БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Ю. С. Харин И. А. Бодягин Е. В. Вечерко

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ТЕОРИИ ИНФОРМАЦИИ

Допущено
Министерством образования Республики Беларусь
в качестве учебного пособия для студентов
учреждений высшего образования
по специальностям «Компьютерная безопасность»,
«Прикладная криптография»

МИНСК
БГУ
2018

УДК 519.72(075.8)
ББК 22.18я73-1
Х20

Р е ц е н з е н т ы :
кафедра высшей математики Военной академии Республики Беларусь
(заведующий кафедрой доктор технических наук,
профессор *В. А. Липницкий)*;
доктор физико-математических наук, профессор *В. И. Берник*

ISBN 978-985-566-525-1

ПРЕДИСЛОВИЕ

Тот, кто владеет *информацией*,
тот владеет *миром*.
У. Черчилль

На рубеже XX и XXI вв. значение *информации* (от лат. informatio – разъяснение) в человеческом обществе превзошло значение другого крайне важного фактора – *энергии*, игравшего ключевую роль в XX в. Создание глобального *информационного пространства* (*цифрового мира*), а также национальных и корпоративных (отраслевых) пространств – неизбежная необходимость развития информационно-коммуникационных технологий, носящая необратимый характер. Основные источники информации в настоящее время – это Интернет, социальные сети, корпоративные сети, мобильные устройства, данные с автоматических сенсоров, бизнес-компании, финансовые и медицинские регистры, базы генетических данных. Сегодня наблюдается экспоненциальный рост объемов регистрируемой информации: например, в США к 2020 г. прогнозируется иметь в хранилищах примерно 6500 экзобайт информации (1 экзобайт = 10^{18} байт). В цифровом мире *информация* все отчетливее становится высокоценным товаром, который необходимо производить (порождать), хранить (накапливать), обрабатывать (анализировать), транспортировать (передавать) и защищать.

Теория информации – наука, исследующая *информацию*, а также процессы ее хранения, преобразования и передачи математическими методами. Возникновение теории информации стало возможным после того, как было осознано, что несмотря на смысловую разнородность информации, ее количество, абстрагируясь, можно задать числом так же, как можно выразить числом расстояние, время, массу, энергию и другие физические величины.

В последние годы область применения теории информации значительно расширилась, особенно в связи с разработкой компьютерных систем кодирования и защиты информации.

Изучение методов теории информации является необходимым условием подготовки высококвалифицированных специалистов в области информационных технологий и особенно защиты информации. За рубежом лекции по теории информации читают

студентам, обучающимся по специальностям «Математика», «Прикладная математика», «Информатика», «Информационные технологии и защита информации».

В 1997 г. в высших учебных заведениях Республики Беларусь открыта специализация «Математическое и программное обеспечение криптографии и анализа данных», в 2002 г. — специальность «Компьютерная безопасность», а в 2012 г. — «Прикладная криптография». Для подготовки специалистов этих направлений и предназначено учебное пособие «Математические основы теории информации».

При написании данной книги учтен опыт преподавания курса «Теория информации» в Белорусском государственном университете, а также в ведущих зарубежных учебных и научных центрах. Учебное пособие состоит из 9 глав. В главах 1—3 рассматриваются основные понятия теории информации. Глава 4 посвящена стационарным источникам сообщений и их энтропийным свойствам. Марковские источники сообщений, в том числе с использованием новых малопараметрических моделей цепей Маркова высокого порядка, подробно изучаются в главе 5. В главе 6 излагается Шенноновский подход к криптографической защите информации, главах 7—9 — теория кодирования дискретных сообщений. Наряду с теоретическим материалом в учебном пособии приведены более 130 упражнений и 30 тестовых заданий, а также методические указания, решения, ответы к тестам и упражнениям. Литература, представленная в библиографических ссылках, будет полезна студентам не только в процессе обучения, но и для научно-исследовательской работы.

Настоящее издание составляет основу учебно-методического комплекса по дисциплине «Теория информации» для студентов при выполнении учебной программы первой и второй ступени высшего образования. Оно также будет полезно специалистам в области прикладной математики и информационных технологий, желающим познакомиться с методами теории информации.

Авторство в учебном пособии распределено следующим образом: Ю. С. Харин — гл. 1, 2, 3, 4, 5, 6; И. А. Бодягин — гл. 7, 9, разд. 2.3—2.5, 3.3, 3.4, 4.5—4.7, 6.3—6.5, 8.11; Е. В. Вечерко — гл. 8, разд. 1.8—1.10, 5.9, 5.10.

Авторы признательны рецензентам: доктору физико-математических наук профессору В. И. Бернику, доктору технических наук профессору В. А. Липницкому за замечания и рекомендации по улучшению пособия.

Предложения и замечания по содержанию учебного пособия просьба направлять по адресу: кафедра математического моделирования и анализа данных, факультет прикладной математики и информатики, Белорусский государственный университет, пр. Независимости, 4, Минск 220030, Республика Беларусь; тел./факс: +375 17 2095104; e-mail: kharin@bsu.by.

ОСНОВНЫЕ ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

$\lvert A\rvert$	–	модуль числа A, или мощность множества A, или определитель матрицы A, или длина слова A
$\mathbb{N}$	–	множество натуральных чисел
$\mathbb{Z}$	–	кольцо целых чисел
$\mathbb{R}$	–	поле вещественных чисел
$\mathcal{R}$	–	скорость передачи информации
C^*	–	пропускная способность канала связи
A^n	–	множество слов, состоящих из n символов, в алфавите A
A^*	–	множество слов конечной длины в алфавите A
ДСК(p)	–	двоичный симметричный канал с параметром p
$A \times B$	–	декартово произведение множеств A и B
$\mathfrak{D}$	–	декодер общего вида
$\mathfrak{D}_\rho$	–	декодер в ближайшее кодовое слово
$\mathfrak{D}_L$	–	декодер на основе метода максимального правдоподобия
$\mathfrak{D}_S$	–	декодер на основе таблицы стандартного расположения
$d_{\mathcal{C}}$	–	минимальное кодовое расстояние блокового кода $\mathcal{C}$
$\mathbb{F}_q$	–	конечное поле Галуа из q элементов
V_n	–	n-мерное векторное пространство над полем $\mathbb{F}_q$
$\oplus$	–	операция сложения в поле $\mathbb{F}_2$
$\parallel$	–	конкатенация
$[x]$, $\lfloor x \rfloor$	–	целая часть числа $x \in \mathbb{R}$, наибольшее целое $n \leqslant x$
$\lceil x \rceil$	–	наименьшее целое $n \geqslant x$
$\mathsf{I}\{A\}$	—	индикатор наступления события A

$\mathbf{1}_A\{x\} = \mathsf{I}\{x \in A\}$ – индикаторная функция множества A

$\delta_{ij} = \mathsf{I}\{i = j\}$ – символ Кронекера

$\mathsf{H}\{\xi\}$ – энтропия Шеннона случайной величины ξ

$\mathsf{H}\{\xi|\eta\}$ – условная энтропия случайной величины ξ при условии случайной величины η

$\mathsf{I}\{\xi,\eta\}$ – взаимная информация случайных величин ξ и η

$\mathsf{P}\{A\}$ – вероятность наступления случайного события A

$\mathsf{E}\{\xi\}$ – математическое ожидание случайной величины ξ

$\mathsf{D}\{\xi\}$ – дисперсия случайной величины ξ

$\mathsf{Cov}\{\xi, \eta\}$ – ковариация случайных величин ξ, η

$\mathfrak{L}\{\xi\}$ – закон распределения вероятностей случайной величины ξ

$O(\varepsilon), o(\varepsilon)$ – символы Ландау

$\binom{n}{m} = \dfrac{n!}{m!(n-m)!}$ – число сочетаний из n по m

$w(c)$ – вес Хэмминга слова c

$\mathrm{ord}(\cdot)$ – порядок элемента в группе

ЦМ – цепь Маркова

ОЦМ – однородная цепь Маркова

ДВР – дискретный временной ряд

РРСП – равномерно распределенная случайная последовательность

ИДС – источник дискретных сообщений

ИНС – источник непрерывных сообщений

п. н. – почти наверное

$\square$ – конец доказательства

ВВЕДЕНИЕ

Предмет теории информации

В основе теории информации лежит статистическое описание источников сообщений и каналов связи, а также базирующееся на нем измерение количества информации между сообщениями по Шеннону, при котором количество информации определяется только вероятностными свойствами сообщений.

Предметом изучения теории информации являются:

– математические модели источников сообщений и их так называемые энтропийные свойства, характеризующие, насколько следующий порождаемый символ непредсказуем (случаен) при известных предшествующих;

– математические модели, методы и алгоритмы хранения, обработки, передачи и защиты информации, теоретические свойства указанных моделей, методов и алгоритмов;

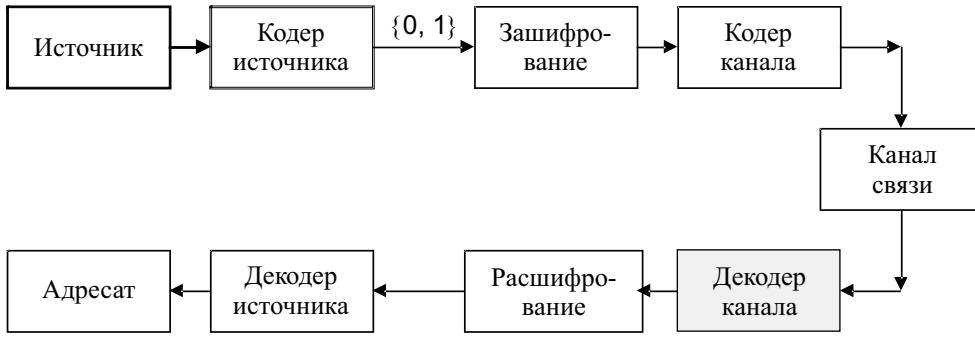
– математические модели и свойства каналов передачи сообщений (каналов связи). К наиболее значимым свойствам относят пропускную способность канала связи.

Теория информации — важный инструмент анализа различных технических систем: телеметрических, защиты информации, передачи речи, изображений или других данных, управления, «data mining» и др. Теория информации позволяет ответить на вопросы о предельной возможности перечисленных систем, определить, в какой мере проектируемая система уступает теоретически возможной.

Основные понятия

Введем определения основных понятий, используемых в теории информации применительно к системе передачи сообщений, обобщенная структурная схема которой представлена на с. 8.

Источник – это лицо или устройство, порождающее некоторое сообщение, подлежащее передаче в точку приема адресата. Наиболее распространены в настоящее время следующие типы сообщений: аудиосигналы (речь, музыка), изображения, видеопоследовательности.



Кодер источника – устройство, преобразующее передаваемые сообщения во взаимно-однозначно соответствующие им *элементарные последовательности двоичных символов*.

Зашифрование – математическое преобразование открытого сообщения в шифрованное сообщение с помощью функции зашифрования.

Кодер канала (передатчик) – устройство, преобразующее двоичную последовательность шифрованного сообщения в передаваемый по каналу *сигнал*, соответствующий передаваемому сообщению.

Канал связи – техническое устройство или физическая среда, используемые для передачи сигнала (например, провода, коаксиальный кабель, эфир с заданной полосой радиочастот); шумы (помехи), действующие в канале, могут приводить к искажению сигнала.

Декодер канала (приемник) – устройство, преобразующее полученный на выходе канала сигнал в двоичную последовательность.

Расшифрование – математическое преобразование, обратное зашифрованию.

Декодер источника – устройство, преобразующее двоичную последовательность в сообщение.

Адресат – это лицо или устройство, которому предназначено передаваемое сообщение.

Существуют два основных класса источников:

– класс *дискретных источников*, порождающих дискретные последовательности символов из конечного алфавита;

– класс *непрерывных источников*, порождающих функции времени.

Приведем примеры важных прикладных задач, решаемых методами теории информации.

Пример 1 (задача оптимального кодирования). Задан некоторый источник дискретных сообщений. Требуется найти наименьшее число двоичных символов, необходимых для кодирования последовательности сообщений, порождаемых этим источником, и построить алгоритм оптимального кодирования.

Пример 2 (задача криптографии). Требуется построить такую функцию шифрования, чтобы количество информации об исходном сообщении, содержащейся в шифрованном сообщении, было минимально возможным.

История развития теории информации

Приведем краткий обзор истории развития теории информации [4].

Годом рождения теории информации принято считать 1948 г., когда американский ученый К. Шеннон опубликовал статью «Математическая теория связи» [34] о закономерностях передачи информационных сообщений по каналам связи. Однако некоторые теоретические основы были заложены ранее, в первой трети XX в.

Один из базовых законов теории связи был установлен еще в 1924 г. Г. Найквистом и К. Купфмюллером. Этот закон сформулировал критерий того, что информация непрерывного сигнала не будет утрачена и может быть полностью восстановлена при передаче с помощью телеграфного сигнала. К знаковым работам, относящимся к предыстории теории информации, относится работа академика В. А. Котельникова, который в 1933 г. доказал известную теорему, позволяющую представить любую функцию с ограниченной полосой частот F своими отсчетами, взятыми через интервалы времени $1/(2F)$. В 1948 г. этот же вывод независимо был сделан К. Шенноном. Данные результаты оказали огромное влияние на прогресс в области телекоммуникаций, связанных с переходом от аналоговых систем передачи сообщений к цифровым.

В области сокращения избыточности сообщений еще в 1837 г. С. Морзе создал код, который был построен таким образом, что наиболее часто встречающимся в тексте буквам английского алфавита были присвоены наиболее короткие последовательности из точек и тире. Он позволял передавать телеграфные сообщения по каналу связи, используя наименьшее в среднем число символов. Этот код является одним из первых практических примеров статистического кодирования дискретного источника сообщений. При этом он до сих пор находит применение. К. Шеннон в своих работах открыл общие принципы сжатия данных, относящихся к источникам информации произвольной природы.

Понятие количества информации, содержащейся в сообщении, впервые ввел американский ученый Р. Хартли в 1928 г. До него само понятие «сообщение» носило расплывчатый характер. Р. Хартли предполагал, что алфавит сообщения состоит из K равновероятных символов, и предложил определять информацию, содержащуюся в одном передаваемом символе сообщения, величиной $\log K$. Эта идея в дальнейшем была обобщена К. Шенноном на случай неравновероятных символов.

Стоит отметить, что статистический подход к описанию сообщений также впервые был предложен еще до публикации работы К. Шеннона. В. И. Сифоров и Н. Винер в своих работах рассматривали проблемы передачи информации с точки зрения статистики.

В работе К. Шеннона «Математическая теория связи» [34] в 1948 г. были обобщены отдельные разрозненные идеи предшественников. В ней давались четкие математические описания проблем, стоящих перед теорией связи, и указывались возможные математические пути их решения. Поскольку теория информации зародилась при решении задач электросвязи и изначально не считалась

математической теорией, то ее значение для математики не сразу было понятно. На первых порах к идеям К. Шеннона многие относились со скептицизмом. Однако уже в середине 50-х гг. XX в. за разработку математических аспектов теории информации взялись крупные математики. Из американских ученых можно выделить Б. Мак-Миллана, А. Файнстейна, Дж. Вольфовица. В это же время в СССР задачами математической теории информации занимались такие ученые, как А. Н. Колмогоров, А. Я. Хинчин, А. М. Яглом, Р. Л. Стратонович.

Важную роль в области кодирования дискретных источников сообщений сыграли такие математики, как Р. Фано, Д. Хаффман, Л. Крафт. В 1965 г. А. Н. Колмогоров предложил рассматривать «универсальное» кодирование, обозначающее кодирование источников сообщений, статистические характеристики которых априорно неизвестны. Один из наиболее известных алгоритмов универсального кодирования был предложен в 1978 г. израильскими учеными Я. Зивом и А. Лемпелем.

Параллельно активно развивалась теория кодирования источников аналоговых сигналов. В этом направлении можно выделить таких математиков, как П. Элайс, Г. Крамер, Э. Д. Витерби.

Сегодня большинство математических результатов теории информации активно используется во многих прикладных сферах человеческой деятельности, связанных с информацией: мобильная и телефонная связь, телевидение, передача информации посредством Интернета, защита информации. Для компактного хранения данные архивируются с помощью так называемых оптимальных кодов. Все более активно математические методы теории информации применяются в криптологии [14].

Глава 1

ФУНКЦИОНАЛЫ ЭНТРОПИИ И ИНФОРМАЦИИ

1.1. ИСТОЧНИКИ ДИСКРЕТНЫХ СООБЩЕНИЙ И ИХ ВЕРОЯТНОСТНЫЕ МОДЕЛИ

Рассмотрим произвольный источник сообщений. Каждое сообщение представляет собой некоторую последовательность символов (например, букв белорусского алфавита, точек и тире в телеграфии, нулей и единиц в компьютерной логике и т. д.). Отдельный *символ сообщения* обозначим ξ и предположим его числовой величиной, принимающей всевозможные значения из некоторого множества $A \subset \mathbb{R}$, $\xi \in A$.

Множество A значений символа сообщения ξ принято называть *алфавитом сообщений*. Если алфавит A является конечным множеством мощности $2 \leqslant N < \infty$:

$$A = \left\{a^{(1)}, a^{(2)}, \ldots, a^{(N)}\right\},$$

то принято говорить, что имеет место *источник дискретных сообщений* (ИДС). В противном случае говорят об *источнике непрерывных сообщений* (ИНС).

В этом разделе рассматриваются модели ИДС, наиболее часто используемые в современных криптосистемах, а модели ИНС — в разд. 1.5.

Величины $a^{(1)}, \ldots, a^{(N)}$ называют *символами алфавита*, а число N — *мощностью алфавита*. Появление в сообщении любого символа алфавита характеризуется высокой степенью неопределенности. Для математического описания этой неопределенности воспользуемся дискретной вероятностной моделью. Пусть $(\Omega, \mathcal{F}, \mathsf{P})$ — основное вероятностное пространство, описывающее случайный «эксперимент» по появлению (регистрации) символа ξ. Здесь Ω — пространство элементарных событий (исходов случайного эксперимента), $\mathcal{F}$ — σ-алгебра (или алгебра) событий из Ω, $\mathsf{P} : \mathcal{F} \to [0,1]$ — вероятностная мера, определенная на $\mathcal{F}$. Тогда каждому элементарному исходу $\omega \in \Omega$ этого эксперимента ставится в соответствие значение символа $\xi = \xi(\omega) \in A$. Таким образом, символ $\xi = \xi(\omega)$ — дискретная случайная величина, полностью определяемая дискретным распределением вероятностей:

$$\mathsf{P}\{\xi = a\} = p(a),\ a \in A, \tag{1.1}$$

$$0 < p\,(a) < 1,\ \sum_{i=1}^{N} p\left(a^{(i)}\right) = 1. \tag{1.2}$$

В качестве примера в табл. 1.1 представлено распределение вероятностей символов русского алфавита (упорядоченных в порядке убывания $p\left(a^{(i)}\right)$).

Таблица 1.1

Символ	пробел	о	е, ё	а	и	т	н	с
Вероятность	0,175	0,090	0,072	0,062	0,062	0,053	0,053	0,045
Символ	р	в	л	к	м	д	п	у
Вероятность	0,040	0,038	0,035	0,028	0,026	0,025	0,023	0,021
Символ	я	ы	з	ь, ъ	б	г	ч	й
Вероятность	0,018	0,016	0,016	0,014	0,014	0,013	0,012	0,010
Символ	х	ж	ю	ш	ц	щ	э	ф
Вероятность	0,009	0,007	0,006	0,006	0,004	0,003	0,003	0,002

Таким образом, ИДС в случае односимвольного сообщения описывается дискретной вероятностной моделью $\langle A, p\,(a)\rangle$, которую называют *дискретной вероятностной схемой* [14]. Эта модель описывает лишь одиночный случайный символ сообщения. Сообщение, порождаемое ИДС, — это в общем случае последовательность $n \geqslant 1$ случайных символов:

$$\Xi_n = (\xi_1,\, \ldots, \xi_n) \in A^n.$$

При этом полное вероятностное описание ИДС задается вероятностной моделью случайного временного ряда (случайного процесса) Ξ_n с дискретным временем $t \in \mathbb{N}$ и дискретным пространством состояний A ($n = 1, 2, \ldots$):

$$\langle A^n, p_n(a_1,\, \ldots, a_n)\rangle\,,$$
$$p_n(a_1,\, \ldots, a_n) = \mathsf{P}\left\{\xi_1 = a_1,\, \ldots, \xi_n = a_n\right\}, a_1,\, \ldots, a_n \in A\,, \tag{1.3}$$

где $p_n(a_1, \ldots, a_n)$ — n-мерное дискретное распределение вероятностей n-символьного сообщения. Отметим, что n-мерные распределения вероятностей (1.3) удовлетворяют условию самосогласованности ($1 \leqslant k_1 < \ldots < k_m \leqslant n$, $1 \leqslant m < n$):

$$p_m(a_{k_1},\, \ldots, a_{k_m}) = \sum_{a_i \in A,\, i \in \{1,\ldots,n\} \setminus \{k_1,\ldots,k_m\}} p_n(a_1,\, \ldots, a_n)\,. \tag{1.4}$$

ИДС называется *стационарным*, если случайный процесс Ξ_n является стационарным (в узком смысле), т. е. если конечномерные распределения (1.3) инвариантны относительно сдвига начала отсчета времени.

Стационарный ИДС называется *дискретным источником без памяти* (ДИБП), если для любых $a_1, \ldots, a_n \in A$ справедлива факторизация n-мерного распределения вероятностей [14]:

$$p_n(a_1,\, \ldots, a_n) = \prod_{i=1}^{n} p\,(a_i)\,, \tag{1.5}$$

т. е. порождаемые ИДС случайные символы $\xi_1, \ldots, \xi_n$ независимы в совокупности и одинаково распределены.

1.2. ФУНКЦИОНАЛ ЭНТРОПИИ И ЕГО СВОЙСТВА

Пусть ИДС описывается некоторой дискретной вероятностной моделью $\langle A, p(a)\rangle$. Тогда *энтропией* ИДС (или энтропией случайного символа ξ) называется величина, определяемая функционалом [8, 24, 25]:

$$\mathbf{H}\{\xi\} = h\left(p\left(a^{(1)}\right), \ldots, p\left(a^{(N)}\right)\right) ::= \mathbf{E}\{-\log_b p(\xi)\} =$$

$$= -\sum_{i=1}^{N} p\left(a^{(i)}\right) \log_b p\left(a^{(i)}\right), \tag{1.6}$$

где $\mathbf{E}\{\cdot\}$ — символ математического ожидания. Если в (1.6) логарифм берется по основанию $b = 2$, то энтропия измеряется в *битах* (от англ. **bi**nary digi**t**, bit), а если используется натуральный логарифм по основанию $b = e$, то энтропия измеряется в *натах* (от англ. **na**tural digi**t**, nat). В случаях, где это не вызовет неоднозначности, будем опускать основание логарифма. Можно поставить любое требуемое основание b и измерять энтропию в необходимых единицах. В большинстве примеров будем пользоваться логарифмом по основанию $b = 2$ и, следовательно, измерять энтропию в битах.

Заметим, что встречающаяся в (1.6) неопределенность $0 \log 0$ разрешается следующим образом: $0 \log 0 ::= 0$.

Сформулируем и докажем основные свойства функционала энтропии.

Свойство 1.1. Функционал энтропии обладает свойствами непрерывности, симметричности, принимает неотрицательные значения: $\mathbf{H}\{\xi\} \geqslant 0$; он обращается в 0 только для вырожденного распределения:

$$\exists a' \in A,\ p(a') = 1,\ p(a) = 0,\ a \neq a'. \tag{1.7}$$

Доказательство. Поскольку $0 \leqslant p(\xi) \leqslant 1$, то $\eta = -\log p(\xi) \geqslant 0$. Согласно свойству математического ожидания из (1.6) имеем

$$\mathbf{H}\{\xi\} = \mathbf{E}\{\eta\} \geqslant 0,$$

причем $\mathbf{H}\{\xi\} = 0$ тогда и только тогда, когда $\eta \overset{\text{п.н.}}{=} 0$. Последнее соотношение, очевидно, выполняется лишь в случае (1.7). Непрерывность и симметричность следуют из (1.6) и свойств логарифма. □

Заметим, что вырожденное распределение (1.7) соответствует случаю, когда символ ξ не является случайным: $\xi = a'$ с вероятностью 1 ($\xi \overset{\text{п.н.}}{=} \text{const}$).

Свойство 1.2. Энтропия ИДС с алфавитом мощности $N < \infty$ имеет максимальное значение

$$\max_{p(\cdot)} \mathbf{H}\{\xi\} = \log N, \tag{1.8}$$

которое достигается, если дискретное распределение вероятностей $p(\cdot)$ — равномерное, т. е. все N значений символов равновероятны:

$$p(a) = \frac{1}{N},\ a \in A\,. \tag{1.9}$$

Доказательство. Воспользуемся неравенством Йенсена [30]

$$\mathsf{E}\left\{f(\zeta)\right\} \leqslant f\left(\mathsf{E}\left\{\zeta\right\}\right), \tag{1.10}$$

которое справедливо для любой случайной величины ζ и произвольной вогнутой функции $y = f(x)$. Положим, в (1.10)

$$\zeta = \frac{1}{p(\xi)}, f(x) = \log_b x\,, \tag{1.11}$$

причем $f''(x) = -\dfrac{1}{(\ln b)x^2} < 0$, так что $f(\cdot)$ — вогнута. Используя (1.6) и (1.11), имеем

$$\mathsf{E}\left\{\zeta\right\} = \sum_{i=1}^{N} p\left(a^{(i)}\right) \frac{1}{p\left(a^{(i)}\right)} = N, \quad f(\mathsf{E}\left\{\zeta\right\}) = \log N,$$

$$\mathsf{E}\left\{f(\zeta)\right\} = \sum_{i=1}^{N} p\left(a^{(i)}\right) \log \frac{1}{p\left(a^{(i)}\right)} = \mathsf{H}\left\{\xi\right\}.$$

Подставляя эти выражения в (1.10), получим (1.8). Вычислив (1.6) с учетом (1.9), найдем

$$h\left(\frac{1}{N}, \ldots, \frac{1}{N}\right) = -\sum_{i=1}^{N} \frac{1}{N} \log \frac{1}{N} = \log N.$$

□

Отметим, что впервые энтропия была введена Д. Хартли в 1928 г. в виде

$$\mathsf{H}\left\{\xi\right\} = \log N \tag{1.12}$$

для случайного символа ξ с N равновероятными значениями, и поэтому (1.12) иногда называют *энтропией Хартли*. Энтропия в общем виде (1.6) называется *энтропией Шеннона*.

Следствие 1.1. *Чем больше мощность алфавита N, тем больше энтропия Хартли (максимально возможная энтропия).*

Свойство 1.3 (свойство аддитивности). Если случайные символы $\xi_1 \in A_1$, $\xi_2 \in A_2$ сообщения независимы, то совместная энтропия равна сумме энтропий:

$$\mathsf{H}\left\{\xi_1, \xi_2\right\} = \mathsf{H}\left\{\xi_1\right\} + \mathsf{H}\left\{\xi_2\right\}. \tag{1.13}$$

Доказательство. Построим на вероятностном пространстве $(\Omega, \mathcal{F}, \mathsf{P})$ случайный вектор $\Xi_2 = (\xi_1, \xi_2) \in A_1 \times A_2$ с дискретным распределением вероятностей:

$$p_2(a_1, a_2) = \mathsf{P}\left\{\xi_1 = a_1, \xi_2 = a_2\right\}, \sum_{a_1 \in A_1} \sum_{a_2 \in A_2} p_2(a_1, a_2) = 1.$$

В силу независимости ξ_1, ξ_2 (ИДС без памяти) имеем

$$p_2(a_1, a_2) = p(a_1)p(a_2),\ a_1 \in A_1, a_2 \in A_2. \tag{1.14}$$

Тогда из (1.6) и (1.14) следует (с учетом условия нормировки):

$$\begin{aligned}
\mathsf{H}\{\xi_1, \xi_2\} &= -\sum_{a_1 \in A_1} \sum_{a_2 \in A_2} p_2(a_1, a_2) \log p_2(a_1, a_2) = \\
&= -\sum_{a_1 \in A_1} p(a_1) \sum_{a_2 \in A_2} p(a_2) \left(\log p(a_1) + \log p(a_2)\right) = \\
&= -\sum_{a_1 \in A_1} p(a_1) \log p(a_1) - \sum_{a_2 \in A_2} p(a_2) \log p(a_2) = \\
&= \mathsf{H}\{\xi_1\} + \mathsf{H}\{\xi_2\},
\end{aligned}$$

что совпадает с (1.13). □

Следствие 1.2. *Если независимы в совокупности $n \geqslant 2$ случайных символов $\xi_1 \in A_1, \ldots, \xi_n \in A_n$, то их совместная энтропия аддитивна:*

$$\mathsf{H}\{\xi_1, \ldots, \xi_n\} = \mathsf{H}\{\xi_1\} + \ldots + \mathsf{H}\{\xi_n\}. \tag{1.15}$$

Доказательство. Оно состоит в $(n-1)$-кратном применении свойства 1.3. □

Свойство 1.4. Добавление к алфавиту символов одного символа с нулевой вероятностью, а следовательно, и любого количества таких символов не изменяет энтропии ИДС.

Доказательство. Пусть алфавит A расширен, как указано в условии:

$$A = \left\{a^{(1)}, \ldots, a^{(N)}\right\},\ A' = A \cup \left\{a^{(N+1)}\right\},\ \mathsf{P}\left\{\xi = a^{(N+1)}\right\} = 0.$$

Тогда согласно (1.6)

$$\mathsf{H}\{\xi'\} = -\sum_{i=1}^{N+1} p\left(a^{(i)}\right) \log p\left(a^{(i)}\right) = -\sum_{i=1}^{N} p\left(a^{(i)}\right) \log p\left(a^{(i)}\right) = \mathsf{H}\{\xi\}.$$

□

Свойство 1.5. Функция $\mathsf{H}(p_1, \ldots, p_N) = -\sum_{i=1}^{N} p_i \log p_i$, определяемая (1.6), строго вогнута на симплексе

$$S_{N-1} = \left\{(p_1, \ldots, p_N) : p_i \geqslant 0,\ \sum_{i=1}^{N} p_i = 1\right\}.$$

Доказательство. Во-первых, область определения энтропии — симплекс S_{N-1} — выпукла. Во-вторых, $H(\cdot)$ представима в виде суммы функций одной переменной (т. е. сепарабельна): $\mathsf{H}(p_1, \ldots, p_N) = \sum_{i=1}^{N} f(p_i)$, где $f(x) = -x \log x$ —

строго вогнутая функция при $x > 0$, так как $f''(x) = -(\log e)/x < 0$. Поскольку сумма строго вогнутых функций на выпуклом множестве строго вогнута, то доказываемое свойство верно. □

1.3. УСЛОВНАЯ ЭНТРОПИЯ И ЕЕ СВОЙСТВА

Чтобы изучить новые важные свойства энтропии, используемые в криптосистемах, нам понадобится понятие условной энтропии.

Пусть определен случайный вектор символов $\xi = (\xi_j) \in A_1 \times \cdots \times A_n$ с некоторым n-мерным дискретным распределением вероятностей

$$p_n(a_1, \ldots, a_n) = \mathsf{P}\{\xi_1 = a_1, \ldots, \xi_n = a_n\}, \tag{1.16}$$

где $a_1 \in A_1, \ldots, a_n \in A_n$. Пусть задано натуральное число $2 \leqslant k \leqslant n$ и определено $(n-k+1)$-мерное условное распределение вероятностей подвектора $\xi' = (\xi_k, \ldots, \xi_n) \in A_k \times \cdots \times A_n$ при условии, что фиксирован подвектор $\xi = (\xi_1, \ldots, \xi_{k-1}) \in A_1 \times \cdots \times A_{k-1}$:

$$p_{n-k+1}(a_k, \ldots, a_n \mid a_1, \ldots, a_{k-1}) = \frac{p_n(a_1, \ldots, a_n)}{p_{k-1}(a_1, \ldots, a_{k-1})}. \tag{1.17}$$

Здесь использована формула умножения вероятностей.

Условной энтропией подвектора ξ' *при условии*, что фиксирован подвектор ξ, называется функционал

$$\begin{aligned}
&\mathsf{H}\{\xi_k, \ldots, \xi_n \mid \xi_1 = a_1, \ldots, \xi_{k-1} = a_{k-1}\} = \\
&= -\sum_{a_k \in A_k} \cdots \sum_{an \in An} p_{n-k+1}(a_k, \ldots, a_n \mid a_1, \ldots, a_{k-1}) \times \\
&\qquad \times \log p_{n-k+1}(a_k, \ldots, a_n \mid a_1, \ldots, a_{k-1}).
\end{aligned} \tag{1.18}$$

Условной энтропией случайного подвектора символов $\xi' = (\xi_k, \ldots, \xi_n)$ *относительно* случайного подвектора символов $\xi = (\xi_1, \ldots, \xi_{k-1})$ называется функционал, получающийся усреднением (1.18):

$$\begin{aligned}
\mathsf{H}\{\xi' \mid \xi\} &= \sum_{a_1 \in A_1} \cdots \sum_{a_{k-1} \in A_{k-1}} p_{k-1}(a_1, \ldots, a_{k-1}) \times \\
&\times \mathsf{H}\{\xi_k, \ldots, \xi_n \mid \xi_1 = a_1, \ldots, \xi_{k-1} = a_{k-1}\} = \\
&= -\sum_{a_1 \in A_1} \cdots \sum_{an \in An} p_n(a_1, \ldots, a_n) \times \\
&\times \log p_{n-k+1}(a_k, \ldots, a_n \mid a_1, \ldots, a_{k-1}) \geqslant 0.
\end{aligned} \tag{1.19}$$

Продолжим исследование свойств энтропии и условной энтропии с учетом введенных понятий.

Теорема 1.1. *Если подвекторы случайных символов ξ', ξ независимы, то условная энтропия совпадает с безусловной:*

$$\mathsf{H}\{\xi' \mid \xi\} = \mathsf{H}\{\xi'\}. \tag{1.20}$$

Доказательство. В силу независимости ξ', ξ условное распределение вероятностей совпадает с безусловным:

$$p_{n-k+1}\left(a_k,\,\ldots,a_n \mid a_1,\,\ldots,a_{k-1}\right) = p_{n-k+1}(a_k,\,\ldots,a_n)\,.$$

Подставляя это выражение в (1.19) и используя свойство самосогласованности распределений

$$\sum_{a_1\in A_1}\cdots\sum_{a_{k-1}\in A_{k-1}} p_n(a_1,\,\ldots,a_n) = p_{n-k+1}(a_k,\,\ldots,a_n)\,,$$

получим (1.20). □

Теорема 1.2. *Для любой последовательности случайных символов сообщения $\xi_1,\,\ldots,\xi_n$ энтропия обладает свойством иерархической аддитивности:*

$$\begin{aligned}\mathsf{H}\left\{\xi_1,\,\ldots,\xi_n\right\} &= \mathsf{H}\left\{\xi_1\right\} + \mathsf{H}\left\{\xi_2 \mid \xi_1\right\} + \\ &+ \mathsf{H}\left\{\xi_3 \mid (\xi_1,\xi_2)\right\} + \,\ldots\, + \mathsf{H}\left\{\xi_n \mid (\xi_1,\,\ldots,\xi_{n-1})\right\}.\end{aligned} \tag{1.21}$$

Доказательство. Воспользуемся обобщенной формулой умножения вероятностей (свойством иерархической мультипликативности вероятностей):

$$p_n(a_1,\,\ldots,a_n) = p_1(a_1)p_1(a_2 \mid a_1)\,\cdots\,p_1(a_n \mid (a_1,\,\ldots,a_{n-1})).$$

Тогда по определению энтропии с учетом свойств вероятности имеем

$$\begin{aligned}\mathsf{H}\left\{\xi_1,\,\ldots,\xi_n\right\} = &-\sum_{a_1\in A_1} p_1(a_1)\log p_1(a_1) - \\ &-\sum_{a_1\in A_1}\sum_{a_2\in A_2} p_2(a_1,a_2)\log p_1(a_2 \mid a_1) - \,\ldots\, - \\ &-\sum_{a_1\in A_1}\cdots\sum_{a_n\in A_n} p_n(a_1,\,\ldots,a_n)\log p_1(a_n \mid (a_1,\,\ldots,a_{n-1}))\,.\end{aligned}$$

Учитывая (1.19), придем к (1.21). □

Следствие 1.3. *Если случайные символы сообщения $\xi_1,\,\ldots,\xi_n$ независимы в совокупности, то выполняется доказанное в предыдущем пункте свойство аддитивности (1.15):*

$$\mathsf{H}\left\{\xi_1,\,\ldots,\xi_n\right\} = \sum_{i=1}^{n}\mathsf{H}\left\{\xi_i\right\}.$$

Доказательство. Согласно теореме 1.1 условные энтропии совпадают с безусловными:

$$\mathsf{H}\left\{\xi_i \mid (\xi_1,\,\ldots,\xi_{i-1})\right\} = \mathsf{H}\left\{\xi_i\right\},\,\big(i=\overline{1,n}\,\big).$$

Используя с учетом этого факта теорему 1.2, придем к доказываемому. □

Следует заметить, что свойство иерархической аддитивности (1.21) и его следствие порождаются наличием логарифмической функции в определении энтропии; это свойство является ключевым для функционала энтропии.

Теорема 1.3. *Пусть $\xi = (\xi_i) \in A$ — произвольный случайный вектор символов с дискретным распределением вероятностей $p_n(a)$, $a = (a_i) \in A$, а $q_n(a)$, $a \in A$ — некоторое дискретное распределение вероятностей. Тогда справедливо неравенство*

$$J(p_n : q_n) = \sum_{a \in A} p_n(a) \log \frac{p_n(a)}{q_n(a)} \geqslant 0\,. \tag{1.22}$$

Равенство нулю имеет место тогда и только тогда, когда распределение $q_n(\cdot)$ совпадает с $p_n(\cdot)$:

$$q_n(a) = p_n(a),\ a \in A\,. \tag{1.23}$$

Доказательство. Воспользуемся известным неравенством Йенсена, справедливым для произвольной случайной величины ζ и выпуклой вверх функции $f(x)$ [30]:

$$\mathsf{E}\,\{f(\zeta)\} \leqslant f(\mathsf{E}\,\{\zeta\})\,. \tag{1.24}$$

Положим, в (1.24)

$$\zeta = \frac{q_n(\xi)}{p_n(\xi)} \geqslant 0\,,\, f(x) = \log x\,.$$

Тогда, используя условие нормировки, имеем

$$\mathsf{E}\,\{\zeta\} = \sum_{a \in A} p_n(a) \frac{q_n(a)}{p_n(a)} = \sum_{a \in A} q_n(a) = 1,$$

$$\mathsf{E}\,\{f(\zeta)\} = \sum_{a \in A} p_n(a) \log \frac{q_n(a)}{p_n(a)} = -J(p_n : q_n)\,.$$

Подставляя эти выражения в (1.24), получим неравенство

$$J(p_n : q_n) \geqslant 0\,.$$

Как известно, равенство в неравенстве Йенсена имеет место тогда и только тогда, когда $\zeta \equiv \mathrm{const} = c$. В силу условия нормировки константа c может быть равна только единице:

$$c = \frac{q_n(a)}{p_n(a)} \equiv 1\,,$$

что и означает $q_n(\cdot) \equiv p_n(\cdot)$. □

Теорема 1.4. *Условная энтропия не может превосходить безусловную:*

$$\mathsf{H}\left\{\xi' \mid \xi\right\} \leqslant \mathsf{H}\left\{\xi'\right\}. \tag{1.25}$$

Доказательство. Воспользуемся теоремой 1.3 и положим, что в (1.22)

$$p(a) = \mathsf{P}\left\{\xi' = a \mid \xi = b\right\},\ q(a) = \mathsf{P}\left\{\xi' = a\right\},\ a \in A, b \in B.$$

Тогда получим

$$\sum_{a \in A} \mathsf{P}\left\{\xi' = a \mid \xi = b\right\} \log \frac{\mathsf{P}\left\{\xi' = a \mid \xi = b\right\}}{\mathsf{P}\left\{\xi' = a\right\}} \geqslant 0,$$

или (что эквивалентно)

$$-\sum_{a\in A}\mathsf{P}\left\{\xi'=a\mid\xi=b\right\}\log\mathsf{P}\left\{\xi'=a\mid\xi=b\right\}\leqslant$$
$$\leqslant-\sum_{a\in A}\mathsf{P}\left\{\xi'=a\mid\xi=b\right\}\log\mathsf{P}\left\{\xi'=a\right\}.$$

Умножим обе части этого неравенства на $\mathsf{P}\{\xi=b\}\geqslant 0$ и просуммируем по всевозможным $b\in B$ (с учетом формулы умножения вероятностей):

$$-\sum_{a\in A}\sum_{b\in B}\mathsf{P}\left\{\xi'=a\mid\xi=b\right\}\log\mathsf{P}\left\{\xi'=a\mid\xi=b\right\}\leqslant$$
$$\leqslant-\sum_{a\in A}\sum_{b\in B}\mathsf{P}\left\{\xi'=a,\,\xi=b\right\}\log\mathsf{P}\left\{\xi'=a\right\}=$$
$$=-\sum_{a\in A}\mathsf{P}\left\{\xi'=a\right\}\log\mathsf{P}\left\{\xi'=a\right\},$$

что совпадает с (1.25). □

Следствие 1.4. *При добавлении условий условная энтропия не увеличивается:*

$$\mathsf{H}\left\{\xi\mid(\eta,\zeta)\right\}\leqslant\mathsf{H}\left\{\xi\mid\eta\right\}. \tag{1.26}$$

Доказательство. Неравенство (1.26) доказывается аналогично теореме 1.4. □

Следствие 1.5. *Энтропия последовательности случайных символов сообщения $\xi_1,\ldots,\xi_n$ не превосходит суммы энтропий всех этих символов, рассматриваемых по отдельности:*

$$\mathsf{H}\left\{\xi_1,\ldots,\xi_n\right\}\leqslant\sum_{i=1}^{n}\mathsf{H}\left\{\xi_i\right\}. \tag{1.27}$$

Доказательство. Согласно теореме 1.4 справедливы неравенства

$$\mathsf{H}\left\{\xi_i\mid(\xi_1,\ldots,\xi_{i-1})\right\}\leqslant\mathsf{H}\left\{\xi_i\right\}\ \left(i=\overline{2,n}\right).$$

Подставляя их в (1.21), получим (1.27). □

В криптологии дискретные сообщения $\xi\in A$ часто подвергаются *дискретным функциональным преобразованиям* (ДФП):

$$\eta=f(\xi),\xi\in A,\eta\in B, \tag{1.28}$$

где A, B — некоторые конечные множества. Исследуем, как изменяется энтропия сообщения при таких функциональных преобразованиях.

Теорема 1.5. *При дискретном функциональном преобразовании вида (1.28) энтропия не возрастает:*

$$\mathsf{H}\left\{\eta\right\}\leqslant\mathsf{H}\left\{\xi\right\}, \tag{1.29}$$

причем равенство в (1.29) достигается тогда и только тогда, когда ДФП (1.28) — биекция.

Доказательство. Рассмотрим «составное» сообщение $\begin{pmatrix}\xi\\ \eta\end{pmatrix} \in A \times B$. По теореме 1.2

$$\mathbf{H}\{\xi,\eta\} = \mathbf{H}\{\xi\} + \mathbf{H}\{\eta \mid \xi\} = \mathbf{H}\{\eta\} + \mathbf{H}\{\xi \mid \eta\}. \tag{1.30}$$

В силу функциональной зависимости (1.28) условное распределение η при условии, что $\xi = a$ фиксировано, является вырожденным:

$$\mathbf{P}\{\eta = b \mid \xi = a\} = \mathbf{P}\{f(a) = b \mid \xi = a\} = \delta_{f(a),\,b},\ b \in B,$$

где $\delta_{i,\,j}$ — символ Кронеккера. Согласно свойству 1.1 энтропия для вырожденного распределения обращается в 0:

$$\mathbf{H}\{\eta \mid \xi\} = 0.$$

Тогда из последнего равенства (1.30) имеем

$$\mathbf{H}\{\eta\} = \mathbf{H}\{\xi\} - \mathbf{H}\{\xi \mid \eta\}.$$

Поскольку $\mathbf{H}\{\xi \mid \eta\} \geqslant 0$ по свойству энтропии, то отсюда следует (1.28). Равенство в (1.29) будет тогда и только тогда, когда $\mathbf{H}\{\xi \mid \eta\} = 0$. Это возможно лишь в случае функциональной зависимости ξ от η: $\xi = f^{-1}(\eta)$, т. е. когда (1.28) — биекция. □

1.4. АКСИОМАТИЧЕСКОЕ ОПРЕДЕЛЕНИЕ ЭНТРОПИИ

Введенный и исследованный в разд. 1.2 и 1.3 функционал энтропии Шеннона

$$\mathbf{H}\{\xi\} = H_m(p_1, \ldots, p_m) = -\sum_{i=1}^{m} p_i \log p_i \tag{1.31}$$

для ИДС $\langle A, \{p_i\}\rangle$, $p_i = \mathbf{P}\left\{\xi = a^{(i)}\right\}$, $i = (1, \ldots, m)$, обладает всеми свойствами, которые необходимо требовать от количественной меры неопределенности, поэтому энтропию Шеннона $\mathbf{H}\{\xi\}$ используют в криптологии как количественную меру неопределенности сообщения $\xi \in A$. Однако является ли (1.31) единственным функционалом, обладающим изученными свойствами?

Ответ на этот вопрос дает аксиоматический подход, суть которого состоит в следующем. Задается система аксиом минимального размера, и на их основе доказывается единственность функционала энтропии (1.31). Приведем три известные системы аксиом (**A**).

Система аксиом Хинчина

A1. $H_m(p_1, \ldots, p_m)$ — ненулевая непрерывная по $p_1, \ldots, p_m$ в симплексе S_{m-1} функция.

A2. $H_m(p_1, \ldots, p_m)$ — симметричная функция.

A3. $H_{m+1}(p_1, \ldots, p_m, 0) = H_m(p_1, \ldots, p_m)$.

A4. $H_{mn}((q_{11}, \ldots, q_{1m}), \ldots, (q_{n1}, \ldots, q_{nm})) = H_n(p_1, \ldots, p_n) \sum_{i=1}^{n} p_i \times$

$\times H_m\left(\frac{q_{i1}}{p_i}, \ldots, \frac{q_{im}}{p_i}\right)$, если $p_i = q_{i1} + \ldots + q_{im}$ $(i = 1, \ldots, n)$.

A5. $H_m(p_1, \ldots, p_m) \leqslant H(1/m, \ldots, 1/m)$.

Система аксиом Фаддеева

A1′. $H_2(p, 1-p)$ непрерывна при $0 \leqslant p \leqslant 1$ и положительна хотя бы при одном значении p.

A2′. $H_m(p_1, \ldots, p_m)$ — симметричная функция.

A3′. При $m \geqslant 2$ $H_{m+1}((p_1, \ldots, p_{m-1}), q_1, q_2) = H_m(p_1, \ldots, p_m) + p_m H_2\left(\frac{q_1}{p_m}, \frac{q_2}{p_m}\right)$, если $p_m = q_1 + q_2$.

Доказательство эквивалентности этих двух систем аксиом и единственности функционала энтропии вида (1.31) имеется в [8].

Приведем еще одну систему аксиом и обоснование функционала (1.31), следуя [33].

Система аксиом Чечета [33]

A1″(непрерывность). Функция $H_m(p_1, \ldots, p_m)$, определяющая энтропию ИДС $\langle A, (p_1, \ldots, p_m)\rangle$, непрерывна по совокупности переменных в симплексе S_{m-1} для любого $m \geqslant 2$.

A2″(монотонность). Последовательность $F(m) = H\left(\frac{1}{m}, \ldots, \frac{1}{m}\right)$, $m \geqslant 2$, не убывает.

A3″(разложение). Если $m = m_1 + \ldots + m_k$, $m_i \geqslant 2$,

$$(p_1, \ldots, p_m) \equiv \Big((q_{11}, \ldots, q_{1m_1}), \ldots, (q_{k1}, \ldots, q_{km_k})\Big), \quad \alpha_i = \sum_{j=1}^{m_i} q_{ij},$$

$$\text{то } H_m(p_1, \ldots, p_m) = H(\alpha_1, \ldots, \alpha_k) + \sum_{i=1}^{k} \alpha_i H_{m_i}\left(\frac{q_{i1}}{\alpha_i}, \ldots, \frac{q_{im_i}}{\alpha_i}\right).$$

Замечание 1.1. Аксиома **A3″** означает, что энтропия — одна и та же при двух эквивалентных способах фиксации результатов случайного эксперимента. Первый способ состоит в непосредственной фиксации того, какой символ из алфавита символов A мощностью m зарегистрирован. Во втором способе вначале регистрируется одно из k подмножеств алфавита, в которое попал случайный символ, а затем фиксируется номер символа в этом подмножестве.

Теорема 1.6 (единственность функционала энтропии). *Если неотрицательная функция $H_k(\alpha_1, \ldots, \alpha_k)$, заданная на симплексе S_{k-1} для любого $k \geqslant 2$, удовлетворяет системе аксиом $\{\mathbf{A1}'', \mathbf{A2}'', \mathbf{A3}''\}$, то либо $H_k(\cdot) \equiv 0$, либо*

$$H_k(\alpha_1, \ldots, \alpha_k) = -c \cdot \sum_{i=1}^{k} \alpha_i \log \alpha_i, \tag{1.32}$$

где $c > 0$ — положительная константа.

Доказательство. Во-первых, докажем, что при равных значениях аргумента $\alpha_1 = \ldots = \alpha_k = 1/k$ функция $H_k(\cdot)$ имеет специальный вид

$$F(k) ::= H_k(1/k, \ldots, 1/k) = c \log k. \tag{1.33}$$

Пусть $r \geqslant 2$, $n \geqslant 2$, $m = r^n$, $m_1 = \ldots = m_r = r^{n-1}$. Тогда согласно $\mathbf{A3}''$ и обозначению (1.33)

$$F\left(r^n\right) = H_{r^n}\left(r^{-n}, \ldots, r^{-n}\right) = H_r(1/r, \ldots, 1/r) + \\ + \sum_{i=1}^{r} \frac{1}{r} H_{r^{n-1}}\left(\frac{1}{r^{n-1}}, \ldots, \frac{1}{r^{n-1}}\right) = F(r) + F\left(r^{n-1}\right),$$

и методом математической индукции по n получим равенство

$$F\left(r^n\right) = n \cdot F(r), \tag{1.34}$$

верное также при $n = 1$.

Допустим, что при некотором $r_0 \geqslant 2$ имеет место $F(r_0) = 0$. В этом случае из $\mathbf{A2}''$ следует $F(r) = 0$ для любого $2 \leqslant r \leqslant r_0$. С другой стороны, для любого $r > r_0$ найдется натуральное n: $r \leqslant r_0^n$, откуда в силу $\mathbf{A2}''$ и (1.34)

$$F(r) \leqslant F\left(r_0^n\right) = n \cdot F(r_0) = 0.$$

Таким образом, все члены последовательности $F(1), F(2), F(3), \ldots$ солидарны: либо все положительны, либо все равны нулю.

Установим вид функции $F(\cdot)$ в предположении, что она не равна нулю тождественно. Пусть $s \geqslant r \geqslant 2$, $u \geqslant 1$ — произвольные натуральные числа. Тогда найдется единственное натуральное число n такое, что $r^n \leqslant s^u \leqslant r^{n+1}$. Из $\mathbf{A2}''$ имеем

$$F\left(r^n\right) \leqslant F\left(s^u\right) \leqslant F\left(r^{n+1}\right),$$

поэтому согласно (1.34)

$$nF(r) \leqslant uF(s) \leqslant (n+1)F(r),$$

или после почленного деления на $uF(r)$

$$\frac{n}{u} \leqslant \frac{F(s)}{F(r)} \leqslant \frac{n}{u} + \frac{1}{u}. \tag{1.35}$$

Для логарифмической функции, рассуждая аналогично, получим

$$\frac{n}{u} \leqslant \frac{\log s}{\log r} \leqslant \frac{n}{u} + \frac{1}{u}. \tag{1.36}$$

Из неравенств (1.35), (1.36) следует

$$\left|\frac{F(s)}{F(r)} - \frac{\log s}{\log r}\right| \leqslant \frac{1}{u}, \quad u \geqslant 1.$$

Отсюда из-за произвола u можно сделать вывод:

$$\frac{F(s)}{F(r)} = \frac{\log s}{\log r}, \quad s \geqslant r \geqslant 2,$$

или

$$\frac{F(s)}{\log s} = c = \text{const}, \quad s \geqslant 2,$$

что и доказывает (1.33).

Во-вторых, используя (1.33), покажем справедливость представления (1.32) в произвольной точке $\alpha = (\alpha_1, \ldots, \alpha_k) \in S_{k-1}$. Рассмотрим вначале случай 1, когда $\alpha_1, \ldots, \alpha_k > 0$ — произвольные рациональные положительные числа. Приведем эти рациональные дроби к общему знаменателю:

$$\alpha_i = \frac{m_i}{m}, m_i \geqslant 2, i = 1, \ldots, k, \sum_{i=1}^{k} m_i = m.$$

В силу $\mathbf{A3}''$ имеем

$$F(m) = H_m\left(\frac{1}{m}, \ldots, \frac{1}{m}\right) = H_k\left(\frac{m_1}{m}, \ldots, \frac{m_k}{m}\right) + \\ + \sum_{i=1}^{k} \frac{m_i}{m} H_{m_i}\left(\frac{1}{m_i}, \ldots, \frac{1}{m_i}\right) = H_k(\alpha) + \sum_{i=1}^{k} \alpha_i F(m_i),$$

откуда найдем

$$H_k(\alpha) = F(m) - \sum_{i=1}^{k} \alpha_i F(m_i).$$

Из первой части доказательства либо $H_k(\alpha) \equiv 0$, либо согласно (1.33)

$$H_k(\alpha) = c \log m - \sum_{i=1}^{k} \alpha_i c \log m_i = -c \sum_{i=1}^{k} \alpha_i \log \alpha_i. \tag{1.37}$$

Теперь рассмотрим случай 2, когда все $\alpha_1, \ldots, \alpha_k$ — рациональные и среди них имеется $1 \leqslant s < k$ чисел, равных нулю, так что $\alpha_{\min} = \min\limits_{i:\alpha_i \neq 0} \alpha_i > 0$. Выберем произвольное натуральное $n > s/\left((k-s)\alpha_{\min}\right)$ и построим вспомогательный вектор с положительными рациональными координатами $\alpha^{(n)} = \left(\alpha_1^{(n)}, \ldots, \alpha_k^{(n)}\right)$:

$$\alpha_i^{(n)} = \begin{cases} \dfrac{1}{n}, \text{ если } \alpha_i = 0, \\ \alpha_i - \dfrac{s}{(k-s)n}, \text{ если } \alpha_i > 0. \end{cases}$$

Легко проверить, что этот вектор удовлетворяет условию нормировки и, следовательно, уже рассмотренному случаю 1. Поэтому либо $H_k\left(\alpha^{(n)}\right) \equiv 0$, либо справедливо (1.37):

$$H_k(\alpha^{(n)}) = -c \sum_{i=1}^{k} \alpha_i^{(n)} \log \alpha_i^{(n)}. \tag{1.38}$$

По построению существует

$$\lim_{n \to \infty} \alpha^{(n)} = \alpha,$$

поэтому согласно $\mathbf{A1}''$ и (1.38)

$$H_k(\alpha) = H\left(\lim_{n\to\infty} \alpha^{(n)}\right) = \lim_{n\to\infty} H_k\left(\alpha^{(n)}\right) = -c\sum_{i=1}^{k} \alpha_i \log \alpha_i,$$

что совпадает с (1.32), или $H_k(\alpha) \equiv 0$.

Для завершения доказательства теоремы осталось рассмотреть случай 3, когда среди координат вектора $\alpha \in S_{k-1}$ имеются иррациональные числа. Тогда найдется аппроксимирующая последовательность $\alpha^{(n)} \in S_{k-1}$ с рациональными координатами, и мы придем к (1.32), снова пользуясь аксиомой $\mathbf{A1}''$. □

Отметим в заключение, что в теории информации кроме энтропии Шеннона используются и другие функционалы, характеризующие неопределенность дискретной вероятностной схемы, которые также называют энтропией. Кратко представим эти функционалы [38].

Энтропия Реньи порядка $\alpha \in (0, 1)$:

$$H\left(p_1, \ldots, p_k;\, \alpha\right) = \frac{1}{1-\alpha} \log \sum_{i=1}^{k} p_i^{\alpha}.$$

При $\alpha \to 0$ энтропия по Реньи переходит в энтропию по Хартли, при $\alpha \to 1$ — в энтропию Шеннона.

Энтропия Реньи порядка α типа β:

$$H\left(p_1, \ldots, p_k;\, \alpha, \beta\right) = \frac{1}{2^{1-\beta} - 1}\left(\sum_{i=1}^{k} (p_i^{\alpha})^{(\beta-1)/(\alpha-1)} - 1\right),$$

где $\alpha > 0$, $\beta > 0$, $\alpha \neq \beta$, $\alpha \neq 1$, $\beta \neq 1$.

Энтропия Тсаллиса порядка $q > 1$:

$$H\left(p_1, \ldots, p_k;\, q\right) = \frac{1}{q-1}\left(1 - \sum_{i=1}^{q} p_i^{q}\right).$$

При $q \to 1$ энтропия по Тсаллису переходит в энтропию по Шеннону.

1.5. ИСТОЧНИКИ НЕПРЕРЫВНЫХ СООБЩЕНИЙ И ИХ ЭНТРОПИЙНЫЕ СВОЙСТВА

До сих пор предполагалось, что источник сообщений порождает дискретные символьные последовательности со значениями в дискретном алфавите. В приложениях, однако, встречаются источники непрерывных сообщений $\xi \in A$, где $A \subseteq \mathbb{R}$ — некоторое подмножество мощности континуум. Например, речевой сигнал в каждый момент времени можно рассматривать как величину звукового давления ξ в заданной точке пространства (в данном случае $A = [0, b]$, где b — некоторая максимально допустимая величина давления). Другой пример —

оптические сигналы. Построим обобщенную математическую модель и введем понятие энтропии для этих более сложных случаев.

Пусть (для фиксированного момента времени) сообщение принимает значение, описываемое случайной величиной $\xi \in A$, заданной на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbf{P})$ и имеющей распределение вероятностей (индуцированную вероятностную меру):

$$\mathbf{P}_\xi(B) ::= \mathbf{P}\{\xi(\omega) \in B\}, B \in \mathcal{B}, \tag{1.39}$$

где $\mathcal{B}$ — борелевская σ-алгебра на числовой прямой; при этом $\mathbf{P}_\xi(A) = 1$. Для обобщения понятия энтропии потребуем, чтобы на измеримом пространстве $(\mathbb{R},\mathcal{B})$ кроме $\mathbf{P}_\xi(\cdot)$ была определена еще одна вспомогательная мера $\nu = \nu(B)$, $B \in \mathcal{B}$, такая, что мера $\mathbf{P}_\xi(\cdot)$ абсолютно непрерывна относительно $\nu(\cdot)$. Напомним, что мера $\mathbf{P}_\xi(\cdot)$ называется абсолютно непрерывной относительно $\nu(\cdot)$, если $\forall B \in \mathcal{B}\ \nu(B) = 0 \Rightarrow \mathbf{P}_\xi(B) = 0$.

Согласно известной теореме Радона – Никодима из условия абсолютной непрерывности меры $\mathbf{P}_\xi(\cdot)$ относительно меры $\nu(\cdot)$ вытекает существование борелевской функции $f(x)$, $x \in A$, обозначаемой $\dfrac{d\mathbf{P}_\xi}{d\nu}(x)$ и называемой *производной Радона – Никодима*. Она определена везде в A, за исключением подмножества A_0, где $\nu(A_0) = 0$, а значит, и $\mathbf{P}_\xi(A_0) = 0$. При этом справедливо интегральное представление

$$\mathbf{P}_\xi(B) = \int\limits_B f(x)\nu(dx), B \in \mathcal{B}.$$

Энтропией случайного сообщения ξ с распределением $\mathbf{P}_\xi(\cdot)$ называется величина интеграла Лебега

$$\mathbf{H}\{\xi\} = -\int\limits_A \log \frac{d\mathbf{P}_\xi}{d\nu}(x)\mathbf{P}_\xi(dx). \tag{1.40}$$

В частности, если A — дискретное множество, а $\nu(\cdot)$ — «считающая мера» (например, при $|A| < \infty$, $\nu(B) = |B|$ — мощность множества B), то (1.40) превращается в определение Шеннона:

$$\mathbf{H}\{\xi\} = -\sum_{a \in A} (\log p_\xi(a)) p_\xi(a), \tag{1.41}$$

где $p_\xi(a) ::= \mathbf{P}\{\xi = a\}$, $a \in A$, — дискретное распределение вероятностей случайного символа.

Рассмотрим другой частный случай (1.40), когда случайное сообщение ξ имеет абсолютно непрерывное распределение вероятностей (относительно меры Лебега):

$$\mathbf{P}_\xi(B) = \int\limits_B p_\xi(x)dx, B \in \mathcal{B}, \tag{1.42}$$

с плотностью распределения $p_\xi(x) \geqslant 0$, удовлетворяющей условию нормировки

$$\int\limits_A p_\xi(x)dx = 1.$$

Учитывая (1.42), в качестве меры $\nu(\cdot)$ в (1.40) примем меру Лебега:

$$\nu(B) = \frac{\mathrm{mes}(B)}{\mathrm{mes}(A)}, B \subseteq A\,. \tag{1.43}$$

Предполагается, что $\mathrm{mes}(A) < \infty$.

Тогда из (1.42) и (1.43) следует

$$\frac{d\mathsf{P}_\xi}{d\nu}(x) = p_\xi(x)\,\mathrm{mes}(A)\,,$$

а интеграл Лебега (1.40) выражается через интеграл Римана:

$$\mathsf{H}\{\xi\} = -\log \mathrm{mes}(A) - \int\limits_A p_\xi(x)\log p_\xi(x)\,dx\,. \tag{1.44}$$

Дифференциальной (*относительной*) *энтропией* случайного сообщения $\xi \in A$ с плотностью распределения $p_\xi(x)$ называется значение функционала

$$\mathsf{H}_d\{\xi\} = -\int\limits_A p_\xi(x)\log p_\xi(x)\,dx\,. \tag{1.45}$$

Замечание 1.2. Термин «относительная» показывает, что она вычислена *относительно* меры Лебега (1.43).

Замечание 1.3. Если «доопределить» функцию плотности на всей числовой прямой

$$p_\xi(x) = 0\,,\ x \in \mathbb{R}\backslash A\,,$$

то в определении (1.45) всегда будем полагать интегрирование на $\mathbb{R} = (-\infty, \infty)$.

Замечание 1.4. Аналогично (1.45) определяется энтропия и в ситуации, когда случайный символ $\xi \in \mathbb{R}^N$ — N-мерный и описывается N-мерной плотностью распределения $p_\xi(x)$, $x \in \mathbb{R}^N$.

Отметим, что функционалы (1.44) и (1.45) отличаются на константу

$$\mathsf{H}\{\xi\} = \mathsf{H}_d\{\xi\} - \log \mathrm{mes}(A)\,.$$

Пользоваться функционалом (1.44) менее удобно, чем (1.45), поскольку при $A = \mathbb{R}$ получим $\mathrm{mes}(A) = +\infty$ и $\mathsf{H}\{\xi\} = -\infty$ для любой плотности $p_\xi(\cdot)$, в то время как дифференциальная энтропия (1.45) конечна. Поэтому свойства обобщенной энтропии будем выражать через свойства дифференциальной.

Свойство 1.6. Дифференциальная энтропия не изменяется при сдвиге распределения вероятностей, «зеркальных отражениях» и «перестановках фрагментов».

Доказательство. Проиллюстрируем схему доказательства для преобразований типа «сдвиг»:

$$\tilde{p}_\xi(x) = p_\xi(x - c)\,,$$

где $c \in \mathbb{R}$ — некоторая константа, определяющая величину сдвига. Тогда из (1.45), делая замену переменных $y = x - c$, имеем

$$\mathsf{H}_d\{\xi\} = -\int\limits_{-\infty}^{\infty} p_\xi(x-c)\log p_\xi(x-c)dx = \\ = -\int\limits_{-\infty}^{\infty} p_\xi(y)\log p_\xi(y)dy = \mathsf{H}_d\{\xi\}\,.$$

Для других преобразований доказательство проводится аналогично. □

Условной дифференциальной энтропией случайного символа $\xi \in \mathbb{R}$ относительно случайного символа $\eta \in \mathbb{R}$ называется величина двойного интеграла

$$\mathsf{H}_d\{\xi \mid \eta\} = -\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} p_{\xi,\eta}(x,y)\log p_{\xi\mid\eta}(x \mid y)\,dxdy\,, \tag{1.46}$$

где $p_{\xi,\eta}(x,y)$ — совместная плотность распределения вероятностей случайных символов (ξ,η); $p_{\xi\mid\eta}(x \mid y) = p_{\xi,\eta}(x,y)/p_\eta(y)$ — условная плотность распределения ξ при условии $\eta = y$. Это определение похоже на (1.19), применимое для ИДС.

Свойство 1.7. Справедливо свойство *иерархической аддитивности* дифференциальной энтропии для произвольной системы случайных символов $(\xi,\eta) \in \mathbb{R}^2$:

$$\mathsf{H}_d\{\xi,\eta\} = \mathsf{H}_d\{\eta\} + \mathsf{H}_d\{\xi \mid \eta\} = \mathsf{H}_d\{\xi\} + \mathsf{H}_d\{\eta \mid \xi\}\,. \tag{1.47}$$

Доказательство. Оно проводится аналогично доказательству подобного свойства в дискретном случае с применением формулы умножения плотностей и условия нормировки. □

Лемма 1.1. *Для любых плотностей распределения вероятностей $p(x)$, $q(x)$, $x \in \mathbb{R}^N$, выполняется неравенство*

$$J(p(\cdot) : q(\cdot)) ::= \int\limits_{\mathbb{R}^N} p(x)\log\frac{p(x)}{q(x)}\,dx \geqslant 0\,. \tag{1.48}$$

Доказательство. Оно проводится с использованием неравенства Йенсена по аналогии с дискретным случаем (см. теорему 1.3). □

Пусть $p_X(x)$, $x \in \mathbb{R}$ — произвольная плотность распределения вероятностей, а $a(x,y)$, $x,y \in \mathbb{R}$, — произвольная весовая функция, удовлетворяющая следующим условиям:

$$a(x,y) \geqslant 0, \int\limits_{-\infty}^{\infty} a(x,y)\,dx = \int\limits_{-\infty}^{\infty} a(x,y)\,dy = 1. \tag{1.49}$$

Тогда интегральное преобразование $p_X(\cdot) \to q(\cdot)$, задаваемое соотношением

$$q(y) = \int\limits_{-\infty}^{\infty} a(x,y) p_X(x)\,dx, y \in \mathbb{R}\,, \tag{1.50}$$

называется *преобразованием линейного сглаживания* (*усреднения*), или *линейной фильтрации*.

Замечание 1.5. Если $a(x,y) = \delta(y-x)$ — обобщенная δ-функция Дирака, то преобразование (1.50) становится тождественным: $q(\cdot) \equiv p_X(\cdot)$.

Замечание 1.6. Легко убедиться, что функция $q(y)$, определяемая (1.50), с учетом (1.49) удовлетворяет свойствам плотности распределения вероятностей.

Свойство 1.8. При линейном сглаживании плотности распределения вероятностей дифференциальная энтропия не убывает.

Доказательство. Пусть случайный символ $\xi \in \mathbb{R}$ имеет исходную плотность $p_X(\cdot)$, а $\eta \in \mathbb{R}$ — сглаженная плотность $q(\cdot)$. Вычислим разность их дифференциальных энтропий с учетом (1.50) и (1.49):

$$\Delta H ::= \mathsf{H}_d\{\eta\} - \mathsf{H}_d\{\xi\} = -\int\limits_{-\infty}^{\infty} q(y)\log q(y)\,dy +$$

$$+\int\limits_{-\infty}^{\infty} p_X(x)\log p_X(x)\,dx = \int\limits_{-\infty}^{\infty} p_X(x)\log p_X(x)\,dx \int\limits_{-\infty}^{\infty} a(x,y)\,dy -$$

$$-\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} a(x,y)p_X(x)\log q(y)\,dxdy =$$

$$= \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} a(x,y)p_X(x)\log\frac{a(x,y)p_X(x)}{a(x,y)q(y)}\,dxdy\,.$$

В силу (1.49) $p_1(x,y) = a(x,y)p_X(x) \geqslant 0, q_1(x,y) = a(x,y)q(y) \geqslant 0$ и удовлетворяют условиям нормировки:

$$\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} p_1(x,y)\,dxdy = \int\limits_{-\infty}^{\infty} p_X(x) \int\limits_{-\infty}^{\infty} a(x,y)\,dydx = 1,$$

$$\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} q_1(x,y)\,dxdy = \int\limits_{-\infty}^{\infty} q(y) \int\limits_{-\infty}^{\infty} a(x,y)\,dxdy = 1$$

и, следовательно, являются двумерными плотностями некоторых распределений вероятностей. Тогда согласно лемме 1.1 получим неравенство

$$\Delta H = \int\limits_{\mathbb{R}^2} p_1(x,y) \log \frac{p_1(x,y)}{q_1(x,y)}\, dxdy \geqslant 0\,.$$

□

Свойство 1.9. Пусть $\xi \in \mathbb{R}^N$ — случайный символ с плотностью распределения $p_\xi(x)$, $x \in \mathbb{R}^N$, имеющий дифференциальную энтропию $\mathsf{H}_d\{\xi\}$, а $y = f(x) : \mathbb{R}^N \to \mathbb{R}^N$ — взаимно однозначное непрерывно дифференцируемое функциональное преобразование. Тогда для дифференциальной энтропии случайного символа $\eta = f(\xi)$ справедливо соотношение

$$\mathsf{H}_d\{\eta\} = \mathsf{H}_d\{\xi\} + \mathsf{E}\{\log|J_f(\xi)|\}\,, \tag{1.51}$$

где $J_f(x) = \left|\dfrac{Df(x)}{Dx}\right|$ — якобиан преобразования $y = f(x)$.

Доказательство. Обозначим $x = f^{-1}(y)$ — обратное функциональное преобразование, а $J_{f^{-1}}(y) = \left|\dfrac{Df^{-1}(y)}{Dy}\right|$ — его якобиан. По правилам функционального преобразования многомерных случайных величин

$$p_\eta(y) = p_\xi\big(f^{-1}(y)\big)\Big|J_{f^{-1}}(y)\Big|. \tag{1.52}$$

В таком случае из (1.52) по формуле (1.45)

$$\mathsf{H}_d\{\eta\} = \mathsf{E}\{-\log p_\eta(\eta)\} = \mathsf{E}\big\{-\log p_\xi\big(f^{-1}(\eta)\big)\big\} - \mathsf{E}\left\{\log\left|J_{f^{-1}}(\eta)\right|\right\}.$$

Воспользовавшись свойством математического ожидания функции от случайных величин и свойством якобиана прямого $y = f(x)$ и обратного $x = f^{-1}(y)$ преобразований, получим

$$\Big|J_{f^{-1}}(y)\Big|_{y=f(x)} = |J_f(x)|^{-1},$$

$$\mathsf{H}_d\{\eta\} = \mathsf{E}\{-\log p_\xi(\xi)\} - \mathsf{E}\left\{\log|J_f(\xi)|^{-1}\right\} = \mathsf{H}_d\{\xi\} + \mathsf{E}\{\log|J_f(\xi)|\}\,,$$

что совпадает с (1.51). □

Следствие 1.6. *При взаимно однозначном функциональном преобразовании $y = f(x) : \mathbb{R}^N \to \mathbb{R}^N$ дифференциальная энтропия может возрастать, убывать и оставаться неизменной. Она неизменна тогда и только тогда, когда плотность распределения $p_\xi(\cdot)$ и якобиан преобразования $J_f(x)$ обладают специальным свойством:*

$$\mathsf{E}\{\log|J_f(\xi)|\} = \int\limits_{\mathbb{R}^N} p_\xi(x)\log|J_f(x)|dx = 0\,. \tag{1.53}$$

Следствие 1.7. *Если функциональное преобразование $f(\cdot)$ линейное:*

$$f(x) = Ax + b\,,$$

где $b \in \mathbb{R}^N$ — произвольный вектор; $A = (a_{ij})$ — произвольная невырожденная $(N \times N)$-матрица, то

$$\boldsymbol{H}_d\{\eta\} = \boldsymbol{H}_d\{\xi\} + \log|A|\,. \tag{1.54}$$

Доказательство. Согласно (1.51)

$$J_f(x) = \left|\frac{Dy}{Dx}\right| = A,$$

$$\mathbf{E}\{\log|J_f(\xi)|\} = \mathbf{E}\{\log|A|\} = \log|A|\,.$$

□

Замечание 1.7. Свойство 1.9 определяет существенное различие между энтропией ИДС и дифференциальной энтропией. Как известно, при взаимно однозначных функциональных преобразованиях энтропия ИДС неизменна. Это различие — результат определения дифференциальной энтропии *относительно* меры Лебега.

Следствие 1.8. *В условиях следствия 1.7 дифференциальная энтропия неизменна, если преобразование имеет единичный якобиан: $|A| = 1$.*

Следствие 1.9. *При ортогональном преобразовании случайного сообщения $\xi \in \mathbb{R}^N$*

$$\eta = A\xi\,,\, AA^T = I_N\,,$$

дифференциальная энтропия не изменяется:

$$\boldsymbol{H}_d\{\eta\} = \boldsymbol{H}_d\{\xi\}\,. \tag{1.55}$$

Доказательство. Для ортогонального преобразования $|A| = 1$. Поэтому согласно следствию 1.8 из (1.54) получим (1.55). □

Пример 1.1. Если случайный символ ξ равномерно распределен на отрезке $[a, b]$, т. е. $\mathcal{L}\{\xi\} = R[a, b]$, то

$$\mathbf{H}_d\{\xi\} = \log(b - a)\,.$$

Пример 1.2. Если случайный символ ξ имеет гауссовское распределение $\mathcal{L}\{\xi\} = \mathcal{N}_1\left(\mu\,, \sigma^2\right)$ со средним μ и дисперсией $\sigma^2 > 0$, то

$$\mathbf{H}_d\{\xi\} = \log\sqrt{2\pi e\sigma^2}\,.$$

Пример 1.3. Если случайное N-символьное $(N \geqslant 1)$ сообщение распределено по N-мерному гауссовскому закону распределения $\mathcal{L}\{\xi\} = \mathcal{N}_N(\mu\,, \Sigma)$ с вектором математического ожидания $\mu = (\mu_i)$ и ковариационной $(N \times N)$-матрицей $\Sigma = (\sigma_{ij})$, то

$$\mathbf{H}_d\{\xi\} = \log\sqrt{(2\pi e)^N|\Sigma|}\,.$$

В заключение рассмотрим ситуацию, когда есть стационарный источник непрерывных сообщений, порождающий случайный процесс с дискретным временем:

$$\xi_1, \xi_2, \ldots \in \mathbb{R}\,.$$

Исследуем дифференциальную энтропию отрезка этого процесса длительностью n $(n = 1, 2, \ldots)$:

$$\Xi_n = (\xi_1, \ldots, \xi_n) \in \mathbb{R}^n.$$

Предположим, что исследуемый случайный процесс — *стационарный гауссовский процесс*. Условие гауссовости означает, что для любого n распределение вероятностей случайного n-вектора Ξ_n является n-мерным гауссовским с плотностью

$$p_n(x) = (2\pi)^{-n/2}|\Sigma_n|^{-1/2}\exp\left(-\frac{1}{2}(x-a)^T\Sigma_n^{-1}(x-a)\right), x = (x_i) \in \mathbb{R}^n, \qquad (1.56)$$

где $a = (a_i) \in \mathbb{R}^n$ — n-вектор-столбец математических ожиданий $\mathbf{E}\{\xi_i\} = a_i$; $\Sigma_n = (\sigma_{ij})$ — ковариационная $(n \times n)$-матрица; $\sigma_{ij} = \mathbf{Cov}\{\xi_i, \xi_j\} = \mathbf{E}\{(\xi_i - a_i)\times$ $\times(\xi_j - a_j)\}$ — ковариация случайных величин ξ_i, ξ_j $\left(i, j = \overline{1, n}\right)$.

Согласно примеру 1.3 в случае (1.56)

$$\mathbf{H}_d\{\Xi_n\} = \log\sqrt{(2\pi e)^n|\Sigma_n|},$$

поэтому

$$\frac{\mathbf{H}_d\{\Xi_n\}}{n} = \frac{1}{2n}(n\log(2\pi e) + \log|\Sigma_n|) = \log\sqrt{2\pi e} + \frac{1}{2n}\log|\Sigma_n|. \qquad (1.57)$$

Свойство стационарности гауссовского случайного процесса проявляется в специальных свойствах ковариационной матрицы Σ_n:

а) дисперсия случайного процесса ξ_t не зависит от времени t, т. е.

$$\mathbf{D}\{\xi_t\} = \mathbf{Cov}\{\xi_t, \xi_t\} = \sigma_{tt} = \sigma_0 = \operatorname*{invar}_t;$$

б) ковариация значений случайного процесса ξ_t, $\xi_{t'}$ зависит лишь от разности моментов времени:

$$\sigma_{t,t'} = \mathbf{Cov}\{\xi_t, \xi_{t'}\} = \sigma_{|t-t'|}.$$

В силу указанных следствий стационарности исследуемого случайного процесса имеем

$$D_n = |\Sigma_n| = \begin{vmatrix} \sigma_0 & \sigma_1 & \sigma_2 & \ldots & \sigma_{n-1} \\ \sigma_1 & \sigma_0 & \sigma_1 & \ldots & \sigma_{n-2} \\ \sigma_2 & \sigma_1 & \sigma_0 & \ldots & \sigma_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{n-1} & \sigma_{n-2} & \sigma_{n-3} & \ldots & \sigma_0 \end{vmatrix}, \qquad (1.58)$$

где Σ_n — так называемая *симметричная теплицева матрица*. Математиками Д. Пойа и Г. Сеге было установлено следующее асимптотическое поведение обобщенной дисперсии (1.58) при $n \to \infty$:

$$\lim_{n\to\infty}\log(D_n)^{1/n} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\log S_\xi(\lambda)d\lambda, \qquad (1.59)$$

где

$$S_\xi(\lambda) = \sum_{\tau=-\infty}^{\infty} \sigma_\tau \cos(\lambda\tau), \lambda \in [-\pi, \pi], \tag{1.60}$$

есть *спектральная плотность* случайного процесса ξ_t, определяемая как косинус-преобразование Фурье *ковариационной функции*

$$\sigma_\tau = \sigma_{-\tau} = \mathsf{Cov}\left\{\xi_t, \xi_{t+\tau}\right\}, \tau \in \mathbb{Z}.$$

Введем *нормированную корреляционную функцию*

$$\rho_\tau = \frac{\sigma_\tau}{\sigma_0} = \mathsf{Corr}\left\{\xi_t, \xi_{t+\tau}\right\}, \tau \in \mathbb{Z}, \tag{1.61}$$

и *нормированную спектральную плотность*

$$s_\xi(\lambda) = \frac{S_\xi(\lambda)}{\sigma_0} = \sum_{\tau=-\infty}^{\infty} \rho_\tau \cos(\lambda\tau), \lambda \in [-\pi, \pi]. \tag{1.62}$$

Используя (1.59)–(1.62), найдем удельную энтропию гауссовского стационарного случайного процесса . Из (1.57) с помощью эквивалентных преобразований получим

$$\frac{\mathsf{H}_d\left\{\Xi_n\right\}}{n} = \log\sqrt{2\pi e\sigma_0} + \frac{1}{2n}\log\left|\frac{1}{\sigma_0}\Sigma_n\right|.$$

Согласно (1.58) и (1.61) элементами матрицы $\dfrac{1}{\sigma_0}\Sigma_n$ являются значения нормированной корреляционной функции:

$$d_n = \left|\frac{1}{\sigma_0}\Sigma_n\right| = \begin{vmatrix} \rho_0 & \rho_1 & \rho_2 & \dots & \rho_{n-1} \\ \rho_1 & \rho_0 & \rho_1 & \dots & \rho_{n-2} \\ \rho_2 & \rho_1 & \rho_0 & \dots & \rho_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \dots & \rho_0 \end{vmatrix}, \ \rho_0 = 1.$$

Тогда в силу (1.59) и (1.62)

$$\lim_{n\to\infty} \log(d_n)^{1/n} = \frac{1}{2\pi}\int\limits_{-\pi}^{\pi} \log s_\xi(\lambda)\, d\lambda.$$

Получим выражение энтропии источника $h = \lim\limits_{n\to\infty} \mathsf{H}_d\left\{\Xi_n\right\}/n$ (в гл. 4 подробно рассматриваются свойства этой величины, называемой удельной энтропией).

$$h = \lim_{n\to\infty}\frac{\mathsf{H}_d\left\{\Xi_n\right\}}{n} = \log\sqrt{2\pi e\sigma_0} + \frac{1}{4\pi}\int\limits_{-\pi}^{\pi}\log s_\xi(\lambda)d\lambda =$$

$$= \mathsf{H}_d\left\{\xi_1\right\} + \frac{1}{4\pi}\int\limits_{-\pi}^{\pi}\log s_\xi(\lambda)d\lambda, \tag{1.63}$$

где $\mathbf{H}_d\{\xi_1\} = \log\sqrt{2\pi e\sigma_0}$ — энтропия единичного гауссовского случайного символа с дисперсией σ_0.

В (1.63) первое слагаемое характеризует энтропию единичного (отдельного, изолированного) символа, а второе — зависимость символов в последовательности и выражается через нормированную спектральную плотность $s_\xi(\lambda)$ или, что эквивалентно, через нормированную корреляционную функцию.

Пример 1.4. ξ_t — последовательность независимых гауссовских символов. Тогда

$$\mathbf{Cov}\{\xi_t, \xi_{t'}\} = \sigma_0\delta_{t,\,t'}.$$

При этом $\rho_\tau = \delta_{\tau,\,0}$, $s_\xi(\lambda) = 1 \equiv \text{const}$, $\lambda \in [-\pi, \pi]$. Поскольку согласно (1.60) второе слагаемое — нулевое, получим установленный выше результат:

$$\mathbf{H}_d\{\Xi_n\} = n\mathbf{H}\{\xi_1\}, h = \mathbf{H}_d\{\xi_1\}.$$

Пример 1.5. Пусть имеет место марковская корреляционная зависимость: $\sigma_\tau = \sigma_0 e^{-\alpha|\tau|}$, $\alpha > 0$. Тогда

$$D_n = \sigma_0^n(1-\rho^2)^{n-1},$$

где $\rho = e^{-\alpha}$. Поэтому из (1.63) имеем

$$h = \mathbf{H}_d\{\xi_1\} + \log\sqrt{1-\rho^2} = \mathbf{H}_d\{\xi_1\} + \log\sqrt{1-e^{-2\alpha}}.$$

1.6. ОПТИМИЗАЦИЯ ФУНКЦИОНАЛА ЭНТРОПИИ НА КЛАССЕ ВЕРОЯТНОСТНЫХ РАСПРЕДЕЛЕНИЙ

Для криптологических применений важно исследовать случаи экстремальных значений функционала энтропии. Как видно из примеров 1.1–1.3 в разд. 1.5, дифференциальная энтропия изменяется от $-\infty$ до $+\infty$. Для ИДС $\xi \in A$, как было установлено в разд. 1.2, минимальное значение, равное нулю ($\mathbf{H}_{\min} = 0$), достигалось для вырожденного дискретного распределения вероятностей, а максимальное значение энтропии, равное $\mathbf{H}_{\max} = \log|A|$, — для дискретного равномерного распределения вероятностей. Как видим, имеются существенные различия дискретного и непрерывного случаев. Чтобы максимальные значения дифференциальной энтропии были конечны, будем осуществлять ее максимизацию на заданном ограниченном классе вероятностных распределений $\mathcal{P}$:

$$\mathbf{H}_d\{\xi\} = -\int_{-\infty}^{\infty} p(x)\log p(x)dx \to \max_{p(\cdot)\in\mathcal{P}}. \tag{1.64}$$

Исследуем три наиболее часто встречающихся в прикладных задачах класса абсолютно непрерывных вероятностных распределений.

Класс $\mathcal{P}_1(a,b)$:

$$\mathcal{P}_1(a,b) = \Big\{ p(x), x \in \mathbb{R}:$$

$$p(x) \geqslant 0,\ \int\limits_{-\infty}^{\infty} p(x)dx = 1;\ p(x) = 0, x \notin [a,b] \Bigg\} \tag{1.65}$$

есть семейство одномерных плотностей распределения с конечным носителем $[a,b]$, $-\infty < a < b < \infty$.

Класс $\mathcal{P}_2(a,\sigma^2)$:

$$\mathcal{P}_2(a,\sigma^2) = \Bigg\{ p(x), x \in \mathbb{R} \,:\, p(x) \geqslant 0,\ \int\limits_{-\infty}^{\infty} p(x)dx = 1,$$

$$\int\limits_{-\infty}^{\infty} xp(x)dx = a,\ \int\limits_{-\infty}^{\infty} (x-a)^2 p(x)dx \leqslant \sigma^2 \Bigg\} \tag{1.66}$$

есть семейство одномерных плотностей с конечными моментами первого и второго порядков: заданным математическим ожиданием (средним) $\mathsf{E}\{\xi\} = a$ и ограниченной дисперсией $\mathsf{D}\{\xi\} \leqslant \sigma^2$.

Класс $\mathcal{P}_3(n,\mu,\Sigma)$:

$$\mathcal{P}_3(n,\mu,\Sigma) = \Bigg\{ p(x), x \in \mathbb{R}^n \,:\, p(x) \geqslant 0,\ \int\limits_{\mathbb{R}^n} p(x)\,dx = 1,$$

$$\int\limits_{\mathbb{R}^n} xp(x)dx = \mu\,,\ \int\limits_{\mathbb{R}^n} (x-\mu)(x-\mu)^T p(x)dx = \Sigma \Bigg\} \tag{1.67}$$

есть семейство n-мерных плотностей распределения с фиксированным n-вектором математического ожидания $\mu = (\mu_i)$ и невырожденной $(n \times n)$-ковариационной матрицей $\Sigma = (\sigma_{ij})$, $|\Sigma| \neq 0$.

Теорема 1.7. *Для любого случайного символа* $\xi \in \mathbb{R}$ *с плотностью распределения* $p(\cdot) \in \mathcal{P}_1(a,b)$ *дифференциальная энтропия удовлетворяет неравенству*

$$\mathsf{H}_d\{\xi\} \leqslant \log(b-a)\,, \tag{1.68}$$

причем верхняя граница, т. е. максимум дифференциальной энтропии по классу $\mathcal{P}_1(a,b)$, *достигается в случае равномерного на* $[a,b]$ *распределения вероятностей* $R[a,b]$ *с плотностью*

$$p^*(x) = \frac{1}{b-a}\mathbf{1}_{[a,\,b]}(x)\,, x \in \mathbb{R}\,, \tag{1.69}$$

где $\mathbf{1}_A(x)$ *— индикаторная функция множества* A.

Доказательство. Будем решать экстремальную задачу (1.64), (1.65) при $\mathcal{P} = \mathcal{P}_1(a,b)$. Без учета условия неотрицательности $p(\cdot) \geqslant 0$ в (1.65) эта задача эквивалентна следующей задаче вариационного исчисления с ограничением

типа равенства

$$\begin{gathered}\mathsf{H}_d\{\xi\} = -\int\limits_a^b p(x)\log p(x)\,dx \to \max\limits_{p(\cdot)},\\ \int\limits_a^b p(x)\,dx = 1.\end{gathered} \tag{1.70}$$

Для решения задачи (1.70) применим метод неопределенных множителей Лагранжа. Для этого составим функционал Лагранжа

$$L(p(\cdot),\lambda) = \int\limits_a^b \left(-p(x)\log p(x) + \lambda p(x)\right)dx\,,$$

где λ — неопределенный множитель Лагранжа, и запишем необходимое условие максимума:

$$\begin{cases}\delta L = \int\limits_a^b (-1 - \log p(x) + \lambda)\delta p(x)\,dx = 0\,,\\ \int\limits_a^b p(x)\,dx = 1,\end{cases} \tag{1.71}$$

где $\delta p(x)$ — вариация функции $p(\cdot)$; δL — первая вариация функционала $L(\cdot)$. Поскольку $\delta p(x)$ — произвольная вариация, то система уравнений (1.71) примет эквивалентный вид:

$$\begin{cases}-1 - \log p(x) + \lambda = 0\,,\\ \int\limits_a^b p(x)\,dx = 1\,.\end{cases} \tag{1.72}$$

Решим первое уравнение (1.72):

$$p(x) = 2^{\lambda-1} = \text{const} = \mu\,, x \in [a,b]\,,$$

и, подставляя это решение во второе уравнение (1.72), получим единственное решение задачи (1.70) в виде (1.69). Заметим, что найденное решение удовлетворяет снятому ограничению $p(\cdot) \geqslant 0$.

Подставим (1.69) в целевую функцию (1.70):

$$\max_{p(\cdot)} \mathsf{H}_d\{\xi\} = \log(b-a)\,,$$

что и означает (1.68). □

Теорема 1.8. *Для любого случайного символа $\xi \in \mathbb{R}$ с плотностью распределения $p(\cdot) \in \mathcal{P}_2(a,\sigma^2)$ дифференциальная энтропия удовлетворяет неравенству*

$$\mathsf{H}_d\{\xi\} \leqslant \log\sqrt{2\pi e\sigma^2}\,, \tag{1.73}$$

причем верхняя граница в (1.73), т. е. максимум дифференциальной энтропии на классе $\mathcal{P}_2(a, \sigma^2)$, достигается в случае гауссовского (нормального) распределения вероятностей $\mathcal{N}_1\left(a, \sigma^2\right)$ с плотностью

$$p^*(x) = n_1\big(x \mid a, \sigma^2\big) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{(x-a)^2}{2\sigma^2}}, x \in \mathbb{R}\,. \tag{1.74}$$

Доказательство. Существует два способа доказательства. Первый основан на решении задачи вариационного исчисления (1.64) при $\mathcal{P} = \mathcal{P}_2(a, \sigma^2)$ методом неопределенных множителей Лагранжа подобно доказательству предыдущей теоремы. Второй — менее громоздкий и излагается ниже.

Получим сначала вспомогательное неравенство — оценку сверху для интеграла

$$-\int\limits_{-\infty}^{\infty} p(x) \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-a)^2}{2\sigma^2}}\right) dx = \log\sqrt{2\pi\sigma^2} + \frac{\log e}{2\sigma^2}\int\limits_{-\infty}^{\infty} (x-a)^2 p(x)\, dx\,.$$

В силу (1.66) последний интеграл равен дисперсии $\mathsf{D}\{\xi\} \leqslant \sigma^2$, поэтому

$$-\int\limits_{-\infty}^{\infty} p(x) \log n_1\big(x \mid a,\sigma^2\big) dx \leqslant \log\sqrt{2\pi e\sigma^2},\ p(\cdot) \in \mathcal{P}_2(a, \sigma^2).$$

Из этого неравенства имеем оценку разности левой и правой частей (1.73):

$$\mathsf{H}_d\{\xi\} - \log\sqrt{2\pi e\sigma^2} \leqslant -\int\limits_{-\infty}^{\infty} p_X(x) \log p_X(x)\, x\, dx +$$

$$+\int\limits_{-\infty}^{\infty} p_X(x) \log n_1\big(x \mid a, \sigma^2\big) dx = -\int\limits_{-\infty}^{\infty} p_X(x) \log\frac{p_X(x)}{n_1(x \mid a, \sigma^2)}\, dx =$$

$$= -J(p(\cdot) : n_1(\cdot))\,.$$

Согласно лемме 1.1 $J(p(\cdot) : n_1(\cdot)) \geqslant 0$, следовательно,

$$\mathsf{H}_d\{\xi\} - \log\sqrt{2\pi e\sigma^2} \leqslant 0\,,$$

что эквивалентно (1.73).

Как установлено в разд. 1.5, для гауссовского распределения (1.74) выполнено

$$\mathsf{H}_d\{\xi^*\} = \log\sqrt{2\pi e\sigma^2}\,,$$

т. е. (1.73) обращается в равенство. □

Теорема 1.9. *Для любой случайной n-символьной последовательности $\Xi_n = (\xi_i) \in \mathbb{R}^n$ с n-мерной плотностью распределения вероятностей $p(\cdot) \in$ $\in \mathcal{P}_3(n, \mu, \Sigma)$ дифференциальная энтропия удовлетворяет неравенству*

$$\mathsf{H}_d\{\Xi_n\} \leqslant \log\sqrt{(2\pi e)^n |\Sigma|}\,, \tag{1.75}$$

причем верхняя граница, т. е. максимум дифференциальной энтропии по классу $\mathcal{P}_3(n, \mu, \Sigma)$*, достигается в случае* n*-мерного гауссовского (нормального) распределения вероятностей* $\mathcal{N}_n(\mu, \Sigma)$ *с плотностью*

$$p^*(x) = n_n(x \,|\, \mu, \Sigma) = (2\pi)^{-n/2}|\Sigma|^{-1/2}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right). \qquad (1.76)$$

Доказательство. Воспользуемся тем же методом доказательства, что и в теореме 1.8. Сначала получим вспомогательное равенство для n-кратного интеграла, используя свойства семейства $\mathcal{P}_3(n, \mu, \Sigma)$ и формулу (1.76):

$$Q = -\int\limits_{\mathbb{R}^n} p_X(x)\log p^*(x)\,dx = \log\sqrt{(2\pi)^n|\Sigma|} +$$

$$+\frac{\log e}{2}\int\limits_{\mathbb{R}^n}(x-\mu)^T\Sigma^{-1}(x-\mu)p_X(x)\,dx =$$

$$= \log\sqrt{(2\pi)^n|\Sigma|} + \frac{\log e}{2}\mathsf{E}\left\{\mathrm{tr}\left(\Sigma^{-1}(x-\mu)(x-\mu)^T\right)\right\} =$$

$$= \log\sqrt{(2\pi)^n|\Sigma|} + \frac{\log e}{2}\,\mathrm{tr}\left(\Sigma^{-1}\mathsf{E}\left\{(x-\mu)(x-\mu)^T\right\}\right) = \log\sqrt{(2\pi e)^n|\Sigma|}\,.$$

Оценим с учетом этого равенства разность левой и правой частей (1.75):

$$\mathsf{H}_d\{\xi\} - \log\sqrt{(2\pi e)^n|\Sigma|} = \mathsf{H}_d\{\xi\} - Q = -\int\limits_{\mathbb{R}^n} p_X(X)\log\frac{p_X(x)}{p^*(x)}\,dx \leqslant 0\,.$$

Здесь использована лемма 1.1. Как установлено в разд. 1.5, для многомерного гауссовского распределения (1.76) выполнено

$$\mathsf{H}_d\{\Xi_n^*\} = \log\sqrt{(2\pi e)^n|\Sigma|}\,,$$

т. е. (1.75) обращается в равенство. □

1.7. КОЛИЧЕСТВО ИНФОРМАЦИИ ПО ШЕННОНУ И ЕГО СВОЙСТВА

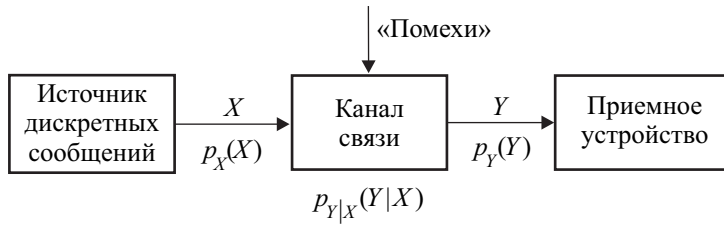
Рассмотрим простейшую схему передачи дискретной информации (см. рисунок). Пусть источник дискретных сообщений порождает случайную символьную последовательность — входной сигнал:

$$X = (x_1, \ldots, x_n) \in A^n,$$

где A — конечный «входной алфавит» символов, $|A| = N < \infty$; n — количество входных символов.

Канал связи, находящийся под воздействием «помех», преобразует входной сигнал X в выходную последовательность (выходной сигнал):

$$Y = (y_1, \ldots, y_{n'}) \in \mathcal{B}^{n'},$$



где $\mathcal{B}$ — конечный «выходной алфавит» символов, $|\mathcal{B}| = N' < \infty$; n' — количество выходных символов. Это преобразование может быть как детерминированным, так и стохастическим.

Обозначим: $p_X(X)$ — дискретное n-мерное распределение вероятностей входного сигнала; $p_Y(Y)$ — дискретное n'-мерное распределение вероятностей выходного сигнала; $p_{Y|X}(Y \mid X)$ — условное распределение вероятностей выходного сигнала Y при условии, что входной сигнал был X;

$$p_{X|Y}(X \mid Y) = \frac{p_{Y|X}(Y \mid X) p_X(X)}{p_Y(Y)} \tag{1.77}$$

есть условное распределение вероятностей входного сигнала X при условии, что выходной сигнал оказался Y.

Количеством информации по Шеннону, содержащейся в случайной символьной последовательности $Y \in \mathcal{B}^{n'}$ относительно входного сообщения $X \in A^n$, называется разность безусловной и условной энтропий

$$\mathbf{I}\{X, Y\} = \mathbf{H}\{X\} - \mathbf{H}\{X \mid Y\}, \tag{1.78}$$

где

$$\mathbf{H}\{X\} = -\sum_{X \in A^n} p_X(X) \log p_X(X) \tag{1.79}$$

есть безусловная энтропия входного сообщения;

$$\mathbf{H}\{X \mid Y\} = -\sum_{\substack{X \in A^n, \\ Y \in \mathcal{B}^{n'}}} p_{X,Y}(X, Y) \log p_{X|Y}(X \mid Y) \tag{1.80}$$

есть условная энтропия входного сообщения X относительно выходного сообщения Y.

Данное определение количества информации как приращения энтропии введено К. Шенноном в 1948 г. Исследуем свойства шенноновского количества информации.

Свойство 1.10. Справедливы следующие эквивалентные выражения количества информации через энтропию:

$$\mathbf{I}\{X, Y\} = \mathbf{H}\{X\} + \mathbf{H}\{Y\} - \mathbf{H}\{X, Y\} = \mathbf{H}\{Y\} - \mathbf{H}\{Y \mid X\}. \tag{1.81}$$

Доказательство. Воспользуемся установленным ранее свойством иерархической аддитивности энтропии составной последовательности $X||Y =$

$= (x_1, \ldots, x_n, y_1, \ldots, y_{n'}) \in A^n \times \mathcal{B}^{n'}$:

$$\mathsf{H}\{X, Y\} = \mathsf{H}\{Y\} + \mathsf{H}\{X \mid Y\} = \mathsf{H}\{X\} + \mathsf{H}\{Y \mid X\}. \tag{1.82}$$

Из (1.82) имеем $\mathsf{H}\{X \mid Y\} = \mathsf{H}\{X, Y\} - \mathsf{H}\{Y\}$. Подставляя это в (1.78), получим первое равенство в (1.81). Второе равенство в (1.81) получается подстановкой в его первую часть второго представления для $\mathsf{H}\{X, Y\}$ из (1.82). □

Свойство 1.11. Функционал шенноновского количества информации обладает свойством симметричности: $\mathsf{I}\{X, Y\} = \mathsf{I}\{Y, X\}$.

Доказательство. Согласно (1.81) имеем симметричное выражение:

$$\mathsf{I}\{X, Y\} = \mathsf{H}\{X\} + \mathsf{H}\{Y\} - \mathsf{H}\{X, Y\} = \mathsf{I}\{Y, X\}. \qquad \square$$

Замечание 1.8. Это свойство показывает, что Y содержит такое же количество информации об X, что и X об Y.

Свойство 1.12. Справедлива следующая формула для вычисления шенноновского количества информации:

$$\mathsf{I}\{X, Y\} = \sum_{\substack{X \in A^n \\ Y \in \mathcal{B}^{n'}}} p_{X,Y}(X, Y) \log \frac{p_{X,Y}(X, Y)}{p_X(X) p_Y(Y)}. \tag{1.83}$$

Доказательство. Вычислим правую часть (1.83), используя условие нормировки, свойство согласованности вероятностных распределений, а также (1.79) и (1.80):

$$\begin{aligned} B = -\mathsf{H}\{X, Y\} - \sum_{X \in A^n} \left(\sum_{Y \in \mathcal{B}^{n'}} p_{X,Y}(X, Y) \log p_X(X) \right) - \\ - \sum_{Y \in \mathcal{B}^{n'}} \left(\sum_{X \in A^n} p_{X,Y}(X, Y) \log p_Y(Y) \right) = \\ = -\mathsf{H}\{X, Y\} + \mathsf{H}\{X\} + \mathsf{H}\{Y\}, \end{aligned}$$

что согласно (1.81) совпадает с левой частью (1.83). □

Свойство 1.13. Количество информации, содержащейся в сообщении X о нем самом, равно энтропии сообщения X:

$$\mathsf{I}\{X, X\} = \mathsf{H}\{X\} = -\sum_{X \in A^n} p_X(X) \log p_X(X). \tag{1.84}$$

Доказательство. Очевидно, что условное распределение X относительно X является вырожденным: $\mathsf{P}\{X = x \mid X = y\} = \delta_{xy}$. Поэтому $\mathsf{H}\{X \mid Y\} \equiv 0$ и из (1.78) следует (1.84). □

Замечание 1.9. Соотношение (1.84) означает, что энтропия, свойства которой исследованы выше, может рассматриваться как частный случай функционала количества информации. Энтропию сообщения X поэтому иногда называют *собственной информацией* об X.

Свойство 1.14. Количество информации удовлетворяет неравенству

$$0 \leqslant \mathsf{I}\{X,Y\} \leqslant \min\{\mathsf{H}\{X\},\mathsf{H}\{Y\}\}. \tag{1.85}$$

Доказательство. По свойствам условной энтропии

$$0 \leqslant \mathsf{H}\{X \mid Y\} \leqslant \mathsf{H}\{X\},$$

поэтому из (1.78) и свойства 1.11 следует неравенство (1.85). □

Свойство 1.15. Количество информации по Шеннону $\mathsf{I}\{X,Y\}$ обращается в 0 тогда и только тогда, когда сообщения X, Y статистически независимы.

Доказательство. По свойству условной энтропии $\mathsf{H}\{X \mid Y\} = \mathsf{H}\{X\}$ тогда и только тогда, когда сообщения X, Y статистически независимы. Тогда из (1.78) получим доказываемый результат.

Другой способ доказательства основан на использовании представления (1.84) и леммы 1.1. □

Свойство 1.16. При функциональных преобразованиях сообщений $\tilde{X} = \varphi(X)$ или $\tilde{Y} = \psi(Y)$ количество информации не может возрасти:

$$\mathsf{I}\{X,Y\} \geqslant \mathsf{I}\{\varphi(X),Y\}; \tag{1.86}$$
$$\mathsf{I}\{X,Y\} \geqslant \mathsf{I}\{X,\psi(Y)\}, \tag{1.87}$$

причем равенства в (1.86) и (1.87) имеют место тогда и только тогда, когда $\varphi(\cdot)$, $\psi(\cdot)$ — биекции.

Доказательство. По определению (1.78)

$$\mathsf{I}\left\{\tilde{X},Y\right\} = \mathsf{H}\{Y\} - \mathsf{H}\left\{Y \mid \tilde{X}\right\}.$$

По свойству условной энтропии $\mathsf{H}\{Y \mid \varphi(X)\} \geqslant \mathsf{H}\{Y \mid X\}$, которое обращается в равенство лишь в случае, когда $\varphi(\cdot)$ — взаимно однозначное функциональное преобразование. Поэтому справедливо неравенство (1.86). Неравенство (1.87) доказывается аналогично. □

Свойство 1.17. Если сообщения Y_1, Y_2 независимы, то выполняется свойство аддитивности количества информации:

$$\mathsf{I}\{X,(Y_1,Y_2)\} = \mathsf{I}\{X,Y_1\} + \mathsf{I}\{X,Y_2\}. \tag{1.88}$$

Доказательство. Воспользуемся формулами (1.81), (1.78) и свойством аддитивности энтропии:

$$\begin{gathered}\mathsf{I}\{X,(Y_1,Y_2)\} = -\mathsf{H}\{X,(Y_1\,Y_2)\} + \mathsf{H}\{X\} + \mathsf{H}\{Y_1,Y_2\} = \\ = -\mathsf{H}\{X,Y_1\} - \mathsf{H}\{X,Y_2\} + 2\mathsf{H}\{X\} + \mathsf{H}\{Y_1\} + \mathsf{H}\{Y_2\} = \\ = \mathsf{I}\{X,Y_1\} + \mathsf{I}\{X,Y_2\},\end{gathered}$$

что совпадает с (1.88). □

Свойство 1.18. Справедливо неравенство

$$\mathsf{I}\{X,(Y_1,Y_2)\} \geqslant \max\{\mathsf{I}\{X,Y_1\},\mathsf{I}\{X,Y_2\}\}.$$

Доказательство. В силу (1.81) и свойств энтропии

$$\mathsf{I}\{X,(Y_1,Y_2)\} = \mathsf{H}\{X\} - \mathsf{H}\{X \mid Y_1,Y_2\} \geqslant \mathsf{I}\{X,Y_i\},\ i=1,2. \quad \square$$

Пример 1.6. Двоичное сообщение описывается однородной цепью Маркова $\xi_1, \xi_2, \ldots \in \{0,1\}$ с дискретным временем, с начальным распределением $\pi = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$ и матрицей вероятностей одношаговых переходов

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix}, \, 0 \leqslant \alpha \leqslant 1.$$

Пусть для некоторого момента времени $t = 1, 2, \ldots$ определены два соседних символа: $X ::= \xi_{t+1}$ (будущий символ = «пропущенный» символ), $Y ::= \xi_t$ (соседний наблюдаемый символ). Оценить количество информации о ξ_{t+1}, содержащееся в ξ_t.

Решение. По свойствам ОЦМ с дискретным временем имеем

$$\pi^* = \pi = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}, p_X(X) = p_Y(Y) \equiv \frac{1}{2}, \, X, Y \in \{0,1\},$$

$$p_{X,Y}(X,Y) = \frac{1}{2} \begin{cases} 1-\alpha, \text{ если } Y = X, \\ \alpha, \text{ если } Y \neq X. \end{cases}$$

Поэтому

$$\mathsf{I}\{\xi_{t+1}, \xi_t\} = \sum_{X,Y=0}^{1} \frac{1}{2}\Big((1-\alpha)\delta_{Y,X} + \alpha(1-\delta_{Y,X})\Big) \log\Big(2\big((1-\alpha)\delta_{Y,X} + \\ + \alpha(1-\delta_{Y,X})\big)\Big) = 1 + \alpha\log\alpha + (1-\alpha)\log(1-\alpha) = 1 + h(\alpha),$$

где $h(\alpha) = -\alpha\log\alpha - (1-\alpha)\log(1-\alpha)$ — функция, исследованная в примере из разд. 5.7.

Рассмотрим систему передачи информации, в которой действует источник непрерывных сообщений $X = (x_1, \ldots, x_n) \in \mathbb{R}^n$, а принимаемое сообщение $Y = (y_1, \ldots, y_{n'}) \in \mathbb{R}^{n'}$ также непрерывно.

Обозначим: $p_X(X)$ — плотность распределения вероятностей входного сообщения X; $p_Y(Y)$ — плотность распределения вероятностей выходного сообщения Y; $p_{Y|X}(Y \mid X)$ — условная плотность распределения выходного сообщения Y (при входном сообщении X);

$$p_{X|Y}(X \mid Y) = \frac{p_X(X)p_{Y|X}(Y \mid X)}{\int\limits_{\mathbb{R}^n} p_X(X')p_{Y|X}(Y \mid X')dX'} \tag{1.89}$$

есть условная плотность распределения входного сообщения X (при выходном сообщении Y); $p_{X,Y}(X,Y)$ — совместная плотность распределения;

$$\mathsf{H}_d\{X\} = -\int\limits_{\mathbb{R}^n} p_X(X)\log p_X(X)dX \tag{1.90}$$

есть безусловная дифференциальная энтропия входного случайного сообщения;

$$\mathsf{H}_d\{X \mid Y\} = -\int\limits_{\mathbb{R}^{n'}}\int\limits_{\mathbb{R}^n} p_{X,Y}(X,Y)\log p_{X|Y}(X \mid Y)dXdY \tag{1.91}$$

есть условная дифференциальная энтропия входного сигнала X относительно выходного сообщения Y.

Понятие количества информации при этом вводится аналогично дискретному случаю, рассмотренному выше.

Количеством информации по Шеннону, содержащейся в случайном выходном сигнале $Y \in \mathbb{R}^{n'}$, относительно случайного входного сообщения $X \in \mathbb{R}^n$ называется разность безусловной и условной дифференциальных энтропий (1.90) и (1.91):

$$\mathsf{I}\{X,Y\} = \mathsf{H}_d\{X\} - \mathsf{H}_d\{X \mid Y\}. \tag{1.92}$$

Свойства шенноновского количества информации (1.92) для дискретного случая сохраняют силу и для источника непрерывных сообщений. В частности, аналогично свойству 1.12 удобна следующая формула для вычислений шенноновского количества информации:

$$\mathsf{I}\{X,Y\} = \int\limits_{\mathbb{R}^{n'}}\int\limits_{\mathbb{R}^n} p_{X,Y}(X,Y)\log\frac{p_{X,Y}(X,Y)}{p_X(X)p_Y(Y)}\,dXdY \geqslant 0. \tag{1.93}$$

Пример 1.7. Пусть $X \in \mathbb{R}$, $Y \in \mathbb{R}$ — одномерные непрерывные сообщения (входной и выходной сигналы), имеющие совместное двумерное нормальное (гауссовское) распределение вероятностей:

$$\mathcal{L}\{X,Y\} = \mathcal{N}_2(\mu,\Sigma),$$

где $\mu = \begin{pmatrix}\mu_X\\ \mu_Y\end{pmatrix} \in \mathbb{R}^2$ — вектор-столбец математических ожиданий, а

$$\Sigma = \begin{pmatrix}\sigma_{XX} & \sigma_{XY}\\ \sigma_{XY} & \sigma_{YY}\end{pmatrix}$$

есть ковариационная матрица входного и выходного сигналов.

Проводя вычисления согласно (1.93), получим

$$\mathsf{I}\{X,Y\} = -\log\sqrt{1-\rho^2_{X,Y}}, \tag{1.94}$$

где $\rho_{X,Y} = \sigma_{XY}/\sqrt{\sigma_{XX}\sigma_{YY}}$ — коэффициент корреляции входного и выходного сообщений.

