL} \cdot 2^{-L R_0(P_X)} = 2^{-L(R_0(P_X) - R)}. \end{aligned} \quad (9.26)$$

В качестве $P_X(x)$ следует выбирать распределение, максимизирующее экспоненту Бхаттачария $R_0(P_X)$. Тогда оценка примет вид

$$E_{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M} \text{Pr}_{\text{er},s} \leq 2^{-L(R_0 - R)}, \quad (9.27)$$

где R_0 задается формулами (9.18), (9.20).

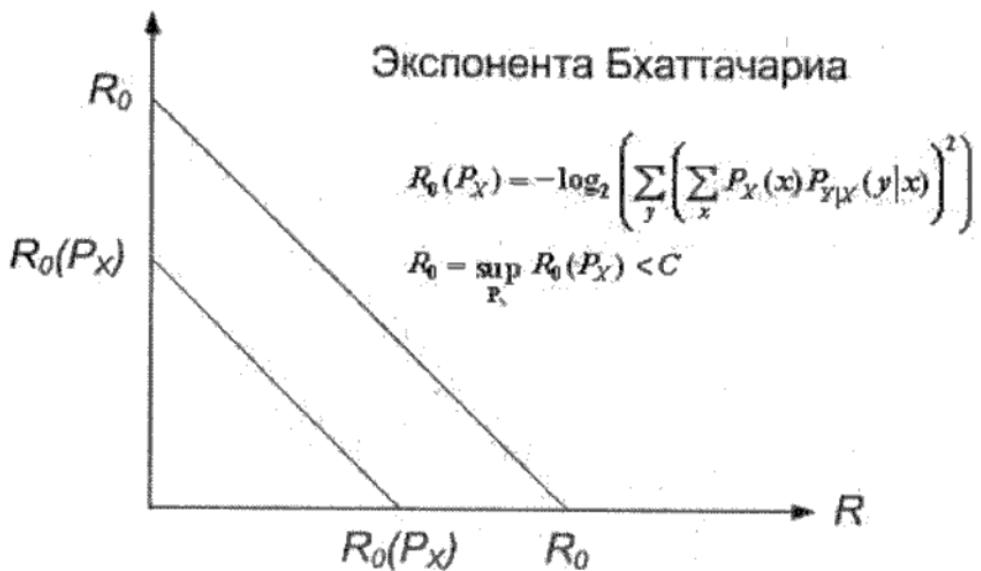


Рис. 9.1. Зависимость экспоненты Бхаттачария вероятности ошибки от скорости передачи

На рис. 9.1 показана зависимость экспоненты вероятности ошибки от скорости передачи. Она положительна при $R < R_0$. При этих скоростях вероятность ошибочного декодирования экспоненциально убывает с ростом длины блока L .

Так как в общем случае $R_0 < C$, то доказательство теоремы Шеннона еще не завершено.

Г л а в а 10

Граница Галлагера

10.1. Вывод основного неравенства

Начальный этап анализа не отличается от вывода границы Бхаттачария. Рассматривается передача $M = 2^{LR}$ сообщений, где R – скорость передачи. Выбирается код $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M$, состоящий из M кодовых векторов длины L .

Пусть выбраны оптимальные по максимальному правдоподобию области декодирования $D_i, i = 1, 2, \dots, M$.

Предположим, что для передачи выбран кодовый вектор \mathbf{x}_s . Тогда вероятность ошибочного декодирования этого вектора равна

$$\Pr_{\text{er},s} = \sum_{\mathbf{y} \in D_s^c} P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}_s) = \sum_{\mathbf{y} \in \mathbf{Y}} P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}_s) \varphi_s(\mathbf{y}), \quad (10.1)$$

где для распространения суммирования на все выходное пространство снова использована характеристическая функция множества D_s^c , дополнительного к D_s :

$$\varphi_s(\mathbf{y}) = \begin{cases} 1, & \text{если } \mathbf{y} \in D_s^c; \\ 0, & \text{если } \mathbf{y} \in D_s. \end{cases} \quad (10.2)$$

Отличие от метода Бхаттачария заключается в выборе функции $c_s(\mathbf{y})$. В методе Галлагера эта функция равна

$$c_s(\mathbf{y}) = \left(\frac{\sum_{j=1, j \neq s}^M P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}_j)^{\frac{1}{1+\rho}}}{P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}_s)^{\frac{1}{1+\rho}}} \right)^{\rho}, \quad (10.3)$$

где ρ – вещественный параметр, который далее будет выбираться из промежутка $0 \leq \rho \leq 1$.

Эта функция удовлетворяет условию $c_s(\mathbf{y}) \geq \varphi_s(\mathbf{y}), \forall \mathbf{y}$, так как в области D_s^c в числителе (10.3) хотя бы одно слагаемое не меньше знаменателя. Действительно, если бы для некоторого u каждое слагаемое в числителе было бы меньше знаменателя, то такой вектор был бы отнесен в множество D_s , а не в множество D_s^c .

Подставив в соотношение (10.1) функцию $c_s(\mathbf{y})$ вместо $\varphi_s(\mathbf{y})$, получим после элементарных преобразований оценку

$$\Pr_{\text{er},s} \leq \sum_{\mathbf{y} \in \mathbf{Y}} P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}_s)^{\frac{1}{1+\rho}} \left(\sum_{j=1, j \neq s}^M P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}_j)^{\frac{1}{1+\rho}} \right)^\rho. \quad (10.4)$$

Полученная граница называется границей Галлагера.

Рассмотрим теперь *случайный* код $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M$. Пусть координаты X_{ij} , $i = 1, 2, \dots, M; j = 1, 2, \dots, L$, случайных кодовых векторов являются независимыми случайными величинами с одинаковым распределением $P_X(x)$. Тогда случайные кодовые векторы будут независимы и иметь одинаковое распределение

$$P_{\mathbf{X}}(\mathbf{x}) = P_X(x_1)P_X(x_2)\dots P_X(x_L),$$

где $\mathbf{x} = (x_1, x_2, \dots, x_L)$ – значение кодового вектора.

Найдем среднее значение оценки (10.1). Под знаком первой суммы в правой части стоит произведение двух множителей, первый из которых зависит только от случайного вектора \mathbf{X}_s , а второй зависит от остальных случайных кодовых векторов, но не зависит от \mathbf{X}_s . Поэтому усреднение этих множителей можно проводить отдельно. Обозначив среднее значение первого множителя через $f(\mathbf{y}, \rho)$, получим

$$f(\mathbf{y}, \rho) = E_{\mathbf{X}_s} P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}_s)^{\frac{1}{1+\rho}} = \sum_{\mathbf{x}} P_{\mathbf{X}}(\mathbf{x}) P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x})^{\frac{1}{1+\rho}}. \quad (10.5)$$

Второй множитель имеет вид Z^ρ , где

$$Z = \sum_{j=1, j \neq s}^M P_{Y|X}(y | X_j)^{\frac{1}{1+\rho}}.$$

Так как функция Z^ρ является вогнутой в интервале $Z \geq 0$ при $\rho \leq 1$, то в соответствии с неравенством Гёльдера–Иенсена (см. Приложение, (Б.7)) имеем

$$\begin{aligned} E(Z^\rho) &\leq (EZ)^\rho = \\ &= \left(\sum_{j=1, j \neq s} E_{X_j} P_{Y|X}(y | X_j)^{\frac{1}{1+\rho}} \right)^\rho = \\ &= \left(\sum_{j=1, j \neq s} f(y, \rho) \right)^\rho = \\ &= (M-1)^\rho (f(y, \rho))^\rho. \end{aligned} \tag{10.6}$$

Таким образом, усреднение оценки (10.4) приводит к следующей верхней границе для средней вероятности ошибочного декодирования:

$$\begin{aligned} \overline{Pr}_{er,s} &= EPr_{er,s} \leq \sum_{y \in Y} (M-1)^\rho (f(y, \rho))^{1+\rho} \leq \\ &\leq M^\rho \sum_{y \in Y} \left(\sum_x P_X(x) P_{Y|X}(y | x)^{\frac{1}{1+\rho}} \right)^{1+\rho}. \end{aligned} \tag{10.7}$$

Если в (10.7) перейти от суммирования по всем возможным векторам x и y к $2L$ -кратному суммированию по всем координатам и учесть, что все слагаемые можно представить как по-координатные произведения, то оценку можно преобразовать к виду

$$\begin{aligned} \overline{Pr}_{er} &\leq M^\rho \left\{ \sum_{y \in V} \left(\sum_{x \in U} P_X(x) P_{Y|X}(y | x)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right\}^L = \\ &= 2^{-L(E_0(\rho, P_X) - \rho R)}, \end{aligned} \tag{10.8}$$

где введена функция Галлагера:

$$E_0(\rho, P_X) = -\log_2 \sum_{y \in \mathcal{V}} \left(\sum_{x \in \mathcal{U}} P_X(x) P_{Y|X}(y | x)^{\frac{1}{1+\rho}} \right)^{1+\rho}. \quad (10.9)$$

Выражение

$$E_G(R, \rho, P_X) = E_0(\rho, P_X) - \rho R \quad (10.10)$$

называется экспонентой вероятности ошибки.

Максимум этого выражения по ρ

$$E_G(R, P_X) = \max_{0 \leq \rho \leq 1} \{E_0(\rho, P_X) - \rho R\} \quad (10.11)$$

называется субоптимальной экспонентой Галлагера.

Выражение, оптимизированное как по ρ , так и по распределению $P_X(x)$,

$$E_G(R) = \sup_{P_X} \max_{0 \leq \rho \leq 1} \{E_0(\rho, P_X) - \rho R\} \quad (10.12)$$

называется оптимальной экспонентой Галлагера.

Окончательно верхняя граница для вероятности ошибочного декодирования имеет вид

$$\overline{\Pr_{\text{er}}} \leq 2^{-LE_G(R)}. \quad (10.13)$$

10.2. Функция Галлагера и экспонента вероятности ошибки

Перечислим основные свойства функции Галлагера (10.9).

- Для любых распределений $P_X(x)$

$$E_0(\rho, P_X)|_{\rho=0} = 0.$$

Доказательство. Подставим значение $\rho = 0$ в формулу (10.9). Получим соотношение

$$\begin{aligned} E_0(\rho, P_X)|_{\rho=0} &= -\log_2 \sum_{y \in \mathcal{V}} \left(\sum_{x \in \mathcal{U}} P_X(x) P_{Y|X}(y|x) \right) = \\ &= -\log_2 \sum_{y \in \mathcal{V}} P_Y(y) = -\log_2 1 = 0. \end{aligned}$$

2. Для любых распределений $P_X(x)$ и $\rho > 0$

$$E_0(\rho, P_X) \begin{cases} > 0, & \text{если } C > 0; \\ = 0, & \text{если } C = 0. \end{cases}$$

Доказательство. Запишем функцию Галлагера в виде

$$E_0(\rho, P_X) =$$

$$= -\log_2 \sum_{y \in \mathcal{V}} \left(\sum_{x \in \mathcal{U}} P_X(x)^{\frac{\rho}{1+\rho}} (P_X(x) P_{Y|X}(y|x))^{\frac{1}{1+\rho}} \right)^{1+\rho}.$$

Применим неравенство Коши–Гельдера (Б.6) к внутренней сумме, полагая $\frac{1}{p} = \frac{\rho}{1+\rho}$, $\frac{1}{q} = \frac{1}{1+\rho}$:

$$\begin{aligned} &\sum_{x \in \mathcal{U}} P_X(x)^{\frac{\rho}{1+\rho}} (P_X(x) P_{Y|X}(y|x))^{\frac{1}{1+\rho}} \leq \\ &\leq \left\{ \sum_{x \in \mathcal{U}} P_X(x) \right\}^{\frac{\rho}{1+\rho}} \left\{ \sum_{x \in \mathcal{U}} P_X(x) P_{Y|X}(y|x) \right\}^{\frac{1}{1+\rho}} = \\ &= P_Y(y)^{\frac{1}{1+\rho}}. \end{aligned}$$

Следовательно,

$$E_0(\rho, P_X) \geq -\log_2 \sum_{y \in \mathcal{V}} P_Y(y) = -\log_2 1 = 0.$$

Равенство будет достигаться только в случае, если для всех x выполняются равенства $P_X(x) P_{Y|X}(y|x) = c P_X(x)$, где c – некоторая константа. Это значит, что $P_{Y|X}(y|x) = c$, т.е. строки матрицы переходных вероятностей $P_{Y|X}(y|x)$ одинаковы. Другими словами, выход канала не зависит от входа, т.е. его пропускная способность равна $C = 0$.

3. Производная функции Галлагера по ρ в точке $\rho = 0$ равна средней взаимной информации $I(Y; X)$ между выходом и входом канала для входного распределения $P_X(x)$:

$$\frac{\partial E_0(\rho, P_X)}{\partial \rho} \Big|_{\rho=0} = I(Y; X) = H(Y) - H(Y | X). \quad (10.14)$$

Доказательство. Используя при малых ρ приближения $A^{\frac{1}{1+\rho}} \approx A - \rho A \ln A$, $B^{1+\rho} \approx B + \rho B \ln B$, приближенно получаем

$$\begin{aligned} & \sum_{y \in \mathcal{V}} \left(\sum_{x \in \mathcal{U}} P_X(x) P_{Y|X}(y | x)^{\frac{1}{1+\rho}} \right)^{1+\rho} \approx \\ & \approx \sum_{y \in \mathcal{V}} \left(\sum_{x \in \mathcal{U}} P_X(x) P_{Y|X}(y | x) (1 - \rho \ln P_{Y|X}(y | x)) \right)^{1+\rho} \approx \\ & \approx \sum_{y \in \mathcal{V}} \left(P_Y(y) - \rho \sum_{x \in \mathcal{U}} P_X(x) P_{Y|X}(y | x) \ln P_{Y|X}(y | x) \right)^{1+\rho} \approx \\ & \approx \sum_{y \in \mathcal{V}} (P_Y(y) - \rho \sum_{x \in \mathcal{U}} P_X(x) P_{Y|X}(y | x) \ln P_{Y|X}(y | x)) \times \\ & \quad \times (1 + \rho \ln P_Y(y)) \approx \\ & \approx (1 + \rho \sum_{y \in \mathcal{V}} P_Y(y) \ln P_Y(y) - \\ & \quad - \rho \sum_{y \in \mathcal{V}} \sum_{x \in \mathcal{U}} P_X(x) P_{Y|X}(y | x) \ln P_{Y|X}(y | x)) = \\ & = (1 - \rho(H(Y) - H(Y | X))). \end{aligned} \quad (10.15)$$

Отсюда следует, что

$$E_0(\rho, P_X) = -\log_2(1 - \rho(H(Y) - H(Y | X))) \quad (10.16)$$

и соотношение (10.14).

4. Производная функции Галлагера по ρ – положительная в интервале $[0, 1]$ убывающая функция, причем

$$\frac{\partial^2 E_0(\rho, P_X)}{\partial \rho^2} \leq 0. \quad (10.17)$$

Перейдем к изучению свойств экспоненты Галлагера. Параметр $0 \leq \rho \leq 1$ и распределение $P_X(x)$ находятся в нашем распоряжении. Сначала изучим максимум экспоненты вероятности

ошибки по параметру ρ , т.е. субоптимальную экспоненту Галлагера $E_G(R, P_X)$.

Приравнивая к нулю производную экспоненты вероятности ошибки (10.10), найдем уравнение для точки экстремума:

$$\frac{\partial E_0(\rho, P_X)}{\partial \rho} - R = 0. \quad (10.18)$$

Выясним, когда это уравнение имеет решение.

На рис. 10.1 изображен график первой производной $\frac{\partial E_0(\rho, P_X)}{\partial \rho}$ как функции ρ в интервале $[0, 1]$. Как следует из рисунка, единственная точка экстремума определяется уравнением (10.18) в том случае, если скорость передачи R расположена в интервале $[R_c(P_X), I(Y; X)]$, где величина $R_c(P_X)$, называемая *критической скоростью* для входного распределения P_X , определяется формулой

$$R_c(P_X) = \frac{\partial E_0(\rho, P_X)}{\partial \rho} \Big|_{\rho=1}. \quad (10.19)$$

Эта точка соответствует максимуму в силу свойства (10.17).

Зависимость субоптимальной экспоненты от скорости передачи в этом диапазоне удобно описывать параметрической кривой

$$E_G(R, P_X) = E_0(\rho, P_X) - \rho \frac{\partial E_0(\rho, P_X)}{\partial \rho};$$

$$R = \frac{\partial E_0(\rho, P_X)}{\partial \rho}, \quad 0 \leq \rho \leq 1. \quad (10.20)$$

При скоростях $R \leq R_c(P_X)$ максимум экспоненты вероятности ошибки достигается в точке $\rho = 1$ и равен

$$E_G(R, P_X) = E_0(1, P_X) - R. \quad (10.21)$$

Заметим, что величина

$$E_0(1, P_X) = -\log_2 \sum_{y \in \mathcal{V}} \left(\sum_{x \in \mathcal{U}} P_X(x) P_{Y|X}(y | x)^{\frac{1}{2}} \right)^2 = R_0(P_X) \quad (10.22)$$

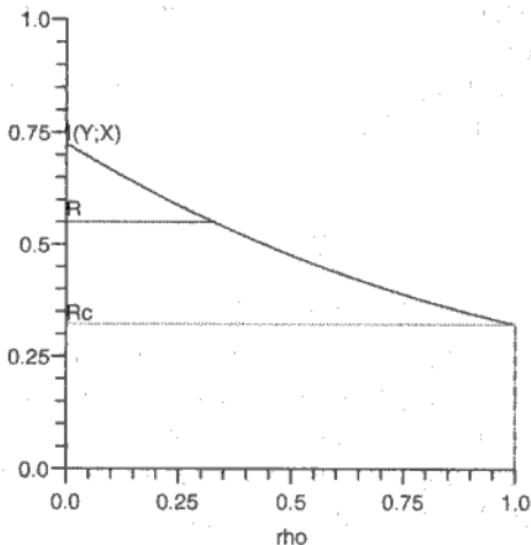


Рис. 10.1. Первая производная $\frac{\partial E_0(\rho, q_x)}{\partial \rho}$ как функция ρ

совпадает с экспонентой Бхаттачария $R_0(P_X)$ из (9.18) для распределения $P_X(x)$.