Рассмотрим ситуацию, когда одно из сообщений X, Y непрерывно, а другое — дискретно. Пусть, например, имеется источник непрерывных сообщений $X \in \mathbb{R}^n$ с плотностью распределения $p_X(X)$, а канал связи является цифровым, поэтому $Y \in \mathcal{B}$ — дискретное выходное сообщение с дискретным распределением вероятностей $p_Y(Y)$, где $\mathcal{B}$ — дискретное множество. Обозначим через $p_{X|Y}(X \mid Y)$ условную плотность распределения случайного входного сигнала X при условии, что выходное сообщение — Y. Аналогично (1.93) получим формулу для

вычисления шенноновского количества информации:

$$\mathsf{I}\{X,Y\} = \sum_{Y\in\mathcal{B}} p_Y(Y) \int\limits_{\mathbb{R}^n} p_{X|Y}(X \mid Y) \log \frac{p_{X|Y}(X \mid Y)}{p_X(X)}\, dX\,. \tag{1.95}$$

Свойства функционала (1.95) совпадают со свойствами, доказанными ранее.

1.8. ЗАДАНИЯ ДЛЯ ТЕСТОВ

1.1. Количество взаимной информации удовлетворяет неравенству:
а) $\min\{\mathsf{H}\{X\},\mathsf{H}\{Y\}\} \leqslant \mathsf{I}\{X,Y\} \leqslant \max\{\mathsf{H}\{X\},\mathsf{H}\{Y\}\}$;
б) $0 \leqslant \mathsf{I}\{X,Y\} \leqslant \mathsf{H}\{X\} - \mathsf{H}\{Y|X\}$;
в) $0 \leqslant \mathsf{I}\{X,Y\} \leqslant \min\{\mathsf{H}\{X\},\mathsf{H}\{Y\}\}$;
г) $0 \leqslant \mathsf{I}\{X,Y\} \leqslant \mathsf{H}\{X,Y\} - \mathsf{H}\{Y\} - \mathsf{H}\{X\}$;
д) $0 \leqslant \mathsf{I}\{X,Y\} \leqslant \min\{\mathsf{H}\{Y|X\},\mathsf{H}\{X|Y\}\}$.

1.2. Для любой последовательности случайных символов сообщения $\xi_1, \ldots, \xi_n$ энтропия обладает свойством иерархической аддитивности:
а) $\mathsf{H}\{\xi_1, \ldots, \xi_n\} = \mathsf{H}\{\xi_1\} + \mathsf{H}\{\xi_2|\xi_1\} + \mathsf{H}\{\xi_3|\xi_2\} + \ldots + \mathsf{H}\{\xi_n|\xi_{n-1}\}$;
б) $\mathsf{H}\{\xi_1, \ldots, \xi_n\} = \mathsf{H}\{\xi_1\} + \mathsf{H}\{\xi_2\} + \mathsf{H}\{\xi_3\} + \ldots + \mathsf{H}\{\xi_n\}$;
в) $\mathsf{H}\{\xi_1, \ldots, \xi_n\} = n\mathsf{H}\{\xi_1\}$;
г) $\mathsf{H}\{\xi_1, \ldots, \xi_n\} = \mathsf{H}\{\xi_1\} + \mathsf{H}\{\xi_2|\xi_1\} + \mathsf{H}\{\xi_3|\xi_2,\xi_1\} + \ldots + \mathsf{H}\{\xi_n|\xi_1, \ldots, \xi_{n-1}\}$;
д) $\mathsf{H}\{\xi_1, \ldots, \xi_n\} = \mathsf{H}\{\xi_1\} + \mathsf{H}\{\xi_1|\xi_2\} + \mathsf{H}\{\xi_2|\xi_3\} + \ldots + \mathsf{H}\{\xi_{n-1}|\xi_n\}$.

1.3. Если сообщения Y_1, Y_2 независимы, то выполняется свойство аддитивности взаимной энтропии:
а) $\mathsf{I}\{X,(Y_1,Y_2)\} = \mathsf{H}\{X,Y_1\} + \mathsf{H}\{X,Y_2\} - \mathsf{H}\{Y_1,Y_2\}$;
б) $\mathsf{I}\{X,(Y_1,Y_2)\} = \mathsf{I}\{X,Y_1\} + \mathsf{I}\{X,Y_2\}$;
в) $\mathsf{I}\{X,(Y_1,Y_2)\} = \mathsf{H}\{Y_1\} + \mathsf{H}\{Y_2\} - \mathsf{H}\{Y_1,Y_2|X\}$;
г) $\mathsf{I}\{X,(Y_1,Y_2)\} = \mathsf{I}\{X\} + \mathsf{I}\{X,Y_1\} + \mathsf{I}\{X,Y_2\}$;
д) $\mathsf{I}\{X,(Y_1,Y_2)\} = \mathsf{I}\{Y_1,Y_2\} - \mathsf{I}\{X,Y_1,Y_2\}$.

1.4. Для стационарного источника дискретных сообщений без памяти не выполняется:
а) $\mathsf{I}\{\xi_n,\xi_{n-1}\} = 0$;
б) $\mathsf{H}\{\xi_1, \ldots, \xi_{n+m}\} < \mathsf{H}\{\xi_1, \ldots, \xi_n\} + \mathsf{H}\{\xi_{n+1}, \ldots, \xi_{n+m}\}$;
в) $\mathsf{H}\{\Xi_n\} = \mathsf{H}\{\xi_1\} + \ldots + \mathsf{H}\{\xi_n\}$;
г) $\mathsf{I}\{\Xi_n,\Xi_{n+m}\} = n\mathsf{H}\{\xi_1\}$;
д) $\mathsf{H}\{\xi_n|\xi_{n-1}, \ldots, \xi_1\} = \mathsf{H}\{\xi_n\}$.

1.5. Для условной энтропии случайной величины $\xi_2 \in A$ при условии случайной величины $\xi_1 \in A$ не выполняется:
а) $0 \leqslant \mathsf{H}\{\xi_2|\xi_1\} \leqslant \mathsf{H}\{\xi_2\}$;
б) $\mathsf{H}\{\xi_2|\xi_1\} = \mathsf{I}\{\xi_2,\xi_2\} - \mathsf{I}\{\xi_1,\xi_2\}$;
в) $\mathsf{H}\{\xi_2|\xi_1\} \leqslant \mathsf{H}\{\xi_2\} - \mathsf{I}\{\xi_1,\xi_2\}$;

г) $\mathsf{H}\{\xi_2|\xi_1\} = -\sum\limits_{i,j\in A} \mathbf{P}\{\xi_1 = i, \xi_2 = j\} \log \mathbf{P}\{\xi_2 = j|\xi_1 = i\}$;

д) $\mathsf{H}\{\xi_2|\xi_1\} = \mathsf{H}\{\xi_1, \xi_2\} - \mathsf{H}\{\xi_2\}$.

1.9. РЕШЕНИЕ ТИПОВЫХ ЗАДАЧ

Задача 1.1. Пусть заданы две независимые случайные величины ξ_1, ξ_2, имеющие равномерное дискретное распределение вероятностей на множестве $\{1, 2, \ldots, 6\}$. Случайная величина $\eta = \xi_1 + \xi_2$ есть сумма двух величин ξ_1, ξ_2. Найти $\mathsf{I}\{\xi_1, \xi_1\}$, $\mathsf{I}\{\eta, \eta\}$, $\mathsf{I}\{\xi_1, \eta\}$.

Решение. Для взаимной информации $\mathsf{I}\{\xi_1, \xi_1\}$ справедливо соотношение

$$\mathsf{I}\{\xi_1, \xi_1\} = \mathsf{I}\{\xi_2, \xi_2\} = \mathsf{H}\{\xi_1\} - \mathsf{H}\{\xi_1|\xi_1\} = \mathsf{H}\{\xi_1\} = \log 6 = \log 2 + \log 3.$$

Установим распределение вероятностей случайной величины $\eta \in \{2, 3, \ldots, 12\}$:

$$p_k = \mathbf{P}\{\eta = k\} = \mathbf{P}\{\xi_1 + \xi_2 = k\} =$$

$$= \sum_{1\leqslant i,j\leqslant 6,\ i+j=k} \mathbf{P}\{\xi_1 = i\}\mathbf{P}\{\xi_2 = k - i|\xi_1 = i\} =$$

$$= \sum_{1\leqslant i,j\leqslant 6,\ i+j=k} \mathbf{P}\{\xi_1 = i\}\mathbf{P}\{\xi_2 = k - i\} = \sum_{1\leqslant i,j\leqslant 6,\ i+j=k} \frac{1}{36}.$$

Построим табл. 1.2, определяющую $\eta = \xi_1 + \xi_2$.

Таблица 1.2

ξ_2 \ ξ_1	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Следовательно, распределение вероятностей случайной величины η имеет вид

$$p = \frac{1}{36}\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 5 & 4 & 3 & 2 & 1 \end{pmatrix}.$$

Для взаимной информации $\mathsf{I}\{\eta,\eta\}$ получим соотношение

$$\mathsf{I}\{\eta,\eta\} = \mathsf{H}\{\eta\} = -2\left(\frac{1}{36}\log\frac{1}{36} + \frac{2}{36}\log\frac{2}{36} + \frac{3}{36}\log\frac{3}{36} + \frac{4}{36}\log\frac{4}{36} + \right.$$
$$\left. + \frac{5}{36}\log\frac{5}{36}\right) - \frac{6}{36}\log\frac{6}{36} = -\frac{4}{36}\log 2 - \frac{6}{36}\log 3 - \frac{16}{36}\log 2 - \frac{10}{36}\log 5 -$$
$$-\frac{6}{36}\log 6 + \frac{4}{36}(1+2+3+4+5)\log 6 + \frac{12}{36}\log 6 = -\frac{20}{36}\log 2 - \frac{6}{36}\log 3 -$$
$$-\frac{10}{36}\log 5 + \frac{66}{36}\log 2 + \frac{66}{36}\log 3 = \frac{1}{36}(46\log 2 + 60\log 3 - 10\log 5).$$

Совместное распределение вероятностей случайных величин η, ξ_1 запишем как

$$\mathbf{P}\{\eta = i, \xi_1 = j\} = \mathbf{P}\{\xi_1 = j\}\mathbf{P}\{\xi_2 = i - j | \xi_1 = j\} = \begin{cases} 1/36,\ 1 \leqslant i - j \leqslant 6, \\ 0,\ \text{иначе.} \end{cases}$$

Количество взаимной информации $\mathsf{I}\{\eta, \xi_1\}$ имеет вид

$$\mathsf{I}\{\eta, \xi_1\} = \mathsf{H}\{\xi_1\} + \mathsf{H}\{\eta\} - \mathsf{H}\{\eta,\xi_1\} = \log 6 + \frac{1}{36}(46\log 2 + 60\log 3 -$$
$$-10\log 5) - \left(-36 * \frac{1}{36}\log\frac{1}{36}\right) = \frac{1}{36}(46\log 2 + 60\log 3 - 10\log 5) -$$
$$-\log 2 - \log 3 = \frac{1}{36}(10\log 2 + 24\log 3 - 10\log 5).$$

Задача 1.2. Пусть независимые случайные величины ξ_1, ξ_2 имеют бернуллиевское распределение вероятностей с параметром $p = 1/2$. Случайная величина η порождается функциональным преобразованием $\eta = (\xi_1 + 1)^2 - \xi_2$. Найти $\mathsf{H}\{\xi_1\}$, $\mathsf{H}\{\eta\}$, $\mathsf{H}\{\xi_1|\eta\}$, $\mathsf{H}\{\eta|\xi_1\}$.

Решение. По определению $\mathsf{H}\{\{\}\xi_1\} = \mathsf{H}\{\{\}\xi_2\} = \log 2$. Значения случайной величины η в зависимости от вектора (ξ_1, ξ_2) заданы в табл. 1.3.

Таблица 1.3

(ξ_1, ξ_2)	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
η	1	0	4	3

Распределение вероятностей случайной величины $\eta \in \{0,1,3,4\}$ имеет вид

$$p = \frac{1}{4}\begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}.$$

Следовательно, $\mathsf{H}\{\eta\} = \log 4 = 2\log 2$.

Найдем совместное распределение вероятностей случайных величин ξ_1, η:

$$\mathbf{P}\{\xi_1 = i, \eta = j\} =$$
$$= \mathbf{P}\{\xi_1 = i\}\mathbf{P}\{\eta = j|\xi_1 = i\} = \frac{1}{2}\mathbf{P}\{\xi_2 = (i+1)^2 - j|\xi_1 = i\} =$$
$$= \frac{1}{2}\mathbf{P}\{\xi_2 = (i+1)^2 - j\} = \begin{cases} 1/4,\ (i+1)^2 - j \in \{0{,}1\}, \\ 0,\ (i+1)^2 - j \notin \{0{,}1\}. \end{cases}$$

Энтропия $\mathsf{H}\{\xi_1, \eta\}$ случайных величин ξ_1, η равна $2\log 2$. В результате для условных энтропий $\mathsf{H}\{\xi_1|\eta\}$, $\mathsf{H}\{\eta|\xi_1\}$ справедливы равенства

$$\mathsf{H}\{\xi_1|\eta\} = \mathsf{H}\{\xi_1, \eta\} - \mathsf{H}\{\eta\} = 2\log 2 - 2\log 2 = 0,$$
$$\mathsf{H}\{\eta|\xi_1\} = \mathsf{H}\{\xi_1, \eta\} - \mathsf{H}\{\xi_1\} = 2\log 2 - \log 2 = \log 2.$$

1.10. ЗАДАЧИ И УПРАЖНЕНИЯ

1.1. $\xi \in \{0, 1\}$ — двоичный случайный символ с распределением вероятностей Бернулли: $\mathbf{P}\{\xi = 1\} = 1 - \mathbf{P}\{\xi = 0\} = p$. Вычислить энтропию $\mathsf{H}\{\xi\}$, построить график зависимости энтропии $\mathsf{H}\{\xi\}$ от элементарной вероятности $p \in [0, 1]$ и исследовать эту функцию на экстремум.

1.2. Пусть ξ_1,$\xi_2 \in \{0,1\}$ имеют дискретное равномерное распределение вероятностей, т. е. $\mathbf{P}\{\xi_i = 1\} = 1 - \mathbf{P}\{\xi_1 = 0\} = \dfrac{1}{2}$, $i \in \{0, 1\}$, а случайная величина η есть функциональное преобразование величин ξ_1, ξ_2: $\eta = \xi_1 + \xi_2$, $\eta \in \{0, 1, 2\}$. Случайные величины ξ_1, ξ_2 независимы. Найти $\mathsf{H}\{\xi_1\}$, $\mathsf{H}\{\eta\}$, $\mathsf{H}\{\xi_1|\eta\}$, $\mathsf{H}\{\eta|\xi_1\}$.

1.3. ξ_1 имеет дискретное равномерное распределение вероятностей на множестве $\{1, 2, \ldots, 6\}$, а случайная величина ξ_2 принимает значение 0, если ξ_1 четное, и 1, если нечетное. Определить $\mathsf{H}\{\xi_2\}$, $\mathsf{H}\{\xi_2|\xi_1\}$, $\mathsf{H}\{\xi_1|\xi_2\}$.

1.4. Пусть $\xi \in \{0, 1, \ldots, N-1\}$ — случайный символ, причем элементарная вероятность $p_0 = \mathbf{P}\{\xi = 0\} = \varepsilon$ фиксирована. Найти максимум энтропии $\mathsf{H}\{\xi\}$ при произвольных $p_i = \mathbf{P}\{\xi = i\}$, $i \in \{1, \ldots, N-1\}$.

1.5. Дискретная случайная величина ξ задана распределением вероятностей $\mathbf{P}\{\xi = i\} = \dfrac{1}{2^i}$, $i = 1, 2, \ldots$. Вычислить энтропию $\mathsf{H}\{\xi\}$.

1.6. Случайные символы ξ_1, ξ_2 в сообщении зависимы. Известно, что $\mathsf{H}\{\xi_1\} = 8$ битов, $\mathsf{H}\{\xi_2\} = 12$ битов. Какие значения может принимать условная энтропия $\mathsf{H}\{\xi_2|\xi_1\}$, если $\mathsf{H}\{\xi_1|\xi_2\}$ изменяется в максимально возможных пределах?

1.7. Дискретные случайные величины $\xi_1, \xi_2 \in \{0, 1\}$ зависимы и их совместное распределение вероятностей имеет вид

$$p(\xi_1,\xi_2) = \begin{pmatrix} 1/3 & 1/6 \\ 1/6 & 1/3 \end{pmatrix}.$$

Определить $\mathsf{H}\{\xi_1\}$, $\mathsf{H}\{\xi_2\}$, $\mathsf{H}\{\xi_1|\xi_2\}$, $\mathsf{H}\{\xi_2|\xi_1\}$.

1.8. Доказать, что количество взаимной информации $\mathsf{I}\{x, y\} = \mathsf{H}\{x\} - \mathsf{H}\{x|y\}$ обладает свойством симметричности. Найти $\mathsf{I}\{x, x\}$.

1.9. Дискретная случайная величина ξ_1 имеет равномерное распределение вероятностей на множестве $\{-1, 0, 1\}$, а дискретная случайная величина ξ_2 — на

множестве $\{0,1,2\}$. Случайные величины ξ_1, ξ_2 независимы. Вычислить $\mathsf{H}\{\eta\}$, где $\eta = \xi_1^2 + \xi_2$ есть функциональное преобразование ξ_1, ξ_2. Найти $\mathsf{H}\{\eta\}$, $\mathsf{I}\{\xi_1, \eta\}$, $\mathsf{I}\{\xi_2, \eta\}$.

1.10. Случайная величина $\xi_1 \in \{0,1\}$ имеет равномерное дискретное распределение вероятностей. Случайная величина $\xi_2 \in \{0,1\}$ задается условным распределением вероятностей

$$P = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{pmatrix}, \; p_{ij} = \mathbf{P}\{\xi_2 = j | \xi_1 = i\}, \; i,j \in \{0,1\}.$$

Установить, в каких случаях безусловная энтропия $\mathsf{H}\{\xi_2\}$ достигает максимального значения.

1.11. Для трех дискретных случайных символов ξ_1, ξ_2, ξ_3 энтропии одинаковы: $\mathsf{H}\{\xi_1\} = \mathsf{H}\{\xi_2\} = \mathsf{H}\{\xi_3\} = h$. Вычислить количество взаимной информации $\mathsf{I}\{\xi_1, \xi_2, \xi_3\}$, если:

1) $\mathsf{H}\{\xi_1, \xi_2, \xi_3\} = 3h$;
2) $\mathsf{H}\{\xi_1, \xi_2, \xi_3\} = h$.

1.12. Имеется случайная последовательность двоичных символов $\xi_1, \ldots, \xi_{2n} \in \{-1, 1\}$. Известно, что возможно появление только таких реализаций, для которых $\xi_1 + \ldots + \xi_{2n} = 0$, причем все они равновероятны. Определить $\mathsf{H}\{\xi_1\}$, $\mathsf{I}\{\xi_1, \xi_2\}$, $\mathsf{I}\{(\xi_1, \ldots, \xi_{2n-1}), \xi_{2n}\}$.

1.13. Случайная величина $\xi_1 \in \{0,1\}$ имеет распределение вероятностей $\pi_1 = \mathbf{P}\{\xi_1 = 1\} = 1 - \mathbf{P}\{\xi_1 = 0\}$, а случайная величина $\xi_2 \in \{0,1\}$ задается условным распределением вероятностей

$$P = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{pmatrix}, \; p_{ij} = \mathbf{P}\{\xi_2 = j | \xi_1 = i\}, \; i,j \in \{0,1\}.$$

Найти $\mathsf{H}\{\xi_2 | \xi_1\}$. Использовать обозначения: $h(p) = -p \log p - (1-p) \times \times \log(1-p)$ — энтропия бернуллиевского закона распределения вероятностей с параметром $p \in [0,1]$.

1.14. ξ — случайная величина с плотностью распределения вероятностей $f(x)$, а $\eta = a\xi + b$, где a, b — фиксированные параметры, причем $a \neq 0$. Вычислить $\mathsf{H}_d\{\eta\}$.

1.15. Случайный символ ξ на $[a,b]$ имеет равномерное распределение вероятностей $\mathcal{L}\{\xi\} = \mathcal{R}[a,b]$. Определить дифференциальную энтропию $\mathsf{H}_d\{\xi\}$.

1.16. Случайный символ ξ имеет гауссовское распределение вероятностей $\mathcal{L}\{\xi\} = \mathcal{N}_1(\mu, \sigma^2)$. Найти дифференциальную энтропию $\mathsf{H}_d\{\xi\}$.

1.17. Пусть $\xi \geqslant 0$ — случайная величина с экспоненциальным распределением вероятностей $\mathcal{L}\{\xi\} = E(\lambda)$, плотность которого равна $p_\xi(x) = \lambda e^{-\lambda x}$, $x \geqslant 0$. Вычислить дифференциальную энтропию $\mathsf{H}_d\{\xi\}$.

1.18. $\xi \in \{0, 1, 2, \ldots\}$ — случайная величина с геометрическим распределением вероятностей $\mathcal{L}\{\xi\} = G(p)$: $p_i = \mathbf{P}\{\xi = i\} = (1-p)^i p$. Определить энтропию $\mathsf{H}\{\xi\}$ и исследовать на экстремум ее зависимость от параметра $p \in [0, 1]$.

1.19. Пусть $\xi \in \{0, 1, \ldots, N-1\}$ — случайная величина с биномиальным распределением вероятностей $\mathcal{L}\{\xi\} = Bi(N-1, p)$: $p_i = \mathbf{P}\{\xi = i\} = C_{N-1}^i \times$ $\times (1-p)^{N-1-i} p^i$. Найти энтропию $\mathsf{H}\{\xi\}$.

1.20. Даны значения $\mathsf{H}\{x\}$ и $\mathsf{H}\{y\}$. В каких пределах может меняться $\mathsf{I}\{x,y\}$ при изменении $\mathsf{H}\{x,y\}$ от минимального до максимально возможного значений?

1.21. Вычислить дифференциальную энтропию случайной величины, заданной функцией распределения вероятностей

$$F(x) = \begin{cases} 0, \, x \leqslant 0, \\ x^2, \, 0 \leqslant x \leqslant 1, \\ 1, \, 1 < x. \end{cases}$$

1.22. Совместное распределение вероятностей случайных величин x, y задано матрицей

$$p(x,y) = \frac{1}{8} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Определить энтропии $\mathsf{H}\{x\}$, $\mathsf{H}\{y\}$, $\mathsf{H}\{x|y\}$, $\mathsf{H}\{y|x\}$, $\mathsf{H}\{x, y\}$.

1.23. Имеются два дискретных троичных источника с независимыми элементами. На выходе каждого источника появляются сообщения одинаковой длины — по 15 элементов. Количество различных элементов в сообщении каждого источника постоянно. Сообщения каждого источника отличаются только порядком элементов. Известны два типичных сообщения: 021202120212021 — первого источника и 012101201101201 — второго. Элемент какого источника несет в среднем большее количество информации?

1.24. Эксперимент x состоит в случайном выборе целого числа от 1 до 1050. Эксперимент y — определение остатков от деления этого числа на 5 и 7. Найти энтропии $\mathsf{H}\{x\}$, $\mathsf{H}\{y\}$, $\mathsf{H}\{x|y\}$.

1.25. Определить количество информации $\mathsf{I}\{x, y\}$, если совместное распределение вероятностей случайных величин x, y имеет вид

$$p(x,y) = \frac{1}{9} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Глава 2

ДИСКРЕТНЫЕ ИСТОЧНИКИ СООБЩЕНИЙ И ИХ ВЕРОЯТНОСТНЫЕ МОДЕЛИ

2.1. ДИСКРЕТНЫЕ ВРЕМЕННЫЕ РЯДЫ, ИХ МОДЕЛИ И ВЕРОЯТНОСТНЫЕ ХАРАКТЕРИСТИКИ

При функционировании любой криптографической системы в ее узлах и на выходе формируются хаотические последовательности символов из некоторого дискретного множества. Удобной математической моделью для их описания является *дискретная случайная последовательность*, или *дискретный временной ряд* (ДВР).

Определение 2.1. *Дискретный временной ряд x_t есть упорядоченная по индексу $t \in \mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ последовательность дискретных случайных величин*

$$x_t = x(\omega,t) \in A, \omega \in \Omega, t \in \mathbb{Z}, \tag{2.1}$$

определенных на одном и том же вероятностном пространстве $(\Omega, \mathcal{F}, \mathsf{P})$ и принимающих значения из некоторого дискретного множества A, где Ω — пространство элементарных событий; $\mathcal{F}$ — σ-алгебра подмножеств из Ω, называемых случайными событиями; $\mathsf{P}(\cdot)$ — вероятностная мера, определенная на $\mathcal{F}$.

В (2.1) индекс $t \in \mathbb{Z}$ интерпретируется как дискретное время, а A называется *пространством состояний* (*алфавитом*) временного ряда. В существующих криптографических системах алфавит A конечен, поэтому далее без потери общности условимся полагать

$$A = \{0, 1, \ldots, N-1\},$$

где $2 \leqslant N < +\infty$ — мощность алфавита A, т. е. число различных возможных значений дискретного временного ряда x_t. Если $N = 2$, то $x_t \in A = \{0, 1\}$ принято называть *двоичным* (*бинарным*) *временным рядом.*

Определение 2.2. *Отсчетом ДВР в некоторый фиксированный момент времени $t = s \in \mathbb{Z}$ называется дискретная случайная величина $x_s \in A$. Реализация (траектория) ДВР — это упорядоченная совокупность всех отсчетов при фиксированном элементарном событии (исходе случайного эксперимента) $\omega \in \Omega$:*

$$X_\omega = \{x(\omega, t) : t \in \mathbb{Z}\}.$$

Совокупность всех реализаций $X = \{X_\omega : \omega \in \Omega\}$ называется ансамблем реализаций ДВР.

Следуя аксиоматике теории вероятностей [35], построим вероятностное пространство $(\Omega, \mathcal{F}, \mathsf{P})$, на котором в дальнейшем и будем рассматривать дискретные временные ряды и порожденные ими случайные события. В качестве пространства элементарных событий Ω примем пространство всевозможных бесконечных последовательностей символов из алфавита A:

$$\Omega = A^\infty;$$

элементарное событие

$$\omega = (\ldots, \omega_{-1}, \omega_0, \omega_1, \ldots) \in \Omega -$$

это бесконечная последовательность символов $\omega_t \in A$, $t \in \mathbb{Z}$.

Определение 2.3. *Пусть m — произвольное натуральное число, $t_1 < t_2 < \ldots < t_m$ — произвольный набор m упорядоченных моментов времени $t_1, \ldots, t_m \in \mathbb{Z}$, $a_1, \ldots, a_m \in A$ — произвольный набор m символов из алфавита A. Цилиндрическим множеством (цилиндром) называется множество (случайное событие)*

$$C_{t_1,\ldots,t_m}(a_1, \ldots, a_m) = \left\{\omega \in \Omega : \omega_{t_i} = a_i, i = 1, \ldots, m\right\} \subseteq \Omega. \qquad (2.2)$$

Построим σ-алгебру $\mathcal{F}$ как бесконечную систему множеств, полученных применением к системе всех цилиндрических множеств C_* операций разности, счетного объединения и пересечения. В результате имеем измеримое пространство $(\Omega, \mathcal{F})$. Для построения вероятностной меры $\mathsf{P}(\cdot)$ на этом измеримом пространстве $(\Omega, \mathcal{F})$ воспользуемся теоремой Колмогорова о продолжении вероятностной меры [35] (изученной ранее в курсе теории вероятностей и математической статистики). Для этого достаточно задать вероятностную меру на множестве $C_* \in \mathcal{F}$, т. е. на всевозможных цилиндрических множествах вида (2.2):

$$\mathsf{P}\left(C_{t_1,\ldots,t_m}(a_1, \ldots, a_m)\right) = \mathsf{P}\left\{\omega_{t_1} = a_1, \ldots, \omega_{t_m} = a_m\right\}.$$

Например, для ДИБП с алфавитом $A = \{0, 1\}$ имеем

$$\mathsf{P}\left(C_{t_1,\ldots,t_m}(a_1, \ldots, a_m)\right) = \prod_{i=1}^{m} p^{a_i}(1-p)^{1-a_i} = p^{\sum\limits_{i=1}^{m} a_i}(1-p)^{m-\sum\limits_{i=1}^{m} a_i},$$

где $p \in [0, 1]$ — параметр модели, обычно называемый вероятностью успеха, т. е. события $\left\{\omega_{t_i} = 1\right\}$.

Построение вероятностных моделей простого марковского источника и марковского источника с глубиной памяти $s \geqslant 1$ приведено в гл. 5 данного учебного пособия.

Вероятностные модели ДВР задаются на основе их вероятностных характеристик. Определим основные вероятностные характеристики ДВР.

Определение 2.4. *Пусть зафиксированы произвольное натуральное число $n \in \mathbb{N}$ и упорядоченные моменты времени $t_1 < t_2 < \ldots < t_n$ $(t_1, \ldots, t_n \in \mathbb{Z})$. Совместное дискретное распределение вероятностей отсчетов $x_{t_1}, \ldots, x_{t_n} \in A$:*

$$P_n(a_1, \ldots, a_n; t_1, \ldots, t_n) = \mathsf{P}\left\{x_{t_1} = a_1, \ldots, x_{t_n} = a_n\right\}, a_1, \ldots, a_n \in A, \quad (2.3)$$

называется n-мерным распределением вероятностей временного ряда x_t.

Функция (2.3) обладает следующими свойствами:

— область значений $P_n(\cdot)$ есть $[0, 1]$;

— *свойство нормировки* —

$$\sum_{a_1, \ldots, a_n \in A} P_n(a_1, \ldots, a_n; t_1, \ldots, t_n) \equiv 1;$$

— *свойство согласованности* $(1 \leqslant k < n)$ —

$$\sum_{a_{k+1}, \ldots, a_n \in A} P_n(a_1, \ldots, a_n; t_1, \ldots, t_n) \equiv P_k(a_1, \ldots, a_k; t_1, \ldots, t_k).$$

Заметим, что в последнем соотношении k-мерное распределение вероятностей $P_k(\cdot)$ называется *маргинальным распределением вероятностей* по отношению к исходному $P_n(\cdot)$.

Известно [35], что семейство всевозможных конечномерных распределений вероятностей (2.3) однозначно задает *вероятностную модель ДВР*. Отметим еще, что в силу конечности N ДВР x_t имеет ограниченные моменты любого порядка $s > 0 : \mathsf{E}\{x_t^s\} < +\infty$.

Определение 2.5. *Математическим ожиданием (средним) и дисперсией ДВР* x_t *называются функции времени*

$$m_t = \mathsf{E}\{x_t\} = \sum_{a \in A} a P_1(a; t), t \in \mathbb{Z},$$

$$d_t = \mathsf{D}\{x_t\} = \mathsf{E}\left\{(x_t - m_t)^2\right\} = \sum_{a \in A} (a - m_t)^2 P_1(a; t) \geqslant 0, t \in \mathbb{Z}, \quad (2.4)$$

соответственно.

Определение 2.6. *Ковариационная и корреляционная функция ДВР* x_t *являются функциями двух переменных:*

$$\sigma(t_1,t_2) = \mathsf{Cov}\left\{x_{t_1}, x_{t_2}\right\} = \mathsf{E}\left\{(x_{t_1} - m_{t_1})(x_{t_2} - m_{t_2})\right\} =$$
$$= \sum_{a_1, a_2 \in A} (a_1 - m_{t_1})(a_2 - m_{t_2}) P_2(a_1, a_2; t_1, t_2), t_1, t_2 \in \mathbb{Z},$$

$$\rho(t_1, t_2) = \mathsf{Corr}\left\{x_{t_1}, x_{t_2}\right\} = \frac{\sigma(t_1, t_2)}{\sqrt{d_{t_1} d_{t_2}}} \in [-1, +1], t_1, t_2 \in \mathbb{Z}, \quad (2.5)$$

соответственно.

Отметим, что $\sigma(t, t) = d_t$, $t \in \mathbb{Z}$.

Определение 2.7. *Дискретный временной ряд* x_t *называется стационарным в узком смысле (сильно стационарным), если любое его n-мерное распределение вероятностей (2.3) инвариантно относительно сдвига времени, т. е.*

для любого $n \in \mathbb{N}$, *любых* $t_1, \ldots, t_n \in \mathbb{Z}$ ($t_1 < t_2 < \ldots < t_n$) *и любого сдвига времени* $\tau \in \mathbb{Z}$ *выполняется соотношение*

$$P_n(a_1, \ldots, a_n;\, t_1 + \tau, \ldots, t_n + \tau) =$$

$$= P_n(a_1, \ldots, a_n;\, t_1, \ldots, t_n), a_1, \ldots, a_n \in A. \tag{2.6}$$

Соотношение (2.6) означает, что для стационарного в узком смысле дискретного временного ряда n-мерное распределение вероятностей $P_n(\cdot;\, t_1, \ldots, t_n)$ зависит лишь от взаиморасположения моментов времени $t_1, \ldots, t_n$ (т. е. от разностей $t_2 - t_1, t_3 - t_2, \ldots, t_n - t_{n-1}$). В частности, из (2.6) следует, что одномерное распределение (при $n = 1$) вообще не зависит от времени:

$$P_1(a;\, t) \equiv P_1(a), \tag{2.7}$$

а двумерное распределение (при $n = 2$) зависит лишь от $t_2 - t_1$:

$$P_2(a_1, a_2;\, t_1, t_2) \equiv P_2(a_1, a_2;\, t_2 - t_1). \tag{2.8}$$

Определение 2.8. *Дискретный временной ряд* x_t *называется стационарным в широком смысле (слабо стационарным), если его математическое ожидание не зависит от времени* t:

$$m_t \equiv m, m \in \mathbb{R}, \tag{2.9}$$

а его ковариационная функция зависит лишь от разности моментов времени $t_2 - t_1 \in \mathbb{Z}$:

$$\sigma(t_1, t_2) \equiv \sigma(t_2 - t_1), \tag{2.10}$$

где $\sigma(u)$, $u \in \mathbb{Z}$, — *некоторая четная неотрицательно определенная функция.*

Теорема 2.1. *Если* N *конечно, то всякий стационарный в узком смысле ДВР* x_t *является одновременно и стационарным в широком смысле дискретным временным рядом.*

Доказательство. Как отмечалось выше, в силу конечности N для ДВР x_t существуют ограниченные моменты любого порядка. Следовательно, существуют моменты первого и второго порядка (2.4), (2.5). Поэтому из (2.7) вытекает (2.9), а из (2.8) следует (2.10). □

Отметим, что, как видно из теоремы 2.1, из стационарности в узком смысле следует стационарность в широком смысле; обратное, вообще говоря, неверно.

Как известно [30], ковариационная функция $\sigma(u)$ стационарного в широком смысле ДВР x_t характеризует уровень линейной стохастической зависимости отсчетов x_t и x_{t+u}, отстоящих на u единиц времени. На практике временные ряды имеют «затухающую память» [13]:

$$P_2(a_1, a_2;\, t_1, t_2) - P_1(a_1;\, t_1)P_1(a_2;\, t_2) \to 0,\ a_1, a_2 \in A,$$

при $|t_2 - t_1| \to +\infty$, влекущую в силу (2.5), (2.8) убывание ковариационной функции стационарного ДВР:

$$\sigma(u) \to 0, |u| \to +\infty. \tag{2.11}$$

Определение 2.9. *Пусть (2.11) выполняется в усиленном смысле, так что сходится ряд*

$$\sum_{u=-\infty}^{+\infty} |\sigma(u)| < +\infty. \tag{2.12}$$

Спектральной плотностью $S(\lambda)$, $\lambda \in [-\pi, \pi]$, стационарного в широком смысле ДВР x_t, имеющего ковариационную функцию $\sigma(u)$, $u \in \mathbb{Z}$, называется *дискретное преобразование Фурье* (ДПФ) от ковариационной функции

$$S(\lambda) = \frac{1}{2\pi} \sum_{u=-\infty}^{+\infty} \sigma(u)\, e^{i\lambda u} \equiv \frac{1}{2\pi} \sum_{u=-\infty}^{+\infty} \sigma(u) \cos(\lambda u), \tag{2.13}$$

где i — мнимая единица.

Отметим некоторые свойства спектральной плотности $S(\cdot)$:

— неотрицательность: $S(\lambda) \geqslant 0$, $\lambda \in [-\pi, \pi]$;

— четность $S(-\lambda) = S(\lambda)$, $\lambda \in [-\pi, \pi]$;

— взаимнооднозначная связанность $S(\cdot)$ и $\sigma(\cdot)$ парой ДПФ (2.13), и

$$\sigma(u) = \int_{-\pi}^{\pi} S(\lambda) e^{i\lambda u} d\lambda \equiv \int_{-\pi}^{\pi} S(\lambda) \cos(\lambda\tau) d\lambda, u \in \mathbb{Z}.$$

Как уже отмечалось, вероятностная модель ДВР полностью определяется заданием системы конечномерных распределений вероятностей:

$$\mathcal{P} = \{P_n(a_1, \ldots, a_n; t_1, \ldots, t_n) : a_1, \ldots, a_n \in A; t_1, \ldots, t_n \in \mathbb{Z}\}. \tag{2.14}$$

Такая система распределений вероятностей $\mathcal{P}$ бесконечна, поэтому для ее задания необходимы некоторые конструктивные функциональные соотношения. Эти функциональные соотношения для $\mathcal{P}$ устанавливают конкретный вид вероятностной модели ДВР и будут рассматриваться в последующих разделах данной главы.

2.2. РАВНОМЕРНО РАСПРЕДЕЛЕННАЯ СЛУЧАЙНАЯ ПОСЛЕДОВАТЕЛЬНОСТЬ И ЕЕ СВОЙСТВА

Простейшей вероятностной моделью ДВР является *равномерно распределенная случайная последовательность* (РРСП), которая в криптологии иногда называется *чисто случайной последовательностью* [30].

Определение 2.10. *РРСП — это случайная последовательность* $x_1, x_2, \ldots, x_t, x_{t+1}, \ldots$ *со значениями в дискретном множестве* $A = \{0, 1, \ldots, N-1\}$, *определенная на вероятностном пространстве* $(\Omega, \mathcal{F}, \mathsf{P})$ *и удовлетворяющая двум свойствам — 2.1 и 2.2.*

Свойство 2.1. Для любого $n \in \mathbb{N}$ и произвольных значений индексов $t_1, \ldots, t_n \in \mathbb{Z}$, $t_1 < \ldots < t_n$, случайные величины $x_{t_1}, \ldots, x_{t_n} \in A$ независимы в совокупности

$$\mathsf{P}\left\{x_{t_1} = i_1, \ldots, x_{t_n} = i_n\right\} = \prod_{k=1}^{n} \mathsf{P}\left\{x_{t_k} = i_k\right\}, i_1, \ldots, i_n \in A.$$

Свойство 2.2. Для любого номера $t \in \mathbb{N}$ случайная величина x_t имеет дискретное равномерное на A распределение вероятностей

$$\mathsf{P}\{x_t = i\} = \frac{1}{N}, \, i \in A.$$

Из базовых свойств 2.1, 2.2 и определений легко доказываются следующие дополнительные свойства.

Свойство 2.3. Если $\{x_t\}$ — РРСП, то для любого $n \in \mathbb{N}$ и любой фиксированной последовательности индексов $t_1, \ldots, t_n \in \mathbb{Z}$, $t_1 < \ldots < t_n$, n-мерное дискретное распределение вероятностей случайного вектора $(x_{t_1}, x_{t_2}, \ldots, x_{t_n}) \in A^n$ является равномерным:

$$P_n(i_1, \ldots, i_n; \, t_1, \ldots, t_n) = \mathsf{P}\left\{x_{t_1} = i_1, \ldots, x_{t_n} = i_n\right\} = \frac{1}{N^n}, i_1, \ldots, i_n \in A.$$

Свойство 2.4. Если $x_t \in A$ — элемент РРСП, то $\forall k \in \mathbb{N}$ справедливы следующие выражения его начального и центрального моментов k-го порядка («моменты РРСП»):

$$\alpha_k = \mathsf{E}\left\{x_t^k\right\} = \frac{1}{N(k+1)} \sum_{l=0}^{k} \binom{k+1}{l} B_l N^{k+1-l},$$

$$\mu_k = \mathsf{E}\left\{(x_t - \alpha_1)^k\right\} = k! \sum_{l=0}^{k} \frac{B_l}{l!} \sum_{s=1}^{k+1-l} \frac{N^{s-1}(-(N-1)/2)^{k+1-l-s}}{s!(k+1-l-s)!},$$

где $\{B_l\}$ — числа Бернулли [11]. В частности, математическое ожидание $m_t = \mathsf{E}\{x_t\} = \dfrac{N-1}{2}$, а дисперсия $d_t = \mathsf{D}\{x_t\} = \dfrac{N^2-1}{12}$.

Свойство 2.5. Для ковариационной функции и спектральной плотности РРСП $\{x_t\}$ справедливы следующие выражения:

$$r(\tau) = \mathsf{E}\{(x_t - \alpha_1)(x_{t+\tau} - \alpha_1)\} = \frac{N^2-1}{12}\delta_{\tau,0}, \, \tau \in \mathbb{Z},$$

$$S(\lambda) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{+\infty} r(\tau)\cos(\lambda\tau) = \frac{N^2-1}{24\pi}, \, \lambda \in [-\pi, \, +\pi],$$

где $\delta_{i,j}$ — символ Кронекера.

Свойство 2.6 (воспроизводимость при прореживании). Для любой фиксированной последовательности моментов времени $t_1, \ldots, t_n \in \mathbb{Z}$, $t_1 < \ldots < t_n < t_{n+1} < \ldots$, при прореживании РРСП $\{x_t\}$ возникает подпоследовательность

$$y_1 = x_{t_1}, \ldots, y_n = x_{t_n}, \, y_{n+1} = x_{t_{n+1}}, \ldots,$$

также являющаяся РРСП.

Свойство 2.7 (воспроизводимость при суммировании). Если $x_t \in A$ — РРСП, а $\xi_t \in A$ — произвольная неслучайная либо случайная последовательность, не зависящая от $\{x_t\}$, то случайная последовательность

$$y_t = (x_t + \xi_t) \bmod N$$

также является РРСП.

Свойство 2.8. Если $\{x_t\}$ — РРСП, то $\forall n \in \mathbb{N}$ количество информации по Шеннону, содержащейся в отрезке последовательности $X_n = (x_1, \ldots, x_n) \in A^n$ о будущем элементе x_{n+1}, равно нулю:

$$\mathsf{I}\{x_{n+1}, X_n\} = 0,$$

поэтому для любого алгоритма прогнозирования $\hat{x}_{n+1} = f(X_n)$: $A^n \to A$ вероятность ошибки прогнозирования не может быть меньше, чем для «угадывания по жребию»:

$$\mathsf{P}\{\hat{x}_{n+1} \neq x_{n+1}\} \geqslant 1 - \frac{1}{N}\,.$$

Свойство 2.9. Если $\{x_t\}$ — РРСП, то для любого $k \in \mathbb{N}$ и произвольной интегрируемой борелевской функции k переменных $y = f(z_1, \ldots, z_k)$, $z_1, \ldots, z_k \in \in \mathbb{R}$, при $n \to \infty$ имеет место сходимость почти наверное:

$$\frac{1}{n}\sum_{\tau=1}^{n} f(x_{(\tau-1)k+1}, \ldots, x_{\tau k}) \xrightarrow{\text{п.н.}} \frac{1}{N^k} \sum_{i_1, \ldots, i_k \in A} f(t_{i_1}, \ldots, i_{i_k}).$$

Свойство 2.10. Если $\{x_t\}$ — равномерно распределенная последовательность порядка $k = \infty$ в смысле Г. Вейля [9, 28], то $\{x_t\}$ — РРСП.

2.3. ЗАДАНИЯ ДЛЯ ТЕСТОВ

2.1. Отсчет дискретного временного ряда — это:
а) случайная величина;
б) функция от параметра t;
в) множество значений этого временного ряда;
г) совокупность реализаций;
д) мощность множества значений.

2.2. Свойство согласованности n-мерного распределения вероятностей:
а) $P_n(\cdot) \in [0, 1]$;
б) $\sum_{a_1, \ldots, a_n \in A} P_n(a_1, \ldots, a_n; t_1, \ldots, t_n) \equiv 1$;
в) для $1 \leqslant k < n$

$$\sum_{a_{k+1}, \ldots, a_n \in A} P_n(a_1, \ldots, a_n; t_1, \ldots, t_n) \equiv P_k(a_1, \ldots, a_k; t_1, \ldots, t_k);$$

г) $P_n(a_1, \ldots, a_n; t_1, \ldots, t_n) = \prod_{i=1}^{n} P_1(a_i; t_i)$;
д) $\mathsf{P}\{(x_{t_1}, \ldots, x_{t_k}) \in B\} = \sum_{a_1, \ldots, a_k \in B} P_k(a_1, \ldots, a_k; t_1, \ldots, t_k)$.

2.3. Какое из указанных ниже соотношений не обязательно выполнено для дискретного стационарного в узком смысле временного ряда:

а) $P_1(a,t) = P_1(a)$;

б) $P_n(a_1, \ldots, a_n; t_1, \ldots, t_n) = \prod_{i=1}^{n} P_1(a_i; t_i)$;

в) $m(t) = m$;

г) $D(t) = D$;

д) $\sigma(t_1, t_2) = \sigma(t_2 - t_1)$.

2.4. Какое свойство, вообще говоря, не выполнено для спектральной плотности дискретного временного ряда:

а) $S(\lambda) = (1/2\pi) \sum\limits_{u=-\infty}^{+\infty} \sigma(u) \cos(\lambda u)$;

б) $S(\lambda) = (1/2\pi) \sum\limits_{u=-\infty}^{+\infty} \sigma(u) \sin(\lambda u)$;

в) $S(\lambda) \geqslant 0$, $\lambda \in [-\pi, \pi]$;

г) $S(-\lambda) = S(\lambda)$, $\lambda \in [-\pi, \pi]$;

д) $\sigma(u) = \int\limits_{-\pi}^{\pi} S(\lambda) \cos(\lambda\tau) d\lambda$, $u \in \mathbb{Z}$.

2.5. Каким свойством не обладает РРСП x_t с множеством значений $\{0, 1, 2\}$:

а) $y_t = (x_t + 1) \bmod 3$ является РРСП;

б) корреляция $r(\tau) = 2/3\delta_{\tau,0}$;

в) дисперсия $D = 2/3$;

г) математическое ожидание $m = 2/3$;

д) спектральная плотность $S(\lambda) = 1/3\pi$.

2.4. РЕШЕНИЕ ТИПОВЫХ ЗАДАЧ

Задача 2.1. Пусть $x_t \in \{0, 1\}$ — дискретная двоичная случайная последовательность. Известно, что $\mathsf{P}\{x_1 = 0\} = 1/3$, $\mathsf{P}\{x_1 = 1\} = 2/3$, кроме того, для элементов последовательности выполнено рекуррентное соотношение: $x_{t+1} = x_t \oplus 1$. Найти математическое ожидание $m(t)$, дисперсию $D(t)$, ковариационную функцию $\sigma(t_1, t_2)$. Является ли случайная последовательность x_t стационарной в широком (узком) смысле? При каком начальном распределении вероятностей случайная последовательность x_t будет стационарной в широком смысле?

Решение. Из заданного рекуррентного соотношения имеем

$$x_3 = x_2 \oplus 1 = x_1 \oplus 1 \oplus 1 = x_1.$$

Более того, можно легко доказать, что

$$x_k = \begin{cases} x_1 \oplus 1, \ k \text{ — четное;} \\ x_1, \ k \text{ — нечетное.} \end{cases}$$

Таким образом, необходимо рассмотреть два случая.

1) Число k нечетное, тогда $x_k = x_1$, отсюда получим

$$\mathbf{P}\{x_k = 0\} = \mathbf{P}\{x_1 = 0\} = 1/3, \quad \mathbf{P}\{x_k = 1\} = \mathbf{P}\{x_1 = 1\} = 2/3.$$

Зная распределение вероятностей, вычислим математическое ожидание и дисперсию:

$$m(k) = 0 \cdot \mathbf{P}\{x_k = 0\} + 1 \cdot \mathbf{P}\{x_1 = 1\} = 2/3;$$

$$D(k) = (0 - \frac{2}{3})^2 \cdot \mathbf{P}\{x_k = 0\} + (1 - \frac{2}{3})^2 \cdot \mathbf{P}\{x_1 = 1\} = \frac{4}{9} \cdot \frac{1}{3} + \frac{1}{9} \cdot \frac{2}{3} = 2/9.$$

2) Число k четное. Заметим, что для любых $x, y \in \{0, 1\}$ и момента времени k имеет место следующее соотношение для условной вероятности:

$$\mathbf{P}\{x_k = x | x_1 = y\} = \begin{cases} 0, \, x = y; \\ 1, \, x \neq y. \end{cases}$$

Тогда

$$\mathbf{P}\{x_k = 0\} = \mathbf{P}\{x_k = 0 | x_1 = 0\} \mathbf{P}\{x_1 = 0\} +$$
$$+\mathbf{P}\{x_k = 0 | x_1 = 1\} \mathbf{P}\{x_1 = 1\} = \mathbf{P}\{x_1 = 1\} = 2/3,$$
$$\mathbf{P}\{x_k = 1\} = \mathbf{P}\{x_1 = 0\} = 1/3.$$

Зная распределение вероятностей, вычислим математическое ожидание и дисперсию:

$$m(k) = 0 \cdot \mathbf{P}\{x_k = 0\} + 1 \cdot \mathbf{P}\{x_1 = 1\} = 1/3;$$

$$D(k) = (0 - \frac{1}{3})^2 \cdot \mathbf{P}\{x_k = 0\} + (1 - \frac{2}{3})^2 \cdot \mathbf{P}\{x_1 = 1\} = \frac{1}{9} \cdot \frac{2}{3} + \frac{4}{9} \cdot \frac{1}{3} = 2/9.$$

Чтобы найти ковариационную функцию, необходимо знать все попарные совместные распределения вероятностей отсчетов $\mathbf{P}\left\{x_{t_1} = x_1, x_{t_2} = x_2\right\}$. Всего возможно четыре случая таких совместных распределений в зависимости от четности t_1 и t_2. Разберем один из возможных вариантов, три остальных рассматриваются аналогично.

Пусть t_1 нечетное число, t_2 — четное, тогда

$$\mathbf{P}\left\{x_{t_1} = 0, x_{t_2} = 0\right\} = \mathbf{P}\left\{x_{t_1} = 0\right\} \mathbf{P}\left\{x_{t_2} = 0 | x_{t_1} = 0\right\} =$$
$$= \mathbf{P}\{x_1 = 0\} \mathbf{P}\left\{x_{t_2} = 0 | x_1 = 0\right\} = 0,$$

$$\mathbf{P}\left\{x_{t_1} = 0, x_{t_2} = 1\right\} = \mathbf{P}\left\{x_{t_1} = 0\right\} \mathbf{P}\left\{x_{t_2} = 1 | x_{t_1} = 0\right\} =$$
$$= \mathbf{P}\{x_1 = 0\} \mathbf{P}\left\{x_{t_2} = 1 | x_1 = 0\right\} = \mathbf{P}\{x_1 = 0\} = 1/3,$$

$$\mathbf{P}\left\{x_{t_1} = 1, x_{t_2} = 0\right\} = \mathbf{P}\left\{x_{t_1} = 1\right\} \mathbf{P}\left\{x_{t_2} = 0 | x_{t_1} = 1\right\} =$$
$$= \mathbf{P}\{x_1 = 1\} \mathbf{P}\left\{x_{t_2} = 0 | x_1 = 1\right\} = \mathbf{P}\{x_1 = 1\} = 2/3,$$

$$\mathbf{P}\left\{x_{t_1} = 1, x_{t_2} = 1\right\} = \mathbf{P}\left\{x_{t_1} = 1\right\} \mathbf{P}\left\{x_{t_2} = 1 | x_{t_1} = 1\right\} =$$
$$= \mathbf{P}\{x_1 = 1\} \mathbf{P}\left\{x_{t_2} = 1 | x_1 = 1\right\} = 0.$$

Тогда ковариация вычисляется следующим образом:

$$\mathbf{Cov}\left\{x_{t_1}, x_{t_2}\right\} = (0 - 2/3)(1 - 1/3)1/3 + (1 - 2/3)(0 - 1/3)2/3 = -2/9.$$

Произведя аналогичные вычисления для оставшихся трех случаев, получим

$$\sigma(t_1, t_2) = (-1)^{t_2 - t_1} 2/9.$$

Заметим, что данный временной ряд не является стационарным ни в широком, ни в узком смысле, так как в зависимости от t изменяется его математическое ожидание и распределение $p_t(x) = \mathbf{P}\{x_t = x\}$. Однако нетрудно заметить, что если начальное распределение заменить на $\mathbf{P}\{x_1 = 0\} = \mathbf{P}\{x_1 = 1\} = 1/2$, то распределение станет стационарным в узком и широком смысле.

Ответ: $m(t) = \{2/3, t$ — нечетное$; 1/3, t$ — четное$\}$, $D(t) = 2/9$, $\sigma(t_1, t_2) = (-1)^{t_2 - t_1} 2/9$.

Задача 2.2. У дискретной двоичной случайной последовательности $x_t \in \{0, 1\}$ известны условные распределения вероятностей $p(x_1|x_2) = \mathbf{P}\{x_t = x_1 | x_{t-1} = x_2\}$: $p(0|0) = p(1|0) = 1/2$, $p(0|1) = p$, $p(1|1) = 1 - p$, $\forall t > 1$, где $p \in [0, 1]$ — параметр. Найти такое начальное распределение $p_1(x) = \mathbf{P}\{x_1 = x\}$ вероятностей отсчета x_1 в зависимости от значения параметра p, чтобы распределение вероятностей отсчета x_2 совпадало с распределением вероятностей отсчета x_1. Будет ли при таком начальном распределении случайная последовательность стационарной в широком смысле?

Решение. Обозначим $\mathbf{P}\{x_1 = 0\} = \alpha$, тогда $\mathbf{P}\{x_1 = 1\} = 1 - \alpha$. Вычислим вероятность того, что $\mathbf{P}\{x_2 = 0\}$. По формуле полной вероятности получим

$$\mathbf{P}\{x_2 = 0\} = \mathbf{P}\{x_1 = 0\}\, p_{2,1}(0|0) + \mathbf{P}\{x_1 = 1\}\, p_{2,1}(0|1) = \frac{\alpha}{2} + (1 - \alpha)p.$$

Для того чтобы распределение вероятностей отсчета x_2 совпадало с распределением вероятностей отсчета x_1, необходимо и достаточно $\mathbf{P}\{x_2 = 0\} = \mathbf{P}\{x_1 = 0\} = \alpha$. Отсюда уравнение относительно α:

$$\alpha/2 + (1 - \alpha)p = \alpha \Rightarrow \alpha = \frac{p}{p + 1/2}.$$

Поскольку отсчеты x_3 и x_2 связаны между собой таким же соотношением, каким и x_2 с x_1, то для найденного α распределения отсчетов x_3 и x_2 тоже совпадают. В общем, по индукции можно доказать, что при таком начальном распределении вероятностей распределения всех отсчетов будут одинаковыми. Из этого следует, что математическое ожидание и дисперсия также будут одинаковыми.

Рассмотрим ковариационную функцию. Пусть $t_2 > t_1$, по формуле полной вероятности получим

$$\sigma(t_1, t_2) =$$

$$= \sum_{x_{t_2}, x_{t_2 - 1}, \ldots, x_{t_1 + 1}, x_{t_1} = 0}^{1} (x_{t_2} - m)(x_{t_1} - m) p(x_{t_2}|x_{t_2 - 1}) \ldots p(x_{t_1 + 1}|x_{t_1}) p(x_{t_1}).$$

Очевидно, что ковариационная функция зависит только от разности моментов времени, следовательно, временной ряд x_t стационарный в широком смысле.

Ответ: $\mathbf{P}\{x_1 = 0\} = p/(p + 0{,}5)$, $\mathbf{P}\{x_1 = 1\} = 0{,}5/(p + 0{,}5)$.

2.5. ЗАДАЧИ И УПРАЖНЕНИЯ

2.1. Источник без памяти задан алфавитом $A = \{0, 1\}$ и распределением вероятностей его символов $p(x) = \mathsf{P}\{x_t = x\}$: $p(0) = p$, $p(1) = 1 - p$, $p \in [0, 1]$. Вычислить математическое ожидание и дисперсию суммы n первых символов, порожденных источником.

2.2. Источник без памяти задан алфавитом $A = \{0, 1, 2\}$ и распределением вероятностей его символов $p(x) = \mathsf{P}\{x_t = x\}$: $p(0) = p$, $p(1) = q$, $p(2) = 1 - p - q$, $p, q \in [0, 1]$, $p + q \leqslant 1$. Определить математическое ожидание и дисперсию суммы n первых символов, порожденных источником.

2.3. Пусть для дискретной двоичной случайной последовательности задано начальное распределение вероятностей $p_1(x) = \mathsf{P}\{x_1 = x\}$: $p_1(0) = 1/4$, $p_1(1) = 3/4$. Найти математическое ожидание $m(t)$, дисперсию $D(t)$, ковариационную функцию $\sigma(t_1, t_2)$, установить, при каком начальном распределении вероятностей случайная последовательность x_t будет стационарной в широком смысле (если такое распределение вероятностей существует), если:

а) $x_{t+1} = x_t \oplus u_t$, где $u_t \in \{0, 1\}$ — последовательность независимых одинаково распределенных случайных величин с распределением вероятностей $\mathsf{P}\{u_t = 0\} = \mathsf{P}\{u_t = 1\} = 1/2$;

б) $x_{t+1} = u_t \cdot x_t$, где u_t — последовательность независимых одинаково распределенных случайных величин с распределением вероятностей $\mathsf{P}\{u_t = 0\} = \mathsf{P}\{u_t = 1\} = 1/2$.

2.4. Дискретная случайная последовательность $x_t \in \{0, 1, 2\}$ задана своим начальным распределением вероятностей $p_1(x) = \mathsf{P}\{x_1 = x\}$: $p_1(0) = 1/2$, $p_1(1) = 1/3$, $p_1(2) = 1/6$. Вычислить математическое ожидание $m(t)$, дисперсию $D(t)$, выяснить, при каком начальном распределении вероятностей случайная последовательность x_t будет стационарной в широком смысле (если такое распределение вероятностей существует):

а) $x_{t+1} = (x_t + 1) \bmod 3$;

б) $x_{t+1} = (x_t + u_t) \bmod 3$, где $u_t \in \{0, 1, 2\}$ — последовательность независимых одинаково распределенных случайных величин с распределением вероятностей $\mathsf{P}\{u_t = 0\} = \mathsf{P}\{u_t = 1\} = \mathsf{P}\{u_t = 2\} = 1/3$;

в) $x_{t+1} = (u_t \cdot x_t) \bmod 3$, где u_t — последовательность независимых одинаково распределенных случайных величин с распределением вероятностей $\mathsf{P}\{u_t = 0\} = \mathsf{P}\{u_t = 1\} = \mathsf{P}\{u_t = 2\} = 1/3$.

2.5. Пусть для дискретной случайной последовательности $x_t \in \{0, 1, 2\}$ задано начальное распределение вероятностей $p_1(x) = \mathsf{P}\{x_1 = x\}$: $p_1(0) = 1/2$, $p_1(1) = 1/8$, $p_1(2) = 3/8$. Найти математические ожидания $m(1)$, $m(2)$, $m(3)$ и дисперсии $D(1)$, $D(2)$, $D(3)$ случайной последовательности x_t в моменты времени $t = 1, 2, 3$, если:

а) $x_{t+1} = (x_t + 1) \bmod 3$;

б) $x_{t+1} = (2x_t + u_t) \bmod 3$, где $u_t \in \{0, 1, 2\}$ — последовательность независимых одинаково распределенных случайных величин с распределением вероят-

ностей $\mathbf{P}\{u_t = 0\} = 2/3$, $\mathbf{P}\{u_t = 1\} = 1/6$, $\mathbf{P}\{u_t = 2\} = 1/6$;

в) $x_{t+1} = (u_t \cdot x_t + 1) \bmod 3$, где u_t — последовательность независимых одинаково распределенных случайных величин с распределением вероятностей $\mathbf{P}\{u_t = 0\} = 1/8$, $\mathbf{P}\{u_t = 1\} = 3/4$, $\mathbf{P}\{u_t = 2\} = 1/8$;

г) $x_{t+1} = (u_t \cdot x_t + \beta_t) \bmod 3$, где u_t, β_t — две последовательности независимых одинаково распределенных случайных величин с распределениями вероятностей $\mathbf{P}\{u_t = 0\} = 1/4$, $\mathbf{P}\{u_t = 1\} = 1/2$, $\mathbf{P}\{u_t = 2\} = 1/4$, $\mathbf{P}\{\beta_t = 0\} = 1/2$, $\mathbf{P}\{\beta_t = 1\} = 1/4$, $\mathbf{P}\{\beta_t = 2\} = 1/4$;

д) $x_{t+1} = (\beta_t(u_t \cdot x_t) + (1 - \beta_t)(x_t + u_t)) \bmod 3$, где u_t, β_t — две последовательности независимых одинаково распределенных случайных величин с распределениями вероятностей $\mathbf{P}\{u_t = 0\} = 1/5$, $\mathbf{P}\{u_t = 1\} = 2/5$, $\mathbf{P}\{u_t = 2\} = 2/5$, $\mathbf{P}\{\beta_t = 0\} = 1/3$, $\mathbf{P}\{\beta_t = 1\} = 2/3$.

2.6. Для дискретной случайной последовательности $x_t \in \{1, 2, 3\}$ задано начальное распределение вероятностей $p_1(x) = \mathbf{P}\{x_1 = x\}$: $p_1(1) = 0{,}7$, $p_1(2) = 0{,}2$, $p_1(3) = 0{,}1$. Кроме того, известны условные распределения вероятностей $p(x_1|x_2) = \mathbf{P}\{x_t = x_1|x_{t-1} = x_2\}$ для любого $t > 1$, которое удобно представить в виде матрицы $\{P_{ij}\}$, $P_{ij} = p(x_j|x_i)$:

$$P = \begin{pmatrix} 0{,}1 & 0{,}5 & 0{,}4 \\ 0{,}6 & 0{,}2 & 0{,}2 \\ 0{,}3 & 0{,}4 & 0{,}3 \end{pmatrix}.$$

Определить:

1) распределение по состояниям в момент $t = 2$;

2) вероятность того, что в моменты $t = 0, 1, 2, 3$ случайная последовательность принимала значения 1, 3, 3, 2 соответственно.

2.7. У дискретной двоичной случайной последовательности $x_t \in \{0, 1\}$ известны условные распределения вероятностей $p(x_1|x_2) = \mathbf{P}\{x_t = x_1|x_{t-1} = x_2\}$ для любого $t > 1$. Найти такое начальное распределение $p_1(x) = \mathbf{P}\{x_1 = x\}$ вероятностей отсчета x_1, чтобы распределения вероятностей отсчетов x_1 и x_2 совпадали. Будет ли при таком начальном распределении случайная последовательность стационарной в широком смысле?

а) $p(0|0) = 0{,}7$, $p(1|0) = 0{,}3$, $(0|1) = p$, $p(1|1) = 1 - p$;

б) $p(0|0) = 0{,}9$, $p(1|0) = 0{,}1$, $(0|1) = 1 - p$, $p(1|1) = p$;

в) $p(0|0) = q$, $p(1|0) = 1 - q$, $(0|1) = 1 - p$, $p(1|1) = p$.

2.8. Доказать, что если $\{x_t\}$ — РРСП, то для любого $n \in \mathbb{N}$ и любой фиксированной последовательности индексов $t_1, \ldots, t_n \in \mathbb{Z}$, $t_1 < \ldots < t_n$, n-мерное дискретное распределение вероятностей случайного вектора $(x_{t_1}, x_{t_2}, \ldots, x_{t_n}) \in A^n$ является равномерным:

$$P_n(i_1, \ldots, i_n; t_1, \ldots, t_n) = \mathbf{P}\left\{x_{t_1} = i_1, \ldots, x_{t_n} = i_n\right\} = \frac{1}{N^n}, \forall i_1, \ldots, i_n \in A.$$

2.9. Доказать, что если $x_t \in A$ — РРСП, а $\xi_t \in A$ — произвольная неслучайная либо случайная последовательность, не зависящая от $\{x_t\}$, то случайная

последовательность

$$y_t = (x_t + \xi_t) \bmod N$$

также является РРСП.

2.10. Доказать, что если $\{x_t\}$ — РРСП, то $\forall n \in \mathbb{N}$ количество информации по Шеннону, содержащейся в отрезке последовательности $X_n = (x_1, \ldots, x_n) \in A^n$, о будущем элементе x_{n+1} равно нулю:

$$\mathsf{I}\{x_{n+1}, X_n\} = 0.$$

Глава 3

ПРЕОБРАЗОВАНИЯ СЛУЧАЙНЫХ ПРОЦЕССОВ. ЭРГОДИЧНОСТЬ СЛУЧАЙНЫХ ПРОЦЕССОВ

3.1. ОПРЕДЕЛЕНИЕ ЭРГОДИЧНОСТИ

Пусть на вероятностном пространстве $(\Omega, \mathcal{F}, \mathsf{P})$ определена дискретная вероятностная схема $\langle A,\, p\,(\cdot)\rangle$, порождающая случайную последовательность $x_t \in A$, $t \in \mathbb{Z}$, и $a = (\ldots, a_{-1}, a_0, a_1, \ldots) \in A^\infty$ — произвольная реализация этой случайной последовательности. Обозначим через T оператор сдвига на одну позицию влево членов последовательности:

$$Ta = (\ldots, a_0, a_1, a_2, \ldots) \in A^\infty. \tag{3.1}$$

T^k — это k-кратное применение оператора сдвига, приводящее к сдвигу на k позиций влево членов последовательности. Если $B \subset A^\infty$ — некоторое подмножество реализаций, то

$$TB = \left\{b' \in A^\infty : b' = Tb, b \in B\right\} \subset A^\infty.$$

Заметим, что выполняются свойства: 1) $TA^\infty = A^\infty$; 2) если $B \in \mathcal{F}$, то $TB \in \mathcal{F}$.

Пусть далее $f(\cdot) : A^\infty \to R$ — некоторая $\mathcal{F}$-измеримая действительнозначная функция, суммируемая относительно меры $p(\cdot)$, т. е. такая, для которой существует математическое ожидание (интеграл Лебега) случайной величины $\xi = f(\ldots, x_{-1}, x_0, x_1, \ldots)$:

$$\mathsf{E}\left\{f(\ldots, x_{-1}, x_0, x_1, \ldots)\right\} = \int\limits_{A^\infty} f(a)p(da). \tag{3.2}$$

Соотношение (3.2) означает, что существует математическое ожидание для случайной величины $\xi \in \mathbb{R}$, являющейся результатом применения функционального преобразования $f(\cdot)$ к случайной последовательности $\{x_t : t \in \mathbb{Z}\}$.

Определение 3.1. *Стационарный источник дискретных сообщений $\langle A,\, p\,(\cdot)\rangle$ называется эргодическим, если любое измеримое инвариантное к сдвигу T событие (множество последовательностей $B \in \mathcal{F}$, порожденных этим ИДС) имеет вероятность либо 1, либо 0:*

$$TB = B \Rightarrow p(B) \in \{0, 1\}. \tag{3.3}$$

Проверка условия (3.3) затруднительна. Существует более удобное эквивалентное определение эргодичности, основанное на *эргодической теореме Биркгофа* [8].

3.2. ТЕОРЕМА БИРКГОФА. ЭРГОДИЧНОСТЬ ДИСКРЕТНОГО ИСТОЧНИКА БЕЗ ПАМЯТИ

Теорема 3.1 (теорема Биркгофа). *Для любого ИДС* $\langle A, p(\cdot)\rangle$ *и любой суммируемой функции* $f(\cdot): A^k \to \mathbb{R}$ *существует почти всюду предел*

$$\lim_{l\to\infty} \frac{1}{l} \sum_{k=0}^{l-1} f\left(T^k a\right) = h(a),$$

где функция $h(a)$ *инвариантна относительно сдвига* T *для всех* $a \in A^\infty$*, для которых этот предел существует.*

Определение 3.2. *Стационарный ИДС* $\langle A, p(\cdot)\rangle$ *называется эргодическим, если почти всюду выполняется равенство*

$$h(a) = \mathsf{E}\left\{f(\ldots, x_{-1}, x_0, x_1, \ldots)\right\},$$

т. е.

$$\lim_{l\to\infty} \frac{1}{l} \sum_{k=0}^{l-1} f\left(T^k a\right) \overset{\text{п.н.}}{=} \mathsf{E}\left\{f(\ldots, x_{-1}, x_0, x_1, \ldots)\right\}. \tag{3.4}$$

Соотношение (3.4) означает, что ИДС $\langle A, p(\cdot)\rangle$ является эргодическим, если для любого $k \in \mathbb{N}$ и любой измеримой суммируемой функции $f_k(\cdot): A^k \to \mathbb{R}$ имеет место сходимость почти наверное последовательности средних арифметических к математическому ожиданию

$$\frac{1}{l} \sum_{i=0}^{l-1} f_k\left(x_{i+1}, \ldots, x_{i+k}\right) \xrightarrow{\text{п.н.}} \mathsf{E}\left\{f_k(x_1, \ldots, x_k)\right\} = \mu_k. \tag{3.5}$$

Практическая ценность свойства эргодичности (3.5) состоит в том, что удается построить сильно состоятельную оценку математического ожидания μ_k (правая часть (3.5)) с помощью *скользящего среднего по времени* (левая часть (3.5)):

$$\hat{\mu}_k = \frac{1}{T} \sum_{i=0}^{T-1} f_k\left(x_{i+1}, \ldots, x_{i+k}\right)$$

по наблюдаемому фрагменту $x_1, \ldots, x_T$ достаточно большой длины T.

Теорема 3.2. *Всякий ДИБП является эргодическим.*

Доказательство. Проверим справедливость предельного соотношения (3.5). Пусть $f_k(\cdot): A^k \to \mathbb{R}$ — произвольная измеримая суммируемая функция, $k \in \mathbb{N}$. Без потери общности выберем l кратным k: $l = m \cdot k$. Представим левую часть (3.5) в эквивалентном виде:

$$\frac{1}{l}\sum_{i=0}^{l-1} f_k\left(x_{i+1},\,\ldots,x_{i+k}\right) \equiv \frac{1}{k}\sum_{s=1}^{k}\left(\frac{1}{m}\sum_{j=0}^{m-1} f_k\left(x_{jk+s+1},\,\ldots,x_{jk+s+k}\right)\right) = \frac{1}{k}\sum_{s=1}^{k} F_s, \tag{3.6}$$

где

$$F_s = \frac{1}{m}\sum_{j=0}^{m-1} f_k\left(x_{jk+s+1},\,\ldots,x_{jk+s+k}\right). \tag{3.7}$$

По построению F_s согласно (3.7) все слагаемые в этой сумме независимы и одинаково распределены, так как $\{x_t\}$ — независимые одинаково распределенные случайные величины. Поскольку функции $f_k(\cdot)$ являются суммируемыми, к случайной последовательности средних арифметических (3.7) применим усиленный закон больших чисел (теорема Колмогорова):

$$F_s = \frac{1}{m}\sum_{j=0}^{m-1} f_k\left(x_{jk+s+1},\,\ldots,x_{jk+s+k}\right) \xrightarrow[m\to\infty]{\text{п.н.}} \mu_k. \tag{3.8}$$

Согласно (3.6), (3.8) при каждом $k \in \mathbb{N}$ имеем

$$\frac{1}{k}\sum_{s=1}^{k} F_s \xrightarrow[m\to\infty]{\text{п.н.}} \mu_k,$$

что доказывает справедливость (3.5). □

3.3. РЕШЕНИЕ ТИПОВЫХ ЗАДАЧ

Задача 3.1. Имеется две урны. В первой урне лежат 2 белых и 1 черный шар. Во второй урне лежат 1 белый и 2 черных шара. Производится следующий эксперимент. Случайно выбирается одна из двух урн, причем первая — с вероятностью p. Затем из выбранной урны осуществляется последовательное извлечение шаров с возвращением, при этом $x_t = 1$, если в момент времени t экспериментатор достал белый шар, иначе $x_t = 0$. Будет ли случайная последовательность x_t эргодической?

Решение. Экспериментатор подсчитывает число извлечений белых шаров как $s_n = \dfrac{1}{n}\sum_{i=1}^{n} x_t$. Если в начале эксперимента выбрана первая урна, то по усиленному закону больших чисел s_n почти наверное сходится к $2/3$, в противном случае — к $1/3$. При этом математическое ожидание $\mathsf{E}\{x_t\} = 2p/3 + (1-p)/3 = (1+p)/3$ при любом значении $p \in (0,1)$ отличается от возможных пределов s_n и, значит, случайная последовательность x_t не является эргодической.

Из усиленного закона больших чисел следует, что при $p = 1$ или $p = 0$ случайная последовательность x_t будет эргодической.

Ответ: При $p \in (0,1)$ x_t не эргодическая случайная последовательность, при $p = 1$ или $p = 0$ — эргодическая.

3.4. ЗАДАЧИ И УПРАЖНЕНИЯ

3.1. Имеется две урны. В первой урне лежат a_1 белых и b_1 черных шаров. Во второй — a_2 белых и b_2 черных шаров. Производится следующий эксперимент. Случайно выбирается одна из двух урн, причем первая — с вероятностью p. Затем из выбранной урны осуществляется последовательное извлечение шаров с возвращением, при этом $x_t = 1$, если в момент времени t экспериментатор достал белый шар, иначе $x_t = 0$. Будет ли случайная последовательность x_t эргодической, если:

а) $a_1 = 1$, $b_1 = 3$, $a_2 = 2$, $b_2 = 3$, $p = 2/3$;

б) $a_1 = 2$, $b_1 = 5$, $a_2 = 3$, $b_2 = 4$, $p = 4/7$;

в) $a_1 = b_1 = 1$, $a_2 = b_2 = 2$, $p = 1/2$?

3.2. Для дискретного временного ряда $x_t \in \{0, \ldots, n-1\}$ заданы начальное распределение вероятностей $\pi(x) = \mathsf{P}\{x_1 = x\}$ и условные распределения $p(y|x) = \mathsf{P}\{x_2 = y | x_1 = x\}$, которые удобно представить в виде матрицы $\{P_{ij}\}$, $P_{ij} = p(x_j|x_i)$. Доказать, что дискретный временной ряд будет эргодическим, если:

а) $n = 2$, $\pi(0) = 1/2$, $\pi(1) = 1/2$, $P = \begin{pmatrix} 1/4 & 3/4 \\ 1/4 & 3/4 \end{pmatrix}$;

б) $\pi(0) = 1/n, \ldots, \pi(n-1) = 1/n$, $P = \begin{pmatrix} p_1 & p_2 & \cdots & p_n \\ \vdots & \vdots & \ddots & \vdots \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}$, $p_1 + \ldots + p_n = 1$.

3.3. Рассмотрим двоичный эргодический источник, выбирающий сообщения из множества $\{0, 1\}$. Указать функцию $f()$ для построения строго состоятельной оценки вероятности:

а) появления пары сообщений $(1, 0)$;

б) появления k единиц в последовательности длиной n.

Глава 4

СТАЦИОНАРНЫЕ ИСТОЧНИКИ СООБЩЕНИЙ И ИХ ЭНТРОПИЙНЫЕ СВОЙСТВА

4.1. УДЕЛЬНАЯ ЭНТРОПИЯ СТАЦИОНАРНОЙ СИМВОЛЬНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ

Пусть рассматривается стационарный ИДС (см. разд. 1.1) с некоторым алфавитом $A = \left\{a^{(1)}, \ldots, a^{(N)}\right\}$, порождающий (генерирующий) n-символьные случайные сообщения $\Xi_n = (\xi_1, \ldots, \xi_n) \in A^n$. Исследуем асимптотическое поведение энтропии $\mathsf{H}\{\Xi_n\}$ при $n \to \infty$.

По свойству стационарности (в узком смысле) n-мерное дискретное распределение вероятностей

$$p_n(a_1, \ldots, a_n) = \mathsf{P}\{\xi_1 = a_1, \ldots, \xi_n = a_n\},\ a_1, \ldots, a_n \in A, \tag{4.1}$$

не меняется при сдвиге $\tau \geqslant 1$ начала отсчета времени:

$$\mathsf{P}\{\xi_{\tau+1} = a_1, \ldots, \xi_{\tau+n} = a_n\} = p_n(a_1, \ldots, a_n) = \operatorname*{invar}_{\tau}. \tag{4.2}$$

В силу (4.1) и (4.2) энтропия n-символьного «сдвинутого» сообщения $\tilde{\Xi}_{n,\tau} = (\xi_{\tau+1}, \ldots, \xi_{\tau+n}) \in A^n$ не зависит от τ:

$$\mathsf{H}\left\{\tilde{\Xi}_{n,\tau}\right\} = \mathsf{H}\{\Xi_n\} = \operatorname*{invar}_{\tau}, \tag{4.3}$$

поэтому в дальнейшем при исследовании энтропии можно не делать различий между случайными сообщениями Ξ_n и $\tilde{\Xi}_{n,\tau}$.

Удельной энтропией (*плотностью энтропии*) стационарного ИДС называется предел

$$h = \lim_{n\to\infty} \frac{\mathsf{H}\{\Xi_n\}}{n}, \tag{4.4}$$

если он существует.

Согласно определению (4.4) h представляет собой энтропию, приходящуюся на один символ и вычисленную по бесконечно длинному случайному сообщению. Если $h > 0$, то с увеличением длины сообщения $n \to \infty$ энтропия растет линейно:

$$\mathsf{H}\{\Xi_n\} \sim hn. \tag{4.5}$$

Символ $\sim$ означает, что $\mathsf{H}\{\Xi_n\}/(hn) \to 1$. Выясним условия существования предела в (4.4), т. е. асимптотического поведения энтропии (4.5).

Пусть $\Xi_{n-1} \in A^{n-1}$ и $\Xi_n = (\Xi_{n-1}||\xi_n) \in A^n$ — случайные сообщения длиной $n-1$ и n соответственно (здесь $||$ — знак конкатенации (присоединения)). Обозначим условную энтропию символа ξ_n относительно случайного сообщения Ξ_{n-1}, состоящего из $n-1$ предыдущих символов ($n = 1, 2, \ldots$):

$$\mathbf{H}\{\xi_n \mid \Xi_{n-1}\} = -\sum_{a_1,\ldots,a_n \in A} p_n(a_1, \ldots, a_n) \log p_1(a_n \mid a_1, \ldots, a_{n-1}). \tag{4.6}$$

При этом в случае $n = 1$ полагается $\mathbf{H}\{\xi_1 \mid \Xi_0\} ::= \mathbf{H}\{\xi_1\}$.

Теорема 4.1. *Для произвольного стационарного ИДС числовая последовательность условных энтропий $\mathbf{H}\{\xi_n \mid \Xi_{n-1}\}$, $n = 1, 2, \ldots$, определяемых (4.6), имеет конечный предел*

$$\mathbf{H}\{\xi \mid \Xi_\infty\} ::= \lim_{n\to\infty} \mathbf{H}\{\xi_n \mid \Xi_{n-1}\}. \tag{4.7}$$

Доказательство. Докажем, что данная последовательность не возрастает и ограничена снизу. Действительно, по свойству стационарности (4.3) $\forall i, j \in \mathbb{N}$

$$\mathbf{H}\{\xi_j \mid \xi_{j-i}, \ldots, \xi_{j-1}\} = \mathbf{H}\{\xi_n \mid \xi_{n-i}, \ldots, \xi_{n-1}\}.$$

Поэтому рассматриваемая последовательность совпадает с последовательностью

$$\mathbf{H}\{\xi_n\}, \mathbf{H}\{\xi_n \mid \xi_{n-1}\}, \ldots, \mathbf{H}\{\xi_n \mid \xi_1, \ldots, \xi_{n-1}\}.$$

В силу свойств условной энтропии

$$\mathbf{H}\{\xi_n\} \geqslant \mathbf{H}\{\xi_n \mid \xi_{n-1}\} \geqslant \cdots \geqslant \mathbf{H}\{\xi_n \mid \xi_1, \ldots, \xi_{n-1}\} \geqslant 0.$$

Как известно, любая невозрастающая ограниченная снизу числовая последовательность имеет предел, который обозначается согласно (4.7). □

Теорема 4.2. *Для произвольного стационарного ИДС удельная энтропия (4.4) существует и совпадает с предельным значением (4.7):*

$$h = \mathbf{H}\{\xi \mid \Xi_\infty\}. \tag{4.8}$$

Доказательство. Во-первых, докажем, что числовая последовательность

$$h_n ::= \frac{\mathbf{H}\{\Xi_n\}}{n}, n = 1, 2, \ldots, \tag{4.9}$$

из определения удельной энтропии (4.4) $\left(h ::= \lim\limits_{n\to\infty} h_n\right)$ является невозрастающей и ограниченной снизу. Для $(n+1)$-символьного сообщения $\Xi_{n+1} = (\Xi_n||\xi_{n+1}) \in A^{n+1}$ по свойству иерархической аддитивности энтропии имеем

$$\mathbf{H}\{\Xi_{n+1}\} = \mathbf{H}\{\Xi_n\} + \mathbf{H}\{\xi_{n+1} \mid \Xi_n\}. \tag{4.10}$$

Согласно свойству условной энтропии (следствие 1.4) и свойству стационарности (4.3) имеем оценку для второго слагаемого в (4.10):

$$\mathbf{H}\{\xi_{n+1} \mid \Xi_n\} = \mathbf{H}\{\xi_{n+1} \mid \xi_1, \ldots, \xi_n\} \leqslant \mathbf{H}\{\xi_{n+1} \mid \xi_2, \ldots, \xi_n\} =$$

$$= \mathbf{H}\{\xi_n \mid \xi_1, \ldots, \xi_{n-1}\} = \mathbf{H}\{\xi_n \mid \Xi_{n-1}\}.$$

Подставляя ее в (4.10), получим неравенство

$$\mathbf{H}\{\Xi_{n+1}\} \leqslant \mathbf{H}\{\Xi_n\} + \mathbf{H}\{\xi_n \mid \Xi_{n-1}\}. \tag{4.11}$$

В силу свойства иерархической аддитивности энтропии (теорема 1.2) и свойств условной энтропии

$$\mathbf{H}\{\Xi_n\} = \sum_{i=1}^{n} \mathbf{H}\{\xi_i \mid \Xi_{i-1}\} \geqslant n\mathbf{H}\{\xi_n \mid \Xi_{n-1}\},$$

так что

$$\mathbf{H}\{\xi_n \mid \Xi_{n-1}\} \leqslant \frac{1}{n}\mathbf{H}\{\Xi_n\} = h_n. \tag{4.12}$$

Подставляя (4.12) в (4.11), получим

$$0 \leqslant \mathbf{H}\{\Xi_{n+1}\} \leqslant \left(1 + \frac{1}{n}\right)\mathbf{H}\{\Xi_n\} = \frac{n+1}{n}\mathbf{H}\{\Xi_n\}.$$

Разделив обе части этого неравенства на $n+1$ и использовав обозначение (4.9), найдем

$$0 \leqslant h_{n+1} \leqslant h_n,$$

т. е. $\{h_n\}$ — невозрастающая числовая последовательность, ограниченная снизу. Следовательно, ее предел (4.4) — удельная энтропия h — существует.

Во-вторых, покажем справедливость равенства (4.8) для этого предела. Для произвольного $1 \leqslant k < n$ по свойствам энтропии с учетом (4.9)

$$h_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{H}\{\xi_i \mid \Xi_{i-1}\} \equiv \frac{1}{n}\sum_{i=1}^{k} \mathbf{H}\{\xi_i \mid \Xi_{i-1}\} +$$

$$+\frac{1}{n}\sum_{i=k+1}^{n} \mathbf{H}\{\xi_i \mid \Xi_{i-1}\} \leqslant \frac{k}{n}\mathbf{H}\{\xi_1\} + \frac{n-k}{n}\mathbf{H}\{\xi_{k+1} \mid \Xi_k\}. \tag{4.13}$$

Воспользуемся произволом k и выберем $k = k(\varepsilon)$ таким, чтобы для любого наперед заданного $\varepsilon > 0$ выполнялось неравенство

$$\mathbf{H}\{\xi_{k+1} \mid \Xi_k\} - \mathbf{H}\{\xi \mid \Xi_\infty\} \leqslant \frac{\varepsilon}{2}. \tag{4.14}$$

Это всегда можно сделать, так как по теореме 4.1

$$\mathbf{H}\{\xi_{k+1} \mid \Xi_k\} \to \mathbf{H}\{\xi \mid \Xi_\infty\} + 0.$$

По выбранному таким образом k определим $\bar{n} = \bar{n}(k, \varepsilon)$, чтобы при любом $n > \bar{n}$ выполнялось неравенство

$$\frac{k}{n}\mathbf{H}\{\xi_1\} \leqslant \frac{\varepsilon}{2}. \tag{4.15}$$

Тогда из (4.13)–(4.15) заключим, что $\forall \varepsilon > 0$ $\exists \bar{n}_1 = \bar{n}_1(\varepsilon)$ такое, что для всех $n > \bar{n}_1$ справедлива оценка сверху:

$$h_n \leqslant \mathbf{H}\{\xi \mid \Xi_\infty\} + \varepsilon. \tag{4.16}$$

С другой стороны, имеем оценку снизу:

$$h_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{H}\{\xi_i \mid \Xi_{i-1}\} \geqslant \mathbf{H}\{\xi_n \mid \Xi_{n-1}\} \geqslant \mathbf{H}\{\xi \mid \Xi_\infty\}. \tag{4.17}$$

Объединяя (4.16) и (4.17), получим

$$\mathbf{H}\{\xi \mid \Xi_\infty\} \leqslant h_n \leqslant \mathbf{H}\{\xi \mid \Xi_\infty\} + \varepsilon.$$

Поскольку ε — произвольное положительное число, то из этих неравенств сделаем вывод, что последовательность h_n имеет предел

$$h = \lim_{n\to\infty} h_n = \mathbf{H}\{\xi \mid \Xi_\infty\},$$

т. е. выполняется (4.8). □

Из теорем 4.1 и 4.2 следует *асимптотическое разложение энтропии* n-символьного сообщения:

$$\mathbf{H}\{\Xi_n\} = nh + o(n). \tag{4.18}$$

Уточним остаточный член в (4.18).

Теорема 4.3. *Для энтропии произвольной дискретной стационарной случайной последовательности $\Xi_n \in A^n$ при увеличении количества символов $n \to \infty$ справедлива асимптотика*

$$\mathbf{H}\{\Xi_n\} = nh + 2b + o(1), \tag{4.19}$$

где

$$b = \frac{1}{2}\lim_{m,n\to\infty} (\mathbf{H}\{\Xi_m\} + \mathbf{H}\{\Xi_n\} - \mathbf{H}\{\Xi_{m+n}\}) \geqslant 0. \tag{4.20}$$

Доказательство. Обозначим

$$B_{mn} = \mathbf{H}\{\Xi_m\} + \mathbf{H}\{\Xi_n\} - \mathbf{H}\{\Xi_{m+n}\},\ m, n \in \mathbb{N}. \tag{4.21}$$

Во-первых, заметим, что B_{mn} — симметричная функция относительно m, n.

Во-вторых, по свойству иерархической аддитивности и свойству стационарности из (4.7) имеем

$$\begin{aligned} B_{mn} &= \mathbf{H}\{\Xi_m\} + \mathbf{H}\{\xi_{m+1}, \ldots, \xi_{m+n}\} - \\ &-(\mathbf{H}\{\Xi_m\} + \mathbf{H}\{\xi_{m+1}, \ldots, \xi_{m+n}\} \mid \Xi_m) = \\ &= \mathbf{H}\{\xi_{m+1}, \ldots, \xi_{m+n}\} - \mathbf{H}\{\xi_{m+1}, \ldots, \xi_{m+n} \mid \Xi_m\}. \end{aligned} \tag{4.22}$$

В силу известного свойства энтропии для $\eta = (\xi_{m+1}, \ldots, \xi_{m+n})$

$$\mathbf{H}\{\eta\} \geqslant \mathbf{H}\{\eta \mid \xi_1\} \geqslant \cdots \geqslant \mathbf{H}\{\eta \mid \xi_1, \ldots, \xi_m\}$$

при фиксированном n и растущем m энтропия

$$\mathbf{H}\{\xi_{m+1}, \ldots, \xi_{m+n} \mid \Xi_m\}$$

не возрастает. Это означает, что при фиксированном n последовательность B_{mn} неубывающая по $m = 1, 2, \ldots$. Аналогичное свойство монотонности выполняется для B_{mn} по n, так как B_{mn} симметрична относительно m, n.

Согласно установленному свойству монотонности существует предел (4.20) — конечный или бесконечный.

Чтобы доказать (4.19), воспользуемся (4.20), (4.22) и свойством стационарности:

$$2b = \lim_{m,\, n\to\infty} \left(\mathbf{H}\left\{\Xi_n\right\} - \mathbf{H}\left\{\xi_{m+1},\, \ldots,\, \xi_{m+n} \mid \xi_1,\, \ldots,\, \xi_m\right\}\right). \tag{4.23}$$

Применим еще раз свойство иерархической аддитивности энтропии:

$$\mathbf{H}\left\{\xi_{m+1},\, \ldots,\, \xi_{m+n} \mid \xi_1,\, \ldots,\, \xi_m\right\} = \\ = \sum_{i=1}^{n} \mathbf{H}\left\{\xi_{m+i} \mid \xi_1,\, \ldots,\, \xi_{m+i-1}\right\}. \tag{4.24}$$

Устремим $m \to \infty$ в выражении (4.24) и воспользуемся теоремой 4.2 ($i = \overline{1, n}$):

$$\mathbf{H}\left\{\xi_{m+i} \mid \xi_1,\, \ldots,\, \xi_{m+i-1}\right\} \to h, m \to \infty.$$

В результате из (4.24) получим

$$\lim_{m\to\infty} \mathbf{H}\left\{\xi_{m+1},\, \ldots,\, \xi_{m+n} \mid \xi_1,\, \ldots,\, \xi_m\right\} = nh. \tag{4.25}$$

Переходя в (4.23) к пределу при $m \to \infty$ и используя (4.25), имеем

$$\lim_{n\to\infty} \left(\mathbf{H}\left\{\Xi_n\right\} - nh\right) = 2b,$$

что эквивалентно (4.19). □

Асимптотическое разложение энтропии (4.19) имеет следующую содержательную интерпретацию. Главный член разложения nh — это n-кратная удельная энтропия. Второй член разложения $2b$ — энтропия, обусловленная краевыми (граничными) эффектами. Третий член $o(1)$ — остаточный член разложения.

Следствие 4.1. *Для стационарного ИДС без памяти соотношение (4.19) обращается в точное равенство*

$$\mathbf{H}\left\{\Xi_n\right\} = nh, \tag{4.26}$$

где

$$h = \mathbf{H}\left\{\xi_1\right\} = -\sum_{a_1 \in A} p_1(a_1) \log p_1(a_1)$$

есть энтропия единичного случайного символа.

Доказательство. Соотношение (4.26) следует из свойств условной энтропии. Кроме того, $\mathbf{H}\left\{\Xi_{m+n}\right\} = \mathbf{H}\left\{\Xi_m\right\} + \mathbf{H}\left\{\Xi_n\right\}$, поэтому согласно (4.20) граничные эффекты отсутствуют: $b = 0$. □

4.2. АСИМПТОТИЧЕСКИЕ ЭНТРОПИЙНЫЕ СВОЙСТВА ИСТОЧНИКА ДИСКРЕТНЫХ СООБЩЕНИЙ БЕЗ ПАМЯТИ

Пусть рассматривается стационарный ИДС без памяти (ДИБП) с алфавитом $A = V = \{0, 1\}$, порождающий случайную последовательность независимых в совокупности, одинаково распределенных двоичных символов: $\Xi_n =$

$= (\xi_1, \ldots, \xi_n), n = 1, 2, \ldots$, где $\xi_i \in \{0, 1\}$ — двоичный случайный символ с распределением вероятностей Бернулли $\left(i = \overline{1, n}\right)$:

$$\mathsf{P}\{\xi_i = 1\} = p, \mathsf{P}\{\xi_i = 0\} = q = 1 - p\,, \tag{4.27}$$

где $0 < p < 1/2$ — вероятность появления 1 (случай $p > 1/2$ сводится к этому случаю переобозначением символов $0 \leftrightarrow 1$).

Количество реализаций Ξ_n равно 2^n. В силу схемы независимых испытаний порождения символов распределение вероятностей Ξ_n задается соотношением

$$p_n(a) = p_n(a_1, \ldots, a_n) ::= \mathsf{P}\{\Xi_n = a\} = p^b q^{n-b}, b = \sum_{i=1}^{n} a_i\,, \tag{4.28}$$

или

$$\log p_n(a) = n \log q - \sum_{i=1}^{n} a_i \log \frac{q}{p} = n \log q - b \log \frac{q}{p}\,,$$

где $a = (a_i) \in V_n$, $b = \sum_{i=1}^{n} a_i$ — суммарное количество 1 в двоичном n-векторе a, или значение случайной величины

$$\eta_n = \sum_{i=1}^{n} \xi_i = |\Xi_n|^2. \tag{4.29}$$

Значения вероятностей (4.28) существенно меняются при изменении величины b. Отношение наибольшей из этих вероятностей к наименьшей равно

$$\kappa_n = \frac{\max\limits_a p_n(a)}{\min\limits_a p_n(a)} = \frac{q^n}{p^n} = \left(\frac{q}{p}\right)^n > 1, \tag{4.30}$$

и эта величина экспоненциально увеличивается с ростом n.

С учетом (4.27) и (4.29) вычислим моменты первого и второго порядка для случайной величины η_n:

$$\mathsf{E}\{\eta_n\} = n\mathsf{E}\{\xi_1\} = np < n/2, \quad \mathsf{D}\{\eta_n\} = n\mathsf{D}\{\xi_1\} = np\,q\,. \tag{4.31}$$

В силу (4.29) уклонение случайного числа единиц в Ξ_n от его среднего значения

$$\zeta_n = \eta_n - np \tag{4.32}$$

имеет нулевые среднее и среднеквадратическое отклонения

$$\sigma_{\zeta n} = \sqrt{\mathsf{D}\{\zeta_n\}} = \sqrt{npq}\,,$$

а для относительного уклонения

$$\delta_n = \zeta_n / n \tag{4.33}$$

имеем

$$\mathsf{E}\{\delta_n\} = 0, \sigma_{\delta_n} = \sqrt{\mathsf{D}\{\delta_n\}} = \sqrt{\frac{p\,q}{n}}\,. \tag{4.34}$$

Как видно из (4.34), среднеквадратичное отклонение для случайной величины δ_n при $n \to \infty$ убывает как $1/\sqrt{n}$. Если $|b_n - np| = c\sqrt{npq}$, $c > 0$, то различие наибольшей и наименьшей вероятностей на этом подмножестве значений Ξ_n по-прежнему достаточно велико: $\kappa_n = (q/p)^{c\sqrt{npq}}$. Однако эта величина растет при $n \to \infty$ значительно медленнее, чем $(1/p)^{c\sqrt{npq}}$. Следовательно, справедливо неравенство

$$\log \frac{p_n(a)}{p_n(a')}\bigg|_{|a'|^2=|a|^2-c\sqrt{npq}} \ll \log \frac{1}{p_n(a)}\,.$$

Сформулируем это свойство в виде теоремы.

Теорема 4.4 (первая теорема Шеннона об асимптотической равнораспределенности). *Множество всех 2^n реализаций определенного выше двоичного случайного вектора $\Xi_n \in V_n$ можно разбить на два непересекающихся подмножества:*

$$V_2^n = A_n \cup B_n\,,\, A_n \cap B_n = \varnothing\,, \tag{4.35}$$

так что при $n \to \infty$ выполняются свойства:

а) *множество A_n имеет исчезающе малую вероятность*

$$\mathsf{P}\{\Xi_n \in A_n\} = \sum_{a \in A_n} p_n(a) \to 0\,; \tag{4.36}$$

б) *реализации из множества B_n становятся относительно равновероятными:*

$$\left|\frac{\log p_n(a) - \log p_n(a')}{\log p_n(a)}\right| \to 0,\ a, a' \in B_n\,. \tag{4.37}$$

Доказательство. Воспользуемся неравенством Чебышева (относительно дисперсий) для случайной величины η_n, определяемой (4.29) с учетом (4.31):

$$\forall \varepsilon > 0\ :\ \mathsf{P}\{|\eta_n - np| \geqslant \varepsilon\} \leqslant \frac{\mathsf{D}\{\eta_n\}}{\varepsilon^2} = \frac{npq}{\varepsilon^2}\,.$$

Полагая $\varepsilon = n^{3/4}$, получим

$$\mathsf{P}\left\{|\eta_n - np| \geqslant n^{3/4}\right\} \leqslant \frac{pq}{\sqrt{n}}\,. \tag{4.38}$$

Построим разбиение множества V_n следующим образом:

$$A_n = \left\{a = (a_i) \in V_n\ :\ \left|\sum_{i=1}^n a_i - np\right| \geqslant n^{3/4}\right\},$$

$$B_n = V_n \setminus A_n = \left\{a = (a_i) \in V_n\ :\ \left|\sum_{i=1}^n a_i - np\right| < n^{3/4}\right\}. \tag{4.39}$$

Тогда из (4.38) и (4.39) имеем

$$\mathbf{P}(A_n) = \mathsf{P}\{\Xi_n \in A_n\} \leqslant \frac{pq}{\sqrt{n}} \to 0, n \to \infty\,. \tag{4.40}$$

Следовательно, выполняется (4.36).

В силу (4.39), если $a \in B_n$, то

$$np - n^{3/4} < \sum_{i=1}^{n} a_i < np + n^{3/4},$$

поэтому согласно (4.28)

$$n \log q - \log \frac{q}{p}\Big(np + n^{3/4}\Big) < \log p_n(a) < n \log q - \log \frac{q}{p}\Big(np - n^{3/4}\Big)$$

или

$$\begin{gathered} nq \log q + np \log p - n^{3/4} \log (q/p) < \log p_n(a) < \\ < nq \log q + np \log p + n^{3/4} \log (q/p), a \in B_n, \\ n(q \log q + p \log p) - n^{3/4} \log (q/p) < \log p_n(a) < \\ < n(q \log q + p \log p) + n^{3/4} \log (q/p)\,. \end{gathered} \tag{4.41}$$

Отсюда имеем

$$\forall\, a, a' \in B_n \,:\, \big|\log p_n(a) - \log p_n(a')\big| < 2n^{3/4} \log \frac{q}{p}\,, \tag{4.42}$$

$$-\log p_n(a) = |\log p_n(a)| > -n(p \log p + q \log q) - n^{3/4} \log \frac{q}{p}\,. \tag{4.43}$$

Из неравенств (4.42) и (4.43) следует

$$\begin{gathered} \left|\frac{\log p_n(a) - \log p_n(a')}{\log p_n(a)}\right| < \frac{2n^{3/4} \log (q/p)}{-n(p \log p + q \log q) - n^{3/4} \log (q/p)} = \\ = \frac{2n^{-1/4} \log (q/p)}{-p \log p - q \log q - n^{-1/4} \log (q/p)} \to 0,\ n \to \infty,\ a, a' \in B_n\,, \end{gathered}$$

что совпадает с (4.37). □

Теорема 4.5 (вторая теорема Шеннона об асимптотическом поведении мощности высоковероятностного множества). *Пусть $B_n \subset V_n$ — высоковероятное подмножество реализаций, определенное в теореме 4.4. Тогда мощность этого подмножества $M_n = |B_n|$ при $n \to \infty$ удовлетворяет асимптотике*

$$\frac{\log M_n}{n} \to \mathsf{H}\,\{\xi_1\} = -\,(p \log p + q \log q)\,. \tag{4.44}$$

Доказательство. Согласно (4.28) и (4.39)

$$\mathsf{P}\,\{\Xi_n \in B_n\} = 1 - \mathsf{P}\,\{\Xi \in A_n\} = \sum_{a \in B_n} p_n(a)\,. \tag{4.45}$$

В силу (4.40) из (4.45) имеем

$$1 \geqslant \mathsf{P}\,\{\Xi_n \in B_n\} \geqslant 1 - \frac{p\,q}{\sqrt{n}}\,. \tag{4.46}$$

С другой стороны, с учетом (4.41)

$$p_n(a) < 2^{n(p \log p + q \log q) + n^{3/4} \log (q/p)} = 2^{-n\mathsf{H}\{\xi_1\} + n^{3/4} \log (q/p)}.$$

Поэтому, используя второе равенство в (4.45), найдем

$$\mathsf{P}\{\Xi_n \in B_n\} < M_n 2^{-n\mathbf{H}\{\xi_1\}+n^{3/4}\log(q/p)}. \tag{4.47}$$

Из (4.46) и (4.47) имеем

$$M_n 2^{-n\mathbf{H}\{\xi_1\}+n^{3/4}\log(q/p)} > 1 - \frac{p\,q}{\sqrt{n}}.$$

Следовательно,

$$M_n > \left(1 - \frac{p\,q}{\sqrt{n}}\right) 2^{n\mathbf{H}\{\xi_1\}-n^{3/4}\log(q/p)}. \tag{4.48}$$

Аналогично (4.48) получим оценку снизу для M_n. Воспользуемся в (4.41) и (4.46) левыми неравенствами:

$$1 \geqslant M_n 2^{-n\mathbf{H}\{\xi_1\}-n^{3/4}\log(q/p)}.$$

Отсюда

$$M_n \leqslant 2^{n\mathbf{H}\{\xi_1\}+n^{3/4}\log(q/p)}. \tag{4.49}$$

Логарифмируя (4.48) и (4.49), выполняя деление на n и объединяя результаты в совместное неравенство, получим двустороннюю оценку для $\log M_n/n$:

$$\mathbf{H}\{\xi_1\} - n^{-1/4}\log(q/p) < \frac{\log M_n}{n} \leqslant \mathbf{H}\{\xi_1\} + n^{-1/4}\log(q/p).$$

Устремляя $n \to \infty$, придем к (4.44). □

Теоремы 4.4 и 4.5 легко обобщаются для стационарного ИДС без памяти, у которого алфавит A имеет мощность $N = |A| > 2$. При этом $\Xi_n \in A^n$ принимает одно из N^n возможных различных значений. Согласно теоремам 4.4 и 4.5 внимания заслуживают лишь $M_n \approx 2^{n\mathbf{H}\{\xi_1\}}$ реализаций, которые можно считать равновероятными. Если распределение вероятностей $p_1(a_1)$, $a_1 \in A$, является равномерным: $p_1(a_1) = \mathsf{P}\{\xi_1 = a_1\} = 1/N, a_1 \in A$, то $\mathbf{H}\{\xi_1\} = \log N$ (формула Хартли) и $2^{n\mathbf{H}\{\xi_1\}} = N^n$, т. е. $M_n = |A|^n$. Однако если распределение вероятностей отлично от равномерного, то $\mathbf{H}\{\xi_1\} < \log N$ и доля заслуживающих внимания реализаций

$$u_n = \frac{|B_n|}{|A|^n} = \frac{2^{n\mathbf{H}\{\xi_1\}}}{N^n} = \left(\frac{2^{\mathbf{H}\{\xi_1\}}}{N}\right)^n = \left(2^{\mathbf{H}\{\xi_1\}-\log N}\right)^n \tag{4.50}$$

неограниченно уменьшается с ростом n. Таким образом, подавляющее большинство реализаций при этом несущественные, и их можно отбросить. Этот факт лежит в основе теории кодирования сообщений и широко используется в криптологии.

Пример 4.1. Пусть рассматривается стационарный ИДС без памяти с двоичным алфавитом $A = V$ $(N = 2)$ и вероятностью появления единичного символа $0 \leqslant p \leqslant 1/2$. Тогда согласно теореме 4.4 «высоковероятное множество»

двоичных последовательностей имеет вид

$$B_n = \left\{ a = (a_1, \ldots, a_n) \in V_n \,:\, \left| \frac{1}{n} \sum_{i=1}^{n} a_i - p \right| < \frac{1}{n^{1/4}} \right\}.$$

Например, при $n = 10^4$ множество B_n состоит из двоичных последовательностей, для которых доля единиц заключена в следующих пределах: $p - 0{,}1 < \frac{1}{n} \sum_{i=1}^{n} a_i < p + 0{,}1$. При $p = 0{,}1$ количество единиц в таких двоичных последовательностях $0 \leqslant \sum_{i=1}^{n} a_i \leqslant 2000$. Доля этих «заслуживающих внимания» (по теореме 4.4) последовательностей согласно (4.50) составляет величину

$$u_n = \left(\frac{2^{-(p \log p + (1-p) \log(1-p))}}{2} \right)^n = (F(p))^n,$$

где $F(p) = 2^{-(p \log p + (1-p) \log(1-p) + 1)}$. Некоторые значения этой функции представлены в таблице.

p	0,1	0,2	0,3	0,4	0,5
$F(p)$	0,692	0,825	0,921	0,980	1,000

Например, при $p = 0{,}1$ и $n = 100$ доля «заслуживающих внимания» реализаций составляет $u_n = 2^{-0{,}531n} \approx 10^{-16}$.

4.3. ЭНТРОПИЙНАЯ УСТОЙЧИВОСТЬ СЛУЧАЙНЫХ СИМВОЛЬНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Установленные в предыдущем пункте асимптотические свойства разбиения множества реализаций для стационарных ИДС без памяти могут быть обобщены на класс зависимых случайных символьных последовательностей Ξ_n. При этом обобщении нам понадобится понятие *энтропийной устойчивости случайной последовательности.*

Случайная n-символьная последовательность $\Xi_n = (\xi_1, \ldots, \xi_n) \in A^n$ с n-мерным дискретным распределением вероятностей

$$p_n(a) = \mathbf{P}\{\Xi_n = a\}, a = (a_i) \in A^n \tag{4.51}$$

и неубывающей по n энтропией $0 < \mathbf{H}\{X\} < \infty$, где

$$\mathbf{H}\{X\} = \mathbf{E}\{-\log p_n(\Xi_n)\} = -\sum_{a \in A^n} p_n(a) \log p_n(a), \tag{4.52}$$

называется *энтропийно устойчивой*, если при увеличении числа символов ($n \to \infty$) имеет место сходимость по вероятности

$$\frac{-\log p_n(\Xi_n)}{\mathbf{H}\{X\}} \xrightarrow{\mathbf{P}} 1, \tag{4.53}$$

т. е. $\forall\varepsilon > 0\ \exists\bar{n} = \bar{n}(\varepsilon)$ такое, что при любом $n \geqslant \bar{n}(\varepsilon)$ выполняется неравенство

$$\mathsf{P}\left\{\left|\frac{-\log p_n(\Xi_n)}{\mathsf{H}\{X\}} - 1\right| \geqslant \varepsilon\right\} < \varepsilon; \tag{4.54}$$

при этом функция $\bar{n}(\varepsilon)$ — монотонно убывающая.

Отметим, что для реальных прикладных задач криптологии энтропия возрастает с ростом n: $\mathsf{H}\{X\} \to \infty$. Сформулируем важнейший результат, полученный в 1967 г. Р. Л. Стратоновичем.

Теорема 4.6 (обобщенная теорема Стратоновича). *Если $\Xi_n \in A^n$ — произвольная случайная n-символьная последовательность, удовлетворяющая свойству энтропийной устойчивости, то ее множество N^n реализаций A^n можно разбить на два непересекающихся подмножества A_n и B_n таким образом, что при $n \to \infty$ выполняются следующие асимптотические свойства:*

1) *суммарная вероятность реализаций подмножества A_n исчезающе мала:*

$$\mathsf{P}\{\Xi_n \in A_n\} = \sum_{a \in A_n} p_n(a) \to 0\,; \tag{4.55}$$

2) *реализации высоковероятного подмножества B_n становятся асимптотически относительно равновероятными в следующем смысле:*

$$\left|\frac{\log p_n(a) - \log p_n(a')}{\log p_n(a)}\right| \to 0,\ a, a' \in B_n\,; \tag{4.56}$$

3) *количество $M_n = |B_n|$ реализаций (мощность) множества B_n связано с энтропией (последовательности) $\mathsf{H}\{X\}$ асимптотическим соотношением*

$$\frac{\log M_n}{\mathsf{H}\{X\}} \to 1\,. \tag{4.57}$$

Доказательство. Оно состоит из трех частей.

1) В силу свойства энтропийной устойчивости (4.53) имеем неравенство (4.54), в котором для натурального m примем $\varepsilon = 1/m$. Тогда для любых $m = 1, 2, \ldots, n,\ n \geqslant \bar{n}(1/m)$ из (4.54) получим

$$\mathsf{P}\left\{\left|\frac{-\log p_n(\Xi_n)}{\mathsf{H}\{X\}} - 1\right| \geqslant \frac{1}{m}\right\} < \frac{1}{m}\,. \tag{4.58}$$

Согласно указанному выше свойству монотонности $\bar{n}(\varepsilon)$ выберем $m = m_n \in \mathbb{N}$ так, чтобы $\bar{n}(1/m_n) \leqslant n \leqslant \bar{n}(1/(m_n+1))$, и определим подмножества A_n и B_n, используя (4.58):

$$A_n = \{a = (a_i) \in A^n\ :\ |-\log p_n(a)/\mathsf{H}\{X\} - 1| \geqslant 1/m_n\},$$

$$B_n = A^n \setminus A_n\,. \tag{4.59}$$

С учетом (4.58) и (4.59) получим (4.55):

$$\mathsf{P}\{\Xi_n \in A_n\} < \frac{1}{m_n} \to 0,\ \mathsf{P}\{\Xi_n \in B_n\} > 1 - \frac{1}{m_n} \to 1,$$

поскольку очевидно, что $m_n \to \infty$ при $n \to \infty$.

2) Вследствие (4.59) $\forall a \in B_n$

$$\left(1-\frac{1}{m_n}\right)\mathbf{H}\{X\} < -\log p_n(a) < \left(1+\frac{1}{m_n}\right)\mathbf{H}\{X\} \tag{4.60}$$

или

$$1-\frac{1}{m_n} < \frac{-\log p_n(a)}{\mathbf{H}\{X\}} < 1+\frac{1}{m_n}\,. \tag{4.61}$$

В силу (4.60) $\forall a, a' \in B_n$

$$\left|\frac{\log p_n(a) - \log p_n(a')}{-\log p_n(a)}\right| \leqslant \frac{2\mathbf{H}\{X\}/m_n}{-\log p_n(a)}\,.$$

Кроме того, из 4.61 имеем

$$\frac{\mathbf{H}\{X\}}{-\log p_n(a)} < \frac{1}{1-1/m_n},$$

поэтому

$$\left|\frac{\log p_n(a) - \log p_n(a')}{-\log p_n(a)}\right| < \frac{2/m_n}{1-1/m_n} = \frac{2}{m_n-1} \to 0\,,$$

что доказывает (4.55).

3) Для доказательства (4.57) запишем (4.60) $\forall a \in B_n$ в эквивалентном виде:

$$2^{-(1+1/m_n)\mathbf{H}\{X\}} < p_n(a) < 2^{-(1-1/m_n)\mathbf{H}\{X\}}\,. \tag{4.62}$$

Поскольку

$$1 \geqslant \mathbf{P}\{\Xi_n \in B_n\} = \sum_{a\in Bn} p_n(a) \geqslant 1-\frac{1}{m_n}\,,$$

то с учетом (4.62) (как при доказательстве теоремы 4.5)

$$\left(1-\frac{1}{m_n}\right)2^{(1-1/m_n)\mathbf{H}\{X\}} < M_n < 2^{(1+1/m_n)\mathbf{H}\{X\}}\,.$$

Прологарифмировав и разделив части этого неравенства на $\mathbf{H}\{X\}$, получим

$$1-\frac{1}{m_n}+\frac{\log(1-1/m_n)}{\mathbf{H}\{X\}} < \frac{\log M_n}{\mathbf{H}\{X\}} < 1+\frac{1}{m_n}\,.$$

Поскольку $\mathbf{H}\{X\}$ с ростом n не убывает, а m_n неограниченно возрастает, то из этих неравенств при $n\to\infty$ следует (4.57). □

Заметим, что проверка основного условия теоремы 4.6 (энтропийной устойчивости (4.53)) на практике затруднительна. В связи с этим сформулируем ряд легко проверяемых достаточных условий, влекущих выполнение свойства (4.53).

Теорема 4.7. *Если существует равный нулю предел*

$$\lim_{n\to\infty}\frac{\mathbf{D}\{\log p_n(\Xi_n)\}}{(\mathbf{H}\{X\})^2} = 0\,, \tag{4.63}$$

то случайная символьная последовательность Ξ_n является энтропийно устойчивой.

Доказательство. Воспользуемся неравенством Чебышева (относительно дисперсий) для произвольной случайной величины $\xi \in \mathbb{R}^1$ и произвольного $\varepsilon > 0$:

$$\mathsf{P}\{|\xi - \mathsf{E}\{\xi\}| \geqslant \varepsilon\} \leqslant \frac{\mathsf{D}\{\xi\}}{\varepsilon^2}\,.$$

Полагая $\xi = -\log p_n(\Xi_n)/\mathsf{H}\{X\}$, используя (4.63) и учитывая $\mathsf{E}\{\xi\} = 1$ при $n \to \infty$, получим

$$\mathsf{P}\left\{\left|\frac{-\log p_n(\Xi_n)}{\mathsf{H}\{X\}} - 1\right| \geqslant \varepsilon\right\} \to 0\,,$$

что по определению сходимости по вероятности и означает (4.53). □

Замечание 4.1. Фигурирующая в (4.63) величина

$$\delta_n^2 = \frac{\mathsf{D}\{\log p_n(\Xi_n)\}}{(\mathsf{H}\{X\})^2} = \frac{\mathsf{D}\{-\log p_n(\Xi_n)\}}{(\mathsf{E}-\log p_n(\Xi_n))^2} \geqslant 0$$

есть квадрат коэффициента вариации случайной величины $-\log p_n(\Xi_n)$.

Таким образом, условие (4.63) означает стремление к нулю коэффициента вариации: $\delta_n \to 0$.

Теорема 4.8. *Если энтропия случайной символьной последовательности при $n \to \infty$ бесконечно возрастает ($\mathsf{H}\{X\} \to \infty$) и существует ограниченный верхний предел*

$$\overline{\lim_{n\to\infty}} \frac{\mathsf{D}\{\log p_n(\Xi_n)\}}{\mathsf{H}\{X\}} \leqslant C < +\infty\,, \tag{4.64}$$

то такая символьная последовательность Ξ_n энтропийно устойчива.

Доказательство. Из (4.64) следует, что $\forall \varepsilon > 0$ найдется такой номер $\bar{n} = \bar{n}(\varepsilon)$, что $\forall n \geqslant \bar{n}$ справедлива оценка сверху:

$$\frac{\mathsf{D}\{\log p_n(\Xi_n)\}}{(\mathsf{H}\{X\})^2} \leqslant \frac{C+\varepsilon}{\mathsf{H}\{X\}}\,.$$

Поскольку по условию $\mathsf{H}\{X\} \to \infty$, то при $n \to \infty$ правая часть этого неравенства стремится к нулю, поэтому выполняется (4.63). Применяя теорему 4.6, получим доказываемое. □

Теорема 4.9. *Если случайная символьная последовательность Ξ_n такая, что для нее существует положительная удельная энтропия*

$$h = \lim_{n\to\infty} \frac{\mathsf{H}\{X\}}{n} > 0 \tag{4.65}$$

и так называемая удельная дисперсия

$$d ::= \lim_{n\to\infty} \frac{\mathsf{D}\{\log p_n(\Xi_n)\}}{n} \geqslant 0,\ d < \infty\,, \tag{4.66}$$

то Ξ_n энтропийно устойчива.

Доказательство. Из (4.65) и (4.66) следуют асимптотические соотношения:

$$\mathsf{H}\{X\} = hn + o(n),\ \mathsf{D}\{\log p_n(\Xi_n)\} = dn + o(n)\,.$$

Поэтому при $n \to \infty$ существует конечный предел:

$$\frac{\mathbf{D}\{\log p_n(\Xi_n)\}}{\mathbf{H}\{X\}} = \frac{d + o(1)}{h + o(1)} \to \frac{d}{h} \geqslant 0\,.$$

В силу (4.64) и теоремы 4.8 получим доказываемое. □

Следствие 4.2. *Если последовательность Ξ_n состоит из n независимых в совокупности одинаково распределенных невырожденных случайных символов (т. е. порождается ДИБП), то она является энтропийно устойчивой.*

Доказательство. По условию следствия имеем $\log p_n(a) = \sum_{i=1}^{n} \log p_1(a_i)$, поэтому

$$\mathbf{H}\{X\} = \mathbf{E} - \log p_n(\Xi_n) = n\mathbf{H}\{\xi_1\}, 0 < \mathbf{H}\{\xi_1\} < \infty;$$

$$\mathbf{D}\{\log p_n(\Xi_n)\} = n\mathbf{D}\{\log p_1(\xi_1)\}, 0 < \mathbf{D}\{\log p_1(\xi_1)\} < \infty\,.$$

В результате согласно (4.65) и (4.66)

$$h = \mathbf{H}\{\xi_1\} > 0,\ d = \mathbf{D}\{\log p_1(\xi_1)\} < \infty\,,$$

а значит, выполнено третье достаточное условие энтропийной устойчивости, выражаемое теоремой 4.9. □

Замечание 4.2. Понятие энтропийной устойчивости можно ввести и для отдельной случайной величины ξ, если $\forall \varepsilon > 0$:

$$\mathbf{P}\left\{\left|\frac{-\log p_1(\xi)}{\mathbf{H}\{\xi\}} - 1\right| < \varepsilon\right\} > 1 - \delta\,.$$

В качестве иллюстрации применения теоремы 4.6 рассмотрим сжимающее кодирование сообщений дискретного источника. Построим такое взаимно однозначное отображение (кодирование) множества A^n сообщений длиной n в некоторое множество двоичных кодовых слов различной длины, чтобы средняя длина кодового слова была как можно меньше. Теория такого кодирования будет подробно изложена в гл. 7. Здесь будет рассмотрен пример кодирования на основе свойства асимптотической равнораспределенности.

Пусть $m = 2$, $A = \{0, 1\}$, $\varepsilon > 0$, $\delta > 0$, и выбрано такое подмножество B_n, что $\mathbf{P}\{\Xi \in B_n\} \geqslant 1 - \delta$.

Обозначим через d наименьшее целое число с условием $2^d \geqslant |B_n|$. Очевидно, что все последовательности из множества B_n можно пронумеровать d-разрядными двоичными числами. В качестве кодового слова $\varphi(a)$ для последовательности $a \in B_n$ возьмем ее d-разрядный номер, к которому слева приписан символ 0. Для последовательности $a \in A^n \backslash B_n$ в качестве кодового слова $\varphi(a)$ зададим саму последовательность a, к которой слева приписан символ 1. Очевидно, что по первому символу кодового слова $\varphi(a)$ можно однозначно определить, каким способом оно получено, и восстановить a.

Длину кодового слова $\varphi(a)$ обозначим через $|\varphi(a)|$. Оценим коэффициент сжатия μ_n, который по определению полагаем равным отношению средней дли-

ны кодового слова для случайной последовательности из A^n к n — длине исходной последовательности:

$$\mu_n = \frac{1}{n}\mathsf{E}\left\{|\varphi(a)|\right\}.$$

Из определения математического ожидания для дискретной случайной величины

$$\mu_n = \frac{1}{n}\left((d+1)\mathsf{P}\left\{\Xi \in B^n\right\} + (n+1)(1 - \mathsf{P}\left\{\Xi \in B^n\right\})\right).$$

Из свойства асимптотической равнораспределенности следует

$$n(h-\varepsilon) + \log(1-\delta) \leqslant d \leqslant n(h+\varepsilon) + 1,$$

поэтому

$$(1-\delta)\left(h - \varepsilon + \frac{\log_2(1-\delta)}{n}\right) + \frac{1}{n} \leqslant \mu_n \leqslant h + \varepsilon + \delta + \frac{2}{n}.$$

Из последнего неравенства видно, что при достаточно малых ε, δ и достаточно больших n коэффициент сжатия μ_n можно сделать сколь угодно близким к предельной энтропии h.

4.4. ИНФОРМАЦИОННАЯ ДИВЕРГЕНЦИЯ

Пусть на некотором множестве $\mathcal{X} \subseteq \mathbb{R}$ заданы две дискретные случайные величины ξ и η так, что известны их распределения $p(\cdot)$ и $q(\cdot)$ соответственно:

$$p(x) = \mathsf{P}\left\{\xi = x\right\} q(x) = \mathsf{P}\left\{\eta = x\right\}, x \in \mathcal{X}.$$

Определение 4.1. *Информационная дивергенция (расстояние Кульбака – Лейблера) — это несимметричная мера удаленности друг от друга двух вероятностных распределений $p()$ и $q()$, задаваемая соотношением*

$$D(p,q) = \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)}. \tag{4.67}$$

Следует заметить, что основание логарифма в приведенных формулах, подобно ситуации с измерением энтропии, определяет единицу измерения дивергенции.

Данная мера расстояния в теории информации может интерпретироваться как величина потерь информации при замене истинного распределения $p()$ на распределение $q()$.

Утверждение 4.1. Информационная дивергенция всегда неотрицательна $D(p,q) \geqslant 0$, причем $D(p,q) = 0$ тогда и только тогда, когда $p(x) = q(x)$, $\forall x \in \mathcal{X}$.

Доказательство. Преобразуем величину $-D(p,q)$ и воспользуемся неравенством Йенсена:

$$-D(p,q) = -\sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)} = \sum_{x\in\mathcal{X}} p(x)\log\frac{q(x)}{p(x)} \leqslant$$

$$\leqslant \log \sum_{x\in\mathcal{X}} q(x) = \log 1 = 0.$$

Равенство в данном неравенстве Йенсена для нелинейной логарифмической функции достигается тогда и только тогда, когда $q(x)/p(x) = \alpha$, $\forall x \in \mathcal{X}$. Отсюда из свойства нормировки распределений вероятностей

$$q(x) = \alpha p(x) \;\Rightarrow\; \sum_{x\in\mathcal{X}} q(x) = \alpha \sum_{x\in\mathcal{X}} p(x) \Rightarrow \alpha = 1. \qquad \square$$

В общем случае информационная дивергенция несимметрична, т. е. $D(p,q) \neq \neq D(q,p)$.

Упражнение 4.1. Привести пример таких распределений $p(\cdot)$ и $q(\cdot)$, чтобы для них информационная дивергенция была несимметричной.

Замечание 4.3. Понятие информационной дивергенции можно обобщить и на случай непрерывных распределений. В этом случае считается, что для случайных величин ξ и η известны их плотности распределения вероятностей $p()$ и $q()$ соответственно, а (4.67) перепишется следующим образом:

$$D(p,q) = \int\limits_{\mathcal{X}} p(x) \log \frac{p(x)}{q(x)} dx.$$

4.5. ЗАДАНИЯ ДЛЯ ТЕСТОВ

4.1. Удельная энтропия случайной последовательности ξ_t — это:

а) $\lim\limits_{n\to\infty} \mathbf{H}\{\xi_n|\Xi_{n-1}\}$;
б) $\lim\limits_{n\to\infty} (\mathbf{H}\{\Xi_n\} - \mathbf{H}\{\Xi_{n-1}\})$;
в) $\lim\limits_{n\to\infty} (\mathbf{H}\{\Xi_n\} - \mathbf{H}\{\xi_1\})$;
г) $\lim\limits_{n\to\infty} \dfrac{\mathbf{H}\{\Xi_n\}}{n}$;
д) $\lim\limits_{n\to\infty} \dfrac{\mathbf{H}\{\xi_n\}}{n}$.

4.2. Что не утверждается в первой теореме Шеннона об асимптотической равнораспределенности для стационарного ИДС без памяти:

а) $V_2^n = A_n \cup B_n$;
б) $A_n \cap B_n = \emptyset$;
в) $\mathbf{P}\{\Xi_n \in A_n\} \to 0$;
г) $\forall a,a' \in B_n \left|\dfrac{\log p_n(a) - \log p_n(a')}{\log p_n(a)}\right| \to 0$;
д) $M_n = |B_n|$, $\dfrac{\log M_n}{n} \to \mathbf{H}\{\xi_1\}$?

4.3. Мощность высоковероятностного множества $M_n = |B_n|$ удовлетворяет асимптотике:

а) $\log \dfrac{M_n}{n} \to \mathbf{H}\{\Xi_n\}$;
б) $\dfrac{\log M_n}{n} \to \mathbf{H}\{\xi_1\}$;
в) $\dfrac{M_n}{n} \to \mathbf{H}\{\Xi_n\}$;
г) $\log \dfrac{M_n}{n} \to \mathbf{H}\{\xi_1\}$;
д) $\dfrac{M_n}{\log n} \to \mathbf{H}\{\xi_1\}$.

4.4. Последовательность называется энтропийно устойчивой, если:

а) $\dfrac{-\log p_n(\Xi_n)}{\mathbf{H}\{\Xi_n\}} \xrightarrow{\mathbf{P}} 1$;

б) $\mathbf{H}\{\Xi_n\} = \text{const}$;

в) $h > 0$;

г) $\dfrac{-\log p_n(\xi_n)}{\mathbf{H}\{\Xi_n\}} \xrightarrow{\mathbf{P}} 1$;

д) $\dfrac{\mathbf{H}\{\xi_n|\Xi_{n-1}\}}{\mathbf{H}\{\Xi_n\}} \xrightarrow{\mathbf{P}} 1$.

4.5. На множестве $\mathcal{X}$ заданы два дискретных распределения вероятностей $p(\cdot)$ и $q(\cdot)$. Дивергенция — это:

а) $D(p,q) = \sum\limits_{x\in\mathcal{X}} p(x)\dfrac{p(x)}{q(x)}$;

б) $D(p,q) = \sum\limits_{x\in\mathcal{X}} p(x)\log p(x)$;

в) $D(p,q) = \sum\limits_{x\in\mathcal{X}} p(x)\log q(x)$;

г) $D(p,q) = \sum\limits_{x\in\mathcal{X}} (p(x)\log p(x) - p(x)\log q(x))$;

д) $D(p,q) = \sum\limits_{x\in\mathcal{X}} \dfrac{p(x)}{q(x)}$.

4.6. РЕШЕНИЕ ТИПОВЫХ ЗАДАЧ

Задача 4.1. Пусть дискретный временной ряд $x_t \in \{0,1\}$ таков, что для любого $t > 1$ имеет место соотношение $\mathbf{H}\{x_t|x_{t-1}, \ldots, x_1\} = \mathbf{H}\{x_t|x_{t-1}\} = \mathbf{H}\{x_2|x_1\}$. Кроме того, задано начальное распределение $\pi(x) = \mathbf{P}\{x_1 = x\}$ и условные распределения $p(y|x) = \mathbf{P}\{x_2 = y|x_1 = x\}$: $\pi(0) = 1/3$, $\pi(1) = 2/3$, $p(0|0) = 1/3$, $p(1|0) = 2/3$, $p(0|1) = 3/4$, $p(1|1) = 1/4$. Вычислить удельную энтропию дискретного временного ряда x_t.

Решение. Преобразуем энтропию n-символьной последовательности согласно условию задачи, используя свойство иерархической аддитивности:

$$\mathbf{H}\{x_1, \ldots, x_n\} = \mathbf{H}\{x_1\} + \mathbf{H}\{x_2|x_1\} + \mathbf{H}\{x_3|x_2, x_1\} + \ldots +$$

$$+ \mathbf{H}\{x_n|x_{n-1}, \ldots, x_1\} = \mathbf{H}\{x_1\} + (n-1)\mathbf{H}\{x_2|x_1\}.$$

Тогда для удельной энтропии получим соотношение

$$h = \lim_{n\to\infty} \frac{1}{n}\mathbf{H}\{x_1, \ldots, x_n\} = \lim_{n\to\infty} \frac{1}{n}\left(\mathbf{H}\{x_1\} + (n-1)\mathbf{H}\{x_2|x_1\}\right) = \mathbf{H}\{x_2|x_1\}.$$

Таким образом, для того чтобы найти удельную энтропию в данном случае, необходимо вычислить условную энтропию $\mathbf{H}\{x_2|x_1\}$. По определению условной

энтропии

$$\mathsf{H}\{x_2|x_1\} = -\sum_{x,y=0}^{1} p(x,y)\log p(y|x) = -\sum_{x,y=0}^{1} p(y|x)\pi(x)\log p(y|x).$$

Подставляя известные значения, получим ответ:

$$\mathsf{H}\{x_2|x_1\} = -\left(\frac{1}{3}\cdot\frac{1}{3}\log\frac{1}{3} + \frac{1}{3}\cdot\frac{2}{3}\log\frac{2}{3} + \frac{2}{3}\cdot\frac{3}{4}\log\frac{3}{4} + \frac{2}{3}\cdot\frac{1}{4}\log\frac{1}{4}\right).$$

Ответ: $h = -\left(\frac{1}{9}\log\frac{1}{3} + \frac{2}{9}\log\frac{2}{3} + \frac{1}{2}\log\frac{3}{4} + \frac{1}{6}\log\frac{1}{4}\right)$.

Задача 4.2. Пусть рассматривается стационарный ИДС без памяти с двоичным алфавитом и вероятностью появления единичного символа, равной 0,2. Согласно первой теореме Шеннона об асимптотической равнораспределенности описать множество относительно равновероятных двоичных последовательностей B_n. Найти отношение числа относительно равновероятных двоичных последовательностей к количеству всевозможных n-символьных реализаций при $n = 100$.

Решение. Согласно первой теореме Шеннона об асимптотической равнораспределенности множество относительно равновероятных двоичных последовательностей B_n имеет вид

$$B_n = \left\{a = (a_1, \ldots, a_n) \in V_n : \left|\frac{1}{n}\sum_{i=1}^{n} a_i - p\right| < \frac{1}{n^{1/4}}\right\},$$

т. е. B_n состоит из таких последовательностей, для которых доля единиц заключена в следующих пределах:

$$0{,}2 - \frac{1}{\sqrt{10}} < \frac{1}{100}\sum_{i=1}^{n} a_i < 0{,}2 + \frac{1}{\sqrt{10}},$$

$$0 \leqslant \sum_{i=1}^{n} a_i \leqslant 51.$$

Доля же этих «заслуживающих внимания» последовательностей может быть найдена по формуле

$$u_n = \frac{|B_n|}{2^n} = \left(\frac{2^{-p\log p-(1-p)\log(1-p)}}{2}\right)^n = 2^{-n(p\log p+(1-p)\log(1-p)+1)}.$$

Подставляя значения для p и n, найдем, что $u_{100} \approx 2^{-27}$.

Ответ: $B_n = \left\{a \in V_n : 0 \leqslant \sum\limits_{i=1}^{n} a_i \leqslant 51\right\}$, $u_{100} \approx 2^{-27}$.

4.7. ЗАДАЧИ И УПРАЖНЕНИЯ

4.1. Дискретный временной ряд $x_t \in \{0,1\}$ таков, что для любого $t > 1$ вероятность $\mathsf{P}\{x_t = 0\} = 1/2^t$. Вычислить удельную энтропию.

4.2. Дискретный временной ряд $x_t \in \{0,1\}$ таков, что для любого $t > 1$ имеет место соотношение $\mathsf{H}\{x_t|x_{t-1}, \ldots, x_1\} = \mathsf{H}\{x_t|x_{t-1}\} = \mathsf{H}\{x_2|x_1\}$. Кроме того, задано начальное распределение $\pi(x) = \mathsf{P}\{x_1 = x\}$ и условные распределения $p(y|x) = \mathsf{P}\{x_2 = y|x_1 = x\}$. Найти удельную энтропию дискретного временного ряда x_t, если:

а) $\pi(0) = 1/5$, $\pi(1) = 4/5$, $p(0|0) = 1/3$, $p(1|0) = 2/3$, $p(0|1) = 1/2$, $p(1|1) = 1/2$;

б) $\pi(0) = 1/2$, $\pi(1) = 1/2$, $p(0|0) = 1/5$, $p(1|0) = 4/5$, $p(0|1) = 4/7$, $p(1|1) = 3/7$;

в) $\pi(0) = 0{,}2$, $\pi(1) = 0{,}8$, $p(0|0) = 0{,}2$, $p(1|0) = 0{,}8$, $p(0|1) = 0{,}7$, $p(1|1) = 0{,}3$;

г) $\pi(0) = 0{,}4$, $\pi(1) = 0{,}6$, $p(0|0) = 0{,}75$, $p(1|0) = 0{,}25$, $p(0|1) = 0{,}5$, $p(1|1) = 0{,}55$.

4.3. Пусть для дискретного временного ряда $x_t \in \{0, \ldots, n-1\}$ заданы начальное распределение вероятностей $\pi(x) = \mathsf{P}\{x_1 = x\}$ и условные распределения $p(y|x) = \mathsf{P}\{x_2 = y|x_1 = x\}$, которые удобно представить в виде матрицы $\{P_{ij}\}$, $P_{ij} = p(x_j|x_i)$. Вычислить удельную энтропию, если:

а) $n = 2$, $\pi(0) = 1/4$, $\pi(1) = 3/4$, $P = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$;

б) $n = 2$, $\pi(0) = 1/2$, $\pi(1) = 1/2$, $P = \begin{pmatrix} 1/4 & 3/4 \\ 1/4 & 3/4 \end{pmatrix}$;

в) $n = 3$, $\pi(0) = 1/3$, $\pi(1) = 1/3$, $\pi(2) = 1/3$, $P = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 1/2 & 1/3 & 1/6 \\ 1/2 & 1/3 & 1/6 \end{pmatrix}$;

г) $\pi(0) = 1/n, \ldots, \pi(n-1) = 1/n$, $P = \begin{pmatrix} p_1 & p_2 & \cdots & p_n \\ p_1 & p_2 & \cdots & p_n \\ \vdots & \vdots & \ddots & \vdots \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}$, $p_1 + \ldots + p_n = 1$.

4.4. Рассматривается стационарный ИДС без памяти с двоичным алфавитом и вероятностью появления единичного символа, равной p. Согласно первой теореме Шеннона об асимптотической равнораспределенности опишите множество относительно равновероятных двоичных последовательностей B_n. Найти отношение числа относительно равновероятных двоичных последовательностей к количеству всевозможных n-символьных реализаций, если:

а) $p = 0{,}1$, $n = 10^4$;

б) $p = 0{,}3$, $n = 10^4$;

в) $p = 0{,}4$, $n = 10^2$;

г) $p = 0{,}25$, $n = 10^3$.

4.5. Для стационарного источника сообщений $x_t = \{x_1, x_2, \ldots\}$ доказать, что $\lim\limits_{n\to\infty} \frac{1}{n} \mathbf{I}\{(x_1, \ldots, x_n), (x_{n+1}, \ldots, x_{2n})\} = 0$.

4.6. Пусть $x_t = \{x_1, x_2, \ldots\}$ — стационарный источник сообщений, f и g — функции от одной и двух переменных соответственно, $y_t = f(x_t)$, $z_t = g(x_t, x_{t-1})$. Обозначим через h_x, h_y и h_z удельные энтропии источников x_t, $y_t = \{y_1, y_2, \ldots\}$ и $z_t = \{z_1, z_2, \ldots\}$ соответственно. Доказать неравенство $h_y \leqslant h_x$. Каково соотношение между h_z и h_x?

Глава 5

МАРКОВСКИЕ ИСТОЧНИКИ СООБЩЕНИЙ И ИХ СВОЙСТВА

5.1. ЦЕПЬ МАРКОВА И ЕЕ СВОЙСТВА

В теории информации широкое применение получил класс ДВР, обладающих марковскими свойствами: цепи Маркова с дискретным временем.

Определение 5.1. *ДВР $x_t \in A$, $t \in \mathbb{Z}$, определенный на вероятностном пространстве $(\Omega, \mathcal{F}, \mathsf{P})$, называется цепью Маркова (ЦМ) с пространством состояний A, если для любого $n \in \mathbb{N}$ и любых $t_1, \ldots, t_{n+1} \in \mathbb{Z}$ таких, что $t_1 < t_2 < \ldots < t_{n+1}$, выполняется марковское свойство*

$$\mathsf{P}\left\{x_{t_{n+1}} = i_{n+1} | x_{t_n} = i_n, \ldots, x_{t_1} = i_1\right\} = \mathsf{P}\left\{x_{t_{n+1}} = i_{n+1} | x_{t_n} = i_n\right\}, \\ i_1, \ldots, i_{n+1} \in A. \tag{5.1}$$

Соотношение (5.1) означает, что условное распределение вероятностей состояния $x_{t_{n+1}} \in A$ в будущий момент времени t_{n+1} при условии, что известна некоторая предыстория состояний процесса $\{x_{t_n} = i_n, \ldots, x_{t_1} = i_1\}$ в предыдущие моменты времени, зависит на самом деле не от всей этой предыстории, а лишь от состояния процесса $\{x_{t_n} = i_n\}$ в самый близкий к t_{n+1} прошлый момент времени t_n.

Пусть $t = 0$ — начальный момент времени; $p(0) = (p_0(0), \ldots, p_{N-1}(0))' \in \mathbb{R}^N$ — вектор-столбец *начального распределения вероятностей состояний* цепи Маркова x_t:

$$p_i(0) = P_1(i;\, 0) = \mathsf{P}\{x_0 = i\}\,,\ i \in A,\ \sum_{i \in A} p_i(0) = 1; \tag{5.2}$$

$P(t) = (p_{ij}(t)) \in \mathbb{R}^{N \times N}$ — *матрица вероятностей одношаговых переходов*

$$p_{ij}(t) = \mathsf{P}\{x_{t+1} = j | x_t = i\},\ i, j \in A, t \in \mathbf{N}_0 = \mathbf{N} \cup \{0\}. \tag{5.3}$$

Матрица $P(t)$ относится к классу стохастических матриц, так как обладает следующими свойствами:

$$p_{ij}(t) \in [0, 1],\ i, j \in A, \quad \sum_{j \in A} p_{ij}(t) = 1,\ i \in A\,.$$

Теорема 5.1. *Для любого $n \in \mathbf{N}_0$ $(n+1)$-мерное распределение вероятностей цепи Маркова x_t, $t \geqslant 0$, однозначно выражается через начальное распределение вероятностей $p(0)$ и матрицу вероятностей одношаговых переходов $P(0)$:*

$$P_{n+1}(i_0, i_1, \ldots, i_n; 0, 1, \ldots, n) = \mathbf{P}\{x_0 = i_0, x_1 = i_1, \ldots, x_n = i_n\} = \\ = p_{i_0}(0) \prod_{t=0}^{n-1} p_{i_t, i_{t+1}}(t). \tag{5.4}$$

Доказательство. Воспользуемся (2.3) и обобщенной формулой умножения вероятностей

$$P_{n+1}(i_0, \ldots, i_n; 0, \ldots, n) = \\ = \mathbf{P}\{x_0 = i_0\} \cdot \prod_{t=0}^{n-1} \mathbf{P}\{x_{t+1} = i_{t+1} | x_t = i_t, \ldots, x_0 = i_0\}.$$

Применяя к правой части марковское свойство (5.1) и обозначения (5.2), (5.3), придем к (5.4). □

Заметим, что, используя (5.4) и свойство согласованности из п. 2.1, легко получить m-мерные распределения для любой цепочки моментов времени $0 \leqslant \leqslant t_1 < t_2 < \ldots < t_m$.

Определение 5.2. *Если матрица вероятностей одношаговых переходов $P(t)$, задаваемая (5.3), не зависит от времени $t \in \mathbb{Z}$,*

$$P(t) \equiv P = (p_{ij}) \in \mathbb{R}^{N \times N},$$

где P — некоторая стохастическая матрица, то x_t называется однородной цепью Маркова (ОЦМ), в противном случае — неоднородной.

Обозначим по аналогии с (5.2), (5.3) распределение вероятностей состояний ОЦМ в момент времени $t \in \mathbb{Z}$:

$$p(t) = (p_0(t), \ldots, p_{N-1}(t))', p_i(t) = P_1(i; t) = \mathbf{P}\{x_t = i\}, i \in A;$$

матрицу вероятностей s-шаговых переходов ($s \in \mathbb{N}$):

$$P^{(s)} = \left(p_{ij}^{(s)}\right), p_{ij}^{(s)} = \mathbf{P}\{x_{t+s} = j | x_t = i\}, \ i, j \in A.$$

Следствие 5.1. *Для ОЦМ x_t справедливы следующие формулы:*

$$P_{n+1}(i_0, i_1, \ldots, i_n; 0, 1, \ldots, n) = p_{i_0}(0) \prod_{t=0}^{n-1} p_{i_t, i_{t+1}};$$

$$P^{(s+m)} = P^{(s)} \cdot P^{(m)}, s, m \in \mathbb{N} \text{ (формула Колмогорова – Чепмена)};$$

$$P^{(t)} = P^t, p(t) = (P^t)' p(0).$$

Введем ряд необходимых понятий по классификации состояний ОЦМ [30].

Определение 5.3. *Состояние $i \in A$ называется несущественным, если найдется состояние $j \neq i$ и $s \in \mathbb{N}$ такие, что $p_{ij}^{(s)} > 0$, но $p_{ji}^{(t)} = 0$ для любого $t \in \mathbb{N}$; в противном случае состояние называется существенным.*

Определение 5.4. *Состояние $j \in A$ называется достижимым из состояния $i \in A$, если существует $s \in \mathbf{N}_0$ такое, что $p_{ij}^{(s)} > 0$ (обозначается $i \to j$); если $i \to j$, а $j \to i$, то состояния i, j называются сообщающимися (обозначаются $i \leftrightarrow j$). Бинарное отношение $i \leftrightarrow j$ разбивает множество всех существенных состояний на непересекающиеся неразложимые классы сообщающихся состояний: $S_1, \ldots, S_L \subset A$; если некоторый класс S_l включает в себя единственное состояние $a^* \in A$, то это состояние называется поглощающим (при попадании в него ОЦМ навсегда остается в этом состоянии).*

Определение 5.5. *ОЦМ, множество состояний которой A образует один класс существенных сообщающихся состояний, называется неразложимой.*

Определение 5.6. *Пусть $d_i = НОД\left\{n \in \mathbb{N} : p_{ii}^{(n)} > 0\right\}$. Если $d_i > 1$, то состояние $i \in A$ называется периодическим с периодом d_i; если $d_i = 1$, то состояние $i \in A$ называется непериодическим.*

Обозначим

$$q_i(n) = \mathbf{P}\left\{x_n = i, x_{n-1} \neq i, \ldots, x_1 \neq i | x_0 = i\right\}, \; i \in A; \; P_i = \sum_{m=1}^{+\infty} q_i(n),$$

где $q_i(n)$ — вероятность события, состоящая в том, что ОЦМ x_t, выйдя из начального состояния i, впервые вернется в него на n-м шаге; P_i — вероятность того, что ОЦМ, выйдя из состояния i, когда-нибудь в него вернется.

Определение 5.7. *Состояние $i \in A$ называется возвратным, если $P_i = 1$, в противном случае — невозвратным.*

Теорема 5.2 (критерий возвратности). *Состояние $i \in A$ возвратно тогда и только тогда, когда расходится ряд*

$$Q_i = \sum_{n=1}^{+\infty} p_{ii}^{(n)} = +\infty.$$

Для невозвратного состояния i вероятность возврата $P_i = Q_i/(1 + Q_i)$.

Доказательство. Достаточно воспользоваться формулой полной вероятности и построить производящие функции для $\left\{p_{ii}^{(n)} : n \in \mathbb{N}\right\}$ и $\{p_i(n) : n \in \mathbb{N}\}$ [30]. □

Определение 5.8. *Обозначим $\mu_i = \sum_{n=1}^{+\infty} n p_i(n) \geqslant 1$ — среднее время возвращения в i-е состояние ОЦМ, начавшей свое движение из i-го начального состояния. Состояние $i \in A$ называется положительным, если $\mu_i^{-1} > 0$, т. е. среднее время возвращения конечно: $\mu_i < +\infty$, и нулевым, если $\mu_i^{-1} = 0$ ($\mu_i = +\infty$).*

Теорема 5.3 (теорема солидарности). *В неприводимой ОЦМ все состояния солидарно обладают одним и тем же свойством: если хотя бы одно состояние возвратно, то и все возвратны; если хотя бы одно периодично с некоторым периодом d, то и все периодичны с периодом d; если хотя бы одно со-*

стояние непериодично (т. е. $d = 1$), то и все состояния непериодичны (при этом ОЦМ называется непериодической).

Теорема 5.4 (теорема о циклических подклассах). *Если ОЦМ x_t — неприводимая и периодическая с некоторым периодом $d > 1$, то множество A ее состояний разбивается на d циклических подклассов: $D_0, D_1, \ldots, D_{d-1}$ таких, что с вероятностью единица за один шаг ОЦМ переходит из некоторого класса D_k $(k < d-1)$ в класс D_{k+1}, а из класса D_{d-1} — в D_0.*

Доказательство теорем 5.3, 5.4 приведено в [30].

Подводя итог, представим классификацию состояний цепи Маркова графически (рис. 5.1, 5.2) [35].

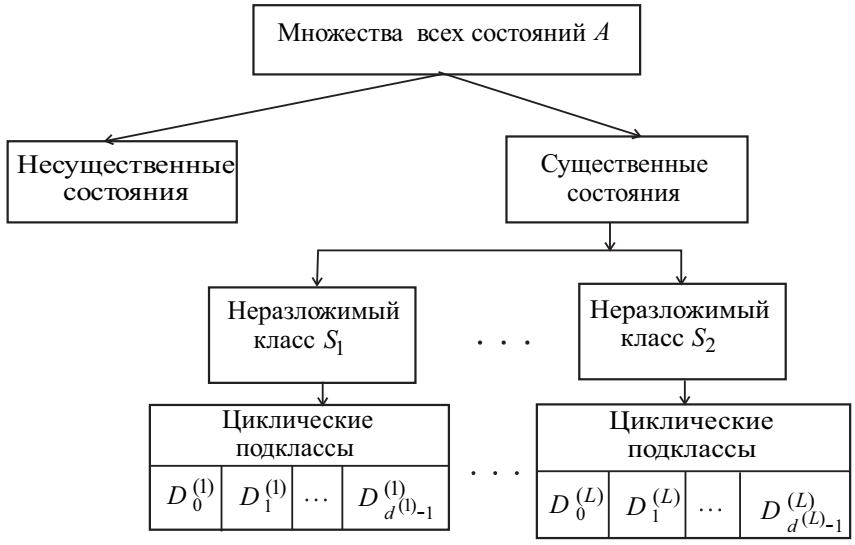


Рис. 5.1. Классификация состояний ОЦМ по арифметическим свойствам вероятностей $\left\{p_{ij}^{(n)}\right\}$

Определение 5.9. *ОЦМ x_t, $t \in \mathbf{N}_0$, называется эргодической, если для любых $i, j \in A$ существуют независимые от i положительные пределы*

$$\lim_{n\to+\infty} p_{ij}^{(n)} = \pi_j^* > 0,$$

при этом вектор-столбец $\pi^* = \left(\pi_0^*, \pi_1^*, \ldots, \pi_{N-1}^*\right)'$ называется стационарным распределением вероятностей ОЦМ.

Стационарное распределение вероятностей ОЦМ x_t является единственным решением системы линейных алгебраических уравнений

$$\begin{cases} P'\pi = \pi, \\ \sum\limits_{i=0}^{N-1} \pi_i = 1. \end{cases} \tag{5.5}$$

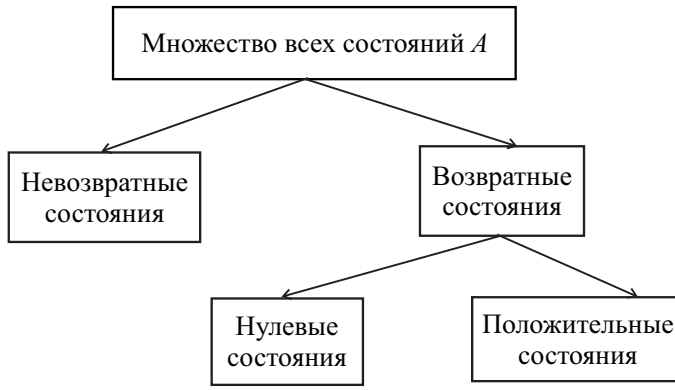


Рис. 5.2. Классификация состояний ОЦМ по асимптотическим свойствам вероятностей $\left\{p_{ii}^{(n)}\right\}$

Теорема 5.5 (критерий эргодичности). *Свойство эргодичности ОЦМ x_t эквивалентно каждому из следующих утверждений* [35]*:*

1) *ОЦМ x_t неразложима и непериодическая;*

2) *для любых $i, j \in A$ существуют положительные пределы*

$$\lim_{n\to+\infty} p_{ij}^{(n)} = \frac{1}{\mu_j} > 0,$$

где μ_j задано в определении 5.8;

3) *найдется такое достаточно большое число шагов n_0, что для всех $n \geqslant n_0$ все элементы матрицы P^n положительны:*

$$\min_{i,\,j\ \in A} p_{ij}^{(n)} > 0.$$

Теорема 5.6. *Пусть для ОЦМ x_t, $t \in \mathbf{N}_0$, выполнено условие эргодичности и $\pi^* = (\pi_i^*)$ — стационарное распределение вероятностей, удовлетворяющее (5.5). Тогда если начальное распределение вероятностей совпадает со стационарным*

$$p(0) = \pi^*, \tag{5.6}$$

то справедливы два результата:

1) *распределение вероятностей $p(t) \equiv \pi^*$ и не зависит от времени t;*

2) *ОЦМ x_t является стационарным в узком смысле ДВР.*

Доказательство. В силу последнего соотношения следствия 5.1, равенств (5.5) и (5.6) придем к справедливости первого утверждения

$$p(t) = (P^t)'P(0) = (P^t)'\pi^* = (P^{t-1})'P'\pi^* = (P^{t-1})'\pi^* = \ \ldots\ = \pi^*.$$

Для доказательства второго утверждения достаточно воспользоваться определением 2.7 и аналогично (5.4) вычислить $(n+1)$-мерное распределение для ОЦМ

x_t при произвольном сдвиге времени $\tau \in \mathbb{N}$ с учетом только что доказанного результата:

$$\begin{aligned}
&\mathsf{P}\{x_\tau = i_0, x_{\tau+1} = i_1, \ldots, x_{\tau+n} = i_n\} = \\
&= \mathsf{P}\{x_\tau = i_0\} \cdot \prod_{t=0}^{n-1} \mathsf{P}\{x_{\tau+t+1} = i_{t+1} | x_{\tau+t} = i_t\} = \\
&= \pi^*_{i_0} \prod_{t=0}^{n-1} p_{i_t, i_{t+1}} = \mathsf{P}\{x_0 = i_0, x_1 = i_1, \ldots, x_n = i_n\}.
\end{aligned}$$

Полученное равенство и означает стационарность в узком смысле. □

Заметим, что в силу определения 5.9 существует предел $\lim\limits_{t\to+\infty} P^t = \Pi^*$, где $\Pi^* = \left(\pi^* \vdots \pi^* \vdots \ldots \vdots \pi^*\right)'$ — $(N \times N)$-матрица, все строки которой одинаковы и совпадают с $\pi^{*\prime}$. Поэтому из следствия 5.1 заключим, что при любом начальном распределение $p(0)$ имеет место сходимость (с экспоненциальной скоростью):

$$p(t) \to \pi^* \text{ при } t \to +\infty.$$

Следовательно, каково бы ни было начальное распределение вероятностей $p(0)$, если исключить из рассмотрения начальный фрагмент $\{x_0, x_1, \ldots, x_T\}$, то при достаточно большом T (которое можно легко оценить) ОЦМ $\{x_{T+1}, x_{T+2}, \ldots\}$ мало отличается от стационарного ДВР.

Отметим еще, что методы компьютерного моделирования цепей Маркова представлены в [32, 37].

В заключение приведем некоторые полезные формулы расчета важнейших вероятностных характеристик для эргодических ОЦМ [10]. Примем следующие обозначения: $D = \operatorname{diag}\{1/\pi^*_0, \ldots, 1/\pi^*_{N-1}\}$ — диагональная $(N\times N)$-матрица; $\mathbf{I}_N$ — единичная $(N\times N)$-матрица; $Z = (z_{ij}) = (\mathbf{I}_N - P + \Pi^*)^{-1}$ — так называемая фундаментальная $(N \times N)$-матрица; Z_{diag} — матрица, полученная из Z заменой всех внедиагональных элементов нулями; $\mathbb{1}_N$ — $(N \times 1)$-вектор-столбец, все элементы которого равны 1; $\mathbb{1}_{N\times N}$ — $(N \times N)$-матрица, все элементы которой равны 1; $\bar{P} = (\bar{p}_{ij})$ — матрица вероятностей одношаговых переходов для обращенной стационарной цепи Маркова, т. е. для ОЦМ, наблюдаемой в обратном времени; $M = (m_{ij})$, m_{ij} — математическое ожидание случайного времени 1-го достижения состояния j, исходя из состояния i $(i, j \in A)$; $W = (w_{ij})$, w_{ij} — дисперсия случайного времени 1-го достижения j из i; $y_i^{(n)}$ — суммарное случайное время пребывания в состоянии i за первые n шагов; $C = (c_{ij})$,

$$c_{ij} = \lim_{n\to+\infty} n^{-1} \mathsf{Cov}\left\{y_i^{(n)}, y_j^{(n)}\right\},\ i, j \in A.$$

Справедливы следующие полезные соотношения [10]:

$$\bar{P} = DP'D^{-1},$$
$$M = (\mathbf{I}_N - Z + \mathbb{1}_{N\times N} Z_{\mathrm{diag}})\, D, \quad \bar{M} = M - M_{\mathrm{diag}},$$
$$W = M\,(2Z_{\mathrm{diag}} D - \mathbf{I}_N) + 2\,(ZM - \mathbb{1}_{N\times N}(ZM)_{\mathrm{diag}})\,,$$
$$c_{ij} = \pi_i^* z_{ij} + \pi_j^* z_{ji} - \pi_i^* \delta_{ij} - \pi_i^* \pi_j^*,\ i, j \in A,$$
$$m_{ii} = 1/\pi_i^*,\, w_{ii} = 2z_{ii}(\pi_i^*)^{-2} = (\pi_i^*)^{-1},$$
$$\pi_0^{*\prime} M = \mathbb{1}_N' Z_{\mathrm{diag}} D;\ M\pi^* = C\mathbb{1}_N,$$
$$P = \mathbf{I}_N + (D - \mathbb{1}_{N\times N})\,(\bar{M})^{-1}.$$

В работе [10] также приводятся полезные формулы вычисления вероятностных характеристик для ОЦМ при наличии поглощающих состояний.

5.2. ЦЕПЬ МАРКОВА ПОРЯДКА s

В теории информации для моделирования ДВР с глубиной памяти $s \in \mathbb{N}$ используется *цепь Маркова порядка* s (ЦМ(s)), обобщающая модель простой цепи Маркова из разд. 5.1.

Определение 5.10. *Цепь Маркова* $x_t \in A$, $t \in \mathbb{N}$, *порядка* s *с пространством состояний* A, *определенная на вероятностном пространстве* $(\Omega, \mathcal{F}, \mathbf{P})$ *и временной области* $\mathbb{N}$, *характеризуется обобщенным марковским свойством* [7]*:*

$$\begin{aligned} &\mathbf{P}\,\{x_{t+1} = i_{t+1} | x_t = i_t,\, \ldots, x_1 = i_1\} = \\ &= \mathbf{P}\,\{x_{t+1} = i_{t+1} | x_t = i_t,\, \ldots, x_{t-s+1} = i_{t-s+1}\} = \\ &= p_{i_{t-s+1}, \ldots, i_t, i_{t+1}}(t), t \geqslant s, i_1, \ldots, i_{t+1} \in A. \end{aligned} \tag{5.7}$$

Соотношение (5.7) означает, что условное распределение вероятностей будущих состояний $x_{t+1} \in A$ при фиксированной предыстории зависит не от всей этой предыстории, а лишь от ближайшей на глубину s предыстории $(x_t, \ldots, x_{t-s+1}) \in A^s$. Если $s = 1$, то (5.7) эквивалентно соотношению (5.1) и ЦМ(1) называют *простой цепью Маркова*. Если $s = 0$, то ЦМ(0) является схемой независимых испытаний.

Цепь Маркова ЦМ(s) характеризуется s-мерным начальным распределением вероятностей

$$\pi_{i_1, \ldots, i_s} = \mathbf{P}\,\{x_1 = i_1, \ldots, x_s = i_s\}\,, i_1, \ldots, i_s \in A,$$

и $(s+1)$-мерной матрицей вероятностей одношаговых переходов в момент времени $t \geqslant s$:

$$\begin{aligned} P(t) &= \left(p_{i_1, \ldots, i_{s+1}}(t)\right),\ p_{i_1, \ldots, i_{s+1}}(t) = \\ &= \mathbf{P}\,\{x_{t+1} = i_{t+1} | x_t = i_t,\, \ldots, x_{t-s+1} = i_{t-s+1}\}\,,\ i_1, \ldots, i_{s+1} \in A\,. \end{aligned} \tag{5.8}$$

Если $P(t)$ не зависит от времени t: $P(t) = P = (p_{i_1,\ldots,i_{s+1}})$, $i_1, \ldots, i_{s+1} \in A$, то имеем однородную цепь Маркова s-го порядка (ОЦМ(s)). Матрица P удовлетворяет условиям нормировки

$$\sum_{i_{s+1}\in A} p_{i_1,\ldots,i_{s+1}} \equiv 1.$$

Оказывается, расширением пространства состояний ОЦМ(s) можно привести к простой цепи Маркова. Примем обозначения: $X_t = (x_{t-s+1}, \ldots, x_t) \in A^s$ — s-предыстория к моменту t ($t \geqslant s$); $< X >= x_{t-s+1}N^{s-1} + \ldots + x_{t-1}N + x_t \in \in \{0, 1, \ldots, N^s - 1\}$ — число, N-ичная запись которого есть X_t.

Теорема 5.7. *Если x_t, $t \geqslant 1$, есть ОЦМ(s) с матрицей вероятностей одношаговых переходов $P = (p_{i_1,\ldots,i_{s+1}})$, $i_1, \ldots, i_{s+1} \in A$, то ДВР $y_t =< X_t >$ — простая цепь Маркова с пространством состояний $\mathbf{B} = \{0, 1, \ldots, N^s - 1\}$ и матрицей вероятностей одношаговых переходов $P^X = \left(p^X_{<I>,<J>}\right)$, $I = (i_0, \ldots, i_{s-1})$, $J = (j_0, \ldots, j_{s-1}) \in A^s$:*

$$p^X_{<I>,<J>} = \begin{cases} p_{i_0,\ldots,i_{s-1},j_{s-1}}, \text{ если } j_0 = i_1, j_1 = i_2, \ldots, j_{s-2} = i_{s-1}, \\ 0 \text{ в противном случае.} \end{cases} \tag{5.9}$$

Доказательство. Достаточно проверить марковское свойство первого порядка для ДВР X_t с использованием (5.8) и принятых обозначений. □

Теорема 5.7 позволяет перенести всю теорию простых цепей Маркова из разд. 5.1 на ОЦМ(s). В частности, получим один из критериев эргодичности ОЦМ(s).

Следствие 5.2. *ОЦМ(s) эргодична тогда и только тогда, когда найдется такое число $n_0 \in \mathbb{N}$, что для любого $n \geqslant n_0$*

$$\min_{I,\,J\in A^s} \left(\left(P^X\right)^n\right)_{<I>,<J>} > 0,$$

где матрица P^X определяется (5.9).

В заключение отметим, что число независимых параметров, определяющих матрицу вероятностей одношаговых переходов $P = (p_{i_1,\ldots,i_{s+1}})$ для ОЦМ(s), равно $D_{\text{ОЦМ}(s)} = N^s(N-1) = Q\left(N^{s+1}\right)$.

При увеличении глубины памяти s число параметров экспоненциально возрастает согласно таблице. Для идентификации такой модели требуется наблюдать реализацию $x_1, x_2, \ldots, x_n$ не всегда доступной на практике длительности $n > D_{\text{ОЦМ}(s)}$.

s	1	2	8	16	32	64	128	256
$D_{\text{ОЦМ}(s)}$	2	4	256	$6{,}4 \cdot 10^4$	$4{,}3 \cdot 10^9$	$1{,}8 \cdot 10^{19}$	$3{,}4 \cdot 10^{38}$	$1{,}2 \cdot 10^{77}$

В связи с этим в криптологии начинают активно использоваться так называемые *малопараметрические модели цепей Маркова высокого порядка* [1, 27, 31], для которых матрица P задается числом параметров $D \ll D_{\text{ОЦМ}(s)}$. Некоторые

из таких малопараметрических моделей рассматриваются в следующих разделах данной главы.

5.3. МОДЕЛЬ ДЖЕКОБСА – ЛЬЮИСА

Определение 5.11. *Модель Джекобса – Льюиса* [42] *для ДВР* x_t, $t \geqslant 1$, *определяется стохастическим разностным уровнем порядка* $s \geqslant 2$ *со случайным запаздыванием:*

$$x_t = \mu_t x_{t-\eta_t} + (1-\mu_t)\xi_t,\ t > s, \tag{5.10}$$

где $\{x_1, \ldots, x_s\}$, $\{\xi_t, \eta_t, \mu_t : t > s\}$ — *независимые в совокупности дискретные случайные величины на* $(\Omega, \mathcal{F}, \mathbf{P})$ *с распределениями вероятностей*

$$\mathbf{P}\{\xi_t = i\} = \pi_i, i \in A, \sum_{\substack{i\in A\\ s}} \pi_i = 1,$$

$$\mathbf{P}\{\eta_t = j\} = \lambda_j, j \in \{1, \ldots, s\}, \sum_{j=1} \lambda_j = 1, \lambda_s \neq 0,$$

$$\mathbf{P}\{\mu_t = 1\} = 1 - \mathbf{P}\{\mu_t = 0\} = \rho,$$

$$\mathbf{P}\{x_k = i\} = \pi_i,\ i \in A, k \in \{1, \ldots, s\}. \tag{5.11}$$

Модель (5.10), (5.11) дает вероятностное описание криптографическому генератору случайной последовательности x_t, схема которого представлена на рис. 5.3.

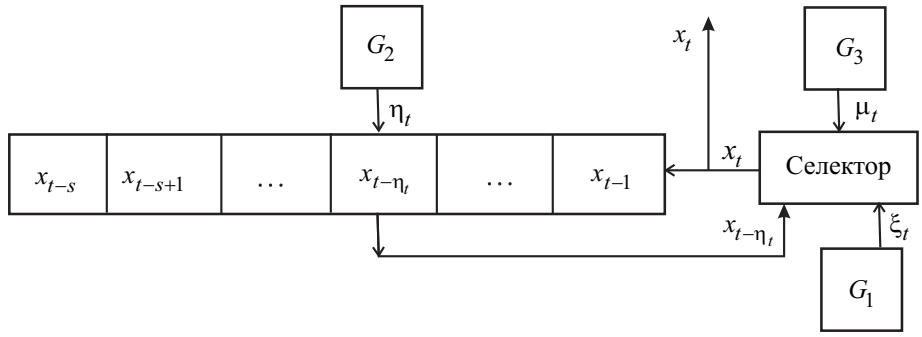


Рис. 5.3. Криптографический генератор последовательности x_t

Генератор состоит из трех простейших генераторов G_1 (двоичной последовательности ξ_t), G_2 (последовательности η_t), G_3 (двоичной последовательности μ_t), регистра сдвига, на каждом такте сдвигающего содержимое s своих ячеек на одну позицию влево, теряя последний бит x_{t-s}, и селектора, выбирающего один из своих входных сигналов, $x_{t-\eta_t}$ или ξ_t, в зависимости от значения μ_t.

Теорема 5.8. *ДВР* x_t, *определяемый* (5.10), (5.11), *является ОЦМ*(s) *с начальным распределением вероятностей* $\pi_{i_1,\ldots,i_s} = \pi_{i_1} \cdots \pi_{i_s}$ *и* $(s+1)$*-мерной*

матрицей вероятностей одношаговых переходов $P(\pi, \lambda, \rho) = (p_{i_1,\ldots,i_{s+1}})$[26]:

$$p_{i_1,\ldots,i_{s+1}} = (1-\rho)\pi_{i_{s+1}} + \rho \sum_{j=1}^{s} \lambda_j \delta_{i_{s-j+1},i_{s+1}},\ i_1, \ldots, i_{s+1} \in A. \tag{5.12}$$

Модель (5.10) предложена английскими статистиками П. Джекобсом и П. Льюисом в 1978 г. Малопараметрическое представление (5.12) $(s+1)$-мерной матрицы вероятностей одношаговых переходов для модели Джекобса – Льюиса характеризуется числом параметров

$$D_{\mathrm{JL}} = N + s - 1.$$

Очевидно, что вместо экспоненциальной зависимости числа параметров для общей модели ЦМ(s) ($D_{\mathrm{ОЦМ}(s)} = N^s(N-1)$) для модели Джекобса – Льюиса число параметров D_{JL} линейно зависит от s. Это создает существенный выигрыш в вычислительной сложности алгоритмов идентификации модели.

5.4. MTD-МОДЕЛЬ РАФТЕРИ

Эта модель предложена американским статистиком А. Рафтери в 1985 г. [44].

Определение 5.12. *MTD-модель (от англ. mixture transition distribution) Рафтери для ДВР* x_t, $t \geqslant 1$, *задается следующим малопараметрическим представлением* $(s+1)$*-мерной матрицы вероятностей одношаговых переходов* $P = (p_{i_1,\ldots,i_{s+1}})$ *на основе смеси вероятностных распределений:*

$$p_{i_1,\ldots,i_{s+1}} = \sum_{j=1}^{s} \lambda_j \cdot q_{i_j,i_{s+1}}, i_1, \ldots, i_{s+1} \in A, \tag{5.13}$$

где $Q = (q_{i,k})$ *— двумерная стохастическая* $(N \times N)$*-матрица,* $i, k \in A$; $\lambda = (\lambda_1, \ldots, \lambda_s)'$ *— s-вектор-столбец дискретного распределения вероятностей смеси, для которого* $\lambda_1 > 0$, $\lambda_2, \ldots, \lambda_s \geqslant 0$, $\lambda_1 + \ldots + \lambda_s = 1$.

Эта модель имеет $D_{\mathrm{MTD}} = N(N-1)/2 + s - 1$ параметров. Обобщением модели (5.13) является MTDg-модель [44], в которой для каждого из s прошлых моментов времени используется «своя» матрица вероятностей переходов

$$p_{i_1,\ldots,i_{s+1}} = \sum_{j=1}^{s} \lambda_j q^{(j)}_{i_j,i_{s+1}},\ i_1, \ldots, i_{s+1} \in A, \tag{5.14}$$

где $Q^{(j)} = \left(q^{(j)}_{i,k}\right)$ — j-я двумерная стохастическая матрица, соответствующая запаздыванию $s-j$, при этом число параметров $D_{\mathrm{MTDg}} = s(N(N-1)/2+1)-1$.

Приведем две теоремы о вероятностных свойствах моделей (5.13), (5.14) [26].

Теорема 5.9. *Для того чтобы ОЦМ*(s)*, определяемая моделью* (*5.13*)*, была эргодической, необходимо и достаточно, чтобы существовало число* $n_0 \in \mathbf{N}_0$ *такое, что при любом* $n \geqslant n_0$ *все элементы матрицы* Q^n *положительны.*

Примем обозначения: $\Pi^* = \left(\pi^*_{i_1,\ldots,i_s}\right)$, $i_1, \ldots, i_s \in A$, — s-мерное стационарное распределение вероятностей эргодической цепи; $\pi^* = (\pi^*_0, \ldots,$

$\pi^*_{N-1}\big)'$ — одномерное (маргинальное) стационарное распределение вероятностей.

Теорема 5.10. *Если x_t — эргодическая ОЦМ(s), удовлетворяющая модели (5.14), то для ее стационарного распределения вероятностей имеет место соотношение*

$$\pi^*_{i_1,\,\ldots,\,i_s} = \prod_{l=0}^{s-1}\left(\pi^*_{i_{s-1}} + \sum_{j=l+1}^{s}\lambda_j\left(q^{(j)}_{i_{j-l},\,i_{s-l}} - \sum_{r=0}^{N-1} q^{(j)}_{r,\,i_{s-l}}\pi^*_r\right)\right).$$

Следствие 5.3. *Если для модели (5.13) выполнено условие эргодичности, то для стационарного двумерного маргинального случайного вектора $(x_{t-m}, x_t)' \in A^2$, $1 \leqslant m \leqslant s$, справедливо соотношение*

$$\pi^*_{k,\,i}(m) = \pi^*_k\pi^*_i + \pi^*_k\lambda_{s-m+1}\,(q_{k,i} - \pi^*_i),\ i,k \in A.$$

5.5. ЦЕПЬ МАРКОВА С ЧАСТИЧНЫМИ СВЯЗЯМИ ЦМ(s, r)

Эта модель разработана в Белорусском государственном университете в 2003 г. [29]. Пусть, как и в разд. 5.2, x_t — однородная ЦМ s-го порядка на $(\Omega, \mathcal{F}, \mathbf{P})$ с некоторой $(s+1)$-мерной матрицей вероятностей переходов $P = (p_{i_1,\,\ldots,\,i_{s+1}})$, $i_1, \ldots, i_{s+1} \in A$; $r \in \{1, \ldots, s\}$ — параметр, который называется числом связей; $M^0_r = \big(m^0_1, \ldots, m^0_r\big) \in \mathrm{M}$ — произвольный целочисленный r-вектор с упорядоченными компонентами $1 = m^0_1 < m^0_2 < \ldots < m^0_r \leqslant s$, который называется шаблоном связей; M — множество всевозможных таких векторов с r компонентами, имеющее мощность $\mathrm{K} = |\mathrm{M}| = C^{r-1}_{s-1}$; $Q^0 = (q^0_{j_1,\,\ldots,\,j_r,\,j_{r+1}})$, $j_1, \ldots, j_{s+1} \in A$, — некоторая $(r+1)$-мерная стохастическая матрица.

Определение 5.13. *Цепь Маркова x_t называется цепью Маркова s-го порядка с r частичными связями* [29] *и обозначается ЦМ(s, r), если ее вероятности одношаговых переходов имеют вид*

$$p_{i_1,\,\ldots,\,i_s,\,i_{s+1}} = q^0_{i_{m^0_1},\,\ldots,\,i_{m^0_r},\,i_{s+1}},\quad i_1, \ldots, i_{s+1} \in A. \tag{5.15}$$

Соотношение (5.15) означает, что вероятность перехода процесса в состояние i_{s+1} в момент времени $t > s$ зависит не от всех s предыдущих состояний процесса $i_1, \ldots, i_s$, а лишь от r избранных состояний $i_{m^0_1}, \ldots, i_{m^0_r}$.

Таким образом, вместо $D_{\text{ЦМ}(s)} = N^s(N-1)$ параметров модель (5.15) полностью определяется $D_{\text{ЦМ}(s,\,r)} = N^r(N-1) + r - 1$ параметрами Q^0, M^0_r. Выигрыш в числе параметров может оказаться весьма существенным: например, если $N = 2$, $s = 32$, $r = 3$, то $D_{\text{ЦМ}(s)} \approx 4{,}3 \cdot 10^9$, в то время как $D_{\text{ЦМ}(s,\,r)} = 10$.

Заметим, что если $s = r$, $M^0_r = (1, \ldots, s)$, то $P = Q^0$ и ЦМ(s, s) есть полносвязная цепь Маркова s-го порядка: ЦМ(s, s) = ЦМ(s). Конструктивным примером ЦМ(s, s) является бинарная ($N = 2$) авторегрессия s-го порядка с r ненулевыми коэффициентами, частным случаем которой является линейная рекуррентная последовательность над кольцом Z_2, порожденная многочленом

степени s с r ненулевыми коэффициентами.

Примем обозначения: $J_s = (j_1, \ldots, j_s) = (J_{s-1}, j_s) \in A^s$ — мультииндекс s-го порядка; δ_{J_s, J'_s} — символ Кронекера для мультииндексов J_s, J'_s; $S_t(X_n; M_r) = \big(x_{t+m_1-1}, \ldots, x_{t+m_r-1}\big) \in A^r$ — функция $A^n \times \mathrm{M} \to A^r$, которую условимся называть селектором r-го порядка с параметрами $M_r \in \mathrm{M}$ и $t \in \{1, \ldots, n-s+1\}$.

Теорема 5.11. *ЦМ(s, r), определяемая (5.15), эргодична тогда и только тогда, когда найдется целое число $l \geqslant\limits_{s+l} 0$ такое, что*

$$\min_{J_s,\, J'_s \in A^s} \sum_{K_l \in A^l} \prod_{i=1}^{s+l} q^0_{S_i\left((J_s,\, K_l,\, J'_s);\, M^0_{r+1}\right)} > 0\,.$$

Доказательство основано на преобразовании ЦМ(s, r) в специальную векторную цепь Маркова первого порядка $<X>$ (см. разд. 5.2).

5.6. ДРУГИЕ МАЛОПАРАМЕТРИЧЕСКИЕ МОДЕЛИ ЦЕПЕЙ МАРКОВА ВЫСОКОГО ПОРЯДКА

В этом разделе дадим краткое описание еще некоторых малопараметрических моделей ОЦМ(s).

Определение 5.14. *Дискретная авторегрессия порядка s ($DAR(s)$) задается следующим стохастическим разностным уравнением* [18, 26]*:*

$$x_t = (\theta_1 x_{t-1} + \ldots + \theta_s x_{t-s} + \xi_t) \bmod N,\ t \geqslant s,$$

где $\theta_1, \ldots, \theta_s \in A$ — коэффициенты авторегрессии, причем $\theta_s \neq 0$; $\{\xi_t\}$ — независимые одинаково распределенные на A случайные величины с некоторым дискретным распределением вероятностей

$$\mathbf{P}\,\{\xi_t = k\} = p_k, k \in A.$$

Число параметров модели DAR(s):

$$D_{\mathrm{DAR}(s)} = N + s - 1.$$

Определение 5.15. *Модель дискретного скользящего среднего порядка q ($DMA(q)$) имеет вид $x_t = (\alpha_0 \xi_t + \alpha_1 \xi_{t-1} + \ldots + \alpha_q \xi_{t-q}) \bmod N$, $t \geqslant q$, где $\alpha_0 = 1$, $\alpha_1, \ldots, \alpha_q \in A$ — коэффициенты скользящего среднего.*

Обобщением моделей DAR(s) и DMA(q) является модель дискретной авторегрессии и скользящего среднего (DARMA(s, q)):

$$x_t = (\theta_1 x_{t-1} + \ldots + \theta_s x_{t-s} + \xi_t + \alpha_1 \xi_{t-1} + \ldots + \alpha_q \xi_{t-q}) \bmod N;$$

число параметров этой модели

$$D_{\mathrm{DARMA}(s,\, q)} = N + s + q - 1.$$

Определение 5.16. *Цепь Маркова переменного порядка определяется следующим малопараметрическим представлением $(s+1)$-мерной матрицы одношаговых переходов* [40] $P = \left(p_{i_1, \ldots, i_{s+1}}\right)$:

$$p_{i_1, \ldots, i_{s+1}} = q_{i_{s-l+1}, \ldots, i_{s+1}},$$

где $l = l(i_1, \ldots, i_s) : A^s \to \{1, \ldots, s\}$ — *некоторая заданная дискретная функция.*

Если $l(i_1, \ldots, i_s) \equiv s$, то получим полносвязную цепь Маркова ЦМ(s). Задание функции $l(\cdot)$ эквивалентно заданию так называемой *контекстной функции*:

$$c(i_1, \ldots, i_s) = (i_{s-l+1}, \ldots, i_s) : A^s \to A^l,$$

являющейся аналогом функции-селектора для модели ЦМ(s, r) (см. подразд. 5.5). Контекстная функция определяет *контекстное дерево*:

$$\tau = \{u : u = c(i_1, \ldots, i_s), (i_1, \ldots, i_s) \in A^s\}.$$

5.7. ЭНТРОПИЙНЫЕ ХАРАКТЕРИСТИКИ МАРКОВСКИХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

В системах обработки и защиты информации символы, образующие сообщение $\Xi_n = (\xi_1, \ldots, \xi_n) \in A^n$, обычно стохастически зависимы. Одной из распространенных моделей стохастической зависимости символов печатных текстов, речевых сообщений и изображений является *однородная цепь Маркова* (ОЦМ) с дискретным временем и конечным пространством состояний $A = \{a^{(1)}, \ldots, a^{(N)}\}$, $N < \infty$ (см. разд. 5.1).

Пусть случайная символьная последовательность, порожденная источником дискретных сообщений $\xi_1, \ldots, \xi_n, \ldots \in A$, — однородная цепь Маркова. Это означает, что распределение вероятностей будущих значений (состояний) при фиксированных настоящих и прошлых значениях не зависит от прошлых значений:

$$\mathbf{P}\{\xi_{t+1} = a_{t+1} \mid \xi_t = a_t, \ldots, \xi_1 = a_1\} = p_{a_t, a_{t+1}}, \tag{5.16}$$

$a_1, \ldots, a_{t+1} \in A$, $t = 1, 2, \ldots$. Известно также, что все конечномерные распределения и все вероятностные характеристики ОЦМ полностью выражаются лишь через вектор-столбец начальных вероятностей

$$\pi = \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_N \end{pmatrix}, \pi_i = \mathbf{P}\left\{\xi_1 = a^{(i)}\right\}, i = \overline{1, N}, \tag{5.17}$$

и через $(N \times N)$-матрицу вероятностей одношаговых переходов

$$P = (p_{ij}),$$

$$p_{ij} = \mathbf{P}\left\{\xi_{t+1} = a^{(j)} \mid \xi_t = a^{(i)}\right\}, i, j = \overline{1, N}\ (t = 1, 2, \ldots). \tag{5.18}$$

Вероятности (5.17) и (5.18) удовлетворяют условиям нормировки

$$\sum_{i=1}^{N} \pi_i = 1, \ \sum_{j=1}^{N} p_{ij} = 1 \ \left(i = \overline{1, N}\right). \tag{5.19}$$

Для получения дальнейших результатов введем N-вектор-столбец *стационарных вероятностей* $\pi^* = (\pi_i^*)$, $i = \overline{1, N}$, являющийся решением системы линейных алгебраических уравнений:

$$\begin{cases} \sum\limits_{i=1}^{n} \pi_i p_{ij} = \pi_j, \ \left(i = \overline{1, N}\right), \\ \pi_1 + \cdots + \pi_N = 1\,. \end{cases} \tag{5.20}$$

Обозначим энтропию стационарного распределения вероятностей:

$$\mathsf{H}^*\{\xi_1\} = -\sum_{i=1}^{N} \pi_i^* \log \pi_i^*. \tag{5.21}$$

Теорема 5.12. *Если случайная символьная последовательность является ОЦМ со стационарным начальным распределением π^* и матрицей вероятностей одношаговых переходов P, то энтропия n-символьного сообщения Ξ_n равна*

$$\mathsf{H}\{\Xi_n\} = \mathsf{H}^*\{\xi_1\} + (n-1)h\,, \tag{5.22}$$

где h — удельная энтропия, определяемая соотношением

$$h = -\sum_{i=1}^{N} \pi_i^* \sum_{j=1}^{N} p_{ij} \log p_{ij}\,. \tag{5.23}$$

Доказательство. В силу марковского свойства (5.16) имеем следующее представление для n-мерного дискретного распределения вероятностей:

$$\begin{gathered} p_n\big(a^{(i_1)}, \ldots, a^{(i_n)}\big) = \mathsf{P}\left\{\xi_1 = a^{(i_1)}\right\} \times \\ \times \prod_{t=1}^{n-1} \mathsf{P}\left\{\xi_{t+1} = a^{(i_{t+1})} \mid \xi_t = a^{(i_t)}, \ldots, \xi_1 = a^{(i_1)}\right\} = \\ = \pi_{i_1} \prod_{t=1}^{n-1} p_{i_t, i_{t+1}} \ (i_1, \ldots, i_n \in \{1, \ldots, N\})\,. \end{gathered} \tag{5.24}$$

Тогда с учетом (5.24) энтропия n-символьного сообщения $\Xi_n = (\xi_1, \ldots, \xi_n)$ примет вид

$$\mathsf{H}\{\Xi_n\} = -\sum_{i_1, \ldots, i_n = 1}^{N} p_n\Big(a^{(i_1)}, \ldots, a^{(i_n)}\Big)\left(\log \pi_{i_1} + \sum_{t=1}^{n-1} \log p_{i_t, i_{t+1}}\right).$$

По свойству согласованности многомерных вероятностных распределений

$$\begin{gathered} \mathsf{H}\{\Xi_n\} = -\sum_{i_1=1}^{N} \pi_{i_1} \log \pi_{i_1} - \\ -\sum_{t=1}^{n-1} \sum_{i_t, i_{t+1}=1}^{N} \mathsf{P}\left\{\xi_t = a^{(i_t)}, \xi_{t+1} = a^{(i_{t+1})}\right\} \log p_{i_t, i_{t+1}} = \end{gathered}$$

$$= -\sum_{i=1}^{N} \pi_i \log \pi_i - \sum_{t=1}^{n-1} \sum_{i_t,\, i_{t+1}=1}^{N} \mathsf{P}\left\{\xi_t = a^{(i_t)}\right\} p_{i_t,\, i_{t+1}} \log p_{i_t,\, i_{t+1}} \,. \qquad (5.25)$$

По условию теоремы начальное распределение ОЦМ совпадает с ее стационарным распределением: $\pi_i = \pi_i^*$ $\left(i = \overline{1, N}\right)$. Тогда из (5.20) легко показать, что одномерное распределение ОЦМ не изменяется с течением времени:

$$\mathsf{P}\left\{\xi_t = a^{(i)}\right\} = \pi_i^* \,,\, t = 1, 2, \ldots \ \left(i = \overline{1, N}\right).$$

Учитывая это в (5.25) и используя обозначения (5.23), получим

$$\mathsf{H}\left\{\Xi_n\right\} = \mathsf{H}^*\left\{\xi_1\right\} + \sum_{t=1}^{n-1} \sum_{i=1}^{N} \left(-\pi_i^* \sum_{j=1}^{N} p_{ij} \log p_{ij} \right) = \mathsf{H}^*\left\{\xi_1\right\} + (n-1)h \,,$$

что совпадает с (5.22).

Найдем удельную энтропию, используя определение и (5.22):

$$\lim_{n\to\infty} \frac{\mathsf{H}\left\{\Xi_n\right\}}{n} = \lim_{n\to\infty} \frac{\mathsf{H}^*\left\{\xi_1\right\}}{n} + h \lim_{n\to\infty} \left(1 - \frac{1}{n}\right) = h \,,$$

что и требовалось доказать. □

Отметим, что если нарушено условие теоремы о стационарности ОЦМ $\pi_i = \pi_i^*$ $\left(i = \overline{1, N}\right)$, то формула (5.25) примет вид

$$\mathsf{H}\left\{\Xi_n\right\} = \mathsf{H}\left\{\xi_1\right\} + \sum_{t=1}^{n-1} \sum_{i,\, j=1}^{N} \left(-\pi_i^{(t)} p_{ij} \log p_{ij} \right), \qquad (5.26)$$

где

$$\pi_i^{(t)} = \mathsf{P}\left\{\xi_t = a^{(i)}\right\}.$$

Из теории цепей Маркова известно, что при $t \to \infty$ и выполнении некоторых ограничений на P согласно эргодической теореме имеет место экспоненциальная сходимость распределения вероятностей $\pi^{(t)} = \left(\pi_i^{(t)}\right)$ к стационарному распределению $\pi^* = (\pi_i^*)$:

$$\pi_i^{(t)} = \pi_i^* + O\left(\rho^t\right),\ i = \overline{1, N},$$

где $0 < \rho < 1$, поэтому из (5.26) имеем

$$\frac{\mathsf{H}\left\{\Xi_n\right\}}{n} = \frac{\mathsf{H}\left\{\Xi_1\right\}}{n} + \frac{1}{n}\left((n-1)h + \sum_{t=1}^{n-1} O\left(\rho^t\right) \right) \to h \,, n \to \infty \,.$$

Таким образом, удельная энтропия ОЦМ даже при нарушении условия стационарности определяется (5.23), хотя соотношение (5.22) при этом не выполняется.

Следствие 5.4. *В условиях теоремы 5.12 соотношение (4.19) обращается в точное равенство*

$$\mathsf{H}\left\{\Xi_n\right\} = nh + 2b \,, \qquad (5.27)$$

где

$$2b = \sum_{i,\,j=1}^{N} \pi_i^* p_{ij} \log \frac{p_{ij}}{\pi_j^*} \,. \tag{5.28}$$

Доказательство. Представляя (5.22) в виде (5.27), найдем

$$2b = \mathsf{H}^* \{\xi_1\} - h \,.$$

Тогда из (5.23) следует

$$2b = -\sum_{i=1}^{N} \pi_i^* \log \pi_i^* + \sum_{i=1}^{N} \pi_i^* \sum_{j=1}^{N} p_{ij} \log p_{ij} = \sum_{i=1}^{N} \pi_i^* \sum_{j=1}^{N} p_{ij} \log \frac{p_{ij}}{\pi_j^*} \,.$$

Используя (5.20), можно показать, что это выражение совпадает с правой частью (5.28). □

Пример 5.1. Рассмотрим двоичную (бинарную) ОЦМ $\xi_1, \xi_2, \ldots \in A = \{0, 1\}$, т. е. $N = 2$, $a^{(1)} = 0$, $a^{(2)} = 1$. Вероятностные характеристики ОЦМ

$$\pi = \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix}, \; P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix},$$

где $\alpha, \beta \in [0, 1]$. Заметим, что при $\alpha + \beta = 1$, $\pi_1 = 1 - \alpha$, $\pi_2 = \alpha$ получим схему независимых испытаний. Для нахождения стационарного распределения π^* согласно (5.20) имеем систему

$$\begin{cases} \pi_1^*(1-\alpha) + \pi_2^* \beta = \pi_1^* \,, \\ \pi_1^* + \pi_2^* = 1. \end{cases}$$

Решение ее единственно:

$$\pi^* = \begin{pmatrix} \dfrac{\beta}{\alpha+\beta} \\[2ex] \dfrac{\alpha}{\alpha+\beta} \end{pmatrix}.$$

По формулам (5.21), (5.23) найдем

$$\mathsf{H}^* \{\xi_1\} = -\frac{\beta}{\alpha+\beta} \log \frac{\beta}{\alpha+\beta} - \frac{\alpha}{\alpha+\beta} \log \frac{\alpha}{\alpha+\beta} \,,$$

$$h = -\frac{\beta}{\alpha+\beta} (\alpha \log \alpha + (1-\alpha) \log (1-\alpha)) -$$

$$-\frac{\alpha}{\alpha+\beta} (\beta \log \beta + (1-\beta) \log (1-\beta)) \,.$$

По соотношению (5.28) получим

$$2b = 2\frac{\alpha\beta}{\alpha+\beta}\log\beta + \log\left(1+\frac{\alpha}{\beta}\right) + \frac{\beta}{\alpha+\beta}(1-\alpha)\log(1-\alpha) +$$

$$+\frac{\alpha}{\alpha+\beta}(1-\beta)\log(1-\beta)\,.$$

При $\beta = \alpha$ (симметричная ОЦМ) найдем $\pi_i^* = 1/2\,,$

$$\mathbf{H}^*\{\xi_1\} = \log 2 = 1\,,$$

$$h = -((1-\alpha)\log(1-\alpha) + \alpha\log\alpha)\,,$$

$$2b = \log 2 + (1-\alpha)\log(1-\alpha) + \alpha\log\alpha = 1 - h\,.$$

При $\alpha = 1/2$ имеем схему независимых испытаний (удельная энтропия максимальна). При $\alpha \to 0$ или $\alpha \to 1$ удельная энтропия $h \to 0$, что согласуется с фактом уменьшения неопределенности двоичных сообщений.

5.8. ТЕОРЕМА МАК-МИЛЛАНА ДЛЯ ДИСКРЕТНОГО ЭРГОДИЧЕСКОГО ИСТОЧНИКА

Пусть $\langle A,\, p(\cdot)\rangle$ — определенный на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbf{P})$ произвольный дискретный эргодический источник, свойства которого рассмотрены в гл. 3. Оказывается, для такого ИДС справедливо свойство асимптотической равнораспределенности, установленное ранее в гл. 4 для ДИБП. Этому обобщению и посвящен данный раздел.

Примем обозначения: $a^{(l)} = \left(a_1^{(l)}, \ldots, a_l^{(l)}\right) \in A^l$ — произвольную цепочку $l \in \mathbb{N}$ символов из алфавита A, $p\left(a^{(l)}\right) \in [0,1]$ — вероятность появления этой цепочки.

Определение 5.17. *Величину*

$$H_l = -\frac{1}{l}\sum_{a^{(l)}\in A^l} p\left(a^{(l)}\right)\log p\left(a^{(l)}\right) \tag{5.29}$$

условимся [33] *называть l-энтропией на знак, а условную энтропию*

$$H^{(l)} = H\left(a_l|a_1,\,\ldots,a_{l-1}\right) = -\sum_{a^{(l)}\in A^l} p\left(a^{(l)}\right)\log p\left(a_l|a_1,\,\ldots,a_{l-1}\right) - \tag{5.30}$$

шаговой l-энтропией ИДС.

Лемма 5.1. *Шаговая l-энтропия и l-энтропия на знак являются невозрастающими последовательностями, имеющими предел:*

$$\lim_{l\to\infty} H_l = \lim_{l\to\infty} H^{(l)} = H^\infty,$$

где H^∞ — энтропия ИДС (определенная в разд. 4.1 как удельная энтропия).

Доказательство. Результаты этой леммы вытекают из доказательства теорем 4.1, 4.2. □

Исследуем асимптотическое поведение l-мерного дискретного распределения вероятностей l-фрагмента $(x_{t+1}, \ldots, x_{t+l}) \in A^l$, порождаемого дискретным эргодическим источником $\langle A, p(\cdot)\rangle$, при $l \to \infty$. Выберем произвольное натуральное число $1 \leqslant m \leqslant l$ и введем в рассмотрение вспомогательную функцию на A^l:

$$q_m\left(a^{(l)}\right) = p\left(a^{(m)}\right) \prod_{k=m+1}^{l} p\left(a_k^{(l)} | a_{k-1}^{(l)}, \ldots, a_{k-m}^{(l)}\right), \quad a^{(l)} \in A^l. \tag{5.31}$$

Лемма 5.2. *Функция, определяемая (5.31), является:*

1) некоторым распределением вероятностей над A^l;

2) l-мерным распределением вероятностей случайной последовательности $\{x_t\}$, если $\{x_t\}$ — однородная цепь Маркова порядка, не превосходящего m.

Доказательство. Для доказательства первого свойства достаточно проверить условие нормировки.

Второе свойство вытекает из марковского свойства m-го порядка и обобщенной формулы умножения вероятностей. □

Из леммы 5.2 следует, что функцию $q_m\left(a^{(l)}\right)$ можно рассматривать как аппроксимацию истинного l-мерного распределения вероятностей $p\left(a^{(l)}\right)$, получающуюся пренебрежением памяти процесса $\{x_t\}$ на глубину большую, чем m.

Лемма 5.3. *Для дискретного эргодического источника при произвольном $m \in \mathbb{N}$ справедливо предельное соотношение*

$$\lim_{l\to\infty} \frac{1}{l} \log q_m\left(a^{(l)}\right) = -H^{(m+1)}. \tag{5.32}$$

Доказательство. Прологарифмируем (5.31):

$$\log q_m\left(a^{(l)}\right) = \log p\left(a^{(m)}\right) + \sum_{k=m+1}^{l} \log p\left(a_k^{(l)} | a_{k-1}^{(l)}, \ldots, a_{k-m}^{(l)}\right). \tag{5.33}$$

Функция $\log p\left(a_k^{(l)} \mid a_{k-1}^{(l)}, \ldots, a_{k-m}^{(l)}\right)$, определенная на A^{m+1}, с учетом (5.30) — суммируемая функция, имеющая конечное математическое ожидание:

$$\sum_{a_k^{(l)}, \ldots, a_{k-m}^{(l)} \in A} p\left(a_{k-m}^{(l)}, \ldots, a_k^{(l)}\right) \log p\left(a_k^{(l)} | a_{k-1}^{(l)}, \ldots, a_{k-m}^{(l)}\right) < \infty.$$

Поэтому к ней применима теорема Биркгофа (см. разд. 3.2), согласно которой почти всюду

$$\lim_{l\to\infty} \frac{1}{l-m} \sum_{k=m+1}^{l} \log p\left(a_k^{(l)} | a_{k-1}^{(l)}, \ldots, a_{k-m}^{(l)}\right) =$$

$$= -H\left(a_{m+1} | a_1, \ldots, a_m\right) = -H^{(m+1)}. \tag{5.34}$$

Разделим (5.33) почленно на l и перейдем к пределу при $l \to \infty$. Поскольку $p\left(a^{(m)}\right)$ не зависит от l, то первое слагаемое в этом соотношении стремится к нулю, а второе с учетом (5.34) приводит к (5.32). □

Введем в рассмотрение функцию-срезку:

$$(x)^+ = \begin{cases} x, & \text{если } x \geqslant 0, \\ 0, & \text{если } x < 0. \end{cases}$$

При этом $|x| = 2(x)^+ - x$.

Лемма 5.4 (об аппроксимации l-мерных распределений). *Для дискретного эргодического источника, произвольных $\varepsilon > 0$, $\delta > 0$, достаточно большого m и любого $l > m$ справедлива следующая оценка ошибки аппроксимации:*

$$\mathsf{P}\left\{\left|\frac{1}{l}\log q_m\left(x^{(l)}\right) - \frac{1}{l}\log p\left(x^{(l)}\right)\right| > \varepsilon\right\} \leqslant \delta, \tag{5.35}$$

где $x^{(l)} = (x_1, \ldots, x_l) \in A^l$ — l-фрагмент случайной последовательности $\{x_t\}$, порождаемой источником.

Доказательство. Оценим левую часть (5.35) сверху, используя неравенство Чебышева (относительно математического ожидания) [30]:

$$D_{lm} = \mathsf{P}\left\{\left|\frac{1}{l}\log q_m\left(x^{(l)}\right) - \frac{1}{l}\log p\left(x^{(l)}\right)\right| > \varepsilon\right\} \equiv$$

$$\equiv \mathsf{P}\left\{\left|\log\frac{q_m\left(x^{(l)}\right)}{p\left(x^{(l)}\right)}\right| > l\varepsilon\right\} \leqslant \frac{1}{l\varepsilon}\mathsf{E}\left\{\left|\log\frac{q_m\left(x^{(l)}\right)}{p\left(x^{(l)}\right)}\right|\right\} \equiv$$

$$\equiv \frac{1}{l\varepsilon}\left(2\mathsf{E}\left\{\left(\log\frac{q_m\left(x^{(l)}\right)}{p\left(x^{(l)}\right)}\right)^+\right\} - \mathsf{E}\left\{\log\frac{q_m\left(x^{(l)}\right)}{p\left(x^{(l)}\right)}\right\}\right).$$

К первому слагаемому в скобках применим известное неравенство

$$(\log x)^+ \leqslant \frac{x\log e}{e}, \quad x \geqslant 0.$$

Получим

$$D_{lm} \leqslant \frac{1}{l\varepsilon}\left(\frac{2\log e}{e}\mathsf{E}\left\{\frac{q_m\left(x^{(l)}\right)}{p\left(x^{(l)}\right)}\right\} - \mathsf{E}\left\{\log\frac{q_m\left(x^{(l)}\right)}{p\left(x^{(l)}\right)}\right\}\right) \equiv$$

$$\equiv \frac{2\log e}{el\varepsilon} - \frac{1}{l\varepsilon}\mathsf{E}\left\{\log\frac{q_m\left(x^{(l)}\right)}{p\left(x^{(l)}\right)}\right\}.$$

Вычислим последнее слагаемое, используя (5.29)–(5.31):

$$\mathsf{E}\left\{\log\frac{q_m\left(x^{(l)}\right)}{p\left(x^{(l)}\right)}\right\} =$$

$$= \sum_{a^{(l)}\in A^l} p\left(a^{(l)}\right)\log\frac{p\left(a^{(m)}\right)\prod\limits_{k=m+1}^{l} p\left(a_k^{(l)}|a_{k-1}^{(l)},\ldots,a_{k-m}^{(l)}\right)}{p\left(a^{(l)}\right)} =$$

$$= lH_l - mH_m - (l-m)H^{(m+1)}.$$

В итоге получим оценку левой части (5.35):

$$D_{lm} \leqslant \frac{1}{\varepsilon}\left(\frac{2\log e}{el} + \frac{m}{l}\left(H_m - H^{(m+1)}\right) + \left(H^{(m+1)} - H_l\right)\right). \tag{5.36}$$

Воспользуемся произволом в выборе m и проанализируем три слагаемых, выделенных в (5.36). Для удобства обозначим $\delta' = \delta\cdot\varepsilon$.

Поскольку $l > m$, то выбором m можно добиться, чтобы первое слагаемое в (5.36) было ограничено сверху:

$$\frac{2\log e}{e}\cdot\frac{1}{l} < \frac{\delta'}{3}.$$

Пусть это выполняется при $m > m_1(\delta')$.

Затем в силу леммы 5.1 найдется $m_2(\delta') < \infty$ такое, что при $m > m_2(\delta')$ второе слагаемое в (5.36) ограничено:

$$\frac{m}{l}\left(H_m - H^{(m+1)}\right) < \frac{\delta'}{3}.$$

Наконец, снова с учетом леммы 5.1 имеем $\lim\limits_{m\to\infty} H^{(m+1)} = \lim\limits_{l\to\infty} H_l = H^\infty$, поэтому найдется $m_3(\delta') < \infty$ такое, что при $m > m_3(\delta')$ третье слагаемое в (5.36) меньше:

$$\left|H^{(m+1)} - H_l\right| < \frac{\delta'}{3}.$$

Выбирая $m_+(\delta') = \max\left(m_1(\delta'), m_2(\delta'), m_3(\delta')\right) < \infty$ и подставляя полученные оценки в (5.36), имеем

$$D_{lm} < \frac{1}{\varepsilon}\left(\frac{\delta'}{3} + \frac{\delta'}{3} + \frac{\delta'}{3}\right) = \delta.$$

□

Теорема 5.13 (теорема Мак-Миллана). *Пусть случайная последовательность $\{x_t\}$ порождена дискретным эргодическим источником $\langle A, p(\cdot)\rangle$. Тогда для произвольных $\varepsilon > 0$, $\delta > 0$, существует натуральное число $l_0 = l_0(\varepsilon,\delta)$ такое, что при $l > l_0$ справедливо неравенство*

$$\mathsf{P}\left\{\left|\frac{1}{l}\log\frac{1}{p\left(x^{(l)}\right)} - H^\infty\right| > \varepsilon\right\} < \delta, \tag{5.37}$$

где H^∞ — энтропия источника (удельная энтропия).

Доказательство. В силу леммы 5.1 имеем

$$\lim_{m\to\infty} H^{(m)} = H^\infty.$$

По определению предела для любого $\varepsilon > 0$ существует натуральное число $m_0 = m_0(\varepsilon) < \infty$ такое, что при $m > m_0$

$$\mathbf{P}\left\{\left|H^{(m)} - H^\infty\right| > \varepsilon\right\} \leqslant \delta', \tag{5.38}$$

причем из-за неслучайности $H^{(m)}$ здесь можно положить $\delta' = 0$.

Из леммы 5.3 для любого $\varepsilon > 0$ найдется $l_0 = l_0(\varepsilon) < \infty$ такое, что при $l > l_0$

$$\mathbf{P}\left\{\left|\frac{1}{l}\log q_m\left(x^{(l)}\right) + H^{(m+1)}\right| > \varepsilon\right\} \leqslant \delta'. \tag{5.39}$$

Из леммы 5.4 для любого $\varepsilon > 0$ для $m > m_0$, $l > m$ получим

$$\mathbf{P}\left\{\left|\frac{1}{l}\log q_m\left(x^{(l)}\right) - \frac{1}{l}\log p\left(x^{(l)}\right)\right| > \varepsilon\right\} \leqslant \delta'. \tag{5.40}$$

Оценим левую часть (5.37), используя (5.38)–(5.40) и свойства вероятности:

$$\mathbf{P}\left\{\left|-\frac{1}{l}\log p\left(x^{(l)}\right) - H^\infty\right| > \varepsilon\right\} \equiv \mathbf{P}\left\{\left|-\frac{1}{l}\log p\left(x^{(l)}\right) + \frac{1}{l}\log q_m\left(x^{(l)}\right) -\right.\right.$$

$$\left.\left. -\frac{1}{l}\log q_m\left(x^{(l)}\right) - H^{(m+1)} + H^{(m+1)} - H^\infty\right| > \varepsilon\right\} \leqslant$$

$$\leqslant \mathbf{P}\left\{\left|-\frac{1}{l}\log p\left(x^{(l)}\right) + \frac{1}{l}\log q_m\left(x^{(l)}\right)\right| > \varepsilon\right\} +$$

$$+\mathbf{P}\left\{\left|\frac{1}{l}\log q_m\left(x^{(l)}\right) + H^{(m+1)}\right| > \varepsilon\right\} + \mathbf{P}\left\{\left|H^{(m+1)} - H^\infty\right| > \varepsilon\right\} \leqslant 3\delta'.$$

Полагая $\delta' = \delta/3$, получим, что для достаточно больших l будет выполняться неравенство (5.37). □

Выводы, сделанные в гл. 4 для ДИБП, оказывается, справедливы и для эргодических ИДС.

Имеет место свойство информационной устойчивости. Собственная информация (энтропия) l-фрагмента последовательности, порожденной эргодическим ИДС, допускает асимптотику при $l \to \infty$:

$$H\left\{x^{(l)}\right\} = \mathbf{I}\left\{x^{(l)}, x^{(l)}\right\} \approx \log\frac{1}{p\left(a^{(l)}\right)} \approx lH^\infty,$$

т. е. выполняется свойство асимптотической равнораспределенности:

$$p\left(a^{(l)}\right) \approx a^{-lH^\infty},$$

где a — основание логарифма. Это свойство позволяет оценить количество последовательностей, порожденных эргодическим ИДС:

$$N_l \approx \frac{1}{p\left(a^{(l)}\right)} \approx a^{lH^{\infty}}.$$

5.9. РЕШЕНИЕ ТИПОВЫХ ЗАДАЧ

Задача 5.1. Пусть дана двоичная ОЦМ $\xi_1,\xi_2,\ldots \in A = \{0,1\}$, т. е. $N = 2$, $a^{(1)} = 0$, $a^{(2)} = 1$. Вероятностные характеристики ОЦМ

$$\pi = \begin{pmatrix}\pi_1\\ \pi_2\end{pmatrix}, P = \begin{pmatrix}1-\alpha & \alpha\\ \beta & 1-\beta\end{pmatrix},$$

где $\alpha, \beta \in [0,1]$. Найти удельную энтропию h.

Решение. Для нахождения стационарного распределения π^* запишем систему

$$\begin{cases}\pi_1^*(1-\alpha) + \pi_2^*\beta = \pi_1^*,\\ \pi_1^* + \pi_2^* = 1.\end{cases}$$

Решение ее единственно:

$$\pi^* = \left(\frac{\beta}{\alpha+\beta} \quad \frac{\alpha}{\alpha+\beta}\right)'.$$

По формулам для удельной энтропии

$$\mathbf{H}^*\{\xi_1\} = -\frac{\beta}{\alpha+\beta}\log\frac{\beta}{\alpha+\beta} - \frac{\alpha}{\alpha+\beta}\log\frac{\alpha}{\alpha+\beta},$$

$$h = -\frac{\beta}{\alpha+\beta}(\alpha\log\alpha + (1-\alpha)\log(1-\alpha)) -$$

$$-\frac{\alpha}{\alpha+\beta}(\beta\log\beta + (1-\beta)\log(1-\beta)).$$

При $\beta = \alpha$ (симметричная ОЦМ) найдем $\pi_i^* = 1/2$,

$$\mathbf{H}^*\{\xi_1\} = \log 2,$$

$$h = -((1-\alpha)\log(1-\alpha) + \alpha\log\alpha),$$

$$2b = \log 2 + (1-\alpha)\log(1-\alpha) + \alpha\log\alpha = 1 - h.$$

При $\alpha = 1/2$ имеем схему независимых испытаний (удельная энтропия максимальна). При $\alpha \to 0$ или $\alpha \to 1$ удельная энтропия $h \to 0$, что согласуется с фактом уменьшения неопределенности двоичных сообщений.

Ответ: $h = -\frac{\beta}{\alpha+\beta}(\alpha\log\alpha + (1-\alpha)\log(1-\alpha)) - \frac{\alpha}{\alpha+\beta}(\beta\log\beta + (1-\beta)\times$ $\times\log(1-\beta))$.

Задача 5.2. Двоичное сообщение описывается однородной цепью Маркова $\xi_1, \xi_2, \ldots \in \{0,1\}$ с дискретным временем, с начальным распределением $\pi = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$ и матрицей вероятностей одношаговых переходов

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix}, \, 0 \leqslant \alpha \leqslant 1.$$

Пусть для некоторого момента времени $t = 1, 2, \ldots$ определены два соседних символа: $X ::= \xi_{t+1}$ (будущий символ = «пропущенный» символ), $Y ::= \xi_t$ (соседний наблюдаемый символ). Оценить количество информации о ξ_{t+1}, содержащееся в ξ_t.

Решение. По свойствам ОЦМ с дискретным временем имеем

$$\pi^* = \pi = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}, p_X(X) = p_Y(Y) \equiv \frac{1}{2}, \, X, Y \in \{0,1\},$$

$$p_{X,Y}(X,Y) = \frac{1}{2} \begin{cases} 1-\alpha, \text{ если } Y = X, \\ \alpha, \text{ если } Y \neq X. \end{cases}$$

Поэтому

$$\mathsf{I}\{\xi_{t+1},\xi_t\} = \sum_{X,Y=0}^{1} \frac{1}{2}\Big((1-\alpha)\delta_{Y,X} + \alpha(1-\delta_{Y,X})\Big) \log\Big(2\big((1-\alpha)\delta_{Y,X} +$$

$$+ \alpha(1-\delta_{Y,X})\big)\Big) = 1 + \alpha\log\alpha + (1-\alpha)\log(1-\alpha) = 1 - h(\alpha)\,,$$

где $h(\alpha) = -\alpha\log\alpha - (1-\alpha)\log(1-\alpha)$.

Ответ: $\mathsf{I}\{\xi_{t+1},\xi_t\} = 1 + \alpha\log\alpha + (1-\alpha)\log(1-\alpha)$.

5.10. ЗАДАЧИ И УПРАЖНЕНИЯ

5.1. Задана однородная двоичная цепь Маркова 1-го порядка $\Xi_n \in V_n$, $V = \{0,1\}$ с матрицей вероятностей одношаговых переходов

$$P = \begin{pmatrix} 1/2 & 1/2 \\ 2/5 & 3/5 \end{pmatrix}.$$

Найти матрицу $P^{(2)}$ вероятностей переходов за $\tau = 2$ шага, $(P^{(2)})_{ij} = \mathbf{P}\{\xi_t = j | \xi_{t-2} = i\}$, $i,j \in V$.

5.2. Пусть $\Xi_n = (\xi_1, \ldots, \xi_n) \in \{0,1\}^n$ — однородная стационарная цепь Маркова 1-го порядка с матрицей вероятностей одношаговых переходов $P = P(\varepsilon) = \begin{pmatrix} \varepsilon & 1-\varepsilon \\ 1-\varepsilon & \varepsilon \end{pmatrix}$, $\varepsilon \in [0,1]$. Определить $\mathsf{H}\{\xi_t | \xi_{t-2}\}$.

5.3. $\Xi_n = (\xi_1, \ldots, \xi_n) \in \{0,1\}^n$ — однородная стационарная цепь Маркова 1-го порядка с матрицей вероятностей одношаговых переходов P и стационарным распределением вероятностей π. Вычислить $\mathsf{H}\{\xi_t|\xi_{t-\tau}\}$, $\tau \geqslant 1$.

5.4. Колода, состоящая из 26 красных карт и 26 черных, хорошо перетасована. Поочередно извлекают карту за картой без возвращения. Пусть a_k — цвет k-й извлеченной карты. Найти $\mathsf{H}\{a_1\}$, $\mathsf{H}\{a_2\}$.

5.5. Для однородной стационарной цепи Маркова 1-го порядка $\Xi_n \in A^n$, $|A| = 3$ с матрицей вероятностей одношаговых переходов

$$P = \frac{1}{6}\begin{pmatrix} 3 & 3 & 0 \\ 2 & 0 & 4 \\ 2 & 4 & 0 \end{pmatrix}.$$

Определить стационарное распределение вероятностей π, вычислить условную энтропию $\mathsf{H}\{\xi_{t+1}|\xi_t\}$ и удельную энтропию h.

5.6. Марковский источник Ξ_n задается алфавитом символов $A = \{a_1, a_2, a_3, a_4\}$, равномерным начальным распределением вероятностей $\mathbf{P}\{\xi_1 = a_i\} = \dfrac{1}{4}$, $i \in \in \{1,2,3,4\}$ и матрицей вероятностей одношаговых переходов

$$P = \frac{1}{8}\begin{pmatrix} 4 & 2 & 1 & 1 \\ 2 & 1 & 1 & 4 \\ 1 & 1 & 4 & 2 \\ 1 & 4 & 2 & 1 \end{pmatrix}.$$

Появление очередного символа в последовательности, порождаемой источником, зависит только от одного предыдущего символа. Найти энтропию $\mathsf{H}\{\Xi_n\}$, условную энтропию $\mathsf{H}\{\xi_{t+1}|\xi_t\}$ и удельную энтропию h.

5.7. $\Xi_n = (\xi_1, \ldots, \xi_n) \in \{0,1\}^n$ — однородная стационарная цепь Маркова 1-го порядка с матрицей вероятностей одношаговых переходов $P = P(\varepsilon) = \begin{pmatrix} \varepsilon & 1-\varepsilon \\ 1-\varepsilon & \varepsilon \end{pmatrix}$, $\varepsilon \in [0,1]$. Вычислить $\mathsf{I}\{\xi_{t+1},\xi_t\}$, $\mathsf{H}\{\xi_t|\xi_t + \xi_{t-1}\}$.

5.8. $\Xi_n = (\xi_1, \ldots, \xi_n) \in \{0,1\}^n$ — однородная стационарная цепь Маркова 1-го порядка с матрицей вероятностей одношаговых переходов $P = P(\varepsilon) = \begin{pmatrix} \varepsilon & 1-\varepsilon \\ 1-\varepsilon & \varepsilon \end{pmatrix}$, $\varepsilon \in [0,1]$. Определить удельную энтропию h.

5.9. $\Xi_n = (\xi_1, \ldots, \xi_n) \in \{0,1\}^n$ — однородная стационарная цепь Маркова 1-го порядка с матрицей вероятностей одношаговых переходов $P = P(\varepsilon) = \dfrac{1}{2}\begin{pmatrix} 1+\varepsilon & 1-\varepsilon \\ 1-\varepsilon & 1+\varepsilon \end{pmatrix}$, $\varepsilon \in [-1,1]$. Найти $\mathsf{I}\{\xi_{t+1},\xi_t\}$, $\mathsf{H}\{\xi_t|\xi_t \oplus \xi_{t-1}\}$.

5.10. Пусть стационарный источник дискретных сообщений порождает случайную символьную последовательность $\Xi_n = (\xi_1, \ldots, \xi_n) \in A^n$, обладающую

марковским свойством s-го порядка ($s > 0$):

$$\mathbf{P}\{\xi_t = a_t | \xi_{t-1}, \ldots, \xi_1 = a_1\} = \mathbf{P}\{\xi_t = a_t | \xi_{t-1}, \ldots, \xi_{t-s} = a_{t-s}\},$$
$$a_1, \ldots, a_t \in A,\ t > s.$$

Определить условную энтропию $\mathsf{H}\{\xi_n | \Xi_{n-1}\}$, $n > s$, и удельную энтропию h.

5.11. $\Xi_n = (\xi_1, \ldots, \xi_n) \in V^n$ — однородная стационарная двоичная цепь Маркова s-го порядка ($s > 0$) с $r \leqslant s$ частичными связями, где π — стационарное распределение вероятностей, $M_r^0 = (m_1^0 = 1, m_2^0, \ldots, m_r^0)$ — шаблон связей, $Q = (q_{j_1, \ldots, j_r, j_{r+1}})$ — $(r+1)$-мерная стохастическая матрица такая, что

$$\mathbf{P}\{\xi_t = i_{s+1} | \xi_{t-1} = i_s, \ldots, \xi_{t-s} = i_1\} = q_{i_{m_1^0}, \ldots, i_{m_r^0}, i_{s+1}}.$$

Вычислить удельную энтропию h.

Глава 6

ШЕННОНОВСКИЙ ПОДХОД К ШИФРОВАНИЮ

6.1. ШЕННОНОВСКИЕ МОДЕЛИ КРИПТОСИСТЕМ

В разд. 6.1, 6.2 будет рассмотрено применение шенноновской теории информации к построению математических моделей криптосистем и оценке стойкости простейших симметричных криптосистем [14, 34].

В настоящее время криптосистемы принято разделять на два класса: симметричные (одноключевые) и асимметричные (двухключевые). Общая схема симметричной криптосистемы приведена на рис. 6.1.

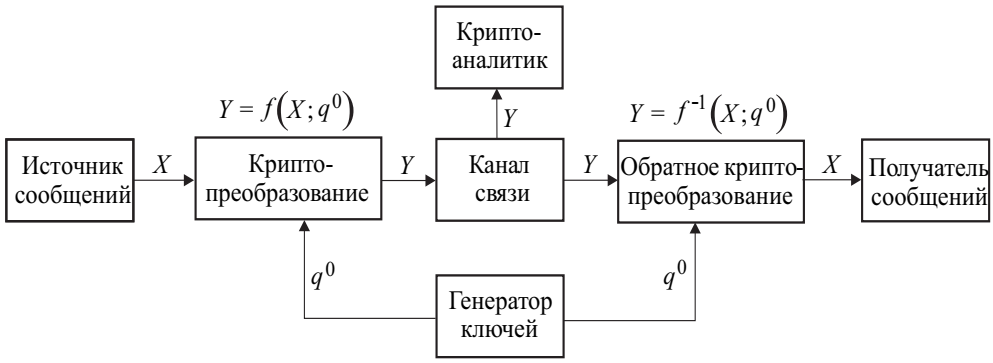


Рис. 6.1. Общая схема криптосистемы

В данной главе приняты следующие обозначения: $X = (x_1, \ldots, x_n)$ — исходное сообщение, генерируемое источником сообщений и представляющее собой последовательность n символов алфавита мощностью $\nu \geqslant 2$:

$$x_i \in A_\nu = \{0, 1, \ldots, \nu - 1\}, \quad i = \overline{1, n}\,, \; X \in A_\nu^n\,;$$

$\theta^0 = (\theta_1^0, \theta_2^0, \ldots, \theta_L^0) \in A_\mu^L$ — истинное значение ключа, L символов которого принимают значения из алфавита A_μ, $\mu \geqslant 2$; $Y = (y_1, y_2, \ldots, y_n) \in A_\nu^n$ — шифртекст (криптограмма, выходная последовательность, зашифрованное сообщение), получающийся применением к X криптографического дискретного функционального преобразования (операции шифрования):

$$Y = f(X;\theta^0):\ V_\nu^n \times V_\mu^L \to V_\nu^n \tag{6.1}$$

или

$$y_i = f_i(x_1, \ldots, x_n;\theta^0)\,,\ i = \overline{1,n}\,.$$

Функция $f(\cdot)$ в (6.1) такова, что при любом фиксированном значении ключа θ^0 она является взаимооднозначным функциональным преобразованием (биекцией), и обратное преобразование (расшифрование, дешифрование) единственно и восстанавливает исходный текст:

$$X = f^{-1}(Y;\theta^0)\,. \tag{6.2}$$

Особенностью симметричных (одноключевых) систем является симметричное использование одного и того же ключа θ^0 отправителем и получателем (этот ключ секретный и поставляется абонентам специальным конфиденциальным способом). Отсюда и название данного класса криптосистем — «одноключевые».

Приведем математические модели элементарных криптосистем, которые потребуются при оценке их стойкости, а также при решении задач криптоанализа и построения стандартных криптосистем.

1. *Подстановка символов алфавита.* Пусть определены подстановка на множестве $\{1, 2, \ldots, \nu\}$

$$\begin{pmatrix} 1 & 2 & \ldots & \nu \\ s_1 & s_2 & \ldots & s_\nu \end{pmatrix} \tag{6.3}$$

и ν-вектор ($L = \nu = \mu$)

$$\theta^0 = (\theta_1^0, \ldots, \theta_\nu^0)\,,\ \theta_i^0 = s_i - 1,\ i = \overline{1,\nu}\,,$$

задающий перестановку символов алфавита A_ν.

Тогда шифр простой подстановки — это криптографическое преобразование вида (6.1), осуществляемое поэлементно:

$$y_t = f_t(x_t;\theta^0) ::= \theta^0_{x_t+1}\,,\ t = 1, 2, \ldots\ . \tag{6.4}$$

Обратное преобразование (6.2) при этом будет иметь вид

$$x_t = \bar\theta^0_{y_t+1} =:: f_t^{-1}(y_t;\theta^0)\,, \tag{6.5}$$

где $\bar\theta_i^0 = \bar s_i - 1$, $\bar s_i$ — элементы подстановки, обратной (6.3):

$$\begin{pmatrix} 1 & 2 & \ldots & \nu \\ \bar s_1 & \bar s_2 & \ldots & \bar s_\nu \end{pmatrix}.$$

2. *Перестановка символов с периодом T.* Пусть $T \in N$ — некоторый заданный период и определена некоторая подстановка на множестве $\{1, 2, \ldots, T\}$:

$$\begin{pmatrix} 1 & 2 & \ldots & T \\ s_1 & s_2 & \ldots & s_T \end{pmatrix}.$$

Как и в предыдущем преобразовании, с помощью этой подстановки определяется ключ $\theta^0 = (\theta_1^0, \ldots, \theta_T^0)$, $\theta_i^0 = s_i$, $i = \overline{1,T}$.

Криптопреобразование (6.1) осуществляется следующим образом. Исходное сообщение X разбивается на блоки символов длиной T, и внутри каждого блока производится перестановка символов в соответствии с заданным ключом θ^0. Для произвольного номера символа $t = (i-1)T + \tau$, где $i \in \{1, 2, \ldots\}$, $\tau \in \{1, 2, \ldots, T\}$, такое преобразование запишется в виде

$$y_t = x_{(i-1)T+\theta^0_\tau}\,. \tag{6.6}$$

Например, если $T = 5$, а $\theta^0 = (2, 3, 1, 5, 4)$, то сообщение

$$X = (x_1, x_2, x_3, x_4, x_5 | x_6, x_7, x_8, x_9, x_{10} | \ldots)$$

переходит в шифртекст

$$Y = (x_2, x_3, x_1, x_5, x_4 | x_7, x_8, x_6, x_{10}, x_9 | \ldots)\,.$$

Легко убедиться, что обратное преобразование (по отношению к (6.6)) имеет вид

$$x_t = y_{(i-1)T+\bar\theta^0_\tau}\,, \tag{6.7}$$

где

$$\begin{pmatrix} 1 & 2 & \ldots & T \\ \bar\theta^0_1 & \bar\theta^0_2 & \ldots & \bar\theta^0_T \end{pmatrix}$$

есть подстановка, обратная θ^0.

Для усиления стойкости криптопреобразования (6.6) к криптоанализу используют композицию нескольких перестановок с различными периодами. Если сделано $L \geqslant 2$ перестановок типа (6.6) с периодами $T_1, \ldots, T_L$, то составная перестановка, очевидно, будет иметь период

$$T = \text{НОК}(T_1, T_2, \ldots, T_L)\,.$$

Следовательно, если периоды $\{T_1, \ldots, T_L\}$ — взаимно простые числа, то достигается максимальный период $T_{\max} = T_1 \cdots T_L$.

3. *Шифр Виженера и его модификации.* Как и в предыдущей криптосистеме, исходный текст X разбивается на блоки длиной T. Ключ θ^0 представляет собой фиксированный набор символов исходного алфавита A_ν ($\mu = \nu$):

$$\theta^0 = (\theta^0_1, \ldots, \theta^0_T),\ \theta^0_i \in V_\nu\,,\ i = \overline{1, T}\,.$$

Криптофункция для произвольного номера $t = (i-1)T + \tau$, $i \in \{1, 2, \ldots\}$, $\tau \in \{1, 2, \ldots, T\}$, задается с помощью вычетов по модулю ν:

$$y_t = (x_t + \theta^0_\tau) \mod \nu\,. \tag{6.8}$$

Это криптопреобразование иногда называется преобразованием циклического сдвига с периодом T. Обратное преобразование по отношению к (6.8)

$$x_t = (y_t + \nu - \theta^0_\tau) \mod \nu\,. \tag{6.9}$$

Пример шифртекста, построенного с помощью шифра Виженера с периодом $T = 6$ и ключом $\theta^0 = (4, 3, 2, 5, 1, 3)$, можно найти в романе Ж. Верна «Жангада» [5].

Приведем ряд частных случаев криптопреобразования Виженера, известных с древних времен.

Ш и ф р Ц е з а р я — это частный случай преобразования Виженера с периодом $T = 1$ и ключом $\theta^0 \in A_\nu$:

$$y_t = (x_t + \theta^0) \mod \nu,\ t = 1, 2, \ldots . \tag{6.10}$$

При этом согласно (6.10) каждый символ (буква) исходного текста заменяется символом, циклически сдвинутым на фиксированное количество мест θ^0 по алфавиту A_ν.