На рис. 10.2 приведена субоптимальная экспонента Галлагера для некоторого распределения входных символов $P_X(x)$.

При $R \leq R_c(P_X)$ субоптимальная экспонента Галлагера представляет собой прямую линию $R_0(P_X) - R$, совпадающую с границей Бхаттачария. При $R_c(P_X) \leq R$ она идет выше границы Бхаттачария и остается положительной вплоть до значения, равного взаимной информации $I(Y; X)$ между выходом и входом канала.

Оптимизация при каждом значении R по входному распределению $P_X(x)$ приводит к оптимальной экспоненте Галлагера, называемой также экспонентой случайного кодирования (рис. 10.3).

Оптимальная экспонента Галлагера также состоит из двух частей. Прямолинейный участок совпадает с соответствующей

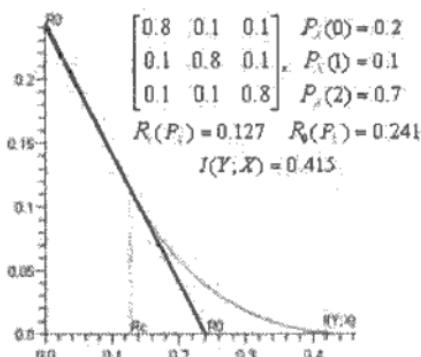


Рис. 10.2. Субоптимальная экспонента Галлагера как функция скорости кода R

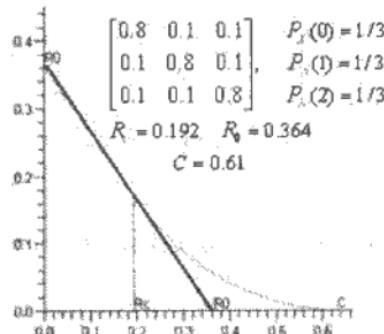


Рис. 10.3. Оптимальная экспонента Галлагера как функция скорости кода R

оптимальной границей Бхаттачария вплоть до критической скорости

$$R_c = \sup_{P_X} \frac{\partial E_0(\rho, P_X)}{\partial \rho} \Big|_{\rho=1}.$$

Далее она идет выше границы Бхаттачария и остается положительной вплоть до значения, равного

$$\sup_{P_X} I(Y; X) = C,$$

т.е. до пропускной способности канала.

Проведенные исследования оценок вероятности ошибочного декодирования позволяют утверждать, что при скоростях передачи $0 \leq R \leq C$ вероятность ошибки может быть сделана сколь угодно малой за счет увеличения длины передаваемых сообщений L . Это завершает доказательство теоремы Шеннона для канала с шумом.

Пример 10.1. Приведем расчетные формулы для двоичного симметричного канала с матрицей переходных вероятностей

$$P_{Y|X} = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}, \quad 0 < p < \frac{1}{2}, \quad (10.23)$$

где p – вероятность ошибки при передаче бита.

Здесь удобно вместо параметра ρ использовать параметр

$$\delta = \frac{p^{\frac{1}{1+\rho}}}{p^{\frac{1}{1+\rho}} + (1-p)^{\frac{1}{1+\rho}}}.$$

Пропускная способность:

$$C = 1 - h(p) = 1 + p \log_2 p + (1-p) \log_2 (1-p).$$

Оптимальное входное распределение: $P_X(0) = P_X(1) = \frac{1}{2}$.

Длина входного кодового вектора: L .

Экспонента Бхаттакария: $R_0 = 1 - 2 \log(\sqrt{p} + \sqrt{1-p})$.

Критическая скорость: $R_c = 1 - h(p_c)$, $p_c = \frac{\sqrt{p}}{\sqrt{p} + \sqrt{1-p}}$.

Оптимальная экспонента Галлагера 1:

$$E_G(R) = R_0 - R, \text{ если } 0 \leq R \leq R_c.$$

Оптимальная экспонента Галлагера 2: Параметрически

$$E_G(R) = -\delta \log_2 p - (1-\delta) \log_2 (1-p) - h(\delta); \\ R = 1 - h(\delta), \quad p \leq \delta \leq p_c, \text{ для } R_c \leq R \leq C.$$

Вероятность ошибочного декодирования: $\overline{\Pr}_{\text{er}} \leq 2^{-L E_G(R)}$.

Г л а в а 11

Непрерывные источники и непрерывные каналы

11.1. Непрерывные сообщения и их информационные характеристики

До сих пор рассматривались случайные величины, принимающие конечное число значений. С этого момента будут рассматриваться величины X , принимающие значения в некотором интервале (A, B) , возможно, бесконечном в обе стороны. Такая случайная величина описывается функцией *плотности вероятности* $p_X(x) \geq 0$ с условием нормировки

$$\int_A^B p_X(x) dx = 1.$$

Совместное распределение двух случайных величин X, Y задается двумерной плотностью вероятности $p_{XY}(x, y) \geq 0$, где

$$\int_A^B \int_C^D p_{XY}(x, y) dx dy = 1.$$

Из двумерной плотности можно найти одномерные плотности

$$p_X(x) = \int_C^D p_{XY}(x, y) dy,$$

$$p_Y(y) = \int_A^B p_{XY}(x, y) dx,$$

а также условные плотности вероятности

$$p_{Y|X}(y | x) = \frac{p_{XY}(x,y)}{p_X(x)},$$

$$p_{X|Y}(x | y) = \frac{p_{XY}(x,y)}{p_Y(y)}.$$

Рассмотрим теоретико-информационные функции.

Дифференциальная энтропия случайной величины X определяется равенством:

$$H(X) = \int_A^B p_X(x) \log_2 \frac{1}{p_X(x)} dx. \quad (11.1)$$

Дифференциальная энтропия непрерывной величины может принимать как положительные, так и отрицательные значения, а также может быть неограниченной величиной.

Пример 11.1. Пусть X случайная величина, равномерно распределенная в интервале $(0, A)$:

$$p_X(x) = \frac{1}{A}, \quad 0 < x < A.$$

Тогда дифференциальная энтропия этой величины равна

$$H(X) = \int_0^A \frac{1}{A} \log_2 A dx = \log_2 A \rightarrow \begin{cases} < 0, & \text{если } A < 1; \\ = 0, & \text{если } A = 1; \\ > 0, & \text{если } A > 1. \end{cases}$$

Условная дифференциальная энтропия непрерывной величины Y при заданном значении случайной величины $X = x$ определяется формулой

$$H(Y | X = x) = \int_C^D p_{Y|X}(y | x) \log_2 \frac{1}{p_{Y|X}(y | x)} dy. \quad (11.2)$$

Условная дифференциальная энтропия непрерывной величины Y при заданной случайной величине X определяется формулой

$$H(Y | X) = \int_A \int_C^{B D} p_{XY}(x, y) \log_2 \frac{1}{p_{Y|X}(y | x)} dx dy. \quad (11.3)$$

Среднее количество взаимной информации, которое содержит одна из непрерывных случайных величин $\{X, Y\}$ относительно другой, определяется по аналогии с дискретным случаем с помощью формулы:

$$\begin{aligned} I(X; Y) &= H(Y) - H(Y | X) = H(X) - H(X | Y) = \\ &= \int_A \int_C^{B D} p_{XY}(x, y) \log_2 \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)} dx dy. \end{aligned} \quad (11.4)$$

Как и в дискретном случае, с помощью неравенства логарифмов доказывается, что $I(X; Y) \geq 0$ и что $I(X; Y) = 0$ тогда и только тогда, когда величины $\{X, Y\}$ независимы, т.е. $p_{XY}(x, y) = p_X(x)p_Y(y)$.

11.2. Гауссовские случайные величины и их информационные характеристики

Случайная величина X называется гауссовской, если ее плотность вероятности $\varphi_X(x)$ имеет вид

$$\varphi_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left(-\frac{(x-m)^2}{2\sigma_X^2}\right), \quad -\infty < x < \infty, \quad (11.5)$$

где m и σ_X – параметры распределения.

Нетрудно убедиться, что

$$\begin{aligned} \frac{1}{\sqrt{2\pi}\sigma_X} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-m)^2}{2\sigma_X^2}\right) dx &= 1, \\ \frac{1}{\sqrt{2\pi}\sigma_X} \int_{-\infty}^{\infty} x \exp\left(-\frac{(x-m)^2}{2\sigma_X^2}\right) dx &= m, \\ \frac{1}{\sqrt{2\pi}\sigma_X} \int_{-\infty}^{\infty} (x-m)^2 \exp\left(-\frac{(x-m)^2}{2\sigma_X^2}\right) dx &= \sigma_X^2, \end{aligned}$$

так что параметр m – это среднее значение гауссовой случайной величины, а параметр σ_X^2 – дисперсия гауссовой случайной величины. Величина σ_X называется среднеквадратическим отклонением.

Вычислим дифференциальную энтропию гауссовой случайной величины, обозначив ее $H_{\text{Gauss}}(X)$. Так как

$$\log_2 \frac{1}{\varphi_X(x)} = \log_2 \sqrt{2\pi\sigma_X} + \frac{(x-m)^2}{2\sigma_X^2} \log_2 e,$$

то

$$\begin{aligned} H_{\text{Gauss}}(X) &= \int_{-\infty}^{\infty} \varphi_X(x) \log_2 \frac{1}{\varphi_X(x)} dx = \\ &= \int_{-\infty}^{\infty} \varphi_X(x) \left\{ \sqrt{2\pi\sigma_X} + \frac{(x-m)^2}{2\sigma_X^2} \log_2 e \right\} dx = \\ &= \sqrt{2\pi\sigma_X} + \frac{1}{2} \log_2 e = \frac{1}{2} \log_2 (2\pi e \sigma_X^2). \end{aligned} \tag{11.6}$$

Пусть X – произвольная случайная величина с плотностью вероятности $p_X(x)$, средним значением m и дисперсией σ_X^2 , определенная в интервале $(-\infty, \infty)$.

Лемма 11.1. *Дифференциальная энтропия случайной величины X удовлетворяет неравенству*

$$H(X) \leq H_{\text{Gauss}}(X) = \frac{1}{2} \log_2 (2\pi e \sigma_X^2),$$

причем равенство достигается тогда и только тогда, когда

$$p_X(x) = \varphi_X(x) = \frac{1}{\sqrt{2\pi\sigma_X}} \exp\left(-\frac{(x-m)^2}{2\sigma_X^2}\right).$$

Доказательство. Сначала докажем, что

$$H(X) = \int_{-\infty}^{\infty} p_X(x) \log_2 \frac{1}{p_X(x)} dx \leq \int_{-\infty}^{\infty} p_X(x) \log_2 \frac{1}{\varphi_X(x)} dx. \tag{11.7}$$

Запишем разность левой и правой частей в виде

$$\begin{aligned} & \int_{-\infty}^{\infty} p_X(x) \log_2 \frac{1}{p_X(x)} dx - \int_{-\infty}^{\infty} p_X(x) \log_2 \frac{1}{\varphi_X(x)} dx = \\ &= \int_{-\infty}^{\infty} p_X(x) \log_2 \frac{\varphi_X(x)}{p_X(x)} dx \end{aligned}$$

и используем неравенство логарифма:

$$\begin{aligned} & \int_{-\infty}^{\infty} p_X(x) \log_2 \frac{\varphi_X(x)}{p_X(x)} dx \leq \int_{-\infty}^{\infty} p_X(x) \left(\frac{\varphi_X(x)}{p_X(x)} - 1 \right) dx = \\ &= \int_{-\infty}^{\infty} (\varphi_X(x) - p_X(x)) dx = 0. \end{aligned}$$

Это доказывает соотношение (11.7). Равенство достигается только при $p_X(x) = \varphi_X(x)$.

Далее имеем

$$\begin{aligned} & \int_{-\infty}^{\infty} p_X(x) \log_2 \frac{1}{\varphi_X(x)} dx = \\ &= \int_{-\infty}^{\infty} p_X(x) \left\{ \sqrt{2\pi}\sigma_X + \frac{(x-m)^2}{2\sigma_X^2} \log_2 e \right\} dx = \\ &= \sqrt{2\pi}\sigma_X + \frac{1}{2} \log_2 e = \frac{1}{2} \log_2 (2\pi e \sigma_X^2) = H_{\text{Gauss}}(X), \end{aligned}$$

что и завершает доказательство.

11.3. Непрерывный канал без памяти с дискретным временем

На рис. 11.1 приведена модель передачи сообщений по непрерывному каналу без памяти с дискретным временем. В некоторый момент времени на вход канала подается сигнал x , представляющий собой значение непрерывной случайной величины X . На выходе канала регистрируется сигнал y , представляющий собой значение выходной непрерывной случайной величины Y .

Канал полностью описывается плотностью переходной вероятности $p_{Y|x}(x | y)$. Для полного описания системы передачи необходимо задать плотность вероятности входного сигнала

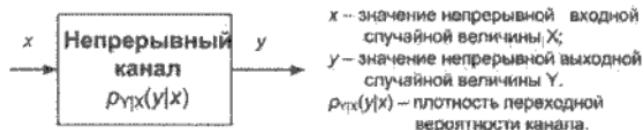


Рис. 11.1. Непрерывный канал с дискретным временем

$p_X(x)$. Тогда это позволяет найти совместную плотность входного и выходного сигналов $p_{XY}(x, y) = p_X(x)p_{Y|X}(x | y)$. В свою очередь, из нее можно найти маргинальную плотность выходного сигнала $p_Y(y) = \int_{-\infty}^{\infty} p_{XY}(x, y) dx$.

Все эти величины необходимы для расчета характеристик системы передачи сообщений. Важнейшей из них является средняя взаимная информация между выходом и входом

$$I(Y; X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{XY}(x, y) \log_2 \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)}.$$

Эта величина должна быть возможно большей. В распоряжении проектировщика есть единственное средство для ее увеличения – это выбор плотности вероятности $p_X(x)$ входного сигнала.

Пропускной способностью непрерывного канала с дискретным временем называется величина

$$C = \sup_{p_X} I(Y; X) = \sup_{p_X} (H(Y) - H(Y | X)). \quad (11.8)$$

Далее будут рассматриваться непрерывные каналы с аддитивным гауссовским шумом.

11.4. Канал с аддитивным гауссовым шумом

На рис. 11.2 показан канал с дискретным временем. В каждый момент времени выход канала $Y = X + Z$ равен сумме входного сигнала X и шума Z .

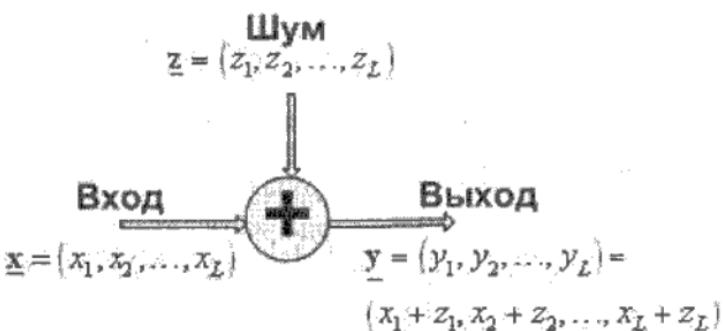


Рис. 11.2. Канал с гауссовым шумом и дискретным временем

Предполагается, что X – случайная величина с нулевым средним, дисперсией σ_X^2 и плотностью вероятности $p_X(x)$; что Z – гауссовская случайная величина с нулевым средним, дисперсией σ_Z^2 и плотностью

$$\varphi_Z(z) = \frac{1}{\sqrt{2\pi}\sigma_Z} \exp\left(-\frac{z^2}{2\sigma_Z^2}\right)$$

и что X и Z независимы. Из этих предположений следует, что выходная случайная величина Y имеет нулевое среднее и дисперсию $\sigma_Y^2 = \sigma_X^2 + \sigma_Z^2$.

Найдем пропускную способность такого канала.

Запишем формулу для переходной плотности вероятности:

$$\begin{aligned} p_{Y|X}(y|x) &= p_{Y|X}(x+z|x) = p_Z(y-x) = \\ &= \frac{1}{\sqrt{2\pi}\sigma_Z} \exp\left(-\frac{(y-x)^2}{2\sigma_Z^2}\right). \end{aligned} \tag{11.9}$$

Таким образом, выход Y при фиксированном x является гауссовой случайной величиной со средним значением, равным x , и с дисперсией σ_Z^2 .