В качестве примера приведем шифртекст длиной $n = 41$:

$$Y = \mathit{PELCGBYBTLVFPELCGBTENCULNAQPELCGBNANYLFVF},$$

полученный с помощью криптопреобразования Цезаря (6.10) при $\nu = 26$, $\theta^0 = 13$ из исходного отрывка английского текста:

$$X = \mathit{CRYPTOLOGYISCRYPTOGRAPHYANDCRYPTOANALYSIS}.$$

Иногда рассматривают обратный шифр Цезаря:

$$y_t = (\theta^0 + \nu - x_t) \mod \nu . \tag{6.11}$$

Ш и ф р Б о ф о р а — это модификация T-периодического шифра Виженера (6.8):

$$y_t = (\theta^0_\tau + \nu - x_t) \mod \nu,\ t = 1, 2, \ldots . \tag{6.12}$$

Повторное применение $L \geqslant 2$ шифров Виженера называется составным шифром Виженера. Пусть есть L шифров Виженера, которые имеют периоды $T_1, \ldots, T_L$ и ключи $\theta^0_1 = (\theta^0_{11}, \ldots, \theta^0_{1T_1})$, …, $\theta^0_L = (\theta^0_{L1}, \ldots, \theta^0_{LT_L})$. Если через $\{\theta^0_{it} : t = 1, 2, \ldots\}$ обозначить ключ θ^0_τ, многократно периодически повторяемый с периодом T_i, то L-составной шифр Виженера имеет вид

$$y_t = (x_t + \theta^0_{1t} + \ \ldots\ + \theta^0_{Lt}) \mod \nu,\ t = 1, 2, \ldots\ . \tag{6.13}$$

Легко показать, что L-составной шифр Виженера можно рассматривать как обычный шифр Виженера с периодом $T = \text{НОК}\{T_1, \ldots, T_L\}$.

4. *Криптопреобразование Вернама* (*поточный шифр*). Криптопреобразование Вернама — специальный частный случай криптопреобразования Виженера (6.8), когда длина используемого ключа равна длине передаваемого сообщения n:

$$y_t = (x_t + \theta^0_t) \mod \nu,\ t = \overline{1, n}\,. \tag{6.14}$$

Обратное криптопреобразование имеет вид

$$x_t = (y_t + \nu - \theta^0_t) \mod \nu,\ t = \overline{1, n}\,.$$

В качестве ключа $\theta^0 = (\theta^0_1, \ldots, \theta^0_n) \in A^n_\nu$ применяется реализация последовательности n независимых, одинаково распределенных на A_ν случайных величин либо некоторая реализация текста. Ключ θ^0 такого типа (используемый всего один раз) в литературе называется бегущей строкой, одноразовым блокнотом (one-time pad) или гаммой. Иногда в качестве $\{\theta^0_t\}$ применяется псевдослучайная последовательность, порождаемая специальным программным датчиком.

5. *Биграммная* (N*-граммная*) *подстановка.* Это криптопреобразование основано на том же принципе, что и простая подстановка (6.4). Однако в отличие от (6.4) в данном случае вместо подстановки одного символа ($A_\nu \leftrightarrow A_\nu$) осуществляется подстановка пар символов (биграмм): $A_\nu^2 \leftrightarrow A_\nu^2$ либо набора N соседних символов (N-грамм): $A_\nu^N \leftrightarrow A_\nu^N$. Ключ θ^0 в биграммной подстановке представляет собой $(\nu \times \nu)$-матрицу, (i,j)-й элемент которой — биграмма (i',j'), заменяющая биграмму (i,j).

В заключение отметим, что если имеются две произвольные криптосистемы T и R, то их часто можно комбинировать для получения новой криптосистемы $S = f(T,R)$.

Наиболее часто используются два следующих типа оператора комбинирования $f(\cdot)$:

— произведение криптосистем

$$S = f_1(T,R) = TR, S = f_2(T,R) = RT,$$

причем, вообще говоря, $TR \neq RT$. В существующих стандартных криптосистемах произведение шифров используется весьма часто. Например, после подстановки применяют транспозицию или после транспозиции — код Виженера;

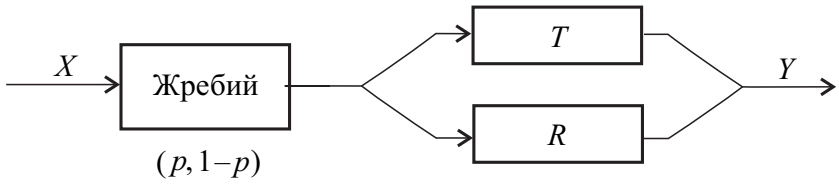


Рис. 6.2. Взвешенная сумма криптосистем

— взвешенная сумма криптосистем

$$S = pT + (1-p)R\,, p \in [0,1].$$

Выбор преобразования T осуществляется с вероятностью p, а преобразования R — с вероятностью $1-p$. Эта функция комбинирования поясняется схемой, приведенной на рис. 6.2.

6.2. ТЕОРЕТИКО-ИНФОРМАЦИОННЫЕ ОЦЕНКИ СТОЙКОСТИ СИММЕТРИЧНЫХ КРИПТОСИСТЕМ

Исследуем общие вопросы устойчивости симметричных криптосистем к криптоанализу, используя теоретико-информационный подход Шеннона. Свойство устойчивости криптосистемы к криптоанализу принято называть криптостойкостью.

Прежде всего рассмотрим вопросы, связанные с потенциальной (т. е. максимально возможной) криптостойкостью. Насколько устойчива криптосистема, если криптоаналитик не ограничен временем и вычислительными средствами для анализа шифртекстов? Имеет ли шифртекст Y единственное решение (т. е.

однозначно ли определяет ключ θ^0)? Какой должна быть минимальная длина шифртекста $n_{\min}$, чтобы решение стало единственным? Существуют ли криптосистемы, в которых нельзя найти единственное решение независимо от длины n исследуемого шифртекста? Существуют ли криптосистемы, в которых криптоаналитик не получает никакой информации, сколько бы шифртекстов он ни перехватил?

Для решения этих проблем К. Шеннон применил аппарат математической теории информации. Основное модельное предположение К. Шеннона об исходном сообщении X состоит в том, что язык источника сообщений может рассматриваться как некоторый вероятностный процесс, порождающий дискретную последовательность символов в соответствии с некоторой системой вероятностей (о моделях сообщений см. в разд. 2). Поэтому исходное сообщение X К. Шеннон предполагал случайным n-вектором с дискретным распределением вероятностей:

$$\mathsf{P}\left\{X = X^{(i)}\right\} = q_i,\ i = \overline{1, \nu^n}\,,$$
$$q_1 + q_2 + \ldots + q_{\nu^n} = 1\,, \tag{6.15}$$

где $X^{(i)} \in A_\nu^n$ — i-й возможный вариант исходного n-символьного сообщения из алфавита A_ν. Ключ $\theta^0 = (\theta_1^0, \ldots, \theta_m^0)$ также предполагается (генерируется) случайным вектором, не зависящим от X, с дискретным распределением вероятностей:

$$\mathsf{P}\left\{\theta^0 = \theta^{(j)}\right\} = p_j\,,\ j = \overline{1, \mu^m}\,,\ \sum_{j=1}^{\mu^m} p_j = 1\,, \tag{6.16}$$

где $\theta^{(j)} \in A_\mu^m$ — j-й возможный вариант ключевой m-символьной последовательности в алфавите A_μ; p_j — *априорная вероятность ключа* $\theta^{(j)}$.

Симметричная криптосистема называется *совершенно криптостойкой*, если апостериорное распределение вероятностей исходного случайного сообщения X при регистрации случайного шифртекста $Y = f(X; \theta^0)$ совпадает с априорным распределением вероятностей:

$$\mathsf{P}\left\{X = X^{(i)} |\ Y = Y^{(l)}\right\} = \mathsf{P}\left\{X = X^{(i)}\right\} = q_i,\ i, l = \overline{1, \nu^n}\,. \tag{6.17}$$

Смысл условия (6.17) в том, что хотя криптоаналитик и имеет шифртекст, он не добавляет ему информации о переданном сообщении.

Теорема 6.1. *Необходимое и достаточное условие совершенной криптостойкости состоит в том, что условное распределение вероятностей шифртекста $Y \in A_\nu^n$ при фиксированном сообщении $X \in A_\nu^n$ не зависит от X:*

$$\mathsf{P}\left\{Y = Y^{(l)} |\ X = X^{(i)}\right\} = \mathsf{P}\left\{Y = Y^{(l)}\right\},\quad i, l = \overline{1, \nu^n}\,. \tag{6.18}$$

Доказательство. По формуле Байеса имеем

$$\mathsf{P}\left\{X = X^{(i)} |\ Y = Y^{(l)}\right\} = \frac{q_i \mathsf{P}\left\{Y = Y^{(l)} |\ X = X^{(i)}\right\}}{\mathsf{P}\left\{Y = Y^{(l)}\right\}}\,.$$

Очевидно, что (6.17) выполняется тогда и только тогда, когда

$$\frac{\mathsf{P}\left\{Y = Y^{(l)} |\ X = X^{(i)}\right\}}{\mathsf{P}\left\{Y = Y^{(l)}\right\}} = 1,$$

что совпадает с (6.18). □

Следствие 6.1. *Если выполняется условие совершенной криптостойкости, то количество информации по Шеннону, содержащейся в шифртексте Y об исходном сообщении X, равно нулю:*

$$\mathsf{I}\{X, Y\} = \mathsf{I}\{Y, X\} = 0\,. \tag{6.19}$$

Доказательство. Вычислим энтропию исходного сообщения X и условную энтропию X относительно шифртекста Y с учетом (6.17):

$$\mathsf{H}\{X\} = -\sum_{i=1}^{\nu^n} q_i \log q_i\,,$$

$$\mathsf{H}\{X|Y\} = -\sum_{i=1}^{\nu^n}\sum_{l=1}^{\nu^n} \mathsf{P}\left\{X = X^{(i)}, Y = Y^{(l)}\right\} \times$$

$$\times \log \mathsf{P}\left\{X = X^{(i)} |\ Y = Y^{(l)}\right\} = -\sum_{i,\,l=1}^{\nu^n} \mathsf{P}\left\{X = X^{(i)}, Y = Y^{(l)}\right\} \times$$

$$\times \log \mathsf{P}\left\{X = X^{(i)}\right\} = -\sum_{i=1}^{\nu^n} q_i \log q_i = \mathsf{H}\{X\}\,.$$

Тогда по определению количества информации имеем

$$\mathsf{I}\{X,Y\} = \mathsf{H}\{X\} - \mathsf{H}\{X|Y\} = 0\,.$$

Второе равенство в (6.19) вытекает из свойства симметричности функционала количества информации. □

Следствие 6.2. *Пусть $\{Y^{(l)} : l = \overline{1, \nu^n}\} = A_\nu^n$ — множество всевозможных шифртекстов, порождаемых криптофункцией $Y = f(X;\,\theta)$;*

$$\mathcal{J}_{il} = \left\{j :\ f\left(X^{(i)}, \theta^{(j)}\right) = Y^{(l)}\right\} \subset N \tag{6.20}$$

есть подмножество номеров ключей, переводящих исходный текст $X^{(i)}$ в один и тот же шифртекст $Y^{(l)}$. Чтобы $f(\cdot)$ удовлетворяла свойству совершенной криптостойкости, необходимо и достаточно, чтобы выполнялось свойство

$$\sum_{j \in \mathcal{J}_{il}} p_j = a_l = \operatorname*{invar}_{1 \leqslant i \leqslant \nu^n}\,. \tag{6.21}$$

Доказательство. Прежде всего отметим, что множество индексов $\mathcal{J}_{il}$ непусто в силу биективности $f(\cdot)$. Вычислим и сравним левую и правую части (6.18) с учетом (6.20) и свойства независимости θ^0 и X:

$$\mathsf{P}\left\{Y = Y^{(l)} |\ X = X^{(i)}\right\} = \mathsf{P}\left\{f(X^{(i)}; \theta^0) = Y^{(l)} |\ X = X^{(i)}\right\} =$$

$$= \sum_{j\in\mathfrak{J}_{il}} \mathsf{P}\left\{\theta^0 = \theta^{(j)} |\ X = X^{(i)}\right\} = \sum_{j\in\mathfrak{J}_{il}} \mathsf{P}\left\{\theta^0 = \theta^{(j)}\right\} = \sum_{j\in\mathfrak{J}_{il}} p_j\,, \tag{6.22}$$

$$\mathsf{P}\left\{Y = Y^{(l)}\right\} = \mathsf{P}\left\{f(X;\theta^0) = Y^{(l)}\right\} =$$

$$= \sum_{i,\,j} q_i p_j \delta_{f(X^{(i)};\,\theta^{(j)}),\,Y^{(l)}} = \sum_{i=1}^{\nu^n} q_i \sum_{j\in\mathfrak{J}_{il}} p_j\,. \tag{6.23}$$

Сравнивая (6.22) и (6.23), заключим справедливость (6.21). □

Теорема 6.2. *Необходимым условием выполнения свойства совершенной криптостойкости является справедливость следующих неравенств энтропий:*

$$\mathsf{H}\left\{\theta^0\right\} \geqslant \mathsf{H}\left\{X\right\}, \tag{6.24}$$

$$\mathsf{H}\left\{\theta^0\right\} \geqslant \mathsf{H}\left\{Y\right\}. \tag{6.25}$$

Доказательство. Как было установлено при доказательстве теоремы 6.1, справедливо равенство

$$\mathsf{H}\left\{X\right\} = \mathsf{H}\left\{X|Y\right\}. \tag{6.26}$$

Воспользуемся свойством иерархической аддитивности энтропии:

$$\mathsf{H}\left\{Y, X, \theta^0\right\} = \mathsf{H}\left\{Y\right\} + \mathsf{H}\left\{X|Y\right\} + \mathsf{H}\left\{\theta^0|X,Y\right\} =$$

$$= \mathsf{H}\left\{Y, \theta^0, X\right\} = \mathsf{H}\left\{Y\right\} + \mathsf{H}\left\{\theta^0|Y\right\} + \mathsf{H}\left\{X|\theta^0,Y\right\}. \tag{6.27}$$

Поскольку при фиксированном шифртексте Y и ключе θ^0 исходный текст $X = f^{-1}(Y;\,\theta^0)$ неслучаен, то $\mathsf{H}\left\{X|\theta^0,Y\right\} = 0$. Поэтому из уравнения (6.27) найдем $\mathsf{H}\left\{X|Y\right\} = \mathsf{H}\left\{\theta^0|Y\right\} - \mathsf{H}\left\{\theta^0|X,Y\right\}$. По свойствам энтропии отсюда следует

$$\mathsf{H}\left\{X|Y\right\} \leqslant \mathsf{H}\left\{\theta^0|Y\right\} \leqslant \mathsf{H}\left\{\theta^0\right\}.$$

Используя это в (6.26), придем к (6.24). Неравенство (6.25) доказывается аналогично. □

Теорему 6.2 в криптологии называют пессимистическим утверждением К. Шеннона, так как она требует, чтобы энтропия (неопредленность) ключа θ^0 была не меньше энтропии исходного текста X (или шифртекста Y). Поскольку распределение $\{q_i\}$ исходного текста X может быть произвольным, а $\max\limits_{\{q_i\}} \mathsf{H}\left\{X\right\} = \log \nu^n$, то неравенство (6.24) примет вид

$$\mathsf{H}\left\{\theta^0\right\} \geqslant \log \nu^n.$$

Для его выполнения в случае $\mu = \nu$ требуется, чтобы длина ключа m была не меньше длины шифруемого текста: $m \geqslant n$. Для практики наиболее интересен случай самых коротких ключей: $m = n$.

Теорема 6.3. *Если* $\mu{=}\nu$, $m{=}n$ *и для любых* $i, l \in \{1, 2, \ldots, \nu^n\}$ *уравнение*

$$f\left(X^{(i)}; \theta^{(j)}\right) = Y^{(l)} \tag{6.28}$$

имеет единственное решение $j = j_{il}$ (т. е. $\mathfrak{J}_{il} = \{j_{il}\}$ — одноточечное множество), то необходимым и достаточным условием совершенной криптостойкости является равновероятность используемых ключей

$$p_j \equiv \text{const} = \frac{1}{\boldsymbol{\nu}^n}\,,\ j = \overline{1, \boldsymbol{\nu}^n}. \tag{6.29}$$

Доказательство. Согласно (6.21)

$$p_{j_{il}} = a_l \ \forall\, i, j, l\,. \tag{6.30}$$

Очевидно (в силу биекции), что j_{il} зависит от i и l так, что при изменении $i \in \{1, 2, \ldots, \boldsymbol{\nu}^n\}$ индекс j_{il} «пробегает» $\boldsymbol{\nu}^n$ всевозможных значений. Поэтому (6.30) невозможно без выполнения (6.29). □

Следствие 6.3. *Криптопреобразование Вернама (6.14) при условии равновероятности ключей (6.26) обладает свойством совершенной криптостойкости.*

Доказательство. В силу (6.14) уравнение (6.28) имеет единственное решение для любых $X^{(i)}, Y^{(l)}$: $\theta^{(j)} = \left(Y^{(l)} + \boldsymbol{\nu} - X^{(i)}\right) \bmod \boldsymbol{\nu}$, где вычеты вычисляются покомпонентно. Поэтому из (6.26) и теоремы 6.3 получим доказываемый результат. □

Данное следствие объясняет, почему для передачи и защиты наиболее важной информации широко используются поточные криптосистемы, базирующиеся на криптопреобразовании Вернама. Следствие объясняет также, почему к качеству генерации ключевой последовательности $\theta^0 = (\theta_1^0, \theta_2^0, \ldots, \theta_n^0)$ (гаммы) предъявляются столь высокие требования.

Рассмотрим еще одну важную характеристику криптосистем (связанную с криптостойкостью), введенную К. Шенноном, — расстояние единственности U. Для этого воспользуемся следующими соотношениями:

$$\begin{aligned}
&\mathbf{H}\left\{\theta^0, X\right\} = \mathbf{H}\left\{\theta^0\right\} + \mathbf{H}\left\{X\right\} \ \text{(по условию независимости } \theta^0, X),\\
&\mathbf{H}\left\{\theta^0 | X\right\} = \mathbf{H}\left\{Y | X\right\} \ \text{(по свойствам криптосистемы)},\\
&\mathbf{H}\left\{\theta^0, X\right\} = \mathbf{H}\left\{Y, X\right\} \ \text{(вытекает из предыдущего)},\\
&\mathbf{H}\left\{Y | X\right\} = \mathbf{H}\left\{Y\right\} + \mathbf{H}\left\{X | Y\right\} \ \text{(по свойству энтропии)}.
\end{aligned}$$

Отсюда получим выражение для условной энтропии исходного текста X относительно наблюдаемого шифртекста Y:

$$\mathbf{H}\left\{X | Y\right\} = \mathbf{H}\left\{X\right\} - \left(\mathbf{H}\left\{Y\right\} - \mathbf{H}\left\{\theta^0\right\}\right).$$

Поскольку условная энтропия неотрицательна, имеем неравенство

$$\mathbf{H}\left\{X | Y\right\} = \mathbf{H}\left\{X\right\} - \left(\mathbf{H}\left\{Y\right\} - \mathbf{H}\left\{\theta^0\right\}\right) \geqslant 0\,. \tag{6.31}$$

Расстояние единственности — такая минимальная длина шифртекста Y (и исходного текста), при которой исчезает неопределенность в исходном тексте X при наблюдении шифртекста Y:

$$U = \min\{n:\ \mathbf{H}\left\{X | Y\right\} = 0\}\,. \tag{6.32}$$

Следуя М. Хеллману (см. разд. 4.2, 4.3), построим оценки энтропий, входящих в (6.31) и (6.32):

$$\mathbf{H}\{X\} = n\log\nu_X,\ \ \mathbf{H}\{Y\} = n\log\nu_Y,\ \ \mathbf{H}\left\{\theta^0\right\} = \log|\Theta|\,, \tag{6.33}$$

где ν_X (ν_Y) — число, подбираемое так, что приближенно ν_X^n (ν_Y^n) реализаций исходных текстов X (шифртекстов Y) имеют вероятности, значительно отличающиеся от нуля, а остальные реализации — пренебрежимо малую вероятность; $|\Theta|$ — мощность пространства используемых ключей; $\nu_X \leqslant \nu_Y < \nu$.

Подставляя (6.33) в (6.31) и (6.32), найдем приближенное выражение для расстояния единственности:

$$U = \frac{\log|\Theta|}{\log(\nu_Y/\nu_X)}\,. \tag{6.34}$$

Для реальных криптосистем обычно оказывается, что шифртекст Y имеет распределение, близкое к равномерному, поэтому $\nu_Y \approx \nu$. Тогда формула (6.34) примет следующий вид:

$$U = \frac{\log|\Theta|}{k}\,,\ k = \log\frac{\nu}{\nu_X} > 0\,,$$

где k — коэффициент, характеризующий избыточность языка. Например, для текстов на английском, немецком и французском языках этот коэффициент приблизительно одинаков и равен $k \approx 0{,}53$.

В целом отметим, что основные задачи криптографии в настоящее время заключаются в обеспечении:

— конфиденциальности;

— проверки авторства;

— целостности.

Для этого активно используются методы теории информации [14].

6.3. ЗАДАНИЯ ДЛЯ ТЕСТОВ

6.1. Какой из перечисленных ниже шифров не является шифром подстановки символов алфавита:

а) аффинный; б) простой замены;
в) перестановочный; г) Цезаря;
д) Бофора?

6.2. Какой из шифров — поточный:

а) Хилла; б) Вернама;
в) перестановочный; г) Цезаря;
д) Бофора?

6.3. В каком из шифров ключ не всегда можно записать с помощью символов алфавита открытых текстов:

а) Виженера; б) Вернама;
в) перестановочный; г) Цезаря;
д) Бофора?

6.4. Согласно «пессимистическому утверждению Шеннона», если криптосистема совершенна:

а) $\mathbf{H}\{X\} \geqslant \mathbf{H}\{X|Y\}$;
б) существует такая константа C, что расстояние единственности $U \leqslant C$;
в) $\mathbf{I}\{X,Y\} > 0$;
г) $\exists z \in \Theta$, что $\mathbf{P}\{\theta = z\} = 0$;
д) $\mathbf{H}\{\theta^0\} \geqslant \mathbf{H}\{X\}$, $\mathbf{H}\{\theta^0\} \geqslant \mathbf{H}\{Y\}$.

6.5. Расстояние единственности — это такое:

а) наименьшее n, что $\mathbf{H}\{X|Y\} = 0$, $X, Y \in A_{\nu}^{n}$;
б) наименьшее n, что $\mathbf{H}\{X\} = 0$, $X \in A_{\nu}^{n}$;
в) наибольшее n, что $\mathbf{H}\{X|Y\} > 0$, $X, Y \in A_{\nu}^{n}$;
г) наименьшее n, что $\mathbf{H}\{\theta|X\} = 0$, $X \in A_{\nu}^{n}$, $\theta \in \Theta$;
д) наибольшее n, что $\mathbf{H}\{Y\} > 0$, $Y \in A_{\nu}^{n}$.

6.4. РЕШЕНИЕ ТИПОВЫХ ЗАДАЧ

Задача 6.1. Зашифровать сообщение «ПРИМЕР» с помощью шифра Цезаря с ключом $\theta^0 = 20$ для русского алфавита.

Решение. Напомним, что функция зашифрования одного символа сообщения в шифре Цезаря определяется следующим образом:

$$y_t = (x_t + \theta^0) \bmod \nu, t = 1, 2, \ldots . \tag{6.35}$$

Для русского алфавита $\nu = 33$. Для того чтобы воспользоваться формулой (6.35), необходимо все символы естественного языка перевести в числа. Традиционно это осуществляется таким образом: первой букве алфавита ставится в соответствие число 0, второй — 1 и т. д., последней — число $\nu - 1$. В таблице показано соответствие чисел и символов русского алфавита.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
А	Б	В	Г	Д	Е	Ё	Ж	З	И	Й	К	Л	М	Н	О	П
17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	
Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я	

Согласно таблице сообщение «ПРИМЕР» представляется в виде последовательности чисел: $x_t = (16, 17, 9, 13, 5, 17)$.

Применим к x_t функцию зашифрования (6.35):

$$y_1 = (x_1 + 20) \bmod 33 = (16 + 20) \bmod 33 = 36 \bmod 33 = 3, \ldots .$$

Таким образом, $y_t = (3, 4, 29, 0, 25, 4)$. Чтобы получить результирующий шифртекст, необходимо с помощью таблицы номера символов y_t снова перевести в символы естественного алфавита.

Ответ: «ГДЬАШД».

Задача 6.2. Функция зашифрования аффинного шифра задается соотношением

$$y_t = (a^0 x_t + b^0) \bmod \nu, t = 1, 2, \ldots, \tag{6.36}$$

где ключом является пара $(a^0, b^0) \in \mathbb{Z}_\nu$, причем НОД чисел $(a^0, \nu) = 1$. Был получен шифртекст «СЦПГЬЦ». Расшифровать это сообщение, если известно, что использовался ключ $a^0 = 5$, $b^0 = 4$.

Решение. Чтобы найти исходное значение x_t в (6.36), необходимо выполнить следующие преобразования:

$$y_t = (a^0 x_t + b^0) \bmod \nu \Rightarrow x_t = (y_t - b^0)(a^0)^{-1} \bmod \nu, t = 1, 2, \ldots . \tag{6.37}$$

Заметим, что поскольку $(a^0, \nu) = 1$, то существует единственный обратный элемент $(a^0)^{-1}$. Для нахождения обратного элемента можно воспользоваться теоремой Эйлера:

$$a^{\varphi(\nu)} = 1 \bmod \nu \Rightarrow a^{-1} = a^{\varphi(\nu)-1} \bmod \nu.$$

В данном случае $a^0 = 5$, $\varphi(33) = 20$, откуда $a^{-1} = a^{19} \bmod 33 = 20 \bmod 33$. Далее, как и при решении предыдущей задачи, поставим в соответствие всем символам шифртекста числа согласно таблице. Получим $y_t = (18, 23, 16, 3, 29, 23)$. Применим к полученной последовательности преобразование (6.37):

$$x_1 = (18 - 4) * 20 \bmod 33 = 280 \bmod 33 = 16, \ldots,$$

$$x_6 = (23 - 4) * 20 \bmod 33 = 380 \bmod 33 = 17.$$

Таким образом, имеем $x_t = (16, 17, 9, 13, 5, 17)$. Преобразовав полученные числа в символы согласно таблице, получим исходное сообщение.

Ответ: «ПРИМЕР».

Задача 6.3. Шифртекст «КТОРПИСЕМИТС» был получен с помощью шифра перестановки символов с ключом $\theta^0 = (1, 5, 6, 2, 4, 3)$. Восстановить исходное сообщение.

Решение. По размеру ключа определим, во-первых, длину периода $T = 6$, во-вторых, подстановку на множестве $\{1, 2, \ldots, 6\}$:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 5 & 6 & 2 & 4 & 3 \end{pmatrix}. \tag{6.38}$$

Для расшифрования исходного сообщения необходимо обратить подстановку (6.38):

$$\sigma^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 4 & 6 & 5 & 2 & 3 \end{pmatrix}.$$

Тогда преобразование расшифрования имеет вид

$$x_t = y_{(i-1)T+\bar{\theta}^0_\tau},$$

где $\bar{\theta}^0_\tau$ — элементы подстановки обратной к (6.38). Таким образом, в нашем случае

$$x_1 = y_{\bar{\theta}^0_1} = y_1 = \text{«К»},\ x_2 = y_{\bar{\theta}^0_2} = y_4 = \text{«Р»}, x_3 = y_{\bar{\theta}^0_3} = y_6 = \text{«И»},$$
$$x_4 = y_{\bar{\theta}^0_4} = y_5 = \text{«П»},\ x_5 = y_{\bar{\theta}^0_5} = y_2 = \text{«Т»}, x_6 = y_{\bar{\theta}^0_6} = y_3 = \text{«О»},$$
$$x_7 = y_{6+\bar{\theta}^0_1} = y_{6+1} = \text{«С»}, \ldots, x_{12} = y_{6+\bar{\theta}^0_6} = y_{6+3} = \text{«М»}.$$

Ответ: «КРИПТОСИСТЕМ».

Задача 6.4. Пусть $A_\nu = A_\mu = \{1,2,3\}$ и функция зашифрования определяется соотношением $E_\theta(x) = L_{\theta,x}$, где

$$L = \begin{pmatrix} 2 & 3 & 1 \\ 3 & 1 & 2 \\ p & q & 3 \end{pmatrix},\ p,q \in A_\nu. \tag{6.39}$$

Определить, при каких значениях p и q описанная криптосистема будет совершенной в предположении, что все ключи равновероятны и для любого $x \in \{1,2,3\}$ вероятность $\mathsf{P}\{X = x\} > 0$.

Решение. Для начала заметим, что матрица (6.39) задает функцию зашифрования только в том случае, если $3 \neq p \neq q \neq 3$, так как в противном случае при $\theta = 3$ получаемое преобразование будет необратимым. Таким образом, имеем два возможных варианта: 1) $p = 1$, $q = 2$; 2) $p = 2$, $q = 1$.

По определению криптосистема называется совершенной, если имеет место равенство

$$\mathsf{P}\{X = x\} = \mathsf{P}\{X = x | Y = y\}.$$

Проверим, выполнено ли это условие в каждом из возможных случаев.

1) Пусть $p = 1$, $q = 2$, тогда матрица (6.39) примет вид

$$L = \begin{pmatrix} 2 & 3 & 1 \\ 3 & 1 & 2 \\ 1 & 2 & 3 \end{pmatrix},$$

отсюда, используя формулу Байеса, получим

$$\mathsf{P}\{X = 1 | Y = 1\} =$$
$$= \frac{\mathsf{P}\{Y = 1 | X = 1\}\,\mathsf{P}\{X = 1\}}{\mathsf{P}\{Y = 1 | X = 1\}\,\mathsf{P}\{X = 1\} + \ldots + \mathsf{P}\{Y = 1 | X = 3\}\,\mathsf{P}\{X = 3\}} =$$
$$= \frac{\mathsf{P}\{\theta = 3\}\,\mathsf{P}\{X = 1\}}{\mathsf{P}\{\theta = 3\}\,\mathsf{P}\{X = 1\} + \mathsf{P}\{\theta = 2\}\,\mathsf{P}\{X = 2\} + \mathsf{P}\{\theta = 1\}\,\mathsf{P}\{X = 3\}} =$$
$$= \frac{1/3\mathsf{P}\{X = 1\}}{1/3\mathsf{P}\{X = 1\} + 1/3\mathsf{P}\{X = 2\} + 1/3\mathsf{P}\{X = 3\}} = \mathsf{P}\{X = 1\}.$$

Аналогично проверяются все остальные пары возможных значений X и Y. Нетрудно проверить, что в этом случае криптосистема будет совершенной.

2) При $p = 2$, $q = 1$ матрица (6.39) примет вид

$$L = \begin{pmatrix} 2 & 3 & 1 \\ 3 & 1 & 2 \\ 2 & 1 & 3 \end{pmatrix}.$$

Из анализа полученной матрицы можно сделать вывод, что $\mathsf{P}\{X = 1|Y = 1\} = 0$, так как $\mathsf{P}\{Y = 1|X = 1\} = 0$. Следовательно, $\mathsf{P}\{X = 1\} \neq \mathsf{P}\{X = 1|Y = 1\}$, поскольку по условию $\mathsf{P}\{X = 1\} > 0$.

Ответ: $p = 1$, $q = 2$.

Задача 6.5. Вычислить расстояние единственности шифра Цезаря для некоторого языка, состоящего из 16 символов с относительной избыточностью языка $R_L = 0{,}4$.

Решение. Для того чтобы получить коэффициент k, характеризующий полную избыточность языка, необходимо умножить коэффициент относительной избыточности на двоичный логарифм размера алфавита. Получим $k = R_L \log_2 16 = 1{,}5$. В таком случае расстояние единственности может быть найдено по формуле

$$U = \frac{\log_2 |\Theta|}{k},$$

где Θ — множество всех возможных ключей. В случае шифра Цезаря на 16-символьном алфавите $\Theta = \mathbb{Z}_{16}$, откуда $|\Theta| = 16$. Тогда окончательно

$$U = \lceil \frac{\log_2 16}{1{,}5} \rceil = \lceil 2\frac{2}{3} \rceil = 3.$$

Ответ: 3.

6.5. ЗАДАЧИ И УПРАЖНЕНИЯ

6.1. Расшифровать сообщение Y, зашифрованное с помощью аффинного шифра $Y_t = (aX_t + b) \bmod 33$, если известен ключ $\theta = (a, b)$:

а) $Y =$«ФРАПФРАВЫР», $a = 13$, $b = 17$;

б) $Y =$«МЫЗИАЦЯЩЩЗП», $a = 19$, $b = 2$;

в) $Y =$«ЭРЛФЬМЪУЭЦ», $a = 4$, $b = 27$;

г) $Y =$«УБШОУЩФЩЫЩЕМЩ», $a = 5$, $b = 1$;

д) $Y =$«МИЧРЖИКЛРГКФЗЧ», $a = 29$, $b = 11$;

е) $Y =$«ЫАФНБЩУЧБ», $a = 10$, $b = 9$.

6.2. Расшифровать сообщение Y, зашифрованное с помощью шифра Виженера, если известен ключ θ:

а) $Y =$«ЮЬУСЭРЙН», $\theta =$«БОБ»;

б) $Y =$«СРУВЕТЩЧГТЬ», $\theta =$«АЛИСА»;

в) Y =«УУЯТЭЪХЕЩЭРУ», θ =«КЛЮЧ»;
г) Y =«ЗЧШВКИВЯЪУФ», θ =«ШИФР»;
д) Y =«РУОВЯСИЩРЪЧ», θ =«ЯЗЫК»;
е) Y =«ВЧВЬУВЫЯЪАЙЕ», θ =«ЧИСЛО».

6.3. С помощью шифра перестановки символов зашифровать сообщение X с ключом θ (если длина сообщения не кратна длине ключа, то добавить в конец сообщения незначащие символы «Ъ»):
а) X =«ПРОСЛУШИВАНИЕ», $\theta = (3, 1, 4, 2)$;
б) X =«ФАКТОРИЗАЦИЯ», $\theta = (4, 2, 1, 3)$;
в) X =«ПОСЛЕДОВАТЕЛЬНОСТЬ», $\theta = (5, 6, 1, 3, 4, 2)$.

6.4. Известно, что сообщение Y было получено с помощью шифра перестановки символов. Расшифровать сообщение если известен ключ θ:
а) Y =«ОСОБНЩЕИЪЕЪЪ», $\theta = (3, 1, 2, 4)$;
б) Y =«АРАНГМЪАМА», $\theta = (3, 5, 1, 2, 4)$;
в) Y =«ЕТСЕРПВАКНОА», $\theta = (4, 6, 5, 2, 3, 1)$.

6.5. Один из примеров биграммного шифра — это шифр Хилла. Ключом является матрица $\theta \in A_{\nu}^{2\times 2}$ размерностью 2×2. Преобразование зашифрования задается как

$$(Y_{2t+1}, Y_{2t+2}) = (X_{2t+1}, X_{2t+2})\theta \bmod \nu.$$

С помощью шифра Хилла зашифровать сообщение X на ключе θ:
а) X =«ТЕОРИЯ», $\theta = \begin{pmatrix} 18 & 8 \\ 14 & 11 \end{pmatrix}$;
б) X =«РАССТОЯНИЕ», $\theta = \begin{pmatrix} 30 & 10 \\ 4 & 7 \end{pmatrix}$;
в) X =«ЕДИНСТВЕННОСТЬ», $\theta = \begin{pmatrix} 31 & 1 \\ 24 & 7 \end{pmatrix}$.

6.6. Известно, что сообщение Y было получено с помощью шифра Хилла. Расшифровать сообщение, если известен ключ θ:
а) Y =«ГВТЪУЫ», $\theta = \begin{pmatrix} 5 & 12 \\ 11 & 28 \end{pmatrix}$;
б) Y =«ЪЕВХЩЧ», $\theta = \begin{pmatrix} 1 & 7 \\ 15 & 22 \end{pmatrix}$;
в) Y =«РФЬАЬВЖДХА», $\theta = \begin{pmatrix} 32 & 3 \\ 11 & 29 \end{pmatrix}$.

6.7. Вычислить количество различных ключей шифра Хилла для алфавита $A_3 = \{0, 1, 2\}$.

6.8. Таблицей шифрования криптосистемы назовем прямоугольную таблицу, строки которой занумерованы всевозможными открытыми тестами, столбцы — ключами шифрования, а на пересечении строки x и столбца θ находится

шифртекст y такой, что $y = E_\theta(x)$. Доказать, что в любом столбце таблицы шифрования все элементы различны.

6.9. Доказать, что таблица шифрования любой совершенной криптосистемы является латинским квадратом, т. е. в каждой строке и каждом столбце такой таблицы все символы обязательно различные.

6.10. Построить таблицу шифрования для следующей системы: $A_3 = \{0,1,2\}$, $\Theta = A_5 = \{0,1,2,3,4\}$, все открытые тексты и ключи равновероятны, т. е. $\mathbf{P}\{X = x\} = 1/3$, $\mathbf{P}\{\theta = z\} = 1/5$ для любых $x \in A_3$, $z \in \Theta$. Шифрование определяется так: $Y = E_\theta(X) = X + \theta \bmod 3$, $X = D_\theta(Y) = Y - \theta \bmod 3$. Найти вероятности $p(y|x) = \mathbf{P}\{Y = y|X = x\}$, $p(x|y) = \mathbf{P}\{X = x|Y = y\}$.

6.11. Определить, при каких значениях параметров m и n следующая криптосистема является совершенной. Пусть $X, Y \in A_n = \{0,1,\ldots,n-1\}$, $\Theta = A_m = \{0,1,\ldots,m-1\}$, все открытые тексты и ключи равновероятны, т. е. $\mathbf{P}\{X = x\} = 1/n$, $\mathbf{P}\{\theta = z\} = 1/m$ для любых $x \in A_n$, $z \in \Theta$. Шифрование определяется так: $Y = E_\theta(X) = X + \theta \bmod n$, $X = D_\theta(Y) = Y - \theta \bmod n$. Найти вероятности $p(y|x) = \mathbf{P}\{Y = y|X = x\}$, $p(x|y) = \mathbf{P}\{X = x|Y = y\}$.

6.12. Определить, при каких целых параметрах a и b следующая криптосистема является совершенной. Пусть $n \in \mathbb{N}$, $X, Y, \theta \in A_n = \{0,1,\ldots,n-1\}$, все открытые тексты и ключи равновероятны, т. е. $\mathbf{P}\{X = x\} = \mathbf{P}\{\theta = z\} = 1/n$ для любых $X, \theta \in A_n$. Функция зашифрования $Y = E_\theta(X) = aX + b\theta \bmod n$.

6.13. Криптосистему назовем строго идеальной, если $\mathbf{H}\{\theta\} = \mathbf{H}\{\theta|Y\}$. Верно ли, что любая совершенная криптосистема является строго идеальной?

6.14. Криптосистему назовем строго идеальной, если $\mathbf{H}\{\theta\} = \mathbf{H}\{\theta|Y\}$. Верно ли, что любая строго идеальная криптосистема является совершенной?

6.15. Пусть криптосистема определена так: $X, Y, \theta \in A_2 = \{0,1\}$, $Y = E_\theta(X) = X \oplus \theta$. При каких распределениях $\mathbf{P}\{X = x\}$, $\mathbf{P}\{\theta = z\}$ криптосистема может быть строго идеальной (см. задачи 6.13 — 6.14), но не совершенной? Совершенной, но не строго идеальной?

6.16. Вычислить расстояние единственности для русского (коэффициент относительной избыточности $R_L = 0{,}73$) и английского ($R_L = 0{,}75$) языков:

а) шифра простой замены; б) аффинного шифра;

в) шифра Цезаря; г) шифра «одноразовый блокнот».

Глава 7

ОПТИМАЛЬНОЕ КОДИРОВАНИЕ

7.1. АЛФАВИТНОЕ КОДИРОВАНИЕ

В этой главе рассматриваются методы преобразования (кодирования) последовательностей, порождаемых дискретным источником, с целью их оптимального представления. При этом оптимальность будем понимать в смысле наименьшей средней длины последовательностей, получаемых в результате кодирования. Такое оптимальное представление важно и для хранения данных, поскольку в этом случае можно сохранить наибольшее количество информации, и для передачи данных, так как чем меньше символов передается, тем меньше вероятность того, что сообщение будет передано неверно.

Пусть даны два конечных множества: $A = \left\{a^{(1)}, \ldots, a^{(m)}\right\}$ — алфавит источника сообщений мощности m и $B = \left\{b^{(1)}, \ldots, b^{(D)}\right\}$ — кодовый алфавит мощности D. Обозначим через $B^* = \bigcup_{n \geqslant 1} B^n$ множество всех конечных слов (последовательностей) в алфавите B.

Определение 7.1. *Алфавитным D-ичным кодированием называется произвольное отображение $\varphi : A \to B^*$. При этом образ $\varphi(a^{(i)}) \in B^*$ буквы $a^{(i)} \in A$ называется кодовым словом или результатом кодирования буквы $a^{(i)}$, а длина этого кодового слова обозначается как $l^{(i)} = \left|\varphi(a^{(i)})\right|$.*

Набор кодовых слов $\varphi(A) = \left\{\varphi(a^{(1)}), \ldots, \varphi(a^{(m)})\right\}$ называется D-ичным кодом для алфавита A. Если при этом все кодовые слова имеют одинаковую длину, кодирование φ называется *равномерным* или *блоковым*, в противном случае — *неравномерным*.

Алфавитное кодирование φ можно продолжить на множество $A^* = \bigcup_{n \geqslant 1} A^n$ всех конечных слов в алфавите A по правилу сцепления (приписывания, конкатенации) кодовых слов: $\varphi^* : A^* \to B^*$ и $\forall n \in \mathbb{N}$, $\forall a^n = (a_1, \ldots, a_n) \in A^n$ имеем $\varphi^*(a_1, \ldots, a_n) = \varphi(a_1) || \ldots || \varphi(a_n)$.

Определение 7.2. *Алфавитное кодирование φ (и набор кодовых слов $\varphi(A)$) называется префиксным, если никакое кодовое слово $\varphi(a^{(i)})$ не является началом какого-либо другого кодового слова $\varphi(a^{(j)})$, $i \neq j$, т. е. если $\varphi(a^{(j)}) = b_1 b_2 \ldots b_{l^{(j)}}$ $\left(b_t \in B,\ t \in \{1, \ldots, l^{(j)}\}\right)$, то для любого $k \in \{1, \ldots, l^{(j)}\}$ не существует такого $a^{(i)} \in A$, $i \neq j$, что $\varphi(a^{(i)}) = b_1 b_2 \ldots b_k$.*

Определение 7.3. *Алфавитное кодирование* φ *(и набор кодовых слов* $\varphi(A)$*) называется суффиксным, если никакое кодовое слово* $\varphi(a^{(i)})$ *не является окончанием какого-либо другого кодового слова* $\varphi(a^{(j)})$, $i \neq j$, *т. е. если* $\varphi(a^{(j)}) = b_1 b_2 \dots b_{l(j)}$ $\left(b_t \in B,\ t \in \{1, \dots, l^{(j)}\}\right)$, *то для любого* $k \in \{1, \dots, l^{(j)}\}$ *не существует такого* $a^{(i)} \in A$, $i \neq j$, *что* $\varphi(a^{(i)}) = b_k b_{k+1} \dots b_{l(j)}$.

Замечание 7.1. Из приведенных выше двух определений следует свойство: если алфавитное кодирование φ является префиксным или суффиксным, то отображение φ инъективное.

Определение 7.4. *Алфавитное кодирование* φ *(и набор кодовых слов* $\varphi(A)$*) называется однозначно декодируемым или разделимым, если отображение* φ^* *является инъективным.*

Можно дать и другое определение однозначно декодируемого кодирования. Алфавитное кодирование φ (и набор кодовых слов $\varphi(A)$) называется однозначно декодируемым, если для любой последовательности $a^n = (a_1, \dots, a_n)$ соответствующая кодовая последовательность $\varphi^*(a_1, \dots, a_n)$ единственным образом разбивается на кодовые слова $\varphi(a_1) || \dots || \varphi(a_n)$ из кода $\varphi(A)$.

Упражнение 7.1. Доказать, что оба определения однозначно декодируемого кодирования эквивалентны, т. е. из условий первого определения следуют условия второго определения и наоборот.

Пример 7.1. Пусть $m = D = 2$, $A = B = \{0,1\}$. В табл. 7.1 приведено несколько различных вариантов алфавитного кодирования, обладающих различными свойствами.

Таблица 7.1

Примеры кодирования

A	φ_1	φ_2	φ_3	φ_4	φ_5
0	0	0	0	00	0
1	01	10	010	11	00
Свойства					
Префиксное	$-$	$+$	$-$	$+$	$-$
Суффиксное	$+$	$-$	$-$	$+$	$-$
Однозначно декодируемое	$+$	$+$	$+$	$+$	$-$

Выясним, как связаны между собой префиксное (суффиксное) и однозначно декодируемое кодирование.

Утверждение 7.1. Всякое префиксное или суффиксное кодирование является однозначно декодируемым. Обратное утверждение неверно.

Доказательство. Пусть алфавитное кодирование φ префиксное. Рассмотрим произвольную последовательность $a^n = (a_1, \dots, a_n) \in A^*$ и покажем, что соответствующую кодовую последовательность $\varphi^*(a_1, \dots, a_n)$ можно единственным образом разбить на кодовые слова $\varphi(a_1) || \dots || \varphi(a_n)$, тогда по определению алфавитное кодирование φ будет однозначно декодируемым.

Предположим, что кодовую последовательность $\varphi^*(a_1, \ldots, a_n)$ можно разбить на кодовые слова еще одним способом: $\varphi(a'_1)||\ldots||\varphi\left(a'_{n'}\right)$, $a'_i \in A$, $i \in \{1, \ldots, n'\}$ и $\exists i \leqslant \min\{n, n'\}$, что $a_i \neq a'_i$. Пусть k равно минимальному i такому, что $a_i \neq a'_i$. Обозначим $l = |\varphi(a_k)|$, $l' = |\varphi(a'_k)|$. Получим

$$\varphi(a_1)||\ldots||\varphi(a_n) = \varphi^*(a_1, \ldots, a_n) = \varphi\left(a'_1\right)||\ldots||\varphi\left(a'_{n'}\right).$$

Отбросим первые совпадающие кодовые слова:

$$\varphi(a_k)||\ldots||\varphi(a_n) = \varphi\left(a'_k\right)||\ldots||\varphi\left(a'_{n'}\right). \tag{7.1}$$

Рассмотрим три возможные ситуации:

1) $l = l'$, тогда из (7.1) следует, что $\varphi(a_k) = \varphi(a'_k)$, т. е. φ не инъективно (согласно предположению $a_k \neq a'_k$), а это противоречит свойству префиксного алфавитного кодирования;

2) $l < l'$. В этом случае из (7.1) следует, что код $\varphi(a_k)$ является началом кода $\varphi(a'_k)$, а это противоречит определению префиксного алфавитного кодирования;

3) $l > l'$. Если так, то из (7.1) следует, что код $\varphi(a'_k)$ — начало кода $\varphi(a_k)$, а это снова противоречит определению префиксного алфавитного кодирования.

Таким образом, во всех трех случаях мы получаем противоречие, а значит, исходное предположение неверно и, следовательно, разложение кодовой последовательности на кодовые слова единственно.

Для суффиксного алфавитного кодирования доказательство производится аналогично, за исключением того, что рассмотрение кодовых слов надо проводить справа налево.

В качестве примера однозначно декодируемого кодирования, которое не является ни префиксным, ни суффиксным, можно взять кодирование φ_3 из примера 7.1. □

7.2. КОДОВЫЕ ДЕРЕВЬЯ

В данном разделе предполагается, что студент знаком с основами теории графов. Если это не так, то хотя здесь и излагаются определения и свойства, но для лучшего понимания материала сначала следует ознакомиться с соответствующей теорией [2, 11, 15] и только затем продолжить изучение данного раздела.

Определение 7.5. *Пусть V – конечное множество, E – множество неупорядоченных пар элементов из V, тогда упорядоченная пара $G = (V, E)$ называется графом или неориентированным графом. Элементы множества V — вершины графа, а элементы множества E – ребра.*

Для неориентированного графа записи $\{u, v\}$ и $\{v, u\}$ $(u, v \in V)$ обозначают одно и тоже ребро, соединяющее вершины u и v. *Смежные* вершины — две вершины графа, соединенные ребром. Вершины, которые соединены ребром, называются его *концами*. Если вершина является концом ребра, то будем говорить, что ребро *выходит* из вершины. Число ребер, выходящих из вершины v, — *степень* вершины v. Вершина степени 0 — *изолированная*, степени 1 — *висячая*.

Определение 7.6. *Путь в графе между вершинами u и v — это такая последовательность чередующихся вершин и ребер графа u, $\{u,w_1\}$, w_1, $\{w_1,w_2\}$, $\ldots$, $\{w_{k-1},w_k\}$, w_k, $\{w_k,v\}$, v, где каждое ребро соединяет вершины последовательности, между которыми оно находится. Цепь в графе — это путь, все ребра которого различны. Цепь графа называется простой, если все ее вершины, кроме, возможно, крайних, различны. Цикл — это цепь в графе, в котором первая и последняя вершина совпадают.*

Определение 7.7. *Граф связный, если в нем между любыми двумя вершинами существует путь.*

Определение 7.8. *Дерево — это связный граф, не содержащий циклов.*

Упражнение 7.2. Доказать, что для графа G, имеющего n вершин и m ребер, следующие утверждения эквивалентны:

— G — дерево;

— G — связный граф и $m = n - 1$;

— G — граф без циклов и $m = n - 1$;

— любые две несовпадающие вершины графа G соединяет единственная простая цепь.

Определение 7.9. *Дерево называется корневым, если в нем выделена некоторая вершина $v_0 \in V$ — корень.*

Для корневого дерева можно определить понятие *уровня* вершины. Нулевой уровень содержит единственную вершину — корень v_0. Пусть определены уровни $0 \leqslant i \leqslant k$, тогда $(k+1)$-й уровень содержит все вершины, смежные с вершинами уровня k и не принадлежащие уровню $k-1$. *Высота вершины* — уровень, которому принадлежит вершина. *Высота корневого дерева* — максимальный уровень среди всех вершин этого дерева.

Обозначим через V_k множество всех вершин корневого дерева G, принадлежащих k-му уровню.

Если вершины u и v связаны ребром и при этом $u \in V_{k-1}$, $v \in V_k$, то вершина u — *предок* вершины v, а вершина v — *потомок* вершины u; при этом ребро (u,v) называется исходящим из вершины u. Если вершина v не имеет потомков, то это *концевая или висячая* вершина, или *лист*.

Если G — корневое дерево, $v \in V_k$, то поддеревом G_v с корнем v называется такой граф, у которого множество вершин состоит из вершины v, всех его потомков $u_1, \ldots, u_c$ на $(k+1)$-уровне, всех потомков вершин $u_1, \ldots, u_c$ на $(k+2)$-уровне и т. д. Заметим, что две вершины дерева G_v связаны ребром тогда и только тогда, когда они связаны ребром в исходном дереве G.

Определение 7.10. *Корневое дерево G называется D-ичным, если любая его вершина имеет не более D потомков.*

Для D-ичного дерева индуктивно определим *D-ичную разметку*, т. е. сопоставим каждой вершине u метку $\mu(u)$, представляющую собой некоторое слово в алфавите $B = \{b^{(1)}, \ldots, b^{(D)}\}$. Вначале для каждой неконцевой вершины u произвольным образом пометим все исходящие из нее ребра различными символами алфавита B; это можно сделать, поскольку число потомков у любой вершины u

не превосходит D. По определению метка $\mu(v_0)$ корня v_0 есть пустое слово. Если вершина u является предком вершины v и метка $\mu(u)$ уже определена, а ребро (u, v) помечено символом $b \in B$, то метка $\mu(v)$ вершины v получается из метки $\mu(u)$ приписыванием справа символа b, $\mu(v) = \mu(u)||b$. Очевидно, для вершин k-го уровня метка имеет длину k. Полученное в результате дерево с метками будем назвать *размеченным*.

Пример 7.2. На рис. 7.1 приведены размеченные корневые деревья:

— двоичное, высотой два, с пятью вершинами, в том числе тремя листьями ($а$);

— троичное, высотой три, с десятью вершинами, в том числе шестью листьями ($б$).

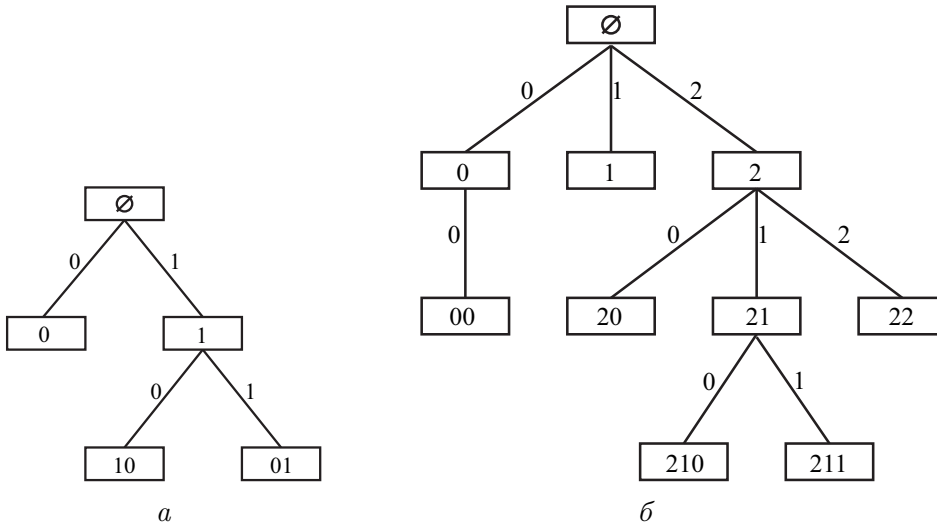


Рис. 7.1. Примеры двух размеченных деревьев: $а$ — двоичное; $б$ — троичное

Определение 7.11. *Полным называется такое D-ичное размеченное дерево G высотой L ($L \geqslant 1$), в котором каждая вершина на уровнях $0 \leqslant k \leqslant L-1$ имеет ровно D потомков, а все D^L вершины уровня L являются листьями.*

В качестве меток вершин полного D-ичного размеченного дерева G встречаются все возможные слова в алфавите B, имеющие длину не более L.

Теорема 7.1. 1) *Если G — D-ичное размеченное корневое дерево, то множество меток всех его листьев образует префиксный D-ичный код.*

2) *Если $\varphi : A \to B^*$ — D-ичное префиксное кодирование, то существует такое D-ичное размеченное корневое дерево G, для которого множество меток всех листьев совпадает с кодом $\varphi(A)$.*

Доказательство. 1) Если в D-ичном размеченном корневом дереве имеется только один лист, то утверждение теоремы доказано. Поэтому пусть G — D-ичное размеченное корневое дерево, имеющее хотя бы два листа u и v. Выберем два произвольных листа u и v. Если предположить, что метка $\mu(u)$ является

префиксом метки $\mu(v)$, то из определения размеченного дерева следует, что вершина v принадлежит поддереву G_u, поэтому вершина u — не лист, противоречие.

2) Пусть $\varphi : A \to B^*$ — D-ичное префиксное кодирование с длинами кодовых слов $l^{(i)} = |a^{(i)}|$, $L = \max\{l^{(1)}, \ldots, l^{(m)}\}$. Рассмотрим полное D-ичное размеченное дерево G' высотой L. Оно содержит в качестве меток все возможные слова в алфавите B, имеющие длину не более L, в том числе все кодовые слова кода $\varphi(A)$.

Применим к дереву G' алгоритм удаления листьев, состоящий в следующем. На первом шаге алгоритма последовательно просматриваются листья дерева G' на уровне L. Если метка $\mu(v)$ листа v принадлежит коду $\varphi(A)$, то этот лист остается без изменений; если же не принадлежит, то удаляется из дерева G' лист v и ребро, связывающее v с его предком на предшествующем уровне дерева. При этом хотя бы один лист на уровне L не будет удален, так как в коде $\varphi(A)$ имеется хотя бы одно кодовое слово длиной L. Если для некоторой вершины u на предшествующему уровне все ее потомки были удалены, то вершина u становится листом. На втором шаге алгоритма последовательно просматриваем новые листья, образовавшиеся на уровне $L-1$ и проделываем с ними описанную выше операцию.

Перемещаясь таким образом от уровня к уровню, завершим алгоритм тогда, когда на очередном шаге невозможно удалить ни один лист. Легко показать, что в итоге получим искомое дерево G. □

Теорема 7.1 устанавливает взаимно однозначное соответствие между множествами слов в алфавите B, в которых никакое слово не является префиксом другого слова, с одной стороны, и размеченными D-ичными деревьями — с другой. Однако множество слов с указанным свойством — еще не префиксный код, поскольку множество неупорядочено. Чтобы из данного множества слов получить алфавитное префиксное кодирование, или префиксный код, необходимо задать взаимно однозначное соответствие между символами алфавита A и кодовыми словами. Такое соответствие можно задать и в терминах размеченного D-ичного дерева.

Определение 7.12. *D-ичным кодовым деревом называется такое размеченное D-ичное дерево G, для которого задано взаимно однозначное соответствие между символами алфавита A и листьями дерева G.*

Таким образом, префиксное кодирование и соответствующее кодовое дерево G можно рассматривать как эквивалентные описания одного и того же объекта.

7.3. НЕРАВЕНСТВА КРАФТА И МАК-МИЛЛАНА

В данном разделе будут доказаны несколько неравенств связывающих длины кодовых слов префиксного и однозначно декодируемого кодов с размером алфавита B, из которых будет следовать несколько важных результатов.

Теорема 7.2 (неравенство Крафта). *Пусть D, m, $l^{(1)}, \ldots, l^{(m)}$ — натуральные числа. Для того чтобы существовало D-ичное префиксное алфавитное*

кодирование $\varphi : A \to B^*$ *с длинами кодовых слов* $l^{(i)} = |\varphi(a^{(i)})|$, $i \in \{1, \ldots, m\}$, *необходимо и достаточно, чтобы выполнялось неравенство*

$$\sum_{i=1}^{m} D^{-l^{(i)}} \leqslant 1. \tag{7.2}$$

Доказательство. Необходимость. Пусть существует D-ичное префиксное алфавитное кодирование $\varphi : A \to B^*$ с длинами кодовых слов $l^{(i)} = |\varphi(a^{(i)})|$. Покажем, что выполнено неравенство (7.2). Обозначим $L = \max\{l^{(1)}, \ldots, l^{(m)}\}$.

Рассмотрим полное D-ичное размеченное дерево G высотой L. Число листьев такого дерева равно D^L. Выделим такие вершины v_i, $1 \leqslant i \leqslant m$, дерева G, что их метки $\mu(v_i)$ совпадают с кодовыми словами $\varphi(a^{(i)})$. Заметим, что все листья поддеревьев G_{v_i} являются листьями и дерева G. Кроме того, поскольку алфавитное кодирование φ префиксное, то различные поддеревья G_{v_i} и G_{v_j}, $i \neq j$, не имеют общих листьев. В силу всего сказанного сумма количества листьев всех поддеревьев G_{v_i} не должна превышать общее количество листьев дерева G, равное D^L. Каждое поддерево G_{v_i} — полное D-ичное дерево высотой $L - l^{(i)}$ и, следовательно, имеет ровно $D^{L-l^{(i)}}$ листьев. Таким образом, имеем неравенство

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

Сократив на D^L, получим требуемое неравенство (7.2).

Достаточность. Покажем, что если натуральные числа $D, m, l^{(1)}, \ldots, l^{(m)}$ таковы, что выполнено неравенство (7.2), то существует D-ичное префиксное алфавитное кодирование $\varphi : A \to B^*$ с длинами кодовых слов $l^{(i)}$. Для этого согласно теореме 7.1 достаточно построить D-ичное размеченное корневое дерево G с m листьями, метки которых имеют длины $l^{(1)}, \ldots, l^{(m)}$.

Пусть $L = \max\{l^{(1)}, \ldots, l^{(m)}\}$. Обозначим через w_l количество чисел набора $l^{(1)}, \ldots, l^{(m)}$ равных l, $l \in \{1, \ldots, L\}$, т. е. $w_l = \sum_{i=1}^{m} \delta_{l,l^{(i)}}$. Тогда неравенство (7.2) можно переписать в виде

$$\sum_{l=1}^{L} w_l D^{-l} \leqslant 1. \tag{7.3}$$

В новых обозначениях построим D-ичное размеченное корневое дерево G с w_1 листьями на первом уровне, w_2 листьями на втором уровне и т. д., w_L листьями на L-м уровне.

Согласно определению числа L, $w_L > 0$, поэтому последнее слагаемое в (7.3) строго положительно, вследствие чего

$$\forall k \in \{1, \ldots, L-1\} : \sum_{l=1}^{k} w_l D^{-l} < 1$$

или после элементарных преобразований

$$\forall k \in \{1, \ldots, L-1\}: \; D^k - \sum_{l=1}^{k} w_l D^{k-l} > 0. \tag{7.4}$$

Опишем построение требуемого дерева. Рассмотрим полное D-ичное размеченное дерево G' высотой L. Рассмотрим D вершин первого уровня. Объявим w_1 из этих вершин листьями, а остальные $D - w_1$ вершин — промежуточными вершинами, из которых будут исходить ребра к вершинам второго уровня. При этом если вершина v объявлена листом, то из дерева G' удалим поддерево G'_v, за исключением самой вершины v (поскольку из листа не может исходить ни одного ребра). Из (7.4) при $k = 1$ получим $D - w_1 > 0$. Таким образом, на первом уровне есть хотя бы одна промежуточная вершина, и, значит, на следующих уровнях дерева G' остались еще вершины.

На рис. 7.2 изображен пример полного дерева G' для $D = 2$ и $l^{(1)} = 1, l^{(2)} = 2, l^{(3)} = 3, L = 3$ и, следовательно, $w_1 = w_2 = w_3 = 1$. Для упрощения понимания метки вершин и ребер не отображены. На первом шаге вершина A объявлена листом. При этом поддерево, для которого вершина была корнем, удалено. На рис. 7.2 удаленные ребра изображены штриховой линией.

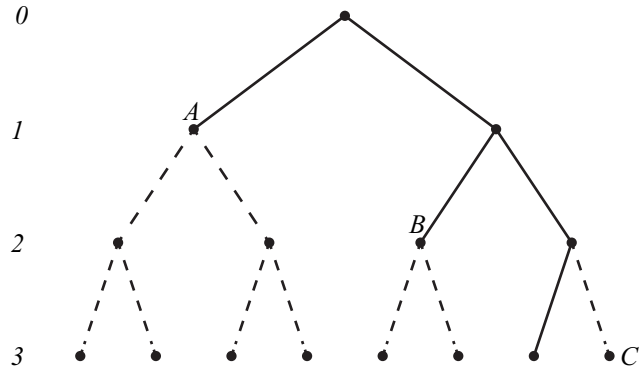


Рис. 7.2. Пример построения кодового дерева при $D = 2$ и $l^{(1)} = 1, l^{(2)} = 2, l^{(3)} = 3$; *0 – 3* — уровни вершин

На втором шаге w_2 из оставшихся на втором уровне вершин объявляем листьями, удаляем исходящие их этих вершин поддеревья, а остальные вершины объявляем промежуточными и повторяем так до уровня $L - 1$. На рис. 7.2 на втором шаге вершина B объявлена листом. Покажем, что такое построение всегда возможно. Пусть на всех уровнях с первого до $(j - 1)$ включительно $(j \in \{2, \ldots, L-1\})$ объявлены листья, а на $(j-1)$ объявлены и промежуточные вершины согласно описанному алгоритму и пусть количество промежуточных вершин больше нуля. Рассмотрим вершины j-го уровня. По алгоритму объявим w_j вершин листьями, удалим исходящие их этих вершин поддеревья, а остальные вершины объявим промежуточными. Посчитаем, сколько вершин на j-м уровне останется промежуточными. Если бы до этого момента ни одна вершина

не была объявлена листом, то всего промежуточных вершин было бы D^j. Однако каждый лист первого уровня уменьшает это количество на D^{j-1}, второго уровня на D^{j-2} и т. д. Это можно пояснить следующим образом: если v — лист l-го уровня, то удаляемое полное поддерево на уровне j имеет высоту $j-l$ и, следовательно, на j-м уровне оно бы имело D^{j-l} вершин. Например, на рис. 7.2 на уровне $j=2$ промежуточных вершин $2^2-2^1-2^0=1$. Учитывая количество листьев каждого из уровней, получим, что количество промежуточных вершин

$$W_j = D^j - \sum_{l=1}^{j} w_l D^{j-l},$$

а эта величина в силу (7.4) при $k=j$ будет строго больше нуля. Это означает, что и на $j+1$ уровне остались еще вершины и можно продолжать построение до $L-1$ уровня включительно.

На последнем L-м уровне поступаем аналогично. Единственное отличие заключается в том, что промежуточных вершин остается

$$W_L = D^L - \sum_{l=1}^{L} w_l D^{L-l}.$$

Согласно (7.3) $W_L \geqslant 0$.

Если $W_L = 0$, то лишних (промежуточных) вершин на L-м уровне нет, т. е. искомое дерево построено.

Если $W_L > 0$, то остались лишние вершины на L-м уровне. В этом случае, как и при доказательстве теоремы 7.1, начиная с уровня L, будем идти от уровня к уровню по направлению к корню и удалять все необъявленные листьями вершины и ребра, связывающие эти вершины с предками. В итоге требуемое D-ичное размеченное корневое дерево построено. Так, на рис. 7.2 вершина C не объявлена листом и поэтому была удалена. □

Замечание 7.2. Из данной теоремы не следует, что любой код с длинами кодовых слов, удовлетворяющими (7.2), является префиксным. Например, множество двоичных кодовых слов 0, 01, 11 удовлетворяет неравенству (7.2), но данное алфавитное кодирование не является префиксным.

Теорема 7.3 (неравенство Мак-Миллана). *Если алфавитное кодирование $\varphi : A \to B^*$ с длинами кодовых слов $l^{(i)} = |\varphi(a^{(i)})|$ является однозначно декодируемым, то справедливо неравенство (7.2).*

Доказательство. Для натуральных n и l обозначим через $C_{n,l}$ множество таких последовательностей $a^n = (a_1, \ldots, a_n) \in A^n$, для которых кодовая последовательность $\varphi^*(a^n)$ имеет длину l. Пусть $c_{n,l}$ — количество элементов множества $C_{n,l}$. По условию отображение $\varphi : A \to B^*$ инъективно и при этом $\varphi^*(C_{n,l}) \subseteq B^l$, следовательно,

$$c_{n,l} \leqslant |B^l| = D^l. \tag{7.5}$$

Для произвольного натурального n, учитывая (7.5), рассмотрим выражение

$$\left(\sum_{i=1}^{m} D^{-l^{(i)}}\right)^n = \sum_{1\leqslant i_j,\,\ldots,\,i_n\leqslant m} D^{-(l^{(i_1)}+\cdots+l^{(i_n)})} =$$

$$= \sum_{(a^{(i_1)},\,\ldots,\,a^{(i_n)})\in A^n} D^{-|\varphi^*(a^{(i_1)},\,\ldots,\,a^{(i_n)})|} = \sum_{l=1}^{nL} c_{n,l} D^{-l} \leqslant \sum_{l=1}^{nL} 1 = nL,$$

где $L = \max\{l^{(1)}, \ldots, l^{(m)}\}$. Таким образом, для произвольного натурального n имеет место неравенство

$$\left(\sum_{i=1}^{m} D^{-l^{(i)}}\right)^n / nL \leqslant 1. \tag{7.6}$$

Отсюда сделаем вывод, что $\sum\limits_{i=1}^{m} D^{-l^{(i)}} \leqslant 1$, так как в противном случае при достаточно большом n мы получим противоречие с (7.6). □

Следствие 7.1. *Если алфавитное кодирование $\varphi : A \to B^*$ является однозначно декодируемым, то существует префиксное алфавитное кодирование $\varphi_1 : A \to B^*$ с таким же набором длин кодовых слов, т. е. $|\varphi(a^{(i)})| = |\varphi_1(a^{(i)})|$, $\forall i \in \{1, \ldots, m\}$.*

7.4. НАСЫЩЕННОЕ КОРНЕВОЕ ДЕРЕВО И ЕГО СВОЙСТВА

Определение 7.13. *D-ичное корневое дерево G называется насыщенным, если любая неконцевая вершина имеет ровно D потомков, за исключением, может быть, одной вершины на предпоследнем уровне, называемой особой вершиной, которая имеет D_0 потомков, $2 \leqslant D_0 < D$.*

Отметим, что дерево, состоящее из одной вершины — корня, насыщенное. Также из определения следует, что при $D = 2$ у насыщенных деревьев особых вершин не существует.

Пример 7.3. На рис. 7.3 приведены примеры деревьев:
— насыщенное, $D = 2$ ($а$);
— ненасыщенное, $D = 3$ ($б$);
— насыщенное, $D = 3$ ($в$).

Лемма 7.1. *Для любых целых чисел $D \geqslant 2$, $m \geqslant 1$ существует насыщенное D-ичное корневое дерево с m листьями.*

Доказательство. Доказательство проведем по методу математической индукции по числу листьев m для каждого фиксированного числа D. При $m = 1$ искомым деревом является дерево, состоящее из одной корневой вершины.

Пусть имеется насыщенное D-ичное корневое дерево G_m с m листьями. Докажем, что тогда существует и насыщенное D-ичное корневое дерево G_{m+1}, име-

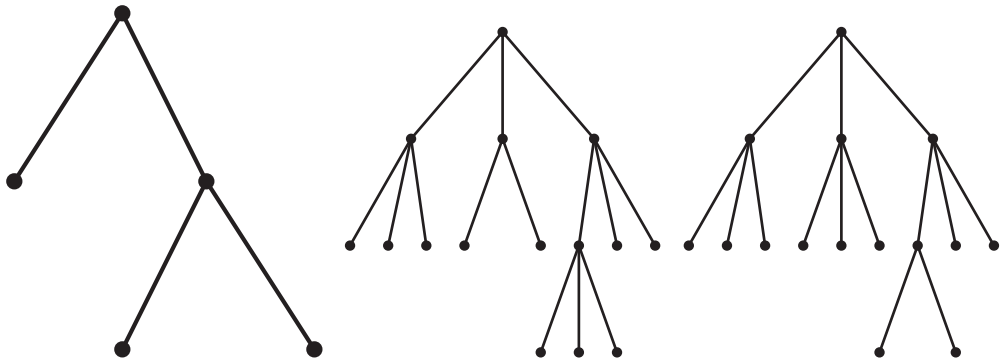


Рис. 7.3. Насыщенные (*а, в*) и ненасыщенные (*б*) корневые деревья

ющее $m+1$ лист. Если в дереве G_m есть особая вершина, то добавим к ней еще одного потомка, и искомое дерево G_{m+1} построено. Если в дереве G_m нет особой вершины, то выберем произвольную концевую вершину v на последнем уровне и добавим к ней двух потомков. Вершина v станет особой, а число листьев увеличится на одно, т. е. искомое дерево G_{m+1} снова построено. □

В насыщенном D-ичном корневом дереве G для заданного m можно определить, есть ли в G особая вершина (если есть, то сколько у нее потомков) и выполнено ли равенство в неравенстве Крафта.

Лемма 7.2. *В насыщенном D-ичном корневом дереве без особых вершин число листьев имеет вид $m = 1 + k(D-1)$, $k \in \mathbb{N}_0$ ($\mathbb{N}_0 = \mathbb{N} \cup \{0\}$).*

Доказательство. Обозначим через L высоту насыщенного дерева G. Доказательство проведем по методу математической индукции по высоте дерева L.

При $L = 0$ имеется дерево, состоящее из одной вершины — корня, который в это же время является и листом, поэтому $m = 1$, $k = 0$. База индукции доказана.

Предположим, что для $L > 0$ и для всех насыщенных деревьев высотой, меньше L, утверждение леммы выполнено. Рассмотрим произвольное насыщенное D-ичное дерево G высотой L. Поскольку $L > 0$, то в дереве G имеется по крайней мере два уровня (нулевой и первый), а это значит, что у корневой вершины (на нулевом уровне) обязательно есть ровно D потомков, обозначим их $v_1, \ldots, v_D$. Число листьев дерева G можно посчитать как сумму листьев поддеревьев $G_{v_1}, \ldots, G_{v_D}$. Заметим, что каждое из поддеревьев — это D-ичное корневое дерево без особых вершин высотой, меньше L, а значит, для подсчета числа их вершин можно воспользоваться предположением индукции. Отсюда получим, что общее количество листьев дерева G равно

$$m = 1 + k_1(D-1) + \ldots + 1 + k_D(D-1) =$$
$$= D + (k_1 + \ldots + k_D)(D-1) = 1 + (1 + k_1 + \ldots + k_D)(D-1). \qquad \square$$

Для $D \geqslant 2$, $m \geqslant 1$ обозначим через m_0 такое наименьшее целое число вида $m_0 = 1 + k(D-1)$, $k \in \mathbb{N}_0$, что $m_0 \geqslant m$. Тогда условие леммы 7.2 можно переформулировать так: если в насыщенном D-ичном корневом дереве с m листьями

нет особой вершины, то $m_0 = m$.

Лемма 7.3. *Пусть G — насыщенное D-ичное корневое дерево с m листьями, у которого есть особая вершина. Тогда $m_0 > m$.*

Доказательство. По условию леммы в дереве G есть особая вершина u на уровне $L-1$, удалим ее D_0 потомков, $2 \leqslant D_0 < D$, а саму особую вершину u сделаем листом. Получим новое насыщенное дерево G' с

$$m' = m - (D_0 - 1) < m \tag{7.7}$$

листьями и без особой вершины. Согласно лемме 7.2 у дерева G' число вершин $m' = 1 + k'(D-1) = m'_0$. Из (7.7) следует, что в исходном дереве G число листьев

$$m = m'_0 + (D_0 - 1) < m'_0 + (D - 1) = m_0. \tag{7.8}$$

Неравенство в (7.8) следует из определения особой вершины в насыщенном дереве ($D_0 < D$), а равенство в (7.8) — из того, что число $m'_0 + (D-1)$ — наименьшее целое, не меньшее m. □

Следствие 7.2. *В насыщенном D-ичном корневом дереве с m листьями нет особой вершины тогда и только тогда, когда $m_0 = m$.*

Доказательство. Необходимость следует из леммы 7.2. Достаточность докажем методом от противного. Пусть $m_0 = m$, но в G есть особая вершина. Тогда согласно лемме 7.3 должно быть выполнено неравенство $m_0 > m$, противоречие. □

Следствие 7.3. *В насыщенном D-ичном корневом дереве с m листьями есть особая вершина тогда и только тогда, когда $m_0 > m$.*

Доказательство. Доказательство проводится аналогично доказательству следствия 7.2. □

Лемма 7.4. *Пусть G — насыщенное D-ичное корневое дерево с m листьями, высоты которых равны $l^{(1)}, l^{(2)}, \ldots, l^{(m)}$, $K = \sum_{i=1}^{m} D^{-l^{(i)}}$. Если $m_0 = m$, то в дереве G особой вершины нет и $K = 1$, а если $m_0 > m$, то особая вершина в дереве G есть, число ее потомков равно $D_0 = m - m_0 + D$, а $K < 1$.*

Доказательство. Проведем по методу математической индукции по m для фиксированного D.

Если $m = 1$, то $m_0 = 1$ и выполнено равенство $m = m_0$. При этом G – корневое дерево, состоящее из одной вершины – корня. В дереве G нет особых вершин и $K = D^0 = 1$. База индукции доказана.

Пусть $m > 1$ и для всех деревьев G, имеющих меньше m листьев, лемма доказана. Докажем лемму и для дерева G, имеющего m листьев. Обозначим через $L = \max\{l^{(1)}, \ldots, l^{(m)}\}$ высоту дерева G.

Если $m_0 > m$, то согласно следствию 7.3 это равносильно тому, что в дереве G есть особая вершина u на уровне $L-1$. Как и при доказательстве леммы 7.3, удалим D_0 потомков вершины u, $2 \leqslant D_0 < D$, а саму особую вершину u сделаем листом. Получим новое насыщенное дерево G', для которого выполнены соотношения (7.7) и (7.8). Тогда соответствующая K величина K' дерева G' равна $K' = K - D_0 D^{-L} + D^{-(L-1)}$. Поскольку в G' выполнено $m' = m'_0$, то

согласно предположению индукции $K' = 1$, следовательно,

$$K = 1 + D_0 D^{-L} - D^{-(L-1)} = 1 - D^{-L}(D - D_0) < 1,$$

и, кроме того,

$$m' = m - D_0 + 1,\ m'_0 = m_0 - D + 1 \Rightarrow$$
$$\Rightarrow m - D_0 + 1 = m' = m'_0 = m_0 - D + 1.$$

Из последнего равенства после простейших преобразований найдем $D_0 = m - m_0 + D$.

При $m_0 = m$ согласно следствию 7.2 в дереве G нет особой вершины. Тогда возьмем произвольную неконцевую вершину u на уровне $L - 1$; эта вершина имеет ровно D потомков. Удалим из дерева G D потомков вершины u, а саму вершину u сделаем листом. Получим новое насыщенное дерево G' с $m' = m - D + l < m$ листьями, не имеющее особой вершины. Соответствующая величина K' для дерева G' находится из соотношения

$$K' = K - DD^{-L} + D^{-(L-1)} = K.$$

Как и в предыдущем случае, используя лемму 7.2, можно показать, что число листьев в дереве G' без особой вершины равно

$$m' = 1 + (k - 1)(D - 1) = m'_0.$$

Значит, согласно предположению индукции $K' = 1$, так что $K = 1$. □

Пример 7.4. Если $m = 9$, $D = 4$, то из условия $m_0 = 1 + 3k \geqslant 9$ получим $m_0 = 10 > 9$, поэтому особая вершина есть, и число ее потомков равно $D_0 = m - m_0 + D = 9 - 10 + 4 = 3$. Если $m = 7, D = 3$, то из условия $m_0 = 1 + 2k \geqslant 7$ следует $m_0 = 7$, поэтому особой вершины нет.

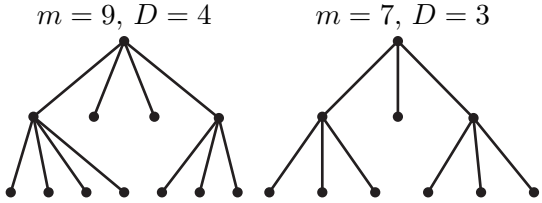


Рис. 7.4. Насыщенные корневые деревья

Возможные насыщенные деревья с данными параметрами приведены на рис. 7.4.

7.5. СРЕДНЯЯ ДЛИНА ОПТИМАЛЬНОГО КОДА

Рассмотрим ситуацию, когда правило кодирования φ^* применяется к последовательностям $\Xi_n = (\xi_1, \ldots, \xi_n)$, порождаемым дискретным источником без памяти $\langle A, p(\cdot)\rangle$, $p(a) = \mathsf{P}\{\xi = a\}$, $a \in A$. В дальнейшем, если не будет оговорено иначе, будем полагать, что распределение $p(\cdot)$ задано и фиксировано.

Определение 7.14. *Пусть заданы алфавит источника* $A = \{a^{(1)}, \ldots, a^{(m)}\}$, *кодовый алфавит* $B = \{b^{(1)}, \ldots, b^{(D)}\}$ *и распределение* $p(a) = \mathsf{P}\{\xi = a\}$, $a \in A$. *Тогда средней длиной кодового слова при алфавитном кодировании* $\varphi : A \to B^*$ *(или средней длиной кода* $\varphi(A)$*) называется величина*

$$l^{\varphi} = \mathsf{E}\{|\varphi(\xi)|\} = \sum_{i=1}^{m} p\left(a^{(i)}\right) l^{(i)}, \tag{7.9}$$

где ξ *— выходной символ ИДС без памяти* $\langle A, p(\cdot)\rangle$, $l^{(i)} = \left|\varphi(a^{(i)})\right|$, $i \in \{1, \ldots, m\}$ *— длины кодовых слов.*

Поскольку длины кодовых слов — натуральные числа $\left(l^{(i)} \geqslant 1\right)$, то для любого кодирования φ верно неравенство $l^{\varphi} \geqslant 1$.

Для краткости записи введем обозначение: $p_i = p\left(a^{(i)}\right)$.

Определение 7.15. *Алфавитное кодирование* $\varphi_0 : A \to B^*$ *и код* $\varphi_0(A)$ *оптимальные, если* φ_0 *однозначно декодируемо и при этом средняя длина* l^{φ_0} *минимальна.*

Замечание 7.3. Для последовательности, порождаемой дискретным источником без памяти, средняя длина кодовой последовательности $\varphi^*(A^n)$ может быть вычислена следующим образом:

$$\mathsf{E}\{|\varphi^*(\Xi_n)|\} = \mathsf{E}\left\{\sum_{i=1}^{n} |\varphi(\xi_i)|\right\} = \sum_{i=1}^{n} \mathsf{E}\{|\varphi(\xi_i)|\} = n\mathsf{E}\{|\varphi(\xi)|\}. \tag{7.10}$$

Согласно (7.10) средняя длина кодовой последовательности $\varphi^*(\Xi_n)$ будет минимальна в случае, когда алфавитное кодирование φ является оптимальным.

Замечание 7.4. Из следствия 7.1 к неравенству Мак-Миллана вытекает, что если оптимальное кодирование существует, то существует и префиксное кодирование с таким же набором длин кодовых слов, а поскольку любое префиксное кодирование однозначно декодируемо по утверждению 7.1, то это префиксное кодирование также оптимальное, поэтому далее при описании оптимальных кодов можно ограничиться только рассмотрением префиксных кодов.

Упражнение 7.3. Доказать, что в случае $m \leqslant D$ оптимальное кодирование φ_0 существует и $l^{\varphi_0} = 1$.

Из определения оптимального кодирования еще не следует, что такое кодирование существует. Действительно, класс всевозможных кодирований бесконечен, и неясно, достигается ли точная нижняя грань $\inf_{\varphi} l^{\varphi}$ на каком-либо кодировании φ. Следующее утверждение дает ответ на данный вопрос.

Утверждение 7.2. Оптимальное кодирование φ_0 существует.

Доказательство. Для доказательства этого утверждения покажем, что можно выбрать подмножество кодов, на котором достигается точная нижняя грань $\inf_{\varphi} l^{\varphi}$ и величина l^{φ} принимает конечное число значений. В этом случае обязательно найдется такое кодирование φ_0, при котором достигается минимум l^{φ}.

Пусть L — это такое наименьшее число, что $m \leqslant D^L$. Тогда существует равномерное кодирование φ, все длины кодовых слов которого равны L. Следовательно, $\inf_\varphi l^\varphi \leqslant L$. А значит, инфимум не изменится, если из рассмотрения исключить все кодирования φ, для которых $l^\varphi > L$. Таким образом, остается рассмотреть лишь кодирования, для которых выполнено соотношение

$$l^\varphi = \sum_{i=1}^{m} p_i l^{(i)} \leqslant L.$$

Если $p_i = 0$, то значение $l^{(i)}$ несущественно. Если же $p_i > 0$, то $l^{(i)} \leqslant L/p_i \leqslant L/p$, где p наименьшее из чисел $p_1, \ldots, p_m$, не равных нулю. При таких ограничениях средняя длина l^φ может принимать лишь конечное число различных значений. Таким образом, искомое подмножество кодирований построено. □

Теорема 7.4. *Если алфавитное кодирование φ однозначно декодируемо, то справедливо неравенство*

$$l^\varphi \geqslant \frac{\mathsf{H}\{\xi\}}{\log D}, \tag{7.11}$$

где ξ — выходной символ ИДС без памяти $\langle A, p(\cdot)\rangle$. Равенство достигается тогда и только тогда, когда все положительные вероятности p_i имеют вид $p_i = D^{-l^{(i)}}$.

Доказательство. Рассмотрим разность

$$\mathsf{H}\{\xi\} - l^\varphi \log D = -\sum_{i=1}^{m} p_i \log p_i - (\log D) \sum_{i=1}^{m} p_i l^{(i)} =$$

$$= -\sum_{i=1}^{m} p_i \log p_i + \sum_{i=1}^{m} p_i \log D^{-l^{(i)}} = \sum_{i=1}^{m} p_i \log \frac{D^{-l^{(i)}}}{p_i}.$$

При этом можно считать, что суммирование идет лишь по тем слагаемым, для которых $p_i > 0$. Поскольку если $p_i = 0$, то соответствующие слагаемые в $\mathsf{H}\{\xi\}$ и l^φ будут равны нулю. Применяя неравенство Йенсена, а затем неравенство Мак-Миллана, получим

$$\mathsf{H}\{\xi\} - l^\varphi \log D \leqslant \log \sum_{i=1}^{m} p_i \frac{D^{-l^{(i)}}}{p_i} = \log \sum_{i=1}^{m} D^{-l^{(i)}} \leqslant \log 1 = 0. \tag{7.12}$$

Докажем второе утверждение теоремы. Если $p_i = D^{-l^{(i)}}$, то

$$\mathsf{H}\{\xi\} - l^\varphi \log D = \sum_{i=1}^{m} p_i \log \frac{D^{-l^{(i)}}}{p_i} = \sum_{i=1}^{m} p_i \log 1 = 0.$$

Обратно пусть имеет место равенство в (7.11), тогда выполнено и равенство в (7.12). Равенство в неравенстве Йенсена достигается при условии

$$\frac{D^{-l^{(1)}}}{p_1} = \cdots = \frac{D^{-l^{(m)}}}{p_m} = \alpha,\ \alpha \in \mathbb{R}.$$

Если $\alpha > 1$, то $\sum_{i=1}^{m} D^{-l^{(i)}} = \alpha \sum_{i=1}^{m} p_i = \alpha > 1$, что противоречит неравенству Мак-Миллана.

Если $\alpha < 1$, то $\log \alpha < 0$ и, следовательно,

$$\mathsf{H}\{\xi\} - l^{\varphi} \log D = \sum_{i=1}^{m} p_i \log \frac{D^{-l^{(i)}}}{p_i} = \sum_{i=1}^{m} p_i \log \alpha < 0,$$

что противоречит предположению о равенстве в (7.11). Остается вариант $\alpha = 1$, из которого следует $p_i = D^{-l^{(i)}}$. Таким образом, второе утверждение теоремы доказано. □

Теорема 7.5. *Если распределение $p(\cdot)$ невырожденное, то существует такое префиксное кодирование φ, для которого справедливо неравенство*

$$l^{\varphi} < 1 + \frac{\mathsf{H}\{\xi\}}{\log D}, \tag{7.13}$$

где ξ — выходной символ ИДС без памяти $\langle A, p(\cdot)\rangle$.

Доказательство. Если $p_i > 0$, то обозначим через $l^{(i)}$ наименьшее натуральное число с условием $l^{(i)} \geqslant -\log_D p_i$. Поскольку распределение $p(\cdot)$ невырожденное, то $p_i < 1$, отсюда

$$0 \leqslant l^{(i)} - 1 < -\log_D p_i \leqslant l^{(i)}, \tag{7.14}$$

или, что то же самое,

$$D^{-l^{(i)}} \leqslant p_i < D^{-l^{(i)}+1}. \tag{7.15}$$

Из последнего неравенства получим, что

$$\sum_{i:p_i>0} D^{-l^{(i)}} \leqslant \sum_{i=1}^{m} p_i = 1. \tag{7.16}$$

Рассмотрим случай, что все вероятности $p_1, \ldots, p_m$ положительны, тогда согласно неравенству Крафта из (7.16) следует существование префиксного кодирования с длинами кодовых слов $|\varphi(a^{(i)})| = l^{(i)}$, $i \in \{1, \ldots, m\}$. Из (7.14) следует, что $l^{(i)} < 1 - \log_D p_i$, откуда по свойству логарифма

$$l^{\varphi} = \sum_{i:p_i>0} p_i l^{(i)} = \sum_{i=1}^{m} p_i l^{(i)} < \sum_{i=1}^{m} p_i \left(1 - \log_D p_i\right) =$$

$$= 1 - \sum_{i=1}^{m} p_i \frac{\log p_i}{\log D} = 1 + \frac{\mathsf{H}\{\xi\}}{\log D}. \tag{7.17}$$

Пусть теперь не все p_i положительны и t — это количество индексов i, $i \in \{1, \ldots, m\}$, что $p_i = 0$. Не нарушая общности, будем считать, что $p_{m-t+1} = \ldots = p_m = 0$. Обозначим через $K = \sum_{i:p_i>0} D^{-l^{(i)}}$.

Если $K < 1$, то выберем настолько большое натуральное число l, чтобы выполнялось неравенство $K + tD^{-l} \leqslant 1$, и зададим $l^{(m-t+1)} = \ldots = l^{(m)} = l$. Тогда для набора $l^{(1)}, \ldots, l^{(m)}$ согласно неравенству Крафта существует префиксное кодирование с заданными длинами кодовых слов, для которых оценка (7.17) тоже справедлива.

Если $K = 1$, то из неравенств (7.15) и (7.16) следует, что все положительные p_i имеют вид $p_i = D^{-l^{(i)}}$, откуда $l^{(i)} = -\log_D p_i$. Имеем цепочку равенств

$$\sum_{i:p_i>0} p_i l^{(i)} = \sum_{i:p_i>0} -p_i \log_D p_i = \frac{\mathsf{H}\{\xi\}}{\log D}. \tag{7.18}$$

Далее поступим следующим образом. Согласно сделанным предположениям $p_1 > 0$, тогда, заменив в сумме $K = \sum_{i:p_i>0} D^{-l^{(i)}}$ число $l^{(1)}$ на $l^{(1)} + 1$, получим

$$K_1 = D^{-(l^{(1)}+1)} + \sum_{i:i>1,p_i>0} D^{-l^{(i)}} < K = 1.$$

Тогда, как и для случая $K < 1$, можно выбрать настолько большое натуральное число l, чтобы выполнялось неравенство $K_1 + tD^{-l} \leqslant 1$. Зададим $l^{(m-t+1)} = \ldots = l^{(m)} = l$. Теперь для набора $l^{(1)} + 1, l^{(1)}, \ldots, l^{(m)}$ согласно неравенству Крафта существует префиксное кодирование φ с заданными длинами кодовых слов. Вычислим среднюю длину кода φ, воспользовавшись (7.18):

$$l^{\varphi} = p_1(l^{(1)} + 1) + \sum_{i>1:p_i>0} p_i l^{(i)} = p_1 + \sum_{i:p_i>0} p_i l^{(i)} = p_1 + \frac{\mathsf{H}\{\xi\}}{\log D}. \tag{7.19}$$

Поскольку распределение $p(\cdot)$ невырожденное, то $p_1 < 1$, откуда из (7.19) получим требуемое соотношение. □

Замечание 7.5. Требование невырожденности в формулировке теоремы 7.5 существенно. Если распределение $p(\cdot)$ вырождено, то $\mathsf{H}\{\xi\} = 0$ и неравенство (7.13) примет вид $l^{\varphi} < 1$, что неверно, так как для любого кодирования $l^{\varphi} \geqslant 1$.

Однако для вырожденного распределения существует такое префиксное кодирование φ, что $l^{\varphi} = 1$. Построим его. Не нарушая общности, можно считать, что $p_1 = 1$ и $p_i = 0, i \in \{2, \ldots, m\}$. Тогда сопоставим $\varphi(a^{(i)})$ последовательность из $(i-1)$ символов $b^{(1)}$ и затем одного символа $b^{(2)}$. Очевидно, что построенный код является префиксным, $\varphi(a^{(1)}) = b^{(2)}$ и $l^{\varphi} = 1$.

Следствие 7.4. *Средняя длина оптимального алфавитного кодирования* φ_0 *удовлетворяет двойному неравенству*

$$\frac{\mathsf{H}\{\xi\}}{\log D} \leqslant l^{\varphi_0} \leqslant 1 + \frac{\mathsf{H}\{\xi\}}{\log D}. \tag{7.20}$$

Доказательство. Левое неравенство (7.20) следует из теоремы 7.4. В случае невырожденного распределения $p(\cdot)$ согласно теореме 7.5 существует префиксное алфавитное кодирование, для которого выполнена правая часть неравенства (7.20), а значит, это неравенство выполнено и для оптимального алфавитного

кодирования. В случае вырожденного распределения $p(\cdot)$ правая часть неравенства (7.20) следует из замечания 7.5. □

Кратко рассмотрим случай, когда имеется произвольный дискретный источник сообщений $\langle A^n, p_n()\rangle$, где $p_n(a_1, \ldots, a_n) = \mathsf{P}\{\xi_1 = a_1, \ldots, \xi_n = a_n\}$ — n-мерное дискретное распределение вероятностей n-символьного сообщения.

Откажемся от продолжения алфавитного кодирования $\varphi : A \to B^*$ на последовательности символов алфавита A по правилу сцепления, а вместо этого будем кодировать целые n-символьные блоки. Другими словами, для любого $n \geqslant 1$ рассмотрим алфавит A^n с заданным на нем распределением $p_n()$, и пусть φ_n — соответствующее этому распределению *оптимальное* D-ичное кодирование $\varphi_n : A^n \to B^*$ со средней длиной кодового слова, равной

$$l^{\varphi_n} = \mathsf{E}\{|\varphi_n(\Xi_n)|\} = \sum_{a^n \in A^n} p_n(a^n)\,|\varphi_n(a^n)|\,.$$

В данном случае интерес представляет среднее число символов кодового алфавита B, приходящихся на один символ алфавита A при большой длине блока n, т. е. величина

$$\bar{l}^{(n)} = \frac{l^{\varphi_n}}{n}.$$

Теорема 7.6 (теорема о кодировании при отсутствии шума). *Если для дискретного источника сообщений существует предельная энтропия h, то существует предел*

$$\lim_{n\to\infty} \bar{l}^{(n)} = \frac{h}{\log D}.$$

Доказательство. Идея доказательства заключается в следующем. В данном случае также верно следствие 7.4 в виде

$$\frac{\mathsf{H}\{\Xi_n\}}{\log D} \leqslant l^{\varphi_n} \leqslant 1 + \frac{\mathsf{H}\{\Xi_n\}}{\log D}.$$

Разделим почленно на n и подставим $h_n = \mathsf{H}\{\Xi_n\}/n$. Получим двойное неравенство

$$\frac{h_n}{\log D} \leqslant \bar{l}^{(n)} \leqslant \frac{1}{n} + \frac{h_n}{\log D}.$$

Поскольку по условию теоремы существует предел $\lim_{n\to\infty} h_n = h$, то левая и правая часть сходятся к величине $h/\log D$, а значит, и заключенная между ними величина $\bar{l}^{(n)}$ тоже сходится к $h/\log D$. □

7.6. СВОЙСТВА ОПТИМАЛЬНОГО КОДА

В данном разделе рассмотрим еще ряд свойств оптимального кодирования и кодовых деревьев, соответствующих оптимальному кодированию. Результаты затем будут использоваться для обоснования корректности методов построения оптимальных кодов.

Предположим, что правило кодирования φ^* применяется к последовательностям, порождаемым дискретным источником без памяти $\langle A, p(\cdot)\rangle$, $A = \{a^{(1)}, \ldots, a^{(m)}\}$, $(m \geqslant 2)$, $p_i = p(a^{(i)}) = \mathbf{P}\{\xi = a^{(i)}\}$, $i \in \{1, \ldots, m\}$.

Лемма 7.5. *Пусть $\varphi : A \to B^*$ — оптимальное D-ичное алфавитное кодирование с длинами кодовых слов $l^{(i)} = |\varphi(a^{(i)})|$, $i \in \{1, \ldots, m\}$, соответствующее распределению $p(\cdot)$. Если $p_i > p_j$, то $l^{(i)} \leqslant l^{(j)}$.*

Доказательство. Допустим противное, т. е. $p_i > p_j$ и $l^{(i)} > l^{(j)}$. Тогда построим новое кодирование φ_1 следующим образом: $\varphi_1(a^{(k)}) = \varphi(a^{(k)})$, $\forall k \in \in \{1, \ldots, m\}\backslash\{i,j\}$, $\varphi_1(a^{(i)}) = \varphi(a^{(j)})$, $\varphi_1(a^{(j)}) = \varphi(a^{(i)})$ (другими словами, поменяем коды символов $a^{(i)}$ и $a^{(j)}$ местами). Рассмотрим разность средних длин кодовых слов:

$$l^{\varphi} - l^{\varphi_1} = l^{(i)}p_i + l^{(j)}p_j - l^{(j)}p_i - l^{(i)}p_j = (l^{(i)} - l^{(j)})(p_i - p_j) > 0. \tag{7.21}$$

Таким образом, получим, что $l^{\varphi} > l^{\varphi_1}$, а это противоречит утверждению леммы о том, что φ — оптимальный код. □

Замечание 7.6. Операцию, осуществленную при доказательстве леммы 7.5, назовем *перестановкой кодовых слов*. Из (7.21) следует, что при перестановке кодовых слов символов $a^{(i)}$ и $a^{(j)}$, для которых либо $p_i = p_j$, либо $l^{(i)} = l^{(j)}$, средняя длина кода не изменяется.

Следствие 7.5. *Пусть $\varphi : A \to B^*$ — оптимальное D-ичное алфавитное кодирование с длинами кодовых слов $l^{(i)} = |\varphi(a^{(i)})|$, $i \in \{1, \ldots, m\}$, соответствующее распределению $p(\cdot)$. Если $l^{(i)} > l^{(j)}$, то $p_i \leqslant p_j$.*

Доказательство. Предположим противное, т. е. пусть $l^{(i)} > l^{(j)}$ и $p_i > p_j$. Тогда согласно лемме 7.5 $l^{(i)} \leqslant l^{(j)}$, противоречие. □

Лемма 7.6. *Если $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_m$, то существует оптимальное D-ичное алфавитное кодирование φ, для которого длины кодовых слов удовлетворяют неравенствам*

$$l^{(1)} \leqslant l^{(2)} \leqslant \ldots \leqslant l^{(m)}. \tag{7.22}$$

Доказательство. Разобьем множество всех символов алфавита A на подмножества $A = A_1 \sqcup A_2 \sqcup \ldots \sqcup A_t$ $(1 \leqslant t \leqslant m)$ такие, что символы, принадлежащие одному подмножеству, имеют одинаковую вероятность появления, т. е. $\forall l \in \{1, \ldots, t\}$, $\forall a^{(i)}, a^{(j)} \in A_l$, $i \neq j$, имеем $p_i = p_j$. Примем $m_l = |A_l|$, $l \in \{1, \ldots, t\}$. Тогда, учитывая упорядоченность вероятностей, получим

$$A_1 = \left\{a^{(1)}, \ldots, a^{(m_1)}\right\}, A_2 = \left\{a^{(m_1+1)}, \ldots, a^{(m_1+m_2)}\right\}, \ldots,$$

$$A_t = \left\{a^{(m-m_t+1)}, \ldots, a^{(m)}\right\}.$$

Выберем произвольное оптимальное кодирование φ_0', оно существует согласно утверждению 7.2. Из упорядоченности вероятностей и леммы 7.5 следует соотношение

$$\forall l, k \in \{1, \ldots, t\},\ l < k,\ \forall a^{(i)} \in A_l,\ \forall a^{(j)} \in A_k,\ l^{(i)} \leqslant l^{(j)}. \tag{7.23}$$

В каждом подмножестве A_l переставим кодовые слова так, чтобы в них было выполнено неравенство (7.22), и получим новое кодирование φ_0. Согласно замечанию 7.6 кодирование φ_0 тоже оптимальное. Учитывая (7.23), заключим, что φ_0 является искомым кодированием. □

Пусть $\varphi : A \to B^*$ — некоторое D-ичное префиксное алфавитное кодирование и G — соответствующее кодовое дерево, т. е. размеченное D-ичное корневое дерево с m листьями, которым сопоставлены кодовые слова $\varphi(a^{(1)}), \ldots, \varphi(a^{(m)})$ для символов $a^{(1)}, \ldots, a^{(m)}$ с приписанными им вероятностями $p_1, \ldots, p_m$. Рассмотрим две операции, позволяющие получить новое D-ичное префиксное алфавитное кодирование φ' и соответствующее кодовое дерево G'.

Пусть u — лист дерева G, находящийся на уровне $l \geqslant 2$, и метка $\mu(u) = b^{(i_1)}b^{(i_2)}\ldots b^{(i_l)}$ листа u является кодовым словом $\varphi(a^{(j)})$ для символа $a^{(j)}$ с приписанной ему вероятностью p_j.

Операция 1 (удаление одиночного листа). Предположим, что лист u — единственный потомок вершины v, находящейся на уровне $(l-1)$ и имеющей метку $\mu(v) = b^{(i_1)}b^{(i_2)}\ldots b^{(i_{l-1})}$. Построим новое кодирование φ', которое для всех символов, кроме $a^{(j)}$, будет совпадать с кодированием φ, а $\varphi'(a^{(j)}) = \mu(v) = b^{(i_1)}b^{(i_2)}\ldots b^{(i_{l-1})}$, т. е. из дерева G удалим лист u и ребро (v, u), тогда вершина v станет листом.

Вычислим разность $l^{\varphi} - l^{\varphi'} = (l - l + 1)p_j = p_j \geqslant 0$. Таким образом, $l^{\varphi} = l^{\varphi'}$ тогда и только тогда, когда $p_j = 0$.

Операция 2 (перенесение листа к ненасыщенной вершине). Выберем вершину v, расположенную на уровне $s \leqslant l-1$, имеющую менее D потомков (назовем такую вершину *ненасыщенной*); $\mu(v) = b^{(i_1)}b^{(i_2)}\ldots b^{(i_s)}$ — метка вершины v и символ $b \in B$ не используются в качестве меток ребер от вершины v к потомкам. Удалим из графа G лист u и ребро, соединявшее этот лист с потомком, а затем добавим лист u в качестве потомка вершины v, а ребру (v, u) припишем метку b. Тогда новое кодирование φ' будет отличаться от φ только тем, что $\varphi'(a^{(j)}) = \mu(u) = b^{(i_1)}b^{(i_2)}\ldots b^{(i_s)}b$.

Вычислим разность $l^{\varphi} - l^{\varphi'} = (l - (s+1))p_j = ((l-1) - s)p_j \geqslant 0$. Таким образом, $l^{\varphi} = l^{\varphi'}$ тогда и только тогда, когда $p_j = 0$ либо $s = l - 1$.

Замечание 7.7. Заметим, что при использовании операции 1 и 2 средняя длина кодового слова может только уменьшиться. Поэтому применение данных операций к кодовому дереву, соответствующему оптимальному кодированию, оставляет неизменным среднюю длину кодового слова и, следовательно, позволяет строить новые кодовые деревья, соответствующие некоторым другим оптимальным кодам.

Лемма 7.7. *Существует такое оптимальное D-ичное префиксное алфавитное кодирование φ'_0, для которого кодовое дерево G' насыщенное.*

Доказательство. Рассмотрим некоторое оптимальное кодирование φ_0 и соответствующее кодовое дерево G. Пусть L — высота кодового дерева G, s — количество листьев, находящихся на уровне L. Математической индукцией по

L докажем, что существует алгоритм, который за конечное число операций 1 и 2 из оптимального кодового дерева G позволяет получить оптимальное насыщенное кодовое дерево G'.

В случае $L = 1$ обязательно выполнено неравенство $2 \leqslant m \leqslant D$ и кодовое дерево G уже является насыщенным, у которого при $m < D$ особая вершина — корень.

Предположим, что $L > 1$ и для всех деревьев высотой, меньшей L, показано существование алгоритма построения оптимального насыщенного кодового дерева G'. Рассмотрим кодовое дерево G, имеющее высоту L. Покажем, что для $\forall s \in \{1, \ldots, D^L\}$ существует искомый алгоритм.

Если $s = 1$, то на уровне L имеется одиночный лист u, который соответствует некоторому символу $a \in A$. С помощью операции 1 удалим одиночный лист u. Как отмечалось ранее, применение данной операции приводит к новому оптимальному кодированию, однако соответствующее кодовое дерево уже будет иметь высоту меньше L, а для такого дерева согласно индукционному предположению искомый алгоритм существует.

Рассмотрим случай, когда $s > 1$ и в графе G на уровне $t < L - 1$ есть ненасыщенная вершина w. Выберем произвольный лист u на уровне L, соответствующий некоторому символу $a \in A$. Перенесем выбранный лист к вершине w, используя операцию 2. Новый код также будет оптимальным, при этом величина s уменьшилась на 1. Будем повторять описанную операцию до тех пор, пока либо s не станет равной 1, либо все ненасыщенные вершины не окажутся только на уровне $L - 1$. Если s станет равной единице, то поступим так же как и в предыдущем случае.

Остается разобрать ситуацию, когда $s > 1$ и в графе G вершины $v_1, \ldots, v_k$ на уровне $L - 1$ ненасыщенные, а на всех предыдущих уровнях все вершины насыщенные; вершина v_i имеет m_i потомков ($i \in \{1, \ldots, k\}$). Тогда $1 \leqslant m_i < D$, $\sum_{i=1}^{k} m_i = s_0 < s$. Поделим s_0 на D с остатком: $s_0 = qD + r$. Поскольку $m_i < D$, то $q < k$. Далее поступим следующим образом: будем переносить потомков вершин $v_2, \ldots, v_k$ вершине v_1, используя операцию 2, до тех пор, пока либо v_1 не станет насыщенной вершиной, либо у вершин $v_2, \ldots, v_k$ не закончатся потомки. Затем будем переносить потомков вершин $v_3, \ldots, v_k$ вершине v_2 и т. д. В результате получим, что вершины $v_1, \ldots, v_q$ окажутся насыщенными, у вершины v_{q+1} ($q+1 \leqslant k$) будет r потомков, а остальные вершины $v_{q+2}, \ldots, v_k$ превратятся в листья. Если $r = 0$, то вершина v_{q+1} — лист и построенное дерево является насыщенным. Если $2 \leqslant r < D$, то вершина v_{q+1} — особая, но построенное дерево все равно насыщенное. Если $r = 1$, то, используя операцию 1, удалим у вершины v_{q+1} одиночный лист, при этом вершина v_{q+1} сама станет листом, и мы снова получим насыщенное дерево. □

Упражнение 7.4. Пусть распределение $p(\cdot)$ таково, что $\forall i \in \{1, \ldots, m\}$, $p_i > 0$, кодовое дерево G соответствует оптимальному D-ичному префиксному алфавитному кодированию.

а) Может ли в графе G существовать ненасыщенная вершина на не предпоследнем уровне?

б) Может ли в графе G найтись вершина, у которой ровно один потомок?

в) Могут ли в графе G на предпоследнем уровне все вершины быть ненасыщенными?

Определение 7.16. *Префиксное алфавитное кодирование* φ *называется приведенным, если соответствующее кодовое дерево* G *насыщено и, кроме того:*

1) *если в дереве* G *на предпоследнем уровне есть особая вершина, то ее* D_0 *потомкам соответствуют символы алфавита* A *с* D_0 *наименьшими вероятностями из набора* $\{p_1, \ldots, p_m\}$*;*

2) *если в дереве* G *нет особой вершины, то имеется неконцевая вершина на предпоследнем уровне,* D *потомкам которой соответствуют символы алфавита* A *с* D *наименьшими вероятностями из набора* $\{p_1, \ldots, p_m\}$.

Пусть m_0 — такое наименьшее натуральное число вида $m_0 = 1 + k(D-1)$, что $m_0 \geqslant m$. Положим,

$$S = \begin{cases} D, \, m = m_0, \\ m - m_0 + D, \, m < m_0, \end{cases}$$

т. е. S — это либо количество потомков особой вершины, если такая вершина существует в насыщенном D-ичном дереве с m листьями, либо число D, если особой вершины в насыщенном дереве не существует.

Замечание 7.8. По лемме 7.4 определение приведенного префиксного алфавитного кодирования можно переформулировать следующим образом: префиксное алфавитное кодирование φ называется приведенным, если S наименее вероятным символам алфавита A соответствуют кодовые слова одинаковой наибольшей длины, которые различаются лишь последним символом.

Лемма 7.8. *Существует приведенное оптимальное* D*-ичное префиксное алфавитное кодирование* φ.

Доказательство. Согласно лемме 7.7 существует оптимальное D-ичное префиксное алфавитное кодирование φ', которому соответствует насыщенное кодовое дерево G'.

Пусть L — высота дерева G'. Тогда по следствию 7.5 на этом уровне расположены листья, которым приписаны символы алфавита A с наименьшими вероятностями из набора $\{p_1, \ldots, p_m\}$. Если переставить эти символы между собой, то средняя длина кода не изменится, а дерево останется насыщенным. Поскольку таким способом можно получить любую перестановку этих символов, то можно получить и перестановку, требуемую для того, чтобы соответствующее кодирование стало приведенным. □

Рассмотрим еще одну операцию над приведенным кодированием.

Операция 3 (редукция приведенного кодирования). Пусть $m > D$, $\varphi : A \to B^*$ — приведенное D-ичное префиксное алфавитное кодирование. Предположим, что кодовые слова $\varphi(a^{(m-S+1)}), \ldots, \varphi(a^{(m)})$ для S символов $a^{(m-S+1)}, \ldots, a^{(m)}$ с наименьшими вероятностями имеют одинаковую длину L, общий префикс $b^{(j_1)}, \ldots, b^{(j_{L-1})}$ и различаются между собой только в последнем

символе.

Рассмотрим алфавит $A^{(1)} = \{a^{(1)}, \ldots, a^{(m-S)}, \sigma\}$, отличающийся от алфавита A удалением S символов $a^{(m-S+1)}, \ldots, a^{(m)}$ и добавлением одного символа σ, а также заданное на алфавите $A^{(1)}$ распределение вероятностей $p^{(1)}$ такое, что $\forall i \in \{1, \ldots, m-S\}$ выполнено $p^{(1)}(a^{(i)}) = p_i^{(1)} = p(a^{(i)}) = p_i$ и $p^{(1)}(\sigma) = \alpha = \sum_{i=m-S+1}^{m} p(a^{(i)})$. Зададим кодирование $\varphi^{(1)} : A^{(1)} \to B^*$ соотношениями

$$\varphi^{(1)}(a^{(i)}) = \varphi(a^{(i)}),\ i \in \{1, \ldots, m-S\},$$
$$\varphi^{(1)}(\sigma) = b^{(j_1)}, \ldots, b^{(j_{L-1})}.$$

Кодирование $\varphi^{(1)}$ будем называть редуцированным по отношению к кодированию φ.