Так как дифференциальная энтропия гауссовой случайной величины зависит только от дисперсии, но не зависит от среднего (см. (11.6)), то условная дифференциальная энтропия $H(Y|X)$ равна энтропии $H_{\text{Gauss}}(Z)$ гауссовой случайной величины Z и не зависит от входной плотности вероятности $p_X(x)$:

$$H(Y | X) = H_{\text{Gauss}}(Z) = \frac{1}{2} \log_2(2\pi e \sigma_Z^2). \quad (11.10)$$

В этих условиях средняя взаимная информация между выходом и входом записывается как

$$I(X; Y) = H(Y) - H(Y | X) = H(Y) - \frac{1}{2} \log_2(2\pi e \sigma_Z^2). \quad (11.11)$$

Чтобы найти пропускную способность, надо найти максимум дифференциальной энтропии $\sup_{p_X} H(Y)$ при условии, что случайная величина Y имеет нулевое среднее и дисперсию $\sigma_Y^2 = \sigma_X^2 + \sigma_Z^2$. Из леммы 11.1 следует, что

$$H(Y) \leq H_{\text{Gauss}}(Y) = \frac{1}{2} \log_2(2\pi e \sigma_Y^2) = \frac{1}{2} \log_2(2\pi e (\sigma_X^2 + \sigma_Z^2))$$

и что максимум достигается, если величина Y имеет гауссовское распределение

$$\varphi_Y(y) = \frac{1}{\sqrt{2\pi(\sigma_X^2 + \sigma_Z^2)}} \exp\left(-\frac{y^2}{2(\sigma_X^2 + \sigma_Z^2)}\right).$$

Но этого можно добиться, если распределение входной случайной величины X также выбрать гауссовским:

$$p_X(x) = \varphi_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left(-\frac{x^2}{2\sigma_X^2}\right).$$

Из теории вероятностей известно, что сумма независимых гауссовых величин является гауссовой случайной величиной со средним, равным сумме средних, и дисперсией, равной сумме дисперсий.

Окончательно пропускная способность канала с дискретным временем и аддитивным гауссовским шумом равна

$$C = \frac{1}{2} \log_2(2\pi e(\sigma_X^2 + \sigma_Z^2)) - \frac{1}{2} \log_2(2\pi e\sigma_Z^2) = \frac{1}{2} \log_2\left(1 + \frac{\sigma_X^2}{\sigma_Z^2}\right), \quad (11.12)$$

где σ_X^2 – мощность входного сигнала, σ_Z^2 – мощность шума в канале, σ_X^2/σ_Z^2 – отношение сигнал/шум по мощности.

Канал с дискретным временем называется каналом без памяти, если переходная плотность вероятности для векторов $\mathbf{x} = (x_1, x_2, \dots, x_L)$ и $\mathbf{y} = (y_1, y_2, \dots, y_L)$ равна произведению покомпонентных плотностей:

$$p_{Y_1 Y_2 \dots Y_L | X_1 X_2 \dots X_L}(y_1, y_2, \dots, y_L | x_1, x_2, \dots, x_L) = \prod_{j=1}^L p_{Y|X}(y_j | x_j).$$

Другими словами, шумы в различные моменты действуют независимо.

Пусть $M = 2^{RL}$ – число передаваемых векторов длины L , где R – скорость передачи. Пусть C – пропускная способность рассматриваемого канала. Сформулируем без доказательства теоремы Шеннона.

Теорема 11.1. *Если скорость передачи $R < C$, то существуют такие входные векторы $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M$ длины L , что вероятности ошибочного декодирования могут быть сделаны пре-небрежимо малыми:*

$$p_{\text{er},i} \rightarrow 0 \text{ при } L \rightarrow \infty; i = 1, 2, \dots, M.$$

Если скорость передачи $R > C$, то ни для какого набора входных сигналов вероятности ошибок декодирования не могут быть сделаны малыми при любых способах обработки:

$$p_{\text{er},i} \not\rightarrow 0, \text{ при } L \rightarrow \infty.$$

11.5. Система параллельных каналов с аддитивным гауссовым шумом

Теперь рассмотрим систему параллельных каналов без памяти с дискретным временем и с аддитивными гауссовыми шумами (рис. 11.3).

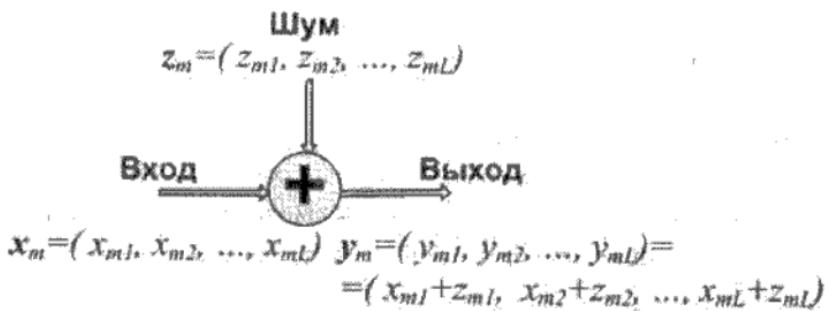
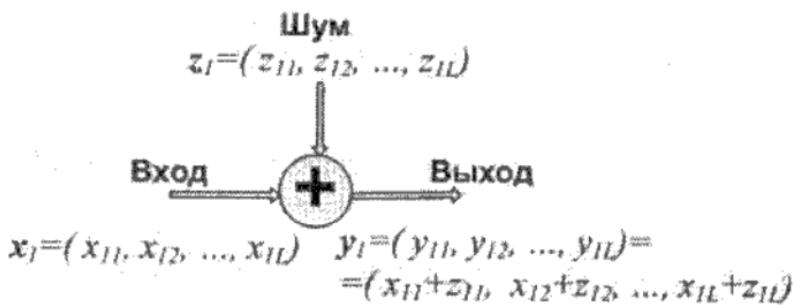


Рис. 11.3. Система параллельных каналов с гауссовым шумом и дискретным временем

Рассмотрим сначала случай однократного использования канала ($L = 1$).

Будем считать, что входным сигналом \mathbf{X} , выходным сигналом \mathbf{Y} и шумовым сигналом \mathbf{Z} являются случайные вектор-столбцы размера m :

$$\begin{aligned}\mathbf{X} &= (X_1, X_2, \dots, X_m)^{\text{tr}}, \\ \mathbf{Y} &= (Y_1, Y_2, \dots, Y_m)^{\text{tr}}, \\ \mathbf{Z} &= (Z_1, Z_2, \dots, Z_m)^{\text{tr}},\end{aligned}$$

где обозначение \mathbf{A}^{tr} означает транспонирование вектора или матрицы.

Предполагается, что шумовые компоненты Z_i , $i = 1 \dots m$, являются независимыми гауссовскими величинами с нулевыми средними и дисперсиями $\sigma_{Z_i}^2$, $i = 1 \dots m$, вообще говоря, различными.

Выходной вектор получается из входного вектора добавлением шума: $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$.

Канал задается переходной плотностью вероятности

$$p_{\mathbf{Y}|\mathbf{X}}(Y_1, Y_2, \dots, Y_m | X_1, X_2, \dots, X_m) = \prod_{i=1}^m p_{Y_i|X_i}(y_i | x_i),$$

где

$$p_{Y_i|X_i}(y_i | x_i) = \frac{1}{\sqrt{2\pi\sigma_{Z_i}^2}} \exp\left(-\frac{(y_i - x_i)^2}{2\sigma_{Z_i}^2}\right), \quad i = 1 \dots m. \quad (11.13)$$

Плотность вероятности $p_{\mathbf{X}}(\mathbf{x})$ входного вектора будет выбираться при вычислении пропускной способности этого канала. В частности, все компоненты вектора \mathbf{X} имеют нулевые средние. Дисперсии (мощности) каждой компоненты $\sigma_{X_i}^2$, $i = 1 \dots m$, будут выбираться оптимальным образом. Для краткости обозначим $E_i = \sigma_{X_i}^2$. Однако сумма мощностей, обозначаемая E_0 , задается заранее и является одним из ограничений для системы связи:

$$E_1 + E_2 + \dots + E_m = E_0. \quad (11.14)$$

Оценим среднюю взаимную информацию между выходом \mathbf{Y} и входом \mathbf{X} :

$$I(\mathbf{Y}; \mathbf{X}) = H(\mathbf{Y}) - H(\mathbf{Y} | \mathbf{X}).$$

В силу цепного равенства для энтропии имеем

$$H(\mathbf{Y}) = H(Y_1 Y_2 \dots Y_m) \leq \sum_{i=1}^m H(Y_i).$$

В силу (11.13) имеем

$$H(\mathbf{Y} \mid \mathbf{X}) = \sum_{i=1}^m H(Y_i \mid X_i).$$

Таким образом,

$$\begin{aligned} I(\mathbf{Y}; \mathbf{X}) &\leq \sum_{i=1}^m H(Y_i) - \sum_{i=1}^m H(Y_i \mid X_i) = \sum_{i=1}^m (H(Y_i) - H(Y_i \mid X_i)) = \\ &= \sum_{i=1}^m I(Y_i; X_i) \leq \frac{1}{2} \sum_{i=1}^m \log_2 \left(1 + \frac{E_i}{\sigma_{Z_i}^2}\right). \end{aligned}$$

Следовательно, пропускная способность системы параллельных каналов с аддитивными гауссовским шумами достигается оптимальным выбором мощностей $E_i \geq 0$ для каждого канала с условием (11.14):

$$C = \max_{\substack{E_1 \geq 0, \dots, E_m \geq 0 \\ E_1 + E_2 + \dots + E_m = E_0}} \left\{ \frac{1}{2} \sum_{i=1}^m \log_2 \left(1 + \frac{E_i}{\sigma_{Z_i}^2}\right) \right\}. \quad (11.15)$$

Для решения этой задачи будем считать, без ограничения общности, что мощности шумов упорядочены по возрастанию:

$$\sigma_{Z_1}^2 \leq \sigma_{Z_2}^2 \leq \dots \leq \sigma_{Z_m}^2.$$

Сначала рассмотрим случай, когда общая мощность E_0 достаточно велика (будет уточнено ниже). Тогда метод множителей Лагранжа с учетом только ограничения (11.14) приведет к условиям

$$E_i + \sigma_{Z_i}^2 = B, \quad i = 1, 2, \dots, m, \quad (11.16)$$

где B – некоторая константа, определяемая из условия (11.14):

$$B = \frac{E_0 + \sum_{i=1}^m \sigma_{Z_i}^2}{m}.$$

Отсюда получаем

$$E_i = B - \sigma_{Z_i}^2 = \frac{E_0 + \sum_{k=1}^m \sigma_{Z_k}^2}{m} - \sigma_{Z_i}^2, \quad i = 1, 2, \dots, m. \quad (11.17)$$

Если теперь вспомнить об условиях $E_i \geq 0$, то решение имеет физический смысл при условии, что для наиболее зашумленного m -го канала

$$E_m = \frac{E_0 + \sum_{k=1}^m \sigma_{Z_k}^2}{m} - \sigma_{Z_m}^2 \geq 0,$$

т.е.

$$E_0 \geq m\sigma_{Z_m}^2 - \sum_{k=1}^m \sigma_{Z_k}^2. \quad (11.18)$$

Решение (11.16) – (11.17) имеет интересную геометрическую интерпретацию, называемую методом "наполнения водой". Предположим, что резервуар имеет профиль дна ($\sigma_{Z_1}^2, \sigma_{Z_2}^2, \dots, \sigma_{Z_m}^2$), как показано на рис. 11.4.

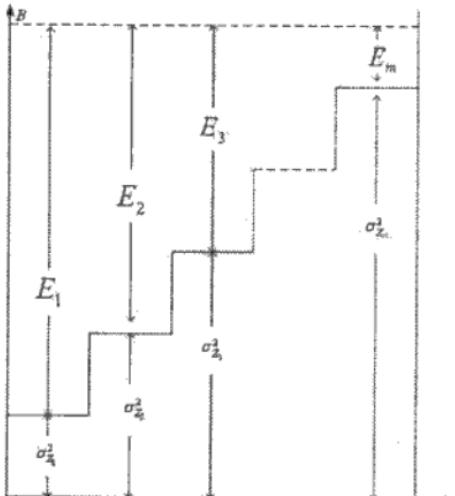


Рис. 11.4. Метод "наполнения водой": E_0 велико

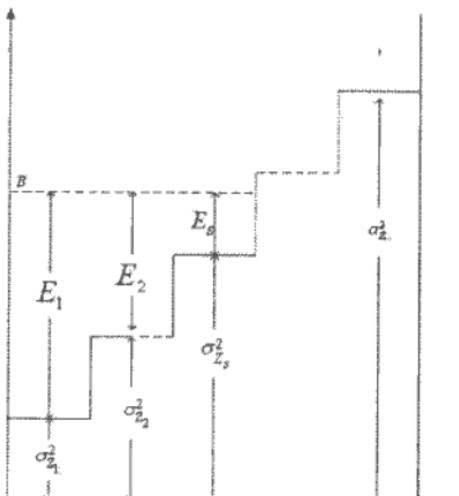


Рис. 11.5. Метод "наполнения водой": E_0 мало

Заполним этот резервуар водой, объем которой равен общей мощности E_0 . Тогда верхний край воды установится на уровне

B , а расстояние от верхнего края до дна определяет мощность E_i , отводимую каналу i .

Этот метод позволяет найти решение и для общего случая, когда не выполняется неравенство (11.18) (см. рис. 11.5). Верхний уровень воды B будет расположен выше $\sigma_{Z_s}^2$, но ниже $\sigma_{Z_{s+1}}^2$ для некоторого s . Это число определяется из условий:

$$E_0 \geq s\sigma_{Z_s}^2 - \sum_{k=1}^s \sigma_{Z_k}^2,$$

$$E_0 \leq (s+1)\sigma_{Z_{s+1}}^2 - \sum_{k=1}^{s+1} \sigma_{Z_k}^2.$$

Отсюда следует, что

$$B = \frac{E_0 + \sum_{k=1}^s \sigma_{Z_k}^2}{s},$$

$$E_i = B - \sigma_{Z_i}^2 = \frac{E_0 + \sum_{k=1}^s \sigma_{Z_k}^2}{s} - \sigma_{Z_i}^2, \quad i = 1, \dots, s,$$

$$E_i = 0, \quad i = s+1, \dots, m.$$

Пропускная способность системы параллельных каналов равна

$$C = \frac{1}{2} \sum_{i=1}^s \log_2 \frac{B}{\sigma_{Z_i}^2}.$$

Важным частным случаем является система параллельных каналов с одинаковыми шумовыми характеристиками:

$$\sigma_{Z_1}^2 = \sigma_{Z_2}^2 = \dots = \sigma_{Z_m}^2 = \sigma_Z^2.$$

Тогда $E_i = E_0/m$, $i = 1, 2, \dots, m$. Пропускная способность в этом случае равна

$$C = \frac{m}{2} \log_2 \left(1 + \frac{E_0}{m\sigma_Z^2} \right). \quad (11.19)$$

В случае, когда система параллельных каналов используется $L > 1$ раз, набор входных сигналов состоит из $M = 2^{RL}$ матриц размера $m \times L$:

$$\mathbf{X}_s = \begin{pmatrix} x_{11}^{(s)} & x_{12}^{(s)} & \dots & x_{1L}^{(s)} \\ x_{21}^{(s)} & x_{22}^{(s)} & \dots & x_{2L}^{(s)} \\ \dots & \dots & \dots & \dots \\ x_{m1}^{(s)} & x_{m2}^{(s)} & \dots & x_{mL}^{(s)} \end{pmatrix}, \quad s = 1, 2, \dots, M.$$

Первая строка матрицы соответствует передаче по первому каналу, вторая строка — по второму каналу и т.д.

Сформулируем без доказательства теоремы Шеннона для этого случая.

Теорема 11.2. *Если скорость передачи $R < C$, то существуют такие входные матрицы $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M$ длины L , что вероятности ошибочного декодирования могут быть сделаны пре-небрежимо малыми:*

$$p_{\text{per},i} \rightarrow 0 \text{ при } L \rightarrow \infty; \quad i = 1, 2, \dots, M.$$

Если скорость передачи $R > C$, то ни для какого набора входных матриц вероятности ошибок декодирования не могут быть сделаны малыми при любых способах обработки:

$$p_{\text{per},i} \not\rightarrow 0, \text{ при } L \rightarrow \infty.$$

11.6. Каналы с непрерывным временем и аддитивным белым гауссовским шумом

В этом разделе рассматриваются каналы, в которых входные $X(t)$ и выходные $Y(t)$ сигналы являются случайными процессами, заданными на некотором интервале времени $(0, T)$. Полное статистическое описание таких сигналов очень сложно. Вместо этого будут рассматриваться только сигналы, которые могут быть разложены в ряд по ортонормальным функциям. Их статистические свойства описываются совместными распределениями вероятностей коэффициентов. Неформально, такой подход

позволяет свести задачи с непрерывным временем к задачам с дискретным временем, подобным рассмотренным в предыдущей главе.

В приложениях большинство физических сигналов являются функциями с ограниченным спектром. В следующем разделе рассмотрено разложение в ряд таких сигналов.

11.6.1. Теорема Котельникова

Пусть задана на бесконечном интервале $-\infty < t < \infty$ функция $f(t)$. Будем предполагать, что существует преобразование Фурье от этой функции:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \exp(-i\omega t) dt. \quad (11.20)$$

В теории информации и других технических приложениях преобразование Фурье называется *спектром* функции. Если переменная t интерпретируется как время, то параметр ω называется *круговой частотой*, а связанный с ней параметр $f = \omega/2\pi$ называется *частотой в герцах*.

Обратное преобразование Фурье имеет следующий вид:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) \exp(i\omega t) d\omega. \quad (11.21)$$

Говорят, что функция $f(t)$ является функцией с *ограниченным спектром*, если существует *конечное* значение круговой частоты $\Omega > 0$ такое, что

$$F(\omega) = 0 \text{ для всех } |\omega| > \Omega. \quad (11.22)$$

Большинство физических сигналов, используемых для передачи сообщений, являются функциями с ограниченным спектром.