Лемма 7.9. *Пусть φ — приведенное D-ичное префиксное алфавитное кодирование и $\varphi^{(1)}$ — редуцированное кодирование. Кодирование φ оптимально для распределения $p(\cdot)$ тогда и только тогда, когда кодирование $\varphi^{(1)}$ оптимально для распределения $p^{(1)}$.*

Доказательство. Учитывая определение операции редукции, вычислим разность

$$l^{\varphi} - l^{\varphi^{(1)}} = \sum_{i=1}^{m} l^{(i)} p_i - \sum_{i=1}^{m-S} l^{(i)} p_i - (L-1)\alpha = \sum_{i=m-S+1}^{m} l^{(i)} p_i - (L-1)\alpha =$$

$$= L \sum_{i=m-S+1}^{m} p_i - (L-1)\alpha = L\alpha - (L-1)\alpha = \alpha.$$

Отсюда получим

$$l^{\varphi} = l^{\varphi^{(1)}} + \alpha. \tag{7.24}$$

Далее доказательство в обе стороны проведем методом от противного. Пусть кодирование φ оптимально для распределения $p(\cdot)$, однако кодирование $\varphi^{(1)}$ не оптимально для распределения $p^{(1)}$. Тогда существует оптимальное D-ичное префиксное алфавитное кодирование $\psi^{(1)} : A^{(1)} \to B^*$. В силу оптимальности имеет место неравенство $l^{\psi^{(1)}} < l^{\varphi^{(1)}}$. Зададим кодирование $\psi : A \to B^*$ следующим образом:

$$\psi(a^{(i)}) = \psi^{(1)}(a^{(i)}),\ i \in \{1, \ldots, m-S\},$$
$$\psi(a^{(m-S+j)}) = \psi^{(1)}(\sigma) b^{(j)},\ j \in \{1, \ldots, S\}.$$

Поскольку по определению величины S для нее выполнено неравенство $S \leqslant D$, то описанное задание ψ корректно. Тогда согласно (7.24)

$$l^{\psi} = \sum_{i=1}^{m-S} p_i |\psi(a^{(i)})| + \sum_{i=m-S+1}^{m} p_i (|\psi^{(1)}(\sigma)) + 1| =$$

$$= \sum_{i=1}^{m-S} p_i |\psi^{(1)}(a^{(i)})| + \alpha(|\psi^{(1)}(\sigma)| + 1) = l^{\psi^{(1)}} + \alpha < l^{\varphi^{(1)}} + \alpha = l^{\varphi}.$$

Получили противоречие с тем, что кодирование φ оптимальное.

Пусть кодирование $\varphi^{(1)}$ оптимально для распределения $p^{(1)}$, но предположим, что кодирование φ не оптимально для распределения $p(\cdot)$. Тогда согласно лемме 7.8 существует приведенное оптимальное D-ичное префиксное алфавитное кодирование $\psi : A \to B^*$. Вследствие оптимальности имеет место неравенство $l^{\psi} < l^{\varphi}$. Поскольку кодирование ψ приведенное, применим к нему операцию редукции и получим новое кодирование $\psi^{(1)} : A^{(1)} \to B^*$. Аналогично (7.24) можно показать, что $l^{\psi} = l^{\psi^{(1)}} + \alpha$, отсюда имеем

$$l^{\psi^{(1)}} + \alpha = l^{\psi} < l^{\varphi} = l^{\varphi^{(1)}} + \alpha \Rightarrow l^{\psi^{(1)}} < l^{\varphi^{(1)}}.$$

Последнее неравенство приводит к противоречию с тем, что кодирование $\varphi^{(1)}$ оптимальное. □

7.7. АЛГОРИТМ ПОСТРОЕНИЯ ПРЕФИКСНЫХ КОДОВ ФАНО

В следующих разделах будут разобраны несколько основных алгоритмов построения алфавитных префиксных кодов. Исторически одним из первых таких алгоритмов является алгоритм Фано. Он весьма простой и при этом позволяет получать коды с достаточно малой средней длиной — в большинстве случаев применения алгоритма Фано получается либо оптимальный код, либо код, близкий к оптимальному. Описание данного алгоритма приведем для случая построения двоичных кодов ($D = 2$), поскольку, во-первых, двоичные коды наиболее важны для практических приложений и, во-вторых, на примере двоичного кодового алфавита легче понять идеи, положенные в основу этих алгоритмов.

Предположим, что на алфавите источника сообщений $A = (a^{(1)}, \ldots, a^{(m)})$, $(m \geqslant 2)$ задано распределение вероятностей $p(\cdot)$, $p_i = p(a^{(i)}) = \mathsf{P}\{\xi = a^{(i)}\}$, причем вероятности символов упорядочены в невозрастающем порядке, т. е. $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_m$.

Алгоритм Фано ($D = 2, B = \{0, 1\}$).

1) Выберем наименьшее число k, $1 \leqslant k < m$, чтобы величина

$$\left| \sum_{i=1}^{k} p_i - \sum_{i=k+1}^{m} p_i \right|$$

была наименьшей. Разобьем множество A на два подмножества $A_0 = \{a^{(1)}, \ldots, a^{(k)}\}$, $A_1 = \{a^{(k+1)}, \ldots, a^{(m)}\}$.

2) Рассмотрим подмножество $A_{i_1, \ldots, i_t}$, $i_1, \ldots, i_t \in B$. Если оно состоит из одного элемента a, то определим код этого символа как $\varphi(a) = i_1 i_2 \ldots i_t$.

Если подмножество $A_{i_1, \ldots, i_t} = \{a_l, \ldots, a_s\}$ состоит более чем из одного эле-

мента ($s > l$), то выберем наименьшее k, $l \leqslant k < s$ такое, чтобы величина

$$\left| \sum_{i=l}^{k} p_i - \sum_{i=k+1}^{s} p_s \right|$$

была наименьшей. Разобьем множество $A_{i_1,\ldots,i_t}$ на два подмножества $A_{i_1,\ldots,i_t,0} = \{a_l, \ldots, a_k\}$, $A_{i_1,\ldots,i_t,1} = \{a_{k+1}, \ldots, a_s\}$.

3) Шаг 2 повторяется до тех пор, пока все подмножества не станут одноэлементными.

Пример 7.5. Рассмотрим алфавит $A = \{a, b, c, d, e, f\}$ с вероятностями $p = \{0{,}36; 0{,}18; 0{,}18; 0{,}12; 0{,}09; 0{,}07\}$. Последовательное разбиение подмножеств представим в виде корневого дерева (рис. 7.5).

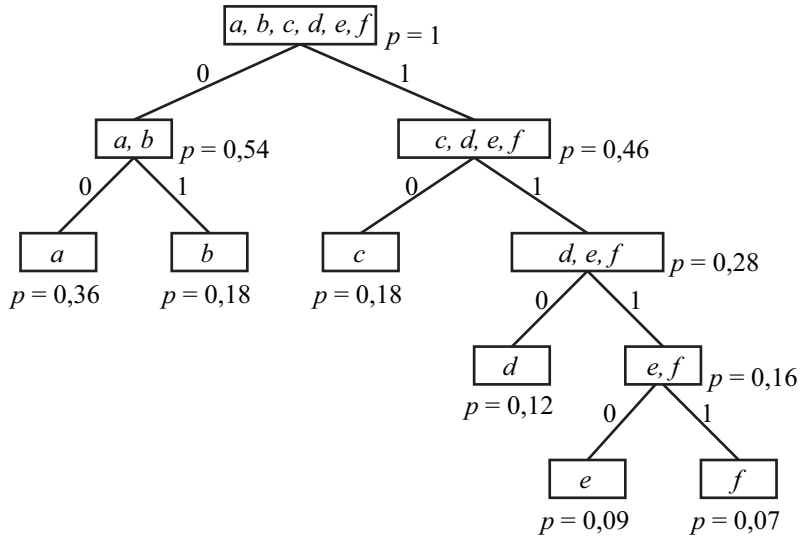


Рис. 7.5. Пример алгоритма Фано

В результате получится кодирование $\varphi(a) = 00$, $\varphi(b) = 01$, $\varphi(c) = 10$, $\varphi(d) = 110$, $\varphi(e) = 1110$, $\varphi(f) = 1111$ со средней длиной кодового слова $l^{\varphi} = (0{,}36 + 0{,}18 + 0{,}18) \cdot 2 + 0{,}12 \cdot 3 + (0{,}09 + 0{,}07) \cdot 4 = 2{,}44$ бита.

Замечание 7.9. В описанном алгоритме на каждом шаге предлагается выбрать наименьшее k. Это требование нельзя опустить, так как в противном случае может возникнуть неоднозначность в выборе значения k, но это требование можно заменить на любое другое, гарантирующее однозначность выбора k. Например, можно выбрать наибольшее значение k.

Упражнение 7.5. Докажите, что описанный алгоритм действительно задает префиксное алфавитное кодирование.

Замечание 7.10. Данный алгоритм легко обобщается на случай D-ичных кодов ($D > 2$). Единственным отличием будет то, что на каждом шаге множества нужно будет разбивать не на два, а не более чем на D наиболее равновероятных подмножеств (если изначально множество содержало не менее D

элементов, то оно разбивается ровно на D подмножеств, иначе оно разбивается на m одноэлементных подмножеств).

7.8. АЛГОРИТМ ПОСТРОЕНИЯ ПРЕФИКСНЫХ КОДОВ ШЕННОНА

Еще одним простым в реализации, но позволяющим получать коды с достаточно малой средней длиной является алгоритм Шеннона.

Как и в предыдущем разделе, предположим, что на алфавите источника сообщений $A = (a^{(1)}, \ldots, a^{(m)})$, $(m \geqslant 2)$ задано распределение вероятностей $p(\cdot)$, $p_i = p(a^{(i)}) = \mathsf{P}\left\{\xi = a^{(i)}\right\}$, причем вероятности символов упорядочены в невозрастающем порядке, т. е. $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_m$. Кроме того, будем считать, что все вероятности положительны, т. е. $p_i > 0$, $\forall i \in \{1, \ldots, m\}$.

Алгоритм Шеннона ($\boldsymbol{D = 2, B = \{0, 1\}}$).

1) Найдем натуральные числа $l^{(i)}$, $i \in \{1, \ldots, m\}$, такие, что выполнены неравенства

$$\frac{1}{2^{l^{(i)}}} \leqslant p_i < \frac{1}{2^{l^{(i)}-1}}. \tag{7.25}$$

2) Вычислим накопленные вероятности:

$$P_1 = 0,\ P_2 = p_1,\ P_3 = p_1 + p_2,\ \ldots, P_m = \sum_{i=1}^{m-1} p_i.$$

3) Для каждого $i \in \{1, \ldots, m\}$ определим код $\varphi(a^{(i)})$ как $l^{(i)}$ символов, стоящих после запятой в двоичном представлении числа P_i.

4) Операция усечения. Для всех i, пробегающих все значения от 1 до m, в кодовом слове $\varphi(a^{(i)})$ справа налево последовательно удаляются символы, пока код $\varphi(a^{(i)})$ не является префиксом ни для какого другого кодового слова $\varphi(a^{(j)})$, $i \neq j$.

Пример 7.6. Выполнение алгоритма Шеннона приведено в табл. 7.2.

Таблица 7.2

Шаги алгоритма Шеннона

	0	1	2	3	4
A	p_i	$l^{(i)}$	P_i	$\varphi_1()$	$\varphi_2()$
a	0,36	2	0,00	00	00
b	0,18	3	0,36 = 0,010	010	01
c	0,18	3	0,54 = 0,100	100	100
d	0,12	4	0,72 = 0,1011	1011	101
e	0,09	4	0,84 = 0,1101	1101	110
f	0,07	4	0,93 = 0,1110	1110	111

В этом примере φ_1 означает кодирование, которое получено на третьем шаге до применения усечения, а φ_2 — кодирование после применения операции усе-

чения. Таким образом, в результате получим кодирование: $\varphi(a) = 00$, $\varphi(b) = 01$, $\varphi(c) = 100$, $\varphi(d) = 101$, $\varphi(e) = 110$, $\varphi(f) = 111$. Средняя длина кодового слова при таком кодировании равна $l^{\varphi_2} = 2 \cdot (0{,}36 + 0{,}18) + 3 \cdot (0{,}18 + 0{,}12 + 0{,}9 + 0{,}7) = = 1{,}08 + 1{,}38 = 2{,}46$. На рис. 7.6 представлено соответствующее кодовое дерево.

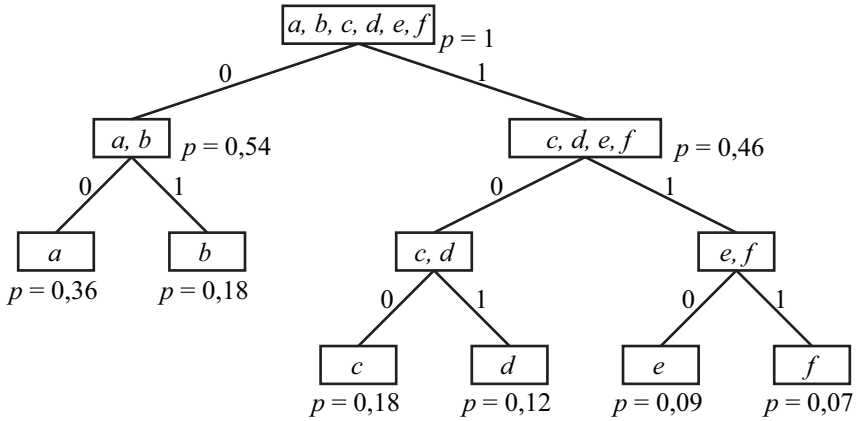


Рис. 7.6. Кодовое дерево для кода Шеннона

Утверждение 7.3. Алгоритм Шеннона задает префиксное алфавитное кодирование.

Доказательство. Очевидно, что если после третьего шага алгоритма был получен префиксный код, то после выполнения четвертого шага он останется префиксным. Поэтому докажем, что на третьем шаге действительно определяется префиксный код.

Выберем два различных символа $a^{(i)}$ и $a^{(j)}, i < j$, алфавита A. Согласно модельным предположениям $p_i \geqslant p_j$, тогда из (7.25) имеем, что $l^{(i)} \leqslant l^{(j)}$. А это означает, что только a_i может быть префиксом a_j.

Рассмотрим разность

$$P_j - P_i \geqslant P_{i+1} - P_i = p_i \geqslant \frac{1}{2^{l^{(i)}}}.$$

Следовательно, в двоичной записи P_i и P_j отличаются не позднее чем в $l^{(i)}$-й цифре после запятой, поэтому a_i не может быть префиксом a_j. □

Утверждение 7.4. Для средней длины кодового слова при алфавитном кодировании φ, построенному по алгоритму Шеннона, справедлива оценка

$$l^{\varphi} < \mathbf{H}\{\xi\} + 1,$$

где ξ — выходной символ ИДС без памяти $\langle A, p(a)\rangle$, а энтропия вычисляется в битах ($b = 2$).

Доказательство. Из неравенства (7.25) получим, что $l^{(i)} < -\log_2 p_i + 1$, тогда

$$l^{\varphi} = \sum_{i=1}^{m} l^{(i)} p_i < -\sum_{i=1}^{m} p_i \log_2 p_i + \sum_{i=1}^{m} p_i = \mathbf{H}\{\xi\} + 1. \qquad \square$$

Замечание 7.11. Как и алгоритм Фано, алгоритм Шеннона легко обобщается на случай D-ичных кодов ($D > 2$). В данном алгоритме вместо двоичной системы счисления надо будет рассматривать D-ичную.

7.9. АЛГОРИТМ ПОСТРОЕНИЯ ОПТИМАЛЬНЫХ ПРЕФИКСНЫХ КОДОВ ХАФФМАНА

Пусть на алфавите источника сообщений $A = (a^{(1)}, \ldots, a^{(m)})$, $(m \geqslant 2)$ задано распределение вероятностей $p(\cdot)$, $p_i = p(a^{(i)}) = \mathsf{P}\left\{\xi = a^{(i)}\right\}$, причем вероятности символов упорядочены в невозрастающем порядке, т. е. $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_m$.

Обозначим m_0 — такое наименьшее натуральное число вида $m_0 = 1+k(D-1)$, что $m_0 \geqslant m$. Положим,

$$S = \begin{cases} D, \; m = m_0, \\ m - m_0 + D, \; m < m_0. \end{cases} \tag{7.26}$$

Алгоритм Хаффмана ($D \geqslant 2, B = \{b^{(1)}, \ldots, b^{(D)}\}$).

1-й этап — слияние.

0) $A^{(0)} \leftarrow A$, $j \leftarrow 0$.

1) Пока $m > D$, выполняются шаги 2–4.

2) Осуществить слияние S наиболее маловероятных символов

$$A^{(j)} \to A^{(j+1)} = \left\{a^{(1)}, \ldots, a^{(m-S)}, \sigma\right\},$$

$$p^{(j)}() \to p^{(j+1)}() : p^{(j+1)}(a^{(i)}) = p^{(j)}(a^{(i)}),\; p^{(j+1)}(\sigma) = \sum_{i=m-S+1}^{m} p^{(j)}(a^{(i)}),$$

где $\{a^{(m-S+1)}, \ldots, a^{(m)}\} \to \sigma$.

3) Упорядочить символы алфавита $A^{(j+1)}$ по невозрастанию вероятностей распределения $p^{(j+1)}()$.

4) $m \leftarrow m - S$, $i \leftarrow i + 1$, вычислить новое S согласно (7.26) и возвратиться к шагу 1.

Замечание 7.12. Согласно лемме 7.4, если m_0 такое наименьшее целое число вида $m_0 = 1 + k(D-1)$, что $m_0 \geqslant m$, то

$$m = 1 + (k-1)(D-1) + S - 1.$$

При этом необходимо отметить, что при всех проходах шагов 2–4, начиная со второго, $S = D$. Поэтому на первом проходе мощность алфавита уменьшается на $S-1$, а на каждом из последующих — на $D-1$. Таким образом, в результате слияний получим цепочку

$$A = A^{(0)} \to A^{(1)} \to \ldots \to A^{(k-1)}, \tag{7.27}$$

причем если изначально $m > D$, то $|A^{(k-1)}| = D$.

2-й этап — кодирование.

Пусть на первом этапе алгоритма построена цепочка слияний (7.27).

1) В качестве кодирования $\varphi^{(k-1)} : A^{(k-1)} \to B^*$ выбрать произвольное взаимно однозначное соответствие между алфавитами $A^{(k-1)}$ и B; $j \leftarrow k-2$.

2) Задать кодирование $\varphi^{(j)} : A^{(j)} \to B^*$ следующим образом. Если алфавит $A^{(j+1)}$ был получен из алфавита $A^{(j)}$ слиянием некоторых S символов $a'_1, \ldots, a'_S$ в один новый символ σ', то кодирование $\varphi^{(j)}$ получается из уже построенного кодирования $\varphi^{(j+1)}$ заменой одного равенства $\varphi^{(j+1)}(\sigma') = w \in B^*$ на S равенств $\varphi^{(j)}(a'_i) = w||b^{(i)}$, $i \in \{1, \ldots, S\}$.

3) $j \leftarrow j-1$.

4) Если $j \geqslant 0$, то перейти на шаг 2.

5) $\varphi \leftarrow \varphi^{(0)}$.

Замечание 7.13. Если после первого этапа $m = |A^{(k-1)}| < D$, то это означает, что изначально $m < D$. В этом случае $k = 1$ и алгоритм тоже работает, за исключением того, что на первом шаге надо построить инъективное (а не биективное) отображение из A на B^*.

Пример 7.7. Построим оптимальное троичное кодирование, используя метод Хаффмана, для алфавита $A = \{a, b, c, d, e, f, g, h\}$ с вероятностями $p = \{0{,}2;\ 0{,}14;\ 0{,}13;\ 0{,}13;\ 0{,}12;\ 0{,}12;\ 0{,}08;\ 0{,}08\}$. В данном случае $m = 8$, поэтому наименьшее $m_0 = 1 + 3k \geqslant 8$ равно 9, а значит, $S = m - m_0 + D = 2$, т. е. в насыщенном троичном дереве есть особая вершина с двумя потомками. Шаги алгоритма Хаффмана представлены на рис. 7.7.

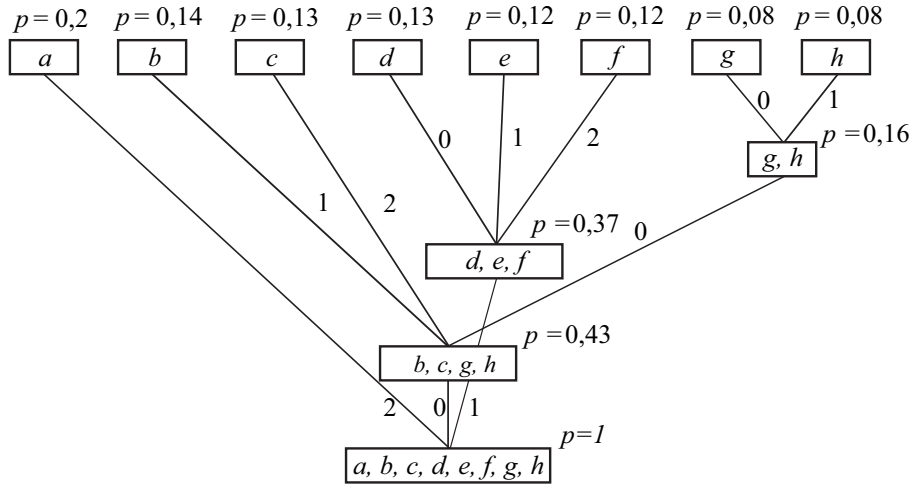


Рис. 7.7. Пример работы алгоритма Хаффмана при $D = 3$

В результате получается кодирование $\varphi(a) = 0$, $\varphi(b) = 10$, $\varphi(c) = 11$, $\varphi(d) = 20$, $\varphi(e) = 21$, $\varphi(f) = 22$, $\varphi(g) = 120$, $\varphi(h) = 121$ со средней длиной кодового слова $l^{\varphi} = 0{,}2 \cdot 1 + (0{,}14 + 0{,}13 + 0{,}13 + 0{,}12 + 0{,}12) \cdot 2 + (0{,}08 + 0{,}08) \cdot 3 = 1{,}96$ троичных символов.

Теорема 7.7. *Для любых m, $D \geqslant 2$ алгоритм Хаффмана строит оптимальное приведенное D-ичное префиксное алфавитное кодирование.*

Доказательство. Пусть $m > D$. Алгоритм Хаффмана на первом этапе строит цепочку слияний алфавитов (7.27), а на втором этапе, уже в обратном порядке, — цепочку кодирований $\varphi = \varphi^{(0)} \leftarrow \varphi^{(1)} \leftarrow \ldots \leftarrow \varphi^{(k-1)}$, или, что то же самое, цепочку кодовых деревьев $G = G^{(0)} \leftarrow G^{(1)} \leftarrow \ldots \leftarrow G^{(k-1)}$.

При этом из описания алгоритма ясно, что все кодовые деревья $G^{(i)}$ в цепочке являются насыщенными (поскольку на каждом шаге, за исключением, быть может, последнего из листов, опускается D потомков), все кодирования $\varphi^{(i)}$ — приведенными (действительно, вершина, из которой исходят ребра в листья с наименьшими вероятностями, всегда находится на предпоследнем уровне), и каждое кодирование $\varphi^{(i)}$ — редуцированным по отношению к кодированию $\varphi^{(i-1)}$, $i \in \{1, \ldots, k-1\}$.

Поскольку очевидно, что $\varphi^{(k-1)}$ — оптимальное по построению, то согласно лемме 7.9 кодирование $\varphi^{(k-2)}$ тоже оптимальное, откуда следует оптимальность $\varphi^{(k-3)}$ и т. д. В конце концов получим, что $\varphi = \varphi^{(0)}$ также оптимальное кодирование.

Если $m \leqslant D$, то первый этап не проводится вовсе, а на втором этапе на первом же шаге задается оптимальное кодирование. □

Приведем алгоритм Хаффмана для случая $D = 2$ как наиболее важного с точки зрения практического применения, при этом сформулируем его в терминах кодовых деревьев. Как и прежде, символы алфавита A упорядочены по невозрастанию вероятностей: $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_m$.

Алгоритм Хаффмана ($D = 2, B = \{0, 1\}$).

1-й этап — построение двоичного дерева. Построим двоичное дерево с m листьями, начиная с листьев и продвигаясь к корню. Возьмем в качестве листьев дерева символы $a^{(1)}, \ldots, a^{(m)}$, которым приписаны вероятности $p_1, \ldots, p_m$.

0) Пусть $A^{(0)} = \{a^{(1)}, \ldots, a^{(m)}\}$ — множество рассматриваемых в данный момент вершин; $j \leftarrow 0$.

1) Выбрать две вершины $a^{(m-1)}$ и $a^{(m)}$ из $A^{(j)}$ с наименьшими вероятностями p_{m-1} и p_m и добавить в дерево новую вершину $\sigma = a^{(m-1)} \cup a^{(m)}$, которой приписать вероятность $p_{m-1} + p_m$. Вершину σ соединить ребрами с вершинами $a^{(m-1)}$ и $a^{(m)}$ и объявить общим предком для этих вершин. Ребро от σ к $a^{(m-1)}$ пометить символом 0, а ребро от σ к $a^{(m)}$ — символом 1. Получен новый набор рассматриваемых вершин $A^{(j+1)} = \{a^{(1)}, \ldots, a^{(m-2)}, \sigma\}$ с соответствующем набором вероятностей $\{p_1, \ldots, p_{m-2}, p_{m-1} + p_m\}$.

2) Отсортировать множество вершин $A^{(j+1)}$ в порядке невозрастания вероятностей; $m \leftarrow m - 1$; $j \leftarrow j + 1$.

3) Если $m > 1$, то перейти к шагу 1.

4) Множество вершин $A^{(m-1)}$ состоит из одного элемента, который объявляется корнем.

2-й этап — кодирование. Чтобы получить кодовое слово $\varphi(a^{(i)})$, последовательно считываются метки ребер на пути от корня дерева к листу $a^{(i)}$.

Пример 7.8. Рассмотрим алфавит с вероятностями из примеров 7.5 и 7.6: $A = \{a, b, c, d, e, f\}$ с вероятностями $p = \{0{,}36;\ 0{,}18;\ 0{,}18;\ 0{,}12;\ 0{,}09;\ 0{,}07\}$. Результат работы алгоритма Хаффмана представлен на рис. 7.8.

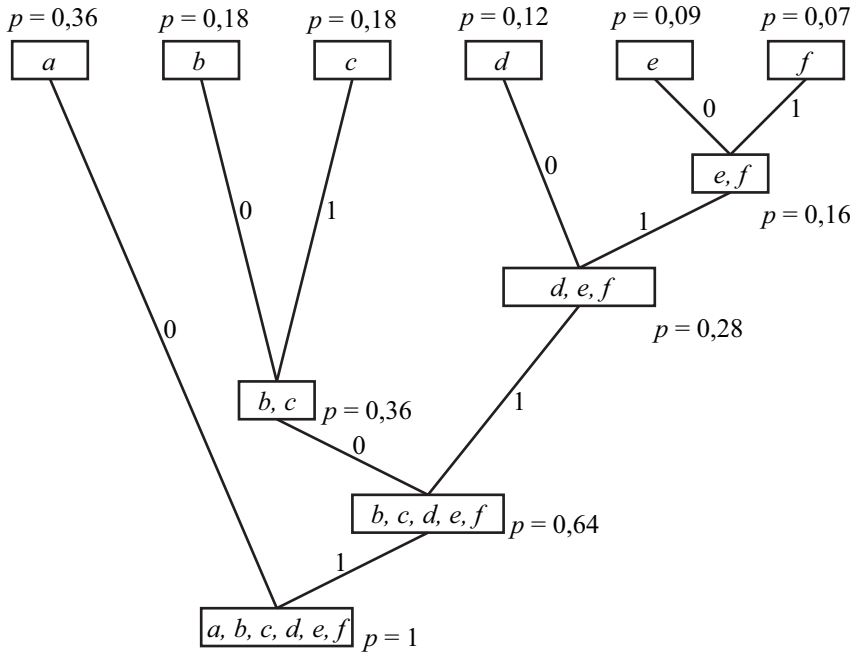


Рис. 7.8. Пример работы алгоритма Хаффмана при $D = 2$

В результате получается кодирование $\varphi(a) = 0$, $\varphi(b) = 100$, $\varphi(c) = 101$, $\varphi(d) = 110$, $\varphi(e) = 1110$, $\varphi(f) = 1111$ со средней длиной кодового слова $l^{\varphi} = 0{,}36 + (0{,}18 + 0{,}18 + 0{,}12) \cdot 3 + (0{,}09 + 0{,}07) \cdot 4 = 2{,}44$ бита.

Результаты примеров 7.5, 7.6, 7.8 можно представить в одной табл. 7.3.

Таблица 7.3

Методы построения префиксных кодов

A	p	Код Фано	Код Шеннона	Код Хаффмана
a	0,36	00	00	0
b	0,18	01	01	100
c	0,18	10	100	101
d	0,12	110	101	110
e	0,09	1110	110	1110
f	0,07	1111	111	1111
l^{φ}	—	2,44	2,46	2,44

Как видно из табл. 7.3, алгоритм Шеннона не всегда дает оптимальное алфавитное префиксное кодирование.

Замечание 7.14. В описанном алгоритме имеется неоднозначность в выборе двух вершин с наименьшими вероятностями, если имеется более двух вершин с одинаковыми наименьшими вероятностями. В приведенном алгоритме в случае такой неоднозначности выбираются две вершины с наибольшими номерами в списке вершин. Однако согласно теореме 7.7 при любом таком соглашении построенный код будет оптимальным.

Пример 7.9. Рассмотрим еще один пример, иллюстрирующий, что алгоритм Фано также не всегда дает оптимальное алфавитное префиксное кодирование. Пусть задан алфавит $A = \{a, b, c, d, e, f\}$ с вероятностями $p = \{0{,}5;\ 0{,}2;\ 0{,}09;\ 0{,}08;\ 0{,}07;\ 0.06\}$. Применив описанные выше алгоритмы, получим коды, приведенные в табл. 7.4.

Таблица 7.4

Методы построения префиксных кодов

A	p	Код Фано	Код Шеннона	Код Хаффмана
a	0,5	0	0	0
b	0,2	100	100	10
c	0,09	101	101	1100
d	0,08	110	1100	1101
e	0,07	1110	1101	1110
f	0,06	1111	111	1111
l^{φ}	—	2,13	2,15	2,1

Упражнение 7.6. Убедиться, что для алфавита A с заданными вероятностями из примера 7.9 алгоритмы Фано, Шеннона и Хаффмана действительно дают коды, записанные в табл. 7.4.

7.10. СЛОВАРНО-ОРИЕНТИРОВАННЫЕ АЛГОРИТМЫ СЖАТИЯ ИНФОРМАЦИИ (МЕТОДЫ ЛЕМПЕЛЯ – ЗИВА)

Кратко рассмотрим еще один менее математически обоснованный, но весьма эффективный на практике подход к кодированию, который обычно называют словарно-ориентированным.

Одним из первых алгоритмов данного подхода является алгоритм LZ77, разработанный математиками Якобом Зивом (Y. Ziv) и Авраамом Лемпелом (A. Lempel). Большинство программ сжатия информации используют какую-либо модификацию LZ77. Одна из причин популярности алгоритмов LZ заключается в их простоте при высокой эффективности сжатия.

Основная идея LZ77 состоит в том, что второе и последующие вхождения некоторой строки символов в сообщении заменяются ссылками на ее первое вхождение. LZ77 использует часть уже просмотренного сообщения как словарь. Чтобы добиться сжатия, он пытается заменить очередной фрагмент сообщения на указатель в содержимое словаря.

LZ77 использует «скользящее» по сообщению окно, разделенное на две неравные части. Первая, бóльшая по размеру, включает уже просмотренную часть сообщения. Вторая, намного меньшая, представляет собой буфер, содержащий еще незакодированные символы входного потока. Обычно размер окна составляет несколько килобайт, а размер буфера — не более ста байт. Алгоритм пытается найти в словаре (бóльшей части окна) фрагмент, совпадающий с содержимым буфера.

Алгоритм LZ77 выдает коды, состоящие из трех элементов:

1) смещение в словаре относительно его начала подстроки, совпадающей с началом содержимого буфера;

2) длина этой подстроки;

3) первый символ буфера, следующий за подстрокой.

Пусть необходимо закодировать слово $M = m_0, \ldots, m_{N-1}$, символы которого принадлежат некоторому алфавиту: $m_i \in A$, $i \in \{0, \ldots, N-1\}$, N — длина кодируемого слова. Пусть заданы K — размер словаря и L — размер буфера ($K > L$). Допишем в начало слова M K искусственных «пустых» символов, т. е. положим $m_{-1} = m_{-2} = \ldots = m_{-K} = «\cdot» \notin A$.

Алгоритм кодирования LZ77.

1) $i \leftarrow 1$.

2) На данном шаге имеем словарь $D = m_{i-K}, \ldots, m_{i-1}$, буфер $B = m_i, \ldots, m_{i+L-1}$. Находится такое максимальное $s \geqslant 1$, что фрагмент $m_i, \ldots, m_{i+s-1}$ целиком содержится в словаре A, т. е. чтобы существовало $j \geqslant 0$, что $m_{i-K+j} = m_i, \ldots, m_{i-K+j+s-1} = m_{i+s-1}$. Стоит отметить, что на s накладывается два ограничения сверху: $s \leqslant L$ (фрагмент не может превосходить размер буфера) и $i + s < N - 1$ (последний символ должен быть закодирован особым образом).

3) Если такие s и j были найдены, то возвращается код (j, s, m_{i+s}), $i \leftarrow i + s + 1$; иначе возвращается код $(0, 0, m_i)$ и $i \leftarrow i + 1$.

4) Если $i < N - 1$, то перейти к шагу 2.

5) Иначе $i = N - 1$. В этом случае записывается код $(0, 0, m_{N-1})$. Последний символ не ищется в словаре, поскольку за ним нет символов.

На практике длина кода вычисляется следующим образом: длина подстроки (фрагмента) не может быть больше размера буфера, а смещение не может быть больше размера словаря минус один. Следовательно, длина двоичного кода смещения будет округленным в большую сторону $\log_2(K)$, а длина двоичного кода для длины подстроки будет округленным в большую сторону $\log_2(L + 1)$. Предполагается, что символ можно закодировать 8 битами (например, ASCII+).

Пример 7.10. Закодируем с помощью LZ77 сообщение $M =$ «красная краска». Положим, $K = 8$, $L = 5$. Результат работы алгоритма представлен в табл. 7.5.

Длина полученного кода равна $9 \cdot (3 + 3 + 8) = 126$ против $14 \cdot 8 = 112$ бит исходной длины строки. В примере 7.10 длина закодированного слова получилась больше длины исходного слова. Это связано с тем, что данный алгоритм не всегда дает выигрыш в длине (особенно для таких коротких сообщений).

Таблица 7.5

Пример кодирования LZ77

Словарь	Буфер	Код
«·· · · · · · ··»	«красн»	(0, 0, к)
«· · · · · · · к»	«расна»	(0, 0, р)
«· · · · ·· кр»	«асная»	(0, 0, а)
«· · · · · кра»	«сная »	(0, 0, с)
«· · ·· крас»	«ная к»	(0, 0, н)
«· · · красн»	«ая кр»	(5, 1, я)
«· красная»	« крас»	(0, 0, ‘ ’)
«красная »	«краск»	(0, 4, к)
«ая краск»	«а»	(0, 0, а)

Рассмотрим алгоритм восстановления исходного сообщения $M = m_0, \ldots, m_{N-1}$ по последовательности кодов. Пусть задан K — размер словаря. Будем считать, что символы $m_{-1} = m_{-2} = \ldots = m_{-K} = «·» \notin A$ уже восстановлены и равны искусственным «пустым» символам.

Алгоритм декодирования LZ77.

1) $i \leftarrow 0$.

2) На данном шаге имеем словарь $D = m_{i-K}, \ldots, m_{i-1}$. Из входного потока считывается очередной код (j, s, m) и восстанавливаются $s + 1$ символы: $m_i \leftarrow m_{i-K+j}, \ldots, m_{i+s-1} \leftarrow m_{i-K+j+s-1}$, $m_{i+s} \leftarrow m$.

3) $i \leftarrow i + s + 1$.

4) Если входной поток кодов непуст, то перейдем к шагу 2.

Пример 7.11. Декодируем сообщение, которое было закодировано в примере 7.10. Результат алгоритма декодирования представлен в табл. 7.6.

Таблица 7.6

Пример декодирования LZ77

Словарь	Входной код	Печать
«·· · · · · · ··»	(0, 0, к)	«к»
«· · · · · · · к»	(0, 0, р)	«р»
«· · · · ·· кр»	(0, 0, а)	«а»
«· · · · · кра»	(0, 0, с)	«с»
«· · ·· крас»	(0, 0, н)	«н»
«· · · красн»	(5, 1, я)	«ая»
«· красная»	(0, 0, ‘ ’)	« »
«красная »	(0, 4, к)	«краск»
«ая краск»	(0, 0, а)	«а»

Алгоритм кодирования LZ77 имеет ряд очевидных недостатков:

1) невозможность кодирования подстрок, отстоящих друг от друга на расстоянии, большем длины словаря;

2) длина подстроки, которую можно закодировать, ограничена размером буфера.

При этом если механически чрезмерно увеличивать размеры словаря и буфера, то это приведет лишь к снижению эффективности кодирования, так как с их ростом будут увеличиваться и длины кодов для смещения, и длины фрагментов, что сделает коды для коротких подстрок недопустимо большими. Кроме того, резко возрастет время работы алгоритма кодирования.

В 1978 г. авторами LZ77 был разработан алгоритм LZ78, лишенный названных недостатков. LZ78 не использует «скользящего» окна, он хранит словарь из уже просмотренных фраз. При старте алгоритма этот словарь содержит только одну пустую строку (строку длиной нуль). Алгоритм считывает символы сообщения до тех пор, пока накапливаемая подстрока входит целиком в одну из фраз словаря. Как только эта строка перестанет соответствовать всем фразам словаря, алгоритм генерирует код, состоящий из индекса строки в словаре, которая до последнего введенного символа содержала входную строку, и символа, нарушившего совпадение. Затем в словарь добавляется введенная подстрока. Если словарь уже заполнен, то из него предварительно удаляется менее всех используемая в сравнениях фраза.

Ключевым для размера получаемых кодов является размер словаря во фразах, потому что каждый код при кодировании по методу LZ78 содержит номер фразы в словаре. Таким образом, коды имеют постоянную длину, равную округленному в большую сторону двоичному логарифму размера словаря плюс, как правило, 8 бит для кодирования символа.

Пример 7.12. Применяя алгоритм LZ78, закодируем и декодируем сообщение «красная краска», используя словарь размером 16 фраз. Кодирование и декодирование представлены в табл. 7.7 и 7.8.

Таблица 7.7

Пример кодирования LZ78

Входная фраза (словарь)	Код	Позиция словаря
«»		0
«к»	(0, к)	1
«р»	(0, р)	2
«а»	(0, а)	3
«с»	(0, с)	4
«н»	(0, н)	5
«ая»	(3, я)	6
« »	(0, ‘ ’)	7
«кр»	(1, р)	8
«ас»	(3, с)	9
«ка»	(1, а)	10

Указатель на любую фразу такого словаря – это число от 0 до 15, для его кодирования достаточно четырех бит. Таким образом, длина полученного кода

равна $10(4+8) = 120$ бит.

Таблица 7.8

Пример декодирования LZ78

Входной код	Печать (словарь)	Позиция словаря
	«»	0
(0, к)	«к»	1
(0, р)	«р»	2
(0, а)	«а»	3
(0, с)	«с»	4
(0, н)	«н»	5
(3, я)	«ая»	6
(0, ‘ ’)	« »	7
(1, р)	«кр»	8
(3, с)	«ас»	9
(1, а)	«ка»	10

Алгоритмы LZ77, LZ78 разработаны математиками и могут использоваться свободно.

В 1984 г. Терри Уэлчем (T. Welch) был модифицирован LZ78 и получен алгоритм LZW.

Алгоритм кодирования LZW.

1) Инициализация словаря всеми возможными односимвольными фразами (обычно это 256 символов расширенного ASCII). Инициализация входной фразы w первым символом сообщения.

2) Считать очередной символ k из кодируемого сообщения.

3) Если k = КОНЕЦ_СООБЩЕНИЯ, то выдать код для w и завершить работу алгоритма; если фраза wk уже есть в словаре, то $w \leftarrow wk$ и перейти на шаг 2; иначе выдать код w, добавить wk в словарь, $w \leftarrow k$ и перейти к шагу 2.

При переполнении словаря, т. е. когда необходимо внести новую фразу в полностью заполненный словарь, из него удаляют либо наиболее редко используемую фразу, либо все фразы, отличающиеся от одинарного символа.

Как и в случае с LZ78, для LZW ключевым для размера получаемых кодов является размер словаря во фразах: LZW-коды имеют постоянную длину, равную округленному в большую сторону двоичному логарифму размера словаря.

Пример 7.13. Закодировать по алгоритму LZW сообщение «красная краска», используя словарь размером 512 фраз. Результаты работы алгоритма приведены в табл. 7.9. Через $ch(x)$ обозначен номер символа x в таблице ASCII+.

Длина полученного кода равна $12 \cdot 9 = 108$ битам.

При декодировании нужно придерживаться следующего правила. Словарь пополняется после считывания первого символа идущего за текущим кода, т. е. из фразы, соответствующей следующему после раскодированного коду, берется первый символ. Это правило позволяет избежать бесконечного цикла при рас-

Таблица 7.9

Пример кодирования LZW

Входная фраза wk (словарь)	Код для w	Позиция словаря
ASCII+		0 – 255
«кр»	ch(к)	256
«ра»	ch(р)	257
«ас»	ch(а)	258
«сн»	ch(с)	259
«на»	ch(н)	260
«ая»	ch(а)	261
«я »	ch(я)	262
« к»	ch( )	263
«кра»	256	264
«аск»	258	265
«ка»	ch(к)	266
«а»	ch(а)	

кодировании сообщений вида $wkwk$, где w – фраза, k – символ. Конкретным примером такого сообщения является любая последовательность трех одинаковых символов, пары которых ранее не встречались.

Пример 7.14. Раскодируем сообщение, закодированное в примере 7.13. Результаты приведены в табл. 7.10.

Таблица 7.10

Пример декодирования LZW

Входной код	Печать	Словарь w	Позиция словаря
	«»	ASCII+	0 – 255
ch(к)	«к»	«кр»	256
ch(р)	«р»	«ра»	257
ch(а)	«а»	«ас»	258
ch(с)	«с»	«сн»	259
ch(н)	«н»	«на»	260
ch(а)	«а»	«ая»	261
ch(я)	«я»	«я »	262
ch( )	« »	« к»	263
256	«кр»	«кра»	264
258	«ас»	«аск»	265
ch(к)	«к»	«ка»	266
ch(а)	«а»		

Алгоритм LZW является запатентованным и, таким образом, представляет собой интеллектуальную собственность.

7.11. СОВРЕМЕННЫЕ ПРОГРАММЫ-АРХИВАТОРЫ

Архиваторы – это программы, использующие алфавитное и словарно-ориентированное кодирование, для уменьшения объема (сжатия) файлов.

Утверждение 7.5. Не существует такого однозначно декодируемого кодирования $\varphi_0 : \{0,1\}^* \to \{0,1\}^*$, что $\forall a \in \{0,1\}^*$ выполнено $|\varphi_0(a)| \leqslant |a|$ и $\exists a_0 \in \{0,1\}^*$, что $|\varphi_0(a_0)| < |a_0|$.

Доказательство. Докажем методом от противного. Предположим, что такое кодирование φ_0 существует. Поскольку оно однозначно декодируемо, то функция $\varphi_0 : \{0,1\}^* \to \{0,1\}^*$ является инъективной, т. е. $\forall a, b \in \{0,1\}^*, a \neq b$, выполнено $\varphi_0(a) \neq \varphi_0(b)$.

Пусть $N = |a_0|$. Рассмотрим все возможные сообщения $a \in \{0,1\}^k$, $1 \leqslant k \leqslant N$. Всего таких сообщений $2 + 4 + \ldots + 2^N = 2^{N+1} - 2$. С другой стороны, посчитаем количество различных закодированных сообщений, которое мы можем получить, кодируя рассматриваемые сообщения. Из определения кодирования φ_0 и равенства $N = |a_0|$ получим, что длины всех закодированных сообщений не будут превосходить N. При этом закодированных сообщений длиной N можно получить строго меньше 2^N, длиной $N - 1$ – не больше 2^{N-1} и т. д., длиной 1 — не больше 2. Таким образом, в сумме возможно получить менее $2^{N+1} - 2$ различных закодированных сообщений, но тогда согласно принципу Дирихле найдутся по крайней мере два таких сообщения $c \in \{0,1\}^{k_1}$, $d \in \{0,1\}^{k_2}$, $c \neq d$, $1 \leqslant k_1, k_2 \leqslant N$, что $\varphi_0(c) = \varphi_0(d)$, а это противоречит инъективности функции φ_0. □

Замечание 7.15. Описанное в условии утверждения 7.5 кодирование, которое уменьшает размер любого сообщения, называется идеальным. И в утверждении 7.5 показано, что идеального кодирования не существует. Однако на практике все равно целесообразно применять различные методы сжатия, так как в большинстве случаев файлы, с которыми работают пользователи (текст, графические изображения), хорошо сжимаются.

Некоторые типичные расширения, соответствующие им программы-архиваторы и методы сжатия данных приведены в табл. 7.11.

Практически все форматы файлов для хранения графической информации используют сжатие данных. В табл. 7.12 приведены типичные расширения графических файлов и поддерживаемые в них методы сжатия данных.

Наибольшую степень сжатия дают двухпроходные алгоритмы, которые последовательно сжимают исходные данные два раза, как правило, один раз применяется словарно-ориентированное кодирование и один раз алфавитное кодирование. Однако такие алгоритмы работают до двух раз медленнее однопроходных при не очень высокой степени увеличения сжатия.

Таблица 7.11

Описание наиболее популярных файлов архивов

Расширение	Программы	Тип кодирования
Z	compress	LZW
arc	ark, pkark	LZW, Хаффмана
zip	zip, pkzip, rar	LZW, LZ77, Хаффмана, Шеннона, Фано
gz	gzip	LZ77, Хаффмана
bz2	bzip2	Бурроуза – Уиллера, Хаффмана
arj	arj	LZ77, Хаффмана
ice, lzh	lha, lharc	LZSS, Хаффмана
pak	pak	LZW
rar	winrar	LZSS, Хаффмана
7-zip	7-zip, winrar	LZMA, LZMA2, PPMD и др.

Таблица 7.12

Обзор наиболее популярных графических файлов

Расширение	Поддерживаемый тип кодирования
gif	LZW
jpeg, jpg	Сжатие с потерями, Хаффмана, арифметическое
bmp, pcx	RLE
tiff, tif	LZW, CCITT/3
png	LZ77, Хаффмана

Программы-архиваторы делятся на сжимающие каждый файл по отдельности и сжимающие все файлы в одном потоке. В последнем случае увеличивается степень сжатия, но одновременно усложняются способы работы с полученным архивом. Например, замена в таком архиве одного файла на его более новую версию может потребовать перекодирования всего архива.

7.12. СЖАТИЕ ДАННЫХ С ПОТЕРЯМИ

Все ранее рассмотренные алгоритмы сжатия информации обеспечивали возможность полного восстановления исходных данных. Но иногда для повышения степени сжатия можно отбрасывать часть исходной информации, т. е. производить сжатие с потерями. В тех случаях, когда сжимается информация, используемая лишь для качественной оценки (это, как правило, аналоговая информация), сжатие с потерями является весьма приемлемым. Оно применяется в основном для трех видов данных: полноцветная графика, звук и видеоинформация.

Сжатие с потерями обычно проходит в два этапа. На первом из них исходная информация приводится (с потерями) к виду, в котором ее можно эффективно сжимать алгоритмами второго этапа сжатия без потерь.

Основная идея сжатия графической информации с потерями заключается в следующем. Каждая точка в картинке характеризуется тремя равнозначными атрибутами: яркостью, цветом и насыщенностью. Но глаз человека не воспринимает эти атрибуты как равные. Он считывает полностью только информацию о яркости и в гораздо меньшей степени — о цвете и насыщенности, что позволяет отбрасывать часть информации о двух последних атрибутах без «видимых» потерь качества изображения.

Для сжатия графических данных с потерями в конце 1980-х гг. установлен один стандарт — формат JPEG (Joint Photographic Experts Group — название объединения его разработчиков). В этом формате можно регулировать степень сжатия, задавать степень потери качества.

Сжатие видеоинформации основано на том, что при переходе от одного кадра к другому на экране обычно практически ничего не меняется. Таким образом сжатая видеоинформация представляет собой запись некоторых базовых кадров и последовательности изменений в них. При этом часть информации может отбрасываться. Сжатую подобным образом информацию можно далее сжимать и другими методами. Хотя существует не один стандарт для сжатия видеоданных, одними из наиболее распространенных являются стандарты MPEG (Motion Picture Experts Group), первый из которых был опубликован в 1988 г.

Для сжатия звуковой информации с потерями существует несколько стандартов. Наиболее широко используемый из них — это MPEG без видеоданных (mp3 — MPEG-1 Layer 3).

7.13. ЗАДАНИЯ ДЛЯ ТЕСТОВ

7.1. Какое кодирование может быть неоднозначно декодируемым:
а) префиксное; б) неравномерное;
в) оптимальное; г) суффиксное;
д) равномерное?

7.2. Какое количество листьев может быть в насыщенном четверичном корневом дереве без особых вершин:
а) 8; б) 9;
в) 10; г) 11;
д) 12?

7.3. Для алфавита A задано распределение вероятностей появления каждого из символов $p(\cdot)$. По этому распределению найдено оптимальное двоичное кодирование φ и вычислена энтропия $\mathbf{H}\{\xi\}$ ИДС без памяти $\langle A, p(\cdot)\rangle$. Какое из соотношений связывает среднюю длину оптимального кода l^{φ} с энтропией $\mathbf{H}\{\xi\}$:
а) $\dfrac{\mathbf{H}\{\xi\}}{\log D} \leqslant l^{\varphi_0} \leqslant 1 + \dfrac{\mathbf{H}\{\xi\}}{\log D}$; б) $l^{\varphi} = D + \mathbf{H}\{\xi\}$;
в) $\dfrac{l^{\varphi}}{\log D} \leqslant \mathbf{H}\{\xi\} \leqslant 1 + \dfrac{l^{\varphi}}{\log D}$; г) $l^{\varphi} < \dfrac{\mathbf{H}\{\xi\}}{D}$;
д) $l^{\varphi} \cdot \mathbf{H}\{\xi\} = 1$?

7.4. Пусть $\varphi : A \to B^*$ – оптимальное D-ичное алфавитное кодирование с длинами кодовых слов $l^{(i)} = |\varphi(a^{(i)})|$, $i \in \{1, \ldots, m\}$, соответствующее распределению $p(\cdot)$. Что следует из того, что $l^{(i)} > l^{(j)}$:

а) $\varphi(a^{(i)}) = \varphi(a^{(j)})$;

б) $p_i > p_j$;

в) такого оптимального кодирования не существует;

г) $l^{(i)} > 2^j$;

д) $p_i \leqslant p_j$?

7.5. Какой из перечисленных алгоритмов на выходе дает оптимальное код:

а) Шеннона; б) Фано;

в) LZW; г) Хаффмана;

д) LZ77?

7.6. Какой из перечисленных алгоритмов возвращает не префиксный код:

а) Хаффмана для $D = 2$; б) LZ78;

в) Фано; г) Хаффмана для $D > 2$;

д) Шеннона?

7.7. В каком из перечисленных форматов данные могут храниться с потерями:

а) gif; б) tiff;

в) png; г) jpeg;

д) bmp?

7.14. РЕШЕНИЕ ТИПОВЫХ ЗАДАЧ

Задача 7.1. Определить, являются ли следующие коды префиксными, суффиксными, однозначно декодируемыми:

а) 010, 0100, 0010;

б) 0, 01, 011, 111;

в) 01, 010, 011, 11, 101.

Решение. а) Данный код не является префиксным, так как кодовое слово 010 — это начало кодового слова 0100. Аналогично кодовое слово 010 — окончание кодового слова 0010, следовательно, код не является суффиксным. Кроме того, код не однозначно декодируемый, для этого рассмотрим кодовую последовательность 0100010. Ее можно получить двумя способами: либо как конкатинация кодовых слов 010 и 0010, либо как конкатинация кодовых слов 0100 и 010.

б) Непосредственной проверкой убеждаемся, что данный код — суффиксный и, следовательно, однозначно декодируемый. При этом код не префиксный, поскольку кодовое слово 0 — начало кодового слова 01.

в) Рассмотрим еще один основанный на построении специального графа способ проверки, является ли заданный код однозначно декодируемым [19].

Определим все последовательности (строки), которые: а) совпадают с началом какого-то кодового слова и одновременно с концом какого-то кодового слова; б) сами не являются кодовыми словами.

В данной задаче это три последовательности: 0 (начало кодового слова 01 и конец кодового слова 010), 1 (начало кодового слова 11 и конец кодового слова 101), 10 (начало кодового слова 101 и конец кодового слова 010); последовательности 01 и 11 не учитываются, потому что они представляют собой кодовые слова. Добавим к этому множеству 0, 1, 10 пустую строку, которую обозначим $\emptyset$. Элементы полученного множества $\emptyset$, 0, 1, 10 будут вершинами графа.

Соединим вершины дугами (направленными ребрами) по следующему правилу: две вершины X и Y соединяются дугой, если последовательная запись кода вершины X, некоторого кодового слова (быть может, пустого) и кода вершины Y дает некоторое новое кодовое слово. Для рассматриваемой задачи дерево будет иметь вид, приведенный на рис. 7.9.

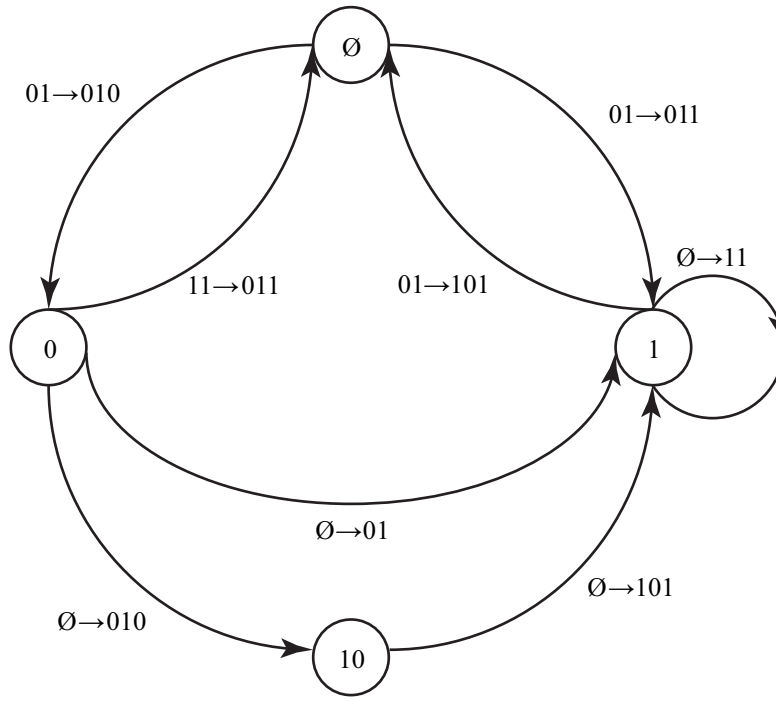


Рис. 7.9. Пример построения специального дерева

На рис. 7.9 надпись «$A \to B$» рядом с дугой, соединяющей вершины X и Y, означает, что если между словами X и Y вписать кодовое слово A, то получим кодовое слово B.

Известно, что код является однозначно декодируемым тогда и только тогда, когда в построенном таким образом графе нет ориентированных циклов, включающих вершину $\emptyset$. В данном случае имеется, например, цикл $\emptyset \to 0 \to \emptyset$. Этому

циклу соответствует сообщение 01011, которое может быть расшифровано как 01 и 011, или 010 и 11.

Ответ:

а) код не является ни префиксным, ни суффиксным, ни однозначно декодируемым;

б) код не является префиксным, однако является суффиксным и однозначно декодируемым;

в) код не является ни префиксным, ни суффиксным, ни однозначно декодируемым.

Задача 7.2. Построить двоичный префиксный код с заданной последовательностью длин кодовых слов 1, 2, 3.

Решение. Для начала проверим, выполнено ли неравенство Крафта: $2^{-1} + 2^{-2} + 2^{-3} = 7/8 \leqslant 1$. Поскольку неравенство Крафта выполнено, то это означает, что префиксный код с заданными длинами кодовых слов существует.

Для нахождения искомого кода построим кодовое дерево от корня к листьям сверху вниз (рис. 7.10).

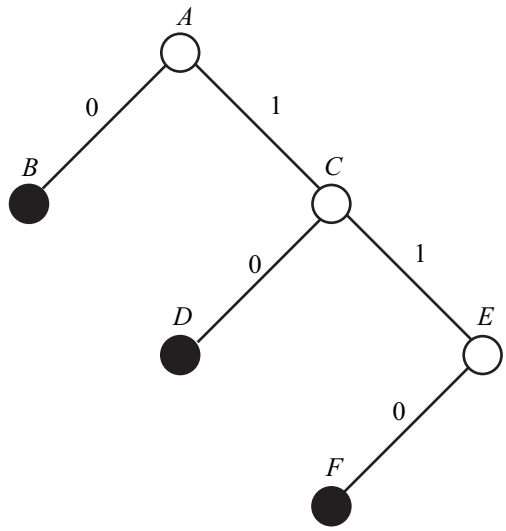


Рис. 7.10. Пример построения кодового дерева

Пусть изначально у нас есть только корень (на рис. 7.10 это вершина A). Добавим к графу две вершины и соединим их ребрами с корнем. Далее на первом уровне получившегося дерева отметим столько вершин, сколько кодовых слов длиной 1 мы должны получить. В данной задаче — одно, поэтому отметим одну вершину B (на рис. 7.10 отмеченные вершины закрашены черным). К каждой из неотмеченных вершин добавим по две дочерние вершины и соединим их ребрами с родительскими. В рассматриваемом примере добавим две дочерние вершины к вершине C. На получившемся втором уровне отметим столько вершин, сколько кодовых слов длиной 2 мы должны получить. В решаемой задаче — одно, поэтому отметим одну вершину D. Затем к каждой из неотмеченных вер-

шин снова добавим по две дочерние вершины и повторим эту процедуру до тех пор, пока не построим уровень, равный длине самого длинного кодового слоя. В рассматриваемой задаче это уровень 3. На этом уровне отметим необходимое количество вершин (на рис. 7.10 одну — F). После чего в обратном порядке снизу вверх удалим лишние неотмеченные висячие вершины. В получившемся дереве расставим метки над ребрами следующим образом. Ребра, выходящие из одной вершины, отметим 0 и 1. В итоге получим искомое кодовое дерево.

Для того чтобы построить D-ичный код, на каждом шаге необходимо добавлять ко всем неотмеченным вершинам по D дочерних вершин, а при расстановке меток ребрам, исходящим из одной вершины, ставить в соответствие метки от 0 до $D-1$.

В таком кодовом дереве каждая отмеченная вершина соответствует одному кодовому слову, а само кодовое слова может быть получено как последовательность меток ребер маршрута от корня к отмеченной вершине. Поскольку маршрут от корня к вершине B есть ребро AB, то кодовое слово для вершины B — это 0. Маршрут к отмеченной вершине D включает в себя два ребра AC и CD, поэтому код этой вершины — 10. И, наконец, маршрут к отмеченной вершине F — $AC \to CE \to EF$, откуда код вершины F равен 110.

Искомый код представляет собой кодовые слова всех отмеченных в кодовом дереве вершин. В решаемой задаче это $\{0, 10, 110\}$. Легко убедиться, что кодовые слова имеют длины 1, 2 и 3 соответственно, что и требовалось в условии задачи.

Ответ: $\{0, 10, 110\}$.

Задача 7.3. Может ли набор чисел 1, 1, 2, 2, 3, 3 быть набором длин кодовых слов однозначно декодируемого 3-ичного кода?

Решение. Проверим необходимое условие однозначно декодируемого D-ичного кода из неравенства Мак-Миллана:

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

Если оно выполнено, тогда согласно неравенству Крафта существует префиксный код, имеющий данные длины кодовых слов, он же является и однозначно декодируемым. Иначе заданный набор чисел не может быть набором длин кодовых слов однозначно декодируемого D-ичного кода. Проверим

$$2 \cdot 3^{-1} + 2 \cdot 3^{-2} + 2 \cdot 3^{-3} = \frac{2}{3} + \frac{2}{9} + \frac{2}{27} = \frac{26}{27} < 1.$$

Необходимое условие выполнено.

Ответ: может.

Задача 7.4. Построить двоичные коды по алгоритму Фано для распределения $p = \{0{,}36;\ 0{,}18;\ 0{,}18;\ 0{,}12;\ 0{,}09;\ 0{,}07\}$ и вычислить среднюю длину кодового слова.

Решение. Для удобства дальнейшего описания алгоритма Фано предположим, что распределение p задано на алфавите $A = \{a, b, c, d, e, f\}$.

Основная идея алгоритма Фано заключается в последовательном разбиении множества символов алфавита на подмножества. Для наглядности такое разбиение будем представлять в виде бинарного дерева (рис. 7.11). Каждой вершине такого дерева будет соответствовать некоторое множество символов и вероятность появления символа из этого множества. Изначально имеется только корень, которому соответствует множество всех символов алфавита и вероятность $p = 1$.

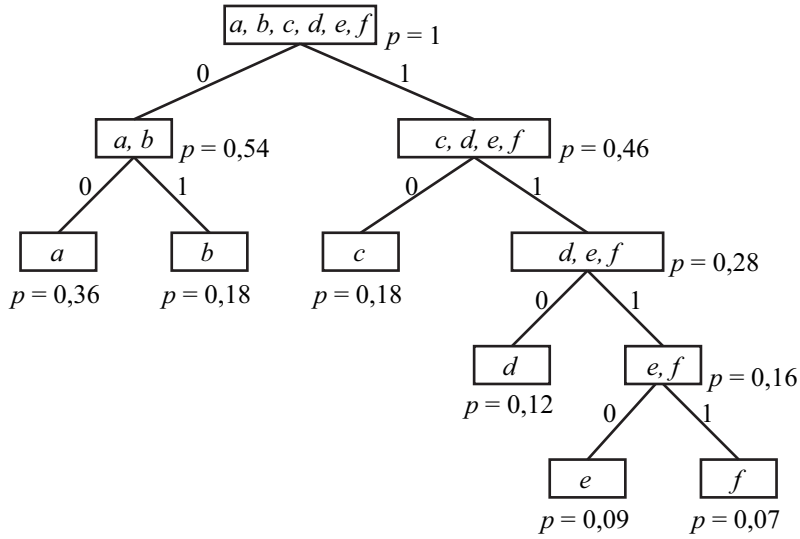


Рис. 7.11. Пример реализации алгоритма Фано

0) Упорядочим элементы алфавита по невозрастанию их вероятностей. В данном случае они так упорядочены изначально.

1) Выберем наименьшее число k, $1 \leqslant k < m$, так, чтобы величина

$$\left| \sum_{i=1}^{k} p_i - \sum_{i=k+1}^{m} p_i \right| \tag{7.28}$$

была наименьшей. При $k = 1$ данная величина равна 0,28, при $k = 2$ – 0,08, при $k = 3$ – 0,44. При большем значении k эта величина будет еще больше. Разобьем множество $\{0{,}36;\ 0{,}18;\ 0{,}18;\ 0{,}12;\ 0{,}09;\ 0{,}07\}$ на два подмножества $A_0 = \{a, b\}$ и $A_1 = \{c, d, e, f\}$, которым поставим в соответствие две новые вершины графа, соединенные ребрами с корнем и имеющие метки 0 и 1 (в первое подмножество попадает k символов, во второе $m - k$).

2) Рассмотрим подмножество $A_0 = \{a, b\}$. Как и на предыдущем шаге, найдем такое k, что величина (7.28) будет минимальной. В данном случае единственное возможное значение $k = 1$. Снова разобьем множество A_0 на два подмножества $A_{00} = \{a\}$ и $A_{01} = \{b\}$ и поставим им в соответствие вершины графа.

3) Рассмотрим подмножество $A_1 = \{c, d, e, f\}$. При $k = 1$ величина (7.28) будет минимальной. Поэтому разобьем множество A_1 на два подмножества

$A_{10} = \{c\}$ и $A_{11} = \{d,e,f\}$ и поставим им в соответствие вершины графа.

4) Продолжим до тех пор, пока каждое подмножество не будет состоять из одного элемента.

Код каждого символа можно определить двумя способами. Во-первых, если символ x принадлежит одноэлементному подмножеству $A_{i_1,\dots,i_k}$, то кодовое слова $\varphi(x) = i_1 \dots i_k$. Во-вторых, построенное бинарное дерево является кодовым, где каждый лист соответствует одному символу.

Бинарное дерево, построенное для данной задачи, изображено на рис. 7.11.

Вычислим среднюю длину построенного двоичного кода. Для этого для каждого символа необходимо найти длину кодового слова, умножить ее на вероятность появления данного символа, после чего сложить все полученные произведения:

$$l^{\varphi} = 0{,}36 \cdot 2 + 0{,}18 \cdot 2 + 0{,}18 \cdot 2 + 0{,}12 \cdot 3 + 0{,}09 \cdot 4 + 0{,}07 \cdot 4 = 2{,}44.$$

Ответ: код {00, 01, 10, 110, 1110, 1111} и его средняя длина 2,44.

Задача 7.5. Построить двоичные коды по алгоритму Шеннона для распределения $p = \{0{,}36;\ 0{,}18;\ 0{,}18;\ 0{,}12;\ 0{,}09;\ 0{,}07\}$ и вычислить среднюю длину кодового слова.

Решение. Для удобства дальнейшего описания алгоритма Шеннона предположим, что распределение p задано на алфавите $A = \{a, b, c, d, e, f\}$.

Пример выполнения алгоритма Шеннона приведен в табл. 7.13.

Таблица 7.13

Шаги алгоритма Шеннона

	0	1	2	3	4
A	p_i	$l^{(i)}$	P_i	$\varphi_1()$	$\varphi_2()$
a	0,36	2	0,00	00	00
b	0,18	3	0,36 = 0,010	010	01
c	0,18	3	0,54 = 0,100	100	100
d	0,12	4	0,72 = 0,1011	1011	101
e	0,09	4	0,84 = 0,1101	1101	110
f	0,07	4	0,93 = 0,1110	1110	111

0) Упорядочим элементы алфавита по невозрастанию их вероятностей. В данной задаче они так упорядочены изначально.

1) Найдем натуральные числа $l^{(i)}$, $i \in \{1, \dots, m\}$, такие, что выполнены неравенства

$$\frac{1}{2^{l^{(i)}}} \leqslant p_i < \frac{1}{2^{l^{(i)}-1}}.$$

Так, для $p_1 = 0{,}36$ имеет место следующее неравенство:

$$\frac{1}{2^2} = 0{,}25 \leqslant 0{,}36 \leqslant 0{,}5 = \frac{1}{2^{2-1}},$$

откуда $l^{(1)} = 2$. Аналогично найдем все остальные $l^{(2)}, \ldots, l^{(6)}$. Они приведены в табл. 7.13 в столбце под номером 1.

2) Вычислим накопленные вероятности:

$$P_1 = 0;\ P_2 = P_1 + p_1 = 0{,}36;\ P_3 = 0{,}36 + 0{,}18 = 0{,}54;$$
$$P_4 = 0{,}54 + 0{,}18 = 0{,}72;\ P_5 = 0{,}72 + 0{,}12 = 0{,}84;$$
$$P_6 = 0{,}84 + 0{,}09 = 0{,}93.$$

3) Для каждого $i \in \{1, \ldots, m\}$ определим код $\varphi(a^{(i)})$ как $l^{(i)}$ символов, стоящих после запятой в двоичном представлении числа P_i. Поскольку $P_1 = 0 = 0_2$, то $\varphi(a) = 00$; $P_2 = 0{,}36 = 0{,}0101_2$, поэтому $\varphi(a) = 010$ и т. д. Перевод в двоичную систему счисления и получившиеся кодовые слова представлены в табл. 7.13 в столбцах под номерами 2 и 3.

4) Операция усечения. Для всех i, пробегающих все значения от 1 до m, в кодовом слове $\varphi(a^{(i)})$ справа налево последовательно удаляются символы, пока код $\varphi(a^{(i)})$ не является префиксом ни для какого другого кодового слова $\varphi(a^{(j)})$, $i \neq j$. Из кода 00 нельзя удалить правый нуль, поскольку в этом случае оставшийся нуль станет префиксом кода 010. Из кода 010 удалим правый 0, поскольку ни один из оставшихся кодов не имеет префикса 01, а вот единицу уже удалить нельзя, так как оставшийся нуль станет префиксом кода 00. Проделав данную операцию со всеми кодовыми словами, получим искомый код. Результаты операции усечения представлены в табл. 7.13 в последнем столбце.

Вычислим среднюю длину кодового слова:

$$l^{\varphi} = 0{,}36 \cdot 2 + 0{,}18 \cdot 2 + 0{,}18 \cdot 3 + 0{,}12 \cdot 3 + 0{,}09 \cdot 3 + 0{,}07 \cdot 3 = 2{,}46.$$

Ответ: код {00, 01, 100, 101, 110, 111} и его средняя длина 2,46.

Задача 7.6. Построить двоичные коды по алгоритму Хаффмана для распределения $p = \{0{,}36;\ 0{,}18;\ 0{,}18;\ 0{,}12;\ 0{,}09;\ 0{,}07\}$ и вычислить среднюю длину кодового слова.

Решение. Для удобства дальнейшего описания алгоритма Хаффмана для построения двоичного кода предположим, что распределение p задано на алфавите $A = \{a, b, c, d, e, f\}$.

1-й этап — построение двоичного дерева. Построим двоичное дерево с $m = 6$ листьями, начиная с листьев и продвигаясь к корню. Возьмем в качестве листьев дерева символы алфавита $a, b, \ldots, f$, которым приписаны вероятности $p_1 = 0{,}36, \ldots, p_6 = 0{,}07$. Дерево, которое мы получим в результате выполнения 1-го этапа алгоритма Хаффмана, изображено на рис. 7.12.

0) Отсортируем множество вершин в порядке невозрастания вероятностей. В нашем случае они отсортированы так изначально.

1) Выберем две вершины e и f с наименьшими вероятностями $p_5 = 0{,}09$ и $p_6 = 0{,}07$ и добавим в дерево новую вершину $\{e, f\}$, которой припишем вероятность $0{,}07 + 0{,}09 = 0{,}16$. Вершину $\{e, f\}$ соединим ребрами с вершинами e и f и объявим общим предком для этих вершин. Ребро от $\{e, f\}$ к e (вершине, чья вероятность была больше) пометим символом 0, а ребро от $\{e, f\}$ к f (вершине,

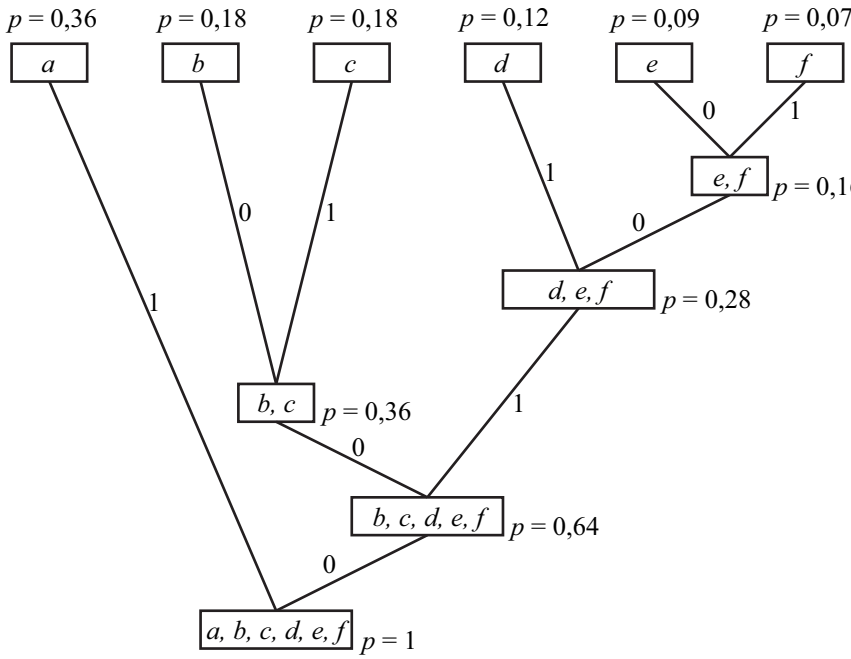


Рис. 7.12. Пример реализации алгоритма Хаффмана для $D = 2$

чья вероятность была меньше) — символом 1. Получим новый набор рассматриваемых вершин $A^1 = \{a, b, c, d, \{e,f\}\}$ с соответствующим набором вероятностей $p = \{0{,}36;\ 0{,}18;\ 0{,}18;\ 0{,}12;\ 0{,}16\}$.

2) Отсортируем множество вершин A^1 в порядке невозрастания вероятностей, получим $A^1 = \{a, b, c, \{e, f\}, d\}$ с соответствующим набором вероятностей $p = \{0{,}36;\ 0{,}18;\ 0{,}18;\ 0{,}16;\ 0{,}12\}$.

3) Выберем две вершины $\{e, f\}$ и d с наименьшими вероятностями $p_4 = 0{,}16$ и $p_5 = 0{,}12$ и добавим в дерево новую вершину $\{d, e, f\}$, которой припишем вероятность $0{,}16 + 0{,}12 = 0{,}28$. Вершину $\{d, e, f\}$ соединим ребрами с вершинами $\{e, f\}$ и d и объявим общим предком для этих вершин. Ребро от $\{d, e, f\}$ к $\{e, f\}$ (вершине, чья вероятность была больше) пометим символом 0, а ребро от $\{d, e, f\}$ к d (вершине, чья вероятность была меньше) — символом 1. Получим новый набор рассматриваемых вершин $A^2 = \{a, b, c, \{d, e, f\}\}$ с соответствующим набором вероятностей $p = \{0{,}36;\ 0{,}18;\ 0{,}18;\ 0{,}28\}$.

И далее повторим описанные шаги до тех пор, пока набор рассматриваемых вершин не станет состоять из одной вершины $\{a, b, c, d, e, f\}$, которая станет корнем искомого дерева.

2-й этап — кодирование.

Чтобы получить кодовое слово $\varphi(x)$ для символа x, последовательно считываем метки ребер на пути от корня дерева к листу x. Например, путь от корня к листу e состоит из ребер $(\{a, b, c, d, e, f\},\ \{b, c, d, e, f\})$, $(\{b, c, d, e, f\},\ \{d, e, f\})$, $(\{d, e, f\},\ \{e, f\})$, $(\{e, f\},\ \{e\})$, имеющих метки 0, 1, 0, 0 соответственно, откуда

$\varphi(e) = 0100$. Коды остальных символов находятся аналогично. В результате получится кодирование $\varphi(a) = 1$, $\varphi(b) = 000$, $\varphi(c) = 001$, $\varphi(d) = 011$, $\varphi(e) = 0100$, $\varphi(f) = 0101$.

Вычислим среднюю длину кодового слова:

$$l^{\varphi} = 0{,}36 \cdot 1 + 0{,}18 \cdot 3 + 0{,}18 \cdot 3 + 0{,}12 \cdot 3 + 0{,}09 \cdot 4 + 0{,}07 \cdot 4 = 2{,}44.$$

Замечание 7.16. В описанном алгоритме имеется неоднозначность в выборе двух вершин с наименьшими вероятностями, если имеется более двух вершин с одинаковыми наименьшими вероятностями. В случае такой неоднозначности выбираются две вершины с наибольшими номерами в списке вершин (вершина, которая является объединением нескольких вершин, имеет номер, равный сумме номеров вершин, входящих в ее состав).

Ответ: код {1, 000, 001, 011, 0100, 0101} и его средняя длина 2,44.

Задача 7.7. Построить троичные коды по алгоритму Хаффмана для распределения $p = \{0{,}2;\ 0{,}14;\ 0{,}13;\ 0{,}13;\ 0{,}12;\ 0{,}12;\ 0{,}08;\ 0{,}08\}$ и вычислить среднюю длину кодового слова.

Решение. Для удобства дальнейшего описания алгоритма Хаффмана для построения двоичного кода предположим, что распределение p задано на алфавите $A = \{a, b, c, d, e, f, g, h\}$. Сформулируем алгоритм Хаффмана для случая $D \leqslant 2$ в терминах кодовых деревьев.

1-й этап — построение D-ичного дерева. Построим троичное (в общем случае D-ичное) дерево с $m = 8$ листьями, начиная с листьев и продвигаясь к корню. Возьмем в качестве листьев дерева символы алфавита $a, b, \ldots, h$, которым приписаны вероятности $p_1 = 0{,}2, \ldots, p_8 = 0{,}08$. Дерево, которое мы получим в результате выполнения 1-го этапа алгоритма Хаффмана, изображено на рис. 7.13.

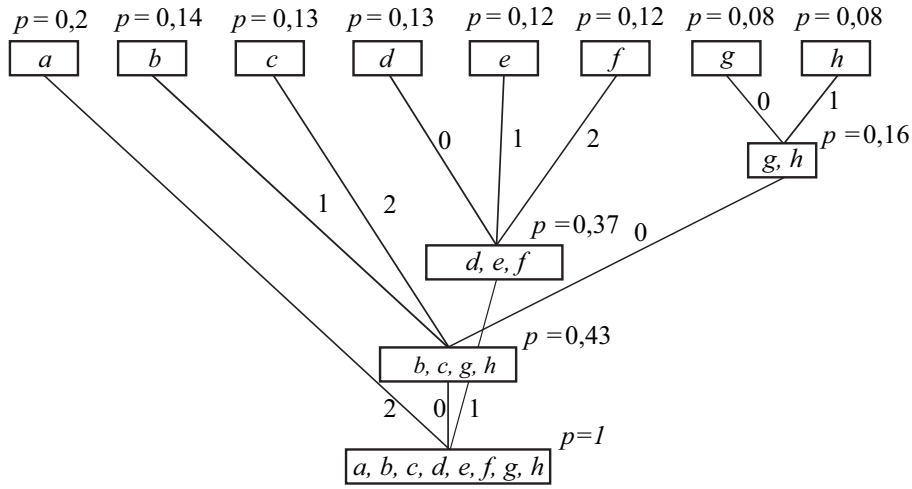


Рис. 7.13. Пример реализации алгоритма Хаффмана для $D > 2$

0) Отсортируем множество вершин в порядке невозрастания вероятностей. В данном примере они отсортированы так изначально.

1) Вычислим m_0 — такое наименьшее натуральное число вида $m_0 = 1+$ $+k(D-1)$, что $m_0 \geqslant m = 8$. По условию $D = 3$, откуда $m_0 = 9$. Вычислим

$$S = \begin{cases} D, & m = m_0, \\ m - m_0 + D, & m < m_0. \end{cases} \tag{7.29}$$

Получим $S = m - m_0 + D = 8 - 9 + 3 = 2$.

2) Выберем $S = 2$ вершин g и h с наименьшими вероятностями $p_7 = 0{,}08$ и $p_6 = 0{,}08$ и добавим в дерево новую вершину $\{g, h\}$, которой припишем вероятность $0{,}08 + 0{,}08 = 0{,}16$. Вершину $\{g, h\}$ соединим ребрами с вершинами g и h и объявим общим предком для этих вершин. Ребро от $\{g, h\}$ к g (вершине, которая расположена левее) пометим символом 0, ребро от $\{g, h\}$ к h (вершине, которая расположена правее) — символом 1. Получим новый набор рассматриваемых вершин $A^1 = \{a, b, c, d, e, f, \{g, h\}\}$ с соответствующим набором вероятностей $p = \{0{,}2;\ 0{,}14;\ 0{,}13;\ 0{,}13;\ 0{,}12;\ 0{,}12;\ 0{,}16\}$.

3) Отсортируем множество вершин A^1 в порядке невозрастания вероятностей, получим $A^1 = \{a, \{g, h\}, b, c, d, e, f\}$ с соответствующим набором вероятностей $p = \{0{,}2;\ 0{,}16;\ 0{,}14;\ 0{,}13;\ 0{,}13;\ 0{,}12;\ 0{,}12\}$.

4) Выберем $D = 3$ вершины d, e и f с наименьшими вероятностями 0,13, 0,12 и 0,12 соответственно. Заметим, что на этом и последующих шагах объединяется D, а не S вершин. Добавим в дерево новую вершину $\{d, e, f\}$, которой припишем вероятность $0{,}13 + 0{,}12 + 0{,}12 = 0{,}37$. Вершину $\{d, e, f\}$ соединим ребрами с вершинами d, e и f и объявим общим предком для этих вершин. Ребро от $\{d, e, f\}$ к d (самой левой из объединяемых вершине) пометим символом 0, ребро от $\{d, e, f\}$ к e (средней вершине) – символом 1, а ребро от $\{d, e, f\}$ к f (самой правой из объединяемых вершине) — символом 2. Получим новый набор рассматриваемых вершин $A^2 = \{a, \{g, h\}, b, c, \{d, e, f\}\}$ с соответствующим набором вероятностей $p = \{0{,}2;\ 0{,}16;\ 0{,}14;\ 0{,}13;\ 0{,}37\}$.

Повторим описанные шаги до тех пор, пока набор рассматриваемых вершин не станет состоять из одной вершины $\{a, b, c, d, e, f, g, h\}$ — корня искомого дерева.

2-й этап — кодирование.

Чтобы получить кодовое слово $\varphi(x)$ для символа x, последовательно считываем метки ребер на пути от корня дерева к листу x. Например, путь от корня к листу g состоит из ребер $(\{a, b, c, d, e, f, g, h\},\ \{b, c, g, h\})$, $(\{b, c, g, h\},\ \{g, h\})$, $(\{g, h\},\ g)$, имеющих метки 0, 0, 0 соответственно. Отсюда $\varphi(g) = 000$. Коды остальных символов находятся аналогично. В результате получится кодирование $\varphi(a) = 2$, $\varphi(b) = 01$, $\varphi(c) = 02$, $\varphi(d) = 10$, $\varphi(e) = 11$, $\varphi(f) = 12$, $\varphi(g) = 000$, $\varphi(h) = 001$.

Вычислим среднюю длину кодового слова:

$$l^{\varphi} = 0{,}2 \cdot 1 + 0{,}14 \cdot 2 + 0{,}13 \cdot 2 + 0{,}13 \cdot 2 + 0{,}12 \cdot 2 + \\ + 0{,}12 \cdot 2 + 0{,}08 \cdot 3 + 0{,}08 \cdot 3 = 1{,}96.$$

Замечание 7.17. В описанном алгоритме имеется неоднозначность в выборе нескольких вершин с наименьшими вероятностями, если есть более двух вершин с одинаковыми наименьшими вероятностями. В случае такой неоднозначности выбираются вершины с наибольшими номерами в списке вершин (вершина, которая является объединением нескольких вершин, имеет номер, равный сумме номеров вершин, входящих в ее состав).

Ответ: код {2, 01, 02, 10, 11, 12, 000, 001} и его средняя длина 1,96.

7.15. ЗАДАЧИ И УПРАЖНЕНИЯ

7.1. Определить, являются ли следующие коды префиксными, суффиксными, однозначно декодируемыми:

а) 00, 01, 10, 11;
б) 0, 01, 001, 0010, 0011;
в) 110, 11, 100, 00, 10;
г) 100, 001, 101, 1101, 11011;
д) 010, 0001, 0110, 1100, 00011, 00110, 11110, 10101;
е) 00, 012, 0110, 0112, 100, 201, 212, 22;
ж) 123, 1234, 5, 421, 2135, 3513, 3512, 5124?

7.2. Построить двоичный код из четырех кодовых слов, не префиксный, не суффиксный, но однозначно декодируемый.

7.3. Известно, что некоторый двоичный код состоит из более чем 2^n кодовых слов, длина каждого из которых не превосходит n. Может ли такой код быть однозначно декодируемым; префиксным?

7.4. Построить двоичный префиксный код с заданной последовательностью длин кодовых слов:

а) 1, 2, 3, 3;
б) 1, 2, 4, 4, 4;
в) 2, 2, 3, 4, 4;
г) 2, 2, 2, 4, 4, 4;
д) 2, 3, 3, 3, 4, 4;
е) 1, 2, 3, 3, 4, 4, 4, 4;
ж) 1, 3, 3, 4, 4, 4, 4;
з) 2, 2, 3, 3, 4, 4, 4, 4;
и) 2, 3, 4, 4, 4.

7.5. Для заданного D построить D-ичный префиксный код с заданной последовательностью длин кодовых слов:

а) $D = 3$, 1, 1, 2, 2, 2, 3;
б) $D = 3$, 1, 1, 2, 3, 3, 3, 3;
в) $D = 4$, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3;
г) $D = 3$, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 4;
д) $D = 4$, 1, 1, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3;
е) $D = 3$, 1, 2, 2, 2, 3, 3, 4, 4.

7.6. Может ли заданный набор чисел быть набором длин кодовых слов однозначно декодируемого D-ичного кода:
а) 1, 2, 2, 2, 3, 3, 3, 3 ($D = 3$);
б) 1, 1, 2, 2, 3, 3, 3 ($D = 3$);
в) 1, 1, 1, 2, 2, 2, 2, 3 ($D = 4$);
г) 1, 2, 2, 2, 2 ($D = 3$);
д) 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3 ($D = 3$);
е) 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3 ($D = 4$)?

7.7. При каком наименьшем D заданный набор чисел может быть набором длин кодовых слов однозначно декодируемого D-ичного кода:
а) 1, 2, 2, 3; б) 2, 2, 2, 4, 4, 4;
в) 1, 1, 2, 2, 3, 3; г) 1, 1, 2, 2, 3, 3, 4, 4, 4, 4?

7.8. При $D = 4$ построить или доказать невозможность построения однозначно декодируемого D-ичного кода, в котором длины кодовых слов не превосходят 6, причем количество кодовых слов длины i равно w_i, $1 \leqslant i \leqslant 6$, если набор $(w_1, \ldots, w_6)$ равен:
а) (3, 3, 3, 3, 4, 0); б) (2, 7, 3, 3, 5, 0);
в) (1, 7, 3, 7,4, 0); г) (0,7, 3,11, 3,4);
д) (1, 10, 5, 11, 3, 5).

7.9. Доказать, что максимальная длина кодового слова в оптимальном коде для m-буквенного алфавита A не превосходит $m - 1$. Показать, что эта верхняя граница достижима.

7.10. Доказать, что сумма длин кодовых слов оптимального двоичного кода для m-буквенного алфавита не превосходит $(m + 2)(m - 1)/2$.

7.11. Построить двоичные коды по алгоритмам Фано, Шеннона и Хаффмана для указанных ниже распределений и вычислить средние длины кодовых слов:
а) {0,4; 0,2; 0,2; 0,2};
б) {0,7; 0,1; 0,1; 0,1};
в) {27/40; 9/40; 3/40; 1/40};
г) {1/3; 1/4; 1/5; 1/6; 1/20};
д) {1/2; 1/4; 1/8; 1/16; 1/16};
е) {0,2; 0,2; 0,2; 0,2; 0,2};
ж) {28/72; 15/72; 12/72; 11/72; 6/72};
з) {0,5; 0,2; 0,1; 0,09; 0,08; 0,03};
и) {0,4; 0,2; 0,1; 0,1; 0,1; 0,1};
к) {0,4; 0,3; 0,1; 0,07; 0,06; 0,04; 0,03};
л) {0,4; 0,3; 0,08; 0,06; 0,04; 0,04; 0,04; 0,04};
м) {0,32; 0,24; 0,20; 0,09; 0,05; 0,04; 0,04; 0,02}.

7.12. Построить коды по алгоритму Хаффмана для указанных ниже распределений при $D \in \{2, 3, 4\}$, вычислить средние длины кодовых слов:
а) {1/2; 1/4; 1/8; 1/16; 1/16};

б) {0,3; 0,2; 0,2; 0,2; 0,1};
в) {3/8; 1/6; 1/8; 1/8; 1/8; 1/12};
г) {1/21; 2/21; 3/21; 4/21; 5/21; 6/21};
д) {0,2; 0,15; 0,05; 0,2; 0,25; 0,15};
е) {0,4; 0,2; 0,1; 0,1; 0,1; 0,1};
ж) {0,3; 0,3; 0,1; 0,1; 0,1; 0,1};
з) {0,3; 0,2; 0,15; 0,15; 0,1; 0,1};
и) {0,24; 0,24; 0,16; 0,16; 0,12; 0,08};
к) {0,3; 0,2; 0,1; 0,1; 0,1; 0,1; 0,1};
л) {0,2; 0,12; 0,08; 0,15; 0,25; 0,1; 0,1};
м) {0,3; 0,25; 0,15; 0,1; 0,1; 0,05; 0,05};
н) {0,49; 0,26; 0,12; 0,04; 0,04; 0,03; 0,02};
о) {0,25; 0,05; 0,1; 0,13; 0,2; 0,12; 0,08; 0,07};
п) {0,21; 0,2; 0,17; 0,16; 0,12; 0,08; 0,04; 0,02};
р) {0,2; 0,15; 0,15; 0,13; 0,12; 0,11; 0,11; 0,03};
с) {0,3; 0,2; 0,15; 0,1; 0,1; 0,08; 0,05; 0,02}.

7.13. Для распределения {1/3; 1/3; 1/4; 1/12} построить два различных двоичных кода с наборами длин кодовых слов (1, 2, 3, 3) и (2, 2, 2, 2). Являются ли построенные коды оптимальными?

7.14. Для распределения {0,3; 0,2; 0,2; 0,1; 0,1; 0,05; 0,05} построить три двоичных кода с наборами длин кодовых слов (2, 2, 2, 3, 4, 5, 5), (2, 2, 3, 3, 3, 4, 4) и (2, 2, 2, 4, 4, 4, 4). Являются ли построенные коды оптимальными?

7.15. Для распределения {0,3; 0,2; 0,15; 0,15; 0,9; 0,06; 0,05} построить два двоичных кода с наборами длин кодовых слов (2, 2, 3, 3, 3, 4, 4) и (1, 2, 4, 4, 4, 5, 5). Являются ли построенные коды оптимальными?

7.16. Для равномерного распределения на алфавите из m символов найти длины кодовых слов двоичного кода Хаффмана и сравнить среднюю длину этого кода с энтропией, если $m = 100$ и в общем случае.

7.17. Построить двоичный код Хаффмана для распределения {1/3; 1/5; 1/5, 2/15; 2/15}. Доказать, что этот код является оптимальным и для равномерного распределения {1/5; 1/5; 1/5; 1/5; 1/5}.

7.18. Пусть дано распределение {0,3; 0,3; 0,2; 0,1; 0,1} и пусть l_i — длины кодовых слов двоичного кода Хаффмана, соответствующего этому распределению, $i = 1, \ldots, 5$. Найти такое распределение вероятностей $q = (q_1, q_2, q_3, q_4, q_5)$ дискретной случайной величины ξ, что $\sum_{i=1}^{5} q_i l_i = \mathbf{H}\{\xi\}$.

7.19. Для распределения $(1/2, 1/4, \ldots, 1/2^{k-2}, 1/2^{k-1}, 1/2^{k-1})$ на множестве из k символов построить оптимальное двоичное кодирование $\varphi_k()$, вычислить его среднюю длину l^{φ_k} и предел $\lim_{k\to\infty} l^{\varphi_k}$.

7.20. Пусть $(l_1, \ldots, l_m)$ — длины кодовых слов двоичного кода Хаффмана, соответствующего распределению $(p_1, \ldots, p_m)$, $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_m$. Доказать, что: а) если $p_1 > 2/5$, то $l_1 = 1$; б) если $p_1 > 1/3$, то $l_1 \leqslant 2$.

7.21. Доказать, что если объем алфавита m не является степенью двойки, то в оптимальном двоичном коде найдутся два слова, имеющие разную длину.

7.22. Построить пример такого оптимального двоичного префиксного кодирования φ для алфавита $A = \{a^{(1)}, \ldots, a^{(m)}, a^{(m+1)}, \ldots, a^{(m+k)}\}$ с распределением $p = (p_1, \ldots, p_m, 0, \ldots, 0)$, что ограничение φ_B кодирования φ на алфавит $B = \{a^{(1)}, \ldots, a^{(m)}\}$ с распределением $(p_1, \ldots, p_m)$ не является оптимальным.

7.23. Построить коды по алгоритму LZ77 с параметрами $K = 8$, $L = 5$ и вычислить длину получившегося кода для следующих текстов, состоящих из латинских букв:

а) abracadabra; б) abbaaababbabbba;
в) enikibeniki; г) messmessage.

7.24. Построить коды по алгоритму LZ78 с размером словаря, равным 16 фраз, и вычислить длину получившегося кода для следующих текстов, состоящих из латинских букв:

а) abracadabra; б) abbaaababbabbba;
в) enikibeniki; г) messmessage.

7.25. Построить коды по алгоритму LZW с размером словаря, равным 512 фраз, и вычислить длину получившегося кода для следующих текстов, состоящих из латинских букв:

а) abracadabra; б) abbaaababbabbba;
в) enikibeniki; г) messmessage.

Глава 8

ЛИНЕЙНЫЕ КОДЫ, ИСПРАВЛЯЮЩИЕ ОШИБКИ

8.1. ОБНАРУЖЕНИЕ И ИСПРАВЛЕНИЕ ОШИБОК

В идеальной системе связи (при отсутствии искажений в канале) символы, которые появляются на выходе устройства, декодирующего сигналы, совпадают с символами, поступающими на вход кодирующего устройства. Однако в реальном канале связи всегда имеются помехи, поэтому целесообразно использование так называемых *корректирующих кодов*, назначение которых состоит в том, чтобы обнаруживать и (или) исправлять ошибки. Для этого перед отправкой сообщения получателю к этому сообщению добавляется некоторая избыточная информация, например контрольная сумма.

Приведем требования, которые предъявляются к преобразованиям кодирования и декодирования:
1) высокая скорость передачи информации;
2) малая вероятность ошибочного декодирования;
3) «приемлемая» трудоемкость алгоритмов кодирования и декодирования;
4) «приемлемая» затрачиваемая память.

Определение 8.1. *Пусть натуральные числа k и n таковы, что $D^k \leqslant q^n$. Блоковым кодированием с входным алфавитом $\mathcal{B} = \{b_1, \ldots, b_D\}$, выходным алфавитом $\mathcal{X} = \{x_1, \ldots, x_q\}$, блоком информационных символов длиной k и блоком кодовых символов длиной n называется произвольное инъективное отображение $f : \mathcal{B}^k \to \mathcal{X}^n$. Множество $\mathcal{C} = f(\mathcal{B}^k) \subseteq \mathcal{X}^n$ называется блоковым кодом, а его элементы — кодовыми словами.*

Предполагается, что введенное выше блоковое кодирование распространяется на множество $\mathcal{B}^*$ всех конечных слов в алфавите $\mathcal{B}$ следующим образом. Если k делит длину $|b|$ входного слова $b \in \mathcal{B}^*$, то это слово разбивается на непересекающиеся блоки длиной k, и каждый такой блок заменяется на блок длиной n в алфавите $\mathcal{X}$ в соответствии с отображением (кодированием) f независимо от других блоков. В общем случае кодирование состоит из шагов:

1) к входному слову $b \in \mathcal{B}^*$ длиной $|b|$ дописывается нулевое слово 0^t длиной t, где t — минимальное неотрицательное целое такое, что $k \mid (|b| + t)$;

2) полученное новое D-ичное слово разбивается на блоки длиной k: $b||0^t = b^{(1)}||b^{(2)}|| \ldots ||b^{(d)}$, $b^{(i)} \in \mathcal{B}^k$;

3) формируется дополнительный блок $b^{(d+1)}$, содержащий представление длиной слова b или числа дописанных нулей t;

4) выполняется кодирование каждого блока $b^{(i)}, i = 1, \ldots, d+1$, с использованием отображения f.

Отметим, что в конкретных случаях данная схема может претерпевать изменения.

Кодовые слова кода $\mathcal{C} \subseteq \mathcal{X}^n$ предназначены для передачи по дискретному каналу связи без памяти $(\mathcal{X}, \mathcal{Y}, \pi)$. Если на вход блокового кодера приходит информационный блок $b = (b_1, \ldots, b_k) \in \mathcal{B}^k$, то полученное кодовое слово $x = f(b) \in \mathcal{C}$ поступает в канал связи, и на выходе канала появляется некоторое слово $y \in \mathcal{Y}^n$. Далее слово y поступает на вход некоторого декодера $\mathfrak{D}$. Если декодер $\mathfrak{D}$ принимает решение, что на входе канала связи было кодовое слово $\tilde{x} \in \mathcal{C}$, то к этому кодовому слову применяется обратное преобразование $\tilde{b} = f^{-1}(\tilde{x})$, и в итоге получается информационный блок $\tilde{b} = f^{-1}(\tilde{x})$, который может и не совпадать с исходным блоком b.

Далее полагаем, что $\mathcal{X} = \mathcal{Y}$. Если в канал связи было передано кодовое слово $x \in \mathcal{X}^n$, из канала принято слово $y \in \mathcal{Y}^n$, которое отличается в s позициях от x, то произошло s ошибок, а $n-s$ символов передано правильно. Обнаружение ошибок при передаче происходит, если принятое из канала слово y не является кодовым, т. е. $y \notin \mathcal{C}$. Далее, когда декодер вычисляет кодовое слово $\tilde{x}$, то происходит исправление ошибок; если при этом оказывается, что результат декодирования $\tilde{x}$ совпадает с отправленным в канал кодовым словом x, то декодер правильно исправил ошибки, в противном случае — декодер ошибся.

В предположении, что все кодовые слова имеют одинаковую вероятность быть переданными по каналу, наилучшим решением на приемнике будет декодирование в такое кодовое слово, которое отличается от полученного в наименьшем числе символов, т. е. декодирование в ближайшее кодовое слово. Алгоритм декодирования в ближайшее кодовое слово основан на расстоянии Хэмминга.

Определение 8.2. *Расстоянием Хэмминга на множестве $\mathcal{X}$ называется отображение $\rho : \mathcal{X}^n \times \mathcal{X}^n \to \mathbb{N} \cup \{0\}$, задаваемое формулой*

$$\rho(a, b) = \sum_{i=1}^{n} \mathsf{I}\{a_i \neq b_i\} = n - \sum_{i=1}^{n} \mathsf{I}\{a_i = b_i\}, \tag{8.1}$$

где $a = a_1 \ldots a_n \in \mathcal{X}^n$, $b = b_1 \ldots b_n \in \mathcal{X}^n$.

Легко показать, что для любых слов $a, b, c \in \mathcal{X}^n$ расстояние Хэмминга обладает следующими свойствами:

1) (неотрицательность) $\rho(a, b) \geqslant 0$, причем равенство имеет место тогда и только тогда, когда $a = b$;

2) (симметричность) $\rho(a, b) = \rho(b, a)$;

3) (неравенство треугольника) $\rho(a, b) + \rho(b, c) \geqslant \rho(c, a)$.

Алгоритм декодирования в ближайшее кодовое слово состоит в том, что если принято слово y, то вычисляются расстояния $\rho(y, x)$ для всех кодовых слов $x \in \mathcal{C}$ и в качестве результата декодирования выбирается кодовое слово, для

которого это расстояние минимально. Если минимальное расстояние достигается для нескольких кодовых слов, то выбирается одно из них по некоторому заранее оговоренному правилу. Декодер в ближайшее кодовое слово будем обозначать $\mathfrak{D}_\rho$.

Отметим, что при использовании двоичного симметричного канала с параметром p декодеры $\mathfrak{D}_L$ (по методу максимального правдоподобия) и $\mathfrak{D}_\rho$ (в ближайшее кодовое слово) эквивалентны. Декодер $\mathfrak{D}_L$ вводится в гл. 9, где более подробно исследуются его свойства.

Лемма 8.1. *Пусть $\mathcal{X} = \mathcal{Y} = \{0,1\}$, канал связи $(\mathcal{X}, \mathcal{Y}, \pi)$ представляет собой ДСК(p), $0 < p < \frac{1}{2}$. Тогда декодеры $\mathfrak{D}_L$ и $\mathfrak{D}_\rho$ эквивалентны.*

Доказательство. По определению ДСК(p) имеем

$$\pi^{(n)}(y|x) = \prod_{i=1}^{n} \pi(y_i|x_i) = (1-p)^{n-s}p^s, \tag{8.2}$$

где $s = \rho(x,y)$. Если $\tilde{x}$ — другое кодовое слово и $\tilde{s} = \rho(\tilde{x}, y)$, то из (8.2) получим

$$\frac{\pi^{(n)}(y|\tilde{x})}{\pi^{(n)}(y|x)} = \frac{(1-p)^{n-\tilde{s}}p^{\tilde{s}}}{(1-p)^{n-s}p^s} = \left(\frac{p}{1-p}\right)^{\tilde{s}-s}.$$

Поскольку $0 < \dfrac{p}{1-p} < 1$, неравенство $\pi^{(n)}(y|\tilde{x}) \geqslant \pi^{(n)}(y|x)$ равносильно неравенству $\tilde{s} \leqslant s$. Это означает, что из двух кодовых слов x и $\tilde{x}$ декодер $\mathfrak{D}_L$ выбирает $\tilde{x}$ тогда и только тогда, когда это же верно и для декодера $\mathfrak{D}_\rho$. □

Для случая $q > 2$ декодеры $\mathfrak{D}_L$ и $\mathfrak{D}_\rho$ не эквивалентны.

Пусть, например, имеет место троичный алфавит $\mathcal{X} = \mathcal{Y} = \{0,1,2\}, q = 3$, и канал связи задается матрицей переходных вероятностей

$$\pi = \begin{pmatrix} p & 1-p & 0 \\ 1-p & p & 0 \\ p & 0 & 1-p \end{pmatrix}.$$

Пусть $n = 3$ и код состоит из двух кодовых слов $x = 011$, $\tilde{x} = 002$. Если $y = 000$, то $\rho(y,x) = 2$, $\rho(y,\tilde{x}) = 1$ и отношение

$$\frac{\pi^{(3)}(y|\tilde{x})}{\pi^{(3)}(y|x)} = \frac{p^3}{p(1-p)^2} = (\frac{p}{1-p})^2$$

не меньше 1 тогда и только тогда, когда $p \geqslant \dfrac{1}{2}$. Отсюда видно, что если $p < \dfrac{1}{2}$, то при $y = 000$ декодеры $\mathfrak{D}_L$ и $\mathfrak{D}_\rho$ дают разные результаты.

Очевидно, что можно предложить пример и для случая $q > 3$. Таким образом, свойством эквивалентности обладают декодеры $\mathfrak{D}_L$ и $\mathfrak{D}_\rho$ именно для двоичного симметричного канала связи. Приведем такой пример: свойство некоррелируемости эквивалентно свойству независимости дискретных случайных величин только в том случае, когда имеет место именно двоичный алфавит. В общем случае свойство некоррелируемости не эквивалентно свойству независимости дискретных случайных величин.

Важным параметром, определяющим способность декодера обнаруживать и исправлять ошибки, является минимальное кодовое расстояние, или кодовое расстояние.

Определение 8.3. *Минимальным кодовым расстоянием блокового кода* $\mathcal{C} \subseteq \mathcal{X}^n$, $|\mathcal{C}| \geqslant 2$, *называется величина*

$$d_{\mathcal{C}} = \min\{\rho(a,b) : \ a,b \in \mathcal{C},\ a \neq b\}. \tag{8.3}$$

Далее речь будет идти о декодере $\mathfrak{D}_\rho$ (в ближайшее кодовое слово).

Лемма 8.2. *Для того чтобы обнаружить все комбинации из* $d-1$ *или меньшего числа ошибок, необходимо и достаточно, чтобы* $d_{\mathcal{C}} \geqslant d$.

Доказательство. Достаточность. Если расстояние между любыми кодовыми словами не меньше, чем d, то никакая комбинация из $d-1$ ошибки не может перевести одно кодовое слово в другое.

Необходимость. Предположим от противного, что $d_{\mathcal{C}} \leqslant d-1$, тогда существует пара кодовых слов $c_1, c_2 \in \mathcal{C}$, такие, что $\rho(c_1,c_2) < d$, и поэтому найдется комбинация из числа ошибок $\rho(c_1,c_2)$, которая может перевести одно кодовое слово $c_1 \in \mathcal{C}$ в другое $c_2 \in \mathcal{C}$. В такой ситуации декодер $\mathfrak{D}_\rho$ не обнаружит наличие ошибок. □

Лемма 8.3. *Для того чтобы исправить все комбинации из* t *или меньшего числа ошибок, необходимо и достаточно, чтобы* $d_{\mathcal{C}} \geqslant 2t+1$.

Доказательство. Достаточность. Любое полученное слово $y \in \mathcal{X}^n$ с $\tau \leqslant t$ ошибками отличается от переданного кодового слова $c \in \mathcal{C}$ в τ позициях. Тогда выберем любое кодовое слово $c_1 \in \mathcal{C}$, такое, что $c_1 \neq c$, и, учитывая неравенство треугольника, получим $\rho(c_1,y) \geqslant \rho(c,c_1) - \tau \geqslant t+1 > \tau$. Значит, слово y отличается не меньше, чем в $\tau+1$ позициях от любого другого кодового слова $c_1 \neq c$. Согласно правилу декодирования в ближайшее кодовое слово декодер $\mathfrak{D}_\rho$ осуществит правильное декодирование.

Необходимость. Если минимальное расстояние меньше, чем $2t+1$, например $d_{\mathcal{C}} = 2t$, то хотя бы в одном случае t-кратная ошибка приведет к такому слову на выходе, которое столь же близко к одному из не передававшихся кодовых слов, как и к переданному кодовому слову. Декодер $\mathfrak{D}_\rho$ может ошибочно принять за переданное кодовое слово то, которое не передавалось. □

Рассмотрим еще один случай, когда входной алфавит $\mathcal{X}$ и выходной алфавит $\mathcal{Y}$ канала связи не совпадают. Предположим, что алфавит $\mathcal{Y}$ состоит из символов алфавита $\mathcal{X} = \{x_1, \ldots, x_q\}$ и еще одного дополнительного символа E, называемого символом стирания. Если на входе канала был символ x_i, а на выходе появился символ $x_j \neq x_i$, то произошло искажение переданного символа; если же на выходе появился символ E, то произошло стирание переданного символа. Введем понятие обобщенного расстояния Хэмминга.

Определение 8.4. *Обобщенным расстоянием Хэмминга на множестве* $\mathcal{Y}^n, \mathcal{Y} = \{x_1, \ldots, x_q, E\}$, *называется отображение* $\tilde{\rho} : \ \mathcal{Y}^n \times \mathcal{Y}^n \to \mathbb{R}$, *задаваемое формулой*

$$\tilde{\rho}(a,b)=\sum_{i=1}^{n}\tilde{I}(a_i,b_i), \tag{8.4}$$

где $a=a_1\dots a_n\in\mathcal{Y}^n$, $b=b_1\dots b_n\in\mathcal{Y}^n$ *и*

$$\tilde{I}(a_i,b_i)=\begin{cases}0, & \text{если } a_i=b_i,\\ 1, & \text{если } a_i\neq b_i,\ a_i\neq E, b_i\neq E,\\ 1/2, & \text{если } a_i\neq b_i \text{ и либо } a_i=E, \text{ либо } b_i=E.\end{cases}$$

Заметим, что обобщенное расстояние Хэмминга также обладает свойствами неотрицательности, симметричности и для него выполняется неравенство треугольника.

Определение декодера $\mathfrak{D}_\rho$ можно обобщить на случай $\mathcal{Y}=\mathcal{X}\cup\{E\}$, заменив метрику ρ на $\tilde{\rho}$.

Лемма 8.4. *Если* $\mathcal{Y}=\mathcal{X}\cup\{E\}$, $d_{\mathcal{C}}=d$ *и* $t_0+2t_1\leqslant d-1$, *то обобщенный декодер* $\mathfrak{D}_{\tilde{\rho}}$ *правильно исправляет любые комбинации из* t_0 *стираний и* t_1 *искажений символов.*

Доказательство. Предположим, что при передаче кодового слова $x\in\mathcal{C}$ в канале связи произошло t_0 стираний и t_1 искажений символов, так что для принятого слова y имеем $\tilde{\rho}(x,y)=t_0/2+t_1\leqslant\dfrac{d-1}{2}$.

Докажем, что для любого другого кодового слова $\tilde{x}\neq x$ справедливо неравенство $\tilde{\rho}(\tilde{x},y)>\dfrac{d-1}{2}$. Действительно, если $\tilde{\rho}(\tilde{x},y)\leqslant\dfrac{d-1}{2}$, то по неравенству треугольника $\tilde{\rho}(\tilde{x},x)\leqslant\tilde{\rho}(\tilde{x},y)+\tilde{\rho}(y,x)\leqslant d-1$, что противоречит условию $d_{\mathcal{C}}=d$.

Таким образом, расстояние $\tilde{\rho}$ от принятого слова y до кодового слова $x\in\mathcal{C}$ меньше, чем до любого другого кодового слова, поэтому декодирование происходит правильно. □

8.2. ЛИНЕЙНЫЕ КОДЫ

Рассмотрим важный класс блоковых кодов — линейные коды, обладающие особой математической структурой.

Определение 8.5. *Пусть* V_n *— n-мерное векторное пространство над полем* $\mathbb{F}_q$. *Линейным блоковым q-ичным кодом длиной n называется произвольное подпространство* $\mathcal{C}\subseteq V_n$.

Если размерность кода $\mathcal{C}$ равна $\dim\mathcal{C}=k$, то $\mathcal{C}$ называют линейным (n,k)-кодом; если, кроме того, $d_{\mathcal{C}}=d$, то код $\mathcal{C}$ — линейный (n,k,d)-код.

Определение 8.6. *Скорость передачи информации есть отношение числа информационных символов к общей длине кодового слова:*

$$\mathcal{R}=\frac{k}{n}. \tag{8.5}$$

Лемма 8.5. *Если* $\mathcal{C}$ *— линейный код, то справедливо равенство*

$$d_{\mathcal{C}}=\min_{c\in\mathcal{C},\ c\neq 0}w(c).$$

Доказательство. Для ненулевого кодового слова $c \in \mathcal{C}$ имеем $w(c) = \rho(c,0) \geqslant$ $\geqslant d_{\mathcal{C}}$. С другой стороны, если для некоторых двух различных кодовых слов $u, v \in \mathcal{C}$ достигается наименьшее значение $\rho(u,v) = d_{\mathcal{C}}$, то в силу линейности кода имеем $u - v \in \mathcal{C}$, $\rho(u,v) = d_{\mathcal{C}} = w(u-v)$. □

Следовательно, кодовое расстояние для линейного кода равно минимальному весу его ненулевых кодовых слов. Это позволяет для линейного кода упростить, по сравнению с произвольным блоковым кодом, нахождение минимального кодового расстояния $d_{\mathcal{C}}$. Действительно, чтобы для некоторого кода $\mathcal{C}$ вычислить $d_{\mathcal{C}}$ по определению, необходимо найти все попарные расстояния между различными кодовыми словами, т. е. перебрать $\binom{|\mathcal{C}|}{2}$ вариантов, а для линейного кода достаточно найти веса ненулевых кодовых слов, т. е. перебрать $|\mathcal{C}|-1$ вариантов.

Пример 8.1. Пусть $q = 2$. Рассмотрим код $\mathcal{C}$, состоящий из двух кодовых слов $c_0 = 00\ldots0$, $c_1 = 11\ldots1$, $|c_i| = n$, $i \in \{0,1\}$. Он обладает большой помехозащищенностью, но очень малой скоростью передачи информации:

$$d_{\mathcal{C}} = n,\ \mathcal{R} = \frac{1}{n}.$$

Введенный код называется кодом кратных повторений.