Учитывая (11.20) и (11.22), имеем

$$f(t) = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} F(\omega) \exp(i\omega t) d\omega, \quad (11.23)$$

где $\Omega = 2\pi F$, F – граничная частота в Гц.

Рассмотрим значения функции $f(t)$ в точках $k/2F$ для всех целых k , $-\infty < k < \infty$. Оказалось, что для функций с ограниченным спектром этих значений достаточно для того, чтобы найти значение этой функции в любой точке временной оси. Это утверждение составляет содержание знаменитой Теоремы Котельникова.

Теорема 11.3. (Теорема Котельникова.) Пусть $f(t)$ – функция с ограниченным спектром с граничной частотой F . Тогда

$$f(t) = \sum_{k=-\infty}^{\infty} f\left(\frac{k}{2F}\right) \frac{\sin \Omega(t - \frac{k}{2F})}{\Omega(t - \frac{k}{2F})}. \quad (11.24)$$

Доказательство. Преобразуем спектр $F(\omega)$ функции $f(t)$ в периодический спектр $PF(\omega)$ путем периодического продолжения в обе стороны по оси частот (рис. 11.6). В результате получим периодическую по ω функцию с периодом 2Ω .



Рис. 11.6. Периодическое продолжение спектра

Заметим, что

$$PF(\omega) \equiv F(\omega), \text{ если } -\Omega < \omega < \Omega. \quad (11.25)$$

Разложим $PF(\omega)$ в ряд Фурье:

$$PF(\omega) = \sum_{k=-\infty}^{\infty} D_k \exp\left(-2\pi i \frac{k\omega}{2\Omega}\right), \quad (11.26)$$

где

$$\begin{aligned} D_k &= \frac{1}{2\Omega} \int_{-\Omega}^{\Omega} PF(\omega) \exp\left(2\pi i \frac{k\omega}{2\Omega}\right) d\omega = \\ &= \frac{2\pi}{2\Omega} \frac{1}{2\pi} \int_{-\Omega}^{\Omega} F(\omega) \exp\left(2\pi i \frac{k\omega}{2\Omega}\right) d\omega = \frac{1}{2F} f\left(\frac{k}{2F}\right). \end{aligned} \quad (11.27)$$

Здесь использованы соотношение (11.25) и равенство (11.23) в точке $t = k/2F$.

Из (11.26) и (11.27) следует

$$PF(\omega) = \sum_{k=-\infty}^{\infty} \frac{1}{2F} f\left(\frac{k}{2F}\right) \exp\left(-2\pi i \frac{k\omega}{2\Omega}\right). \quad (11.28)$$

Снова используем соотношение (11.25) и подставим в равенство (11.23) ряд (11.28):

$$\begin{aligned} f(t) &= \frac{1}{2\pi} \int_{-\Omega}^{\Omega} F(\omega) \exp(i\omega t) d\omega = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} PF(\omega) \exp(i\omega t) d\omega = \\ &= \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \sum_{k=-\infty}^{\infty} \frac{1}{2F} f\left(\frac{k}{2F}\right) \exp i\omega \left(t - \frac{2\pi k}{2\Omega}\right) d\omega = \\ &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \frac{1}{2F} f\left(\frac{k}{2F}\right) \int_{-\Omega}^{\Omega} \exp i\omega \left(t - \frac{k}{2F}\right) d\omega = \\ &= \sum_{k=-\infty}^{\infty} f\left(\frac{k}{2F}\right) \frac{\sin \Omega(t - \frac{k}{2F})}{\Omega(t - \frac{k}{2F})}. \end{aligned} \quad (11.29)$$

11.6.2. Непрерывные каналы с аддитивным белым гауссовским шумом

На рис. 11.7 представлен канал с непрерывным временем и аддитивным белым гауссовским шумом.

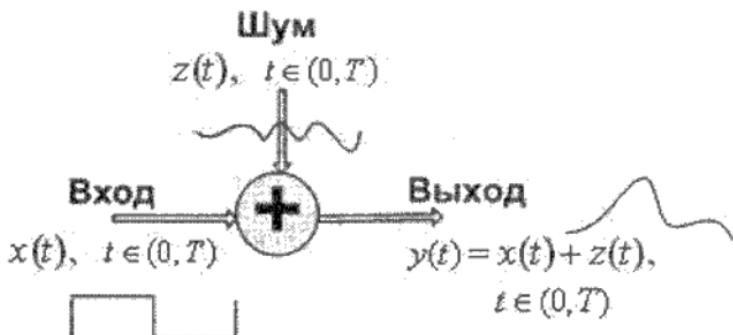


Рис. 11.7. Канал с белым аддитивным гауссовским шумом и непрерывным временем

Выходной сигнал представляет собой сумму входного сигнала и шума:

$$y(t) = x(t) + z(t), \quad (11.30)$$

где сигналы и шум – это некоторые случайные процессы. Формально они определены на всей временной оси, но практически интересен случай, когда сигнал $x(t)$ отличен от нуля только на каком-нибудь интервале, скажем, $(0, T)$.

Шум $z(t)$ представляет собой белый гауссовский шум. Для любого t величина $z(t)$ является гауссовой случайной величиной с нулевым средним. Строгое математическое описание этого процесса является достаточно трудным. Ограничимся его описанием на "физическом" уровне. Рассмотрим некоторую линейную систему, которая описывается импульсной переходной характеристикой $h(t)$. Если на вход системы подается сигнал $u(t)$, то на выходе системы наблюдается сигнал

$$v(t) = \int_{-\infty}^{\infty} u(t - \tau) h(\tau) d\tau.$$

Если вход является случайным процессом, то и выход является случайным процессом.

Шум $z(t)$ называется белым гауссовским шумом, если для любой линейной системы и любого t выходной сигнал

$$v(t) = \int_{-\infty}^{\infty} z(t-\tau) h(\tau) d\tau$$

является гауссовой случайной величиной с нулевым средним и дисперсией

$$Ev(t)^2 = \frac{N_0}{2} \int_{-\infty}^{\infty} h^2(\tau) d\tau.$$

Величина N_0 называется односторонней спектральной плотностью мощности шума $z(t)$.

Найдем пропускную способность канала (11.30). Сначала найдем пропускную способность другого канала, называемого каналом с ограниченной полосой $\Omega = 2\pi F$, а затем используем предельный переход $F \rightarrow \infty$.

Канал с ограниченной полосой $\Omega = 2\pi F$ получается из исходного канала рис. 11.7 добавлением на выходе полосового фильтра, как показано на рис. 11.8.

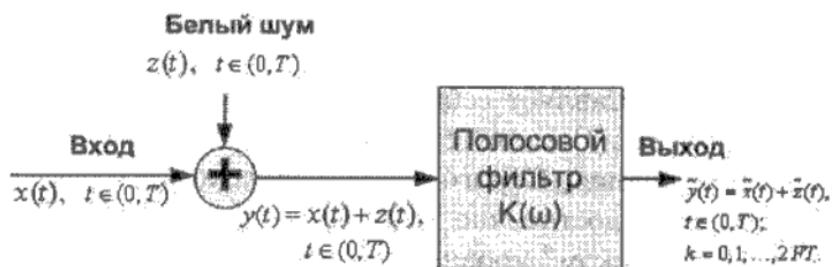


Рис. 11.8. Канал с ограниченной полосой частот

Коэффициент передачи фильтра равен

$$K(\omega) = \begin{cases} 1, & \text{если } |\omega| \leq \Omega; \\ 0, & \text{если } |\omega| > \Omega. \end{cases} \quad (11.31)$$

Рассмотрим подробнее характеристики сигналов в этой системе. Все сигналы на выходе являются функциями с ограниченным спектром и могут быть представлены своими рядами Котельникова:

$$\begin{aligned}\tilde{x}(t) &= \sum_{k=-\infty}^{\infty} \tilde{x}\left(\frac{k}{2F}\right) \frac{\sin \Omega(t-\frac{k}{2f})}{\Omega(t-\frac{k}{2f})}, \\ \tilde{z}(t) &= \sum_{k=-\infty}^{\infty} \tilde{z}\left(\frac{k}{2F}\right) \frac{\sin \Omega(t-\frac{k}{2f})}{\Omega(t-\frac{k}{2f})}, \\ \tilde{y}(t) &= \sum_{k=-\infty}^{\infty} [\tilde{x}\left(\frac{k}{2F}\right) + \tilde{z}\left(\frac{k}{2F}\right)] \frac{\sin \Omega(t-\frac{k}{2f})}{\Omega(t-\frac{k}{2f})}.\end{aligned}\quad (11.32)$$

Сигналы $x(t)$ выбираются при проектировании. Будем считать, что число $2FT \gg 1$ является целым. В качестве отсчетов $x\left(\frac{k}{2F}\right)$, $k = 0, 1, \dots, 2FT$, будем выбирать независимые гауссовские величины с нулевыми средними и одинаковыми дисперсиями $\sigma_k^2 = S$. Все остальные отсчеты положим равными 0. Тогда случайный процесс $x(t)$ описывается конечной суммой Котельникова

$$x(t) = \sum_{k=0}^{2FT} x\left(\frac{k}{2F}\right) \frac{\sin \Omega(t-\frac{k}{2f})}{\Omega(t-\frac{k}{2f})}. \quad (11.33)$$

Кроме того, $\tilde{x}(t) = x(t)$.

Средняя энергия таких сигналов равна

$$\begin{aligned}E_{\text{ave}} &= E \left(\int_{-\infty}^{\infty} x(t)^2 dt \right) = \\ &= \int_{-\infty}^{\infty} \sum_{k=0}^{2FT} E \left(x\left(\frac{k}{2F}\right)^2 \right) \left(\frac{\sin \Omega(t-\frac{k}{2f})}{\Omega(t-\frac{k}{2f})} \right)^2 dt \neq \\ &= ST \left(1 + \frac{1}{2FT} \right).\end{aligned}\quad (11.34)$$

Теоретически такой процесс $x(t)$ не может иметь конечной длительности, но можно считать, что *приближенно* он сосредоточен в интервале $(0, T)$ в том смысле, что средняя энергия E_{rest} той части сигнала, которая находится за пределами интервала $(0, T)$, мала по сравнению с полной энергией. Можно показать, что при $FT \gg 1$

$$E_{\text{rest}} = E \left(\int_{-\infty}^0 x(t)^2 dt + \int_T^{\infty} x(t)^2 dt \right) \leq E_{\text{ave}} \left(\frac{0.1 \ln FT}{FT} + \frac{0.2}{FT} \right), \quad (11.35)$$

так что доля остаточной энергии $E_{\text{rest}}/E_{\text{ave}}$ пренебрежимо мала.

Шум $\tilde{z}(t)$ описывается с помощью отсчетов

$$\tilde{z}(k/2F), \quad -\infty < k < \infty.$$

Так как процесс $\tilde{z}(t)$ получен из белого гауссовского шума $z(t)$ пропусканием через линейный фильтр, то отсчеты $\tilde{z}(k/2F)$ являются независимыми гауссовскими случайными величинами с нулевыми средними и одинаковыми дисперсиями $\sigma_{\tilde{Z}}^2 = N_0/2$, где N_0 – спектральная плотность мощности шума $z(t)$.

На выходе канала для дальнейшей обработки используются только сигналы в интервале $(0, T)$, причем берутся отсчеты сигнала $\tilde{y}(t)$ в моменты времени $t_k = k/2F$, $k = 0, 1, \dots, 2FT$, т.е.

$$\tilde{y}\left(\frac{k}{2F}\right) = \tilde{x}\left(\frac{k}{2F}\right) + \tilde{z}\left(\frac{k}{2F}\right), \quad k = 0, 1, \dots, 2FT.$$

Но это означает, что непрерывный канал (11.30) полностью эквивалентен системе из $m = 2FT + 1$ одинаковых каналов с дискретным временем с аддитивным гауссовским шумом.

Пропускная способность $C_F(T)$ в этом случае равна, как следует из (11.19),

$$C_F(T) = \frac{m}{2} \log_2 \left(1 + \frac{E_0}{m\sigma_{\tilde{Z}}^2} \right) \approx FT \log_2 \left(1 + \frac{ST}{2FTN_0/2} \right) = FT \log_2 \left(1 + \frac{S}{FN_0} \right). \quad (11.36)$$

Величина S/FN_0 представляет собой отношение сигнал/шум.

Формула (11.36) описывает пропускную способность канала с полосой F , если используются сигналы длительности T . Удобно использовать удельную пропускную способность

$$C_F = \frac{C_F(T)}{T} = F \log_2 \left(1 + \frac{S}{FN_0} \right).$$

Найдем теперь удельную пропускную способность канала с белым гауссовским шумом, изображенном на рис. 11.7. Переходя к пределу $F \rightarrow \infty$, получим

$$C_{\infty} = \lim_{F \rightarrow \infty} C_F = \frac{C_F(T)}{T} = \lim_{F \rightarrow \infty} F \log_2 \left(1 + \frac{S}{FN_0}\right) = \frac{S}{N_0} \log_2 e. \quad (11.37)$$

Пусть теперь в этом канале для передачи используются сигналы длительности T . Пусть $M = 2^{RT}$ – число передаваемых сообщений, где R – скорость передачи.

Как и ранее, сформулируем без доказательства теоремы Шеннона для этого случая.

Теорема 11.4. *Если скорость передачи $R < C_{\infty}$, то существуют такие входные сигналы $x_1(t), x_2(t), \dots, x_M(t)$ длительности T , что вероятности ошибочного декодирования могут быть сделаны пренебрежимо малыми:*

$$p_{\text{er},i} \rightarrow 0 \text{ при } T \rightarrow \infty; i = 1, 2, \dots, M.$$

Если скорость передачи $R > C_{\infty}$, то ни для какого набора входных сигналов вероятности ошибок декодирования не могут быть сделаны малыми при любых способах обработки:

$$p_{\text{er},i} \not\rightarrow 0, \text{ при } T \rightarrow \infty.$$

Из теоремы следует интересное наблюдение, носящее фундаментальный характер.

Рассмотрим случай $R = C_{\infty} - \varepsilon$, где $\varepsilon > 0$ малое число. Тогда $M = \lfloor 2^{(C_{\infty}-\varepsilon)T} \rfloor$. Если окажется, что $(C_{\infty}-\varepsilon)T < 1$, то передача невозможна, так как число передаваемых сообщений равно 1.

Но в предельном случае $ST/N_0 \log_2 e = 1$, где ST/N_0 – отношение сигнал/шум. В других единицах, критическое отношение сигнал/шум равно $10 \lg ST/N_0 = -1.6$ дБ.

Если отношение сигнал/шум в канале с белым гауссовским шумом меньше -1.6 дБ, то передача по такому каналу невозможна.

Приложение А

Сведения из теории вероятностей

Напомним основные понятия теории вероятности, касающиеся дискретных случайных величин.

Рассматривается случайный эксперимент с конечным числом исходов. Множество возможных исходов $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ называется *выборочным пространством*, или *пространством исходов*, или *пространством элементарных событий*.

Сами исходы ω_i называются *элементарными событиями*. Событием A называется любое подмножество множества Ω , включая пустое множество \emptyset , называемое *невозможным событием*, и само Ω , называемое *достоверным событием*.

Говорят, что событие A произошло, если исходом случайного эксперимента является одно из элементарных событий, входящих в A .

На пространстве Ω задается *вероятностная мера* P , которая каждому событию A ставит в соответствие неотрицательное число (меру) $P(A)$ между 0 и 1. Вероятностная мера должна удовлетворять условиям

$$P(\Omega) = 1, \tag{A.1}$$

и для любых непересекающихся событий A и B мера их объединения равна сумме мер

$$P(A \cup B) = P(A) + P(B), \quad \text{если } A \cap B = \emptyset. \tag{A.2}$$

Из (A.2) немедленно следует, что $P(\emptyset) = 0$, если выбрать $A = \emptyset$ и $B = \Omega$. Из этого же соотношения следует, что вероятностная мера P полностью определяется числами

$$p_i = P(\{\omega_i\}), \quad i = 1, 2, \dots, n, \quad (\text{A.3})$$

представляющими собой вероятности (меры) элементарных событий. В частности, так как $\Omega = \bigcup_{i=1}^n \{\omega_i\}$, то

$$\sum_{i=1}^n p_i = \sum_{i=1}^n P(\omega_i) = P\left(\bigcup_{i=1}^n \omega_i\right) = P(\Omega) = 1. \quad (\text{A.4})$$

Пара $\{\Omega, P\}$ называется дискретным вероятностным пространством.

Всякая функция $X = X(\omega)$, определенная на пространстве элементарных событий Ω и принимающая значения из некоторого конечного множества $X(\Omega)$, называется *случайной величиной*.

Например, на пространстве $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ можно определить случайные величины X , Y и Z следующим образом:

ω	$X(\omega)$	$Y(\omega)$	$Z(\omega)$
ω_1	2	да	[1 2 2]
ω_2	1	да	[3 4 0]
ω_3	1	нет	[5 6 1]
ω_4	0	нет	[3 6 0]

(A.5)

Для случайной величины X область значений — это множество $X(\Omega) = \{0, 1, 2\}$, случайной величины Y — $Y(\Omega) = \{\text{да, нет}\}$, случайной величины Z — множество

$$Z(\Omega) = \{[1, 2, 2], [3, 4, 0], [5, 6, 1], [3, 6, 0]\}.$$

Распределение вероятностей случайной величины X — это отображение множества $X(\Omega)$ в интервал $[0, 1]$

$$P_X(x) = \Pr(X = x), \quad (\text{A.6})$$

где $\Pr(X = x)$ означает вероятность того, что X примет значение x , т.е. меру события $\{\omega : X(\omega) = x\}$.