Определение 8.7. *Пусть $\mathcal{C} \subseteq V_n$ — линейный код размерности k. Матрица $G = (g_{ij})$ размером $k \times n$, составленная из базисных векторов (слов) $g_i = g_{i1}\ldots g_{in}$, $1 \leqslant i \leqslant k$, подпространства $\mathcal{C}$, называется порождающей матрицей кода $\mathcal{C}$.*

Каждое кодовое слово $c \in \mathcal{C}$ единственным образом представляется в виде линейной комбинации базисных слов:

$$c = u_1 g_1 + \ldots + u_k g_k,$$

где $u_1, \ldots, u_k \in \mathbb{F}_q$, g_i — базисные слова, или в матричном виде

$$c = uG,\ u = u_1 \ldots u_k. \tag{8.6}$$

Формулу (8.6) будем рассматривать как реализацию блокового кодирования f, которое информационному q-ичному слову $u_1u_2\ldots u_k$ ставит в соответствие кодовое q-ичное слово $c_1c_2\ldots c_n$. Таким образом, преобразование кодирования представляет собой умножение слова на порождающую матрицу G линейного кода $\mathcal{C}$.

Введем обозначение: $\mathcal{C}^{\perp} = \{u \in V_n :\ u \perp v\ \forall v \in \mathcal{C}\}$ — множество слов, ортогональных всем словам k-мерного линейного подпространства $\mathcal{C}$, само является линейным подпространством пространства V_n и имеет размерность $n-k$.

Определение 8.8. *Порождающая матрица H для подпространства $\mathcal{C}^{\perp}$ называется проверочной матрицей кода $\mathcal{C}$.*

Определение 8.9. *Ортогональное подпространство $\mathcal{C}^{\perp}$ с порождающей матрицей H называется линейным $(n, n-k)$-кодом, двойственным к линейному (n,k)-коду $\mathcal{C}$.*

Очевидно, что проверочная матрица H линейного (n,k)-кода $\mathcal{C}$ имеет размерность $(n-k) \times n$, а ее $n-k$ строк $h_1, \ldots, h_{n-k}$ образуют базис кода $\mathcal{C}^{\perp}$,

$h_i = h_{i1} \dots h_{in}, 1 \leqslant i \leqslant n-k$. Из определения двойственного кода следует, что каждая строка g_i матрицы G ортогональна всем строкам матрицы H. Поэтому справедливо равенство $GH' = 0$, где 0 есть нулевая матрица размером $k \times (n-k)$.

Лемма 8.6. *Пусть $\mathcal{C}$ — линейный (n,k)-код, $k < n$, H — проверочная матрица кода $\mathcal{C}$ и y — произвольное слово из V_n. Условие $y \in \mathcal{C}$ эквивалентно равенству $yH' = 0^{n-k}$, где 0^{n-k} — нулевое слово длиной $n-k$.*

Доказательство. Следует из определения 8.8. □

Приведем лемму о линейном соотношении столбцов проверочной матрицы.

Лемма 8.7. *Пусть $\mathcal{C}$ — линейный (n,k)-код с проверочной матрицей H. Тогда равенство $d_{\mathcal{C}} = d$ равносильно выполнению условий:*

1) любые $d-1$ столбцов матрицы H линейно независимы;

2) в матрице H существуют d линейно зависимых столбцов.

Доказательство. Условие $cH' = 0^{n-k}$ является необходимым и достаточным для принадлежности слова $c = c_1 \dots c_n$ коду $\mathcal{C}$:

$$c_1 h_1^{\downarrow} + \dots + c_n h_n^{\downarrow} = 0^{n-k}, \tag{8.7}$$

где $h_i^{\downarrow}$ обозначает i-й столбец матрицы H. Если $w(c) = t > 0$, причем ненулевые в слове c — позиции с номерами $i_1, \dots, i_t$, то условие (8.7) примет вид

$$c_{i_1} h_{i_1}^{\downarrow} + \dots + c_{i_t} h_{i_t}^{\downarrow} = 0^{n-k}. \tag{8.8}$$

Это означает, что t столбцов $h_{i_1}^{\downarrow}, \dots, h_{i_t}^{\downarrow}$ матрицы H линейно зависимы.

Если $d_{\mathcal{C}} = d$, то найдется слово $c \in \mathcal{C}$ веса $w(c) = d$, и поэтому некоторые d столбцов матрицы H линейно зависимы. По определению минимального кодового расстояния никакое ненулевое слово c, вес которого меньше d, не является кодовым словом, поэтому никакие нетривиальные линейные комбинации из $d-1$ или меньшего числа столбцов матрицы H не равны нулю. Таким образом, условия 1), 2) теоремы необходимы для выполнения равенства $d_{\mathcal{C}} = d$. Достаточность этих условий доказывается аналогично. □

Введем отношение эквивалентности на множестве линейных кодов.

Определение 8.10. *Два линейных кода $\mathcal{C}_1$ и $\mathcal{C}_2$ называются эквивалентными, если они имеют одинаковую размерность и код $\mathcal{C}_2$ можно получить из кода $\mathcal{C}_1$ при помощи некоторой фиксированной перестановки позиций во всех кодовых словах.*

Отметим, что эквивалентные коды отличаются перестановкой столбцов в порождающей матрице. Очевидно, что при одинаковой перестановке символов в двух словах расстояние Хэмминга между ними не изменится. Отсюда следует, что структура расстояний между кодовыми словами линейного кода не меняется при переходе к эквивалентному коду; в частности, минимальные кодовые расстояния двух эквивалентных кодов равны. Поэтому при декодировании в ближайшее кодовое слово можно считать, что два эквивалентных кода приводят к одинаковым возможностям по обнаружению и исправлению ошибок.

Определение 8.11. *Линейный (n,k)-код $\mathcal{C}$ называется систематическим,*

если первые k символов каждого кодового слова — информационные символы, а последние $n-k$ символов — проверочные.

Порождающая матрица систематического кода имеет вид $G = (I_k|P)$, где I_k — единичная матрица порядка k, а P — произвольная матрица размером $k\times$ $\times(n-k)$. Такая матрица называется приведенной.

Лемма 8.8. *Для любого линейного (n,k)-кода с $k \geqslant 1$ существует эквивалентный ему систематический код.*

Доказательство. Пусть $\mathcal{C}$ — линейный (n,k)-код и G — его порождающая матрица. Применяя алгоритм исключения Гаусса, элементарными преобразованиями строк приведем матрицу G к матрице G_1 ступенчатого вида, где строки начинаются с нескольких нулей, за которыми следует единица, причем первая единица в j-й строке находится в столбце i_j, для $j \in \{1, \ldots, k\}$, $i_1 < i_2 < \ldots < i_k$, а остальные элементы в столбцах $i_1, \ldots, i_k$ равны нулю.

При элементарных преобразованиях строк ранг матрицы сохраняется, а получаемые строки остаются словами кода $\mathcal{C}$. Поэтому строки матрицы G_1 также образуют базис кода $\mathcal{C}$, т. е. матрица G_1 — порождающая для кода $\mathcal{C}$. В матрице G_1 выполним перестановку столбцов: поставим столбец i_1 на первое место, столбец i_2 — на второе место и так до столбца i_k, который ставим на k-е место, а остальные столбцы расположим произвольно на позициях $k+1, \ldots, n$. В результате получим матрицу $G_2 = (I_k|P)$, которая будет порождающей для некоторого систематического (n,k)-кода $\mathcal{C}_1$, эквивалентного коду $\mathcal{C}$. □

Если порождающая матрица кода имеет приведенную форму, то одна из проверочных матриц может быть легко найдена.

Лемма 8.9. *Если $\mathcal{C}$ — систематический линейный (n,k)-код с порождающей матрицей $G = (I_k|P)$, $1 \leqslant l < n$, то проверочная матрица имеет вид*

$$H = (-P^{'}|I_{n-k}), \tag{8.9}$$

где $-P^{'}$ означает матрицу, полученную из P транспонированием и заменой всех элементов на обратные по сложению в поле $\mathbb{F}_q$, а I_{n-k} — единичная матрица порядка $n-k$.

Доказательство. Осуществляется непосредственной проверкой. □

8.3. ДЕКОДИРОВАНИЕ ЛИНЕЙНЫХ КОДОВ

Пространство V_n можно разбить на классы эквивалентных слов, и эти классы называются смежными классами по коду $\mathcal{C}$. Если u — представитель некоторого смежного класса, то этот класс можно записать в виде

$$u + \mathcal{C} = \{u + c : \ c \in \mathcal{C}\}.$$

Если размерность кода $\mathcal{C}$ равна k, то мощность каждого смежного класса будет q^k, а общее число смежных классов — q^{n-k}.

Определение 8.12. *Слово $e = y - c$ называется вектором ошибок, если по каналу связи было отправлено слово c, а получено слово y.*

Очевидно, что задача декодирования, т. е. восстановления кодового слова c по принятому слову y, эквивалентна восстановлению вектора ошибок e. Поскольку ошибки в канале связи могут быть любыми, вектор e может, вообще говоря, принимать произвольные значения из V_n. Однако тот факт, что код $\mathcal{C}$ является линейным подпространством пространства V_n, позволяет существенно сократить перебор возможных вариантов значений слова e. Действительно, условие

$$y - e = c \in \mathcal{C}$$

означает, что подлежащий определению вектор ошибок e лежит в том же смежном классе, что и принятое слово y, тем самым перебор вариантов при поиске вектора e сокращается до одного заданного смежного класса $y + \mathcal{C}$.

Если $e = 0$, то $y = c$ и, следовательно, ошибок при передаче не произошло. Если $e \neq 0^n$, то ненулевые символы вектора ошибок e соответствуют искаженным символам кодового слова c.

Определение 8.13. *Таблицей стандартного расположения для линейного (n, k)-кода $\mathcal{C} = \{c_1 = 0, c_2, \ldots, c_{q^k}\}$ называется таблица $S = (y_{ij})$ размером $q^{n-k} \times q^k$, составленная из слов пространства V_n по следующему правилу:*

1) *в первой строке s_1 таблицы S записаны q^k слов кода $\mathcal{C}$,*

$$s_1 = (y_{11}, \ldots, y_{1q^k}) = (c_1 = 0, c_2, \ldots, c_{q^k});$$

2) *если строки $s_1, \ldots, s_{i-1}$ уже построены, то в качестве слова y_{i1} выбирается произвольное слово из V_n, которое еще не встретилось в построенных первых $i-1$ строках, а остальные элементы i-й строки находятся по формуле*

$$y_{ij} = y_{i1} + c_j,\ 2 \leqslant j \leqslant q^k.$$

Такой способ построения таблицы S гарантирует, что все ее элементы различны. Строками таблицы стандартного расположения являются все различные смежные классы по коду $\mathcal{C}$. В первом столбце таблицы S находятся образующие классов смежности. Отметим также, что таблицу стандартного расположения при заданных ограничениях можно построить не единственным образом.

Определение 8.14. *Пусть $\mathcal{C}$ — линейный код. Декодером $\mathfrak{D}_S$ на основе таблицы стандартного расположения S называется такое отображение $\mathfrak{D}_S : V_n \to \mathcal{C}$, что если принятое слово y совпадает с элементом y_{ij} таблицы S, то $\mathfrak{D}_S(y) = c_j$.*

Заметим, что несмотря на простоту метода декодирования, он неприемлем для реальных линейных кодов.

Лемма 8.10. *Если используется декодер $\mathfrak{D}_S$ на основе таблицы стандартного расположения S для линейного (n, k)-кода, то по полученному слову y будет правильно декодировано переданное слово c тогда и только тогда, когда вектор ошибок $e = c - y$ является образующим смежного класса.*

Доказательство. Если $y - c = y_{i1}$, где y_{i1} — образующий i-го смежного класса, то слово $y = c + y_{i1}$ должно находиться в таблице стандартного расположения S в i-м смежном классе под кодовым словом c и поэтому будет правильно де-

кодирован. Если же слово $y - c$ не является образующим смежного класса, то слово y должно находиться в некотором смежном классе, например j-м, с образующим y_{j1}. Тогда слово y расположено в j-й строке, но не под словом c, потому что $y \neq y_{j1} + c$. □

Существенное уменьшение трудоемкости декодера $\mathfrak{D}_S$ возможно за счет использования понятия синдрома слова, определяемого при фиксированной проверочной матрице.

Определение 8.15. *Пусть $\mathcal{C}$ — линейный код и H — его проверочная матрица. Синдромом $s(y)$ слова $y \in V_n$ называется слово $s(y) = yH'$ длиной $(n-k)$.*

Теорема 8.1. *Пусть $\mathcal{C}$ — линейный (n,k)-код, тогда справедливы следующие утверждения:*

1) *условия $y \in \mathcal{C}$ и $s(y) = 0$ равносильны;*

2) *слова $y, \tilde{y}$ лежат в одном смежном классе по коду $\mathcal{C}$ тогда и только тогда, когда $s(y) = s(\tilde{y})$.*

Доказательство. Первая часть утверждения непосредственно следует из определения синдрома. Вторая часть — из того, что условие $y - \tilde{y} \in \mathcal{C}$ равносильно равенству $(y - \tilde{y})H' = 0$, или, что то же самое, $yH' = \tilde{y}H'$. □

Имеется всего q^{n-k} различных значений синдрома по числу смежных классов, и поэтому синдром можно рассматривать просто как q-ичный $(n-k)$-разрядный номер смежного класса. Таблице стандартного расположения S можем сопоставить новую таблицу $\tilde{S}$ размера $q^{n-k} \times 2$, первый столбец которой совпадает с первым столбцом таблицы S, а во втором столбце записаны синдромы соответствующих смежных классов:

$$\tilde{S} = \begin{pmatrix} y_{11} & s(y_{11}) \\ \vdots & \vdots \\ y_{i1} & s(y_{i1}) \\ \vdots & \vdots \\ y_{q^{n-k},1} & s(y_{q^{n-k},1}) \end{pmatrix}. \tag{8.10}$$

При помощи таблицы $\tilde{S}$ декодер $\mathfrak{D}_S$ можно представить в виде эквивалентного, но более простого алгоритма: для принятого слова y вычислим синдром $s(y)$ и найдем этот синдром во втором столбце таблицы $\tilde{S}$; если $s(y) = y_{i1}$, то, положим, $\mathfrak{D}_S(y) = y - y_{i1}$.

Приведем теорему из [8].

Теорема 8.2. *Пусть $\mathcal{C}$ — (n,k)-код, используемый для передачи по двоичному симметричному каналу, и все кодовые слова имеют одну и ту же вероятность быть переданными. Тогда средняя вероятность правильного декодирования будет наибольшей, если в таблице стандартного расположение каждое образующее слово смежного класса имеет минимальный вес в своем классе.*

Доказательство. Пусть s_{ij} — слово, расположенное в i-й строке и j-м столбце таблицы декодирования. Обозначим через c_{0j} кодовое слово в верхней строке

j-го столбца, а через d_{ij} — кодовое расстояние между s_{ij} и c_{0j} ($d_{ij} = \rho(s_{ij},c_{0j})$). Вероятность правильного декодирования, если было передано слово c_{0j}, равна

$$\sum_{i=0}^{2^{n-k}-1} p^{d_{ij}} q^{n-d_{ij}}, \tag{8.11}$$

где p — вероятность ошибки в канале, $q = 1 - p$ — вероятность правильной передачи в канале.

Поскольку имеется 2^k кодовых слов, которые используются равновероятно, то при усреднении вероятности правильного декодирования по всему коду $\mathcal{C}$ получим

$$p_c = \frac{1}{2^k} \sum_{j=0}^{2^k-1} \sum_{i=0}^{2^{n-k}-1} p^{d_{ij}} q^{n-d_{ij}}. \tag{8.12}$$

Каждому возможному двоичному слову на выходе канала в этой сумме соответствует одно слагаемое; каждое из этих слагаемых принимает максимальное значение, если соответствующее слово декодируется в ближайшее кодовое слово в смысле метрики Хэмминга, так как $p^{d_{ij}} q^{n-d_{ij}}$ — монотонно убывающая функция от расстояния d_{ij}. Из сказанного следует, что вероятность правильного декодирования будет максимальной, если каждое полученное слово будет декодировано в ближайшее кодовое слово.

Предположим теперь, что некоторое слово y расположено в таблице декодирования под кодовым словом c так, что $\rho(y,c) = w$. Допустим, что ближайшее кодовое слово c_1 находится на расстоянии w_1 от y. Пусть f — образующее слово смежного класса, содержащего слово y. Тогда вес слова $f = y - c$ равен w. Элемент $h - g_1 = f + (g - g_1)$ имеет вес w_1 и лежит в том же самом смежном классе. Поскольку предполагалось, что f обладает минимальным весом в своем смежном классе, то $w_1 \geqslant w$, и поэтому y лежит, по крайней мере, так же близко к c, как и к c_1. □

8.4. ГРАНИЦЫ ДЛЯ ПАРАМЕТРОВ КОДА

Важный вопрос при изучении линейных кодов — определение допустимых значений параметров n (длина кодового слова), k (размерность кода) и d (минимальное кодовое расстояние). Если два из трех параметров заданы, то возникает задача оптимизации свойств кода:

1) среди кодов с одинаковыми параметрами n, k лучшим является код, который имеет большее кодовое расстояние d;

2) среди кодов с одинаковыми параметрами n, d лучшим является код, который имеет большее число информационных символов k;

3) среди кодов с одинаковыми параметрами k, d лучшим является код, который имеет меньшую длину n, а следовательно, и меньшее число проверочных символов.

Между рассмотренными параметрами n, k, d существуют известные определенные соотношения, задаваемые границами для минимального кодового расстояния $d_{\mathcal{C}}$ или для скорости передачи информации $\mathcal{R}$.

Теорема 8.3 (верхняя граница Хэмминга). *Для любого линейного q-ичного кода $\mathcal{C}$ с длиной блока n, k информационными символами и кодовым расстоянием $d = 2t + 1$ справедливо неравенство*

$$q^{n-k} \geqslant \sum_{i=0}^{t} \binom{n}{i} (q-1)^i. \tag{8.13}$$

Доказательство. Рассмотрим шары в n-мерном пространстве V_n над полем $\mathbb{F}_q$:

$$K_t(c) = \{y \in V_n : \ \rho(c, y) \leqslant t\}$$

с центрами в кодовых словах $c \in \mathcal{C}$. Очевидно, что

$$\bigcup_{c \in \mathcal{C}} K_t(c) \subseteq V_n.$$

Докажем, что если кодовые слова $c, \tilde{c}$ различны, то шары $K_t(c)$ и $K_t(\tilde{c})$ не пересекаются. Действительно, если два этих шара содержат общий элемент y, то по неравенству треугольника

$$\rho(c, \tilde{c}) \leqslant \rho(c, y) + \rho(y, \tilde{c}) \leqslant 2t < d,$$

что противоречит определению минимального кодового расстояния $d_{\mathcal{C}} = d$.

Следовательно,

$$\bigcup_{c \in \mathcal{C}} |K_t(c)| \leqslant |V_n| = q^n.$$

Заметим, что все q^k шаров $K_t(c)$, $c \in \mathcal{C}$, равномощны, и мощность каждого из них равна

$$|K_t(c)| = \sum_{i=0}^{t} \binom{n}{t} (q-1)^i,$$

поэтому $q^k \sum_{i=0}^{t} \binom{n}{t} (q-1)^i \leqslant q^n$, что равносильно (8.13). □

В формуле (8.13) равенство достигается в том случае, когда $q^k |K_t(c)| = q^n$. Геометрически это означает, что непересекающиеся шары, проведенные вокруг точек, задаваемых кодовыми словами $c_1, \ldots, c_{q^k} \in \mathcal{C}$, заполняют все векторное пространство V_n, и при этом каждое слово из V_n оказывается в одной из сфер.

Определение 8.16. *Линейные коды, для которых выполняется равенство $q^{n-k} = \sum_{i=0}^{t} \binom{n}{i} (q-1)^i$, называются совершенными или плотно упакованными.*

Примером совершенного кода является код кратных повторений.

Отметим, что поскольку в доказательстве границы Хэмминга не использовалось свойство линейности кода, эта граница остается верной и для любого блокового кода $\mathcal{C}$, если положить $k = \log_q |\mathcal{C}|$.

Теорема 8.4 (граница Синглтона). *Для любого линейного q-ичного (n, k, d)-кода $\mathcal{C}$ справедливо неравенство*

$$d \leqslant n - k + 1. \tag{8.14}$$

Доказательство. Если $k = n$, то $\mathcal{C} = V_n$, $d = 1$, и нужное неравенство выполнено.

Пусть $k < n$. Из леммы (8.7) следует, что если $d_{\mathcal{C}} = d$, то любые $d - 1$ столбцов проверочной матрицы H линейно независимы. Поскольку столбцы матрицы H имеют длину $n - k$, максимальное число линейно независимых столбцов не может быть больше, чем $n - k$. Следовательно, $d - 1 \leqslant n - k$. □

Формула (8.14) справедлива не только для линейного (n, k, d)-кода, но и для любого блокового кода $\mathcal{C}$ длиной n с минимальным кодовым расстоянием $d_{\mathcal{C}} = d$, при $k = \log_q |\mathcal{C}|$.

Чтобы доказать такое обобщение, для всех кодовых слов блокового кода $\mathcal{C}$ рассмотрим их подслова, образованные первыми $n - d + 1$ символами. Все эти подвекторы различны, так как в противном случае если два различных кодовых слова совпадают на первых $n - d + 1$ позициях, то расстояние между этими кодовыми словами не превосходит $d - 1$, что противоречит условию $d_{\mathcal{C}} = d$. Итак, мы имеем $|\mathcal{C}| = q^k$ различных слов длиной $n - d + 1$, и это число не может быть больше q^{n-d+1}. Следовательно, $q^k \leqslant q^{n-d+1}$, поэтому $k \leqslant n - d + 1$.

Определение 8.17. *Линейный q-ичный (n, k, d)-код $\mathcal{C}$ называется кодом с максимально допустимым расстоянием или МДР-кодом, если в неравенстве (8.14) достигается равенство.*

Для линейного (n, k)-кода с минимальным расстоянием d граница Синглтона утверждает, что $n \geqslant d + k - 1$. Граница Грайсмера увеличивает правую часть этого неравенства. Введем обозначения: $N(k, d)$ — минимально возможная длина двоичного линейного кода размерностью k с минимальным кодовым расстоянием d.

Теорема 8.5 (граница Грайсмера). *Пусть $\mathcal{C}$ — (n, k)-код с минимальным расстоянием d. Тогда справедливо неравенство*

$$N(k,d) \geqslant d + N(k - 1, \lceil \frac{d}{2} \rceil). \tag{8.15}$$

Доказательство. Рассмотрим порождающую матрицу G кода $\mathcal{C}$. Без ограничения общности можно считать, что

$$G = \left(\begin{array}{cccccc|cccccc} 0 & 0 & \dots & 0 & 0 & & 1 & 1 & \dots & 1 & 1 & \\ \multicolumn{5}{c}{} & G_1 & \multicolumn{6}{c}{} \end{array} \right),$$

где $(k - 1) \times (N(k, d) - d)$-матрица G_1 имеет ранг $(k - 1)$, в противном случае мы бы обратили в нуль первую строку матрицы G_1 и минимальное расстояние кода $\mathcal{C}$ было бы меньше, чем d.

Пусть матрица G_1 порождает $(N(k, d) - d, k - 1)$-код с минимальным расстоянием d_1. Предположим, что слово $u|v \in \mathcal{C}$, где $w(u) = d_1$. Поскольку $u|\bar{v} \in \mathcal{C}$, то

$$d_1 + w(v) \geqslant d,\ d_1 + d - w(v) \geqslant d,$$

и, складывая эти неравенства, получим, что $2d_1 \geqslant d$ или $d_1 \geqslant \lceil \frac{d}{2} \rceil$. Следовательно,

$$N(k-1, \lceil \frac{d}{2} \rceil) \leqslant N(k,d) - d.$$

□

Теорема 8.6. *Если $\mathcal{C}$ — (n,k)-код с минимальным расстоянием d, то*

$$N(k,d) \geqslant \sum_{i=0}^{k-1} \lceil \frac{d}{2^i} \rceil. \tag{8.16}$$

Доказательство. Многократно применяя лемму 8.5, получим

$$N(k,d) \geqslant d + N(k-1, \frac{d}{2}) \geqslant d + \lceil \frac{d}{2} \rceil + N(k-2, \lceil \frac{d}{4} \rceil) \geqslant \ldots \geqslant$$

$$\geqslant \sum_{i=0}^{k-2} \lceil \frac{d}{2^i} \rceil + N(1, \lceil \frac{d}{2^{k-1}} \rceil) = \sum_{i=0}^{k-1} \lceil \frac{d}{2^i} \rceil.$$

□

Пример 8.2. Найдем число $N(5,7)$, т. е. самый короткий код размерности 5, исправляющий 3 ошибки. Согласно теореме 8.6 имеем

$$N(5,7) \geqslant 7 + \lceil \frac{7}{2} \rceil + \lceil \frac{7}{4} \rceil + \lceil \frac{7}{8} \rceil + \lceil \frac{7}{16} \rceil = 7 + 4 + 2 + 1 + 1 = 15.$$

В действительности существует $(15,5)$-код БЧХ с минимальным расстоянием 7, и поэтому $N(5,7) = 15$.

Теорема 8.7 (верхняя граница Плоткина). *Если существует q-ичный блоковый код длиной n с общим числом кодовых слов M и кодовым расстоянием d, то*

$$d \leqslant \frac{(q-1)n}{q(M-1)} M. \tag{8.17}$$

Доказательство. Оценка (8.17) получается в результате оценки сверху среднего расстояния между кодовыми словами.

Пусть $\mathcal{C} = \{c_1, \ldots, c_M\}$. Обозначим $\gamma_m^{(i)}$ m-ю компоненту i-го кодового слова. Для определения среднего расстояния подсчитаем сумму D всевозможных попарных расстояний между кодовыми словами:

$$D = \sum_{i=1}^{M} \sum_{j=1}^{M} \rho(c_i; c_j) = \sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{m=1}^{n} \rho(\gamma_m^{(i)}, \gamma_m^{(j)}),$$

где $\rho(\gamma_m^{(i)}, \gamma_m^{(j)}) = \begin{cases} 0, & \gamma_m^{(i)} = \gamma_m^{(j)}, \\ 1, & \gamma_m^{(i)} \neq \gamma_m^{(j)}. \end{cases}$

Меняя порядок суммирования, получим

$$D = \sum_{m=1}^{n} \sum_{i=1}^{M} \sum_{j=1}^{M} \rho(\gamma_m^{(i)}, \gamma_m^{(j)}).$$

Обозначим y_m^i — число символов, равных i, среди компонент $\gamma_m^{(1)}, \ldots, \gamma_m^{(M)}$, $y_m^i \geqslant 0$, $\sum_{i=1}^q y_m^i = M$.

Имеет место

$$\sum_{i=1}^{M}\sum_{j=1}^{M} d(\gamma_m^{(i)}, \gamma_m^{(j)}) = \sum_{i=1}^{q}\sum_{j=1}^{q} (1-\delta_{ij})\, y_m^i y_m^j =$$

$$= (y_m^1, \ldots, y_m^q)(1_q - I_q) \begin{pmatrix} y_m^1 \\ \vdots \\ y_m^q \end{pmatrix} = F(y, y),$$

где 1_q — матрица $(q \times q)$, состоящая из единиц.

Для оценки сверху условного максимума квадратичной формы $F(y, y)$, зависящей от неотрицательных целочисленных аргументов, удовлетворяющих уравнениям связи, найдем методом Лагранжа максимум квадратичной формы $F(x, x)$, где $x \in \mathcal{R}^q$.

Естественно, что $\max\limits_{\{y\}} F(y, y) \leqslant \max\limits_{\{x\}} F(x, x)$.

С использованием индукции по q покажем, что

$$\max_{\{x\}} F(x, x) = \frac{M^2(q-1)}{q},\ x_i \geqslant 0,\ \sum_{i=1}^{q} x_i = M. \tag{8.18}$$

Если $q = 2$, то

$$\max_{\{x\}} F(x, x) = (x_1, x_2) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 2x_1x_2,\ x_1 + x_2 = M,$$
$$x_1 \geqslant 0,\ x_2 \geqslant 0.$$

Составим функцию Лагранжа и найдем частные производные:

$$\Phi(x_1, x_2; \lambda) = 2x_1x_2 - \lambda(x_1 + x_2),\ \begin{cases} \dfrac{\partial \Phi}{\partial x_1} = 2x_2 - \lambda = 0, \\ \dfrac{\partial \Phi}{\partial x_2} = 2x_1 - \lambda = 0. \end{cases}$$

Решая систему, найдем, что точка условного максимума равна $(x_1^0, x_2^0) = (M/2, M/2)$, а соответствующее значение квадратичной формы равно $\max\limits_{\{x\}} F(x, x) = \dfrac{M^2}{2}$.

Пусть для $q-1$ справедливо утверждение

$$\max_{\{x\}} F((x_1, \ldots, x_{q-1}), (x_1, \ldots, x_{q-1})) = \frac{M^2(q-2)}{q-1}. \tag{8.19}$$

Покажем справедливость (8.18). Составим функцию Лагранжа

$$\Phi(x_1, \ldots, x_q; \lambda) = \sum_{ij} (1-\delta_{ij})\, x_i x_j - \lambda \sum_{i=1}^{q} x_i$$

и вычислим частные производные

$$\frac{\partial \Phi(x_1, \ldots, x_q; \lambda)}{\partial x_i} = \sum_{j=1}^{q} (1 - \delta_{ij})\, x_j - \lambda = 0,\ 1 \leqslant i \leqslant q. \tag{8.20}$$

Складывая уравнения системы (8.20), получим

$$\sum_{i=1}^{q} x_i = \frac{q\lambda}{q-1}. \tag{8.21}$$

Вычтем из полученного соотношения i-е уравнение системы

$$x_i = (\frac{q}{q-1} - 1)\lambda = \frac{\lambda}{q-1},\ i = 1, \ldots, q. \tag{8.22}$$

Из (8.21) имеем

$$M = \frac{q\lambda}{q-1},\ \lambda = \frac{M(q-1)}{q}. \tag{8.23}$$

Подставляя (8.23) в выражение (8.22), получим координаты стационарной для $\Phi(x;\lambda)$ точки: $x = \left(\frac{M}{q}, \ldots, \frac{M}{q}\right)$. В этой точке значение квадратичной формы $F(x,x)$ равно $M^2\frac{q-1}{q}$.

Поскольку рассматриваемая область значений переменных является ограниченной и замкнутой, то экстремум $F(x,x)$ достигается либо на границе, либо в стационарной точке. На границе области по крайней мере одна из переменных обращается в нуль. Следовательно, мы оказываемся в условиях индуктивного предположения. При этом

$$\max_{\{x\}} F((x_1, \ldots, x_{q-1}), (x_1, \ldots, x_{q-1})) = \frac{M^2(q-2)}{q-1} < \frac{M^2(q-1)}{q}.$$

Следовательно,

$$\max_{\{x\}} F((x_1, \ldots, x_q), (x_1, \ldots, x_q)) = \frac{M^2(q-1)}{q}.$$

Так как число различных пар кодовых слов равно $M^2 - M$, то

$$d(M^2 - M) \leqslant D \leqslant \frac{M^2 n(q-1)}{q}.$$

□

Определение 8.18. *Коды, для которых нестрогое неравенство (8.17) обращается в равенство, называются эквидистантными.*

Обращение в равенство (8.17) означает, что расстояние между двумя любыми кодовыми словами одинаково. Это обстоятельство объясняет введенный выше термин.

Вырожденный пример такого кода — код кратных повторений. Другой простой пример — код $\Sigma_2 = \{(000),(110),(101),(011)\}$. Если отметить вершины единичного куба, соответствующие кодовым словам кода Σ_2 и соединить их, то построенная фигура будет представлять собой симплекс. Это и дает основание назвать код Σ_2 симплексным. Другие примеры эквивидистантных кодов — коды, построенные с использованием матриц Адамара, а также коды, составленные из последовательностей максимальной длины.

Теорема 8.8 (нижняя граница Варшамова – Гилберта). *Пусть натуральные числа $d \geqslant 2$, n и r таковы, что выполняется неравенство*

$$\sum_{j=0}^{d-2} \binom{n}{i} (q-1)^i < q^r. \tag{8.24}$$

Тогда существует q-ичный линейный код $\mathcal{C}$ длиной n, размерностью $k \geqslant n - r$ и с минимальным кодовым расстоянием $d_{\mathcal{C}} \geqslant d$.

Доказательство. Построим над полем $\mathbb{F}_q$ матрицу H размером $r \times n$, в которой любые $d-1$ столбцов линейно независимы, что эквивалентно построению кода с минимальным расстоянием, не меньшим, чем d. Для этого в качестве первого столбца матрицы H выберем произвольное ненулевое слово $h_1^{\downarrow}$ из r-мерного пространства $V_r(q)$. В качестве второго столбца $h_2^{\downarrow}$ возьмем произвольное слово из множества

$$V_r^1 = V_r(q) \backslash \{a h_1^{\downarrow}\},\ a \in \mathbb{F}_q.$$

Если в V_r^1 существует хотя бы одно слово, не являющееся линейной комбинацией $h_1^{\downarrow}$ и $h_2^{\downarrow}$, то выберем одно из таких слов в качестве $h_3^{\downarrow}$.

Предположим, что таким образом выбрано j слов: $h_1^{\downarrow}, \ldots, h_j^{\downarrow}$. По построению этих слов каждый столбец $h_i^{\downarrow}$, $1 \leqslant i \leqslant j$ не является линейной комбинацией никаких $d-2$ или менее столбцов из $h_1^{\downarrow}, \ldots, h_{i-1}^{\downarrow}$.

Поскольку i слов из построенных таким образом слов $h_1^{\downarrow}, \ldots, h_j^{\downarrow}$ можно выбрать $\binom{j}{i}$ способами, а число способов выбора i ненулевых коэффициентов линейной комбинации равно $(q-1)^i$, то число столбцов, являющихся линейными комбинациями $d-2$ или менее столбцов из $h_1^{\downarrow}, \ldots, h_j^{\downarrow}$, не больше $\sum_{i=1}^{d-2} \binom{j}{i}(q-1)^i$. Если выполняется неравенство

$$\sum_{i=1}^{d-2} \binom{j}{i} (q-1)^i < q^r - 1,$$

то среди всех $q^r - 1$ ненулевых слов длиной r существует такое слово, которое не может быть представлено никакой линейной комбинацией указанного вида. Это слово можно выбрать в качестве очередного $h_{j+1}^{\downarrow}$.

К тому моменту, когда очередное слово $h_{n+1}^{\downarrow}$ выбрать из множества $V_r(q)\backslash\{0\}$ в силу его конечности уже нельзя, общее число линейных комбинаций, порож-

денных ранее выбранными $h_1^{\downarrow}, \ldots, h_n^{\downarrow}$, удовлетворяет неравенству

$$\sum_{i=1}^{d-2} \binom{n}{i} (q-1)^i \geqslant q^r - 1. \tag{8.25}$$

Добавление нового слова $h_{n+1}^{\downarrow}$ к числу $h_1^{\downarrow}, \ldots, h_n^{\downarrow}$ приводит к тому, что число линейных комбинаций слов $h_1^{\downarrow}, \ldots, h_{n+1}^{\downarrow}$ впервые превзойдет $q^r - 1$.

Совокупность n слов $h_1^{\downarrow}, \ldots, h_n^{\downarrow}$ образует проверочную матрицу размерностью $r \times n$, в которой любые $d-1$ столбцов линейно независимы. Следовательно, построен код с минимальным кодовым расстоянием по крайней мере d. □

С целью получить асимптотическое представление границ при больших значениях n воспользуемся оценками Г. Чернова для биномиальных коэффициентов [8].

Теорема 8.9. *Пусть число t удовлетворяет неравенству $t < \dfrac{q-1}{q} n$. Тогда справедливы оценки*

$$\frac{q^{nH_q(\frac{t}{n})}}{n+1} \leqslant \binom{n}{t} (q-1)^t \leqslant \sum_{i=0}^{t} \binom{n}{t} (q-1)^i \leqslant q^{nH_q(\frac{t}{n})}, \tag{8.26}$$

где $H_q(x) = x \log_q(q-1) - x \log_q x - (1-x) \log_q(1-x)$.

Доказательство. Введем производящую функцию:

$$F(z) = (1 + (q-1)z)^n = \sum_{i=0}^{n} A_i z^i,\ A_i = \binom{n}{i} (q-1)^i. \tag{8.27}$$

Просуммируем коэффициенты по i:

$$\sum_{i=0}^{t} \binom{n}{i} (q-1)^i = \sum_{i=0}^{t} A_i.$$

При $0 < z \leqslant 1$, $i < t$ справедливо неравенство $z^{i-t} \geqslant 1$, используя которое получим

$$\sum_{i=0}^{t} A_i \leqslant \sum_{i=0}^{t} A_i z^{i-t} = z^{-t} \sum_{i=0}^{t} A_i z^t \leqslant z^{-t} F(z). \tag{8.28}$$

Покажем, что отношение $\dfrac{A_i}{A_{i+1}} = \dfrac{i+1}{(n-i)(q-1)}$ является монотонно возрастающей функцией по i. Дифференцируя по i это отношение, получим

$$\frac{\partial \frac{A_i}{A_{i+1}}}{\partial i} = \frac{n+1}{(q-1)(n-i)^2}. \tag{8.29}$$

При $1 \leqslant i < n$ производная положительна:

$$\frac{\partial \frac{A_i}{A_{i+1}}}{\partial i} > 0.$$

Поскольку по условию теоремы $t < \dfrac{(q-1)n}{q}$, то найдется такое число $\tilde{z}$, $0 \leqslant \leqslant \tilde{z} < 1$, при котором

$$\frac{A_{t-1}}{A_t} \leqslant \tilde{z} \leqslant \frac{A_t}{A_{t+1}},$$

тогда $\dfrac{A_{i-1}}{A_i} \leqslant \dfrac{A_{t-1}}{A_t} \leqslant \tilde{z}$ при $0 \leqslant i < t$ и $\tilde{z} \leqslant \dfrac{A_t}{A_{t+1}} \leqslant \dfrac{A_i}{A_{i+1}}$ при $t < i$.

Докажем, что при $0 \leqslant i \leqslant n$

$$A_t \tilde{z}^t \geqslant A_i \tilde{z}^i. \tag{8.30}$$

При $i = t$ нестрогое неравенство (8.30) обращается в равенство.

Пусть $i < t$. В силу того, что $\dfrac{A_i}{A_{i+1}}$ — монотонно возрастающая функция, для $i < t$ справедливо неравенство

$$\frac{A_i}{A_{i+1}} < \frac{A_{t-1}}{A_t} \leqslant \tilde{z}. \tag{8.31}$$

Рассматривая произведение $(t-i)$ неравенств вида (8.31), получим

$$\frac{A_i}{A_{i+1}} \frac{A_{i+1}}{A_{i+2}} \dots \frac{A_{t-1}}{A_t} = \frac{A_i}{A_t} \leqslant \tilde{z}^{t-i}.$$

Отсюда следует

$$A_i \tilde{z}^i \leqslant A_t \tilde{z}^t. \tag{8.32}$$

Пусть $i > t$. Рассмотрим произведение $(i-t)$ неравенств вида $\dfrac{A_i}{A_{i+1}} \geqslant \tilde{z}$:

$$\frac{A_t}{A_{t+1}} \frac{A_{t+1}}{A_{t+2}} \dots \frac{A_{i-1}}{A_i} = \frac{A_t}{A_i} \geqslant \tilde{z}^{i-t}.$$

Поэтому

$$A_t \tilde{z}^t \geqslant A_i \tilde{z}^i, \tag{8.33}$$

что совместно с неравенством (8.32) доказывает справедливость формулы (8.30). Тогда

$$F(\tilde{z}) = \sum_{i=0}^{n} A_i \tilde{z}^i \leqslant (n+1) A_t \tilde{z}^t. \tag{8.34}$$

Из равенств (8.28) и (8.34) имеем систему неравенств:

$$\min_{0<z\leqslant 1} \frac{z^{-t} F(z)}{n+1} \leqslant A_t < \sum_{i=0}^{t} A_i \leqslant \min_{0<z\leqslant 1} z^{-t} F(z). \tag{8.35}$$

Обозначим $\tau = \frac{t}{n}$ и подсчитаем стационарную точку функции $f(z) = z^{-n\tau} F(z) = = z^{-n\tau}(1 + (q-1)z)^n$. Вычислим производную введенной функции:

$$\frac{\partial f(z)}{\partial z} = -n\tau z^{-n\tau-1}(1 + (q-1)z)^n + n z^{-n\tau}(1 + (q-1)z)^{n-1}(q-1).$$

Из уравнения $\dfrac{\partial f(z)}{\partial z} = 0$ найдем

$$-\tau(1 + (q-1)z) + z(q-1) = 0.$$

Корень z_0 уравнения $\dfrac{\partial f(z)}{\partial z} = 0$ равен

$$z_0 = \frac{\tau}{(q-1)(q-\tau)}. \tag{8.36}$$

При $0 < z < z_0$ имеем производную $\dfrac{\partial f(z)}{\partial z} < 0$. При $z > z_0$ — производная $\dfrac{\partial f(z)}{\partial z} > 0$. Следовательно, корень z_0 является точкой минимума. Подстановкой проверяется, что для $t < \dfrac{(q-1)^n}{q}$ этот корень $z_0 < 1 \in [0{,}1)$.

Значение $f(z)$ в точке z_0 равно

$$f(z_0) = \tau^{-n\tau}(q-1)^{n\tau}(1-\tau)^{n\tau}(1-\tau)^{-n} = q^{nH_q(\tau)}. \tag{8.37}$$

Из равенства (8.37) и системы неравенств (8.35) получим требуемые оценки (8.26). □

Используя теорему 8.9, получим асимптотическое представление границ Хэмминга, Плоткина, Варшамова – Гилберта:

$$\frac{k}{n} \leqslant 1 - H_q(\frac{t}{n}),$$
$$\frac{k}{n} \leqslant 1 - \frac{qd-1}{n(q-1)},$$
$$\frac{k}{n} \geqslant 1 - H_q(\frac{d-2}{n}).$$

В этих формулах $H_q(x)$ — функция, введенная в теореме 8.9, совпадает с энтропией при $q = 2$.

8.5. О ПОДХОДАХ К ПОСТРОЕНИЮ ДРУГИХ КОДОВ

Рассмотрим некоторые подходы к построению других кодов [8].

1) *Добавление общей проверки на четность.* Пусть $\mathcal{C}$ двоичный (n, k) код с кодовым расстоянием d, в котором имеются кодовые слова нечетного веса. Для каждого слова $c = (\gamma_1, \ldots, \gamma_n)$ вычислим

$$\gamma_{n+1} = \bigoplus_{i=1}^{n} \gamma_n. \tag{8.38}$$

Рассмотрим новый код $\tilde{\mathcal{C}} = \{\tilde{c} : (\gamma_1, \ldots, \gamma_n, \gamma_{n+1})\}$, где γ_{n+1} задается соотношением (8.38). Если минимальное кодовое расстояние кода $\mathcal{C}$ определялось словом нечетного веса, то минимальное кодовое расстояние $\tilde{d}$ кода $\tilde{\mathcal{C}}$ будет равно $d + 1$.

Если код $\mathcal{C}$ имеет проверочную матрицу $H((n-k)\times n)$, то код $\tilde{\mathcal{C}}$ будет иметь следующую проверочную матрицу:

$$\tilde{H}=\begin{pmatrix} 1 & \dots & 1 & 1 \\ & & & 0 \\ & H & & \vdots \\ & & & 0 \end{pmatrix}. \tag{8.39}$$

2) *Выкалывание кодовых координат.* Из всех кодовых слов одновременно удаляется одна или более одноименных кодовых координат. Для полученного кода $\tilde{\mathcal{C}}$ при удалении одной координаты новые параметры будут следующим образом связаны с параметрами исходного кода $\tilde{n}=n-1$, $\tilde{k}=k$ либо $k-1$, $\tilde{d}=d$, либо $d-1$.

3) *Построение кода выбрасыванием слов.* Пусть $\mathcal{C}$ содержит слова четного и нечетного веса. Слова нечетного веса образуют класс смежности по множеству слов четного веса. Удалим из $\mathcal{C}$ все кодовые слова нечетного веса. Параметры нового кода $\tilde{\mathcal{C}}$ связаны с параметрами исходного кода: $\tilde{n}=n$, $\tilde{k}=k-1$, $\tilde{d}=d$, $\tilde{d}>d$.

4) *Пополнение кода путем добавления новых кодовых слов.* Предположим, что слово $e=(1,\dots,1)$ не принадлежит коду $\mathcal{C}$. Добавляя к коду $\mathcal{C}$ множество $\mathcal{C}+e$, получим новый код $\tilde{\mathcal{C}}$ со следующими параметрами: $\tilde{n}=n$, $\tilde{k}=k+1$, $\tilde{d}=\min\{d,n-d_1\}$, где d_1 — наибольший вес кодовых слов $\tilde{\mathcal{C}}$.

5) *Укорочение кода.* Выбираются все слова, первый символ которых равен нулю. Код $\tilde{\mathcal{C}}$ составляется из этих слов после удаления из них первого символа: $\tilde{n}=n-1$, $\tilde{k}=k-1$, $\tilde{d}\geqslant d$.

6) *Построение кода с помощью прямой суммы.* Пусть $u\in\mathcal{C}_1$, $v\in\mathcal{C}_2$, $\tilde{\mathcal{C}}=\{(u|v)\}$. Параметры кода $\tilde{\mathcal{C}}$ имеют вид $\tilde{n}=n_1+n_2$, $\tilde{k}=k_1+k_2$, $\tilde{d}=\min\{d_1,d_2\}$.

7) *Построение кода с помощью полупрямой суммы.* Пусть $u\in\mathcal{C}_1$, $v\in\mathcal{C}_2$, $n_1=n_2$, $\tilde{\mathcal{C}}=\{(u|u+v)\}$. Тогда $\tilde{n}=2n_1$.

Лемма 8.11. *Кодовое расстояние кода $\tilde{\mathcal{C}}$, построенного с помощью полупрямой суммы, выражается формулой $\tilde{d}=\min\{2d_1,d_2\}$.*

Доказательство. Пусть $a=(u_1|u_1+v_1)$, $b=(u_2|u_2+v_2)$ — различные кодовые слова в $\tilde{\mathcal{C}}$. Если $v_1=v_2$, то $d(a,b)=2d(u_1,u_2)\geqslant 2d_1$. Если $v_1\neq v_2$, то, учитывая неравенство треугольника, имеем

$$\begin{gathered}\rho(a,b)=w(u_1-u_2)+w(u_1-u_2+v_1-v_2)\geqslant\\ \geqslant w(u_1-u_2)+w(v_1-v_2)-w(u_1-u_2)=w(v_1-v_2)\geqslant d_2.\end{gathered}$$

□

8) *Произведение кодов.* Пусть $\mathcal{C}_1$ — систематический (n_1,k_1)-код, $\mathcal{C}_2$ — систематический (n_2,k_2)-код. Запишем $K=k_1k_2$ информационных символов, составляющих слово, подлежащее кодированию, в виде матрицы:

$$A=\begin{pmatrix} a_{11} & \dots & a_{1k_1} \\ \vdots & \ddots & \vdots \\ a_{k_21} & \dots & a_{k_2k_1} \end{pmatrix}. \tag{8.40}$$

Строки матрицы (8.40) закодируем кодом $\mathcal{C}_1$. В результате к каждой строке матрицы будет добавлено $n_1 - k_1$ проверочных символов, удовлетворяющих проверочным соотношениям кода $\mathcal{C}_1$:

$$A_{\mathcal{C}_1} = \begin{pmatrix} a_{11} & \dots & a_{1k_1} & a_{1k_1+1} & \dots & a_{1n_1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{k_2 1} & \dots & a_{k_2 k_1} & a_{k_2 k_1+1} & \dots & a_{k_2 n_1} \end{pmatrix}. \tag{8.41}$$

Далее каждый столбец получившейся матрицы $A_{\mathcal{C}_1}$ закодируем с помощью линейного систематического (n_2, k_2)-кода $\mathcal{C}_2$:

$$A_{\mathcal{C}_1 \times \mathcal{C}_2} = \begin{pmatrix} a_{11} & \dots & a_{1k_1} & \dots & a_{1n_1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{k_2 1} & \dots & a_{k_2 k_1} & \dots & a_{k_2 n_1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n_2 1} & \dots & a_{n_2 k_1} & \dots & a_{n_2 n_1} \end{pmatrix}. \tag{8.42}$$

Выписывая элементы этой матрицы по строчкам, получим кодовое слово, соответствующее исходному информационному слову.

Проводя эту процедуру для каждого информационного слова, построим линейный код $\tilde{\mathcal{C}} = \mathcal{C}_1 \times \mathcal{C}_2$, называемый произведением кодов. Очевидно, что $n = n_1 n_2$.

Лемма 8.12. *Минимальное кодовое расстояние кода, построенного по правилу произведения, имеет вид* $d \geqslant d_1 d_2$.

Доказательство. Слово, принадлежащее $\tilde{\mathcal{C}}$ и записанное в виде матрицы (8.42), должно иметь по крайней мере d_1 ненулевых элементов в каждой строке матрицы (8.42) и хотя бы d_2 ненулевых элементов в каждом столбце этой матрицы, если в строке или столбце есть ненулевые элементы. Следовательно, любое кодовое слово имеет не менее $d_1 d_2$ ненулевых элементов. □

8.6. ТОЖДЕСТВА МАК-УИЛЬЯМС

Рассмотрим (n, k)-код $\mathcal{C}$. Обозначим посредством A_i число таких кодовых слов, вес каждого из которых равен i. Числа A_i при $i = 0, 1, \dots, n$ неотрицательные, целые, удовлетворяют соотношениям

$$0 \leqslant A_i \leqslant n, \ \sum_{i=0}^{n} A_i = |\mathcal{C}|.$$

Определение 8.19. *Набор* $(A_0, A_1, \dots, A_n)$ *называется весовым спектром кода.*

Весовые спектры некоторых из рассмотренных выше кодов:

1) код кратных повторений, $A_0 = A_n = 1$ и $A_i = 0$, $i = 1, \dots, n-1$;

2) симплексный код, двойственный коду Хэмминга. Пусть длина кодового слова равна $n = 2^k - 1$. Тогда $A_0 = 1$, $A_{2^{k-1}} = 2^k - 2$, $A_i = 0$ при $i \neq 0, 2^{k-1}, 2^k$.

Весовой спектр является важной характеристикой корректирующих свойств кода. Через компоненты A_i, могут быть выражены: вероятность ошибки декодирования, вероятность непринятия решения и т. д. Нахождение весового спектра связано с подсчетом веса каждого кодового слова. В некоторых случаях нахождение спектра кода $\mathcal{C}$ можно осуществить, используя известный весовой спектр кода $\mathcal{C}^{\perp}$. В теореме Мак-Уильямс приведены соотношения, позволяющие осуществить подобный пересчет.

Лемма 8.13. *Обозначим* $U+V=\{w:\ w=\alpha u+\beta v\}$, *где* $u\in U$, $v\in V$, *а* α,β *— элементы поля. Тогда*

$$\dim(U\cap V)+\dim(U+V)=\dim U+\dim V. \tag{8.43}$$

Лемма 8.14. *Если* $U^{\perp}$ *— ортогональное пространство для* U, $V^{\perp}$ *— ортогональное пространство для* V, *то* $U^{\perp}\cap V^{\perp}$ *— ортогональное пространство для* $U+V$.

Теорема 8.10. *Пусть* $\mathcal{C}$ *—* (n,k)*-код над* $\mathbb{F}_q$, $\mathcal{C}^{\perp}$ *—* $(n,n-k)$*-код, двойственный к* $\mathcal{C}$, $(A_0,\ldots,A_n)$,$(B_0,\ldots,B_n)$ *— весовые спектры кодов соответственно. Тогда*

$$q^k\sum_{i=0}^{n}B_ix^i=\sum_{j=0}^{n}A_j(1-x)^j(1+(q-1)x)^{n-j}. \tag{8.44}$$

Доказательство. Теорему докажем согласно [8]. Сначала получим ряд тождеств, эквивалентных соотношению (8.44). Затем докажем справедливость одного из установленных тождеств. Соотношение (8.44) эквивалентно

$$\sum_{i=0}^{n}B_ix^i=q^{-k}\sum_{j=0}^{n}\sum_{s=0}^{j}\sum_{t=0}^{n-j}A_j\binom{j}{s}(-1)^sx^s\binom{n-j}{t}(q-1)^tx^t.$$

Чтобы исключить индекс t, положим $i=s+t$. Тогда $t=i-s$.

Заметим, что $\binom{n}{s}=0$, когда $s>n$, $s<0$. Это дает возможность расширить пределы суммирования у внутренних сумм, формально включив в суммы нулевые слагаемые:

$$\sum_{i=0}^{n}B_ix^i=q^{-k}\sum_{j=0}^{n}\sum_{s=0}^{j}\sum_{i=s}^{n-j-s}A_j\binom{j}{s}\binom{n-j}{i-s}(-1)^s(q-1)^{i-s}x^i=$$

$$=q^{-k}\sum_{j=0}^{n}\sum_{s=0}^{j}\sum_{i=0}^{n-}A_j\binom{j}{s}\binom{n-j}{i-s}(-1)^s(q-1)^{i-s}x^i.$$

Изменим порядок суммирования:

$$\sum_{i=0}^{n}B_ix^i=q^{-k}\sum_{i=0}^{n}x^i\left(\sum_{j=0}^{n}A_j\sum_{s=0}^{n}\binom{j}{s}\binom{n-j}{i-s}(-1)^s(q-1)^{i-s}\right).$$

Приравнивая коэффициенты при одинаковых степенях x, получим соотношение, эквивалентное (8.44):

$$B_i = q^{-k} \sum_{j=0}^{n} A_j \sum_{s=0}^{n} \binom{j}{s} \binom{n-j}{i-s} (-1)^s (q-1)^{i-s}.$$

Найдем третье соотношение. Положим, $x = 1 + y$. После подстановки новой переменной y в (8.44) имеем

$$q^k \sum_{i=0}^{n} B_i (1+y)^i = \sum_{j=0}^{n} A_j (-y)^j (q + (q-1)y)^{n-j}.$$

Сделаем ряд эквивалентных преобразований:

$$\sum_{i=0}^{n} B_i \sum_{m=0}^{i} \binom{i}{m} y^m = q^{-k} \sum_{j=0}^{n} A_j (-y)^j \sum_{k=0}^{n-j} \binom{n-j}{k} y^k (q-1)^k q^{n-j-k}.$$

Чтобы исключить индекс k, положим, что $k + j = m$. Тогда $k = m - j$.

$$\sum_{i=0}^{n} B_i \sum_{m=0}^{i} \binom{i}{m} y^m = q^{-k} \sum_{j=0}^{n} \sum_{m=j}^{n} A_j \binom{n-j}{m-j} (-1)^j (q-1)^{m-j} q^{n-m} y^m.$$

Расширяя пределы суммирования и изменяя порядок суммирования, получим

$$\sum_{m=0}^{n} \sum_{i=0}^{n} B_i \binom{i}{m} y^m = q^{-k} \sum_{m=0}^{n} \left(\sum_{j=0}^{n} A_j \binom{n-j}{m-j} (-1)^j (q-1)^{m-j} q^{n-m} \right) y^m.$$

Приравняем коэффициенты при одинаковых степенях y:

$$\sum_{i=0}^{n} B_i \binom{i}{m} = q^{n-m-k} \sum_{j=0}^{n} A_j \binom{n-j}{m-j} (-1)^j (q-1)^{m-j}. \tag{8.45}$$

Наконец, полагая $x = (1+y)^{-1}$ в соотношении (8.44), имеем

$$q^k \sum_{i=0}^{n} B_i \frac{1}{(1+y)^i} = \sum_{j=0}^{n} A_j (1 - \frac{1}{1+y})^j (1 + \frac{q-1}{q+y})^{n-j},$$

$$q^k \sum_{i=0}^{n} B_i (1+y)^{n-i} = \sum_{j=0}^{n} A_j y^j (y+q)^{n-j},$$

$$q^k \sum_{i=0}^{n} B_i \sum_{m=0}^{n-i} \binom{n-i}{m} y^m = \sum_{j=0}^{n} A_j y^j \sum_{k=0}^{n-j} \binom{n-j}{k} y^k q^{n-j-k}.$$

Чтобы исключить индекс k, положим, $k + j = m$. Тогда

$$\sum_{i=0}^{n} B_i \sum_{m=0}^{n-i} \binom{n-i}{m} y^m = q^{-k} \sum_{j=0}^{n} \sum_{m=j}^{n} A_j \binom{n-j}{m-j} y^m q^{n-m}.$$

Расширяя пределы суммирования и изменяя порядок суммирования, придем к равенству

$$\sum_{m=0}^{n}\left(\sum_{i=0}^{n}B_i\binom{n-i}{m}\right)y^m=\sum_{m=0}^{n}\left(q^{n-m-k}\sum_{j=0}^{n}A_j\binom{n-j}{m-j}\right)y^m.$$

Замечая, что $\binom{n-j}{m-j}=\binom{n-j}{n-m}$, приравнивая коэффициенты при одинаковых степенях y, получим

$$\sum_{i=0}^{n}B_i\binom{n-i}{m}=q^{n-k-m}\sum_{j=0}^{n}A_j\binom{n-j}{m-j}. \tag{8.46}$$

Перейдем ко второй части теоремы и докажем тождество (8.46).

Введем в рассмотрение n-множество $N=(1,2,\ldots,n)$ и два его подмножества: m-множество $S=(s_1,\ldots,s_m)$ и $n-m$-множество $T=(t_1,\ldots,t_{n-m})$, $S\cup T=N$, $S\cap T=\emptyset$, $1\leqslant s_i\leqslant n$, $1\leqslant t_j\leqslant n$.

Обозначим F_s — подпространство, составленное из всех слов V_n, символы которых на позициях $s_1,\ldots,s_m$ могут быть ненулевыми, а символы на позициях $t_1,\ldots,t_{n-m}$ — обязательно нулевые. Аналогично F_t — подпространство, составленное из всех слов V_n, символы которых на позициях $t_1,\ldots,t_{n-m}$ могут быть ненулевыми, а символы на позициях $s_1,\ldots,s_m$ — обязательно нулевые. Тогда F_t — ортогональное пространство для F_s и наоборот.

Введем в рассмотрение пространства $G\cap F_s$ и $G^\perp\cap F_t$. По лемме 8.14 пространством, ортогональным $G\cap F_s$, является $G^\perp+F_t$. Обозначим $\dim(G\cap F_s)=d_s$, и $\dim(G^\perp\cap F_t)=d_t$. Поскольку $(G\cap F_s)^\perp=G^\perp\cap F_t$, то $\dim\left(G^\perp\cap F_t\right)=n-d_s$. Из леммы 8.13

$$\dim(G^\perp+F_t)+\dim(G^\perp\cap F_t)=\dim(G^\perp)+\dim(F_t).$$

Подставляя в это соотношение значения размерностей соответствующих пространств, получим

$$d_t=d_s+n-k-m. \tag{8.47}$$

Рассмотрим пару, образованную m-множеством $S=(s_1,\ldots,s_m)$ и словом $v\in G\cap F_s$. Если v — произвольное слово $G\cap F_s$, то в множестве $\{(S,v),\ v\in G\cap F_s\}$ содержится q^{d_s} пар. Если S меняется произвольно, то общее число таких пар равно $\sum_{\{S\}}q^{d_s}$.

Пусть слово $u\in G$ имеет вес j. Среди $n-j$ нулевых символов выберем $n-m$, $m\geqslant j$. Они определят некоторое $(n-m)$-множество $T=(t_1,\ldots,t_{n-m})$. Задание множества T однозначно определяет множество S. Так как множество T мы можем выбрать $\binom{n-j}{n-m}$ способами, то в силу взаимной однозначности таким же будет число множеств S при фиксированном m. Следовательно,

$$\sum_{\{S\}}q^{d_s}=\sum_{j=0}^{n}A_j\binom{n-j}{n-m}. \tag{8.48}$$

Подобным же образом рассматривая $G^{\perp}$ и множество T, придем к соотношению

$$\sum_{\{T\}} q^{d_t} = \sum_{i=0}^{n} B_i \binom{n-i}{m}. \tag{8.49}$$

С учетом (8.47) и того обстоятельства, что каждое множество T определяет соответствующее множество S, имеем

$$\sum_{\{T\}} q^{d_t} = \sum_{\{T\}} q^{d_s+n-k-m} = \sum_{\{S\}} q^{d_s+n-k-m} = q^{n-k-m} \sum_{\{S\}} q^{d_s}.$$

Используя формулы (8.48) и (8.49), получим

$$\sum_{i=0}^{n} B_i \binom{n-i}{m} = q^{n-k-m} \sum_{j=0}^{n} A_j \binom{n-j}{n-m}.$$

Из последнего соотношения заключим

$$q^k \sum_{i=0}^{n} B_i (1+y)^{n-i} = \sum_{j=0}^{n} A_j y^j (1+y)^{n-j}.$$

При подстановке $y = \dfrac{1-x}{x}$ придем к соотношению (8.44). □

8.7. ЦИКЛИЧЕСКИЕ КОДЫ

В классе линейных кодов, в свою очередь, выделяется важный подкласс циклических кодов, для которых алгоритмы кодирования и декодирования основываются на теории конечных полей и многочленов над конечными полями.

Определение 8.20. *Линейный q-ичный (n,k)-код $\mathcal{C}$ называется циклическим, если для любого кодового слова $c = c_0 c_1 \dots c_{n-1}$ его циклический сдвиг $c' = c_{n-1} c_0 \dots c_{n-2}$ также принадлежит коду $\mathcal{C}$.*

Существует взаимно однозначное соответствие между словами из V_n и многочленами над $\mathbb{F}_q$ степени не выше $n-1$:

$$c = c_0 c_1 \dots c_{n-1} \leftrightarrow c_0 + c_1 x + \dots + c_{n-1} x^{n-1}.$$

Обозначим: $\mathbb{F}_q[x]$ — кольцо многочленов над полем $\mathbb{F}_q$; $R_n = \mathbb{F}_q[x]/(x^n - 1)$ — факторкольцо, элементами которого являются классы многочленов, сравнимых по модулю $x^n - 1$. Элементы факторкольца R_n будем обозначать $[c(x)]_{x^n-1}$, или кратко $[c(x)]$. В каждом классе $[c(x)]$ содержится многочлен степени не выше $n-1$, и любые два различных многочлена степени не выше $n-1$ лежат в разных классах факторкольца R_n. Поэтому в качестве системы представителей классов из R_n можно взять множество всех различных многочленов степени не выше $n-1$:

$$R_n = \{[c_0 + c_1 x + \dots + c_{n-1} x^{n-1}] : \ c_i \in \mathbb{F}_q\}.$$

Таким образом, имеется взаимно однозначное соответствие между словами из V_n и классами многочленов из R_n:

$$\mathcal{C} \ni c \leftrightarrow [c(x)] \in R_n.$$

Следовательно, $|R_n| = |V_n| = q^n$. Нетрудно также показать, что R_n с операцией сложения и операцией умножения на элементы поля $\mathbb{F}_q$ представляет собой линейное пространство, изоморфное пространству V_n.

Определение 8.21. *Если $\mathcal{C}$ — произвольное подмножество пространства V_n, то соответствующее подмножество классов из R_n будем обозначать через $[\mathcal{C}]$.*

Очевидно, что если $\mathcal{C}$ — линейный код, то множество $[\mathcal{C}]$ с операцией сложения и операцией умножения на элементы поля $\mathbb{F}_q$ представляет собой линейное пространство, изоморфное $\mathcal{C}$.

Лемма 8.15. *Если $c = c_0c_1 \dots c_{n-1} \in V_n$ и c' — циклический сдвиг слова c, то*

$$[c'(x)] = [x][c(x)].$$

Доказательство. Поскольку $[x^n - 1] = [0]$, $[x^n] = [1]$, имеем

$$\begin{aligned}[x][c(x)] &= [x][c_0 + c_1x + \dots + c_{n-1}x^{n-1}] = \\ &= [c_0x + c_1x^2 + \dots + c_{n-2}x^{n-1} + c_{n-1}x^n] = \\ &= [c_{n-1} + c_0x + c_1x^2 + \dots + c_{n-2}x^{n-1}] = [c'(x)].\end{aligned}$$

□

Лемма 8.16. *Если $\mathcal{C}$ — циклический код, $c = c_0c_1 \dots c_{n-1}$ — кодовое слово кода $\mathcal{C}$ и $f(x)$ — произвольный многочлен над $\mathbb{F}_q$, то класс $[f(x)][c(x)]$ принадлежит множеству $[\mathcal{C}]$.*

Доказательство. Если $f(x) = f_0 + f_1x + \dots + f_tx^t$, то

$$[f(x)][c(x)] = \sum_{i=0}^{t} [f_i][x^i][c(x)].$$

Из определения циклического кода и леммы 8.15 по индукции следует, что для любого $i \geqslant 0$ класс $[x^i][c(x)]$ принадлежит множеству $[\mathcal{C}]$. В силу изоморфизма линейных пространств $\mathcal{C}$ и $[\mathcal{C}]$ каждое слагаемое $[f_i][x^i][c(x)]$ лежит в $[\mathcal{C}]$, и поэтому вся сумма лежит в $[\mathcal{C}]$. □

Лемма 8.17. *Пусть $\mathcal{C}$ — циклический код и $g(x)$ — ненулевой многочлен наименьшей степени, для которого $[g(x)] \in [\mathcal{C}]$. Тогда идеал $[\mathcal{C}]$ порождается классом $[g(x)]$ в R_n, а многочлен $g(x)$ делит $x^n - 1$.*

Доказательство. Ясно, что идеал, порожденный классом $[g(x)]$ в факторкольце R_n, является подмножеством множества $[\mathcal{C}]$. Чтобы доказать обратное включение, рассмотрим произвольный класс $[c(x)] \in [\mathcal{C}]$ и разделим $c(x)$ на $g(x)$ с остатком:

$$c(x) = g(x)a(x) + r(x), \ \deg r(x) < \deg g(x).$$

Тогда $[r(x)] = [c(x) - g(x)a(x)] = [c(x)] - [g(x)][a(x)] \in [\mathcal{C}]$, а так как $g(x)$ — ненулевой многочлен наименьшей степени с условием $[g(x)] \in [\mathcal{C}]$, получим, что $r(x) = 0$, т. е. $g(x)$ делит $c(x)$.

Делимость $x^n - 1$ на $g(x)$ доказывается аналогично — достаточно разделить $x^n - 1$ на $g(x)$ с остатком и воспользоваться определением многочлена $g(x)$. $\square$

Доказанное утверждение означает, что любой идеал в факторкольце R_n является главным идеалом, т. е. он порождается некоторым классом $[g(x)]$. Очевидно, многочлен $g(x)$ при этом можно выбрать унитарным, т. е. со старшим коэффициентом, равным 1.

Идеал, порождаемый классом $[g(x)] \in R_n$, обозначим $< [g(x)] >$.

Определение 8.22. *Пусть $\mathcal{C}$ — циклический код и $g(x)$ — унитарный многочлен наименьшей степени, для которого $< [g(x)] >= [\mathcal{C}]$. Тогда $g(x)$ — порождающий многочлен кода $\mathcal{C}$.*

Порождающий многочлен циклического кода определяется однозначно. Действительно, если $g^{'}(x)$ и $g^{''}(x)$ — два порождающих многочлена, то эти многочлены взаимно делят друг друга, но так как они оба унитарные, $g^{'}(x) = g^{''}(x)$.

Теорема 8.11. *Пусть $\mathcal{C}$ — циклический (n, k)-код и $g(x)$ — его порождающий многочлен. Тогда* $\deg g(x) = n - k$.

Доказательство. Обозначим $\deg g(x) = l$. Если $[c(x)] = [c_0 + c_1 x + \ldots + c_{n-1}x^{n-1}]$ — произвольный элемент идеала $[\mathcal{C}]$, то $g(x)$ делит $c(x)$, т. е. найдется такой многочлен $f(x) = f_0 + f_1 x + \ldots + f_{n-l-1}x^{n-l-1}$, что $c(x) = f(x)g(x)$, или

$$[c(x)] = \sum_{i=0}^{n-l-1} [f_i][x^i g(x)].$$

Таким образом, любой класс $[c(x)] \in [\mathcal{C}]$ представляется в виде линейной комбинации $n-l$ классов $[g(x)], [xg(x)], \ldots, [x^{n-l-1}]g(x)$. Эти классы линейно независимы, поскольку никакой многочлен $x^i g(x)$ не может быть линейно выражен через многочлены $g(x), xg(x), \ldots, x^{i-1}g(x)$ меньшей степени. Следовательно, размерность $[\mathcal{C}]$ как линейного пространства над $\mathbb{F}_q$ равна $n-k$. Линейные пространства $\mathcal{C}$ и $[\mathcal{C}]$ изоморфны, $\dim\mathcal{C} = k$, поэтому $n - l = k$, откуда $\deg g(x) = l = n - k$. $\square$

Следствие 8.1. *Пусть $\mathcal{C}$ — циклический (n, k)-код с $k \geqslant 1$ и порождающим многочленом $g(x) = g_0 + g_1 x + \ldots + g_{n-k}x^{n-k}$. Тогда матрица*

$$G = \begin{pmatrix} g_0 & g_1 & \ldots & \ldots & g_{n-k} & 0 & 0 & \ldots & 0 & 0 \\ 0 & g_0 & \ldots & \ldots & g_{n-k-1} & g_{n-k} & 0 & \ldots & 0 & 0 \\ \ldots & & & & & & & & & \\ 0 & 0 & \ldots & g_0 & \ldots & \ldots & \ldots & \ldots & g_{n-k-1} & g_{n-k} \end{pmatrix}$$

размером $k \times n$ является порождающей матрицей кода $\mathcal{C}$.

Доказательство. Из доказательства теоремы 8.11 видно, что множество из k классов $[g(x)], [xg(x)], \ldots, [x^{k-1}g(x)]$ — базис линейного пространства $[\mathcal{C}]$. Следовательно, соответствующие этим классам кодовые слова

$$g_0 g_1 \ldots g_{n-k} 0 \ldots 0, 0 g_0 g_1 \ldots g_{n-k} 0 \ldots 0, \ \ldots, 0 \ldots 0 g_0 g_1 \ldots g_{n-k}$$

образуют базис кода $\mathcal{C}$. $\square$

Определение 8.23. *Пусть* $\mathcal{C}$ *— циклический код с порождающим многочленом* $g(x)$*. Тогда многочлен*

$$h(x) = \frac{x^n - 1}{g(x)}$$

называется проверочным многочленом кода $\mathcal{C}$*.*

Теорема 8.12. *Пусть* $\mathcal{C}$ *— циклический код с проверочным многочленом* $h(x)$*. Класс* $[c(x)]$ *из факторкольца* R_n *принадлежит идеалу* $[\mathcal{C}]$ *тогда и только тогда, когда* $[c(x)h(x)] = [0]$*.*

Доказательство. Если $[c(x)] \in [\mathcal{C}]$, то $c(x)$ делится на порождающий многочлен $g(x)$, т. е. $c(x) = f(x)g(x)$. Следовательно,

$$[c(x)h(x)] = [f(x)g(x)\frac{x^n - 1}{g(x)}][f(x)(x^n - 1)] = [0].$$

С другой стороны, если $[c(x)h(x)] = [0]$, то $x^n - 1$ делит $c(x)h(x)$. Но $x^n - 1 = g(x)h(x)$, поэтому $g(x)$ делит $c(x)$, так что $[c(x)] \in [\mathcal{C}]$. □

Из теоремы 8.12 видно, что проверочный многочлен для циклического кода играет такую же роль, какую играет проверочная матрица в более общем случае линейного кода. Из теоремы 8.11 следует, что проверочный многочлен циклического (n, k)-кода имеет степень k.

Теорема 8.13. *Пусть* $\mathcal{C}$ *— циклический* (n, k)*-код с проверочным многочленом* $h(x) = h_0 + h_1x + \ldots + h_kx^k$*. Тогда двойственный код* $\mathcal{C}^\perp$ *является циклическим* $(n, n - k)$*-кодом с порождающим многочленом* $g^\perp(x) = h_0^{-1}x^k h(\frac{1}{x}) = h_0^{-1}(h_k + h_{k-1}x + \ldots + h_0x^k)$ [33].

Следствие 8.2. *Пусть* $\mathcal{C}$ *— циклический* (n, k)*-код с* $k < n$ *и проверочным многочленом* $h(x) = h_0 + h_1x + \ldots + h_kx^k$*. Тогда матрица*

$$G = \begin{pmatrix} h_k & h_{k-1} & \ldots & \ldots & h_0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & h_k & \ldots & \ldots & h_1 & h_0 & 0 & \ldots & 0 & 0 \\ \ldots & & & & & & & & & \\ 0 & 0 & \ldots & h_k & \ldots & \ldots & \ldots & \ldots & h_1 & h_0 \end{pmatrix}$$

размером $(n - k) \times n$ *является проверочной матрицей кода* $\mathcal{C}$ [33].

Приведем *алгоритм кодирования* для систематического циклического кода. С этой целью введем информационный многочлен

$$a(x) = c_{n-k} + c_{n-k+1}x + \ldots + c_{n-1}x^{k-1}$$

и корректирующий многочлен

$$r(x) = c_0 + c_1x + \ldots + c_{n-k-1}x^{n-k-1}.$$

Тогда кодовый многочлен $c(x) = c_0 + c_1x + \ldots + c_{n-1}x^{n-1}$ представляется в виде

$$c(x) = r(x) + x^{n-k}a(x).$$

Поскольку порождающий многочлен $g(x)$ делит $c(x)$, $c(x) = g(x)f(x)$, получим

$$x^{n-k}a(x) = g(x)f(x) - r(x),$$

из которого следует правило кодирования: чтобы по заданному информационному многочлену $a(x)$ найти кодовый многочлен $c(x)$, нужно найти корректирующий многочлен $r(x)$, представляющий собой взятый со знаком «минус» остаток от деления многочлена $x^{n-k}a(x)$ на порождающий многочлен $g(x)$. Далее остаток от деления произвольного многочлена $t(x)$ на многочлен $g(x)$ будем обозначать в виде $t_g(x)$.

Предположим, что c — переданное в канал связи кодовое слово циклического кода $\mathcal{C}$, e — вектор ошибок и $y = c + e$ — принятое слово. Этим словам соответствуют классы $[c(x)]$, $[e(x)]$ и $[y(x)]$ из факторкольца R_n. Вначале декодер вычисляет остаток $y_g(x)$ (очевидно, что остатки $y_g(x)$ и $e_g(x)$ равны). Если остаток $y_g(x)$ равен нулю, то слово y принадлежит коду $\mathcal{C}$, декодер не обнаруживает ошибок и в качестве результата декодирования выдает $\mathfrak{D}_\rho(y) = y$. Если же $y_g(x) \neq 0$, то $y \notin \mathcal{C}$, декодер обнаруживает наличие ошибок в канале связи и пытается их исправить. Для этого ищется многочлен $e^{'}(x)$ с наименьшим числом ненулевых членов, удовлетворяющий условию $y_g(x) = e^{'}_g(x)$, и результатом декодирования объявляется слово, соответствующий многочлену $y(x) - e^{'}(x)$.

Цифровые логические устройства легко организовать в виде *цепей регистров сдвига*, имитирующих циклические сдвиги и полиномиальную арифметику, используемые в описании циклических кодов, поэтому структура циклических кодов тесно связана со структурой цепей регистров сдвига. В частности, эти цепи используются для реализации процедур кодирования и декодирования [3].

Регистр сдвига представляет собой последовательность элементов памяти, называемых разрядами, каждый из которых содержит один элемент поля $\mathbb{F}_q$. Содержащийся в каждом разряде символ, покидая его, появляется на выходящей из него линии. Каждый разряд снабжен и входящей линией, по которой в него поступает элемент поля $\mathbb{F}_q$. Также используются такие элементы, как умножитель на скаляр, сумматор и умножитель. Умножитель на скаляр является функцией одной входной переменной, он умножает входную переменную на фиксированный элемент поля $\mathbb{F}_q$. Сумматор и умножитель являются функциями двух входных переменных, принимающих значения из $\mathbb{F}_q$.

Регистры сдвига можно использовать для умножения и деления многочленов над $\mathbb{F}_q$, этим и объясняется их частое применение в конструкциях кодеров и декодеров. Цепи регистров сдвига называются также фильтрами.

Пример 8.3. Пусть $n = 7$, $q = 2$. Разложение многочлена $x^7 - 1 = x^7 + 1$ на неприводимые над полем $\mathbb{F}_2$ множители имеет вид

$$x^7 + 1 = (x + 1)(x^3 + x + 1)(x^3 + x^2 + 1).$$

Поскольку порождающий многочлен циклического кода должен делить многочлен $x^7 + 1$, имеется всего $2^3 = 8$ двоичных циклических кодов длиной 7, из которых два кода $(7, 0)$ и $(7, 7)$-коды тривиальны. Рассмотрим циклический код $\mathcal{C}$ с порождающим многочленом $g(x) = 1 + x^2 + x^3$. Для этого кода размерность k равна 4, корректирующий многочлен равен

$$x^7 + 1 = (x + 1)(1 + x + x^3) = 1 + x^2 + x^3 + x^4,$$

а в качестве порождающей и проверочной матрицы можно взять

$$G = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}, \; H = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{pmatrix}.$$

По теореме 8.7 минимальное кодовое расстояние равно $d = 3$. Действительно, все столбцы матрицы H различны, так что любые два столбца линейно независимы, и при этом имеются три линейно зависимых столбца (например, первый, третий и четвертый). Следовательно, декодер $\mathfrak{D}_\rho$ правильно исправляет любые одиночные ошибки.

Все возможные многочлены $e(x)$, соответствующие одиночным ошибкам, и остатки от деления этих многочленов на порождающий многочлен $g(x)$ представлены в табл. 8.1

Таблица 8.1

$e(x)$	$e_g(x)$
1	1
x	x
x^2	x^2
x^3	$1 + x^2$
x^4	$1 + x + x^2$
x^5	$1 + x$
x^6	$x + x^2$

Закодируем информационное слово $a = 1110$. Для этого информационный многочлен $1 + x + x^2$ умножим на x^3, результат $x^3 + x^4 + x^5$ поделим с остатком на $g(x)$ и получим неполное частное $f(x) = 1 + x^2$ и остаток $r(x) = 1$. Кодовому многочлену $c(x) = x^3 a(x) + r(x) = 1x^3 + x^4 + x^5$ отвечает кодовое слово $c = 1001110$.

Далее предположим, что при передаче по каналу связи произошло искажение четвертого символа кодового слова. Другими словами, вектор ошибок равен $e = 0001000$, а многочлен ошибок равен $e(x) = x^3$. Принятому слову $y = 1000110$ отвечает многочлен $y(x) = 1 + x^4 + x^5$, который не делится на $g(x)$, так что декодер обнаруживает наличие ошибок и пытается их исправить. Для этого вычисляется остаток $y_g(x) = 1 + x^2$ и по таблице находится соответствующий этому остатку многочлен ошибок x^3, т. е. происходит правильное декодирование.

8.8. ЧАСТНЫЕ СЛУЧАИ ЛИНЕЙНЫХ КОДОВ

Рассмотрим метод построения двоичного кода, исправляющего одну ошибку. Синдром принятого слова y равен линейной комбинации тех столбцов проверочной матрицы H, номера которых совпадают с номерами искаженных символов, а коэффициенты линейной комбинации равны величинам ошибок. Проверочная

матрица H кода, исправляющего одну ошибку, должна удовлетворять двум следующим ограничениям. Во-первых, матрица H не должна иметь нулевых столбцов. В противном случае ошибка в соответствующем символе не будет влиять на синдром и не будет обнаружена. Во-вторых, все столбцы матрицы H должны быть различными. В противном случае ошибки в соответствующих позициях не будут различаться.

Зададимся числом проверочных символов $r = n - k$, совпадающим с числом строк проверочной матрицы H_r. Чтобы построить код с максимально возможной скоростью передачи, возьмем все допустимые ненулевые двоичные слова из V_r.

Для построения матрицы H_r условимся выписывать их в порядке возрастания чисел, двоичные разложения которых совпадают с рассматриваемыми столбцами. Для $r = 3$ имеем $2^3 - 1 = 7$ допустимых двоичных столбцов, которые согласно договоренности образуют матрицу H_r следующего вида:

$$H_3 = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

Используя строение матрицы H_r, покажем индуктивный метод построения проверочных матриц:

$$H_2 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix},$$

$$H_3 = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ & H_2 & & 0 & & H_2 & \\ & & & 0 & & & \end{pmatrix},$$

$$H_r = \begin{pmatrix} 0 & \cdots & 0 & 1 & 1 & \cdots & 1 \\ & & & 0 & & & \\ & H_{r-1} & & \vdots & & H_{r-1} & \\ & & & 0 & & & \end{pmatrix}.$$

Определение 8.24. *Двоичный код с длиной блока $n = 2^r - 1$, проверочной матрицей H_r $(r \times n)$, образованной всеми ненулевыми словами из $V_r(2)$, расположенными в порядке возрастания чисел, двоичные разложения которых совпадают с рассматриваемыми столбцами, называется двоичным кодом Хэмминга.*

Введенный $(n = 2^r - 1, k = 2^r - 1 - r)$-код будем обозначать $\mathcal{H}_r$. Рассмотрим процедуру декодирования для *кодов Хэмминга*. Пусть произошла ошибка в l-м символе передаваемого кодового слова c. Подсчитаем синдром полученного слова $y = c + e_l$, $e_l = \underbrace{0 \ldots 01}_{l} 0 \ldots 0$:

$$s(y) = Hy' = H(c + e_l)' = He_l' = h_l^{\downarrow}. \tag{8.50}$$

Следовательно, если произошла одна ошибка, то синдром $s(y)$ в двоичной записи указывает номер столбца, в котором произошла ошибка. Соотношение (8.50) задает изящный и очень простой способ декодирования. Из свойств кода

исправлять одну ошибку следует, что кодовое расстояние $d = 3$. Заметим, что слово $1110\ldots0$ принадлежит $\mathcal{H}_r$, что проверяется умножением на проверочную матрицу. Следовательно, $d = 3$.

Теорема 8.14. *Двоичный код $\mathcal{H}_r$ является совершенным.*

Доказательство. Поскольку код $\mathcal{H}_r$ исправляет одну ошибку, то множества $D(c) = \{\tilde{c} :\ d(\tilde{c}, c) \leqslant 1,\ c \in \mathcal{H}_r,\ \tilde{c} \in V_n\}$ не пересекаются, $|D(c)| = n + 1 = 2^r$, $|\mathcal{H}_r| = 2^{2^r-1-r}$.

Число слов пространства V_n, попавших в множества $D(g)$ для $\forall c \in \mathcal{H}_r$, равно 2^{2^r-1}. Это число совпадает с $|V_r(2)|$. Поэтому каждое слово длиной n находится в одном из таких множеств и, следовательно, код $\mathcal{H}_r$ — совершенный. □

Выше были указаны тривиальный код кратных повторений и код Y_r, являющиеся совершенными. Совершенны также q-ичные коды, исправляющие одну ошибку и имеющие длину $n = (q^m - 1)/(q - 1)$, $q = p^d$, p — простое. Это q-ичные коды Хэмминга, нелинейные коды Васильева, коды Шенгейма, двоичный код Голея, исправляющий тройные ошибки, и троичный код Голея, исправляющий двойные ошибки. Известен результат Титвайнена, заключающийся в следующем. Если q — степень простого числа, то не существует других q-ичных совершенных кодов, за исключением перечисленных, комбинаторно эквивалентных им, и тривиальных нелинейных кодов, получающихся сложением каждого кодового слова указанных выше линейных совершенных кодов с некоторым фиксированным словом длиной n.

Определение 8.25. *Двоичным симплексным кодом Σ_r называют код, двойственный коду Хэмминга $\mathcal{H}_r$.*

По определению двойственного кода порождающая матрица G_r кода $\mathcal{H}_r$ — проверочная для кода Σ_r, проверочная матрица H_r кода $\mathcal{H}_r$ — порождающая для Σ_r. Индуктивный метод построения матрицы H_r позволяет перечислять все кодовые слова кода Σ_r:

$$\Sigma_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \Sigma_r = \begin{pmatrix} & 0 & \\ \Sigma_{r-1} & \vdots & \Sigma_{r-1} \\ & 0 & \\ & 1 & \\ \Sigma_{r-1} & \vdots & \bar{\Sigma}_{r-1} \\ & 1 & \end{pmatrix}.$$

Каждая пара кодовых слов кода Σ_r находится друг от друга на одинаковом расстоянии, т. е. Σ_r — эквидистантный код. В коде Σ_r все слова, за исключением нулевого, имеют одинаковый вес, т. е. это равновесный код.

Коды Рида – Маллера представляют собой класс линейных кодов над $\mathbb{F}_2$. Определим этот код через порождающую матрицу, которую будем строить в удобной для декодирования несистематической форме. Прежде всего определим посимвольное произведение двух слов a и b.

Определение 8.26. *Если* $a = a_0, \ldots, a_{n-1}$, $b = b_0, \ldots, b_{n-1}$, *то их произведение равно слову* $ab = a_0b_0, \ldots, a_{n-1}b_{n-1}$.

Определение 8.27. *Порождающая матрица кода Рида – Маллера* r*-го порядка длиной* 2^m *определяется как совокупность блоков*

$$G = \begin{pmatrix} G_0 \\ G_1 \\ \vdots \\ G_r \end{pmatrix},$$

где G_0 *— слово длиной* $n = 2^m$, *состоящее из одних единиц;* G_1 *—* $(m \times 2^m)$*–матрица, содержащая в качестве столбцов все двоичные слова длиной* m*; строки матрицы* G_l, $2 \leqslant l \leqslant r$ *получаются из строк матрицы* G_1 *как всевозможные посимвольные произведения* l *строк из* G_1.

Поскольку существует всего $\binom{m}{l}$ способов выбора l строк, входящих в произведение, то матрица G_l имеет размер $\binom{m}{l} \times 2^m$. Ясно, что для кода Рида – Маллера порядка r справедливо

$$k = 1 + \binom{m}{1} + \ldots + \binom{m}{r},$$
$$n - k = 1 + \binom{m}{1} + \ldots + \binom{m}{m-r-1},$$

что обеспечивается линейной независимостью строк в матрице G. Код Рида – Маллера нулевого порядка является $(n, 1)$-кодом. Это просто код с повторением, который тривиально декодируется с помощью мажоритарного метода. Минимальное расстояние такого кода равно 2^m.

Определение 8.28. *Кодом Рида – Маллера первого порядка называется множество аффинных булевых функций от* n *переменных.*

С кодами Рида – Маллера связана такая важная характеристика булевых функций, как нелинейность.

Пример 8.4. В качестве примера положим $m = 4$, $n = 16$, $r = 3$. Тогда

$$G_0 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} a_0 \end{pmatrix},$$
$$G_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix}.$$

Поскольку G_1 имеет 4 строки, то матрица G_2 состоит $\binom{4}{2}$ строк:

$$G_2 = \begin{pmatrix} 0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1\\ 0&0&0&0&0&0&0&0&0&0&1&1&0&0&1&1\\ 0&0&0&0&0&0&0&0&0&1&0&1&0&1&0&1\\ 0&0&0&0&0&0&1&1&0&0&0&0&0&0&1&1\\ 0&0&0&0&0&1&0&1&0&0&0&0&0&1&0&1\\ 0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1 \end{pmatrix} = \begin{pmatrix} a_1a_2\\ a_1a_3\\ a_1a_4\\ a_2a_3\\ a_2a_4\\ a_3a_4 \end{pmatrix},$$

а матрица G_3 — $\binom{4}{3}$ строк:

$$G_3 = \begin{pmatrix} 0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&1\\ 0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&1\\ 0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&1 \end{pmatrix} = \begin{pmatrix} a_1a_2a_3\\ a_1a_2a_4\\ a_1a_3a_4\\ a_2a_3a_4 \end{pmatrix}.$$

Таким образом порождающая матрица кода Рида – Маллера третьего порядка длиной 16 является $(15, 16)$-матрицей вида

$$G = \begin{pmatrix} G_0\\ G_1\\ G_2\\ G_3 \end{pmatrix}.$$

Эта порождающая матрица задает $(16, 15)$-код над $\mathbb{F}_2$ (на самом деле это просто код с проверкой на четность). Другой код Рида – Маллера может быть получен с использованием этих матриц, если положить $r = 2$. В этом случае порождающая матрица имеет вид

$$G = \begin{pmatrix} G_0\\ G_1\\ G_2 \end{pmatrix}$$

и задает $(16, 11)$-код над $\mathbb{F}_2$ (в действительности это $(15, 11)$-код Хэмминга, расширенный с помощью проверки на четность).

Из определения порождающей матрицы ясно, что код Рида – Маллера r-го порядка может быть получен пополнением кода Рида – Маллера $(r-1)$-го порядка, а код Рида – Маллера $(r-1)$-го порядка получается из кода r-го порядка с помощью выбрасывания. Поскольку код Рида – Маллера r-го порядка содержит код $(r-1)$-го порядка, ясно, что его минимальное расстояние не может быть больше минимального расстояния кода $(r-1)$-го порядка.

Каждая строка матрицы G_l имеет вес 2^{m-l}. Таким образом, любая строка матрицы G имеет четный вес, а сумма двоичных слов четного веса также имеет четный вес. Следовательно, все линейные комбинации строк матрицы G имеют четный вес, а это означает, что все кодовые слова имеют четный вес. Матрица G_r содержит строки весом 2^{m-r}, и, следовательно, минимальный вес кода не может превосходить 2^{m-r}.

Алгоритм декодирования Рида [3] разработан специально для кодов Рида – Маллера. Можно, конечно, использовать процедуру синдромного декодирования, но в данном случае осуществить ее довольно сложно. Алгоритм Рида отличается от большинства алгоритмов декодирования тем, что позволяет восстановить информационные символы прямо из принятого слова и при этом не дает точного значения самой ошибки. В этом алгоритме не используются также промежуточные переменные, например синдром. Алгоритм Рида позволяет исправлять $2^{m-r-1}-1$ ошибок и восстанавливать k информационных символов. Отсюда следует, что минимальное расстояние будет не меньше 2^{m-r-1}, но оно должно быть четным и поэтому будет не меньше 2^{m-r}.

Рассмотрим совершенные циклические *коды Голея*. Заметим, что

$$\left(\binom{23}{0}+\binom{23}{1}+\binom{23}{2}+\binom{23}{3}\right)2^{12}=2^{23}.$$

Это равенство представляет собой необходимое условие существования совершенного исправляющего тройные ошибки $(23,12)$-кода над $\mathbb{F}_2$. Такой код, названный *кодом Голея*, действительно существует. Он удовлетворяет границе Хэмминга со знаком равенства.

Определим код Голея [3] как двоичный циклический код через порождающий многочлен. Пусть $g(x)$ и $\tilde{g}(x)$ — следующие взаимные многочлены:

$$g(x)=x^{11}+x^{10}+x^6+x^5+x^4+x^2+1,$$
$$\tilde{g}(x)=x^{11}+x^9+x^7+x^6+x^5+x+1.$$

Простым вычислением проверяется, что $(x-1)g(x)\tilde{g}(x)=x^{23}-1$, так что в качестве порождающего многочлена циклического $(23,12)$-кода можно использовать как $g(x)$, так и $\tilde{g}(x)$.

Граница Хэмминга показывает, что минимальное расстояние этого кода не может быть больше 7.

Рассмотрим корни многочленов $g(x)$ и $\tilde{g}(x)$ в соответствующем расширении поля $\mathbb{F}_2$. Построим минимальные многочлены некоторых элементов из $\mathbb{F}_{2048}$, которые затем окажутся равными $g(x)$ и $\tilde{g}(x)$. Если α — примитивный элемент поля, то в силу разложения $2047=23\cdot 89$ и элемент поля $\beta=\alpha^{89}$, и обратный ему элемент β^{-1} имеют порядок 23. Пусть $f(x)$ и $\tilde{f}(x)$ — минимальные многочлены элементов β и β^{-1} соответственно.

Минимальный многочлен элемента β равен

$$f(x)=(x-\beta)(x-\beta^2)(x-\beta^4)\cdots(x-\beta^{2^{r-1}}),$$

где все показатели степеней приводятся по модулю 23. Сопряженными являются элементы множества

$$B=\{\beta,\beta^2,\beta^4,\beta^8,\beta^{16},\beta^9,\beta^{18},\beta^{13},\beta^3,\beta^6,\beta^{12}\},$$

которых всего 11, так что степень многочлена $f(x)$ равна 11. Аналогично минимальный многочлен элемента β^{-1} равен

$$\tilde{f}(x)=(x-\beta^{-1})(x-\beta^{-2})(x-\beta^{-4})\cdots(x-\beta^{-2^{r-1}}),$$

а множество сопряженных элементов

$$\tilde{B} = \{\beta^{-1}, \beta^{-2}, \beta^{-4}, \beta^{-8}, \beta^{-16}, \beta^{-9}, \beta^{-18}, \beta^{-13}, \beta^{-3}, \beta^{-6}, \beta^{-12}\}.$$

Вместе множества B и $\tilde{B}$ содержат 22 элемента поля, порядок каждого из которых равен 23, следовательно, $(x-1)f(x)\tilde{f}(x) = x^{23}-1$, причем согласно теореме о единственности разложения это разложение единственно. Поэтому порождающие многочлены $g(x)$ и $\tilde{g}(x)$ представляют собой минимальные многочлены элементов α^{89} и α^{-89} из поля $\mathbb{F}_{2048}$.

Приведем две леммы без доказательств из [3].

Лемма 8.18. *Код Голея не содержит ненулевых кодовых слов веса 4 и менее.*

Лемма 8.19. *Если вес слова Голея четен, то он кратен 4.*

Теорема 8.15. *Код Голея является совершенным кодом, исправляющим три ошибки.*

Доказательство. Необходимо показать, что минимальное расстояние равно по меньшей мере 7. В силу леммы 8.18 оно не меньше 5, а вследствие леммы 8.19 оно не может быть равно 6. Таким образом, необходимо только доказать, что код не содержит слов весом 5.

Рассмотрим кодовое слово $g(x)\tilde{g}(x)$. В действительности $g(x)\tilde{g}(x) = \sum_{k=0}^{n-1} x^k$, так как $(x-1)g(x)\tilde{g}(x) = x^n - 1 = (x-1)\sum_{k=0}^{n} x^k$. Таким образом, коду принадлежит слово, все символы которого равны 1. Прибавление этого кодового слова к любому кодовому слову весом w дает кодовое слово веса $23-w$. Тогда согласно лемме 8.19 в коде не содержится слов с весами, равными 21, 17, 13, 9, 5 и 1. В частности, не содержится слов весом 5. □

Кроме двоичного кода Голея существует также совершенный троичный $(11, 6)$-код Голея с минимальным расстоянием, равным 5. Этими двумя кодами исчерпываются все нетривиальные примеры совершенных кодов, исправляющих более одной ошибки.

8.9. КОДЫ ХЭММИНГА КАК ЦИКЛИЧЕСКИЕ КОДЫ

Покажем, что при определенном порядке следования столбцов в проверочной матрице код Хэмминга — циклический, что позволяет использовать теорию циклических кодов для кодирования и декодирования.

Теорема 8.16. *Семейство эквивалентных кодов Хэмминга $\mathcal{H}_r$ содержит циклический код с порождающим многочленом $g(x)$ над полем $\mathbb{F}_2$, который является минимальным многочленом примитивного элемента α поля $\mathbb{F}_{2^r}$.*

Доказательство. Доказательство теоремы проведем согласно [33]. Пусть α — примитивный элемент поля $\mathbb{F}_{2^r}$ и $g(x) \in \mathbb{F}_2[x]$ — минимальный многочлен элемента α, $\deg g(x) = r$. Поле $\mathbb{F}_{2^r}$ рассмотрим как линейное пространство над $\mathbb{F}_2$, а в качестве базиса этого линейного пространства — множество элементов $1, \alpha, \alpha^2, \ldots, \alpha^{r-1}$. Для произвольного элемента $\beta \in \mathbb{F}_{2^r}$ обозначим через $\varphi(\beta) = \beta^{\downarrow}$ r-мерный столбец, составленный из коэффициентов разложения

элемента β по указанному базису. Отображение $\varphi: \mathbb{F}_{2^r} \to V_n$ — изоморфизм линейных пространств над полем $\mathbb{F}_2$.

Обозначим $n = 2^r - 1$ и рассмотрим матрицу

$$H = ((1)^{\downarrow},(\alpha)^{\downarrow},(\alpha^2)^{\downarrow}, \ldots, (\alpha^{n-1})^{\downarrow})$$

размером $r \times n$. Поскольку $1, \alpha, \alpha^2, \ldots, \alpha^{n-1}$ — все ненулевые элементы поля $\mathbb{F}_{2^r}$, в матрице H по одному разу записаны (в некотором порядке) все ненулевые r-мерные столбцы над $\mathbb{F}_2$; следовательно, матрица H — проверочная матрица кода Хэмминга $\mathcal{H}_r$.

Слово $c = (c_0, c_1, \ldots, c_n) \in V_n$ принадлежит коду $\mathcal{H}_r$ тогда и только тогда, когда $cH' = 0^{n-k}$, или

$$c_0(1)^{\downarrow} + c_1(\alpha)^{\downarrow} + c_2(\alpha^2)^{\downarrow} + \ldots + c_{n-1}(\alpha^{n-1})^{\downarrow} = 0^{\downarrow}.$$

Применяя к этому равенству изоморфизм φ^{-1}, получим в поле $\mathbb{F}_{2^r}$ равенство

$$c_0 + c_1\alpha + c_2\alpha^2 + \ldots + c_{n-1}\alpha^{n-1} = 0,$$

которое означает, что элемент α — корень многочлена $c(x) = c_0 + c_1x + c_2x^2 + \ldots + c_{n-1}x^{n-1}$ над полем $\mathbb{F}_2$. Поскольку $g(x)$ — минимальный многочлен элемента α, $c(x)$ делится на $g(x)$.

Итак, слово c принадлежит коду $\mathcal{H}_r$ тогда и только тогда, когда многочлен $c(x)$ делится на $g(x)$. Это означает, что $\mathcal{H}_r$ — циклический код с порождающим многочленом $g(x)$. □

8.10. КОДЫ БЧХ, ИСПРАВЛЯЮЩИЕ ЗАДАННОЕ ЧИСЛО ОШИБОК

Рассмотрим способ построения циклических кодов с помощью корней порождающего многочлена. Такой способ позволяет строить циклические коды, для которых выполняется важное свойство: минимальное кодовое расстояние не меньше, чем некоторая наперед заданная величина. Это позволяет гарантированно обнаруживать и исправлять заданное число ошибок.

Пусть $\mathcal{C}$ — циклический код с порождающим многочленом $g(x)$, β — корень многочлена $g(x)$ в некотором расширении $\mathbb{F}_{q^r}$ поля $\mathbb{F}_q$. Поскольку порождающий многочлен $g(x)$ делит любой кодовый многочлен $c(x)$, элемент β также корень многочлена $c(x)$.

Определение 8.29. *Выберем целые $m_0, d_0, r \geqslant 1$. Пусть α — примитивный элемент конечного поля $\mathbb{F}_{q^r}$, $g(x)$ — унитарный многочлен над $\mathbb{F}_q$ наименьшей степени, для которого элементы*

$$\alpha^{m_0},\ \alpha^{m_0+1},\ \ldots,\ \alpha^{m_0+d_0-1}$$

поля $\mathbb{F}_{q^r}$ являются корнями, и n — наименьшее общее кратное порядков указанных элементов, $n = \mathrm{НОК}\{\mathrm{ord}(\alpha^i),\ m_0 \leqslant i \leqslant m_0 + d_0 - 1\}$. БЧХ-кодом с параметрами (m_0,d_0) называется q-ичный циклический код длиной n с порождающим многочленом $g(x)$. Число d_0+1 называется конструктивным кодовым расстоянием БЧХ-кода.

Приведем некоторые теоремы из [33].

Теорема 8.17. *Пусть* $\beta_1, \ldots, \beta_s$ *— элементы поля* $\mathbb{F}_{q^r}$ *и* n *есть наименьшее общее кратное порядков* $\operatorname{ord}(\beta_1), \ldots, \operatorname{ord}(\beta_s)$. *Обозначим через* $[\mathcal{C}]$ *множество всех классов* $[c(x)]$ *из факторкольца* R_n, *для которых все элементы* $\beta_1, \ldots, \beta_s$ *— корни многочлена* $c(x)$:

$$[\mathcal{C}] = \{[c(x)] \in R_n : \ c(\beta_1) = \ldots = c(\beta_s) = 0\}.$$

Тогда множество $\mathcal{C}$ *слов из* V_n, *соответствующих классам из* $[\mathcal{C}]$, *является циклическим кодом.*

Доказательство. Вначале покажем, что множество $[\mathcal{C}]$ определено корректно. Если элемент β_i — корень многочлена $c(x)$, то β_i корень и для всех многочленов $f(x)$ из класса $[c(x)]$. Действительно, многочлен $x^n - 1$ делит разность $f(x) - c(x)$, т. е. $f(x) = c(x) + t(x)(x^n - 1)$. Поскольку $\operatorname{ord}(\beta_i)$ делит n, $\beta_i^n = 1$, поэтому

$$f(\beta_i) = c(\beta_i) + t(\beta_i)(\beta_i^n - 1) = 0.$$

Пусть $m_i(x)$ — многочлен над полем $\mathbb{F}_q$, являющийся минимальным многочленом элемента $\beta_i \in \mathbb{F}_{q^r}$, и $g(x)$ — наименьшее общее кратное многочленов $m_1(x), \ldots, m_s(x)$. Очевидно, что элементы $\beta_1, \ldots, \beta_s$ — корни многочлена $g(x)$. Покажем, что множество $[\mathcal{C}]$ совпадает с идеалом $< [g(x)] >$.

Если $[c(x)] \in < [g(x)] >$, то $g(x)$ делит $c(x)$, поэтому $c(\beta_i) = 0$, так что $< [g(x)] > \in [\mathcal{C}]$. С другой стороны, если для некоторого многочлена $c(x)$ над полем $\mathbb{F}_q$ все элементы $\beta_1, \ldots, \beta_s$ — корни, то минимальные многочлены $m_i(x)$ делят $c(x)$, поэтому и их наименьшее общее кратное $g(x)$ также делит $c(x)$. Следовательно, $c(x) \in < [g(x)] >$, так что $[\mathcal{C}] \in < [g(x)] >$. Поскольку множество $[\mathcal{C}]$ — идеал в факторкольце R_n, соответствующее множество $\mathcal{C}$ слов из V_n есть циклический код. □

Теорема 8.18. *Пусть* $\mathcal{C}$ *— q-ичный циклический* (n, k)*-код с порождающим многочленом* $g(x)$ *и* $\alpha^{i_1}, \ldots, \alpha^{i_{n-k}}$ *все корни многочлена* $g(x)$ *в некотором расширении* $\mathbb{F}_{q^r}$ *поля* $\mathbb{F}_q$, *где* α *— примитивный элемент поля* $\mathbb{F}_{q^r}$ *и* $i_1 \leqslant i_2 \leqslant \ldots \leqslant i_{n-k}$. *Тогда минимальное кодовое расстояние* $d_{\mathcal{C}}$ *больше, чем длина наибольшего интервала последовательных (с шагом 1) чисел в ряду* $i_1, i_2, \ldots, i_{n-k}$.

Доказательство. Из условия теоремы и теоремы 8.17 вытекает, что слово $c = (c_0, c_1, \ldots, c_{n-1})$ принадлежит коду $\mathcal{C}$ в том и только том случае, когда все элементы $\alpha^{i_1}, \ldots, \alpha^{i_{n-k}}$ — корни многочлена $c(x) = c_0 + c_1 x + \ldots + c_{n-1} x^{n-1}$, т. е. $c(\alpha^{i_1}) = \ldots = c(\alpha^{i_{n-k}}) = 0$. Это условие можно записать в матричной форме $cF' = 0^{n-k}$, где матрица F над полем $\mathbb{F}_{q^r}$ имеет вид

$$F = \begin{pmatrix} 1 & \alpha^{i_1} & \alpha^{2i_1} & \ldots & \alpha^{(n-1)i_1} \\ 1 & \alpha^{i_2} & \alpha^{2i_2} & \ldots & \alpha^{(n-1)i_2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha^{i_{n-k}} & \alpha^{2i_{n-k}} & \ldots & \alpha^{(n-1)i_{n-k}} \end{pmatrix}.$$

Матрица F есть специальная матрица с элементами из поля $\mathbb{F}_{q^r}$, ее строки могут совпадать, в отличие от проверочной матрицы, в которой строки должны быть

линейно независимыми. Для матрицы F справедливо утверждение: если любые d столбцов матрицы F линейно независимы над полем $\mathbb{F}_q$, то $d_{\mathcal{C}} > d$.

Предположим, что наибольший интервал последовательных (с шагом 1) чисел в ряду $i_1, i_2, \ldots, i_{n-k}$ имеет вид $m_0, m_0+1, \ldots, m_0+d_0-1$. Нам нужно доказать неравенство $d_{\mathcal{C}} > d_0$. Для этого достаточно установить, что любые d_0 столбцов матрицы F линейно независимы над полем $\mathbb{F}_q$. Рассмотрим подматрицу F_1 матрицы F, расположенную в строках $m_0, m_0+1, \ldots, m_0+d_0-1$:

$$F_1 = \begin{pmatrix} 1 & \alpha^{m_0} & \alpha^{2m_0} & \ldots & \alpha^{(n-1)m_0} \\ 1 & \alpha^{(m_0+1)} & \alpha^{2(m_0+1)} & \ldots & \alpha^{(n-1)(m_0+1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha^{(m_0++d_0-1)} & \alpha^{2(m_0+d_0-1)} & \ldots & \alpha^{(n-1)(m_0+d_0-1)} \end{pmatrix}.$$

Если будет доказано, что любые d_0 столбцов матрицы F_1 линейно независимы над $\mathbb{F}_q$, то это же будет верно и для матрицы F. Возьмем в матрице F_1 произвольные d_0 столбцов, например с номерами $j_1 < j_2 < \ldots < j_{d_0}$, и рассмотрим квадратную подматрицу

$$F_2 = \begin{pmatrix} \alpha^{j_1 m_0} & \alpha^{j_2 m_0} & \ldots & \alpha^{j_{d_0} m_0} \\ \alpha^{j_1(m_0+1)} & \alpha^{j_2(m_0+1)} & \ldots & \alpha^{j_{d_0}(m_0+1)} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha^{j_1(m_0++d_0-1)} & \alpha^{j_2(m_0+d_0-1)} & \ldots & \alpha^{j_{d_0}(m_0+d_0-1)} \end{pmatrix}.$$

Покажем, что определитель $\det F_2$ отличен от нуля. Вынося из первого столбца элемент $\alpha^{j_1 m_0}$, из второго столбца — элемент $\alpha^{j_2 m_0}$ и т. д., получим

$$\det F_2 = \alpha^{(j_1+j_2+\ldots+j_{d_0})m_0} \cdot W(\alpha^{j_1}, \ldots, \alpha^{j_{d_0}}),$$

$$W(\alpha^{j_1}, \ldots, \alpha^{j_{d_0}}) = \det \begin{pmatrix} 1 & 1 & \ldots & 1 \\ \alpha^{j_1} & \alpha^{j_2} & \ldots & \alpha^{j_{d_0}} \\ \alpha^{2j_1} & \alpha^{2j_2} & \ldots & \alpha^{2j_{d_0}} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha^{(d_0-1)j_1} & \alpha^{(d_0-1)j_2} & \ldots & \alpha^{(d_0-1)j_{d_0}} \end{pmatrix},$$

где W есть известный определитель Вандермонда. Поскольку элементы $\alpha^{j_1}, \ldots, \alpha^{j_{d_0}}$ различны, $W(\alpha^{j_1}, \ldots, \alpha^{j_{d_0}}) \neq 0$, то $\det F_2 \neq 0$. □

Теорема 8.18 дает возможность явно указать способ построения циклических кодов, для которых минимальное кодовое расстояние будет больше заданной границы. Такой способ был открыт независимо в работах Р. Боуза, Д. Рой-Чоудхури и А. Хоквингема, эти коды называют БЧХ-кодами.

Наиболее важный для практического применения класс БЧХ-кодов получается при $q = 2, m_0 = 1, d_0 = 2t_0$. Для такого двоичного циклического кода $d_{\mathcal{C}} \geqslant 2t_0+1$, поэтому декодер в ближайшее кодовое слово правильно исправляет любые комбинации из t_0 или меньшего числа ошибок.

Теорема 8.19. *Для любых $r, t_0 \geqslant 1$ таких, что $t_0 < n/2, n = 2^r - 1$, существует двоичный циклический код длиной n с конструктивным кодовым расстоянием $2t_0 + 1$, имеющий не более rt_0 проверочных символов.*

Доказательство. Очевидно, что если $\beta \in \mathbb{F}_{q^r}$ — корень многочлена $f(x)$ над полем $\mathbb{F}_q$, то элемент $\beta^{q^i}, i \in \mathbb{N}$, также является корнем $f(x)$.

Рассмотрим двоичный БЧХ-код с параметрами $m_0 = 1, d_0 = 2t_0$. Порождающий многочлен $g(x)$ этого кода имеет корни $\alpha, \alpha^2, \ldots, \alpha^{2t_0}$, где α — примитивный элемент поля $\mathbb{F}_{q^r}$. Длина кодового слова равна $n = \text{НОК}\{\text{ord}(\alpha^i),\ 1 \leqslant i \leqslant 2t_0\} = \text{ord}(\alpha) = 2^r - 1$.

Пусть $m_i(x)$ — многочлен над $\mathbb{F}_2$, являющийся минимальным многочленом элемента $\alpha_i \in \mathbb{F}_{2^r}$; тогда $\deg m_i(x) \leqslant r$. Как видно из доказательства теоремы 8.17, порождающим многочленом БЧХ-кода является $g(x) = \text{НОК}\{m_i(x),\ 1 \leqslant i \leqslant 2t_0\}$. Из замечания в начале доказательства следует, что $m_{2i}(x)$ делит $m_i(x)$, поэтому при вычислении наименьшего общего кратного $g(x)$ достаточно взять только многочлены $m_i(x)$ с нечетными номерами, т. е.

$$g(x) = \text{НОК}\{m_1(x), m_3(x), \ldots, m_{2t_0-1}(x)\}.$$

Поскольку $g(x)$ наименьшее общее кратное t_0 многочленов, степени которых не превосходят r, справедливо неравенство $\deg g(x) \leqslant rt_0$. С другой стороны, $\deg g(x) = n - k$. Согласно определению систематического кода число $n - k$ равно числу проверочных символов в кодовом слове. □

К кодам БЧХ применимы любые методы декодирования циклических кодов. Однако имеются существенно лучшие алгоритмы, разработанные специально для декодирования кодов БЧХ. К ним относится алгоритм декодирования Питерсона – Горенстейна – Цирлера [3], в основу которого входит решение соответствующей системы линейных уравнений. Такую систему можно решить путем обращения матрицы, если она невырождена. Для эффективного решения данной системы без обращения матрицы разработан *алгоритм Берлекэмпа – Месси* [3], в основе которого лежит тот факт, что данная система обладает специальной структурой. В связи с этим реализация алгоритма декодирования Берлекэмпа – Месси состоит в переформулировке задачи в виде задачи построения схемы с использованием регистров сдвига с линейной обратной связью. И новая задача заключается в том, чтобы среди большого числа таких регистров сдвига найти регистр сдвига с наименьшей длиной. Это позволяет определить вектор ошибок, вес которого минимален для принятого слова. Точное описание процедуры декодирования приведено в книге [3].