Из (A.6) следует, что

$$P_X(x) \geq 0 \text{ для всех } x \in X(\Omega) \quad (\text{A.7})$$

и что

$$\sum_{x \in X(\Omega)} P_X(x) = 1, \quad (\text{A.8})$$

где суммирование ведется по всем x из $X(\Omega)$.

Будем говорить, что задана одномерная дискретная случайная X , принимающая L значений, если область значений $X(\Omega)$ содержит L различных элементов, занумерованных в некотором порядке:

$$X(\Omega) = \{x_1, x_2, \dots, x_L\}.$$

Обозначим для краткости соответствующие вероятности

$$p_X(X = x_i) = p_i, \quad i = 1, 2, \dots, L.$$

Соотношение (A.8) примет вид

$$\sum_{i=1}^L p_i = 1. \quad (\text{A.9})$$

Аналогично, двумерной случайной величиной называют пару случайных величин $\{X, Y\}$, где X может принимать значения x_1, x_2, \dots, x_L , а Y — значения y_1, y_2, \dots, y_M .

Совместное (или двумерное) распределение

$$P_{XY}(X = x_i, Y = y_j) = p_{ij}, \quad i = 1, \dots, L; j = 1, \dots, M,$$

— это набор из LM чисел p_{ij} , удовлетворяющих условиям

$$p_{ij} \geq 0, \quad i = 1, 2, \dots, L; j = 1, 2, \dots, M;$$

$$\sum_{i=1}^L \sum_{j=1}^M p_{ij} = 1.$$

Из совместного распределения можно найти маргинальные (одномерные) распределения каждой из величин в отдельности:

$$P_X(X = x_i) = p_i =$$

$$= \sum_{j=1}^M P_{XY}(X = x_i, Y = y_j) = \sum_{j=1}^M p_{ij}, \quad i = 1, \dots, L;$$

$$\sum_{i=1}^L p_i = 1;$$

$$P_Y(Y = y_j) = q_j =$$

$$= \sum_{i=1}^L P_{XY}(X = x_i, Y = y_j) = \sum_{i=1}^L p_{ij}, \quad j = 1, \dots, M;$$

$$\sum_{j=1}^M q_j = 1.$$

Маргинальные распределения определяются из совместного однозначно.

Обратное утверждение не верно: двумерные величины с различными совместными распределениями $\{p_{ij}\}$ могут иметь одинаковые маргинальные распределения.

Если совместное распределение имеет вид

$$P_{XY}(X = x_i, Y = y_j) = p_i q_j$$

для любых пар i, j , то величины X и Y называются независимыми случайными величинами.

В противном случае величины X и Y называются зависимыми.

Для двумерных случайных величин вводится еще один тип распределений — *условные* распределения.

Условное распределение случайной величины X при условии, что задано значение другой случайной величины $Y = y_j$, определяется через совместное и маргинальное распределение следующим образом:

$$P_{X|Y=y_j}(X = x_i | Y = y_j) = \frac{P_{XY}(X = x_i, Y = y_j)}{P_Y(Y = y_j)} = \frac{p_{ij}}{q_j} = p_{i|j}.$$

Это действительно распределение, так как

$$\sum_{i=1}^L p_{i|j} = 1.$$

Всего для заданной пары случайных величин $\{X, Y\}$ можно определить M условных распределений величины X .

Аналогично определяется условное распределение случайной величины Y при заданном значении случайной величины $X = x_i$:

$$P_{Y|X=x_i}(Y = y_j | X = x_i) = \frac{P_{XY}(X = x_i, Y = y_j)}{P_X(X = x_i)} = \frac{p_{ij}}{p_i} = q_{j|i}.$$

Это действительно распределение, так как

$$\sum_{j=1}^M q_{j|i} = 1.$$

Всего для заданной пары случайных величин $\{X, Y\}$ можно определить L условных распределений величины Y .

Если случайная величина X принимает численные значения, то для нее можно определить *математическое ожидание (среднее значение)* по формуле

$$E_X(X) = \sum_{i=1}^L p_i x_i.$$

Аналогично определяют математическое ожидание для любой численной функции $F(X)$ от случайной величины X :

$$E_X(F(X)) = \sum_{i=1}^L p_i F(x_i).$$

Приложение Б

Некоторые неравенства

Б.1. Неравенство логарифма

Следующее элементарное неравенство, называемое *неравенством логарифма*, часто используется в теории информации.

Лемма Б.1. Для $x > 0$

$$\ln x \leq x - 1, \quad (\text{Б.1})$$

причем равенство достигается только для $x = 1$.

Доказательство. Рассмотрим функцию $f(x) = \ln x - x + 1$. Первая производная $f'(x) = 1/x - 1$ имеет единственный корень $x_0 = 1$. Вторая производная $f''(x_0) = -1/x_0^2 = -1$ в этой точке отрицательна. Следовательно, в этой точке функция достигает максимума, равного $f_{\max} = f(x_0) = 0$.

Таким образом, если $x \neq x_0 = 1$, то $f(x) < 0$, т.е. $\ln x < x - 1$, и $\ln x_0 = x_0 - 1 = 0$, если $x = x_0 = 1$. При произвольном основании логарифма, большем 1, неравенство имеет вид

$$\log x \leq (x - 1) \log e.$$

Б.2. Неравенства Гёльдера–Иенсена

Непрерывная функция $f(x)$, определенная на промежутке $[a, b]$, называется *вогнутой* (или *выпуклой вверх*) на этом промежутке, если для любых двух точек x_1 и x_2 из промежутка и любых двух чисел $0 \leq \alpha_1, 0 \leq \alpha_2$ таких, что $\alpha_1 + \alpha_2 = 1$, выполняется неравенство

$$\alpha_1 f(x_1) + \alpha_2 f(x_2) \leq f(\alpha_1 x_1 + \alpha_2 x_2). \quad (\text{Б.2})$$

Другими словами, любая хорда лежит либо под графиком, либо на нем.

Если функция $f(x)$ дважды дифференцируема в интервале (a, b) , то она будет вогнутой тогда и только тогда, когда в этом интервале $f''(x) \leq 0$.

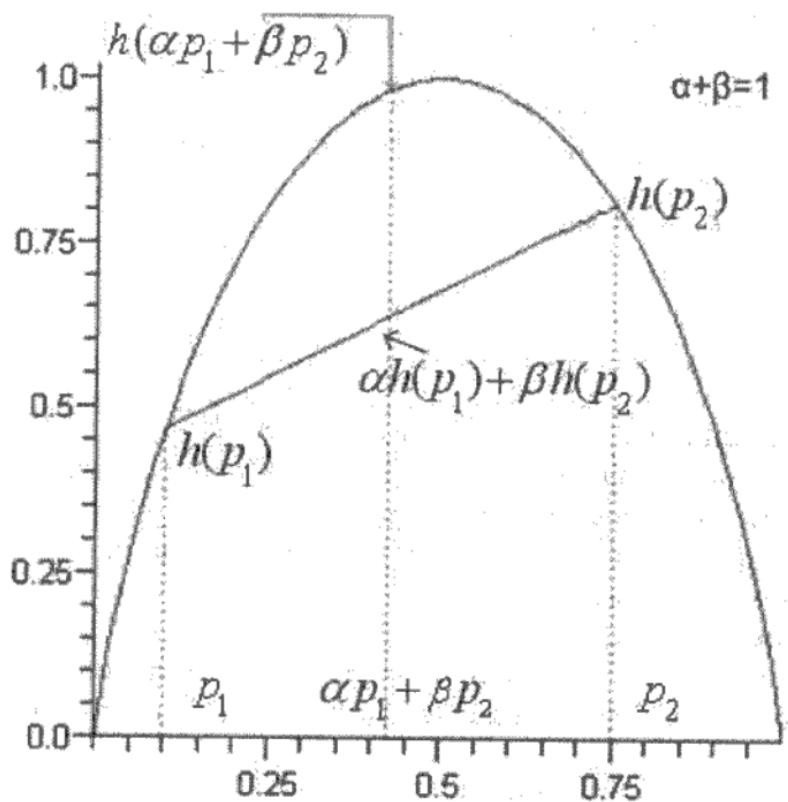


Рис. Б.1. Двоичная энтропия как вогнутая функция

Примером вогнутой функции является функция двоичной энтропии $h(p) = -p \log_2 p - (1-p) \log_2(1-p)$, показанная на рис. Б.1. Её вторая производная отрицательна в интервале $(0, 1)$: $h''(p) = -\frac{\log_2 e}{p(1-p)} < 0$.

Из определения следует более общее неравенство, называемое основным неравенством Гёльдера для вогнутых функций.

Лемма Б.2. Пусть $f(x)$ – вогнутая (выпуклая вверх) на промежутке $[a, b]$ функция. Пусть числа $\alpha_1, \alpha_2, \dots, \alpha_k$ неотрицательны и удовлетворяют условию

$$\alpha_1 + \alpha_2 + \cdots + \alpha_k = 1.$$

Тогда для любого набора из k точек x_1, x_2, \dots, x_k промежутка $[a, b]$ выполняется неравенство

$$\sum_{i=1}^k \alpha_i f(x_i) \leq f\left(\sum_{i=1}^k \alpha_i x_i\right). \quad (\text{Б.3})$$

Равенство в (Б.3) достигается тогда и только тогда, когда $x_1 = x_2 = \cdots = x_k$, либо когда $f(x)$ – линейная функция.

Доказательство. Для $k = 2$ неравенство выполняется по определению. Предположим, что оно верно для всех $i \leq k$ и докажем, что оно верно и для $i = k + 1$. Запишем тождество

$$\sum_{i=1}^{k+1} \alpha_i x_i = \sum_{i=1}^{k-1} \alpha_i x_i + \tilde{\alpha}_k \tilde{x}_k,$$

где

$$\tilde{\alpha}_k = \alpha_k + \alpha_{k+1}; \quad \tilde{x}_k = \frac{\alpha_k}{\alpha_k + \alpha_{k+1}} x_k + \frac{\alpha_{k+1}}{\alpha_k + \alpha_{k+1}} x_{k+1}.$$

Тогда по предположению индукции

$$\begin{aligned} f\left(\sum_{i=1}^{k+1} \alpha_i x_i\right) &= f\left(\sum_{i=1}^{k-1} \alpha_i x_i + \tilde{\alpha}_k \tilde{x}_k\right) \geq \\ &\geq \sum_{i=1}^{k-1} \alpha_i f(x_i) + \tilde{\alpha}_k f(\tilde{x}_k) \geq \\ &\geq \sum_{i=1}^{k-1} \alpha_i f(x_i) + (\alpha_k + \alpha_{k+1}) \left(\frac{\alpha_k}{\alpha_k + \alpha_{k+1}} f(x_k) + \frac{\alpha_{k+1}}{\alpha_k + \alpha_{k+1}} f(x_{k+1}) \right) = \\ &= \sum_{i=1}^k \alpha_i f(x_i), \text{ что и требовалось доказать.} \end{aligned}$$

Вторая часть теоремы очевидна.

Из основного неравенства Гёльдера можно получить много частных неравенств. Одним из таких неравенств является неравенство Коши–Гёльдера.

Лемма Б.3. Пусть a_s, b_s – неотрицательные числа. Пусть $\frac{1}{p} + \frac{1}{q} = 1$, где $p > 1$. Тогда

$$\sum a_s b_s \leq \left(\sum a_s^p \right)^{\frac{1}{p}} \left(\sum b_s^q \right)^{\frac{1}{q}}. \quad (\text{Б.4})$$

Равенство в (Б.4) достигается тогда и только тогда, когда для всех s выполняется равенство $b_s^q = c a_s^p$, где c – некоторая константа.

При $p = 2$ это известное неравенство Коши.

Доказательство. Введем обозначения

$$A = \sum a_s^p, \quad B = \sum b_s^q, \quad \alpha_s = \frac{a_s^p}{A}, \quad \beta_s = \frac{b_s^q}{B}, \quad x_s = \frac{\beta_s}{\alpha_s}.$$

Ясно, что $\sum \alpha_s = 1$, $\sum \beta_s = 1$. Поделим обе части неравенства (Б.4) на правую часть, т.е. на $A^{\frac{1}{p}} B^{\frac{1}{q}}$. Тогда с учетом введенных обозначений доказываемое неравенство сводится к виду

$$\begin{aligned} \sum \frac{a_s}{A^{\frac{1}{p}}} \frac{b_s}{B^{\frac{1}{q}}} &= \sum \left(\frac{a_s^p}{A} \right)^{\frac{1}{p}} \left(\frac{b_s^q}{B} \right)^{\frac{1}{q}} = \\ &= \sum \alpha_s \left(\frac{\beta_s}{\alpha_s} \right)^{\frac{1}{q}} = \sum \alpha_s (x_s)^{\frac{1}{q}} \leq 1. \end{aligned} \quad (\text{Б.5})$$

Так как функция $f(x) = x^{\frac{1}{q}}$ вогнутая при $q > 1$, то из основного неравенства (Б.3) получаем

$$\sum \alpha_s (x_s)^{\frac{1}{q}} \leq \left(\sum \alpha_s x_s \right)^{\frac{1}{q}} = \left(\sum \beta_s \right)^{\frac{1}{q}} = 1.$$

Равенство достигается, когда все x_s одинаковы, т.е. $\alpha_s = \beta_s$ или $b_s^q = c a_s^p$.

Если сделать замену $a_s \rightarrow a_s^{\frac{1}{p}}, b_s \rightarrow b_s^{\frac{1}{q}}$, то неравенство (Б.4) превращается в эквивалентное неравенство

$$\sum a_s^{\frac{1}{p}} b_s^{\frac{1}{q}} \leq \left(\sum a_s \right)^{\frac{1}{p}} \left(\sum b_s \right)^{\frac{1}{q}}, \quad (\text{Б.6})$$

где равенство достигается, если $b_s = c a_s$ для всех s .

Неравенства Гёльдера можно записать в терминах теории вероятностей. Пусть X – случайная величина, принимающая значения x_1, x_2, \dots, x_k с вероятностями $\alpha_1, \alpha_2, \dots, \alpha_k$. Пусть $f(X)$ – функция случайной величины X . Пусть $E_X(f(X)) = \sum \alpha_s f(x_s)$ означает математическое ожидание случайной величины $f(X)$. Тогда для вогнутой функции $f(X)$ верно неравенство

$$E_X(f(X)) \leq f(E_X(X)). \quad (\text{Б.7})$$

В таком виде неравенство справедливо для произвольных числовых случайных величин и называется неравенством Иенсена. Например, если случайная величина X определена в некоторой области A и имеет плотность $p_X(x) \geq 0$, $\int_A p_X(x) dx = 1$, то неравенство Иенсена для вогнутой функции $f(x)$ приобретает вид

$$\int_A p_X(x) f(x) dx \leq f \left(\int_A p_X(x) x dx \right).$$

Впрочем, в литературе часто все перечисленные выше неравенства называют неравенствами Иенсена, или неравенствами Гёльдера, или реже неравенствами Гёльдера–Иенсена.

Приложение В

Задачи и упражнения

B.1. Упражнения к главе 2

Упражнение 1

Условие. Доказать, что нижняя и верхняя границы для энтропии двумерной случайной величины имеют вид

$$0 \leq H(XY) \leq \log_2 L + \log_2 M, \quad (\text{B.1})$$

где L – число значений величины X , M – число значений величины Y .

Используя цепное равенство, запишем энтропию в виде суммы безусловной энтропии и условной энтропии:

$$H(XY) = H(X) + H(Y | X). \quad (\text{B.2})$$

Учтем ограничения на каждое из слагаемых правой части:

$$0 \leq H(X) \leq \log_2 L; \quad 0 \leq H(Y | X) \leq H(Y) \leq \log_2 M. \quad (\text{B.3})$$

Суммирование соответственно правых и левых частей этих неравенств доказывает соотношение, приведенное в условии.

Упражнение 2

Условие. Установить, справедливы или нет следующие неравенства:

1. $H(XY|Z) \geq H(Y|Z);$
2. $H(XYZ) - H(XY) \leq H(YZ) - H(Y).$

В случае справедливости найти условия, когда имеет место точный знак равенства.

Решение.

1. Согласно цепному равенству

$$H(XY|Z) = H(Y|Z) + H(X|ZY) \geq H(Y|Z),$$

так как $H(X|ZY) \geq 0$. Если $H(X|ZY) = 0$, то достигается равенство. В свою очередь это верно, если случайная величина X является детерминированной функцией случайных величин Y, Z .

2. Левая часть второго соотношения согласно цепному равенству равна

$$\begin{aligned} H(XYZ) - H(XY) &= H(XY) + H(Z|XY) - H(XY) = \\ &= H(Z|XY). \end{aligned}$$

Аналогично, правая часть равна

$$H(YZ) - H(Y) = H(Y) + H(Z | Y) - H(Y) = H(Z | Y).$$

Так как $H(Z|XY) \leq H(Z | Y)$, то исходное неравенство в общем случае верно. Точное равенство может достигаться, если случайные величины Z, Y, X образуют цепь Маркова. Тогда $H(Z|XY) = H(Z | Y)$.