Пример 8.5. Построим двоичный БЧХ-код с конструктивным кодовым расстоянием, равным 5. Для этого рассмотрим конечное поле $\mathbb{F}_{16}$, при построении которого используем неприводимый над $\mathbb{F}_2$ многочлен $m_1(x) = 1 + x + x^4$, а поле $\mathbb{F}_{16}$ представим как факторкольцо $\mathbb{F}_2[x]/m_1(x)$, причем примитивный элемент равен $\alpha = x$. Представления для элементов поля в виде степеней элемента α, элементов факторкольца $\mathbb{F}_2[x]/m_1(x)$ и 4-мерных двоичных слов приведены в табл. 8.2

Рассмотрим двоичный БЧХ-код с параметрами $m_0 = 1, d_0 = 4$ как цикли-

Таблица 8.2

Таблица представления элементов поля

0	0	0000
α	x	0100
α^2	x^2	0010
α^3	x^3	0001
α^4	$1+x$	1100
α^5	$x+x^2$	0110
α^6	x^2+x^3	0011
α^7	$1+x+x^3$	1101
α^8	$1+x^2$	1010
α^9	$x+x^3$	0101
α^{10}	$1+x+x^2$	1110
α^{11}	$x+x^2+x^3$	0111
α^{12}	$1+x+x^2+x^3$	1111
α^{13}	$1+x^2+x^3$	1011
α^{14}	$1+x^3$	1001
α^{15}	1	1000

ческий код длиной $2^4-1=15$ с корнями $\alpha, \alpha^2, \alpha^3, \alpha^4$. Многочлен $m_1(x)$ имеет корни $\alpha, \alpha^2, \alpha^4, \alpha^8$, поэтому порождающий многочлен $g(x)$ равен наименьшему общему кратному многочленов $m_1(x)$ и $m_3(x)$, где $m_3(x)$ — минимальный многочлен элемента α^3. Многочлен $m_3(x)$ имеет также корни $\alpha^6, \alpha^9, \alpha^{12}$, поэтому

$$m_3(x)=(x-\alpha^3)(x-\alpha^6)(x-\alpha^9)(x-\alpha^{12}).$$

Для нахождения минимальных многочленов можно использовать некоторые методы из работы [16]. Воспользуемся методом неопределенных коэффициентов для нахождения минимального многочлена $m_3(x)$. Пусть $m_3(x)=x^4+\beta_3x^3+$ $+\beta_2x^2+\beta_1x+1$, где $\beta_i\in F_2$ — неизвестные коэффициенты. Подставив в это выражение корень α^3, получим

$$m_3(\alpha^3)=\begin{pmatrix}1\\1\\1\\1\end{pmatrix}+\beta_3\begin{pmatrix}0\\1\\0\\1\end{pmatrix}+\beta_2\begin{pmatrix}0\\0\\1\\1\end{pmatrix}+\beta_1\begin{pmatrix}0\\0\\0\\1\end{pmatrix}+\begin{pmatrix}1\\0\\0\\0\end{pmatrix}=\begin{pmatrix}0\\0\\0\\0\end{pmatrix}.$$

Решая эту систему, найдем, что $\beta_3=\beta_2=\beta_1=1$. Поэтому $m_3(x)=1+x+$ $+x^2+x^3+x^4$. Следовательно, порождающий многочлен кода имеет вид

$$g(x)=\text{НОК}\{m_1(x),m_3(x)\}=m_1(x)m_3(x)=1+x^4+x^6+x^7+x^8.$$

Скорость передачи информации равна $\mathcal{R}=7/15$. Порождающему многочлену $g(x)$ соответствует кодовое слово $g=(1000101110000000)$ веса 5, так что для рассматриваемого кода истинное и конструктивное кодовые расстояния совпадают: $d_{\mathcal{C}}=d_0+1=5$. Следовательно, декодер в ближайшее кодовое слово правильно исправляет любые одиночные и двойные ошибки.

Пример 8.6. Построим троичный БЧХ-код с конструктивным кодовым расстоянием $d_0+1=5$ и длиной кодового слова $n=8$. Для этого возьмем конечное поле $\mathbb{F}_9$. Всего существует три неприводимых над $\mathbb{F}_3$ многочлена степени 3: x^2+1, x^2+x+2 и x^2+2x+2. Поле $\mathbb{F}_{3^2}$ представим как факторкольцо $\mathbb{F}_3[x]/(x^2+1)$. Найдем примитивный элемент поля $\mathbb{F}_9$. Для этого используем алгоритм нахождения примитивного элемента. Случайным образом выберем элемент из мультипликативной группы поля $\mathbb{F}_9^*$ и проверим, что $\alpha_4=-1=2$. Пусть $\alpha=x$, тогда $\alpha^2=2, \alpha^4=1$, следовательно, $\operatorname{ord}(\alpha)=4$. Далее выберем $\alpha=x+1$, тогда $\alpha^2=x^2+2x+1=2x$, $\alpha^4=x^2=2$, значит, $\alpha=x+1, \operatorname{ord}(\alpha)=8$ — образующий элемент группы $\mathbb{F}_9^*$, т. е. примитивный элемент поля $\mathbb{F}_9\cong\mathbb{F}_3[x]/(x^2+1)$. Представления элементов поля $\mathbb{F}_9$ приведены в табл. 8.3

Таблица 8.3

Таблица представления элементов поля

0	0	00
α	$x+1$	11
α^2	$2x$	02
α^3	$2x+1$	12
α^4	2	20
α^5	$2x+2$	22
α^6	x	01
α^7	$x+2$	21
α^8	1	10

Элементы поля $\mathbb{F}_9$ разобьем на классы:

$$\{0\}=\{8\},\ \{1,3\},\ \{2,6\},\ \{4\},\ \{5,7\}.$$

Так, например, класс $\{2,6\}$ содержит сопряженные элементы α^2, α^6. Поскольку $\alpha^6=x$, то минимальный многочлен элемента α^2 равен $m_2(z)=z^2+1$. Найдем минимальный многочлен примитивного элемента α. Для этого составим систему уравнений

$$m_1(z)=z^2+\beta_1 z+\beta_0,\ m_1(\alpha)=\begin{pmatrix}0\\2\end{pmatrix}+\beta_1\begin{pmatrix}1\\1\end{pmatrix}+\beta_0\begin{pmatrix}1\\0\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}.$$

Получим, что $\beta_1=1$, $\beta_0=2$. Следовательно, $m_1(z)=z^2+z+2$. Так как остался всего один неприводимый над $\mathbb{F}_3$ многочлен степени 2, то $m_5(z)=z^2+2z+2$. Очевидно, что $m_0(z)=z+2$, $m_4(z)=z+1$.

От выбора параметра m_0 зависит скорость передачи информации $\mathcal{R}$. Так, исходя из разбиения на классы, имеем

$$m_0\in\{0,1,4,5\},\ \mathcal{R}=3/8;$$
$$m_0\in\{2,3,6,7\},\ \mathcal{R}=1/8.$$

Выберем $m_0=0$ и рассмотрим БЧХ-код с порождающим многочленом, корни которого — элементы α^0, α^1, α^2, α^3. Следовательно, порождающий многочлен

равен

$$g(x) = \text{НОК}\{m_0(x),m_1(x),m_2(x)\} = m_0(x)m_1(x)m_2(x) =$$
$$= (x+2)(x^2+x+2)(x^2+1) = x^5+2x^3+x^2+x+1.$$

Порождающая матрица имеет вид

$$G = \begin{pmatrix} 1 & 1 & 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 2 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 2 & 0 & 1 \end{pmatrix}.$$

Проверочный многочлен и проверочную матрицу можно записать как

$$h(x) = \frac{x^8-1}{g(x)} = m_4(x)m_5(x) = (x+1)(x^2+2x+2) = x^3+x+2;$$

$$H = \begin{pmatrix} 1 & 0 & 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 2 \end{pmatrix}.$$

Кодовое расстояние и конструктивное кодовое расстояние совпадают: $d_{\mathcal{C}} = 5$.

Важным и широко используемым подмножеством кодов БЧХ являются *коды Рида – Соломона*. Это такие коды БЧХ, у которых мультипликативный порядок алфавита символов кодового слова делится на длину кода. В коде Рида – Соломона, исправляющем t ошибок, порождающий многочлен может быть записан в виде

$$g(x) = (x-\alpha)(x-\alpha^2)\cdots(x-\alpha^{2t}),$$

где α — примитивный элемент. Степень этого многочлена всегда равна $2t$, откуда следует, что параметры кода Рида – Соломона связаны соотношением $n-k=2t$. Коды Рида – Соломона оптимальны в смысле границы Синглтона.

Теорема 8.20. *Код Рида – Соломона имеет кодовое расстояние $n-k+1$ и является кодом с максимальным расстоянием.*

Доказательство. Пусть $d = 2t+1$ — конструктивное расстояние кода. Минимальное расстояние d^* удовлетворяет неравенству

$$d^* \geqslant d = 2t+1 = n-k+1,$$

поскольку для кодов Рида – Соломона $2t = n-k$. Но для любого линейного кода имеет место граница Синглтона $d^* \leqslant n-k+1$. Следовательно, $d^* = n-k+1$ и $d^* = d$. □

Коды Гоппы определяются в терминах многочленов Гоппы $G(z)$. В противоположность циклическим кодам, для которых трудно по порождающему многочлену определить их минимальное расстояние, коды Гоппы обладают тем свойством, что $d \geqslant \deg G(z) + 1$.

Определение кода Гоппы длиной n с символами из поля $\mathbb{F}_q$ опирается на два объекта: многочлен с коэффициентами из поля $\mathbb{F}_{q^m}$ для некоторого фиксированного m, называемого многочленом Гоппы, и подмножество $L = \{\alpha_1, \ldots, \alpha_n\}$

элементов из $\mathbb{F}_{q^m}$ таких, что $G(\alpha_i) \neq 0$ для всех $\alpha_i \in L$. Обычно в качестве L выбирается подмножество всех элементов поля $\mathbb{F}_{q^m}$, которые не являются корнями многочлена $G(z)$.

С каждым словом $a = (a_1, \ldots, a_n)$ над $\mathbb{F}_q$ свяжем рациональную функцию

$$R_a(z) = \sum_{i=1}^{n} \frac{\alpha_i}{z - \alpha_i}. \tag{8.51}$$

Определение 8.30. *Код Гоппы* $\mathcal{G}$ *состоит из всех слов* a *таких, что*

$$R_a(z) \equiv 0 \mod G(z), \tag{8.52}$$

или, что эквивалентно, таких, что $R_a(z) = 0$ *в кольце многочленов* $\mathbb{F}_{q^m}[z]/G(z)$.

Если $G(z)$ неприводим, то $\mathcal{G}$ называется неприводимым кодом Гоппы. Очевидно, что $\mathcal{G}$ — линейный код. Его проверочная матрица может быть найдена из (8.52). Многочлен $z - \alpha_i$ в кольце многочленов по модулю $G(z)$ имеет обратный многочлен, который равен

$$(z - \alpha_i)^{-1} = -\frac{G(z) - G(\alpha_i)}{z - \alpha_i} G^{-1}(\alpha_i).$$

Действительно,

$$-(z - \alpha_i)\frac{G(z) - G(\alpha_i)}{z - \alpha_i} G^{-1}(\alpha_i) \equiv 1 \mod G(z).$$

Следовательно, слово a лежит в коде $\mathcal{G}$ тогда и только тогда, когда

$$\sum_{i=1}^{n} *\alpha_i \frac{G(z) - G(\alpha_i)}{z - \alpha_i} G^{-1}(\alpha_i) = 0. \tag{8.53}$$

Если $G(z) = \sum_{i=0}^{r} g_i z^i$, где $g_i \in \mathbb{F}_{q^m}$ и $g_r \neq 0$, то

$$\frac{G(z) - G(\alpha_i)}{z - \alpha_i} = g_r(z^{r-1} + z^{r-2}\alpha_i + \ldots + \alpha_i^{r-1}) + \\ + g_{r-1}(z^{r-2} + \ldots + \alpha_i^{r-2}) + \ldots + g_2(z + \alpha_i) + g_1.$$

Приравнивая согласно (8.53) нулю коэффициенты при $z^{r-1}, z^{r-2}, \ldots, 1$, видим, что a принадлежит коду $\mathcal{G}$ тогда и только тогда, когда $Ha' = 0$, где

$$H = \begin{pmatrix} g_r G^{-1}(\alpha_1) & \ldots & g_r G^{-1}(\alpha_n) \\ (g_{r-1} + \alpha_1 g_r) G^{-1}(\alpha_1) & \ldots & (g_{r-1} + \alpha_n g_r) G^{-1}(\alpha_n) \\ \vdots & \ddots & \vdots \\ (g_1 + \ldots + \alpha_1^{r-1} g_r) G^{-1}(\alpha_1) & \ldots & (g_1 + \ldots + \alpha_n^{r-1} g_r) G^{-1}(\alpha_n) \end{pmatrix} =$$

$$= \begin{pmatrix} g_r & 0 & 0 & \dots & 0 \\ g_{r-1} & g_r & 0 & \dots & 0 \\ g_{r-2} & g_{r-1} & g_r & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ g_1 & g_2 & g_3 & \dots & g_r \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & \dots & 1 \\ \alpha_1 & \alpha_2 & \dots & \alpha_n \\ \alpha_1^2 & \alpha_2^2 & \dots & \alpha_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^{r-1} & \alpha_2^{r-1} & \dots & \alpha_n^{r-1} \end{pmatrix} \times$$

$$\times \begin{pmatrix} G^{-1}(\alpha_1) & 0 & \dots & 0 \\ 0 & G^{-1}(\alpha_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & G^{-1}(\alpha_n) \end{pmatrix}.$$

Проверочная матрица с элементами из $\mathbb{F}_q$ получается из H заменой каждого элемента матрицы соответствующим столбцом длиной m над $\mathbb{F}_q$.

8.11. О ДРЕВОВИДНОМ КОДИРОВАНИИ: СВЕРТОЧНЫЕ КОДЫ

В данном разделе кратко рассмотрим другой подход к кодированию, исправляющему ошибки, — древовидное кодирование.

Рассмотрим дискретные каналы связи. Пусть поток данных разбивается на короткие блоки длины k_0, которые назовем кадрами информационных символов. Эти кадры обычно включают в себя лишь несколько символов. Кадры информационных символов кодируются кадрами кодового слова длиной n_0 каждый. Однако вместо того, чтобы независимо кодировать отдельные кадры информационных символов в отдельные кадры кодового слова, кодирование каждого кадра информационных символов в отдельный кадр кодового слова производится с учетом предыдущих m кадров информационных символов. Такие коды называются *древовидными*. Среди древовидных кодов наиболее важными являются сверточные коды. Подробно об древовидных и сверточных кодах можно прочитать в [3, 20].

Типичный вид древовидного кода представлен на рис. 8.1. Как уже было отмечено выше, поток входящих информационных символов разбивается на сегменты, которые содержат по k_0 символов и называются кадрами информационных каналов. В кодере хранится m кадров. В течение каждого такта в регистр сдвига вводится новый кадр информационных символов, а кадр информационных символов, дольше остальных хранившихся в нем, выводится из него и сбрасывается. В начале каждого такта кодер по введенному кадру информационных символов и m хранящимся в нем кадрам вычисляет один кадр кодового слова, имеющий длину n_0 символов. Следовательно, каждым k_0 информационным символам соответствует передача по каналу n_0 кодовых символов. Скорость этого древовидного кода определяется соотношением

$$R = \frac{k_0}{n_0}.$$

Так, на рис. 8.1 $k_0 = 3$, $n_0 = 5$, $m = 6$.

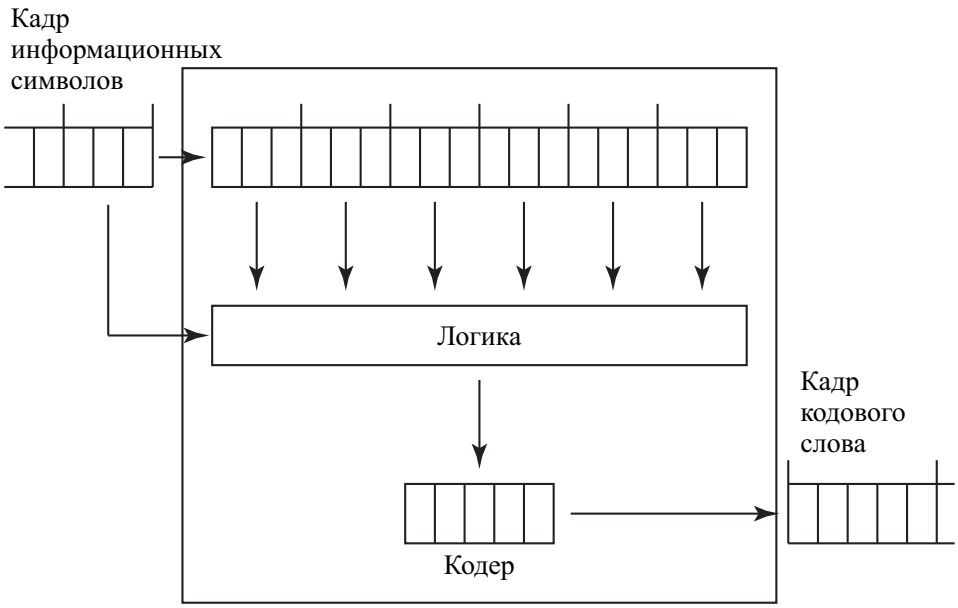


Рис. 8.1. Типичный древовидный кодер в виде регистра сдвига

Важной характеристикой древовидного и, в частности, сверточного кода является величина $\nu = mk_0$. Она называется длиной кодового ограничения. В древовидных кодах используются и другие меры длины. Так, $k = (m+1)k_0$ — информационная длина слова сверточного кода. Соответствующая ей мера кодовых последовательностей называется кодовой длиной блока n. Можно легко показать, что кодовая длина блока — это длина кодового слова, на которой сохраняется влияние одного кадра информационных символов.

Частные случаи древовидных кодов получаются различными комбинациями следующих свойств.

1) *Конечность длины кодового ограничения.* Длина кодового ограничения может быть конечной или бесконечной. Практически древовидные коды всегда имеют конечную длину кодового ограничения. Однако в теоретических исследованиях иногда полезны коды с бесконечной длиной кодового ограничения.

2) *Постоянство во времени.* Если две различные входные последовательности совпадают во всем, но с временных сдвигом на целое число кадров, то соответствующие им кодовые последовательности также совпадают во всем, но с временным сдвигом на то же самое целое число кадров.

3) *Линейность.* Если d_1 и d_2 — две информационные последовательности с кодовыми последовательностями $G(d_1)$ и $G(d_2)$, то $\alpha d_1 + \beta d_2$ соответствует кодовая последовательность

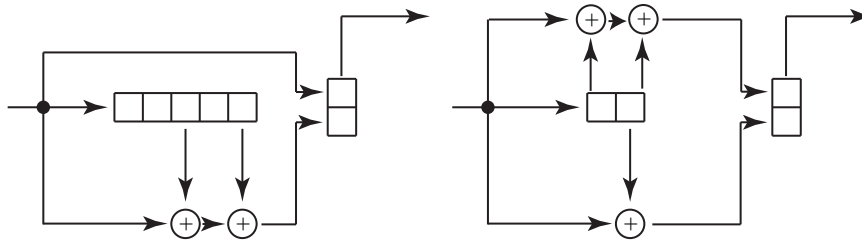
$$G(\alpha d_1 + \beta d_2) = \alpha G(d_1) + \beta G(d_2).$$

4) *Систематичность.* Систематическим является такой код, для которого в выходной последовательности кодовых символов содержится без изменений

породившая ее последовательность информационных символов.

Определение 8.31. *Линейный, постоянный по времени древовидный код, имеющий конечную длину слова $k = (m+1)k_0$, называется сверточным кодом.*

На рис. 8.2 представлены примеры двоичных сверточных кодов.



а *б*

Рис. 8.2. Примеры сверточных кодов: *а*, *б* – возможные сверточные коды

Сверточные коды удобно описывать специальным графом, называемым решеткой, узлы которого находятся в прямоугольной координатной сетке, полубесконечной справа. Число узлов в каждом столбце конечно и не превосходит D^{ν}. Узлы в каждом столбце представляют состояния, в которых может находиться регистр сдвига. Каждый последующий столбец представляет собой набор состояний в следующий момент времени. Конфигурация ребер, соединяющих узлы каждого столбца с узлами столбца справа, одинакова для всех столбцов. Узлы, которые не могут быть достигнуты из верхнего левого узла, обычно не указываются. Каждому ребру приписывается код, который будет выдан кодером при соответствующем изменении состояния.

Такая решетка описывает сверточный код в том смысле, что все пути слева направо по решетке обозначают кодовые слова.

Для примера, изображенного на рис. 8.2, *б*, решетка представлена на рис. 8.3.

Одним из способов декодирования сверточных кодов является *алгоритм декодирования Витерби*. Декодер Витерби итеративно обрабатывает кадр за кадром, двигаясь по решетке, аналогичной используемой кодером, и пытается повторить путь кодера. В любой момент времени декодер не знает, в каком узле находится кодер и поэтому не пытается декодировать этот узел. Вместо этого декодер по принятой последовательности определяет наиболее правдоподобный путь к каждому узлу и определяет расстояние между каждым таким путем (каким бы было выходное сообщение) и принятой последовательностью (каким точно было выходное сообщение). Это расстояние называется мерой расходимости пути. Если все пути в множестве наиболее правдоподобных путей начинаются одинаково, то декодер, как правило, знает начало пути, пройденного кодером.

В следующий такт декодер определяет наиболее правдоподобный путь к каждому из новых узлов этого такта. Но путь в каждый новый узел должен пройти через один из старых узлов. Возможные пути к новому узлу можно получить, продолжая к этому узлу те старые пути, которые можно к нему продолжить. Наиболее правдоподобный путь находится прибавлением приращения

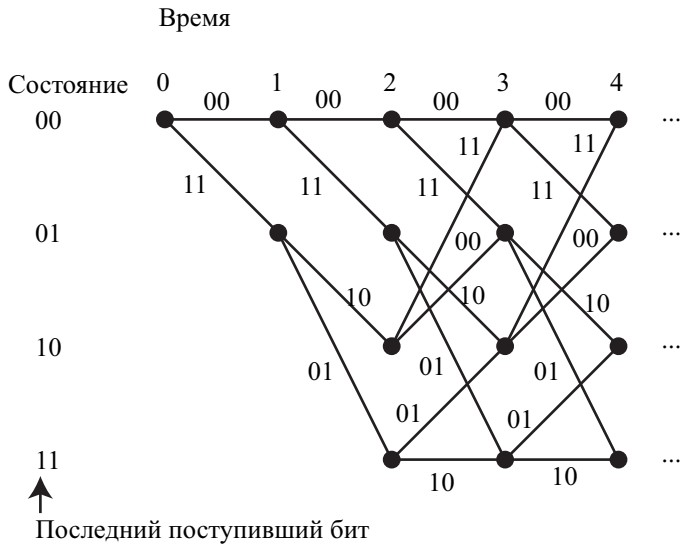


Рис. 8.3. Типичный древовидный кодер в виде регистра сдвига

меры расходимости на продолжениях старых путей к мере расходимости путей, ведущих в старый узел. В каждый новый узел ведет D^{ν} таких путей, и путь с наименьшей мерой расходимости является наиболее правдоподобным путем к новому узлу. Этот процесс повторяется для каждого из новых узлов. В конце итерации декодер знает наиболее правдоподобный путь к каждому из узлов в новом такте.

Рассмотрим множество «выживших» путей, ведущих из вершины 00 в множество узлов в r-й момент времени. Эти пути проходят через один или более узлов в момент времени 1. Если все пути проходят через один и тот же узел в первом временном кадре, то вне зависимости от того, в каком узле кодер находится в r-й момент времени, известен наиболее правдоподобный среди посещенных им в момент времени 1 узлов. Иначе говоря, декодирован первый информационный кадр, хотя еще не принято решение об r-м кадре.

Затем декодер отбрасывает первое ребро и использует новый кадр принятого слова для следующей итерации. Если вновь все «выжившие» пути проходят через один и тот же узел самого старого среди оставшихся кадров, этот информационный кадр декодируется.

Если для данного канала код построен надлежащим образом, то это решение с большой вероятностью будет правильным. Этому, однако, могут помешать следующие обстоятельства. Не все выжившие пути проходят через один и тот же узел. В этом случае процесс декодирования нарушается. Декодер может разрешать неопределенность, используя любое произвольное правило. Другая возможность состоит в том, что декодер не принимает решения, а отмечает этот участок как сегмент кодового слова, который невозможно исправить.

Кроме алгоритма Витерби, для декодирования сверточных кодов используются *метод порогового декодирования, метод последовательного декодирования* и др.

8.12. РЕШЕНИЕ ТИПОВЫХ ЗАДАЧ

Задача 8.1. Для кода $\mathcal{C}$, заданного порождающей матрицей

$$G = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix},$$

найти размерность кода k, мощность кода и проверочную H матрицы. Построить таблицу стандартного расположения, декодировать слова $y_1 = 110101$, $y_2 = 001111$, $y_3 = 010010$ с использованием синдромов.

Решение. Поскольку ранг матрицы G равен 3, то ее строки образуют базис линейного пространства, содержащего кодовые слова длиной $n = 6$, размерность кода $k = 3$, мощность кода $|\mathcal{C}| = 2^k = 2^3$.

Для того чтобы построить проверочную матрицу H, найдем ортогональное подпространство $\mathcal{C}^{\perp}$, имеющее размерность $n-k$. Базис пространства $\mathcal{C}^{\perp}$ может быть задан как

$$\begin{pmatrix} a_1 & a_2 & a_3 & a_1+a_3 & a_1+a_2 & a_2+a_3 \\ b_1 & b_2 & b_3 & b_1+b_3 & b_1+b_2 & b_2+b_3 \\ c_1 & c_2 & c_3 & c_1+c_3 & c_1+c_2 & c_2+c_3 \end{pmatrix},$$

где строки матрицы линейно независимы. Например, одна из проверочных матриц может иметь вид

$$H = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}.$$

Минимальное кодовое расстояние $d = 3$, так как любые два столбца проверочной матрицы H линейно независимы, и при этом имеются три линейно зависимых столбца. Следовательно, код $\mathcal{C}$ может исправлять любую одну ошибку и обнаруживать любые две ошибки при передаче кодового слова.

Построим таблицу стандартного расположения, в первой строке которой записаны кодовые слова из множества $\mathcal{C}$:

$$S = \begin{pmatrix} 000000 & 101011 & 011110 & 110101 & 000111 & 101100 & 011001 & 110010 \\ 100000 & 001011 & 111110 & 010101 & 100111 & 001100 & 111001 & 010010 \\ 010000 & 111011 & 001110 & 100101 & 010111 & 111100 & 001001 & 100010 \\ 001000 & 100011 & 010110 & 111101 & 001111 & 100100 & 010001 & 111010 \\ 000100 & 101111 & 011010 & 110001 & 000011 & 101000 & 011101 & 110110 \\ 000010 & 101001 & 011100 & 110111 & 000101 & 101110 & 011011 & 110000 \\ 000001 & 101010 & 011111 & 110100 & 000110 & 101101 & 011000 & 110011 \\ 100001 & 001010 & 111111 & 010100 & 100110 & 001101 & 111000 & 010011 \end{pmatrix}.$$

Вычислим синдром образующих элементов смежных классов, записанных в пер-

вом столбце S:

$$S' = \begin{pmatrix} 000000 & s(000000) = 000 \\ 100000 & s(100000) = 100 \\ 010000 & s(010000) = 010 \\ 001000 & s(001000) = 001 \\ 000100 & s(000100) = 101 \\ 000010 & s(000010) = 110 \\ 000001 & s(000001) = 011 \\ 100001 & s(100001) = 111 \end{pmatrix}.$$

Для декодирования слов можно воспользоваться одной из таблиц S, S', так как $s(y_1) = 000$, $s(y_2) = 001$, $s(y_3) = 100$, то $\mathfrak{D}_\rho(y_1) = 110101$, $\mathfrak{D}_\rho(y_2) = 001111 - 001000 = 000111$, $\mathfrak{D}_\rho(y_3) = 010010 - 100000 = 110010$.

Ответ: $\mathfrak{D}_\rho(y_1) = 110101$, $\mathfrak{D}_\rho(y_2) = 000111$, $\mathfrak{D}_\rho(y_3) = 110010$.

Задача 8.2. Для двоичного циклического кода $\mathcal{C}$ длиной $n = 7$ с порождающим многочленом $g(x) = 1 + x^2 + x^3$ найти проверочный многочлен, порождающую и проверочную матрицы и минимальное кодовое расстояние. Закодировать информационное слово $a = 1110$ для систематического кода. Декодировать принятое слово $y = 1000110$.

Решение. Разложение многочлена $x^7 - 1 = x^7 + 1$ на неприводимые над полем $\mathbb{F}_2$ множители имеет вид

$$x^7 + 1 = (x + 1)(x^3 + x + 1)(x^3 + x^2 + 1).$$

Поскольку порождающий многочлен циклического кода должен делить многочлен $x^7 + 1$, имеется всего $2^3 = 8$ двоичных циклических кодов длиной 7. Для кода $\mathcal{C}$ размерность k равна 4, а корректирующий многочлен

$$h(x) = \frac{x^7 + 1}{g(x)} = (x + 1)(1 + x + x^3) = 1 + x^2 + x^3 + x^4.$$

В качестве порождающей и проверочной матриц можно взять

$$G = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix},$$

$$H = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{pmatrix}.$$

Минимальное кодовое расстояние $d = 3$, так как любые два столбца проверочной матрицы H линейно независимы, и при этом имеются три линейно зависимых столбца (например, первый, третий и четвертый). Следовательно, декодер $\mathfrak{D}_\rho$ правильно исправляет любые одиночные ошибки.

Закодируем информационное слово $a = 1110$. Для этого информационный многочлен $1 + x + x^2$ умножим на $x^{n-k} = x^3$, результат $x^3 + x^4 + x^5$ разделим с

остатком на $g(x)$ и получим неполное частное $f(x) = 1 + x^2$ и остаток $r(x) = 1$. Кодовому многочлену $c(x) = x^3a(x) + r(x) = 1 + x^3 + x^4 + x^5$ отвечает кодовое слово $c = 1001110$.

Принятому слову $y = 1000110$ отвечает многочлен $y(x) = 1+x^4+x^5$, который не делится на $g(x)$, так что декодер обнаруживает наличие ошибок и пытается их исправить.

Все возможные многочлены $e(x)$, соответствующие одиночным ошибкам, и остатки от деления этих многочленов на порождающий многочлен $g(x)$ представлены в табл. 8.4

Таблица 8.4

$e(x)$	$e_g(x)$
1	1
x	x
x^2	x^2
x^3	$1 + x^2$
x^4	$1 + x + x^2$
x^5	$1 + x$
x^6	$x + x^2$

Далее вычисляется остаток $y_g(x) = 1 + x^2$ и по таблице находится соответствующий этому остатку многочлен ошибок x^3, т. е. $\mathfrak{D}_\rho(y) = 1 + x^3 + x^4 + x^5$.

Задача 8.3. Построить двоичный БЧХ-код, исправляющий 2 ошибки, с длиной кодового блока $n = 2^4 - 1 = 15$. Найти порождающий многочлен, проверочный многочлен и порождающую матрицу.

Решение. Рассмотрим конечное поле $\mathbb{F}_{16}$ как расширение поля $\mathbb{F}_{16}$ корнем α неприводимого над $\mathbb{F}_2$ многочлена $m(x) = 1 + x + x^4$, при этом поле $\mathbb{F}_{16}$ представим как факторкольцо $\mathbb{F}_2[x]/m(x)$, и $\alpha = [x]_{m(x)}$.

Построим двоичный БЧХ-код с параметрами $m_0 = 1, d_0 = 4$ как циклический код длиной $2^4-1 = 15$ с корнями $\alpha, \alpha^2, \alpha^3, \alpha^4$. Многочлен $m(x)$ имеет корни $\alpha, \alpha^2, \alpha^4, \alpha^8$, поэтому порождающий многочлен $g(x)$ равен наименьшему общему кратному многочленов $m(x)$ и $m_3(x)$, где $m_3(x)$ — минимальный многочлен элемента α^3. Многочлен $m_3(x)$ имеет также корни $\alpha^6, \alpha^9, \alpha^{12}$, следовательно,

$$m_3(x) = (x - \alpha^3)(x - \alpha^6)(x - \alpha^9)(x - \alpha^{12}) = 1 + x + x^2 + x^3 + x^4,$$

откуда порождающий многочлен кода можно записать как

$$g(x) = \text{НОК}\{m(x),m_3(x)\} = m(x)m_3(x) = 1 + x^4 + x^6 + x^7 + x^8,$$

а проверочный многочлен —

$$h(x) = \frac{x^{15} + 1}{g(x)} = 1 + x^4 + x^6 + x^7.$$

Порождающему многочлену $g(x)$ соответствует кодовое слово веса 5 $g =$

$= 1000101110000000$, поэтому кодовое расстояние и конструктивное кодовое расстояние совпадают, $d_{\mathcal{C}} = d_0 + 1 = 5$. Порождающая матрица имеет вид

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

Декодер в ближайшее кодовое слово правильно исправляет любые одиночные и двойные ошибки.

8.13. ЗАДАЧИ И УПРАЖНЕНИЯ

8.1. Пусть код $\mathcal{C}$ задан порождающей матрицей

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

Найти размерность кода, мощность кода и кодовое расстояние d.

8.2. Построить код, двойственный к коду кратных повторений $\mathcal{C}$.

8.3. Чему равно количество различных линейных (n, k)-кодов?

8.4. Чему равно количество различных базисов в n-мерном пространстве V_n над полем $\mathbb{F}_2$?

8.5. Информационные символы $(\varepsilon_1, \varepsilon_2)$, $\varepsilon_i \in \{0, 1\}$, $i \in \{1, 2\}$ кодируются двоичным кодом $\mathcal{C}$:

$$\begin{array}{l} a_1 = 00 \\ a_2 = 01 \\ a_3 = 10 \\ a_4 = 11 \end{array}, \quad \mathcal{C} = \left\{ \begin{array}{l} g_1 = 00000 \\ g_2 = 01101 \\ g_3 = 10111 \\ g_4 = 11010 \end{array} \right\}.$$

Проверить, что данный код $\mathcal{C}$ является линейным. Найти размерность кода и построить его порождающую G и проверочную H матрицы. Определить кодовое расстояние и соответствующее ему кодовое слово.

8.6. Код $\mathcal{C}$ задан порождающей матрицей G. Найти проверочную матрицу H и построить таблицу стандартного расположения S, если

$$G = \begin{pmatrix} 11010 \\ 01101 \end{pmatrix}.$$

8.7. По двоичному симметричному каналу передаются слова, принадлежащие коду

$$C = \left\{ \begin{array}{ll} c_1 = 0000 & c_5 = 1001 \\ c_2 = 0011 & c_6 = 1010 \\ c_3 = 0101 & c_7 = 1100 \\ c_4 = 0110 & c_8 = 1111 \end{array} \right\}.$$

Кодовые слова передаются равновероятно. Показать, что код $\mathcal{C}$ линейный. Построить его порождающую матрицу G и проверочную матрицу H. Вычислить кодовое расстояние d.

8.8. По заданной порождающей матрице G двоичного линейного кода определить проверочную матрицу H и параметры n, k, d:

$$G = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \end{pmatrix}.$$

8.9. Для двоичного линейного кода с порождающей матрицей

$$G = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix}$$

найти проверочную матрицу, а также синдромы для слов 01000, 00101, 10010, 11111.

8.10. Для троичного линейного кода с порождающей матрицей

$$G = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 2 & 0 & 1 & 1 \end{pmatrix}$$

декодировать в ближайшее кодовое слово следующие слова: 2121, 1201, 2222.

8.11. Для $(5, 3)$-кода над полем $\mathbb{F}_3$ с порождающей матрицей

$$G = \begin{pmatrix} 2 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 2 & 0 \\ 0 & 0 & 2 & 1 & 1 \end{pmatrix}$$

найти проверочную матрицу H, кодовое расстояние d и синдромы для слов 21100, 02110, 00211, 22010.

8.12. Для двоичного линейного кода с порождающей матрицей

$$G = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

найти проверочную матрицу, декодировать слова 01000, 00101, 10010, 11111.

8.13. Двоичный линейный $(8,5)$-код $\mathcal{C}$ задан порождающей матрицей

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

Показать, что $\mathcal{C}$ — циклический код. Найти его порождающий и проверочный многочлены. Построить проверочную матрицу H кода $\mathcal{C}$. Вычислить кодовое расстояние.

8.14. Убедиться, что $\mathcal{C}$, заданный порождающей матрицей

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

является циклическим. Определить его порождающий многочлен, проверочный многочлены и проверочную матрицу H.

8.15. Выяснить, является ли циклическим $(5,3)$-код над полем $\mathbb{F}_3$ с порождающей матрицей

$$G = \begin{pmatrix} 2 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 2 & 0 \\ 0 & 0 & 2 & 1 & 1 \end{pmatrix}.$$

8.16. Для двоичного циклического кода длиной $n = 7$ с порождающим многочленом $1 + x^2 + x^3 + x^4$ найти проверочный многочлен, порождающую и проверочную матрицы и минимальное кодовое расстояние.

8.17. Доказать, что $(4,2)$-код над полем $\mathbb{F}_3$ с порождающей матрицей

$$G = \begin{pmatrix} 1 & 0 & 2 & 0 \\ 1 & 1 & 2 & 2 \end{pmatrix}$$

является циклическим. Вычислить порождающий многочлен $g(x)$.

8.18. Построить двоичный БЧХ-код с длиной кодового блока $n = 15$ и кодовым расстоянием $d = 7$. Определить порождающий многочлен, проверочный многочлен, порождающую и проверочную матрицы.

Глава 9

ДИСКРЕТНЫЕ (БЕЗ ПАМЯТИ) КАНАЛЫ ПЕРЕДАЧИ ИНФОРМАЦИИ И СВЯЗАННЫЕ С НИМИ ТЕОРЕМЫ КОДИРОВАНИЯ

9.1. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ДИСКРЕТНОГО КАНАЛА СВЯЗИ

Реальный канал связи представляет собой систему устройств для преобразования сигналов и физическую среду, в которой эти сигналы передаются (хранятся). Сигналы, передаваемые по реальным каналам связи, подвергаются искажениям вследствие различных шумов и помех. Физическая природа сигналов и механизм возникновения искажений не будут рассматриваться в данном учебном пособии, предметом рассмотрения станет лишь математическая модель канала связи.

Определение 9.1. *Дискретный канал связи задан, если даны два конечных множества* $\mathcal{X} = \{x^{(1)}, \ldots, x^{(q)}\}$ *(входной алфавит) и* $\mathcal{Y} = \{y^{(1)}, \ldots, y^{(s)}\}$ *(выходной алфавит) и для любого* $n \in \mathbb{N}$, *любых* $(x_1, \ldots, x_n) \in \mathcal{X}^n$, $(y_1, \ldots y_n) \in \mathcal{Y}^n$ *определены переходные вероятности*

$$\pi^{(n)}(y_1, \ldots, y_n | x_1, \ldots, x_n) \geqslant 0$$

так, что справедливо равенство

$$\sum_{(y_1, \ldots y_n) \in \mathcal{Y}^n} \pi^{(n)}(y_1, \ldots, y_n | x_1, \ldots, x_n) = 1.$$

Далее для краткости записи будем использовать обозначения

$$x^n = (x_1, \ldots, x_n) \in \mathcal{X}^n, \quad y^n = (y_1, \ldots, y_n) \in \mathcal{Y}^n.$$

Определение 9.2. *Случайным входным вектором канала является любой случайный вектор* X^n, *определенный на множестве* $\mathcal{X}^n$.

Случайный вектор X^n задается распределением вероятностей $U_{1, \ldots, n}(x^n) = \mathsf{P}\{X^n = x^n\}$.

Определение 9.3. *Случайным выходным вектором канала называется такой случайный вектор* Y^n, *определенный на множестве* $\mathcal{Y}^n$, *что совместное*

распределение вероятностей пары случайных векторов (X^n, Y^n) *имеет вид*

$$\mathsf{P}\{X^n = x^n, Y^n = y^n\} = U_{1,\dots,n}(x^n)\pi^{(n)}(y^n|x^n). \tag{9.1}$$

Из (9.1) следует, что случайный вектор $Y^n \in \mathcal{Y}^n$ имеет распределение вероятностей

$$V_{1,\dots,n}(y^n) = \mathsf{P}\{Y^n = y^n\} = \sum_{x^n \in \mathcal{X}^n} U_{1,\dots,n}(x^n)\pi^{(n)}(y^n|x^n). \tag{9.2}$$

Кроме того, из (9.1) вытекает соотношение

$$\mathsf{P}\{Y^n = y^n | X^n = x^n\} = \pi^{(n)}(y^n|x^n).$$

Таким образом, переходная вероятность $\pi^{(n)}(y^n|x^n)$ есть условная вероятность появления на выходе канала вектора y^n при условии, что на вход канала поступил вектор x^n.

Определение 9.4. *Дискретный канал связи называется каналом без памяти, если для любого* $n \in \mathbb{N}$ *и любых* $(x_1, \dots, x_n) \in \mathcal{X}^n$, $(y_1, \dots, y_n) \in \mathcal{Y}^n$ *переходная вероятность* $\pi^{(n)}(y^n|x^n)$ *представима в виде*

$$\pi^{(n)}(y^n|x^n) = \prod_{i=1}^{n} \pi_i(y_i|x_i). \tag{9.3}$$

Лемма 9.1. *Для дискретного канала без памяти (сокращенно ДКБП), для любого* $k \in \mathbb{N}$ *и любого* $x_k \in \mathcal{X}$ *выполнено соотношение*

$$\sum_{y_k \in \mathcal{Y}} \pi_k(y_k|x_k) = 1.$$

Доказательство. Докажем это утверждение методом математической индукции по k. Пусть $k = 1$, тогда согласно (9.3) имеем $\pi^{(1)}(y^1|x^1) = \pi_1(y_1|x_1)$, откуда по свойству условной вероятности

$$\sum_{y_1 \in \mathcal{Y}} \pi_1(y_1|x_1) = \sum_{y_1 \in \mathcal{Y}} \pi^{(1)}(y^1|x^1) = 1.$$

Пусть для всех $k < K$ утверждение леммы доказано, докажем утверждение для $k = K$. По индукционному предположению

$$\sum_{y_K \in \mathcal{Y}} \pi_K(y_K|x_K) = \sum_{y_K \in \mathcal{Y}} \pi_K(y_K|x_K) \prod_{i=1}^{K-1} \left(\sum_{y_i \in \mathcal{Y}} \pi_i(y_i|x_i) \right) =$$

$$= \sum_{y^K \in \mathcal{Y}^K} \pi_1(y_1|x_1) \cdots \pi_n(y_n|x_n) = \sum_{y^K \in \mathcal{Y}^K} \pi^{(K)}(y^K|x^K) = 1. \qquad \square$$

Через X_k обозначим k-й символ случайного вектора $X^n = (X_1, \dots, X_n)$, а через Y_k — k-й символ случайного вектора $Y^n = (Y_1, \dots, Y_n)$. Пусть $\mathrm{P}^{(k)}(x_k)$ — распределение k-го символа случайного вектора $X^n = (X_1, \dots, X_n)$, т. е.

$$\mathrm{P}^{(k)}(x_k) = \mathsf{P}\{X_k = x_k\} = \sum_{x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n \in \mathcal{X}^{n-1}} U_{1,\dots,n}(x^n).$$

Лемма 9.2. *Для дискретного канала без памяти и для любых $k, n \in \mathbb{N}$, $k \leqslant n$, справедливо соотношение*

$$\mathsf{P}\{X_k = x_k, Y_k = y_k\} = \mathrm{P}^{(k)}(x_k)\pi_k(y_k|x_k).$$

Доказательство. По свойству распределения вероятностей имеем

$$\begin{aligned}
&\mathsf{P}\{X_k = x_k, Y_k = y_k\} = \\
&= \sum_{(x_1,\ldots,x_{k-1},x_{k+1},\ldots,x_n)\in\mathcal{X}^{n-1}} \sum_{(y_1,\ldots,y_{k-1},y_{k+1},\ldots,y_n)\in\mathcal{Y}^{n-1}} \mathsf{P}\{X^n = x^n, Y^n = y^n\} = \\
&= \sum_{(x_1,\ldots,x_{k-1},x_{k+1},\ldots,x_n)\in\mathcal{X}^{n-1}} \sum_{(y_1,\ldots,y_{k-1},y_{k+1},\ldots,y_n)\in\mathcal{Y}^{n-1}} U_{1,\ldots,n}(x^n)\pi^{(n)}(y^n|x^n) = \\
&= \sum_{(x_1,\ldots,x_{k-1},x_{k+1},\ldots,x_n)\in\mathcal{X}^{n-1}} U_{1,\ldots,n}(x^n) \sum_{(y_1,\ldots,y_{k-1},y_{k+1},\ldots,y_n)\in\mathcal{Y}^{n-1}} \prod_{i=1}^{n}\pi_i(y^i|x^i).
\end{aligned} \tag{9.4}$$

Учитывая, что по лемме 9.1 $\sum_{y_i\in\mathcal{Y}}\pi_i(y_i|x_i) = 1$, продолжим (9.4):

$$\begin{aligned}
&\mathsf{P}\{X_k = x_k, Y_k = y_k\} = \\
&= \sum_{(x_1,\ldots,x_{k-1},x_{k+1},\ldots,x_n)\in\mathcal{X}^{n-1}} U_{1,\ldots,n}(x^n)\pi_k(y^k|x^k) = \mathrm{P}^{(k)}(x_k)\pi_k(y^k|x^k).
\end{aligned}$$

□

Замечание 9.1. Если x_k таково, что $\mathrm{P}^{(k)}(x_k) > 0$, то утверждение леммы (9.2) можно переписать в следующем виде:

$$\pi_k(y_k|x_k) = \frac{\mathsf{P}\{X_k = x_k, Y_k = y_k\}}{\mathrm{P}^{(k)}(x_k)} = \mathsf{P}\{Y_k = y_k|X_k = x_k\}.$$

Таким образом, $\pi_k(y_k|x_k)$ можно проинтерпретировать как вероятность того, что на шаге k на выход поступил символ y_k при условии, что на вход канала поступил символ x_k, и, кроме того, эта вероятность не зависит от того, каковы были переходы до шага k и после шага k (это и означает, что канал как бы без памяти). Вероятности $\pi_k(y_k|x_k)$ удобно задавать в виде стохастических матриц размером $q \times s$, где на пересечении i-й строки и j-го столбца находится вероятность $\pi_k(y^{(j)}|x^{(i)})$.

Следствие 9.1. *Для дискретного канала без памяти и для любых $k, n \in \mathbb{N}$ $(k \leqslant n)$ справедливо соотношение*

$$\mathsf{P}\{Y_k = y_k\} = \sum_{x_k\in\mathcal{X}} \mathrm{P}^{(k)}(x_k)\pi(y_k|x_k).$$

Доказательство. Следует из леммы 9.2 и свойства самосогласованности. □

Определение 9.5. *Дискретный канал без памяти называется стационарным, если $\pi_1(\cdot) = \pi_2(\cdot) = \ldots$, т. е. переходные вероятности символов не зависят от номера шага.*

Таким образом, для задания стационарного дискретного канала без памяти необходимо задать тройку $(\mathcal{X}, \mathcal{Y}, \pi)$, где $\mathcal{X} = \{x^{(1)}, \ldots, x^{(q)}\}$ — входной алфавит, $\mathcal{Y} = \{y^{(1)}, \ldots, y^{(s)}\}$ — выходной алфавит и $\pi = (\pi_{ij}) = (\pi(y^{(j)}|x^{(i)}))$ — стохастическая матрица размером $q \times s$.

Лемма 9.3. *Если распределение $U_{1,\ldots,n}$ таково, что координаты случайного вектора X^n независимы, то для стационарного дискретного канала связи без памяти координаты случайного вектора Y^n также будут независимы.*

Доказательство. Согласно условию леммы

$$U_{1,\ldots,n}(x^n) = \prod_{i=1}^{n} P^{(i)}(x_i),$$

откуда

$$V_{1,\ldots,n}(y^n) = \mathsf{P}\{Y^n = y^n\} = \sum_{x^n \in \mathcal{X}^n} U_{1,\ldots,n}(x^n)\pi^{(n)}(y^n|x^n) =$$

$$= \sum_{x^n \in \mathcal{X}^n} \prod_{i=1}^{n} P^{(i)}(x_i)\pi(y_i|x_i) = \prod_{i=1}^{n} \sum_{x_i \in \mathcal{X}} P^{(i)}(x_i)\pi(y_i|x_i) = \prod_{i=1}^{n} \mathsf{P}\{Y_i = y_i\}. \qquad \square$$

Упражнение 9.1. Останется ли верным утверждение леммы 9.3 если опустить условие стационарности дискретного канала связи без памяти?

Канал связи $(\mathcal{X}, \mathcal{Y}, \pi)$ удобно изображать графически в виде двудольного ориентированного графа, в котором одна доля соответствует символам входного алфавита $\mathcal{X}$, вторая — символам выходного алфавита $\mathcal{Y}$, и если $\pi_{ij} > 0$, то из вершины $x^{(i)}$ в вершину $y^{(j)}$ выходит ориентированная дуга с пометкой π_{ij}.

Пример 9.1. *Двоичный симметричный канал* с параметром p, или сокращенно ДСК(p), задается входным алфавитом $\mathcal{X} = \{0, 1\}$, выходным алфавитом $\mathcal{Y} = \{0, 1\}$ и матрицей $\pi = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$, где параметр $0 \leqslant p \leqslant 1$.

Таким образом, при передаче символа по ДСК может происходить правильная передача или искажение (замена на противоположный символ), а параметр p представляет собой вероятность искажения символа при передаче. Графически ДСК(p) можно представить как диаграмму (рис. 9.1).

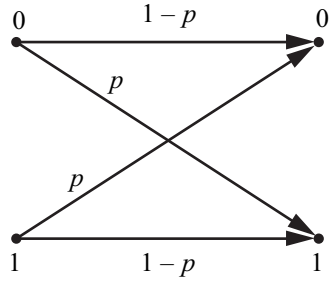


Рис. 9.1. Двоичный симметричный канал с параметром p

Пример 9.2. *Двоичный канал со стиранием* с параметром p задается входным алфавитом $\mathcal{X} = \{0,1\}$, выходным алфавитом $\mathcal{Y} = \{0,E,1\}$ и матрицей $\pi = \begin{pmatrix} 1-p & p & 0 \\ 0 & p & 1-p \end{pmatrix}$, где параметр $0 \leqslant p \leqslant 1$.

При передаче символа по данному каналу может происходить правильная передача или искажение (стирание), выходной символ E (от англ. erase) обозначает результат стирания входного символа, а параметр p представляет собой вероятность стирания. Графически канал со стиранием можно представить как диаграмму (рис. 9.2, *а*).

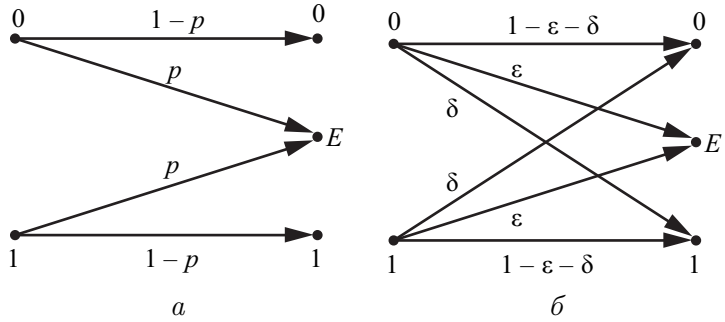


Рис. 9.2. Двоичные каналы: *а* – со стиранием с параметром p; *б* – с искажением и стиранием с параметрами ε и δ

Пример 9.3. *Канал с искажением и стиранием* с двумя параметрами ε, δ задается входным алфавитом $\mathcal{X} = \{0,1\}$, выходным алфавитом $\mathcal{Y} = \{0,E,1\}$ и матрицей $\pi = \begin{pmatrix} 1-\varepsilon-\delta & \varepsilon & \delta \\ \delta & \varepsilon & 1-\varepsilon-\delta \end{pmatrix}$. Параметры канала удовлетворяют неравенствам $\varepsilon \geqslant 0$, $\delta \geqslant 0$, $\varepsilon + \delta \leqslant 1$.

При передаче символа по данному каналу может происходить правильная передача, искажение (на входе 0, а на выходе 1, или наоборот) либо стирание (на выходе символ стирания E). Параметр δ представляет собой вероятность искажения переданного символа, а параметр ε — вероятность стирания. Графически данный канал можно представить как диаграмму (рис. 9.2, *б*).

9.2. ПРОПУСКНАЯ СПОСОБНОСТЬ ДИСКРЕТНОГО КАНАЛА БЕЗ ПАМЯТИ

Определение 9.6. *Пропускной способностью дискретного канала связи без памяти называется величина*

$$C^* = \sup_{n\in\mathbb{N}} \left(\frac{1}{n} \sup_{U_{1,\ldots,n}} \mathsf{I}\{X^n, Y^n\} \right),$$

где внутренний супремум берется по всем возможным распределениям $U_{1,\ldots,n}$ случайного вектора X^n, $\mathsf{I}\{X^n, Y^n\}$ — количество информации по Шеннону, со-

держащейся в случайной символьной последовательности Y^n относительно входной последовательности X^n.

Можно сказать, что пропускная способность определяет наибольшее допустимое количество полезной информации, приходящейся на один символ, которое может быть передано по каналу связи от отправителя получателю.

Замечание 9.2. Величину $\mathsf{I}\{X^n, Y^n\}$ можно рассматривать как действительную функцию, аргументом которой является распределение $U_{1,\dots,n}$ на множестве $\mathcal{X}^n$ (распределение случайного вектора Y^n однозначно определяется распределением $U_{1,\dots,n}(\cdot)$ согласно (9.2)). Поскольку эта функция непрерывна, а класс всех распределений $U_{1,\dots,n}(\cdot)$ на $\mathcal{X}^n$ представляет собой замкнутое и ограниченное множество в евклидовом пространстве $\mathbb{R}^{q^n}$, то для каждого $n \geqslant 1$ по теореме Вейерштрасса точная верхняя грань $\sup_{U_{1,\dots,n}} \mathsf{I}\{X^n, Y^n\}$ достигается на некотором распределении $U^*_{1,\dots,n}(\cdot)$. Такое распределение $U^*_{1,\dots,n}(\cdot)$ (при заданном n) называют *оптимальным входным распределением*, и при этом канал связи используется наилучшим образом, так что в сообщениях длиной n по этому каналу передается максимально возможное количество информации.

Теорема 9.1. *Для стационарного дискретного канала связи без памяти справедливо равенство*

$$C^* = \max_u \mathsf{I}\{X, Y\},$$

где $u(x) = \mathsf{P}\{X = x\}$, и максимум берется по всем возможным распределениям $u(\cdot)$ случайного вектора $X = X^1$ на множестве $\mathcal{X}$.

Доказательство. Представим x^n в виде $x^n = (x^{(i_1)}, \dots, x^{(i_n)})$ и y^n в виде $y^n = (y^{(j_1)}, \dots, y^{(j_n)})$. Условную энтропию $\mathsf{H}\{Y^n|X^n\}$ для стационарного дискретного канала связи без памяти можно записать как

$$\mathsf{H}\{Y^n|X^n\} = -\sum_{x^n\in\mathcal{X}^n,\ y^n\in\mathcal{Y}^n} \mathsf{P}\{X^n = x^n, Y^n = y^n\} \log_2 \pi^{(n)}(y^n|x^n) =$$

$$= -\sum_{k=1}^{n} \sum_{x^n\in\mathcal{X}^n,\ y^n\in\mathcal{Y}^n} U_{1,\dots,n}(x^n)\pi^{(n)}(y^n|x^n) \log_2 \pi(y^{(j_k)}|x^{(i_k)}) =$$

$$= -\sum_{k=1}^{n} \sum_{x^{(i_k)}\in\mathcal{X}} \sum_{y^{(j_k)}\in\mathcal{Y}} \mathrm{P}^{(k)}(x^{(i_k)})\pi(y^{(j_k)}|x^{(i_k)}) \log_2 \pi(y^{(j_k)}|x^{(i_k)}). \qquad (9.5)$$

Поскольку условная энтропия $\mathsf{H}\{Y_k|X_k\}$ k-го символа случайного вектора Y^n при условии X_k имеет вид

$$\mathsf{H}\{Y_k|X_k\} = -\sum_{x^{(i_k)}\in\mathcal{X}} \sum_{y^{(j_k)}\in\mathcal{Y}} \mathrm{P}^{(k)}(x^{(i_k)})\pi(y^{(j_k)}|x^{(i_k)}) \log_2 \pi(y^{(j_k)}|x^{(i_k)}), \qquad (9.6)$$

то, подставив (9.6) в (9.5), получим

$$\mathsf{H}\{Y^n|X^n\} = \sum_{k=1}^{n} \mathsf{H}\{Y_k|X_k\}.$$

Из свойства полуаддитивности энтропии

$$\mathsf{H}\{Y^n\} \leqslant \sum_{k=1}^{n} \mathsf{H}\{Y_k\},$$

следовательно,

$$\mathsf{I}\{X^n, Y^n\} = \mathsf{H}\{Y^n\} - \mathsf{H}\{Y^n|X^n\} \leqslant$$

$$\leqslant \sum_{k=1}^{n} (\mathsf{H}\{Y_k\} - \mathsf{H}\{Y_k|X_k\}) = \sum_{k=1}^{n} \mathsf{I}\{X_k, Y_k\}. \tag{9.7}$$

Заметим, что согласно лемме 9.2 величину $\mathsf{I}\{X_k, Y_k\}$ можно рассматривать как действительную функцию, аргументом которой является распределение $\mathrm{P}^{(k)}$. Тогда для любого распределения $U_{1,\ldots,n}(\cdot)$

$$\sum_{k=1}^{n} \mathsf{I}\{X_k, Y_k\} \leqslant \sum_{k=1}^{n} \max_{\mathrm{P}^{(k)}} \mathsf{I}\{X_k, Y_k\} = n \max_{u} \mathsf{I}\{X, Y\}. \tag{9.8}$$

Докажем, что указанная верхняя граница в (9.8) достигается. Пусть распределение $U_{1,\ldots,n}(\cdot)$ таково, что координаты случайного вектора X^n независимы, тогда согласно лемме 9.3 координаты случайного вектора Y^n также будут независимы и, следовательно, по свойству энтропии неравенство (9.7) обратится в равенство. Пусть $u^*(\cdot)$ — это такое распределение, при котором достигается максимум $\mathsf{I}\{X_k, Y_k\}$. Тогда если выбрать $\mathrm{P}^{(1)}(\cdot) = \mathrm{P}^{(2)}(\cdot) = \ldots = \mathrm{P}^{(n)}(\cdot) = u(\cdot)$, то неравенство (9.8) также обратится в равенство. Таким образом, имеем

$$\mathsf{I}\{X^n, Y^n\} = n \max_{u} \mathsf{I}\{X, Y\},$$

откуда очевидно, что при любом $n \in \mathbb{N}$

$$\frac{1}{n} \sup_{U_{1,\ldots,n}} \mathsf{I}\{X^n, Y^n\} = \max_{u} \mathsf{I}\{X, Y\}.$$

□

В общем случае вычисление пропускной способности C^* для произвольного стационарного канала связи без памяти $(\mathcal{X}, \mathcal{Y}, \pi)$ представляет собой сложную задачу, которую можно сформулировать как задачу поиска максимума неотрицательной функции $\mathsf{I}\{X, Y\}$, зависящей от q действительных переменных $u_1, \ldots, u_n$ ($u_i = \mathsf{P}\{X = x^{(i)}\}$), при условии совокупности ограничений, состоящих из одного равенства $u_1 + \ldots + u_n = 1$ и q неравенств $u_1 \geqslant 0$, $i \in \{1, \ldots, q\}$.

Приведем без доказательства теорему о необходимых и достаточных условиях, при которых набор $u = (u_1, \ldots, u_n)$ обращает в максимум функцию $\mathsf{I}\{X, Y\}$ [33]. Нам понадобится обозначение

$$\mathsf{I}\left\{x^{(i)}, Y\right\} = \sum_{j=1}^{s} \pi_{ij} \log \frac{\pi_{ij}}{\sum_{i=1}^{q} u_i \pi_{ij}},\ x^{(i)} \in \mathcal{X}.$$

Теорема 9.2. *Пусть задан стационарный дискретный канал связи без памяти $(\mathcal{X}, \mathcal{Y}, \pi)$. Распределение $u = (u_1, \ldots, u_n)$ обращает в максимум величину $\mathsf{I}\{X, Y\}$ в том и только том случае, когда существует такое число C, что*

$$\text{если } u_i > 0, \text{ то } \mathsf{I}\left\{x^{(i)}, Y\right\} = C,$$
$$\text{если } u_i = 0, \text{ то } \mathsf{I}\left\{x^{(i)}, Y\right\} \leqslant C.$$

При этом пропускная способность C^* равна числу C.

Если дано входное распределение $u = (u_1, \ldots, u_n)$ и предполагается, что оно оптимально для рассматриваемого канала связи, то условия теоремы 9.2 позволяют легко проверить это предположение и тем самым найти C^*. Если же распределение $u = (u_1, \ldots, u_n)$ неизвестно, то условия теоремы дают систему уравнений относительно распределения u. Эта система уравнений является трансцендентной, и в общем случае способ ее аналитического решения неизвестен. В одних случаях система упрощается, и ее удается решить аналитически. В других для нахождения пропускной способности C^* и оптимального входного распределения применяют численные итерационные алгоритмы.

9.3. СИММЕТРИЧНЫЕ КАНАЛЫ

Рассмотрим некоторые частные случаи, когда вычисление пропускной способности C^* для стационарного дискретного канала связи без памяти $(\mathcal{X}, \mathcal{Y}, \pi)$ существенно упрощается.

Определение 9.7. *Дискретный стационарный канал связи без памяти $(\mathcal{X}, \mathcal{Y}, \pi)$ называется симметричным по входу, если все строки матрицы $\pi \in \in \mathbb{R}^{q\times s}$ являются перестановками одного и того же набора чисел $\pi_1^1, \ldots, \pi_s^1$.*

Утверждение 9.1. Если канал связи симметричен по входу, то

$$C^* \leqslant \log s + \sum_{i=1}^{s} \pi_i^1 \log \pi_i^1. \tag{9.9}$$

Доказательство. Поскольку $|\mathcal{Y}| = s$, то верно неравенство $\mathsf{H}\{Y\} \leqslant \log s$, причем равенство достигается только для равномерного выходного распределения $v = (1/s, \ldots, 1/s)$.

Учитывая, что канал связи симметричен по входу, вычислим

$$\mathsf{H}\{Y|X\} = -\sum_{i=1}^{q}\sum_{j=1}^{s} u_i \pi_{ij} \log \pi_{ij} =$$
$$= -\sum_{i=1}^{q} u_i \sum_{j=1}^{s} \pi_{ij} \log \pi_{ij} = -\sum_{i=1}^{q} u_i \sum_{j=1}^{s} \pi_j^1 \log \pi_j^1 =$$
$$= -\sum_{j=1}^{s} \pi_j^1 \log \pi_j^1 \left(\sum_{i=1}^{q} u_i\right) = -\sum_{j=1}^{s} \pi_j^1 \log \pi_j^1. \tag{9.10}$$

Далее согласно теореме 9.1 имеем

$$C^* = \max_u \mathsf{I}\{X, Y\} = \max_u (\mathsf{H}\{Y\} - \mathsf{H}\{Y|X\}) =$$

$$= \max_u (\log s + \sum_{j=1}^{s} \pi_j^1 \log \pi_j^1).$$

Вследствие того, что выражение в скобках не зависит от распределения u, получим требуемое соотношение. □

Для каналов, симметричных по входу, шум в канале в одинаковой степени нарушает передачу каждого из возможных входных символов.

Определение 9.8. *Дискретный стационарный канал связи без памяти* $(\mathcal{X}, \mathcal{Y}, \pi)$ *называется симметричным по выходу, если все столбцы матрицы* $\pi \in \mathbb{R}^{q \times s}$ *являются перестановками одного и того же набора чисел* $\pi_1^2, \ldots, \pi_q^2$.

Утверждение 9.2. Если канал связи симметричен по выходу, то равномерному распределению $u = (1/q, \ldots, 1/q)$ на входе соответствует равномерное распределение $v = (1/s, \ldots, 1/s)$ на выходе.

Доказательство. Для $1 \leqslant j \leqslant s$

$$v_j = \sum_{i=1}^{q} u_i \pi_{ij} = \frac{1}{q} \sum_{i=1}^{q} \pi_{ij} = \frac{1}{q} \sum_{i=1}^{q} \pi_i^2,$$

т. е. v_j не зависит от j. А так как $\sum_{j=1}^{s} v_j = 1$, то получим, что $v_j = 1/s$. □

Определение 9.9. *Дискретный стационарный канал связи без памяти* $(\mathcal{X}, \mathcal{Y}, \pi)$ *называется симметричным, если он симметричен по входу и по выходу.*

Утверждение 9.3. Если канал связи симметричен, то

$$C^* = \log s + \sum_{i=1}^{s} \pi_i^1 \log \pi_i^1.$$

Доказательство. Равенство в (9.9) достигается для равномерного выходного распределения $v = (1/s, \ldots, 1/s)$. Кроме того, верхняя граница в (9.9) не зависит от распределения u. Тогда если в качестве u взять равномерное распределение $u = (1/q, \ldots, 1/q)$, то согласно утверждению 9.2 выходное распределение v тоже будет равномерным $v = (1/s, \ldots, 1/s)$ и для него будет достигаться равенство в (9.9). □

Пример 9.4. Для ДСК с параметром p имеем $q = s = 2$, и в силу утверждения 9.3 (очевидно, что ДСК(p) является симметричным)

$$C^* = \log_2 s + \sum_{i=1}^{s} \pi_i^1 \log_2 \pi_i^1 = \log_2 2 + p \log_2 p + (1-p) \log_2 (1-p) = 1 - h(p),$$

где $h(p) = -p \log_2 p - (1-p) \log_2 (1-p)$. График зависимости пропускной способности от параметра p представлен на рис. 9.3. В данном примере пропускная способность измеряется в битах ($b = 2$).

График показывает, что полученный результат вполне согласуется с логикой. Действительно, если $p = 0$, то искажения в ДСК не происходят и каждый переданный символ безошибочно воспроизводится на выходе канала, при этом

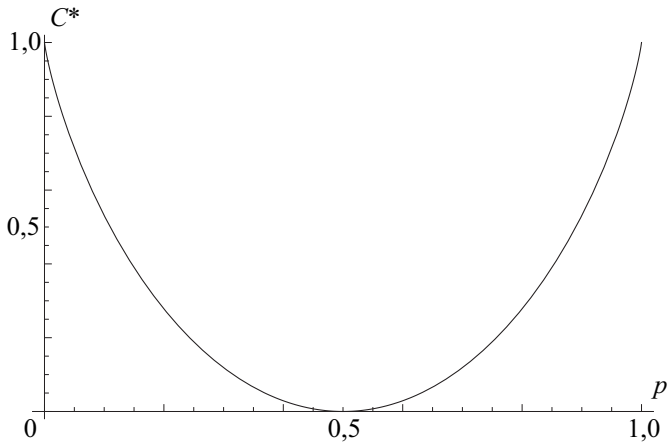


Рис. 9.3. Зависимость пропускной способности ДСК от p

пропускная способность равна $C^* = 1$. Если $p = 1$, то каждый переданный символ наверняка искажается, и поэтому по выходу можно однозначно восстановить вход; в этом случае также $C^* = 1$. Если же $p = 1/2$, то любому переданному символу будет соответствовать случайный выходной символ с равномерным распределением, статистически не зависящий от переданного символа. Разумеется, такой канал не передает никакой полезной информации, и его пропускная способность равна нулю.

Пример 9.5. Рассмотрим двоичный канал со стиранием с параметром p ($b = 2$). Для него имеем $q = 2$, $s = 3$. Этот канал симметричен по входу, но не симметричен по выходу. Пусть задано некоторое входное распределение $u = (\alpha, 1-\alpha)$. Согласно (9.10) получим, что $\mathsf{H}\{Y|X\} = h(p)$ и не зависит от u. Выходное распределение равно $v = ((1-p)\alpha, p, (1-p)(1-\alpha))$. Для энтропии $\mathsf{H}\{Y\}$ по неравенству Йенсена имеем

$$\mathsf{H}\{Y\} = (1-p)\left(\alpha \log_2\left(\frac{1}{(1-p)\alpha}\right) + (1-\alpha)\log_2\left(\frac{1}{(1-p)(1-\alpha)}\right)\right) -$$

$$-p\log_2 p \leqslant (1-p)\log_2\frac{2}{1-p} - p\log_2 p,$$

причем равенство будет достигаться, когда $\alpha = 1/2$. Таким образом,

$$C^*(p) = (1-p)\log_2\frac{2}{1-p} - p\log_2 p + p\log_2 p + (1-p)\log_2(1-p) = (1-p).$$

В этом примере полученный результат вполне согласуется с логикой. Если $p = 0$, то стирания никогда не происходят, каждый переданный символ безошибочно воспроизводится на выходе канала и пропускная способность равна $C^* = 1$. Если $p = 1$, то каждый переданный символ наверняка стирается и по выходу нельзя получить никакой информации о входе. При этом $C^* = 0$. Однако до вычислений трудно было предположить, что пропускная способность линейно зависит от вероятности стирания.

Пример 9.6. Рассмотрим канал с искажением и стиранием с параметрами ε, δ ($b = 2$). Входному распределению $u = (\alpha, 1-\alpha)$ соответствует выходное распределение

$$v = ((1-\delta-\varepsilon)\alpha + \delta(1-\alpha), \varepsilon, (1-\delta-\varepsilon)(1-\alpha) + \delta\alpha).$$

Как и в предыдущем примере, можно показать, что энтропия $\mathbf{H}\{Y\}$ будет принимать свой максимальное значение при $\alpha = 1/2$; при этом распределение v примет вид $v = ((1-\delta)/2, \varepsilon, (1-\delta)/2)$. Учитывая (9.10), вычислим пропускную способность

$$C^*(\varepsilon,\delta) = -2\frac{1-\varepsilon}{2}\log_2\frac{1-\varepsilon}{2} - \varepsilon\log_2\varepsilon + (1-\delta-\varepsilon)\log_2(1-\delta-\varepsilon) + \varepsilon\log_2\varepsilon +$$
$$+\delta\log_2\delta = (1-\varepsilon)(1-\log_2(1-\varepsilon)) + (1-\delta-\varepsilon)\log_2(1-\delta-\varepsilon) +$$
$$+\delta\log_2\delta = (1-\varepsilon)\left(1 - h\left(\frac{\delta}{1-\varepsilon}\right)\right).$$

График зависимости C^* от параметров ε, δ изображен на рис. 9.4.

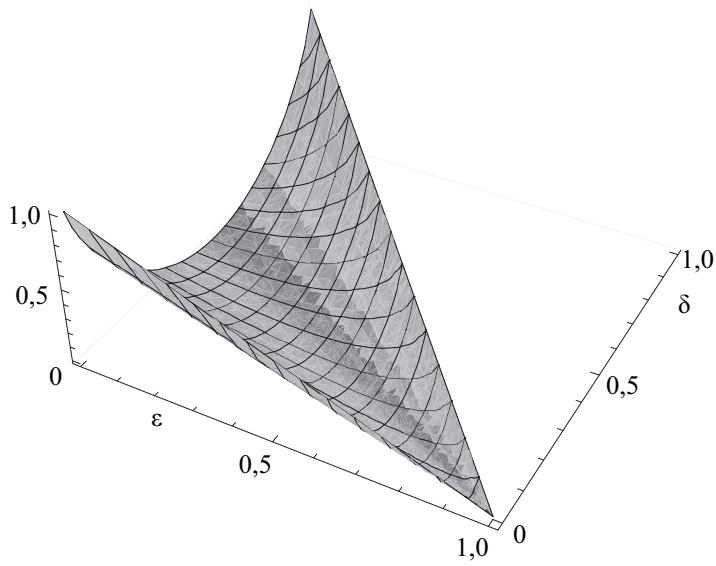


Рис. 9.4. Пропускная способность канала с искажением и стиранием

Рассмотрим некоторые частные случаи. Если $\varepsilon = 0$, то стирания никогда не происходят и придем к случаю ДСК с параметром δ. Подстановка $\varepsilon = 0$ приводит к выражению

$$C^*(0,\delta) = 1 - h(\delta),$$

что совпадает с результатом, полученным в примере 9.4.

Если $\delta = 0$, то искажение никогда не происходит и мы получим случай дискретного канала со стиранием и параметром ε. При $\delta = 0$ имеем

$$C^*(\varepsilon,0) = 1 - \varepsilon,$$

что совпадает с результатом примера 9.5.

Кроме того, на отрезке $0 \leqslant \varepsilon \leqslant 1$, $\varepsilon+\delta = 1$ $h(\delta/(1-\varepsilon)) = 0$, откуда $C^*(\varepsilon,\delta) = 1-\varepsilon$; а на отрезке $0 \leqslant \varepsilon \leqslant 1$, $\varepsilon + 2\delta = 1$ $h(\delta/(1-\varepsilon)) = 1$, откуда $C^*(\varepsilon,\delta) = 0$.

9.4. СОЕДИНЕНИЕ КАНАЛОВ

Если даны два или более каналов связи, то их можно комбинировать различными способами, получая каналы связи с новыми свойствами.

Определение 9.10. *Пусть даны два стационарных дискретных канала связи без памяти* $(\mathcal{X}_1, \mathcal{Y}_1, \pi_1)$ *и* $(\mathcal{X}_2, \mathcal{Y}_2, \pi_2)$, $q_i = |\mathcal{X}_i|$, $s_i = |\mathcal{Y}_i|$, $i = 1,2$. *Тогда канал связи* $(\mathcal{X}, \mathcal{Y}, \pi)$ *называется: 1) последовательным соединением данных каналов, если* $\mathcal{X}_2 = \mathcal{Y}_1$ *и* $\pi = \pi_1 \cdot \pi_2$; *2) параллельным соединением данных каналов, если* $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$, $\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2$ *и переходные вероятности удовлетворяют соотношению*

$$\pi((y_1,y_2)|(x_1,x_2)) = \pi_1(y_1|x_1)\pi_2(y_2|x_2);$$

3) суммой данных каналов, если $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$, $\mathcal{X}_1 \cap \mathcal{X}_2 = \emptyset$, $\mathcal{Y} = \mathcal{Y}_1 \cup \mathcal{Y}_2$, $\mathcal{Y}_1 \cap \mathcal{Y}_2 = \emptyset$ *и матрица* π *имеет блочный вид*

$$\pi = \begin{pmatrix} \pi_1 & 0 \\ 0 & \pi_2 \end{pmatrix}.$$

Последовательное соединение каналов означает, что выходной символ первого канала подается на вход второго канала. При этом равенство $\pi = \pi_1 \cdot \pi_2$ представляет собой просто другую форму записи формулы полной вероятности.

Параллельное соединение каналов означает, что на вход канала поступает пара случайных символов (X_1, X_2), на выходе получается пара случайных символов (Y_1, Y_2), и при этом передача $X_1 \to Y_1$ и $X_2 \to Y_2$ происходит независимо, в соответствии с переходными вероятностями соответствующих каналов.

Сумма каналов означает, что на вход нового канала подается символ либо из алфавита $\mathcal{X}_1$, либо из алфавита $\mathcal{X}_2$, и затем этот символ передается по тому каналу, символом алфавита которого он является.

В следующих трех теоремах устанавливаются соотношения для пропускных способностей последовательного, параллельного соединения и суммы двух каналов связи.

Теорема 9.3. *Если канал связи* $(\mathcal{X}, \mathcal{Y}, \pi)$ *является последовательным соединением двух каналов* $(\mathcal{X}_1, \mathcal{Y}_1, \pi_1)$ *и* $(\mathcal{X}_2, \mathcal{Y}_2, \pi_2)$, *то справедливо неравенство*

$$C^* \leqslant \min\{C_1^*, C_2^*\}.$$

Доказательство. Для начала докажем следующее свойство количества информации по Шеннону: если A, B и C — случайные величины, то

$$\mathsf{I}\{A,(B,C)\} = \mathsf{I}\{A,C\} + \mathsf{I}\{A,B|C\}. \tag{9.11}$$

Для этого выполним несколько преобразований, основанных на свойствах энтропии и определении $\mathsf{I}\{A,B\}$:

$$\mathsf{I}\{A,(B,C)\} = \mathsf{H}\{A\} + \mathsf{H}\{B,C\} - \mathsf{H}\{A,B,C\} =$$

$$= \mathsf{H}\{A\} + \mathsf{H}\{B,C\} - \mathsf{H}\{C\} - \mathsf{H}\{A,B|C\} =$$
$$= \mathsf{H}\{A\} + \mathsf{H}\{C\} + \mathsf{H}\{B|C\} - (\mathsf{H}\{A,C\} - \mathsf{H}\{A|C\}) - \mathsf{H}\{A,B|C\} =$$
$$= (\mathsf{H}\{A\} + \mathsf{H}\{C\} - \mathsf{H}\{A,C\}) + (\mathsf{H}\{A|C\} + \mathsf{H}\{B|C\} - \mathsf{H}\{A,B|C\}) =$$
$$= \mathsf{I}\{A,C\} + \mathsf{I}\{A,B|C\}.$$

Согласно (9.11) можем записать

$$\mathsf{I}\{X_1,Y_1\} + \mathsf{I}\{X_1,Y_2|Y_1\} = \mathsf{I}\{X_1,(Y_1,Y_2)\} =$$
$$= \mathsf{I}\{X_1,Y_2\} + \mathsf{I}\{X_1,Y_1|Y_2\}. \tag{9.12}$$

Далее для $\forall x \in \mathcal{X}_1$, $\forall y \in \mathcal{Y}_1$, $\forall z \in \mathcal{Y}_2$ имеем по определению условной плотности

$$\mathsf{P}\{Y_2 = z|X_1 = x, Y_1 = y\} = \frac{\mathsf{P}\{X_1 = x\}\,\pi_1(y|x)\pi_2(z|y)}{\mathsf{P}\{X_1 = x\}\,\pi_1(y|x)} =$$
$$= \pi_2(z|y) = \mathsf{P}\{Y_2 = z|Y_1 = y\}.$$

Откуда следует, что Y_2 не зависит от X_1 при условии Y_1. Значит,

$$\mathsf{I}\{X_1,Y_2|Y_1\} = 0. \tag{9.13}$$

Кроме того, $\mathsf{I}\{X_1,Y_1|Y_2\} \geqslant 0$, откуда согласно (9.12) получим неравенство

$$\mathsf{I}\{X_1,Y_2\} \leqslant \mathsf{I}\{X_1,Y_1\}. \tag{9.14}$$

Поменяв в соотношении (9.12) величины X_1 и Y_2 местами, имеем

$$\mathsf{I}\{Y_2,X_1\} + \mathsf{I}\{Y_1,Y_2|X_1\} = \mathsf{I}\{Y_2,(Y_1,X_1)\} = \mathsf{I}\{Y_1,Y_2\} + \mathsf{I}\{X_1,Y_2|Y_1\},$$

откуда, учитывая (9.13) и неравенство $\mathsf{I}\{Y_1,Y_2|X_1\} \geqslant 0$, получим

$$\mathsf{I}\{X_1,Y_2\} \leqslant \mathsf{I}\{Y_1,Y_2\}. \tag{9.15}$$

Объединяя результаты (9.14) и (9.15), придем к неравенству

$$\mathsf{I}\{X_1,Y_2\} \leqslant \min\{\mathsf{I}\{Y_1,Y_2\},\mathsf{I}\{X_1,Y_1\}\} \leqslant \min\{C_1^*,C_2^*\}.$$

А так как последнее соотношение верно для любого распределение $u(\cdot)$, то, максимизировав левую часть, получим требуемое неравенство. □

Теорема 9.4. *Если канал связи $(\mathcal{X},\mathcal{Y},\pi)$ является параллельным соединением двух каналов $(\mathcal{X}_1,\mathcal{Y}_1,\pi_1)$ и $(\mathcal{X}_2,\mathcal{Y}_2,\pi_2)$, то справедливо равенство*

$$C^* = C_1^* + C_2^*.$$

Доказательство. Доказательство проводится аналогично доказательству теоремы 9.1. Пусть входная пара случайных величин (X_1,X_2) имеет следующее распределение вероятностей:

$$U_{1,2}(x_1,x_2) = \mathsf{P}\{X_1 = x_1, X_2 = x_2\},$$

тем самым заданы и частные распределения

$$U_1(x_1) = \mathsf{P}\{X_1 = x_1\} = \sum_{x_2 \in \mathcal{X}_2} U_{1,2}(x_1,x_2),$$

$$U_2(x_2) = \mathsf{P}\{X_2 = x_2\} = \sum_{x_1 \in \mathcal{X}_1} U_{1,2}(x_1, x_2).$$

Далее, так же как и при доказательстве теоремы 9.1, получим

$$\mathsf{H}\{Y_1, Y_2 | X_1, X_2\} = \mathsf{H}\{Y_1|X_1\} + \mathsf{H}\{Y_2|X_2\},$$
$$\mathsf{H}\{Y_1, Y_2\} \leqslant \mathsf{H}\{Y_1\} + \mathsf{H}\{Y_2\},$$

причем в последнем неравенстве равенство достигается только если X_1 и X_2 независимы. Следовательно,

$$\mathsf{I}\{(Y_1, Y_2)|(X_1, X_2)\} \leqslant \mathsf{I}\{Y_1|X_1\} + \mathsf{I}\{Y_2|X_2\} \leqslant C_1^* + C_2^*,$$

и равенство достигается, когда случайные величины X_1 и X_2 независимы, а их распределение обращают в максимум величины C_1^*, C_2^* соответственно. □

Теорема 9.5. *Если канал связи $(\mathcal{X}, \mathcal{Y}, \pi)$ является суммой двух каналов $(\mathcal{X}_1, \mathcal{Y}_1, \pi_1)$ и $(\mathcal{X}_2, \mathcal{Y}_2, \pi_2)$, то при $b = 2$ справедливо равенство*

$$2^{C^*} = 2^{C_1^*} + 2^{C_2^*}.$$

Доказательство. Пусть на вход канала связи $(\mathcal{X}, \mathcal{Y}, \pi)$ поступает случайная величина $X \in \mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$ с распределением вероятностей $u = (u_1, \ldots, u_q) \in \mathbb{R}^s$, где $u_i = \mathsf{P}\{X = x^{(i)}\}$, $x^{(i)} \in \mathcal{X}$, а на выходе мы получим случайную величину Y с распределением $v = (v_1, \ldots, v_s) \in \mathbb{R}^s$, $v_i = \mathsf{P}\{Y = y^{(i)}\}$, $y^{(i)} \in \mathcal{Y}$. Причем согласно определению $q = |\mathcal{X}| = |\mathcal{X}_1| + |\mathcal{X}_2| = q_1 + q_2$. Аналогично $s = s_1 + s_2$.

Представим вектор u в виде $u = (u^{(1)}, u^{(2)})$, где

$$u^{(1)} = (u_1, \ldots, u_{q_1}) \in \mathbb{R}^{q_1}, \quad u^{(2)} = (u_{q_1+1}, \ldots, u_q) \in \mathbb{R}^{q_2},$$

и вектор v — как $v = (v^{(1)}, v^{(2)})$, где

$$v^{(1)} = (v_1, \ldots, v_{s_1}) \in \mathbb{R}^{s_1}, \quad v^{(2)} = (v_{s_1+1}, \ldots, v_s) \in \mathbb{R}^{s_2}.$$

Для стационарного временного ряда без памяти имеем

$$v_j = \mathsf{P}\left\{Y = y^{(j)}\right\} = \sum_{x \in \mathcal{X}} \mathsf{P}\{X = x\}\, \mathsf{P}\left\{Y = y^{(j)} | X = x\right\} = \sum_{i=1}^{q} u_i \pi_{ij}$$

или в матричном виде $v = u \cdot \pi$. Учитывая вид матрицы $\pi = \begin{pmatrix} \pi_1 & 0 \\ 0 & \pi_2 \end{pmatrix}$, получим

$$v^{(1)} = u^{(1)}\pi_1, \quad v^{(2)} = u^{(2)}\pi_2.$$

Обозначим через α сумму координат вектора $u^{(1)}$, т. е. $\alpha = \sum_{i=1}^{s_1} u_i$. Поскольку матрица π_1 является стохастической, то сумма элементов вектора v^1 тоже равняется α. Введем в рассмотрение векторы $a^{(1)}$, $a^{(2)}$, $b^{(1)}$, $b^{(2)}$ следующим образом:

$$a^{(1)} = \begin{cases} \dfrac{1}{\alpha} u^{(1)}, & \alpha \neq 0, \\ \left(\dfrac{1}{q_1}, \ldots, \dfrac{1}{q_1}\right) \in \mathbb{R}^{q_1}, & \alpha = 0, \end{cases} \qquad a^{(2)} = \begin{cases} \dfrac{1}{1-\alpha} u^{(2)}, & \alpha \neq 1, \\ \left(\dfrac{1}{q_2}, \ldots, \dfrac{1}{q_2}\right) \in \mathbb{R}^{q_2}, & \alpha = 1, \end{cases}$$

$$b^{(1)} = a^{(1)}\pi_1, \quad b^{(2)} = a^{(2)}\pi_2.$$

В новых обозначениях $u = (\alpha a^{(1)}, (1-\alpha)a^{(2)})$, $v = (\alpha b^{(1)}, (1-\alpha)b^{(2)})$.

Пусть случайные величины X_1 и Y_1 — входной и выходной символ канала $(\mathcal{X}_1, \mathcal{Y}_1, \pi_1)$, а случайные величины X_2, Y_2 — входной и выходной символ канала $(\mathcal{X}_2, \mathcal{Y}_2, \pi_2)$. Тогда можно считать, что символы X_1 и X_2 имеют распределения вероятностей $a^{(1)}$ и $a^{(2)}$ соответственно. Если $\alpha \neq 0$ и $\alpha \neq 1$, то это действительно так. Если $\alpha = 0$, то $\mathsf{P}\{X \in \mathcal{X}_1\} = 0$, и, следовательно, можно задать любое распределение для X_1, для определенности — равномерное. Аналогично для случая $\alpha = 1$. Поскольку символы X_1 и X_2 имеют распределения вероятностей $a^{(1)}$ и $a^{(2)}$, то символы Y_1 и Y_2 — $b^{(1)}$ и $b^{(2)}$ соответственно.

По определению количества информации имеем

$$\mathsf{I}\{X,Y\} = \mathsf{H}\{Y\} - \mathsf{H}\{Y|X\}. \tag{9.16}$$

Вычислим каждое слагаемое в (9.16) по отдельности:

$$\mathsf{H}\{Y\} = -\sum_{j=1}^{s} v_j \log_2 v_j = -\sum_{j=1}^{s_1} v_j \log_2 v_j - \sum_{j=s_1+1}^{s} v_j \log_2 v_j =$$

$$= -\sum_{j=1}^{s_1} \alpha b_j^{(1)} \log_2(\alpha b_j^{(1)}) - \sum_{j=1}^{s_2}(1-\alpha) b_j^{(2)} \log_2((1-\alpha) b_j^{(2)}) =$$

$$= -\alpha \log_2 \alpha - (1-\alpha)\log_2(1-\alpha) - \alpha \sum_{j=1}^{s_1} b_j^{(1)} \log_2 b_j^{(1)} -$$

$$- (1-\alpha) \sum_{j=1}^{s_2} b_j^{(2)} \log_2 b_j^{(2)} = h(\alpha) + \alpha \mathsf{H}\{Y_1\} + (1-\alpha)\mathsf{H}\{Y_2\}, \tag{9.17}$$

где $h(\alpha) = -\alpha \log_2 \alpha - (1-\alpha)\log_2(1-\alpha)$.

Найдем второе слагаемое:

$$\mathsf{H}\{Y|X\} = \sum_{i=1}^{q} u_i \mathsf{H}\{Y|X = x^{(i)}\} =$$

$$= \sum_{i=1}^{q_1} u_i \mathsf{H}\{Y|X = x^{(i)}\} + \sum_{i=q_1+1}^{q} u_i \mathsf{H}\{Y|X = x^{(i)}\} =$$

$$= \alpha \sum_{i=1}^{q_1} a_i^{(1)} \mathsf{H}\{Y|X = x^{(i)}\} + (1-\alpha) \sum_{i=1}^{q_2} a_i^{(2)} \mathsf{H}\{Y|X = x^{(q_1+i)}\} =$$

$$= \alpha \mathsf{H}\{Y_1|X_1\} + (1-\alpha)\mathsf{H}\{Y_2|X_2\}. \tag{9.18}$$

Подставив (9.17) и (9.18) в (9.16), получим

$$\mathsf{I}\{X,Y\} = h(\alpha) + \alpha \mathsf{H}\{Y_1\} + (1-\alpha)\mathsf{H}\{Y_2\} - \alpha \mathsf{H}\{Y_1|X_1\} +$$
$$+ (1-\alpha)\mathsf{H}\{Y_2|X_2\} = h(\alpha) + \alpha \mathsf{I}\{X_1, Y_1\} + (1-\alpha)\mathsf{I}\{X_2, Y_2\} \leqslant$$

$$\leqslant h(\alpha) + \alpha C_1^* + (1-\alpha)C_2^*.$$

Причем равенство достигается, когда при фиксированном α в векторе $u = (\alpha a^{(1)}, (1-\alpha)a^{(2)})$ векторы $a^{(1)}$ и $a^{(2)}$ взяты таким образом, что

$$\max_{a^{(1)}} \mathsf{I}\{X_1, Y_1\} = C_1^*, \quad \max_{a^{(2)}} \mathsf{I}\{X_2, Y_2\} = C_2^*.$$

Таким образом, вычисление пропускной способности суммы двух каналов связи сводится к нахождению максимума функции

$$f(\alpha) = h(\alpha) + \alpha C_1^* + (1-\alpha)C_2^*$$

на промежутке $0 \leqslant \alpha \leqslant 1$. Найдем стационарные точки, для чего решим уравнение

$$f'(\alpha) = \log_2 \frac{1-\alpha}{\alpha} + C_1^* - C_2^* = 0,$$

$$\alpha_0 = \frac{2^{C_1^*}}{2^{C_1^*} + 2^{C_2^*}}.$$

Вычислим значение функции f в этой точке:

$$f(\alpha_0) = -\frac{2^{C_1^*}}{2^{C_1^*} + 2^{C_2^*}} \log_2 \frac{2^{C_1^*}}{2^{C_1^*} + 2^{C_2^*}} - \frac{2^{C_2^*}}{2^{C_1^*} + 2^{C_2^*}} \log_2 \frac{2^{C_2^*}}{2^{C_1^*} + 2^{C_2^*}} +$$

$$+ \frac{C_1^* 2^{C_1^*}}{2^{C_1^*} + 2^{C_2^*}} + \frac{C_2^* 2^{C_2^*}}{2^{C_1^*} + 2^{C_2^*}} = \frac{2^{C_1^*}}{2^{C_1^*} + 2^{C_2^*}} \log_2 (2^{C_1^*} + 2^{C_2^*}) +$$

$$+ \frac{2^{C_2^*}}{2^{C_1^*} + 2^{C_2^*}} \log_2 (2^{C_1^*} + 2^{C_2^*}) = \log_2 (2^{C_1^*} + 2^{C_2^*}).$$

Поскольку это значение больше, чем значения $f(\alpha)$ в крайних точках ($f(0) = C_2^*$, $f(1) = C_1^*$), то $C^* = f(\alpha_0)$. □

Замечание 9.3. Результат теоремы 9.5 легко обобщается на случай произвольного $b > 1$, а именно

$$b^{C^*} = b^{C_1^*} + b^{C_2^*}.$$

9.5. ДЕКОДЕРЫ ОБЩЕГО ВИДА

Пусть даны два конечных множества $\mathcal{X}$ и $\mathcal{Y}$. Положим, что $X^n \in \mathcal{X}^n$ есть случайное слово, поступающее на вход канала связи. Предполагается, что при передаче сообщения по каналу связи в силу наличия помех возможны разного рода искажения сигнала, поэтому обозначим $Y^n \in \mathcal{Y}^n$ – соответствующее случайное слово, полученное на выходе канала. Для описания канала связи зададим вероятности:

$$\pi^{(n)}(y^n | x^n) = \mathsf{P}\{Y^n = y^n | X^n = x^n\},$$

т. е. вероятности того, что на выходе мы получим слово y^n, если на вход канала подавалось слово x^n. В этом случае будем говорить, что задан дискретный канал связи $(\mathcal{X}^n, \mathcal{Y}^n, \pi^{(n)})$.

Определение 9.11. *Произвольное подмножество*

$$\mathcal{C} = \{x^n(1), \ldots, x^n(M)\} \subseteq \mathcal{X}^n$$

называется кодом длиной n и объемом M, состоящим из кодовых слов $x^n(i) = (x_1(i), \ldots, x_n(i))$, $i \in \{1, \ldots, M\}$.

Предположим, что по каналу связи передаются только кодовые слова кода $\mathcal{C}$. Это означает, что распределение $U_{1,\ldots,n}(x^n) = \mathsf{P}\{X^n = x^n\}$ сосредоточено на коде $\mathcal{C}$, т. е.

$$U_{1,\ldots,n}(x^n) \geqslant 0,\ x^n \in \mathcal{C}; \quad U_{1,\ldots,n}(x^n) = 0,\ x^n \notin \mathcal{C}.$$

Определение 9.12. *Скоростью передачи информации по каналу связи $(\mathcal{X}^n, \mathcal{Y}^n, \pi^{(n)})$ — это величина*

$$R_n = \frac{1}{n}\mathsf{H}\{X^n\}.$$

Величина R_n имеет смысл среднего количества информации, приходящегося на один символ передаваемого по каналу кодового слова. В частности, если распределение $U_{1,\ldots,n}(\cdot)$ является равномерным на коде $\mathcal{C}$, то

$$R_n = \frac{\log M}{n}.$$

Определение 9.13. *Декодер общего вида — произвольное отображение $\mathfrak{D} : \mathcal{Y}^n \to W$, где множество решений $W = \mathcal{C} \cup \{\varepsilon\}$ состоит из кодовых слов кода $\mathcal{C}$ и специального элемента ε, имеющего смысл сообщения об ошибке декодирования.*

Пусть $\mathcal{Y}^n = A_1 \sqcup \ldots \sqcup A_M \sqcup A_{M+1}$ — произвольное разбиение множества $\mathcal{Y}^n$ на $M+1$ непересекающихся подмножеств $A_1, \ldots, A_{M+1}$, называемых решающими областями декодера. Если случайный вектор Y^n на выходе канала связи принадлежит решающей области A_i, $1 \leqslant i \leqslant M$, то декодер $\mathfrak{D}$ принимает решение о том, что на входе канала было кодовое слово $X^n = x^n(i)$; если же $Y^n \in A_{M+1}$, то декодер объявляет об ошибке ε, т. е. декодер не может принять решение в пользу какого-либо кодового слова из кода $\mathcal{C}$. Таким образом, чтобы задать декодер $\mathfrak{D}$, необходимо задать разбиение множества $\mathcal{Y}^n$ на $M+1$ подмножество и каждому из этих подмножеств поставить в соответствие один из символов кода $\mathcal{C}$ либо символ ошибки ε.

Определение 9.14. *Предположим, что было передано кодовое слово $X^n = x^n(i) \in \mathcal{C}$, а на выходе канала связи было получено слово $Y^n = y^n \in A_j$, $i \in \{1, \ldots, M\}$, $j \in \{1, \ldots, M+1\}$. Тогда если $i = j$, то принятое слово Y^n правильно декодировано, в противном случае $(i \neq j)$ произошла ошибка декодирования.*

Другими словами, ошибочное декодирование означает, что $\mathfrak{D}(Y^n) \neq X^n$.

Пусть заданы канал связи $(\mathcal{X}^n, \mathcal{Y}^n, \pi^{(n)})$, код $\mathcal{C} \subseteq \mathcal{X}^n$ и декодер $\mathfrak{D}$. Качество полученной системы передачи информации будем измерять вероятностями принятия ошибочных решений при декодировании.

Определение 9.15. *Условной вероятностью ошибочного декодирования при условии, что было передано кодовое слово* $X^n = x^n(i)$*, является величина*

$$\lambda_i = \mathsf{P}\left\{\mathfrak{D}(Y^n) \neq X^n | X^n = x^n(i)\right\}.$$

Средней вероятностью ошибочного декодирования называется величина

$$\lambda = \mathsf{P}\left\{\mathfrak{D}(Y^n) \neq X^n\right\}.$$

Замечание 9.4. Далее, когда будет необходимо подчеркнуть зависимость декодера $\mathfrak{D}$ или средней вероятности ошибочного декодирования λ от длины n кодового слова, будем добавлять к обозначению индекс n и писать $\mathfrak{D}_n$ и $\lambda^{(n)}$.

Утверждение 9.4. Для условных и средней вероятностей ошибочного декодирования справедливы соотношения

$$\lambda_i = 1 - \sum_{y^n \in A_i} \pi^{(n)}(y^n | x^n(i)), \tag{9.19}$$

$$\lambda = 1 - \sum_{i=1}^{M} U_{1,\ldots,n}(x^n(i)) \sum_{y^n \in A_i} \pi^{(n)}(y^n | x^n(i)). \tag{9.20}$$

Доказательство. Вероятность принять решение $\mathfrak{D}(Y^n) = x^n(j)$ при условии, что было передано $X^n = x^n(i)$, равна

$$\mathsf{P}\left\{\mathfrak{D}(Y^n) = x^n(j) | X^n = x^n(i)\right\} = \sum_{y^n \in A_j} \pi^{(n)}(y^n | x^n(i)).$$

Следовательно,

$$\lambda_i = \mathsf{P}\left\{\mathfrak{D}(Y^n) \neq X^n | X^n = x^n(i)\right\} = 1 - \mathsf{P}\left\{\mathfrak{D}(Y^n) = X^n | X^n = x^n(i)\right\} =$$
$$= 1 - \sum_{y^n \in A_i} \pi^{(n)}(y^n | x^n(i)).$$

По формуле полной вероятности

$$\lambda = \sum_{i=1}^{M} \lambda_i U_{1,\ldots,n}(x^n(i)) = \sum_{i=1}^{M} \left(1 - \sum_{y^n \in A_i} \pi^{(n)}(y^n | x^n(i))\right) U_{1,\ldots,n}(x^n(i)) =$$
$$= \sum_{i=1}^{M} U_{1,\ldots,n}(x^n(i)) - \sum_{i=1}^{M} U_{1,\ldots,n}(x^n(i)) \sum_{y^n \in A_i} \pi^{(n)}(y^n | x^n(i)) =$$
$$= 1 - \sum_{i=1}^{M} U_{1,\ldots,n}(x^n(i)) \sum_{y^n \in A_i} \pi^{(n)}(y^n | x^n(i)). \qquad \square$$

Замечание 9.5. В литературе встречаются такие характеристики канала связи, как минимальная и максимальная вероятность ошибочного декодирования [33]:

$$\lambda_{\min} = \min_{1 \leqslant i \leqslant M} \lambda_i, \quad \lambda_{\max} = \max_{1 \leqslant i \leqslant M} \lambda_i.$$

9.6. ПРИМЕРЫ ДЕКОДЕРОВ

В качестве примеров рассмотрим два важных декодера: $\mathfrak{D}_L, \mathfrak{D}_{AP}$.

Декодер $\mathfrak{D}_L$ — декодер по методу максимального правдоподобия. Алгоритм декодера $\mathfrak{D}_L$ состоит в том, что если принято слово y^n, то вычисляются переходные вероятности $\pi^{(n)}(y^n|x^n)$ для всех кодовых слов x^n и в качестве результата декодирования выбирается то кодовое слово, для которого эта вероятность максимальна. Если максимальное значение переходной вероятности достигается для нескольких кодовых слов, то выбирается одно из них по некоторому заранее оговоренному правилу, например наименьшее в смысле лексикографического порядка.

Декодер $\mathfrak{D}_{AP}$ — декодер по методу максимальной апостериорной вероятности. Если $\mathsf{P}\{Y^n = y^n\} > 0$, то определена условная вероятность

$$\mathsf{P}\{X^n = x^n | Y^n = y^n\} = \frac{U_{1,\ldots,n}(x^n)\pi^{(n)}(y^n|x^n)}{\sum\limits_{z^n \in \mathcal{C}} U_{1,\ldots,n}(z^n)\pi^{(n)}(y^n|z^n)}. \tag{9.21}$$

Эту вероятность обычно называют апостериорной вероятностью кодового слова x^n, т. е. вероятностью после опыта (после получения слова y^n из канала связи), в то время как вероятность $\mathsf{P}\{X^n = x^n\} = U_{1,\ldots,n}(x^n)$ называют априорной, т. е. до опыта.

Алгоритм декодера $\mathfrak{D}_{AP}$ состоит в том, что если принято слово y^n, то вычисляются апостериорные вероятности $\mathsf{P}\{X^n = x^n|Y^n = y^n\}$ для всех кодовых слов $x^n \in \mathcal{C}$ и в качестве результата декодирования из них выбирается то, для которого эта вероятность максимальна. Если максимальное значение апостериорной вероятности достигается для нескольких кодовых слов, то, как и выше, выбирается одно из них по некоторому заранее оговоренному правилу.

Отметим, что декодер $\mathfrak{D}_L$, в отличие от декодера $\mathfrak{D}_{AP}$, по определению не зависит от распределения $U_{1,\ldots,n}(x^n)$ на множестве кодовых слов и поэтому является более простым; декодер $\mathfrak{D}_L$ зависит только от кода и от канала связи и может применяться для любого распределения $U_{1,\ldots,n}(x^n)$.

Задача построения декодера заключается в построении решающего правила в дискриминантном анализе. При этом декодер по методу максимального правдоподобия является решающим правилом по методу максимального правдоподобия, а декодер по методу максимальной апостериорной вероятности — байесовским решающим правилом. Байесовское решающее правило и декодер $\mathfrak{D}_{AP}$ оптимальны в смысле минимума средней вероятности ошибочного декодирования.

Утверждение 9.5. Декодер $\mathfrak{D}_{AP}$ обеспечивает минимальную среднюю вероятность λ ошибочного декодирования.

Доказательство. Пусть заданы канал связи $(\mathcal{X}^n, \mathcal{Y}^n, \pi^{(n)})$, код $\mathcal{C} \subseteq \mathcal{X}^n$ и декодер $\mathfrak{D}$ с решающими областями $A_1, \ldots, A_{M+1}$. Запишем вероятность правиль-

ного декодирования в следующем виде:

$$1-\lambda = \mathsf{P}\left\{\mathfrak{D}(Y^n) = X^n\right\} = \sum_{y^n\in\mathcal{Y}^n} \mathsf{P}\left\{Y^n = y^n\right\} \mathsf{P}\left\{\mathfrak{D}(Y^n) = X^n | Y^n = y^n\right\} =$$

$$= \sum_{i=1}^{M} \sum_{y^n\in A_i} \mathsf{P}\left\{Y^n = y^n\right\} \mathsf{P}\left\{X^n = x^n(i) | Y^n = y^n\right\}.$$

Из такого представления ясно, что величина $1-\lambda$ максимальна, если каждая решающая область A_i, $i\in\{1,\ldots,M\}$, декодера $\mathfrak{D}$ состоит из слов y^n, для которых

$$\mathsf{P}\left\{X^n = x^n(i) | Y^n = y^n\right\} \geqslant \mathsf{P}\left\{X^n = x^n(j) | Y^n = y^n\right\}$$

при всех $j\in\{1,\ldots,M\}$. Это условие совпадает с определением декодера $\mathfrak{D}_{AP}$. □

Определение 9.16. *Два декодера $\mathfrak{D}_1$ и $\mathfrak{D}_2$ называются эквивалентными, если $\forall y^n\in\mathcal{Y}^n$ справедливо равенство $\mathfrak{D}_1(y^n) = \mathfrak{D}_2(y^n)$.*

Утверждение 9.6. Если распределение $U_{1,\ldots,n}(x^n)$ на множестве кодовых слов равномерное, то декодеры $\mathfrak{D}_L$ и $\mathfrak{D}_{AP}$ эквивалентны.

Доказательство. Перепишем (9.21) с учетом того, что $U_{1,\ldots,n}(x^n) = 1/M$, $\forall x^n\in\mathcal{C}$,

$$\mathsf{P}\left\{X^n = x^n | Y^n = y^n\right\} = \frac{U_{1,\ldots,n}(x^n)\pi^{(n)}(y^n|x^n)}{\sum\limits_{z^n\in\mathcal{C}} U_{1,\ldots,n}(z^n)\pi^{(n)}(y^n|z^n)} = \frac{\pi^{(n)}(y^n|x^n)}{\sum\limits_{z^n\in\mathcal{C}} \pi^{(n)}(y^n|z^n)}.$$

Поскольку знаменатель в правой части равенства не зависит от x^n, то максимум апостериорной вероятности $\mathsf{P}\left\{X^n = x^n | Y^n = y^n\right\}$ достигается в точке x^n максимума $\pi^{(n)}(y^n|x^n)$. □

9.7. НЕРАВЕНСТВО ФАНО

На неравенстве Фано будет основано доказательство обратной теоремы кодирования для дискретных стационарных каналов. Кроме того, неравенство Фано имеет и самостоятельное значение.

Для краткости записи введем следующее обозначение:

$$\mathsf{P}\left\{x^n(i) | y^n\right\} = \mathsf{P}\left\{X^n = x^n(i) | Y^n = y^n\right\}.$$

Теорема 9.6 (неравенство Фано). *Если выполнено хотя бы одно из условий:*

$$\mathsf{P}\left\{Y^n\in A_{M+1}\right\} = 0 \text{ или } \lambda \leqslant 1 - \frac{1}{M},$$

то справедливо неравенство

$$\mathsf{H}\{X^n | Y^n\} \leqslant h(\lambda) + \lambda\log(M-1), \tag{9.22}$$

где $h(\lambda) = -\lambda\log\lambda - (1-\lambda)\log(1-\lambda)$.

Доказательство. При доказательстве утверждения 9.5 было показано, что

$$1-\lambda=\sum_{i=1}^{M}\sum_{y^n\in A_i}\mathsf{P}\left\{Y^n=y^n\right\}\mathsf{P}\left\{x^n(i)|y^n\right\}.$$

Представим вероятность ошибочного декодирования в следующем виде:

$$\lambda=\mathsf{P}\left\{\mathfrak{D}(Y^n)\neq X^n\right\}=\sum_{y^n\in\mathcal{Y}^n}\mathsf{P}\left\{Y^n=y^n\right\}\mathsf{P}\left\{\mathfrak{D}(y^n)\neq X^n|Y^n=y^n\right\}=$$

$$=\sum_{i=1}^{M}\sum_{y^n\in A_i}\mathsf{P}\left\{Y^n=y^n\right\}\sum_{j\in\{1,\ldots,M\}\setminus\{i\}}\mathsf{P}\left\{x^n(j)|y^n\right\}+\mathsf{P}\left\{Y^n\in A_{M+1}\right\}.$$

Аналогично запишем выражение для энтропии $\mathsf{H}\{X^n|Y^n\}$:

$$\mathsf{H}\{X^n|Y^n\}=-\sum_{x^n\in\mathcal{X}^n,y^n\in\mathcal{Y}^n}\mathsf{P}\left\{X^n=x^n,Y^n=y^n\right\}\log\mathsf{P}\left\{x^n|y^n\right\}=$$

$$=-\sum_{i=1}^{M+1}\sum_{y^n\in A_i}\mathsf{P}\left\{Y^n=y^n\right\}\sum_{j=1}^{M}\mathsf{P}\left\{x^n(j)|y^n\right\}\log\mathsf{P}\left\{x^n(j)|y^n\right\}.$$

Используя полученные выражения, преобразуем следующую разность:

$$\mathsf{H}\{X^n|Y^n\}-h(\lambda)-\lambda\log(M-1)=$$

$$=\mathsf{H}\{X^n|Y^n\}+(1-\lambda)\log(1-\lambda)+\lambda\log\frac{\lambda}{M-1}=$$

$$=-\sum_{i=1}^{M+1}\sum_{y^n\in A_i}\mathsf{P}\left\{Y^n=y^n\right\}\sum_{j=1}^{M}\mathsf{P}\left\{x^n(j)|y^n\right\}\log\mathsf{P}\left\{x^n(j)|y^n\right\}+$$

$$+\sum_{i=1}^{M}\sum_{y^n\in A_i}\mathsf{P}\left\{Y^n=y^n\right\}\mathsf{P}\left\{x^n(i)|y^n\right\}\log(1-\lambda)+$$

$$+\sum_{i=1}^{M}\sum_{y^n\in A_i}\mathsf{P}\left\{Y^n=y^n\right\}\sum_{j\in\{1,\ldots,M\}\setminus\{i\}}\mathsf{P}\left\{x^n(j)|y^n\right\}\log\frac{\lambda}{M-1}+$$

$$+\mathsf{P}\left\{Y^n\in A_{M+1}\right\}\log\frac{\lambda}{M-1}=$$

$$=\sum_{i=1}^{M}\sum_{y^n\in A_i}\mathsf{P}\left\{Y^n=y^n\right\}\mathsf{P}\left\{x^n(i)|y^n\right\}\log\frac{1-\lambda}{\mathsf{P}\left\{x^n(i)|y^n\right\}}+$$

$$+\sum_{i=1}^{M}\sum_{y^n\in A_i}\mathsf{P}\left\{Y^n=y^n\right\}\sum_{j\neq i}\mathsf{P}\left\{x^n(j)|y^n\right\}\log\frac{\lambda}{(M-1)\mathsf{P}\left\{x^n(j)|y^n\right\}}+$$

$$+\sum_{y^n\in A_{M+1}}\mathsf{P}\{Y^n=y^n\}\sum_{j=1}^{M}\mathsf{P}\{x^n(j)|y^n\}\log\frac{\lambda}{(M-1)\mathsf{P}\{x^n(j)|y^n\}}=$$

$$=S_1+S_2+S_3.$$

Оценим каждое из слагаемых по отдельности, используя неравенство $\log p\leqslant(p-1)\log e$:

$$S_1\leqslant\log e\sum_{i=1}^{M}\sum_{y^n\in A_i}\mathsf{P}\{Y^n=y^n\}\left(1-\lambda-\mathsf{P}\{x^n(i)|y^n\}\right),$$

$$S_2\leqslant\log e\sum_{i=1}^{M}\sum_{y^n\in A_i}\mathsf{P}\{Y^n=y^n\}\sum_{j\neq i}\left(\frac{\lambda}{M-1}-\mathsf{P}\{x^n(j)|y^n\}\right)=$$

$$=\log e\sum_{i=1}^{M}\sum_{y^n\in A_i}\mathsf{P}\{Y^n=y^n\}\left(\lambda-1+\mathsf{P}\{x^n(i)|y^n\}\right),$$

$$S_3\leqslant\log e\sum_{y^n\in A_{M+1}}\mathsf{P}\{Y^n=y^n\}\sum_{j=1}^{M}\left(\frac{\lambda}{M-1}-\mathsf{P}\{x^n(j)|y^n\}\right)=$$

$$=\mathsf{P}\{Y^n\in A_{M+1}\}\left(\frac{\lambda M}{M-1}-1\right)\log e.$$

Поскольку верхние оценки для S_1 и S_2 равны по модулю и отличаются лишь знаком, то $S_1+S_2\leqslant 0$, поэтому

$$\mathsf{H}\{X^n|Y^n\}-h(\lambda)-\lambda\log(M-1)\leqslant$$

$$\leqslant\mathsf{P}\{Y^n\in A_{M+1}\}\left(\frac{\lambda M}{M-1}-1\right)\log e. \tag{9.23}$$

Если выполнено $\mathsf{P}\{Y^n\in A_{M+1}\}=0$, то

$$\mathsf{H}\{X^n|Y^n\}-h(\lambda)-\lambda\log(M-1)\leqslant$$

$$\leqslant\mathsf{P}\{Y^n\in A_{M+1}\}\left(\frac{\lambda M}{M-1}-1\right)\log e=0.$$

Если выполнено $\lambda\leqslant 1-1/M$, то $\lambda M/(M-1)\leqslant 1$ и

$$\mathsf{H}\{X^n|Y^n\}-h(\lambda)-\lambda\log(M-1)\leqslant$$

$$\leqslant\mathsf{P}\{Y^n\in A_{M+1}\}\left(\frac{\lambda M}{M-1}-1\right)\log e\leqslant 0. \qquad \square$$

Неравенство (9.23) называют *усиленным неравенством Фано.*

Замечание 9.6. Условие $\mathsf{P}\{Y^n\in A_{M+1}\}=0$ означает, что декодер всегда принимает решение в пользу какого-либо кодового слова и никогда не сообщает об ошибке.