Упражнение 3

Условие. Пусть X, Y, Z – три случайные величины, принимающие конечное число значений. Доказать, что

$$I(Y; Z) \leq I(XY; Z).$$

При каких условиях имеет место знак равенства?

Решение. Используя формулы для взаимной информации, запишем левую и правую части соотношения. Левая часть есть:

$$I(Y; Z) = H(Y) - H(Y | Z).$$

С учетом цепного равенства правая часть есть:

$$\begin{aligned} I(XY; Z) &= H(XY) - H(XY | Z) = \\ &= H(Y) + H(X | Y) - H(Y | Z) - H(X | YZ) = \\ &= I(Y; Z) + H(X | Y) - H(X | YZ). \end{aligned}$$

Так как $H(X | Y) - H(X | YZ) \geq 0$, то записанное в условии соотношение справедливо. Знак равенства имеет место, если величины X, Y, Z связаны в цепь Маркова первого порядка. В этом случае $H(X | Y) = H(X | YZ)$.

Упражнение 4

Условие. Доказать симметрию условной взаимной информации для двух случаев – в условии задано значение третьей случайной величины и в условии задана третья случайная величина:

$$\begin{aligned} I(X; Y | Z = z) &= I(Y; X | Z = z) \\ I(X; Y | Z) &= I(Y; X | Z). \end{aligned} \tag{B.4}$$

Решение. Сначала докажем первое соотношение, используя определение условной взаимной при условии, что задано значение третьей случайной величины:

$$I(X; Y | Z = z) = H(X | Z = z) - H(X | Y, Z = z),$$

где

$$\begin{aligned} H(X | Z = z) &= E_{X|Z=z} \log \frac{1}{P_{X|Z=z}(x|Z=z)} = \\ &= E_{XY|Z=z} \log \frac{1}{P_{X|Z=z}(x|Z=z)} \end{aligned}$$

и

$$H(X | Y, Z = z) = E_{XY|Z=z} \log \frac{1}{P_{X|Y,Z=z}(x | y, Z = z)}.$$

Возьмем разность этих энтропий и под знаком логарифма умножим и разделим на $P_{Y|Z=z}(y | Z = z)$:

$$\begin{aligned}
H(X \mid Z = z) - H(X \mid Y, Z = z) &= \\
&= E_{XY|Z=z} \log \frac{P_{X|Y,Z=z}(x|y,Z=z)}{P_{X|Z=z}(x|Z=z)} \frac{P_{Y|Z=z}(y|Z=z)}{P_{Y|Z=z}(y|Z=z)} = \\
&= E_{XY|Z=z} \log \frac{P_{XY|Z=z}(xy|Z=z)}{P_{X|Z=z}(x|Z=z)P_{Y|Z=z}(y|Z=z)} = \\
&= E_{XY|Z=z} \log \frac{P_{X|Z=z}(x|Z=z)P_{Y|X,Z=z}(y|x,Z=z)}{P_{X|Z=z}(x|Z=z)P_{Y|Z=z}(y|Z=z)} = \\
&= H(Y \mid Z = z) - H(Y \mid X, Z = z) \equiv I(Y; X \mid Z = z).
\end{aligned}$$

Доказательство второго соотношения (B.4) точно такое же, как доказательство первого соотношения. Отличие состоит в том, что вместо заданного значения третьей случайной величины $Z = z$ надо записать Z – эту случайную величину.

Упражнение 5

Условие. Доказать, что условная взаимная информация подчиняется следующим ограничениям:

$$0 \leq I(X; Y \mid Z) \leq \min[H(X \mid Z), H(Y \mid Z)].$$

Решение. Запишем соотношение, используя определение условной взаимной информации и свойство симметрии:

$$I(X; Y \mid Z) = H(X \mid Z) - H(X \mid YZ) = H(Y \mid Z) - H(Y \mid XZ). \quad (\text{B.5})$$

Как было показано ранее (см. глава 2), добавление случайной величины в условие не увеличивает энтропию, но может ее уменьшить. Поэтому знак неравенства относительно нуля в исходном соотношении поставлен правильно. Взаимная информация $I(X; Y \mid Z)$ равна нулю в случае равенства условных энтропий:

$$H(X \mid Z) = H(X \mid YZ) \text{ или } H(Y \mid Z) = H(Y \mid XZ).$$

Это равенство справедливо, если равны следующие условные вероятности

$$P_{X|Z}(x \mid z) = P_{X|Z}(x \mid yz) \text{ или } P_{Y|Z}(y \mid z) = P_{Y|XZ}(y \mid xz).$$

В свою очередь это выполняется, если величины X , Z , Y , а также Y , Z , X , образуют марковскую цепь первого порядка. В этом случае $I(X; Y | Z) = 0$.

Из соотношения (B.5) следует, что максимальное значение условной взаимной информации достигается в том случае, когда условная энтропия $H(Y | XZ)$ равна нулю. Это происходит, если случайная величина X (или Y) есть детерминированная функция случайных величин YZ (или XZ), указанных в условии. Тогда максимальное значение взаимной информации $I(X; Y | Z)$ равно $H(X | Z)$ или $H(Y | Z)$. Поэтому в качестве верхней границы следует использовать минимальное из этих двух значений, что и указано в условии.

Упражнение 6

Условие. Всегда ли верно соотношение

$$I(X; Y | Z) \leq I(X; Y)?$$

Решение. Чтобы ответить на этот вопрос, представим каждую из частей этого неравенства в соответствии с определением. Взаимная информация между случайными величинами X и Y при условии, что задана третья случайная величина Z , равна

$$I(X; Y | Z) = H(X | Z) - H(X | Y, Z).$$

Взаимная информация между случайными величинами X и Y равна

$$I(X; Y) = H(X) - H(X | Y).$$

Сравним правые части этих соотношений, учитывая неравенства между энтропиями:

$$H(X) \geq H(X | Z) \text{ и } H(X | Y) \geq H(X | Y, Z).$$

В первом соотношении первое слагаемое может быть меньше, чем первое слагаемое во втором соотношении, но зато вычитаемое может быть меньше, чем вычитаемое во втором соотношении. Вывод: в общем случае приведенное в условии неравенство несправедливо. В некоторых частных случаях оно может выполняться, в других частных случаях может выполняться противоположное неравенство.

B.2. Задачи и упражнения к главе 3

Задача 1

Условие. Рассмотрим стационарный двоичный марковский источник $U_1 U_2 U_3 \dots$ порядка $m = 1$. Пусть задано двумерное распределение

$$\begin{aligned} P_{U_1 U_2}(U_1 = 0, U_2 = 0) &= q_0; & P_{U_1 U_2}(U_1 = 0, U_2 = 1) &= q_1; \\ P_{U_1 U_2}(U_1 = 1, U_2 = 0) &= q_2; & P_{U_1 U_2}(U_1 = 1, U_2 = 1) &= q_3; \\ q_0 + q_1 + q_2 + q_3 &= 1. \end{aligned} \tag{B.6}$$

Найти энтропию H_∞ .

Решение. Марковский источник первого порядка полностью определяется совместным распределением первых двух величин U_1, U_2 . Сначала выясним требования к этому распределению. Чтобы источник был стационарным, необходимо и достаточно, чтобы одномерные распределения случайных величин U_1 и U_2 совпадали:

$$P_{U_1}(U_1 = 0) = P_{U_2}(U_2 = 0), \quad P_{U_1}(U_1 = 1) = P_{U_2}(U_2 = 1).$$

Следовательно,

$$\begin{aligned} P_{U_1}(U_1 = 0) &= P_{U_1 U_2}(U_1 = 0, U_2 = 0) + P_{U_1 U_2}(U_1 = 0, U_2 = 1) = \\ &= q_0 + q_1 = \\ P_{U_2}(U_2 = 0) &= P_{U_1 U_2}(U_1 = 0, U_2 = 0) + P_{U_1 U_2}(U_1 = 1, U_2 = 0) = \\ &= q_0 + q_2. \end{aligned} \tag{B.7}$$

Отсюда следует:

$$q_1 = q_2, \quad P_U(U = 0) = q_0 + q_1, \quad P_U(U = 1) = q_1 + q_3.$$

Энтропия H_∞ марковского источника первого порядка равна условной энтропии $H(U_2 | U_1)$:

$$\begin{aligned} H_\infty &= H(U_2 | U_1) = H(U_1 U_2) - H(U_1) = \\ &= -q_0 \log_2 q_0 - 2q_1 \log_2 q_1 - q_3 \log_2 q_3 + \\ &+ (q_0 + q_1) \log_2 (q_0 + q_1) + (q_1 + q_3) \log_2 (q_1 + q_3). \end{aligned}$$

В частности, если $q_0 = q_3$, то $2q_0 + 2q_1 = 1$, так что

$$H_\infty = -(2q_0) \log_2(2q_0) - (1 - 2q_0) \log_2(1 - 2q_0).$$

Задача 2

Условие. Для стационарного марковского источника первого порядка $U_1, U_2, \dots, U_n, \dots$, найти предельную энтропию

$$\lim_{n \rightarrow \infty} \frac{H(U_1, U_2, \dots, U_n)}{n} = H_\infty.$$

Решение. Используя цепное равенство, запишем формулу для многомерной энтропии в следующем виде:

$$H(U_1, U_2, \dots, U_n) = H(U_1) + H(U_2 | U_1) + \dots + H(U_n | U_1 U_2 \dots, U_{n-1}).$$

Учтем марковское свойство и стационарность источника:

$$H(U_3 | U_1 U_2) = H(U_3 | U_2) = H(U_2 | U_1);$$

$$H(U_4 | U_1 U_2 U_3) = H(U_4 | U_3) = H(U_2 | U_1);$$

\vdots

$$H(U_n | U_1 U_2 \dots U_{n-1}) = H(U_n | U_{n-1}) = H(U_2 | U_1).$$

В результате получим

$$H(U_1, U_2, \dots, U_n) = H(U_1) + (n-1)H(U_2 | U_1).$$

Разделив обе части этого соотношения на n и устремив n к бесконечности, получим $H_\infty = H(U_2 | U_1)$, т.е. для марковского источника первого порядка предельная энтропия равна условной 1-энтропии.

Задача 3

Условие. Пусть X_1, X_2, X_3, \dots — стационарный марковский источник 1-го порядка. Доказать, что

$$H(X_1 | X_2) \leq H(X_3 | X_1).$$

Для каких источников достигается знак равенства?

Решение. Используя цепное равенство, представим трехмерную энтропию в виде

$$H(X_1, X_2, X_3) = H(X_1) + H(X_3 | X_1) + H(X_2 | X_1, X_3).$$

Так как $H(X_2 | X_1, X_3) \leq H(X_2 | X_1)$, то верно следующее неравенство:

$$H(X_1, X_2, X_3) \leq H(X_1) + H(X_3 | X_1) + H(X_2 | X_1).$$

Используем другое представление трехмерной энтропии через цепное равенство:

$$\begin{aligned} H(X_1, X_2, X_3) &= H(X_2) + H(X_1 | X_2) + H(X_3 | X_1, X_2) = \\ &= H(X_2) + H(X_1 | X_2) + H(X_3 | X_2) = \\ &= H(X_1) + H(X_1 | X_2) + H(X_2 | X_1). \end{aligned}$$

В этих соотношениях использовано свойство марковости источника

$$H(X_3 | X_1, X_2) = H(X_3 | X_2)$$

и свойство стационарности

$$H(X_3 | X_2) = H(X_2 | X_1).$$

Сравнивая правые части неравенства и последнего уравнения, получаем искомое неравенство.

Равенство $H(X_1 | X_2) = H(X_3 | X_1)$ достигается для стационарных источников без памяти. В этом случае $H(X_j | X_i) = H(X_k)$ для любых индексов i, j, k .

B.3. Задачи и упражнения к главе 4

Упражнение 1.

Условие. Установить, какие из следующих кодов являются однозначно декодируемыми:

1. $\{0, 10, 11\}$;
2. $\{0, 01, 11\}$;
3. $\{0101111001, 01011, 110011, 10111, 0111\}$;
4. $\{0, 01\}$.

Решение

1. Код 1 префиксный, поэтому однозначно декодируемый.
2. Код 2 инверсный к коду 1, поэтому он также однозначно декодируемый.
3. Код 3 не является однозначно декодируемым. Для доказательства используем критерий Сардинаса–Патерсона:
 - Кодовое слово 01011 является префиксом кодового слова 0101111001.
 - Суффикс 11001 является префиксом кодового слова 110011.
 - Суффикс 1 является префиксом кодовых слов 110011 и 10111.
 - Суффикс второго из этих кодовых слов, т.е. 10111, совпадает с кодовым словом.
4. Код 4 инверсный к префиксному, поэтому однозначно декодируемый.

Задача 2

Условие. Алфавит источника без памяти U состоит из двух букв, появляющихся с вероятностями $p_1 = 0, 25$; $p_2 = 0, 75$. Выходной алфавит состоит из трех букв. Входная последовательность разбивается на тройки, которые кодируются независимо друг от друга. Построить код Шеннона–Фано. Найти среднюю длину на букву в битах и сравнить с энтропией источника.

Решение. Распределение троек входных букв источника таково:

$$P_1 = p_1^3 = 1/64, P_2 = p_1^2 p_2 = 3/64, P_3 = p_1 p_2 p_1 = 3/64, P_4 = p_1 p_2^2 = 9/64,$$

$$P_5 = p_2 p_1^2 = 3/64, P_6 = p_2 p_1 p_2 = 9/64, P_7 = p_2^2 p_1 = 9/64, P_8 = p_2^3 = 27/64.$$

Длины кодовых слов по Шеннону–Фано:

$$w_1 = \lceil \log_3 64 \rceil = 4, w_2 = \lceil \log_3 64/3 \rceil = 3,$$

$$w_3 = \lceil \log_3 64/3 \rceil = 3, w_4 = \lceil \log_3 64/9 \rceil = 2,$$

$$w_5 = \lceil \log_3 64/3 \rceil = 3, w_6 = \lceil \log_4 64/9 \rceil = 2,$$

$$w_7 = \lceil \log_3 64/9 \rceil = 2, w_8 = \lceil \log_3 64/27 \rceil = 1.$$

Средняя длина в символах:

$$\bar{w} = \frac{1}{3} \sum_{i=1}^8 P_i w_i = \frac{1}{3} \left(\frac{1}{64} \cdot 4 + 3 \cdot \frac{3}{64} \cdot 3 + 3 \cdot \frac{9}{64} \cdot 2 + \frac{27}{64} \cdot 1 \right) = \frac{36}{64} = 0.563.$$

Средняя длина в битах:

$$\bar{W} = \bar{w} \log_2 3 = 0.563 \cdot 1.585 = 0.892.$$

Энтропия источника в битах:

$$H(U) = \frac{1}{4} \log_2 4 + \frac{3}{4} \log_2 (4/3) = 0.415.$$

Задача 3

Условие. Дискретный источник без памяти порождает символы $u_1, u_2, u_3, u_4, u_5, u_6$ с вероятностями $0.05, 0.1, 0.15, 0.2, 0.23, 0.27$ соответственно. Построить оптимальный префиксный троичный код. Найти среднюю длину на букву в битах и сравнить с энтропией источника.

Решение. Код Хаффмана является оптимальным префиксным кодом. Найдем представление $(K-D)(D-2) = q(D-1)+r$, где r — число свободных листьев на последнем ярусе. В нашем случае число входных символов $K = 6$, число выходных символов $D = 3$, поэтому $(6-3) \cdot 1 = 1 \cdot 2 + 1$, так что $r = 1$. Далее используя троичное кодовое дерево и стандартную процедуру Хаффмана (см. главу 4), находим длины кодовых слов:

$$w_1 = 3, w_2 = 3, w_3 = 2, w_4 = 2, \\ w_5 = 1, w_6 = 1.$$

Средняя длина в символах на букву:

$$\bar{w} = \sum_{i=1}^6 P_i w_i = \\ = (0.05 + 0.1) \cdot 3 + (0.15 + 0.2) \cdot 2 + (0.23 + 0.27) \cdot 1 = 1.65.$$

Средняя длина в битах:

$$\overline{W} = \bar{w} \log_2 3 = 1.65 \cdot 1.585 = 2.615.$$

Энтропия источника в битах:

$$H(U) = - \sum_{i=1}^7 p_i \log_2 p_i = 2.421.$$

B.4. Задачи и упражнения к главе 5

Упражнение 1

Условие. Построить код Танстолла при следующих параметрах: число букв входного алфавита $k = 3$, число букв выходного алфавита $D = 2$, длина блока $N = 3$, вероятности $p(0) = 0.7, p(1) = 0.2, p(2) = 0.1$. Нарисовать таблицу для кодирования источника.

Решение. По приведенной классификации кодов код Танстолла характеризуется входными блоками равной длины и выходными блоками одной и той же длины. Сначала определим число расширений q , используя следующую формулу:

$$D^N - k = q(k - 1) + r,$$

где r – остаток, равный количеству неиспользуемых кодовых слов. Подставляя заданные значения параметров k, D, N , получим значения $q = 2, r = 1$.

Далее строим первый ярус (см. главу 5) для 3 входных сообщений с соответствующими вероятностями:

$$p(0) = 0.7 \quad | \quad p(1) = 0.2 \quad | \quad p(2) = 0.1.$$

Наиболее вероятным является сообщение u_0 , поэтому из нулевой ветви дерева делаем первое расширение. Получим 5 сообщений с соответствующими вероятностями:

$$\begin{array}{c|c|c|c} u_0u_0 & u_0u_1 & u_0u_2 & u_1 & u_2 \\ 0.7 \times 0.7 = 0.49 & 0.7 \times 0.2 = 0.14 & 0.7 \times 0.1 = 0.07 & 0.2 & 0.1 \end{array}$$

Наиболее вероятным является сообщение u_0u_0 . Второе расширение делаем из нулевой ветви. Получаем 7 сообщений с соответствующими вероятностями:

$$\begin{array}{ccc} u_0u_0u_0 & u_0u_0u_1 & u_0u_0u_2 \\ 0.7 \times 0.7 \times 0.7 = 0.343 & 0.7 \times 0.2 \times 0.7 = 0.098 & 0.7 \times 0.1 \times 0.7 = 0.049 \\ u_0u_1 & u_0u_2 & u_1 \quad u_2 \\ 0.2 \times 0.7 = 0.14 & 0.1 \times 0.7 = 0.07 & 0.2 \quad 0.1 \end{array}$$

$$P_{\text{all}} = 0.343 + 0.098 + 0.049 + 0.14 + 0.07 + 0.2 + 0.1 = 1.$$

Этим сообщениям ставим в соответствие двоичные кодовые слова длины 3:

$$\begin{aligned} u_0u_0u_0 &\leftrightarrow 000, \quad u_0u_0u_1 \leftrightarrow 001, \quad u_0u_0u_2 \leftrightarrow 010, \quad u_0u_1 \leftrightarrow 011, \\ u_0u_2 &\leftrightarrow 100, \quad u_1 \leftrightarrow 101, \quad u_2 \leftrightarrow 110. \end{aligned}$$

Код Танстолла построен.

Определим характеристики этого кода.

Средняя длина сообщений на входе в троичном алфавите равна $3 \times (0.343 + 0.098 + 0.049) + 2 \times (0.14 + 0.07) + 0.2 + 0.1 = 2.19$.

Средняя длина в битах на входе равна

$$2.19 \cdot \log_2 3 = 2.19 \cdot 1.585 = 3.471.$$

Оценим эффективность кода. Найдем отношение длины блока на выходе кодера к средней длине последовательности сообщений на входе: $\beta = \frac{3}{3.471} = 0.864$. Полученное значение определяет в среднем число двоичных символов на одно сообщение

источника. В среднем длина сообщений на выходе меньше длины на входе в $\frac{1}{\beta} = \frac{3.471}{3} = 1.157$ раз.

Найдем нормированную на число букв алфавита энтропию источника по формуле:

$$\frac{H}{\log_2 3} = \frac{-0.7 \log_2 0.7 - 0.2 \log_2 0.2 - 0.1 \log_2 0.1}{\log_2 3} = \frac{1.157}{1.585} = 0.730.$$

Эффективность кода определим в виде отношения нормированной энтропии к средней длине на выходе на одно сообщение источника: $\eta = \frac{0.730}{0.864} = 0.845$.

B.5. Задачи и упражнения к главе 6

Упражнение 1

Условие. Двоичная последовательность является ε -типической для некоторого распределения вероятностей. При каких условиях инверсная последовательность, т.е. последовательность, полученная заменой 1 на 0 и 0 на 1, будет ε -типической для того же распределения вероятностей?

Решение. Пусть двоичная последовательность состоит из N символов, среди которых n_0 нулевых символов и $n_1 = N - n_0$ единичных символов. Заданы значение ε , вероятность нулевого символа $p(0) = p$ и вероятность единичного символа $p(1) = 1 - p$. По условию задачи эта последовательность является ε -типической. Это значит, что выполняются следующие неравенства:

$$(1 - \varepsilon)pN \leq n_0 \leq (1 + \varepsilon)pN;$$
$$(1 - \varepsilon)(1 - p)N \leq (N - n_0) \leq (1 + \varepsilon)(1 - p)N.$$

В инверсной последовательности число нулевых символов равно числу единичных символов, а число единичных символов равно числу нулевых символов исходной последовательности. Остальные значения, т.е. вероятности появления и ε , остаются прежними. Для того, чтобы инверсная последовательность была ε -типической, должны выполняться следующие неравенства:

$$(1 - \varepsilon)pN \leq (N - n_0) \leq (1 + \varepsilon)pN;$$

$$(1 - \varepsilon)(1 - p)N \leq n_0 \leq (1 + \varepsilon)(1 - p)N.$$

Тривиальное решение $(N - n_0) = n_0$, т.е. число единичных и нулевых символов одно и то же. Это верно для той же области значений p, ε , для которой выполняются заданные неравенства.

Рассмотрим другие возможные варианты соотношений единичных и нулевых символов. Введем обозначение $\gamma = n_0/N$. Прежде всего найдем значения $\gamma = n_0/N$ в зависимости от значений p, ε , при которых выполняются соотношения ε -типичности для исходной последовательности. Перепишем эти соотношения в виде

$$(1 - \varepsilon)p \leq \gamma \leq (1 + \varepsilon)p;$$

$$(1 - \varepsilon)(1 - p) \leq \gamma \leq (1 + \varepsilon)(1 - p).$$

После некоторых элементарных преобразований получаем

$$(1 - \varepsilon)p \leq \gamma \leq (1 + \varepsilon)p; \quad 0 \leq p \leq 1/2;$$

$$(1 + \varepsilon)p - \varepsilon \leq \gamma \leq (1 - \varepsilon)p + \varepsilon; \quad 1/2 \leq p \leq 1.$$

Теперь представим соотношения ε -типичности для инверсной последовательности:

$$(1 - \varepsilon)p \leq (1 - \gamma) \leq (1 + \varepsilon)p;$$

$$(1 - \varepsilon)(1 - p) \leq \gamma \leq (1 + \varepsilon)(1 - p).$$

Так же, как в предыдущем случае находим значения γ , для которых выполняются условия ε -типичности для инверсной последовательности:

$$(1 - p) - \varepsilon p \leq \gamma \leq 1 - (1 - \varepsilon)p; \quad 0 \leq p \leq 1/2;$$

$$(1 - \varepsilon)(1 - p) \leq \gamma \leq (1 + \varepsilon)(1 - p); \quad 1/2 \leq p \leq 1.$$

Соотношения ε -типичности для обеих последовательностей выполняются (области совпадают), если при любом значении ε значения γ и p таковы, что

$$(1-p) - \varepsilon p \leq \gamma \leq (1+\varepsilon)p; 1/2 - \varepsilon/2(1+\varepsilon) \leq p \leq 1/2;$$

$$(1+\varepsilon)p - \varepsilon \leq \gamma \leq (1+\varepsilon)(1-p); 1/2 \leq p \leq 1/2 + \varepsilon/2(1+\varepsilon).$$

Для других значений p , т.е. для $0 \leq p \leq 1/2 - \varepsilon/2(1+\varepsilon)$ и $1/2 + \varepsilon/2(1+\varepsilon) \leq p \leq 1$ подходящих значений γ не существует.

Упражнение 2

Условие. Для двоичного источника без памяти с вероятностью $P_U(u=0) = p$ задано значение $\varepsilon = 0.2$. Для каких p последовательность

$$z = (0010001011001011001000101100)$$

не будет ε -типической?

Решение. Сначала подсчитаем общее число символов в последовательности, а также число нулевых и единичных символов: $N = 28$, $n_0 = 17$, $n_1 = 11$ соответственно. Запишем условия, при которых последовательность является ε -типической:

$$(1 - \varepsilon)pN \leq n_0 \leq (1 + \varepsilon)pN,$$

$$(1 - \varepsilon)(1 - p)N \leq n_1 \leq (1 + \varepsilon)(1 - p)N.$$

Подставим в эти соотношения численные значения:

$$(1 - 0.2)p \cdot 28 \leq 17 \leq (1 + 0.2)p \cdot 28,$$

$$(1 - 0.2)(1 - p) \cdot 28 \leq 11 \leq (1 + 0.2)(1 - p) \cdot 28.$$

Получим соотношения для вероятности нулевого символа, при которых выполняется соотношение ε -типичности:

$$p \geq 17/33.6 = 0.506, \quad p \leq 0.759, \quad p \leq 0.673, \quad p \geq 0.509.$$

Если хотя бы одно из неравенств не выполняется, то последовательность не будет ε -типической. Соотношение ε -типичности не выполняется для значений p в интервале $0 \leq p < 0.509$ или в интервале $0.673 < p \leq 1$.

Задача 3

Условие. Последовательность двоичных символов x порождена источником с вероятностями символов

$$p_0 = p; \quad p_1 = 1 - p.$$

Эта последовательность передается по каналу с переходными вероятностями

$$\begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix}.$$

В результате наблюдается последовательность пар вход-выход:

$$\begin{aligned} x &= 0110011110100101100101111101001010, \\ y &= 0010010110110100100111101111001011. \end{aligned}$$

Задано значение $\varepsilon = 0.2$. Для каких значений p эта последовательность пар является ε -типической?

Решение. Запишем соотношения ε -типичности для последовательности пар:

$$Lp_{ij}(1 - \varepsilon) < r_{ij} < Lp_{ij}(1 + \varepsilon).$$

Здесь введены обозначения: L – общая длина последовательности пар, p_{ij} – совместная вероятность пары ij , r_{ij} – число пар вида ij в этой последовательности, $i = 0, 1$, $j = 0, 1$.

Сначала подсчитаем общее число пар и число различных пар:

$$L = 34, \quad r_{00} = 11, \quad r_{01} = 4, \quad r_{10} = 4, \quad r_{11} = 15.$$

Учтем заданные в матрице значения переходных вероятностей и найдем вероятности различных пар: $p_{ij} = p_i p(j/i)$, где

$$p(0/0) = p(1/1) = 3/4, \quad p(1/0) = p(0/1) = 1/4.$$

Подставим значения вероятностей и значение $\varepsilon = 0.2$ в условия ε -типичности. Получим следующие неравенства:

$$34p \cdot (3/4)(0.8) < 11 < 34p \cdot (3/4)(1.2),$$

$$34p \cdot (1/4)(0.8) < 4 < 34p \cdot (1/4)(1.2),$$

$$34(1-p)(1/4)(0.8) < 4 < 34(1-p)(1/4)(1.2),$$

$$34(1-p)(3/4)(0.8) < 15 < 34(1-p)(3/4)(1.2).$$

В результате численных расчетов получаем соотношения для p :

$$p > 0.359, \quad p > 0.265, \quad p > 0.392, \quad p > 0.412,$$

$$p < 0.588, \quad p < 0.608, \quad p < 0.510, \quad p < 0.539.$$

Из этих неравенств следует, что условия ε -типичности выполняются, если $0.412 < p < 0.510$.

B.6. Задачи и упражнения к главе 8

Упражнение 1

Условие. Дискретный канал без памяти задается матрицей переходных вероятностей

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{8} & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{4} & \frac{1}{8} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{2} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{2} & \frac{1}{8} & \frac{1}{4} \end{bmatrix}$$

Найти пропускную способность этого канала.

Решение. По определению, пропускная способность канала выражена формулой

$$C = \sup_{p_X} \{H(Y) - H(Y|X)\},$$

где верхняя грань берется по входному распределению $p_X(x)$.

1. Исходя из вида матрицы переходных вероятностей, отмечаем, что канал симметричен по входу, так как все строки – перестановки первой строки. Отсюда следует, что условная энтропия $H(Y|X)$ не зависит от входного распределения $p_X(x)$ и равна энтропии первой строки:

$$H(Y|X) = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{8} \log_2 \frac{1}{8} - \frac{1}{4} \log_2 \frac{1}{4} - \frac{1}{8} \log_2 \frac{1}{8} = \frac{7}{4}.$$

Теперь надо найти максимум энтропии $H(Y)$ по входному распределению $p_X(x)$:

$$C = \sup_{p_X} \{H(Y)\} - H(Y|X) = \sup_{p_X} \{H(Y)\} - \frac{7}{4}.$$

2. Очевидно,

$$\sup_{p_X} \{H(Y)\} \leq \log_2 4 = 2.$$

3. Кроме того, отмечаем, что канал симметричен по выходу, так как все столбцы – перестановки первого столбца. Отсюда следует, что при равномерном распределении на входе выход также имеет равномерное распределение. Оптимальным является равномерное распределение, так как при нем достигается максимум

$$\sup_{p_X} \{H(Y)\} = \log_2 4 = 2.$$

Объединяя полученные значения энтропий с соответствующими знаками, получаем пропускную способность:

$$C = 2 - \frac{7}{4} = \frac{1}{4}$$

бит на одно использование канала.

Упражнение 2

Условие. Дискретный канал без памяти со стираниями и ошибками задается матрицей переходных вероятностей

$$\begin{bmatrix} q & p & r \\ p & q & r \end{bmatrix}, \quad q + p + r = 1.$$

Найти пропускную способность этого канала.

Решение. Так как вторая строка матрицы переходных вероятностей является перестановкой элементов первой строки, то канал симметричен по входу. Поэтому условная энтропия $H(Y|X)$ не зависит от входного распределения и равна энтропии первой строки:

$$H(Y|X) = -q \log_2 q - p \log_2 p - r \log_2 r.$$

Теперь надо найти максимум энтропии $H(Y)$. Запишем формулы для распределения $p_Y(y)$, обозначив выходные символы канала 0, 1, 2:

$$\begin{aligned} p_Y(0) &= p_X(0)q + p_X(1)p; \\ p_Y(1) &= p_X(0)p + p_X(1)q; \\ p_Y(2) &= p_X(0)r + p_X(1)r = r. \end{aligned}$$

Из симметрии первых двух соотношений ясно, что входное распределение должно быть равномерным, чтобы максимизировать энтропию $H(Y)$:

$$p_X(0) = p_X(1) = \frac{1}{2}.$$

Следовательно, выходное распределение имеет вид

$$\begin{aligned} p_Y(0) &= \frac{1}{2}q + \frac{1}{2}p = \frac{q+p}{2}, \\ p_Y(1) &= \frac{1}{2}p + \frac{1}{2}q = \frac{q+p}{2}, \\ p_Y(2) &= \frac{1}{2}r + \frac{1}{2}r = r. \end{aligned}$$

Пропускная способность равна

$$\begin{aligned} C &= H(Y) - H(Y|X) = \\ &= -2\frac{q+p}{2} \log_2 \frac{q+p}{2} - r \log_2 r + q \log_2 q + p \log_2 p + r \log_2 r = \\ &= (1-r)\left(1 + \frac{p}{p+q} \log_2 \frac{p}{p+q} + \frac{q}{p+q} \log_2 \frac{q}{p+q}\right). \end{aligned}$$

B.7. Задачи к главе 9

Задача 1

Условие. Дискретный канал без памяти описывается матрицей переходных вероятностей

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}.$$

Входной и выходной алфавиты состоят из символов 0, 1, 2. Сообщения, подлежащие передаче, – это следующие векторы длины N :

$$\mathbf{x}_1 = (0, 0, \dots, 0);$$

$$\mathbf{x}_2 = (1, 1, \dots, 1);$$

$$\mathbf{x}_3 = (2, 2, \dots, 2).$$

На выходе канала получаем вектор \mathbf{y} той же длины, что и на входе.

Найти оптимальные области декодирования для этих сообщений, точные значения для вероятностей ошибок и оценки Бхattachария для вероятностей ошибок.

Решение. Пусть $n_y(0)$, $n_y(1)$, $n_y(2)$ – число нулей, единиц и двоек в векторе \mathbf{y} . Учитывая значения элементов матрицы переходных вероятностей, отметим следующее: на выходе не может появиться вектор \mathbf{y} , в котором одновременно

$$n_y(0) \geq 1, n_y(1) \geq 1, n_y(2) \geq 1.$$

Соответствующие переходные вероятности равны нулю:

$$p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}_i) = 0, i = 1, 2, 3.$$

Для всех остальных векторов

$$p_{Y|X}(y|x_i) = \frac{1}{2^N}, \quad i = 1, 2, 3.$$

Число таких векторов равно $3 \cdot 2^N - 3$. Их можно произвольным образом разбить на три равные части, по $2^N - 1$ векторов в каждой. Все такие области будут оптимальными областями декодирования.

Точные вероятности ошибок равны: $P_{e1} = P_{e2} = P_{e3} = \frac{1}{2^N}$.

Оценки Бхаттачарриа таковы: $\widetilde{P}_{e1} = \widetilde{P}_{e2} = \widetilde{P}_{e3} = \frac{2}{2^N}$.

Задача 2

Условие. Пусть вход X канала без памяти двоичный (0 или 1), а выход Y – троичный (0, 1 или 2). Пусть

$$\begin{aligned} P_{Y|X}(0|0) &= 1/4, \quad P_{Y|X}(1|0) = 1/4, \quad P_{Y|X}(2|0) = 1/2, \\ P_{Y|X}(0|1) &= 1/2, \quad P_{Y|X}(1|1) = 1/4, \quad P_{Y|X}(2|1) = 1/4. \end{aligned}$$

Найти расстояние Бхаттачарриа между 0 и 1. Найти среднюю вероятность ошибки для случайного кода из двух слов длины N при равномерном распределении входных символов.

Решение. Расстояние Бхаттачарриа между 0 и 1 равно

$$\begin{aligned} D_B &= -\log_2 \left(\sum_y \sqrt{P_{Y|X}(y|0)P_{Y|X}(y|1)} \right) = \\ &= -\log_2 \left(\sqrt{\frac{1}{4}\frac{1}{2}} + \sqrt{\frac{1}{4}\frac{1}{4}} + \sqrt{\frac{1}{2}\frac{1}{4}} \right) = 0.0632482. \end{aligned}$$

Экспонента случайного кодирования определяется по формуле

$$\begin{aligned} E_r &= -\log_2 \left(\sum_y \left(\sum_x \sqrt{P_{Y|X}(y|x)} q_X(x) \right)^2 \right) = \\ &= -\log_2 \left(\left(\sqrt{\frac{1}{4}} + \sqrt{\frac{1}{2}} \right)^2 /4 + \left(\sqrt{\frac{1}{4}} + \sqrt{\frac{1}{4}} \right)^2 /4 + \right. \\ &\quad \left. + \left(\sqrt{\frac{1}{2}} + \sqrt{\frac{1}{4}} \right)^2 /4 \right) = 0.0312775. \end{aligned}$$

Оптимальная оценка средней вероятности ошибки равна

$$\overline{P}_{\text{err, opt}} \leq \frac{1}{2} 2^{-ND_B} = \frac{1}{2} 2^{-0.0632482N}$$

Оценка случайного кодирования равна

$$\overline{P}_{\text{err, random}} \leq \frac{1}{2} 2^{-NE_r} = \frac{1}{2} 2^{-0.0312775N}$$

B.8. Задачи к главе 10

Задача 1

Условие. Найти экспоненту вероятности ошибки $E(R, C)$ и пропускную способность для двоичного симметричного канала, характеризуемого вероятностью искажения символа $p = 0.1$ при скорости передачи $R = 0.12$. Оценить вероятность ошибки на блок при длине блока $L = 100$.

Решение. Для решения этой задачи используем расчетные формулы, приведенные в главе 10.

Пропускную способность вычислим с помощью формулы

$$C = 1 - h(p) = 1 + p \log_2 p + (1 - p) \log_2(1 - p).$$

При $p = 0.1$ получаем $C = 0.469$.

Критическая скорость определяется по формуле

$$R_c = 1 - h(p_c), \quad p_c = \frac{\sqrt{p}}{\sqrt{p} + \sqrt{1 - p}}.$$

Подставляя значение $p = 0.1$, получаем $p_c = 0.25$, $h(p_c) = 0.811$, $R_c = 0.189$.

Проверяем соотношение между заданной скоростью передачи и критической скоростью. В данном случае выполняется соотношение: $R \leq R_c$. Это значит, что для оптимальной экспоненты Галлагера надо использовать формулу

$$E_G(R) = R_0 - R, \text{ если } 0 \leq R \leq R_c,$$

где $R_0 = 1 - 2 \log(\sqrt{p} + \sqrt{1-p})$. $R_0 = 0.322$ при $p = 0.1$. Следовательно, $E_G(R) = R_0 - R = 0.322 - 0.120 = 0.202$. Верхняя граница вероятности ошибки на блок равна $2^{-L E_G(R)} = 8.34 \cdot 10^{-7}$.

Задача 2

Условие. Найти экспоненту вероятности ошибки $E(R, C)$ и пропускную способность для двоичного симметричного канала, характеризуемого вероятностью искажения символа $p = 0.1$ при скорости передачи $R = 0.20$. Оценить вероятность ошибки на блок при длине блока $L = 100$.

Решение. Для решения этой задачи используем расчетные формулы, приведенные в главе 10.

Пропускную способность вычислим с помощью формулы

$$C = 1 - h(p) = 1 + p \log_2 p + (1 - p) \log_2(1 - p).$$

При $p = 0.1$ получаем $C = 0.469$.

Критическая скорость определяется по формуле

$$R_c = 1 - h(p_c), \quad p_c = \frac{\sqrt{p}}{\sqrt{p} + \sqrt{1-p}}.$$

Подставляя значение $p = 0.1$, получаем $p_c = 0.25$, $h(p_c) = 0.811$, $R_c = 0.189$.

Проверяем соотношение между заданной скоростью передачи и критической скоростью. В данном случае выполняется соотношение: $R > R_c$. Это значит, что для оптимальной экспонента Галлагера надо использовать параметрическую формулу

$$\begin{aligned} E_G(R) &= -\delta \log_2 p - (1 - \delta) \log_2(1 - p) - h(\delta); \\ R &= 1 - h(\delta), \quad p \leq \delta \leq p_c, \text{ для } R_c \leq R \leq C. \end{aligned}$$

При $R = 0.20$ значение $\delta = 0.243$, $E_G(R) = 0.122$.

Верхняя граница вероятности ошибки на блок равна

$$2^{-L E_G(R)} = 2.08 \cdot 10^{-4}.$$

B.9. Задачи и упражнения к главе 11

Упражнение 1. Спектр функции с ограниченным спектром равен

$$F(\omega) = \begin{cases} \frac{1+\exp(i\pi\frac{\omega}{\Omega})}{2F}, & |\omega| \leq \Omega = 2\pi F; \\ 0, & |\omega| > \Omega = 2\pi F. \end{cases}$$

Найти эту функцию.

Решение. В соответствии с теоремой Котельникова имеем формулу для функции с ограниченным спектром в виде

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) \exp(i\omega t) d\omega.$$

Подставляем заданный спектр в эту формулу и получаем функцию

$$\begin{aligned} f(t) &= \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \frac{1 + \exp(-\frac{i\omega}{2F})}{2F} \exp(i\omega t) d\omega = \\ &= \frac{\sin \Omega t}{\Omega t} + \frac{\sin \Omega(t + \frac{1}{2F})}{\Omega(t + \frac{1}{2F})}. \end{aligned}$$

Упражнение 2. Функция с ограниченным спектром имеет вид

$$f(t) = \frac{\sin \Omega(t + \frac{1}{2F})}{\Omega(t + \frac{1}{2F})} + \frac{\sin \Omega t}{\Omega t}.$$

Найти спектр этой функции.

Решение. Применим теорему Котельникова для нахождения спектра:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \exp(-i\omega t) dt.$$

Подставим заданную функцию $f(t)$ в эту формулу :

$$F(\omega) = \int_{-\infty}^{\infty} \frac{\sin \Omega t}{\Omega t} \exp(-i\omega t) dt + \int_{-\infty}^{\infty} \frac{\sin \Omega(t + \frac{1}{2F})}{\Omega(t + \frac{1}{2F})} \exp(-i\omega t) dt.$$

Каждый из интегралов вычислим в отдельности.

$$\begin{aligned} I_1 &= \int_{-\infty}^{\infty} \frac{\sin \Omega t}{\Omega t} \exp(-i\omega t) dt = \int_{-\infty}^{\infty} \frac{\sin \Omega t}{\Omega t} (\cos \omega t - i \sin \omega t) dt = \\ &= \int_{-\infty}^{\infty} \frac{\sin \Omega t}{\Omega t} \cos \omega t dt - i \int_{-\infty}^{\infty} \frac{\sin \Omega t}{\Omega t} \sin \omega t dt. \end{aligned}$$

Здесь использована формула Эйлера: $\exp(-i\omega t) = \cos \omega t - i \sin \omega t$. Так как интеграл от нечетной функции в симметричных пределах равен нулю, то

$$\int_{-\infty}^{\infty} \frac{\sin \Omega t}{\Omega t} \sin \omega t dt = 0.$$

Преобразуем подынтегральное выражение в I_1 , используя тригонометрические формулы:

$$\sin \Omega t \cos \omega t = \frac{\sin(\Omega + \omega)t + \sin(\Omega - \omega)t}{2}.$$

$$I_1 = \int_{-\infty}^{\infty} \frac{\sin(\Omega + \omega)t}{2\Omega t} dt + \int_{-\infty}^{\infty} \frac{\sin(\Omega - \omega)t}{2\Omega t} dt = I_{11} + I_{12}.$$

Используем табличный интеграл:

$$\int_{-\infty}^{\infty} \frac{\sin at}{t} dt = \pi \operatorname{sign}(a).$$

В результате получим:

$$I_{11} = \int_{-\infty}^{\infty} \frac{\sin(\Omega + \omega)t}{2\Omega t} dt = \begin{cases} \frac{\pi}{2\Omega}, & \omega \geq \Omega, \\ \frac{-\pi}{2\Omega}, & \omega < \Omega, \end{cases}$$

$$I_{12} = \int_{-\infty}^{\infty} \frac{\sin(\Omega - \omega)t}{2\Omega t} dt = \begin{cases} \frac{\pi}{2\Omega}, & \omega < \Omega, \\ \frac{-\pi}{2\Omega}, & \omega \geq \Omega. \end{cases}$$

Суммируя интегралы и учитывая формулу $\Omega = 2\pi F$, получаем:

$$I_1 = I_{11} + I_{12} = \frac{1}{2F}, \quad -2\pi F \leq \omega \leq 2\pi F.$$

Второй интеграл I_2 отличается от интеграла I_1 фиксированным сдвигом по t на $+\frac{1}{2F}$ и, как следствие, дополнительным множителем $\exp(i\frac{\omega}{2F})$:

$$I_2 = \frac{\exp(i\frac{\omega}{2F})}{2F}, \quad -2\pi F \leq \omega \leq 2\pi F.$$

Суммируя интегралы $I = I_1 + I_2$, получаем следующее выражение для спектра:

$$F(\omega) = \begin{cases} \frac{1+\exp(i\pi\frac{\omega}{\Omega})}{2F}, & |\omega| \leq \Omega = 2\pi F; \\ 0, & |\omega| > \Omega = 2\pi F. \end{cases}$$

Задача 1

Условие. Пусть X, Y – независимые одинаково распределенные случайные величины с равномерным распределением в интервале $(-1, 1)$. Пусть $Z = X + Y$. Найти дифференциальные энтропии $H(X), H(Z)$ и взаимную информацию $I(X; Z)$.

Решение.

$$H(X) = - \int_{-1}^1 p_X(x) \log_2(p_X(x)) dx = \int_{-1}^1 \frac{1}{2} \log_2(2) dx = 1.$$

Случайная величина Z имеет треугольное распределение

$$p_Z(z) = \frac{1}{2}(1 - \frac{|z|}{2}), \quad -2 \leq z \leq 2,$$

что приводит к

$$H(Z) = 1 + \frac{\log_2 e}{2}.$$

По определению $I(X; Z) = H(Z) - H(Z|X)$. Но $H(Z|X) = 1$, так как случайная величина Z при каждом заданном значении x имеет равномерное распределение в интервале длины 2. Следовательно,

$$I(X; Z) = 1 + \frac{\log_2 e}{2} - 1 = \frac{\log_2 e}{2}.$$

Задача 2

Условие. Найти пропускную способность системы связи, состоящей из 6 независимых параллельных каналов с дискретным временем и аддитивными гауссовыми шумами с дисперсиями

$$\sigma_1^2 = \sigma_2^2 = 1, \sigma_3^2 = \sigma_4^2 = 3, \sigma_5^2 = 4, \sigma_6^2 = 5,$$

при условии, что суммарная энергия на одну передачу равна $E = 8$. То же, если $E = 16$.

Решение. При оптимальном распределении энергии в параллельных каналах должно выполняться следующее соотношение:

$$B = \sigma_i^2 + E_i, i = \overline{1, s}, s \leq m,$$

где m – число параллельных каналов, $s + 1$ – номер канала, у которого $\sigma_{s+1}^2 \geq B$.

Сначала определим s для $E_0 = 8$:

$$\frac{E_0 + \sum_1^6 \sigma_i^2}{6} = \frac{25}{6} > 4.$$

Значит, канал с номером $i = 6$, у которых $\sigma_5^2 = 5 \geq 4$, энергию не получит. Исключаем его и снова проверяем соотношение:

$$\frac{E_0 + \sum_1^5 \sigma_i^2}{5} = \frac{20}{5} = 4.$$

Теперь исключаем канал с номером $i = 5$, так как у него дисперсия шума равна полученному усредненному значению. Тогда

$$B = \frac{1+1+3+3+8}{4} = 4, s = 4.$$

Распределяем энергию между каналами:

$$E = 8; B = 4; E_1 = E_2 = 3; E_3 = E_4 = 1; E_5 = E_6 = 0.$$

Вычислим пропускную способность по формуле:

$$C = \frac{1}{2} \sum_{i=1}^s \log\left(1 + \frac{E_i}{\sigma_i^2}\right) = \frac{1}{2} \sum_{i=1}^s \log\left(\frac{B}{\sigma_i^2}\right).$$

$$\text{Получим } C = \log_2\left(\frac{B}{1}\right) + \log_2\left(\frac{B}{3}\right) = \log_2(16/3) = 2.415.$$

Для случая $E_0 = 16$ имеем

$$\frac{E_0 + \sum_1^6 \sigma_i^2}{6} = \frac{33}{6} = 5.5.$$

В этом случае $s = m = 6, B = \frac{1+1+3+3+4+5+8}{6} = 5.5$. Энергия распределяется между всеми каналами:

$$E = 16; B = 5.5; \\ E_1 = E_2 = 4.5; E_3 = E_4 = 2.5; E_5 = 1.5; E_6 = 0.5.$$

Пропускная способность в этом случае

$$C = \log_2\left(\frac{B}{1}\right) + \log_2\left(\frac{B}{3}\right) + \frac{1}{2} \log_2\left(\frac{B}{4}\right) + \frac{1}{2} \log_2\left(\frac{B}{5}\right) = 3.632.$$

Окончательно запишем ответ в виде:

$$E = 8; B = 4; E_1 = E_2 = 3; E_3 = E_4 = 1; E_5 = E_6 = 0.$$

$$C = \log_2\left(\frac{B}{1}\right) + \log_2\left(\frac{B}{3}\right) = \log_2\left(\frac{16}{3}\right) = 2.415.$$

$$E = 16; B = 5.5; E_1 = E_2 = 4.5; E_3 = E_4 = 2.5; \\ E_5 = 1.5; E_6 = 0.5.$$

$$C = \log_2\left(\frac{B}{1}\right) + \log_2\left(\frac{B}{3}\right) + \frac{1}{2} \log_2\left(\frac{B}{4}\right) + \frac{1}{2} \log_2\left(\frac{B}{5}\right) = 3.632.$$

Задача 3

Условие. Найти пропускную способность C_m системы связи, состоящей из m независимых параллельных каналов с дискретным временем и аддитивными гауссовыми шумами с одинаковыми дисперсиями σ^2 при условии, что суммарная энергия на одну передачу равна E . Найти предел $C_\infty = \lim_{m \rightarrow \infty} C_m$.

Решение. Запишем общую формулу для пропускной способности C_m системы связи, состоящей из m независимых параллельных каналов:

$$C_m = \frac{1}{2} \sum_{i=1}^m \log_2 \left(1 + \frac{E_i}{\sigma_i^2} \right).$$

Учтем равенство дисперсий $\sigma_i^2 = \sigma^2$, следовательно, равного распределения энергии $E_i = \frac{E}{m}$:

$$C_m = \frac{m}{2} \log_2 \left(1 + \frac{E}{m\sigma^2} \right).$$

Предельное значение пропускной способности при $m \mapsto \infty$ равно

$$C_\infty = \lim_{m \rightarrow \infty} C_m = \frac{E}{2\sigma^2} \log_2 e.$$

Окончательно запишем ответ в виде:

$$C_m = \frac{m}{2} \log_2 \left(1 + \frac{E}{m\sigma^2} \right);$$

$$C_\infty = \frac{E}{2\sigma^2} \log_2 e.$$

Список литературы

- [1] *Shannon C.E.* "A Mathematical Theory of Communication," Bell System Techn. J., 27, 1948, 379-423, 623-656. Русский перевод: Шеннон К. Математическая теория связи // Работы по теории связи и кибернетике: Сб. — М.: ИЛ, 1963.
- [2] *Котельников В.А.* Теория потенциальной помехоустойчивости. — М.: Госэнергоиздат , 1956.
- [3] *Колмогоров А.Н.* Теория передачи информации и теория алгоритмов // Сессия АН СССР по научным проблемам автоматизации производства: Сб. — М.: Наука, 1987.
- [4] *Рябко Б.Я., Фионов А.Н.* Основы современной криптографии. — М.: Научный мир, 2004.
- [5] *Валиев К.А., Кокин А.А.* Квантовые компьютеры: надежды и реальность. — Ижевск: НИЦ РХД, 2001.
- [6] *Колесник В.Д., Полтырев Г.Ш.* Курс теории информации. — М.: Наука, 1982.
- [7] *Галлагер Р.* Теория информации и надежная связь. — М.: Сов. радио, 1974.
- [8] *Фано Р.* Передача информации. Статистическая теория связи. — М.: Мир, 1965.
- [9] *Вернер М.* Основы кодирования: Учебник для вузов/ Пер. с немецкого — М.: Техносфера, 2004.

Учебное издание

*ГАБИДУЛИН Эрнст Мухамедович,
ПИЛИПЧУК Нина Ивановна*

ЛЕКЦИИ
ПО ТЕОРИИ ИНФОРМАЦИИ

Редакторы: *О.П. Котова, Л.В. Себова*
Корректор *В.А. Дружинина*

Подписано в печать 25.09.2007. Формат 60 × 84 1/16. Бумага офсетная.

Печать офсетная. Усл. печ л. 13,5. Уч.- изд. л. 12,0 Тираж 500 экз.

Заказ № 1310.

Государственное образовательное учреждение
высшего профессионального образования
Московский физико-технический институт
(государственный университет)

141700, Московская обл., г. Долгопрудный, Институтский пер., 9

Отпечатано в полном соответствии с качеством
предоставленных диапозитивов
в ГУП МО «Орехово-Зуевская типография»
142603, Московская область, г. Орехово-Зуево, ул. Дзержинского, д. 1

ISBN 5-7417-0197-3

A standard linear barcode representing the ISBN number 5-7417-0197-3.

9 785741 701973