Следствие 9.2. *Если* $\mathsf{P}\{Y^n \in A_{M+1}\} = 0$, *то для обращения неравенства Фано (9.22) в равенство необходимо и достаточно, чтобы были выполнены следующие условия:*

$$\forall i \in \{1, \ldots, M\},\ \forall y^n \in A_i,\ \mathsf{P}\{x^n(i)|y^n\} = 1 - \lambda,$$

$$\mathsf{P}\{x^n(j)|y^n\} = \frac{\lambda}{M-1},\ j \neq i. \tag{9.24}$$

Доказательство. Неравенство $\log p \leqslant (p-1)\log e$ обращается в равенство тогда и только тогда, когда $p = 1$. Таким образом, в доказательстве теоремы 9.6 сумма $S_1 + S_2 = 0$ тогда и только тогда, когда выполнены условия 9.24. Из того, что $\mathsf{P}\{Y^n \in A_{M+1}\} = 0$, следует, что $\forall y^n \in A_{M+1}$ выполнено $\mathsf{P}\{Y^n = y^n\} = 0$, откуда $S_3 = 0$. □

Замечание 9.7. Условия (9.24) следствия 9.2 означают, что для произвольного полученного вектора $\forall y^n \notin A_{M+1}$ декодер может принять решение в пользу правильного кодового слова с вероятностью $1-\lambda$ или в пользу любого из $M-1$ других кодовых слов с одинаковыми вероятностями $\lambda/(M-1)$.

Следствие 9.3. *Если* $\mathsf{P}\{Y^n \in A_{M+1}\} > 0$, *то для обращения неравенства Фано (9.22) в равенство необходимо и достаточно, чтобы были выполнены условия (9.24) и*

$$\lambda = 1 - \frac{1}{M}.$$

Доказательство. При доказательстве следствия 9.2 было показано, что при выполнении условий (9.24) в доказательстве теоремы 9.6 сумма $S_1 + S_2 = 0$. Для достижения равенства в оценке сверху суммы S_3 необходимо и достаточно, чтобы для всех $y^n \in A_{M+1}$ и $1 \leqslant j \leqslant M$ было выполнено условие

$$\mathsf{P}\{x^n(j)|y^n\} = \frac{\lambda}{M-1}.$$

С другой стороны,

$$1 = \sum_{j=1}^{M} \mathsf{P}\{x^n(j)|y^n\} = \frac{\lambda M}{M-1} \Leftrightarrow \lambda = 1 - \frac{1}{M}.$$

Подставив $\lambda = 1 - 1/M$ в (9.23), получим равенство в (9.22). □

Рассмотрим функцию

$$g(\lambda) = h(\lambda) + \lambda \log_2 (M-1)$$

из правой части неравенства Фано при $b = 2$. Имеем $g(0) = 0$, $g(1) = \log_2 (M-1)$.

Найдем стационарные точки

$$g'(\lambda) = \log_2 \frac{1-\lambda}{\lambda} + \log_2 (M-1) = 0 \Leftrightarrow \lambda = 1 - \frac{1}{M}.$$

Получим, что функция $g(\lambda)$ имеет единственную стационарную точку $\lambda_0 = 1 - 1/M$, в которой функция принимает наибольшее значение $g(\lambda_0) = \log_2 M$.

На отрезке $\lambda \in [0, \lambda_0]$ функция $g(\lambda)$ возрастает от 0 до $\log_2 M$, а на отрезке $\lambda \in [\lambda_0, 1]$ — убывает от $\log_2 M$ до $\log_2 (M-1)$. График данной функции изображен на рис. 9.5.

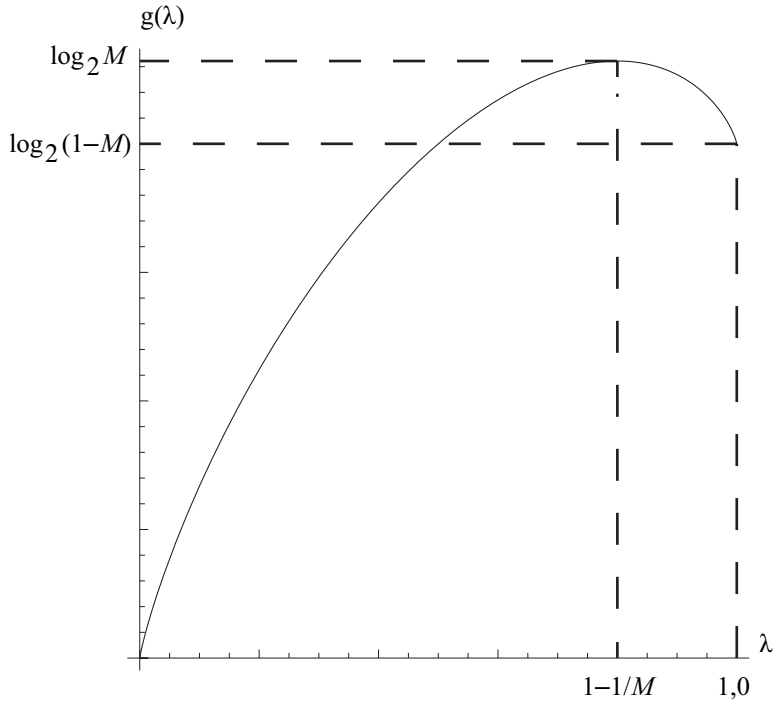


Рис. 9.5. График функции $g(\lambda)$

Замечание 9.8. Пусть на выходе канала связи имеется так называемый случайный декодер, который независимо от принятого из канала вектора $y^n \in \mathcal{Y}^n$ принимает решение, выбирая случайно и равновероятно произвольное кодовое слово из кода $\mathcal{C} = \{x^n(1), \ldots, x^n(M)\}$. При этом, очевидно, правильное декодирование происходит с вероятностью $1/M$, а ошибка декодирования — с вероятностью $\lambda = 1 - 1/M = \lambda_0$. Если декодер не случайный, а устроен некоторым разумным образом, т. е. результат декодирования зависит от принятого вектора и свойств канала связи, то естественно предполагать, что $\lambda \leqslant \lambda_0$.

9.8. ОБРАТНАЯ ТЕОРЕМА КОДИРОВАНИЯ

В данном разделе сформулируем и докажем одну из фундаментальных теорем теории информации. Доказательство проведем согласно [33].

Теорема 9.7 (обратная теорема кодирования). *Пусть* $\varepsilon > 0$ *и* $(\mathcal{X}, \mathcal{Y}, \pi)$ *— стационарный дискретный канал связи без памяти с пропускной способностью* C^*. *Предположим, что для всех* $n \in \mathbb{N}$ *заданы код* $\mathcal{C}^{(n)} \subseteq \mathcal{X}^n$ *и распределение* $U_{1,\ldots,n}$ *на множестве* $\mathcal{X}^n$, *сосредоточенное на коде* $\mathcal{C}^{(n)}$, *что скорость*

передачи $R_n = \mathsf{H}\{X^n\}/n$ *удовлетворяет условию* $R_n \geqslant C^* + \varepsilon$.

Тогда найдется такое число $\delta > 0$, *что для любого* $n \geqslant 1$ *и любого декодера общего вида* $\mathfrak{D}_n$ *средняя вероятность ошибочного декодирования* $\lambda^{(n)}$ *удовлетворяет условию* $\lambda^{(n)} \geqslant \delta$.

Доказательство. По условию теоремы

$$\mathsf{H}\{X^n\} = nR_n \geqslant n(C^* + \varepsilon) > 0.$$

Отсюда, в частности, следует, что число кодовых слов $M_n = |\mathcal{C}^{(n)}| \geqslant 2$. Если бы $M_n = 1$, то это означало, что распределение $U_{1,\ldots,n}$ вырожденное и, следовательно, $\mathsf{H}\{X^n\} = 0$. Тогда для всех $n \in \mathbb{N}$, что $\lambda^{(n)} > 1 - 1/M_n$ выполнено $\lambda > 1/2$. Множество всех таких n обозначим через N_1.

Рассмотрим $n \notin N_1$, для которых $\lambda^{(n)} \leqslant 1 - 1/M_n$. Обозначим $N_2 = \mathbb{N}\backslash N_1$. Согласно неравенству Фано для произвольного декодера $\mathfrak{D}_n$ выполнено

$$\mathsf{H}\{X^n|Y^n\} \leqslant g_n(\lambda^{(n)}) = h(\lambda^{(n)}) + \lambda^{(n)} \log (M_n - 1).$$

Следовательно,

$$\mathsf{I}\{X^n, Y^n\} = \mathsf{H}\{X^n\} - \mathsf{H}\{X^n|Y^n\} \geqslant n(C^* + \varepsilon) - g_n(\lambda^{(n)}).$$

С другой стороны, по определению пропускной способности C^* для любого распределения $U_{1,\ldots,n}(\cdot)$ справедливо

$$\mathsf{I}\{X^n, Y^n\} \leqslant nC^*.$$

Откуда получим, что

$$g_n(\lambda^{(n)}) \geqslant n\varepsilon > 0.$$

Из последнего неравенства, в частности, следует, что по свойствам функции $g(\lambda)$ вероятность ошибочной классификации $\lambda^{(n)} > 0$ для всех $n \in N_2$. Если множество N_2 конечно, тогда в качестве δ можно выбрать наименьший элемент из λ_n, $n \in N_2$. Поэтому далее будем считать, что множество N_2 счетно.

Поскольку $M_n = |\mathcal{C}^{(n)}| \leqslant |\mathcal{X}^n| = q^n$, то

$$g_n(\lambda^{(n)}) \leqslant h(\lambda^{(n)}) + \lambda^{(n)} n \log q.$$

Значит,

$$\varepsilon \leqslant \frac{1}{n} h(\lambda^{(n)}) + \lambda^{(n)} \log q. \tag{9.25}$$

Предположим, что существует такая подпоследовательность $n_k \in N_2$, $k \in \mathbb{N}$, что $\lambda^{(n_k)} \to 0$. Но тогда и правая часть (9.25) тоже стремится к нулю и в некоторый момент k должна стать меньше ε, чего быть не может. Значит, последовательность $\lambda^{(n)}$ такова, что никакая ее подпоследовательность не стремится к нулю. Поэтому найдется искомое $\delta > 0$, что $\lambda^{(n)} > \delta$ для любого $n \in \mathbb{N}$. □

Замечание 9.9. Теорема означает, что если скорость передачи информации больше, чем пропускная способность канала связи, то никаким даже сколь угодно сложным способом кодирования и декодирования нельзя добиться, чтобы вероятность ошибочного декодирования стала меньше определенного положительного числа.

9.9. ПРЯМАЯ ТЕОРЕМА КОДИРОВАНИЯ

Пусть для пары случайных векторов X^n, Y^n задано их совместное распределение

$$p_{X^n,Y^n}(x^n,y^n)=\mathsf{P}\left\{X^n=x^n,Y^n=y^n\right\},\ x^n\in\mathcal{X}^n,\ y^n\in\mathcal{Y}^n,$$

тогда частные распределения могут быть найдены по формулам

$$p_{X^n}(x^n)=\mathsf{P}\left\{X^n=x^n\right\}=\sum_{y^n\in\mathcal{Y}^n}p_{X^n,Y^n}(x^n,y^n),$$

$$p_{Y^n}(y^n)=\mathsf{P}\left\{Y^n=y^n\right\}=\sum_{x^n\in\mathcal{X}^n}p_{X^n,Y^n}(x^n,y^n).$$

Определение 9.17. *Пусть $\varepsilon>0$. Вектор x^n называется ε-типичным, если*

$$\left|\log\frac{1}{p_{X^n}(x^n)}-\mathsf{H}\{X^n\}\right|<n\varepsilon.$$

Вектор y^n называется ε-типичным, если

$$\left|\log\frac{1}{p_{Y^n}(y^n)}-\mathsf{H}\{Y^n\}\right|<n\varepsilon.$$

Пара (x^n,y^n) называется ε-типичной, если ε-типичны векторы x^n, y^n и

$$\left|\log\frac{1}{p_{X^n,Y^n}(x^n,y^n)}-\mathsf{H}\{X^n,Y^n\}\right|<n\varepsilon. \tag{9.26}$$

Множество всех ε-типичных пар обозначим $\mathcal{W}_{n,\varepsilon}$.

Определение 9.18. *Пусть по стационарному каналу связи без памяти $(\mathcal{X},\mathcal{Y},\pi)$ передаются кодовые слова кода $\mathcal{C}=\{x^n(1),\ldots,x^n(M)\}\subseteq\mathcal{X}^n$. Декодером ε-типичных пар назовем такой декодер $\mathfrak{D}_\varepsilon:\mathcal{Y}^n\to\mathcal{C}\cup\{\varepsilon\}$, что если для принятого вектора y^n существует единственное кодовое слово $x^n(i)\in\mathcal{C}$, образующее вместе с y^n ε-типичную пару $(x^n(i),y^n)\in\mathcal{W}_{n,\varepsilon}$, то $\mathfrak{D}_\varepsilon(y^n)=x^n(i)$; в противном случае $\mathfrak{D}_\varepsilon(y^n)=\varepsilon$, т. е. декодер объявляет об ошибке.*

Далее в этом разделе будем полагать, что пары символов $(X_1,Y_1),\ldots,(X_n,Y_n)$ независимы и одинаково распределены, т. е. распределение пары случайных векторов (X^n,Y^n) можно записать как

$$p_{X^n,Y^n}(x^n,y^n)=\mathsf{P}\left\{X^n=x^n,Y^n=y^n\right\}=\prod_{i=1}^n p_{X,Y}(x_i,y_i).$$

В этом случае частные распределения векторов X^n и Y^n имеют вид

$$p_{X^n}(x^n)=\prod_{i=1}^n p_X(x_i),\ x^n\in\mathcal{X}^n,\ p_{Y^n}(y^n)=\prod_{i=1}^n p_Y(y_i),\ y^n\in\mathcal{Y}^n, \tag{9.27}$$

а для энтропий выполнены соотношения

$$\mathsf{H}\{X^n\}=n\mathsf{H}\{X\},\ \mathsf{H}\{Y^n\}=n\mathsf{H}\{Y\},\ \mathsf{H}\{X^n,Y^n\}=n\mathsf{H}\{X,Y\}. \tag{9.28}$$

При сформулированных предположениях о независимости справедливы следующие три леммы о совместной асимптотической равнораспределенности.

Лемма 9.4. *Для стационарного канала связи без памяти при достаточно больших* $n \in \mathbb{N}$

$$\mathsf{P}\{(X^n, Y^n) \in \mathcal{W}_{n,\varepsilon}\} \geqslant 1 - \varepsilon.$$

Доказательство. Обозначим через $\mathcal{X}_{n,\varepsilon}$ и $\mathcal{Y}_{n,\varepsilon}$ множество всех ε-типичных векторов x^n и y^n соответственно, а через $\mathcal{Z}_{n,\varepsilon}$ — множество всех пар (x^n, y^n), для которых выполнено соотношение (9.26). Тогда множество ε-типичных пар представимо в виде

$$\mathcal{W}_{n,\varepsilon} = (\mathcal{X}_{n,\varepsilon} \times \mathcal{Y}_{n,\varepsilon}) \cap \mathcal{Z}_{n,\varepsilon}.$$

Поскольку можно считать, что случайные векторы X^n, Y^n и пара случайных векторов (X^n, Y^n) порождаются источником дискретных сообщений без памяти, который является энтропийно устойчивым, то из обобщенной теоремы Стратоновича следует, что найдутся такие числа n_1, n_2, n_3, что

$$\forall n > n_1 : \ \mathsf{P}\{X^n \notin \mathcal{X}_{n,\varepsilon}\} < \varepsilon/3,$$

$$\forall n > n_2 : \ \mathsf{P}\{Y^n \notin \mathcal{Y}_{n,\varepsilon}\} < \varepsilon/3,$$

$$\forall n > n_3 : \ \mathsf{P}\{(X^n, Y^n) \notin \mathcal{Z}_{n,\varepsilon}\} < \varepsilon/3.$$

Тогда при $n > \max\{n_1, n_2, n_3\}$ имеем

$$\mathsf{P}\{(X^n, Y^n) \notin \mathcal{W}_{n,\varepsilon}\} \leqslant$$

$$\leqslant \mathsf{P}\{(X^n \notin \mathcal{X}_{n,\varepsilon}) \cup (Y^n \notin \mathcal{Y}_{n,\varepsilon}) \cup ((X^n, Y^n) \notin \mathcal{Z}_{n,\varepsilon})\} \leqslant$$

$$\leqslant \mathsf{P}\{X^n \notin \mathcal{X}_{n,\varepsilon}\} + \mathsf{P}\{Y^n \notin \mathcal{Y}_{n,\varepsilon}\} + \mathsf{P}\{(X^n, Y^n) \notin \mathcal{Z}_{n,\varepsilon}\} < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.$$

Перейдя к вероятности обратного события, получим требуемое соотношение. □

Лемма 9.5. *Для стационарного канала связи без памяти при всех* $n \in \mathbb{N}$

$$|\mathcal{W}_{n,\varepsilon}| \leqslant b^{n(\mathbf{H}\{X,Y\}+\varepsilon)},$$

где b *— основание логарифма, соответствующее единице измерения энтропии.*

Доказательство. Если (x^n, y^n) — ε-типичная пара, то согласно (9.26) и (9.28) получим цепочку неравенств:

$$-n\varepsilon < \log_b \frac{1}{p_{X^n,Y^n}(x^n, y^n)} - \mathbf{H}\{X^n, Y^n\} < n\varepsilon,$$

$$n(\mathbf{H}\{X,Y\} - \varepsilon) < \log_b \frac{1}{p_{X^n,Y^n}(x^n, y^n)} < n(\mathbf{H}\{X,Y\} + \varepsilon),$$

$$-n(\mathbf{H}\{X,Y\} - \varepsilon) > \log_b p_{X^n,Y^n}(x^n, y^n) > -n(\mathbf{H}\{X,Y\} + \varepsilon),$$

$$b^{-n(\mathbf{H}\{X,Y\}-\varepsilon)} > p_{X^n,Y^n}(x^n, y^n) > b^{-n(\mathbf{H}\{X,Y\}+\varepsilon)}. \tag{9.29}$$

Применяя правое неравенство (9.29), запишем

$$|\mathcal{W}_{n,\varepsilon}| = \sum_{(x^n,y^n)\in\mathcal{W}_{n,\varepsilon}} 1 < \sum_{(x^n,y^n)\in\mathcal{W}_{n,\varepsilon}} p_{X^n,Y^n}(x^n,y^n) b^{n(\mathbf{H}\{X,Y\}+\varepsilon)} =$$

$$= b^{n(\mathbf{H}\{X,Y\}+\varepsilon)} \mathbf{P}\left\{(X^n,Y^n)\in\mathcal{W}_{n,\varepsilon}\right\} \leqslant b^{n(\mathbf{H}\{X,Y\}+\varepsilon)}. \qquad \square$$

Лемма 9.6. *Если случайные векторы $\tilde{X}^n$ и $\tilde{Y}^n$ независимы и распределены соответственно так же, как векторы X^n и Y^n, то при всех $n\in\mathbb{N}$*

$$\mathbf{P}\left\{(\tilde{X}^n,\tilde{Y}^n)\in\mathcal{W}_{n,\varepsilon}\right\} < b^{-n(\mathbf{I}\{X,Y\}-3\varepsilon)}.$$

Кроме того, при всех достаточно больших $n\in\mathbb{N}$

$$\mathbf{P}\left\{(\tilde{X}^n,\tilde{Y}^n)\in\mathcal{W}_{n,\varepsilon}\right\} > (1-\varepsilon) b^{-n(\mathbf{I}\{X,Y\}+3\varepsilon)}.$$

Доказательство. Если (x^n,y^n) — ε-типичная пара, то аналогично (9.29)

$$b^{-n(\mathbf{H}\{X\}-\varepsilon)} > p_{X^n}(x^n) > b^{-n(\mathbf{H}\{X\}+\varepsilon)}, \tag{9.30}$$

$$b^{-n(\mathbf{H}\{Y\}-\varepsilon)} > p_{Y^n}(y^n) > b^{-n(\mathbf{H}\{Y\}+\varepsilon)}. \tag{9.31}$$

Применяя левые неравенства (9.30) и (9.31) и лемму 9.5, имеем

$$\mathbf{P}\left\{(\tilde{X}^n,\tilde{Y}^n)\in\mathcal{W}_{n,\varepsilon}\right\} = \sum_{(x^n,y^n)\in\mathcal{W}_{n,\varepsilon}} p_{X^n}(x^n) p_{Y^n}(y^n) <$$

$$< \sum_{(x^n,y^n)\in\mathcal{W}_{n,\varepsilon}} b^{-n(\mathbf{H}\{X\}-\varepsilon)} b^{-n(\mathbf{H}\{Y\}-\varepsilon)} = b^{-n(\mathbf{H}\{X\}+\mathbf{H}\{Y\}-2\varepsilon)} |\mathcal{W}_{n,\varepsilon}| \leqslant$$

$$\leqslant b^{-n(\mathbf{H}\{X\}+\mathbf{H}\{Y\}-\mathbf{H}\{X,Y\}-3\varepsilon)} = b^{-n(\mathbf{I}\{X,Y\}-3\varepsilon)}.$$

Первая часть леммы доказана. Применяя правые неравенства (9.30) и (9.31), как и выше, получим

$$\mathbf{P}\left\{(\tilde{X}^n,\tilde{Y}^n)\in\mathcal{W}_{n,\varepsilon}\right\} > b^{-n(\mathbf{H}\{X\}+\mathbf{H}\{Y\}+2\varepsilon)} |\mathcal{W}_{n,\varepsilon}|. \tag{9.32}$$

Применяя левое неравенство (9.29) и лемму 9.4, при достаточно больших $n\in\mathbb{N}$ найдем

$$|\mathcal{W}_{n,\varepsilon}| = \sum_{(x^n,y^n)\in\mathcal{W}_{n,\varepsilon}} 1 > \sum_{(x^n,y^n)\in\mathcal{W}_{n,\varepsilon}} p_{X^n,Y^n}(x^n,y^n) b^{n(\mathbf{H}\{X,Y\}-\varepsilon)} =$$

$$= b^{n(\mathbf{H}\{X,Y\}-\varepsilon)} \mathbf{P}\left\{(X^n,Y^n)\in\mathcal{W}_{n,\varepsilon}\right\} \geqslant (1-\varepsilon) b^{n(\mathbf{H}\{X,Y\}-\varepsilon)}.$$

Подставляя вычисленную оценку в (9.32), получим требуемое соотношение. $\square$

Теорема 9.8 (прямая теорема кодирования). *Пусть $(\mathcal{X},\mathcal{Y},\pi)$ — стационарный канал связи без памяти с пропускной способностью $C^*>0$. Тогда для любых $\alpha>0$, $\beta>0$ при достаточно больших $n\in\mathbb{N}$ существует такой код $\mathcal{C}^{(n)}\subseteq\mathcal{X}^n$ и такой декодер $\mathfrak{D}$, что:*

1) *при равномерном распределении* $U_{1,\dots,n}$ *на* $\mathcal{C}^{(n)}$ *скорость передачи* $R_n = (\log_2 |\mathcal{C}^{(n)}|)/n$ *удовлетворяет двойному неравенству*

$$C^* - \alpha < R_n < C^*,$$

при этом скорость передачи измеряется в битах $b = 2$*;*

2) *максимальная вероятность ошибочного декодирования удовлетворяет неравенству*

$$\lambda_{\max} < \beta.$$

Доказательство. Доказательство проводится согласно [33]. Разобьем его на несколько этапов.

1) *Задание основных параметров.* Выберем произвольное число R такое, что

$$C^* - \alpha < R < C^*, \tag{9.33}$$

и, положим,

$$M_n = [2^{nR}] + 1.$$

Зафиксируем оптимальное распределение $p(x)$ случайного символа $X \in \mathcal{X}$ на входе стационарного канала связи без памяти, для которого достигается пропускная способность канала связи, т. е.

$$\mathsf{I}\{X, Y\} = C^*. \tag{9.34}$$

Выберем число ε так, что

$$0 < \varepsilon < \min\left\{\frac{C^* - R}{3}, \frac{\beta}{4}\right\}, \tag{9.35}$$

и зафиксируем множество $\mathcal{W}_{n,\varepsilon}$ ε-типичных пар.

2) *Построение множества кодов.* Под кодом будем понимать произвольный набор U векторов $x^n(i) = (x_1(i), \dots, x_n(i)) \in \mathcal{X}^n$, $i \in \{1, \dots, M_n\}$. При этом допускается, что некоторые кодовые слова (случайные векторы в наборе) могут совпадать. Это не влияет по существу на результат, но значительно упрощает доказательство.

Источник сообщений выбирает случайное сообщение w, равномерно распределенное на множестве $\{1, \dots, M_n\}$. Выбранному сообщению w кодер ставит в соответствие кодовое слово $X^n = x^n(w)$, которое затем передается по каналу связи $(\mathcal{X}, \mathcal{Y}, \pi)$. К полученному на выходе канала вектору Y^n применяется декодер ε-типичных пар $\mathfrak{D}_\varepsilon$, в результате получается кодовое слово $x^n(v)$ или сообщение ε об ошибке.

Если при декодировании получено сообщение ε об ошибке или получено кодовое слово $x^n(v)$, но $v \neq w$, то произошло ошибочное декодирование. Тогда в силу равномерности распределения w вероятность ошибочного декодирования равна

$$\lambda(U) = \frac{1}{M_n} \sum_{i=1}^{M_n} \lambda_i(U),$$

где

$$\lambda_i(U) = \mathsf{P}\{\mathfrak{D}_\varepsilon(Y^n) \neq X^n | X^n = x^n(i)\} -$$

вероятность ошибочного декодирования при условии, что отправлено сообщение $w = i$.

Будем считать, что символы кодового слова $x^n(i)$ независимы, поэтому согласно заданному распределению

$$Q(x^n(i)) = \mathsf{P}\{X^n = x^n(i)\} = \prod_{j=1}^{n} p(x_j(i)).$$

Кроме того, будем полагать, что кодовые слова в коде тоже независимы, вследствие чего

$$Q(U) = \mathsf{P}\{\mathcal{C} = U\} = \prod_{i=1}^{M_n} Q(x^n(i)).$$

Далее в доказательстве теоремы будем рассматривать множество всевозможных q^{nM_n} кодов $\{U\}$ с заданным распределением вероятностей $Q(U)$. Другими словами, допустим, что последовательность координат кодовых слов случайного кода U порождается дискретным источником без памяти с алфавитом $\mathcal{X}$ и заданным на нем распределением $p(x)$.

3) *Оценка средней вероятности ошибочного декодирования.* Средняя по множеству кодов вероятность ошибочного декодирования имеет вид

$$\bar{\lambda} = \sum_{U} Q(U)\lambda(U), \tag{9.36}$$

где суммирование (усреднение) проводится по всевозможным q^{nM_n} кодам $\{U\}$. Учитывая выражение для $\lambda(U)$, получим

$$\bar{\lambda} = \frac{1}{M_n} \sum_{i=1}^{M_n} \bar{\lambda}_i,$$

где

$$\bar{\lambda}_i = \sum_{U} Q(U)\lambda_i(U),\ i \in \{1, \ldots, M_n\},$$

есть средняя по множеству кодов вероятность ошибочного декодирования при условии, что отправлено сообщение $w = i$.

Пусть $w = 1$. В канал связи отправлено кодовое слово $x^n(1)$, а на выходе получен случайный вектор Y^n. В соответствии с рассматриваемой моделью на множестве кодов случайный вектор Y^n не зависит от кодовых слов $x^n(2), \ldots, x^n(M_n)$.

Обозначим событие, что кодовое слово $X^n(i)$ вместе с вектором Y^n образует ε-типичную пару

$$E_i = \{(x^n(i), Y^n) \in \mathcal{W}_{n,\varepsilon}\},\ i \in \{1, \ldots, M_n\}.$$

Из определения декодера $\mathfrak{D}_\varepsilon$ следует, что ошибочное декодирование при $w = 1$ происходит хотя бы в одном из следующих случаев:

1) отправленное кодовое слова $x^n(1)$ и принятый вектор Y^n не образуют ε-типичную пару (событие $\bar{E}_1$);

2) принятый вектор Y^n образует ε-типичную пару с кодовым словом $X^{(n)}(i)$, $i \in \{2, \ldots, M\}$ (событие $E_2 \cup \ldots \cup E_M$).

Тогда

$$\bar{\lambda}_i = \mathsf{P}\left\{\bar{E}_1 \cup E_2 \cup \ldots \cup E_M\right\} \leqslant \mathsf{P}\left\{\bar{E}_1\right\} + \sum_{i=2}^{M} \mathsf{P}\left\{E_i\right\}.$$

Оценим полученное выражение сверху. Согласно лемме 9.4 при достаточно больших $n \in \mathbb{N}$

$$\mathsf{P}\left\{\bar{E}_1\right\} < \varepsilon.$$

В силу независимости Y^n от кодовых слов $x^n(2), \ldots, x^n(M)$ по лемме 9.6 имеем

$$\mathsf{P}\left\{E_i\right\} < 2^{-n(\mathsf{I}\{X,Y\}-3\varepsilon)}.$$

Следовательно, учитывая значение M_n,

$$\bar{\lambda}_i < \varepsilon + (M-1)2^{-n(\mathsf{I}\{X,Y\}-3\varepsilon)} \leqslant \varepsilon + 2^{-n(\mathsf{I}\{X,Y\}-R-3\varepsilon)}.$$

Аналогичное неравенство можно получить для любого $w = i$, $i \in \{1, \ldots, M_n\}$, поэтому

$$\bar{\lambda} = \frac{1}{M_n}\sum_{i=1}^{M_n} \bar{\lambda}_i \leqslant \varepsilon + 2^{-n(\mathsf{I}\{X,Y\}-R-3\varepsilon)}.$$

В полученной оценке согласно (9.34) и (9.35) множитель при n в показателе степени будет положителен, так как

$$\mathsf{I}\{X,Y\} - R - 3\varepsilon = C^* - R - 3\varepsilon > C^* - R - 3\frac{C^*-R}{3} = 0,$$

а значит, слагаемое $2^{-n(\mathsf{I}\{X,Y\}-R-3\varepsilon)}$ стремится к нулю с ростом n. Следовательно, при достаточно больших n справедливо неравенство

$$\bar{\lambda} < 2\varepsilon. \tag{9.37}$$

4) *Выбор кода.* Ключевой момент доказательства состоит в возможности выбора кода с нужными условиями. Из соотношений (9.36) и (9.37) по принципу Дирихле обязательно найдется такой код U_0, для которого $\lambda(U_0) < 2\varepsilon$. Докажем, что существует такой код U_0, что все его кодовые слова $x^n(i)$ различны и являются ε-типичными.

Вначале докажем, что при достаточно больших n число ε-типичных векторов x^n больше, чем M_n. Поскольку кодовые слова порождаются источником сообщений без памяти, то выполняется свойство асимптотической равнораспределенности. Поэтому можно показать, что при достаточно больших n для мощности множества $\mathcal{X}_{n,\varepsilon}$ ε-типичных векторов x^n справедлива нижняя оценка

$$|\mathcal{X}_{n,\varepsilon}| > (1-\varepsilon)2^{n(\mathsf{H}\{X\}-\varepsilon)}.$$

Согласно выбору M_n

$$M = [2^{nR}] + 1 \leqslant 2^{nR} + 1.$$

Исходя из выбора ε величина $h = C^* - R - \varepsilon > 0$. Кроме того, по свойству энтропии $\mathsf{H}\{X\} \geqslant \mathsf{I}\{X,Y\} = C^*$. Тогда

$$|\mathcal{X}_{n,\varepsilon}| > (1-\varepsilon)2^{n(\mathsf{H}\{X\}-\varepsilon)} \geqslant (1-\varepsilon)2^{n(C^*-\varepsilon)} = (1-\varepsilon)2^{n(R+h)}.$$

За счет выбора достаточно большого n можно сделать так, чтобы выражение $(1-\varepsilon)2^{n(R+h)}$ стало больше $2^{nR}+1$, откуда следует неравенство $|\mathcal{X}_{n,\varepsilon}| > M_n$ для всех достаточно больших n.

Пусть в выбранном коде U_0 кодовое слово $x^n(i)$ не является ε-типичным вектором. Тогда по определению декодера $\mathfrak{D}_\varepsilon$ при посылке в канал кодового слова $x^n(i)$ декодер всегда будет возвращать сообщение об ошибке ε, следовательно, $\lambda_i(U_0) = 1$. Заменим кодовое слово $x^n(i)$ на произвольный ε-типичный вектор из множества $\mathcal{X}_{n,\varepsilon}\backslash U_0$ (из ранее показанного следует, что мощность данного множества положительна, и значит, оно не пустое). Получим новый код, в котором $\lambda_i(U_0)$ (а значит, и $\lambda(U_0)$) могут только уменьшиться.

Пусть в выбранном коде U_0 совпадают два кодовых слова: $x^n(i) = x^n(j)$, $i \neq j$. В этом случае согласно определению декодера $\mathfrak{D}_\varepsilon$ имеет место следующее равенство: $\lambda_i(U_0) = \lambda_j(U_0) = 1$. Заменим кодовое слово $x^n(i)$ на произвольный ε-типичный вектор из множества $\mathcal{X}_{n,\varepsilon}\backslash U_0$. Получим новый код, в котором $\lambda_i(U_0)$, $\lambda_j(U_0)$ (а значит, и $\lambda(U_0)$) могут только уменьшиться.

Используя описанные выше замены, можно получить код U_0, состоящий из различных ε-типичных кодовых слов $x^n(i)$, $i \in \{1, \ldots, M_n\}$. Не ограничивая общности, будем считать, что эти кодовые слова упорядочены по невозрастанию условной вероятности ошибки декодирования, т. е. так, что

$$\lambda_1(U_0) \leqslant \lambda_2(U_0) \leqslant \ldots \leqslant \lambda_{M_n}(U_0).$$

Рассмотрим новый код $\mathcal{C}^{(n)}$, состоящий из первых $[M_n/2]$ кодовых слов кода U_0. Покажем от противного, что выполнено неравенство

$$\lambda_{\max}(\mathcal{C}^{(n)}) = \lambda_{[M/2]}(U_0) < 4\varepsilon.$$

Пусть это не так, т. е. $\lambda_{[M/2]}(U_0) \geqslant 4\varepsilon$, тогда

$$\lambda(U_0) = \frac{1}{M_n}\sum_{i=1}^{M_n}\lambda_i(U_0) \geqslant \frac{1}{M_n}\sum_{i=[M_n/2]+1}^{M_n}\lambda_i(U_0) \geqslant 4\varepsilon\frac{M-[M/2]}{M} \geqslant 2\varepsilon,$$

что противоречит выбору кода U_0.

Таким образом, для кода $\mathcal{C}^{(n)}$ максимальная вероятность ошибочного декодирования, в силу (9.35), удовлетворяет условию

$$\lambda_{\max}(\mathcal{C}^{(n)}) < 4\varepsilon < \beta,$$

а скорость передачи

$$R_n = \frac{1}{n}\log_2\left[\frac{M_n}{2}\right] = \frac{1}{n}\log_2\left[\frac{[2^{nR}+1]}{2}\right]$$

лежит в границах

$$R - \frac{1}{n} + \frac{1}{n}\log_2(1-2^{1-nR}) < R_n \leqslant R - \frac{1}{n} + \frac{1}{n}\log_2(1+2^{-nR}),$$

так как $\forall a \in \mathbb{R}$ выполнено неравенство $a - 1 < [a] \leqslant a$. Поэтому при достаточно больших n, в силу (9.33), скорость передачи R_n удовлетворяет условию

$$C^* - \alpha < R_n < C^*. \qquad \square$$

Замечание 9.10. Теорема означает, что при достаточно большой длине кода можно добиться, чтобы скорость передачи информации была сколь угодно близка к пропускной способности канала связи, не превосходя ее, и при этом вероятность ошибочного декодирования была сколь угодно малой.

Замечание 9.11. Отметим, что в ключевом моменте (выборе кода с нужными условиями из множества всех возможных кодов), а потому и в целом доказательство теоремы неконструктивное: доказывается лишь существование нужного кода, но он не строится явно.

Замечание 9.12. Теорема остается справедливой и для случая $b \neq 2$.

Прямая теорема кодирования является одним из важнейших результатов теории информации, поскольку указывает предельные возможности надежной передачи информации по каналу связи с шумом, но, к сожалению, эта теорема не дает никакого указания на практически приемлемый способ построения кодов и декодеров с нужными свойствами. Эффективное построение кодов и декодеров, свойства которых были бы близки к предельным, описанным в прямой теореме кодирования, остается важной открытой проблемой.

9.10. ЗАДАНИЯ ДЛЯ ТЕСТОВ

9.1. Что необходимо определить для задания дискретного канала связи:

а) входной алфавит $\mathcal{X}$ и выходной алфавит $\mathcal{Y}$;

б) входной алфавит $\mathcal{X}$, $\forall x \in \mathcal{X}$ распределение $U(x) = \mathsf{P}\{X = x\}$ и выходной алфавит $\mathcal{Y}$;

в) входной алфавит $\mathcal{X}$, выходной алфавит $\mathcal{Y}$ и совместное распределение $\mathsf{P}\{X = x, Y = y\}$;

г) входной алфавит $\mathcal{X}$, выходной алфавит $\mathcal{Y}$ и стохастическую матрицу $\pi_{ij} = \{\mathsf{P}\{Y = y^{(j)} | X = x^{(i)}\}\}$;

д) входной алфавит $\mathcal{X}$, выходной алфавит $\mathcal{Y}$ и для любого $n \in \mathbb{N}$ и любых $x^n \in \mathcal{X}^n$, $y^n \in \mathcal{Y}^n$ переходные вероятности $\pi^{(n)}(y^n | x^n)$?

9.2. Какое из приведенных утверждений неверное для дискретного стационарного канала связи без памяти:

а) $C^* = \max\limits_u \mathsf{I}\{X, Y\}$;

б) $C^* = \sup\limits_{n \geqslant 1} \left(\dfrac{1}{n} \inf_{U_{1,\ldots,n}} (\mathsf{H}\{X^n | Y^n\} - \mathsf{H}\{X^n\}) \right)$;

в) $C^* = \max\limits_u (\mathsf{H}\{Y\} - \mathsf{H}\{Y | X\})$;

г) $C^* = \dfrac{1}{n}\mathbf{H}\{X^n\}$;

д) $C^* = \sup\limits_{n\geqslant 1}\left(\dfrac{1}{n}\sup\limits_{U_{1,\ldots,n}}\mathbf{I}\{X^n, Y^n\}\right)$?

9.3. Канал связи $(\mathcal{X}, \mathcal{Y}, \pi)$ называется симметричным, если:

а) равномерному распределению на входе соответствует равномерное распределение на выходе;

б) все строки матрицы $\pi \in \mathbb{R}^{p\times q}$ являются перестановками одного и того же набора чисел $\pi_1^1, \ldots, \pi_q^1$;

в) он симметричен по входу и по выходу;

г) все столбцы матрицы $\pi \in \mathbb{R}^{p\times q}$ являются перестановками одного и того же набора чисел $\pi_1^1, \ldots, \pi_p^1$;

д) матрица $\pi \in \mathbb{R}^{p\times p}$ симметричная, т. е. $\pi^T = \pi$.

9.4. Пусть заданы два канала связи без памяти с пропускными способностями $C_1^* > 2$ и $C_2^* > 2$ соответственно $(C_1^* > C_2^*)$. При какой операции пропускная способность станет максимальной:

а) последовательном соединении каналов;

б) использовании только первого канала;

в) параллельном соединении каналов;

г) сумме каналов;

д) использовании только второго канала?

9.5. Пусть задан код длины n и объемом M. Если решающая область декодера $A_{M+1} = \emptyset$, то:

а) вероятность ошибки декодера равна нулю;

б) он является декодером по методу максимального правдоподобия;

в) декодер никогда не возвращает сообщение об ошибке;

г) декодер имеет наименьшую среднюю вероятность ошибочного декодирования;

д) он является декодером по методу максимальной апостериорной вероятности.

9.6. Какое из перечисленных неравенств называется неравенством Фано:

а) $\mathbf{H}\{X|Y\} \leqslant \mathbf{H}\{X\}$;

б) $\mathbf{H}\{X^n|Y^n\} \leqslant h(\lambda) + \lambda\log(M-1)$;

в) $\dfrac{1}{n}\mathbf{H}\{X^n\} \leqslant C^* + \varepsilon$;

г) $1 + p\log p + (1-p)\log(1-p) \geqslant 0$;

д) $\lambda \leqslant 1 - 1/M$?

9.7. Какой декодер обеспечивает минимальную среднюю вероятность ошибочного декодирования:

а) декодер по методу максимальной апостериорной вероятности;

б) декодер общего вида;

в) декодер по методу максимального правдоподобия;

г) любой декодер эквивалентный декодеру по методу максимального правдоподобия;

д) декодер ε-типичных пар?

9.11. РЕШЕНИЕ ТИПОВЫХ ЗАДАЧ

Задача 9.1. Вычислить в битах пропускную способность и оптимальное входное распределение для стационарного дискретного канала связи без памяти, заданного матрицей π переходных вероятностей:

$$\pi = \begin{pmatrix} 0{,}1 & 0{,}9 \\ 0{,}3 & 0{,}7 \end{pmatrix}.$$

Решение. Известно, что для стационарного дискретного канала связи без памяти пропускная способность может быть найдена следующим образом:

$$C^* = \max_u \mathsf{I}\{X, Y\}.$$

Пусть на вход поступает один из сигналов x_1, x_2, а на выходе мы получим y_1, y_2. Положим, $\mathsf{P}\{X = x_1\} = \gamma$, тогда $\mathsf{P}\{X = x_2\} = 1-\gamma$ и задача сводится к нахождению такого значения γ, при котором $\mathsf{I}\{X, Y\}$ достигает своего максимума.

Для вычисления количества информации $\mathsf{I}\{X, Y\}$ необходимо найти распределение вероятностей случайного выхода Y и совместное распределение вероятностей (X, Y). Найдем их. По формуле полной вероятности

$$\mathsf{P}\{Y = y_1\} = \mathsf{P}\{Y = y_1 | X = x_1\}\,\mathsf{P}\{X = x_1\} +$$
$$+\,\mathsf{P}\{Y = y_1 | X = x_2\}\,\mathsf{P}\{X = x_2\} = 0{,}1\gamma + 0{,}3(1-\gamma) = 0{,}3 - 0{,}2\gamma.$$

Аналогично

$$\mathsf{P}\{Y = y_2\} = 0{,}7 + 0{,}2\gamma.$$

По формуле произведения вероятностей получим

$$\mathsf{P}\{X = x, Y = y\} = \mathsf{P}\{Y = y | X = x\}\,\mathsf{P}\{X = x\},$$
$$\forall x \in \{x_1, x_2\},\ \forall y \in \{y_1, y_2\}.$$

Вычислим $\mathsf{I}\{X, Y\}$ по определению:

$$\mathsf{I}\{X, Y\} = \mathsf{H}\{Y\} - \mathsf{H}\{Y|X\} =$$
$$= -(0{,}3 - 0{,}2\gamma)\log_2(0{,}3 - 0{,}2\gamma) - (0{,}7 + 0{,}2\gamma)\log_2(0{,}7 + 0{,}2\gamma) +$$
$$+\,0{,}1\gamma\log_2 0{,}1 + 0{,}9\gamma\log_2 0{,}9 +$$
$$+\,0{,}3(1-\gamma)\log_2 0{,}3 + 0{,}7(1-\gamma)\log_2 0{,}7. \qquad (9.38)$$

Для нахождения максимума количества информации $\mathsf{I}\{X, Y\}$ вычислим производную по γ и приравняем ее к нулю:

$$\mathsf{I}\{X, Y\}'_\gamma = 0{,}2\log_2(0{,}3 - 0{,}2\gamma) + 0{,}2 - 0{,}2\log_2(0{,}7 + 0{,}2\gamma) - 0{,}2 +$$

$$+0{,}1\log_2 0{,}1 + 0{,}9\log_2 0{,}9 - 0{,}3\log_2 0{,}3 - 0{,}7\log_2 0{,}7 = 0.$$

По свойству логарифма имеем

$$\log_2 \frac{0{,}3 - 0{,}2\gamma}{0{,}7 + 0{,}2\gamma} =$$
$$= \frac{0{,}3\log_2 0{,}3 + 0{,}7\log_2 0{,}7 - 0{,}1\log_2 0{,}1 - 0{,}9\log_2 0{,}9}{0{,}2}. \qquad (9.39)$$

Введем обозначение

$$C = 2^{\frac{0{,}3\log_2 0{,}3+0{,}7\log_2 0{,}7-0{,}1\log_2 0{,}1-0{,}9\log_2 0{,}9}{0{,}2}}.$$

Тогда уравнение (9.39) перепишется в следующем виде:

$$\frac{0{,}3 - 0{,}2\gamma}{0{,}7 + 0{,}2\gamma} = C,$$

откуда после элементарных преобразований

$$\gamma = \frac{0{,}3 - 0{,}7C}{0{,}2(1 + C)} \approx 0{,}533655.$$

Подставив полученное значение γ в (9.38), получим искомую пропускную способность, измеренную в битах $b = 2$:

$$C^* \approx 0{,}0469926.$$

Ответ: $C^* \approx 0{,}0469926$, $\gamma \approx 0{,}533655$.

Задача 9.2. По ДСК(p) передаются равновероятные кодовые слова $x(1) = (00)$ и $x(2) = (11)$. На выходе используется декодер с решающими областями $A_1 = \{(00), (01)\}$, $A_2 = \{(10),(11)\}$. Вычислить вероятности ошибок λ_1, λ_2, λ.

Решение. Напомним, что λ_i – условная вероятность ошибочного декодирования при условии, что было передано кодовое слово $X = x(i)$:

$$\lambda_i = \mathsf{P}\left\{\mathfrak{D}(Y^n) \neq x(i) | X^n = x(i)\right\} = \mathsf{P}\left\{Y \notin A_i | X^n = x(i)\right\};$$

λ — средняя вероятность ошибочного декодирования:

$$\lambda = \mathsf{P}\left\{\mathfrak{D}(Y^n) \neq X^n\right\}.$$

Известно, что эти вероятности могут быть вычислены по следующим формулам:

$$\lambda_i = 1 - \sum_{y^n \in A_i} \pi^{(n)}(y^n | x(i)),$$

$$\lambda = 1 - \sum_{i=1}^{M} \mathsf{P}\left\{X^n = x(i)\right\} \sum_{y^n \in A_i} \pi^{(n)}(y^n | x(i)).$$

Таким образом, для расчета соответствующих вероятностей необходимо найти условные вероятности $\pi^{(2)}(y^2|x(i)) = \mathsf{P}\left\{Y^2 = y^2 | X^2 = x(i)\right\}$. Для нахождения этих условных вероятностей вспомним, что ДСК — двоичный симметричный

канал, по которому с вероятностью $1-p$ один символ передается без искажений, а с вероятностью p при передаче меняется на противоположный. Имеем

$$\pi^{(2)}((00)|(00)) = (1-p)^2,$$

поскольку каждый из символов сообщения передается независимо и вероятность того, что символ не изменился, равна $1-p$. Аналогично

$$\pi^{(2)}((01)|(00)) = p(1-p),$$

так как вероятность, что первый символ не изменится, равна $1-p$, а вероятность, что второй символ изменится на противоположный, — p. Откуда, учитывая $A_1 = \{(00),(01)\}$, получим

$$\lambda_1 = 1 - \mathbf{P}\left\{Y^2 = (00)|X^2 = (00)\right\} - \mathbf{P}\left\{Y^2 = (01)|X^2 = (00)\right\} =$$
$$= 1 - \pi((00)|(00)) - \pi((01)|(00)) = 1 - (1-p)^2 - p(1-p) = p.$$

Аналогично вычислим оставшиеся вероятности, учитывая, что в силу равновероятности кодовых слов $\mathbf{P}\left\{X^2 = x(1)\right\} = \mathbf{P}\left\{X^2 = x(2)\right\} = 1/2$:

$$\lambda_2 = 1 - \mathbf{P}\left\{Y^2 = (10)|X^2 = (11)\right\} - \mathbf{P}\left\{Y^2 = (11)|X^2 = (11)\right\} = p;$$
$$\lambda = 1 - \frac{1}{2}\left(\mathbf{P}\left\{Y^2 = (00)|X^2 = (00)\right\} + \mathbf{P}\left\{Y^2 = (01)|X^2 = (00)\right\}\right) -$$
$$-\frac{1}{2}\left(\mathbf{P}\left\{Y^2 = (10)|X^2 = (11)\right\} + \mathbf{P}\left\{Y^2 = (11)|X^2 = (11)\right\}\right) = p.$$

Ответ: $\lambda_1 = p$, $\lambda_2 = p$, $\lambda = p$.

Задача 9.3. По ДСК(0,2) передаются кодовые слова $x(1) = (00)$ и $x(2) = (11)$ с вероятностями 0,25 и 0,75 соответственно. На выходе используется декодер $\mathfrak{D}_{AP}$ (по методу максимальной апостериорной вероятности). Задать этот декодер таблично и проверить, выполняется ли неравенство Фано для этого случая (не используя известную лемму).

Решение. Напомним, что алгоритм декодера $\mathfrak{D}_{AP}$ состоит в том, что если принят вектор y^n, то вычисляются апостериорные вероятности $\mathbf{P}\{X^n = x^n|Y^n = y^n\}$ для всех кодовых слов $x^n \in \mathcal{C}$ и в качестве результата декодирования выбирается то кодовое слово, для которого эта вероятность максимальна. Апостериорная вероятность находится следующим образом:

$$\mathbf{P}\{X^n = x^n|Y^n = y^n\} = \frac{\mathbf{P}\{X^n = x^n\}\,\pi^{(n)}(y^n|x^n)}{\sum\limits_{z^n\in\mathcal{C}} \mathbf{P}\{X^n = z^n\}\,\pi^{(n)}(y^n|z^n)}.$$

Условные вероятности $\pi(y^n|x(i))$ рассчитываются, как и при решении предыдущей задачи.

Вычислим апостериорные вероятности:

$$\mathbf{P}\left\{X^2 = (00)|Y^2 = (00)\right\} =$$
$$= \frac{\mathbf{P}\left\{X^2 = (00)\right\}\pi^{(2)}((00)|(00))}{\mathbf{P}\left\{X^2 = (00)\right\}\pi^{(2)}((00)|(00)) + \mathbf{P}\left\{X^2 = (11)\right\}\pi^{(2)}((00)|(11))} =$$

$$= \frac{0{,}25 \cdot 0{,}8^2}{0{,}25 \cdot 0{,}8^2 + 0{,}75 \cdot 0{,}2^2} \approx 0{,}842105,$$

$$\mathsf{P}\left\{X^2 = (11) | Y^2 = (00)\right\} = \frac{0{,}75 \cdot 0{,}2^2}{0{,}25 \cdot 0{,}8^2 + 0{,}75 \cdot 0{,}2^2} \approx 0{,}157895.$$

Отсюда получим, что поскольку

$$\mathsf{P}\left\{X^2 = (00) | Y^2 = (00)\right\} > \mathsf{P}\left\{X^2 = (11) | Y^2 = (00)\right\},$$

то $\mathfrak{D}_{AP}((00)) = (00)$.

Вычислив остальные апостериорные вероятности, получим, что решающими областями декодера $\mathfrak{D}_{AP}$ являются

$$A_1 = \{(00)\},\ A_2 = \{(01), (10), (11)\}.$$

Перейдем ко второй части задачи. Неравенство Фано имеет вид

$$\mathsf{H}\{X^n | Y^n\} \leqslant h(\lambda) + \lambda \log_2(M-1),$$

где $h(\lambda) = -\lambda \log_2 \lambda - (1-\lambda)\log_2(1-\lambda)$ — функция двоичной энтропии. В данной задаче $M = 2$, поэтому второе слагаемое в правой части обратится в нуль. Для того чтобы проверить, выполнено ли это неравенство, необходимо найти вероятность ошибочного декодирования λ и условную энтропию $\mathsf{H}\{X^2|Y^2\}$.

Найдем среднюю вероятность ошибочного декодирования (см. решение предыдущей задачи): $\lambda = 0{,}12$.

Вычислим условную энтропию по определению:

$$\mathsf{H}\{X^2|Y^2\} = -\sum_{x^2 \in \mathcal{C}, y^2 \in \mathcal{Y}^2} \mathsf{P}\left\{X^2 = x^2, Y^2 = y^2\right\} \log_2 \mathsf{P}\left\{X^2 = x^2 | Y^2 = y^2\right\}.$$

Подставив в эту формулу найденные апостериорные и совместные вероятности, получим $\mathsf{H}\{X^2|Y^2\} \approx 0{,}449529$. Отсюда

$$0{,}449529 \approx \mathsf{H}\{X^2|Y^2\} < h(\lambda) \approx 0{,}529361.$$

Ответ: решающие области декодера $A_1 = \{(00)\}$, $A_2 = \{(01), (10), (11)\}$.

Задача 9.4. Задать таблично декодер $\mathfrak{D}_L$ (по методу максимального правдоподобия), если кодовые слова кода $x(1) = (00)$ и $x(1) = (11)$ передаются:

а) по ДСК с параметром $p = 0{,}2$;

б) двоичному каналу со стиранием с параметром $p = 0{,}2$.

Решение. Напомним, что алгоритм декодера $\mathfrak{D}_L$ состоит в том, что если принят вектор y^n, то вычисляются переходные вероятности $\pi^{(n)}(y^n|x^n)$ для всех кодовых слов x^n и в качестве результата декодирования выбирается то кодовое слово, для которого эта вероятность максимальна. Если максимальное значение переходной вероятности достигается для нескольких кодовых слов, то выбирается одно из них по некоторому заранее оговоренному правилу, например наименьшее в смысле лексикографического порядка.

а) Как и в предыдущих рассматриваемых задачах, найдем условные вероятности:

$$\pi^{(2)}((00)|(00)) = \mathsf{P}\left\{Y^2 = (00) | X^2 = (00)\right\} = (1-p)^2 = 0{,}64;$$

$$\pi^{(2)}((00)|(11)) = \mathsf{P}\left\{Y^2 = (00)|X^2 = (11)\right\} = p^2 = 0{,}04.$$

Поскольку $\pi^{(2)}((00)|(00)) > \pi^{(2)}((00)|(11))$, то $\mathfrak{D}_L((00)) = (00)$. Вычисляя аналогично остальные условные вероятности, получим, что решающими областями декодера $\mathfrak{D}_L$ являются

$$A_1 = \{(00), (01), (10)\},\ A_2 = \{(11)\}.$$

Отметим, что при расчете условных вероятностей получается $\pi^{(2)}((01)|(00)) = \pi^{(2)}((01)|(11)) = 0{,}16$, поэтому в этом случае декодер выбирает результирующее слово наименьшее в смысле лексикографического порядка.

б) В этом случае выходной алфавит имеет вид $\mathcal{Y} = \{0,\ 1,\ E\}$, где символ E означает, что произошло стирание входного символа. Таким образом, отличие от предыдущего пункта будет заключаться в том, что теперь вместо 4 возможных исходов мы имеем 9. Для удобства запишем получаемые условные вероятности в виде таблицы.

$X\backslash Y$	(00)	(01)	(0E)	(10)	(11)	(1E)	(E0)	(E1)	(EE)
(00)	0,64	0	0,16	0	0	0	0,16	0	0,04
(11)	0	0	0	0	0,64	0,16	0	0,16	0,04

Из приведенной таблицы сделаем вывод:

$$A_1 = \{(00), (0E), (E0), (EE)\},\ A_2 = \{(11), (1E), (E1)\}.$$

Заметим, что для данной модели вероятность из сигнала (00) или сигнала (11) получить сигнал (10) равна нулю, поэтому этот выходной сигнал не включен ни в A_1, ни в A_2. В данном случае можно доопределить множество $A_3 = \{(01), (10)\}$. При этом если $Y^2 \in A_3$, то декодер объявляет об ошибке, т. е. декодер не может принять решение в пользу какого-либо кодового слова из кода $\mathcal{C}$.

Ответ: а) $A_1 = \{(00), (01), (10)\}$, $A_2 = \{(11)\}$; б) $A_1 = \{(00), (0E), (E0), (EE)\}$, $A_2 = \{(11), (1E), (E1)\}$, $A_3 = \{(01), (10)\}$.

9.12. ЗАДАЧИ И УПРАЖНЕНИЯ

9.1. По ДСК с параметром p передается кодовое слово $x^n = (111\ldots1)$ (состоящее из n единиц). Известно, что на выходе было получено слово y^n, в котором на k позициях оказались нули ($0 \leqslant k \leqslant n$). Найти условную вероятность $\mathsf{P}\left\{Y^n = y^n|X^n = x^n\right\}$.

9.2. Найти вероятность того, что при передаче кодового слова $x^n = (111\ldots1)$ (состоящего из n единиц) по ДСК(p) на выходе будет получено сообщение, содержащее k нулей ($0 \leqslant k \leqslant n$).

9.3. Вычислить пропускную способность и оптимальное входное распределение для канала связи, заданного матрицей π переходных вероятностей. Если

матрица содержит параметр p, построить график зависимости пропускной способности от p:

а) $\begin{pmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{pmatrix}$; б) $\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$; в) $\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$; г) $\begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$;

д) $\begin{pmatrix} \frac{3}{4} & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}$; е) $\begin{pmatrix} \frac{3}{4} & \frac{1}{4} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}$; ж) $\begin{pmatrix} \frac{3}{4} & 0 & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{4} & 0 & \frac{3}{4} \end{pmatrix}$; з) $\begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}$;

и) $\begin{pmatrix} \frac{5}{8} & \frac{1}{8} & \frac{1}{4} \\ \frac{5}{16} & \frac{3}{8} & \frac{5}{16} \\ \frac{3}{8} & \frac{1}{16} & \frac{9}{16} \end{pmatrix}$; к) $\begin{pmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \end{pmatrix}$; л) $\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$;

м) $\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$; н) $\begin{pmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \ddots & \vdots \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}$;

о) $\begin{pmatrix} 1-p & p & 0 \\ p & 1-p & 0 \\ p & 0 & 1-p \end{pmatrix}$; п) $\begin{pmatrix} 1-p & p & 0 \\ p & 1-p & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

9.4. Случайный символ X, порождаемый дискретным источником с алфавитом $\mathcal{X} = \{0, 1, 2\}$ и распределением $p = (1/4, 1/4, 1/2)$, передается одновременно по двум каналам связи с матрицами переходных вероятностей

$$\pi_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad \pi_2 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Пусть Y и Z — случайные символы на выходе первого и второго канала соответственно. Вычислить $\mathsf{H}\{X\}$, $\mathsf{H}\{Y\}$, $\mathsf{H}\{Z\}$, $\mathsf{H}\{Y,Z\}$, $\mathsf{I}\{X,Y\}$, $\mathsf{I}\{X,Z\}$, $\mathsf{I}\{X,(Y,Z)\}$.

9.5. Два дискретных симметричных канала (ДСК) с параметрами $p_1 = 1/4$, $p_2 = 1/3$ соединены последовательно. Пусть X и Y — входной и выходной символы первого ДСК, Y и Z —входной и выходной символы второго ДСК, X принимает значения 0 и 1 с вероятностями 0,5. Вычислить величины $\mathsf{I}\{X,Y\}$, $\mathsf{I}\{Y,Z\}$, $\mathsf{I}\{X,Z\}$.

9.6. Вычислить пропускную способность $C^*(p)$ для последовательного соединения из n двоичных симметричных каналов с одинаковым параметром p. Найти предел $\lim_{n\to\infty} C^*(p)$ как функцию от p.

9.7. Для канала связи с алфавитами $\mathcal{X} = \mathcal{Y} = \{0, 1, 2, 3, 4\}$ переходная вероятность $\pi(y|x)$ равна 0,5, если $y = x \pm 1 \pmod 5$, и равна 0 в противном случае. Вычислить пропускную способность.

9.8. К случайному входному символу X, принимающему числовые значения $0, 1, \ldots, 10$, в канале связи прибавляется по модулю 11 независимая случайная помеха Z, принимающая значения 1, 2 и 3 с равными вероятностями, так что на выходе получается $Y = (X + Z) \bmod 11$. Вычислить пропускную способность и оптимальное входное распределение.

9.9. К случайному входному символу X, принимающему числовые значения 0 и 1, в канале связи прибавляется независимая случайная помеха Z, принимающая значения 0 и a ($a \in \mathbb{R}$, $a > 0$) с равными вероятностями, так что на выходе получается $Y = X + Z$. Вычислить пропускную способность как функцию от параметра a.

9.10. Случайный входной символ X, принимающий числовые значения 0 и 1, в канале связи умножается на независимую случайную помеху Z, принимающую значения 0 и 1 с вероятностями p и $1 - p$ соответственно, так что на выходе получается $Y = X \cdot Z$. Вычислить пропускную способность как функцию от параметра p.

9.11. По ДСК(p) передаются равновероятные кодовые слова $x(1) = (000)$ и $x(2) = (111)$. На выходе используется декодер с решающими областями $A_1 = \{(000), (001), (010), (100)\}$, $A_2 = \{(110), (111), (101), (011)\}$. Вычислить вероятности ошибок λ_1, λ_2, λ.

9.12. По ДСК с параметром $p = 0{,}1$ передаются равновероятные кодовые слова $x(1) = (010)$, $x(2) = (101)$. Вычислить среднюю вероятность ошибочного декодирования λ, если на выходе используется декодер с решающими областями: а) $A_1 = \{(010), (000), (110), (011)\}$, $A_2 = \{(101), (001), (111), (100)\}$; б) $A_1 = \{(000), (001), (010), (100)\}$, $A_2 = \{(110), (101), (011), (111)\}$.

Какой из двух описанных декодеров лучше? Зависит ли выбор лучшего декодера от значения параметра p?

9.13. По ДСК(p) передаются кодовые слова $x(1) = (000)$, $x(2) = (111)$ с вероятностями α и $1 - \alpha$ соответственно. На выходе используется декодер $\mathfrak{D}_{AP}$ (по

методу максимальной апостериорной вероятности). Найти решающие области декодера $\mathfrak{D}_{AP}$ и записать неравенство Фано для случая:

а) $p = 0{,}15$, $\alpha = 0{,}4$; б) $p = 0{,}25$, $\alpha = 0{,}5$;

в) $p = 0{,}35$, $\alpha = 0{,}35$; г) $p = 0{,}40$, $\alpha = 0{,}2$.

9.14. Задать таблично декодер $\mathfrak{D}_L$ (по методу максимального правдоподобия), если кодовые слова кода $\{(0000), (0011), (1100), (1111)\}$ передаются:

а) по ДСК с параметром $p < 0{,}5$;

б) по двоичному каналу со стиранием с параметром $p < 0{,}5$.

9.15. Задать таблично декодер $\mathfrak{D}_{AP}$, если равновероятные кодовые слова кода $\{(00000), (01101), (10111), (11010)\}$ передаются по ДСК с параметром $p < 0{,}5$.

9.16. По ДСК с параметром $p = 0{,}1$ передаются четыре кодовых слова $\{(00000), (01111), (10101), (11010)\}$ с вероятностями 0,8; 0,1; 0,05 и 0,05 соответственно. Указать решающие области для декодеров $\mathfrak{D}_L$ и $\mathfrak{D}_{AP}$.

9.17. Дискретный стационарный канал связи без памяти задан входным алфавитом $\mathcal{X} = \{x_1, x_2, x_3\}$, выходным алфавитом $\mathcal{Y} = \{y_1, y_2, y_3\}$ и матрицей вероятностей переходов

$$\pi = \begin{pmatrix} 0{,}5 & 0{,}3 & 0{,}2 \\ 0{,}4 & 0{,}3 & 0{,}3 \\ 0{,}1 & 0{,}9 & 0 \end{pmatrix}.$$

Входные символы канала равновероятны. Декодер определяется условием $\mathfrak{D}(y_i) = x_i$, $i = 1, 2, 3$. Вычислить вероятности ошибок $\lambda_1, \lambda_2, \lambda_3, \lambda$. Записать неравенства Фано.

9.18. Дискретный стационарный канал связи без памяти задан входным алфавитом $\mathcal{X} = \{x_1, x_2, x_3\}$, выходным алфавитом $\mathcal{Y} = \{y_1, y_2, y_3\}$ и матрицей вероятностей переходов

$$\pi = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 1/6 & 1/2 & 1/3 \\ 1/3 & 1/6 & 1/2 \end{pmatrix}.$$

Входные символы канала имеют вероятности 0,5; 0,25 и 0,25 соответственно. На выходе используется декодер $\mathfrak{D}_{AP}$. Вычислить вероятности ошибок $\lambda_1, \lambda_2, \lambda_3, \lambda$. Проверить выполнение неравенства Фано.

ПРИЛОЖЕНИЕ 1

ОТВЕТЫ НА ЗАДАНИЯ ДЛЯ ТЕСТОВ

Глава 1
1.1. – в); 1.2. – г); 1.3. – б); 1.4. – б); 1.5. – д).

Глава 2
2.1. – а); 2.2. – в); 2.3. – б); 2.4. – б); 2.5. – г).

Глава 4
4.1. – г); 4.2. – д); 4.3. – б); 4.4. – а); 4.5. – г).

Глава 6
6.1. – в); 6.2. – б); 6.3. – в); 6.4. – д); 6.5. – а).

Глава 7
7.1. – б); 7.2. – в); 7.3. – а); 7.4. – д); 7.5. – г); 7.6. – б); 7.7. – г).

Глава 9
9.1. – д); 9.2. – г); 9.3. – в); 9.4. – г); 9.5. – в); 9.6. – б); 9.7. – а).

ПРИЛОЖЕНИЕ 2

УКАЗАНИЯ И РЕШЕНИЯ К ЗАДАЧАМ И УПРАЖНЕНИЯМ

Глава 2

2.8. Совместное распределение вычисляется непосредственно с использованием свойства независимости в совокупности.

2.9. Проверяется непосредственным вычислением вероятностей нового временного ряда.

2.10. Следует из свойства количества информации и того, что x_{n+1} и X_n независимы.

Глава 3

3.2. Доказать, что случайная последовательность x_t представляет собой последовательность независимых одинаково распределенных случайных величин.

Глава 4

4.1. По свойству энтропии $\mathsf{H}\{x_1, \ldots, x_n\} \leqslant \mathsf{H}\{x_1\} + \ldots + \mathsf{H}\{x_n\}$. Доказать по индукции, что $-\left(1 - \dfrac{1}{2^n}\right) \log_2 \left(1 - \dfrac{1}{2^n}\right) \leqslant \dfrac{1}{2^n}$, тогда $\mathsf{H}\{x_i\} \leqslant \dfrac{i+1}{2^i}$, откуда $h = \lim\limits_{n\to\infty} \dfrac{1}{n} \mathsf{H}\{x_1, \ldots, x_n\} \leqslant \lim\limits_{n\to\infty} \dfrac{1}{n} \sum\limits_{i=1}^{\infty} \dfrac{i+1}{2^n} = \lim\limits_{n\to\infty} \dfrac{3}{n} = 0.$

4.3. Доказать, что случайные величины x_t независимы в совокупности и, начиная со второго элемента, имеют распределение вероятностей $p_1, \ldots, p_n$.

4.5. Воспользоваться равенством $\mathsf{I}\{X, Y\} = \mathsf{H}\{X\} + \mathsf{H}\{Y\} - \mathsf{H}\{X, Y\}$.

Глава 6

6.6. Для расшифрования сообщения необходимо обратить матрицу. Это можно сделать по формуле

$$\theta^{-1} = (\det \theta)^{-1} \begin{pmatrix} \theta_{22} & -\theta_{12} \\ -\theta_{21} & \theta_{11} \end{pmatrix} \bmod \nu, \ \det \theta = \theta_{11}\theta_{22} - \theta_{12}\theta_{21} \bmod \nu.$$

6.7. Ключ-матрица шифра Хилла должна быть обратима, а для этого согласно формуле, приведенной в решении задачи 6.6, определитель этой матрицы должен быть взаимно прост с количеством ν символов алфавита.

6.8. Доказательство основано на том, что функция зашифрования при фиксированном θ должна быть обратимой.

6.12 – 6.13. Для вычисления энтропии ключа θ можно использовать следующие соотношения:

$$\mathbf{H}\{\theta|Y\} = \sum_y \mathbf{P}\{Y = y\}\, \mathbf{H}\{\theta|y\},$$

$$\mathbf{H}\{\theta|y\} = -\sum_{z\in\Theta} \mathbf{P}\{\theta = z|Y = y\} \log_2 \left(\mathbf{P}\{\theta = z|Y = y\}\right).$$

Глава 7

7.3. Пусть такой код однозначно декодируем. Тогда из условия следует, что $m \geqslant 2^n + 1$ и для любого кодового слова $l^{(i)} \leqslant n$. Проверим, выполнено ли неравенство Мак-Миллана.

$$\sum_{i=1}^{m} 2^{-l^{(i)}} \geqslant 2^{-l^{(2^n+1)}} + \sum_{i=1}^{2^n} 2^{-l^{(i)}} \geqslant 2^{-n} + 2^n 2^{-n} = 2^{-n} + 1 > 1.$$

Получили противоречие с неравенством Мак-Миллана, и, следовательно, такой код не может быть однозначно декодируемым. Аналогично, проверяя неравенство Крафта, можно показать, что данный код не является префиксным.

7.9. Докажем этот факт по индукции. Очевидно, что $D \geqslant 2$. При $m = 2$ оптимальный код будет состоять из односимвольных кодовых слов при любом распределении вероятностей. Тогда максимальная длина кодового слов равна 1 и не превосходит $2 - 1 = 1$, что верно.

Пусть утверждение задачи верно для алфавитов с длинами $2, 3, \ldots, m$. Рассмотрим $(m + 1)$-буквенный алфавит. Тогда для заданного распределения $\{p_i\}_{i=1,\ldots,m}$ существует приведенное оптимальное D-ичное префиксное алфавитное кодирование, для которого можно произвести операцию редукции, т. е. заменить S наименее вероятных символов одним с вероятностью, равной сумме вероятностей заменяемых символов. Не нарушая общности, можно считать, что это символы с номерами $m+1, \ldots, m+2-S$. Редуцированное кодирование для $m+2-S$-буквенного алфавита будет оптимальным, следовательно, так как $m+2-S \leqslant m$, имеем $\max\{l^{(1)}, \ldots, l^{(m+1-S)}, l(\sigma)\} \leqslant m-1$, где символ σ является объединением S наименее вероятных символов. Согласно определению операции редукции имеют место равенства $l^{(m+1)} = l(\sigma)+1, \ldots, l^{(m+2-S)} = l(\sigma)+1$, откуда $\max\{l^{(1)}, \ldots, l^{(m+1)}\} \leqslant m$.

Верхняя граница достижима при $D = 2$ и $p_i = 2^{-i}, i \in \{1, \ldots, m-1\}$, $p_m = 2^{-(m-1)}$. (Проверить самостоятельно.)

7.10. Докажем этот факт по индукции. При $m = 2$ оптимальный код будет иметь вид $\{0, 1\}$ при любом распределении вероятностей. Тогда сумма длин кодовых слов равна 2 и не превосходит $4 \cdot 1/2 = 2$, что верно.

Пусть утверждение задачи верно для m-буквенного алфавита. Рассмотрим $(m + 1)$-буквенный алфавит. Тогда для заданного распределения $\{p_i\}_{i=1,\ldots,m}$ существует приведенное оптимальное 2-ичное префиксное алфавитное кодирование, для которого можно произвести операцию редукции, т. е. заменить два наименее вероятных символа одним с вероятностью, равной сумме вероятно-

стей заменяемых символов. Не нарушая общности, можно считать, что это символы с номерами m и $m+1$. Редуцированное кодирование для m-буквенного алфавита будет оптимальным, следовательно, $l^{(1)} + \ldots + l^{(m-1)} + l(\sigma) \leqslant (m+2)(m-1)/2$, где символ σ — объединение двух наименее вероятных символов, тогда согласно определению операции редукции имеют место равенства $l^{(m)} = l(\sigma) + 1$, $l^{(m+1)} = l(\sigma) + 1$. Согласно задаче 7.9 $\tilde{l}^{(m+1)} \leqslant m-1$, откуда

$$l^{(1)} + \ldots + l^{(m+1)} \leqslant l^{(1)} + \ldots + \tilde{l}^{(m)} + 1 + (m-1) + 1 \leqslant \frac{(m+2)(m-1)}{2} +$$

$$+ m + 1 = \frac{m^2 + m - 2 + 2m + 2}{2} = \frac{m^2 + 3m}{2} = \frac{(m+3)m}{2}.$$

7.13. Указание. Для заданного распределения вероятностей можно применить алгоритм Хаффмана, построить оптимальный двоичный код и найти среднюю длину кодового слова.

7.14. См. предыдущую задачу.

7.15. См. предыдущую задачу.

7.16. Индукцией по m доказать, что суммарная длина всех кодовых слов будет $l_s = m\lfloor \log_2 m \rfloor + 2(m - 2^{\lfloor \log_2 m \rfloor})$, тогда средняя длина кодового слова вычисляется как $l^{\varphi} = \frac{l_s}{m} = \lfloor \log_2 m \rfloor + 2 - \frac{2^{\lfloor \log_2 m \rfloor + 1}}{m}$. Энтропия для равномерного распределения на алфавите из m элементов — $\mathrm{H}\{\xi\} = \log_2 m$. Можно доказать, что $l^{\varphi} - \mathrm{H}\{\xi\}$ всегда положительна и незначительно отличается от нуля.

7.17. Для равномерного распределения применить алгоритм Хаффмана, вычислить его среднюю длину и показать, что эта длина совпадает со средней длиной двоичного кода, построенного для заданного распределения вероятностей и вычисленной для равномерного распределения.

7.18. По определению энтропии дискретной случайной величины имеем

$$\sum_{i=1}^{5} q_i l_i = \mathsf{H}\{\xi\} = -\sum_{i=1}^{5} q_i \log_2 q_i = \sum_{i=1}^{5} q_i \log_2 \frac{1}{q_i}.$$

Выберем q_i так, чтобы для любого i имело место равенство $l_i = \log_2 1/q_i$, что равносильно $2^{l_i} = 1/q_i$, откуда $q_i = 2^{-l_i}$. Остается построить код Хаффмана для заданного в условии задачи распределения и проверить, что величины q_i удовлетворяют условию нормировки.

7.19. Показать, что для заданного распределения оптимальным является код $\varphi_k(a^{(j)}) = \{11\ldots10\ (j-1 \text{ единица}), j < k; 11\ldots11\ (k-1 \text{ единица}),\ j = k\}$. Тогда

$$l^{\varphi_k} = \sum_{j=1}^{k-1} \frac{j}{2^j} + \frac{k-1}{2^{k-1}}.$$

7.20. а) Чтобы доказать, что $l_1 = 1$, надо показать, что в алгоритме Хаффмана до предпоследнего шага в бинарном дереве не появятся вершины, имеющие вероятность больше, чем p_1. Для $m = 2$ и $m = 3$ это утверждение очевидно. Рассмотрим случай, когда $m = 4$. Имеем $p_1 \geqslant p_2 \geqslant p_3 \geqslant p_4$ и $p_1 > 2/5$. На текущем

шаге объединим вершины с вероятностями p_3 и p_4. Пусть у вновь образованной вершины вероятность $p_3 + p_4 > p_1 > 2/5$. Тогда $p_1 + p_3 + p_4 > 4/5$, поэтому $p_2 < 1/5$, но $p_2 \geqslant p_3 \geqslant p_4$, откуда $p_3 + p_4 < 2/5$. Получили противоречие и, значит, $p_3 + p_4 < p_1$. Следовательно, не появятся вершины, имеющие вероятность большую, чем p_1. В случае $m > 4$ доказательство проводится аналогично.

б) Предположим, $l_1 = 3$, тогда при построении бинарного дерева и сортировке вершин по крайней мере два раза должны были появляться вершины, лежащие левее вершины, имеющей вероятность $p_1 > 1/3$. Отметим, что вершины, лежащие левее, имеют бóльшую вероятность. Тогда сумма вероятностей этих двух вершин и вершины с номером один будет больше 1, чего быть не может.

7.21. Пусть длина всех кодовых слов равна l, тогда $l \geqslant \lceil \log_2 m \rceil$. Рассмотрим кодовое дерево. Поскольку m не степень двойки, то l-м уровне находится меньше чем 2^l вершин, а это значит, что обязательно найдется вершина на уровне меньшем, чем l, степень которой равна 1.

Пусть такая вершина v находится на уровне $l-1$, а ее единственным потомком является вершина u. Поменяем вершину u местами с наиболее вероятной вершиной w, при этом средняя длина не изменится, но вершина w станет потомком вершины v. После чего произведем удаление вершины w и ребра (w, u), а вершину v сделаем новым листом. В этом случае средняя длина гарантированно уменьшится, поскольку $p(w) > 0$.

Пусть такая вершина v находится на уровне $s < l-1$. Тогда перенесем самый высоковероятный лист w к ненасыщенной вершине v, т. е. сделаем вершину w потомком v. При этом средняя длина гарантированно уменьшится.

Таким образом, предположив, что длина всех кодовых слов равна l, мы получили противоречие с тем, что код является оптимальным.

7.22. Указание. При построении кодов Хаффмана символы с нулевой вероятностью могут повлиять на коды символов с ненулевой вероятностью.

Глава 9

9.3. В случае если локальный максимум не находится внутри области, в которой мы ищем максимум, то максимум достигается на границе области.

9.6. Требуемый результат можно доказать по методу математической индукции; для вычисления предела при $0 < p < 1$ можно рассмотреть поведение элементов матриц в подпоследовательности $n = 2^k$.

9.7, 9.8. Воспользоваться свойством симметричности по входу и по выходу.

9.15. В случае равновероятности входных символов декодеры $\mathcal{D}_L$ и $\mathcal{D}_{AP}$ эквивалентны. Поэтому в данном случае проще построить декодер $\mathcal{D}_L$.

ПРИЛОЖЕНИЕ 3

ОТВЕТЫ К ЗАДАЧАМ И УПРАЖНЕНИЯМ

Глава 1

1.1. $h(p) = -p\log p - (1-p)\log(1-p)$.

1.2. $\mathsf{H}\{\xi_1\} = \log 2, \mathsf{H}\{\eta\} = \dfrac{3}{2}\log 2, \mathsf{H}\{\xi_1|\eta\} = \dfrac{1}{2}\log 2, \mathsf{H}\{\eta|\xi_1\} = \log 2$.

1.3. $\mathsf{H}\{\xi_2\} = \log 2, \mathsf{H}\{\xi_2|\xi_1\} = 0, \mathsf{H}\{\xi_1|\xi_2\} = \log 3$.

1.4. $\mathsf{H}\{\xi\} \leqslant -\varepsilon\log\varepsilon - (1-\varepsilon)\log\dfrac{1-\varepsilon}{N-1}$, причем равенство достигается тогда и только тогда, когда $p_i = \dfrac{1-\varepsilon}{N-1}$, $i \in \{1,\ldots,N-1\}$.

1.5. $2\log 2$.

1.6. $4 \leqslant \mathsf{H}\{\xi_2|\xi_1\} \leqslant 12$.

1.7. $\mathsf{H}\{\xi_1\} = \mathsf{H}\{\xi_2\} = \log 2$, $\mathsf{H}\{\xi_1|\xi_2\} = \mathsf{H}\{\xi_2|\xi_1\} = \log 3 - (2/3)\log 2$.

1.8. $\mathsf{I}\{x,x\} = \mathsf{H}\{x\}$.

1.9. $\mathsf{H}\{y\} = \dfrac{12\log 3 - 2\log 2}{9}$, $\mathsf{I}\{x_1,y\} = \dfrac{3\log 3 - 2\log 2}{9}$, $\mathsf{I}\{x_2,y\} = \dfrac{3\log 3 + 4\log 2}{9}$.

1.10. $\alpha = \beta$.

1.11. 1) 0; 2) h.

1.12. $\mathsf{H}\{\xi_1\} = \log 2$, $\mathsf{I}\{\xi_1,\xi_2\} = \dfrac{n}{2n-1}\log n + \dfrac{n-1}{2n-1}\log(n-1) - \log(2n-1) + \log 2$, $\mathsf{I}\{(\xi_1,\ldots,\xi_{n-1}),\xi_n\} = \log 2$.

1.13. $(1-\pi_1)h(\alpha) + \pi_1 h(\beta)$.

1.14. $\mathsf{H}_d\{\eta\} = \mathsf{H}_d\{\xi\} + \log|a|$.

1.15. $\log(b-a)$.

1.16. $\log\sqrt{2\pi e\sigma^2}$.

1.17. $\log e - \log\lambda$.

1.18. $-\log p - \dfrac{1-p}{p}\log(1-p)$.

1.19. $\sum\limits_{i=1}^{N-1} C_{N-1}^{i} p^{i}(1-p)^{N-1-i}\log C_{N-1}^{i} - n(p\log p + (1-p)\log(1-p))$.

1.20. $\mathsf{I}\{x,y\} = 0$ при $\mathsf{H}\{x,y\} = \mathsf{H}\{x\} + \mathsf{H}\{y\}$, $\mathsf{I}\{x,y\} = \min\{\mathsf{H}\{x\}, \mathsf{H}\{y\}\}$ при $\mathsf{H}\{x,y\} = \max\{\mathsf{H}\{x\},\mathsf{H}\{y\}\}$.

1.21. $0{,}5\log e - \log 2$.

1.22. $\mathsf{H}\{x\} = \mathsf{H}\{y\} = \dfrac{11}{4}\log 2 - \dfrac{3}{4}\log 3$, $\mathsf{H}\{x|y\} = \mathsf{H}\{y|x\} = \dfrac{1}{4}\log 2 + \dfrac{3}{4}\log 3$, $\mathsf{H}\{x,y\} = 3\log 2$.

1.23. Элемент первого источника несет большее количество информации.

1.24. $\mathsf{H}\{x\} = \log 1050$, $\mathsf{H}\{x\} = \log 35$, $\mathsf{H}\{x|y\} = \log 30$.

1.25. $\mathsf{I}\{x,y\} = 0$.

Глава 2

2.1. np, $np(1-p)$.

2.2. $n(2-q-2p)$, $n(4p+q-(2q+p)^2)$.

2.3. а) $m(1) = 3/4$, $D(1) = 3/16$, $m(t) = 1/2$, $D(t) = 1/4, t > 1$, $\sigma(t_1,t_2) = 0$, стационарен при $\mathsf{P}\{x_1 = 0\} = \mathsf{P}\{x_1 = 1\} = 1/2$; б) $m(t) = 3/2^{t+1}$, $D(t) = \left(1 - \dfrac{3}{2^{t+1}}\right)\dfrac{3}{2^{t+1}}$, $\sigma(t_1,t_2) = \left(1 - \dfrac{3}{2^{t_1+1}}\right)\dfrac{3}{2^{t_2+1}}, t_2 \geqslant t_1$, стационарен при $\mathsf{P}\{x_1 = 0\} = 1$, $\mathsf{P}\{x_1 = 1\} = 0$.

2.4. а) При $t = 3k+1$: $m(t) = 1/3$, $D(t) = 2/3$; при $t = 3k+2$: $m(t) = 7/6$, $D(t) = 17/36$; при $t = 3k$: $m(t) = 7/6$, $D(t) = 29/36$; стационарен при $\mathsf{P}\{x_1 = 0\} = \mathsf{P}\{x_1 = 1\} = \mathsf{P}\{x_1 = 2\} = 1/3$; б) при $t = 1$: $m(t) = 1/3$, $D(t) = 2/3$; при $t > 1$: $m(t) = 1$, $D(t) = 2/3$; стационарен при $\mathsf{P}\{x_1 = 0\} = \mathsf{P}\{x_1 = 1\} = \mathsf{P}\{x_1 = 2\} = 1/3$; в) при $t = 1$: $m(t) = 1/3$, $D(t) = 2/3$; при $t > 1$: $m(t) = \dfrac{2^{t-3}}{3^{t-2}}$,

$$D(t) = \frac{2^{2t-6}}{3^{2t-4}}\left(1 - \frac{2^{t-2}}{3^{t-1}}\right) + \left(1 - \frac{2^{t-3}}{3^{t-1}}\right)^2\frac{2^{t-3}}{3^{t-1}} + \left(2 - \frac{2^{t-3}}{3^{t-1}}\right)\frac{2^{t-3}}{3^{t-1}},$$

стационарен при $\mathsf{P}\{x_1 = 0\} = 1$, $\mathsf{P}\{x_1 = 1\} = \mathsf{P}\{x_1 = 2\} = 0$.

2.5. а) $m(1) = 7/8$, $m(2) = 3/4$, $m(3) = 11/8$, $D(1) = 55/64$, $D(2) = 7/16$, $D(3) = 31/64$; б) $m(1) = 7/8$, $m(2) = 13/16$, $m(3) = 31/32$, $D(1) = 55/64$, $D(2) = 469/768$, $D(3) = 735/1024$; в) $m(1) = 7/8$, $m(2) = 27/32$, $m(3) = 647/512$, $D(1) = 55/64$, $D(2) = 423/1024$, $D(3) = 143055/262144$; г) $m(1) = 7/8$, $m(2) = 131/128$, $m(3) = 2343/2048$, $D(1) = 55/64$, $D(2) = 93889/131072$, $D(3) = 441399757/536870912$; д) $m(1) = 7/8$, $m(2) = 53/60$, $m(3) = 83/90$, $D(1) = 55/64$, $D(2) = 2411/3600$, $D(3) = 1117/1620$.

2.6. 1) (0,385, 0,336, 0,279), 0,0336.

2.7. а) $\left(\dfrac{p}{0{,}3+p}, \dfrac{0{,}3}{0{,}3+p}\right)$; б) $\left(\dfrac{1-p}{1{,}1-p}, \dfrac{0{,}1}{1{,}1-p}\right)$; в) $\left(\dfrac{1-p}{2-q-p}, \dfrac{1-q}{2-q-p}\right)$.

Глава 3

3.1. а) нет; б) нет; в) да.

3.3. а) $f(x_1, x_2) = (x_1 \oplus x_2)(x_1 \oplus 1)$; б) $f(x_1, \ldots, x_n) = x_1 \cdot x_2 \cdots x_k \cdot (x_{k+1} \oplus \oplus 1) \cdots (x_n \oplus 1) + \ldots + (x_1 \oplus 1) \cdot (x_2 \oplus 1) \cdot \ldots \cdot (x_{n-k} \oplus 1) x_{n-k+1} \cdot \ldots \cdot x_n$ (в сумме C_n^k слагаемых, в каждом из которых ровно k множителей имеют вид x_i, а $n-k$ множителей вид $x_j \oplus 1$).

Глава 4

4.1. $h = 0$.

4.2. а) $h = -\left(\dfrac{1}{15}\log_2\dfrac{1}{3} + \dfrac{2}{15}\log_2\dfrac{2}{3} - 4/5\right)$; б) $h = -\left(\dfrac{1}{10}\log_2\dfrac{1}{5} + \dfrac{2}{5}\log_2\dfrac{4}{5} + \right.$ $\left. + \dfrac{2}{7}\log_2\dfrac{4}{7} + \dfrac{3}{14}\log_2\dfrac{3}{7}\right)$; в) $h = -(0{,}04\log_2 0{,}2 + 0{,}16\log_2 0{,}8 + 0{,}56\log_2 0{,}7 + +0{,}24\log_2 0{,}3)$; г) $h = -(0{,}3\log_2 0{,}75 + 0{,}1\log_2 0{,}25 - 0{,}6)$.

4.3. а) $h = 1$; б) $h = \dfrac{1}{2} - \dfrac{3}{4}\log_2\dfrac{3}{4}$; в) $h = \dfrac{1}{2} - \dfrac{1}{3}\log_2\dfrac{1}{3} - \dfrac{1}{6}\log_2\dfrac{1}{6}$; г) $h = -\sum\limits_{i=1}^{n} p_i \log_2 p_i$.

4.4. а) $B_n = \left\{a \in V_n : 0 < \sum\limits_{i=1}^{n} a_i < 2000\right\}$, $u_{10000} \approx 2^{-5310}$; б) $B_n = \left\{a \in V_n : 2000 < \sum\limits_{i=1}^{n} a_i < 4000\right\}$, $u_{10000} \approx 2^{-1187}$; в) $B_n = \left\{a \in V_n : 9 \leqslant \sum\limits_{i=1}^{n} a_i \leqslant 71\right\}$, $u_{100} \approx 2^{-3}$; г) $B_n = = \left\{a \in V_n : 73 \leqslant \sum\limits_{i=1}^{n} a_i \leqslant 427\right\}$, $u_{1000} \approx 2^{-188}$.

4.6. $h_z \leqslant 2h_x$.

Глава 5

5.1. $P^{(2)} = \begin{pmatrix} 9/20 & 11/20 \\ 11/25 & 14/25 \end{pmatrix}$.

5.2. $\mathbf{H}\{\xi_t | \xi_{t-2}\} = -\varepsilon_2 \log \varepsilon_2 - (1-\varepsilon_2)\log(1-\varepsilon_2)$, где $\varepsilon_2 = -2\varepsilon^2 - 2\varepsilon$.

5.3. $\mathbf{H}\{\xi_t | \xi_{t-\tau}\} = -\sum\limits_{i,j \in V} \pi_i (P^\tau)_{ij} \log (P^\tau)_{ij}$.

5.4. $\mathbf{H}\{a_1\} = \mathbf{H}\{a_2\} = \log 2$.

5.5. $\pi = \begin{pmatrix} 2/5 & 9/25 & 6/25 \end{pmatrix}$, $\mathbf{H}\{\xi_{t+1} | \xi_t\} = h = \dfrac{3}{5}\log 3$.

5.6. $\mathbf{H}\{\xi_{t+1} | \xi_t\} = h = \dfrac{7}{4}\log 2$, $\mathbf{H}\{\Xi_n\} = 2\log 2 + (n-1)\dfrac{7}{4}\log 2$.

5.7. $\mathbf{I}\{\xi_{t+1}, \xi_t\} = \log 2 + \varepsilon \log \varepsilon + (1-\varepsilon)\log(1-\varepsilon)$, $\mathbf{H}\{\xi_t | \xi_t + \xi_{t-1}\} = (1-\varepsilon)\log 2$.

5.8. $h = -\varepsilon \log \varepsilon - (1-\varepsilon)\log(1-\varepsilon)$.

5.9. $\mathsf{I}\{\xi_{t+1},\xi_t\} = \dfrac{1}{2}(1+\varepsilon)\log(1+\varepsilon)+\dfrac{1}{2}(1-\varepsilon)\log(1-\varepsilon)$, $\mathsf{H}\{\xi_t|\xi_t \oplus \xi_{t-1}\} = \log 2$.

5.10. $\mathsf{H}\{\xi_n|\Xi_{n-1}\} = h = \mathsf{H}\{\xi_{s+1}|\xi_s,\ldots,\xi_1\}$.

5.11. $h = -\sum\limits_{i_1,\ldots,i_s,i_{s+1}\in V} \pi_{i_1,\ldots,i_s} q_{i_{m_1^0},\ldots,i_{m_r^0},i_{s+1}} \log q_{i_{m_1^0},\ldots,i_{m_r^0},i_{s+1}}$.

Глава 6

6.1. а) МАТЕМАТИКА; б) КРИПТОЛОГИЯ; в) ИНФОРМАЦИЯ; г) РАСПРЕДЕЛЕНИЕ; д) ПРЕОБРАЗОВАНИЕ; е) ШИФРТЕКСТ.

6.2. а) ЭНТРОПИЯ; б) СЕКРЕТНОСТЬ; в) ИЗБЫТОЧНОСТЬ; г) ПОДСТАНОВКА; д) СЛУЧАЙНОСТЬ; е) КОРРЕКТНОСТЬ.

6.3. а) ОПРСШЛИУНВИАЪЕЪЪ; б) ТАФКЗРОИЯЦАИ; в) ЕДПСЛОЕЛОАТВТЬЬОСН.

6.4. а) СООБЩЕНИЕЪЪЪ; б) АНАГРАММАЪ; в) ПЕРЕСТАНОВКА.

6.5. а) ПИМЙПЫ; б) ОЕСИГЮЩХЩЩ; в) УАФЗЧТРДКМЁИЮЧ.

6.6. а) ПЕРИОД; б) СЕКРЕТ; в) ПОДГОТОВКА.

6.7. 48.

6.10. $p(x|y) = p(y|x) = 1/5$, если $(x,y) \in \{(0,2), (1,0), (2,1)\}$; иначе $p(x|y) = p(y|x) = 2/5$ (см. таблицу).

$X\backslash\theta$	0	1	2	3	4
0	0	1	2	0	1
1	1	2	0	1	2
2	2	0	1	2	0

6.11. При $n = m$.

6.12. При $a \neq 0$, $b \neq 0$ и таких, что a и n взаимно простые, т. е. $(a, n) = 1$.

6.13. Неверно.

6.14. Неверно.

6.15. При $\mathsf{P}\{X = x\} = 1/2$ и $\mathsf{P}\{\theta = z\} \neq 1/2$ криптосистема строго идеальна, но не совершенна; при $\mathsf{P}\{X = x\} \neq 1/2$ и $\mathsf{P}\{\theta = z\} = 1/2$ система совершенна, но не строго идеальна.

6.16. а) $U_{\text{rus}} = 34$, $U_{\text{eng}} = 26$; б) $U_{\text{rus}} = U_{\text{eng}} = 3$; в) $U_{\text{rus}} = U_{\text{eng}} = 2$; г) $U_{\text{rus}} = U_{\text{eng}} = \infty$.

Глава 7

7.1. а) Префиксный — нет, суффиксный — нет, однозначно декодируемый — да; б) префиксный — нет, суффиксный — нет, однозначно декодируемый — нет; в) префиксный — нет, суффиксный — нет, однозначно декодируемый — да; г) префиксный — нет, суффиксный — нет, однозначно декодируемый — нет; д) префиксный — нет, суффиксный — нет, однозначно декодируемый — да; е) префиксный — да, суффиксный — нет, однозначно декодируемый — да; ж) префиксный — нет, суффиксный — нет, однозначно декодируемый — да.

7.2. Например, {00, 001, 11, 011}.

7.3. Данный код не может быть ни однозначно декодируемым, ни префиксным.

7.4. Возможные коды: а) {0, 10, 110, 111}; б) {0, 10, 1100, 1101, 1110}; в) {00, 01, 100, 1010, 1011}; г) {00, 01, 10, 1100, 1101, 1110}; д) {00, 010, 011, 100, 1010, 1011}; е) такого кода не существует, не выполнено неравенство Крафта; ж) {0, 100, 101, 1100, 1101, 1110, 1111}; з) {00, 01, 100, 101, 1100, 1101, 1110, 1111}; и) {00, 010, 0110, 0111, 1000}.

7.5. Возможные коды: а) такого кода не существует; б) {0, 1, 20, 210, 211, 212, 220}; в) {0, 1, 20, 21, 22, 23, 30, 31, 320, 321, 322}; г) {00, 01, 02, 10, 11, 120, 121, 122, 200, 201, 202, 2100}; д) такого кода не существует; е) {0, 10, 11, 12, 200, 201, 2020, 2021}.

7.6. а) да; б) да; в) нет; г) да; д) да; е) да.

7.7. а) $D = 3$; б) $D = 2$; в)$D = 3$; г) $D = 4$.

7.8. Например, а) { 0, 1, 2, 30, 31, 32, 330, 331, 332, 3330, 3331, 3332, 33330, 33331, 33332, 33333}; б) нет; в) {0, 10, 11, 12, 13, 20, 21, 22, 230, 231, 232, 2330, 2331, 2332, 2333, 3000, 3001, 3002, 30030, 30031, 30032, 30033}; г) {00, 01, 02, 03, 10, 11, 12, 130, 131, 132, 1330, 1331, 1332, 1333, 2000, 2001, 2002, 2003, 2010, 2011, 2012, 20130, 20131, 20132, 201330, 201331, 201332, 201333}; д) нет.

7.11. а) коды Фано {0, 10, 110, 111}, средняя длина 2; коды Шеннона {00, 01, 10, 11}, средняя длина 2; коды Хаффмана {0, 01, 000, 001}, средняя длина 2;

б) коды Фано {0, 10, 110, 111}, средняя длина 1,5; коды Шеннона {0, 10, 110, 111}, средняя длина 1,5; коды Хаффмана {0, 11, 100, 101}, средняя длина 1,5;

в) коды Фано {0, 10, 110, 111}, средняя длина 1,425; коды Шеннона {0, 10, 1110, 1111}, средняя длина 1,525; коды Хаффмана {0, 10, 110, 111}, средняя длина 1,425;

г) коды Фано {00, 01, 10, 110, 111}, средняя длина 133/60; коды Шеннона {00, 01, 10, 110, 111}, средняя длина 133/60; коды Хаффмана {00, 01, 11, 100, 101}, средняя длина 133/60;

д) коды Фано {0, 10, 110, 1110, 1111}, средняя длина 1,875; коды Шеннона {0, 10, 110, 1110, 1111}, средняя длина 1,875; коды Хаффмана {0, 10, 110, 1110, 1111}, средняя длина 1,875;

е) коды Фано {00, 01, 10, 110, 111}, средняя длина 2,4; коды Шеннона {000, 001, 01, 10, 11}, средняя длина 2,4; коды Хаффмана {01, 10, 11, 000, 001}, средняя длина 2,4;

ж) коды Фано {00, 01, 10, 110, 111}, средняя длина 161/72; коды Шеннона {00, 01, 10, 110, 111}, средняя длина 161/72; коды Хаффмана {1, 000, 001, 010, 011}, средняя длина 160/72;

з) коды Фано {0, 100, 101, 110, 1110, 1111}, средняя длина 2,11; коды Шеннона {0, 100, 101, 110, 1110, 1111}, средняя длина 2,11; коды Хаффмана {0, 11, 1000, 1001, 1010, 1011}, средняя длина 2,1;

и) коды Фано {0, 100, 101, 110, 1110, 1111}, средняя длина 2,4; коды Шеннона {00, 01, 100, 101, 110, 111}, средняя длина 2,4; коды Хаффмана {1, 01, 0000, 0001, 0010, 0011}, средняя длина 2,4;

к) коды Фано {0, 10, 1100, 1101, 1110, 11110, 11111}, средняя длина 2,27; коды Шеннона {00, 01, 10, 110, 1110, 11110, 11111}, средняя длина 2,4; коды Хаффмана {1, 00, 0100, 0101, 0111, 01100, 01101}, средняя длина 2,27;

л) коды Фано {0, 10, 1100, 1101, 11100, 11101, 11110, 11111}, средняя длина 2,36; коды Шеннона {00, 01, 101, 1100, 1101, 11100, 11101, 1111}, средняя длина 2,6; коды Хаффмана {1, 00, 0100, 0111, 01010, 01011, 01100, 01101}, средняя длина 2,36;

м) коды Фано {00, 01, 10, 1100, 1101, 1110, 11110, 11111}, средняя длина 2,54; коды Шеннона {00, 01, 10, 1100, 1101, 1110, 11110, 11111}, средняя длина 2,54; коды Хаффмана {00, 01, 11, 101, 10000, 10001, 10010, 10011}, средняя длина 2,54.

7.12. а) $D = 2$, {0, 10, 110, 1110, 1111}, средняя длина 1,875; $D = 3$, {0, 1, 20, 21, 22}, средняя длина 1,25; $D = 4$, {0, 1, 2, 30, 31}, средняя длина 1,125;

б) $D = 2$, {00, 10, 11, 010, 011}, средняя длина 2,3; $D = 3$, {1, 2, 00, 01, 02}, средняя длина 1,5; $D = 4$, {0, 2, 3, 10, 11}, средняя длина 1,3;

в) $D = 2$, {1, 001, 010, 011, 0000, 0001}, средняя длина $59/24$; $D = 3$, {1, 00, 01, 02, 20, 21}, средняя длина $39/24$; $D = 4$, {0, 2, 3, 10, 11, 12}, средняя длина $32/24$;

г) $D = 2$, {0111, 0110, 010, 11, 10, 00}, средняя длина $51/21$; $D = 3$, {021, 020, 01, 00, 2, 1}, средняя длина $34/21$; $D = 4$, {12, 11, 10, 3, 2, 0}, средняя длина $27/21$;

д) $D = 2$, {10, 001, 0001, 11, 01, 0000}, средняя длина 2,55; $D = 3$, {2, 02, 011, 00, 1, 010}, средняя длина 1,75; $D = 4$, {2, 00, 02, 3, 1, 01}, средняя длина 1,35;

е) $D = 2$, {1, 01, 0000, 0001, 0010, 0011}, средняя длина 2,4; $D = 3$, {0, 2, 11, 12, 100, 101}, средняя длина 1,6; $D = 4$, {0, 2, 3, 10, 11, 12}, средняя длина 1,3;

ж) $D = 2$, {00, 01, 100, 101, 110, 111}, средняя длина 2,4; $D = 3$, {1, 2, 01, 02, 000, 001}, средняя длина 1,6; $D = 4$, {0, 1, 3, 20, 21, 22}, средняя длина 1,3;

з) $D = 2$, {00, 10, 010, 011, 110, 111}, средняя длина 2,5; $D = 3$, {1, 2, 01, 02, 000, 001}, средняя длина 1,7; $D = 4$, {1, 2, 3, 00, 01, 02}, средняя длина 1,35;

и) $D = 2$, {01, 10, 000, 001, 110, 111}, средняя длина 2,52; $D = 3$, {1, 2, 01, 02, 000, 001}, средняя длина 1,72; $D = 4$, {1, 2, 3, 00, 01, 02}, средняя длина 1,36;

к) $D = 2$, {00, 10, 011, 110, 111, 0100, 0101}, средняя длина 2,7; $D = 3$, {1, 00, 01, 02, 20, 21, 22}, средняя длина 1,7; $D = 4$, {1, 2, 3, 00, 01, 02, 03}, средняя длина 1,4;

л) $D = 2$, {11, 100, 0001, 001, 01, 101, 0000}, средняя длина 2,73; $D = 3$, {00, 02, 12, 01, 2, 10, 11}, средняя длина 1,75; $D = 4$, {2, 00, 03, 3, 1, 01, 02}, средняя длина 1,4;

м) $D = 2$, {00, 01, 100, 101, 110, 1110, 1111}, средняя длина 2,55; $D = 3$, {1, 2, 01, 02, 000, 001, 002}, средняя длина 1,65; $D = 4$, {0, 2, 3, 10, 11, 12, 13}, средняя длина 1,3;

н) $D = 2$, {1, 00, 011, 01000, 01001, 01010, 01011}, средняя длина 2,02; $D = 3$, {0, 1, 20, 22, 210, 211, 212}, средняя длина 1,34; $D = 4$, {0, 1, 3, 20, 21, 22, 23},

средняя длина 1,13;

о) $D = 2$, {01, 1011, 0000, 001, 11, 100, 0001, 1010}, средняя длина 2,85; $D = 3$, {2, 101, 11, 01, 00, 02, 12, 100}, средняя длина 1,87; $D = 4$, {1, 011, 02, 3, 2, 00, 03, 010}, средняя длина 1,54;

п) $D = 2$, {10, 11, 000, 001, 011, 0100, 01010, 01011}, средняя длина 2,79; $D = 3$, {2, 00, 01, 02, 10, 11, 120, 121}, средняя длина 1,85; $D = 4$, {1, 2, 3, 00, 01, 02, 030, 031}, средняя длина 1,48;

р) $D = 2$, {11, 000, 001, 011, 100, 101, 0100, 0101}, средняя длина 2,94; $D = 3$, {2, 00, 01, 10, 11, 12, 020, 021}, средняя длина 1,94; $D = 4$, {1, 2, 3, 01, 02, 03, 000, 001}, средняя длина 1,64;

с) $D = 2$, {00, 10, 010, 110, 111, 0110, 01110, 01111}, средняя длина 2,72; $D = 3$, {1, 00, 01, 02, 20, 21, 220, 221}, средняя длина 1,77; $D = 4$, {1, 2, 3, 00, 01, 02, 030, 031}, средняя длина 1,42.

7.13. Например, {0, 10, 110, 111} и {00, 01, 10, 11}. Средняя длина кодового слова 2. Да, являются оптимальными.

7.14. Например, {00, 01, 10, 110, 1110, 11110, 11111}, {00, 01, 100, 101, 110, 1110, 1111} и {00, 01, 10, 1100, 1101, 1110, 1111}. Средняя длина кодового слова 2,6. Да, являются оптимальными.

7.15. Например, {00, 01, 100, 101, 110, 1110, 1111} и {0, 10, 1100, 1101, 1110, 11110, 11111}. Средняя длина первого 2,61, второго – 2,81. Первый является оптимальным, второй – нет.

7.16. При $m = 100$, средняя длина символа равна 6,72, а энтропия $\mathrm{H}\{\xi\} \approx \approx 6{,}6439$.

7.17. {00, 10, 11, 010, 011}.

7.18. (1/4, 1/4, 1/4, 1/8, 1/8).

7.19. 2.

7.22. Например, $m = 2$, $k = 1$, $p_1 = p_2 = 1/2$.

7.23. а) (0,0,a), (0,0,b), (0,0,r), (5,1,c), (3,1,d), (1,2,r), (0,0,a), длина кода 98 бит; б) (0,0,a), (0,0,b), (7,1,a), (4,1,a), (4,2,b), (5,3,b), (0,0,b), (0,0,a) длина кода 112 бит; в) (0,0,e), (0,0,n), (0,0,i), (0,0,k), (6,1,b), (2,3,k), (0,0,i), длина кода 98 бит; г) (0,0,m), (0,0,e), (0,0,s), (7,1,m), (4,3,a), (0,0,g), (0,0,e), длина кода 98 бит.

7.24. а) (0,a), (0,b), (0,r), (1,c), (1,d), (1,b), (3,a), длина кода 84 бита; б) (0,a), (0,b), (2,a), (1,a), (3,b), (5,b), (3,“), длина кода 84 бита; в) (0,e), (0,n), (0,i), (0,k), (3,b), (1,n), (3,k), (3, “), длина кода 92 бита; г) (0,m), (0,e), (0,s), (3,m), (2,s), (3,a), (0,g), (2,“), длина кода 92 бита.

7.25. а) ch(a), ch(b), ch(r), ch(a), ch(c), ch(a), ch(d), 256, 258, длина кода 81 бит; б) ch(a), ch(b), ch(b), ch(a), 259, 258, 261, 257, ch(a), длина кода 81 бит; в) ch(e), ch(n), ch(i), ch(k), ch(i), ch(b), 256, 258, ch(i), длина кода 81 бит; г) ch(m), ch(e), ch(s), ch(s), 256, 258, ch(a), ch(g), ch(e), длина кода 81 бит.

Глава 8

8.1. $k = 4$, $|\mathcal{C}| = 2^k = 2^4$, $d = 2$.

8.2. Одна из порождающих матриц кода $\mathcal{C}^{\perp}$ может быть

$$G_1 = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 1 \end{pmatrix},$$

которая является проверочной матрицей кода $\mathcal{C}$.

8.3. $\dfrac{(2^n-1)(2^n-2)\cdots(2^n-2^{k-1})}{(2^k-1)(2^k-2)\cdots(2^k-2^{k-1})}$.

8.4. $\dfrac{(2^n-1)(2^n-2)\cdots(2^n-2^{n-1})}{n!}$.

8.5. $k=2$, $G = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix}$, $H = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix}$. Кодовое расстояние $d=3$, которому, например, соответствует кодовое слово $g_2 = (01101)$.

8.6. $H = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix}$,

$$S = \begin{pmatrix} 00000 & 10111 & 01101 & 11010 \\ 10000 & 00111 & 11101 & 01010 \\ 01000 & 11111 & 00101 & 10010 \\ 00100 & 10011 & 01001 & 11110 \\ 00010 & 10101 & 01111 & 11000 \\ 00001 & 10110 & 01100 & 11011 \\ 10100 & 00011 & 11001 & 01110 \\ 10001 & 00110 & 11100 & 01011 \end{pmatrix}.$$

8.7. $G = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$, $H = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}$. Кодовое расстояние $d=2$.

8.8. $n=5$, $k=3$, $H = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \end{pmatrix}$, $d=2$.

8.9. $H = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{pmatrix}$, $s(01000) = (01000)H^T = (101)$, $s(01000) = s(00101) = s(10010) = s(11111) = (101)$.

8.10. $\mathcal{D}_{\rho}(2121) = (0121)$, $\mathcal{D}_{\rho}(1201) = (1201)$, $\mathcal{D}_{\rho}(2222) = (2220)$.

8.11. $H = \begin{pmatrix} 0 & 1 & 2 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{pmatrix}$, $d=2$, $s(21100) = s(00211) = s(22010) = (00)$, $s(02110) = (22)$.

8.12. $H = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$, $s(01000) = s(11111) = (010)$, $s(00101) = s(10010) = (101)$. Так как кодовое расстояние d равно 2, то для принятых векторов декодер $\mathcal{D}_\rho$ только обнаруживает ошибку.

8.13. $g(x) = 1 + x + x^2 + x^3$, $h(x) = 1 + x + x^4 + x^5$, $d = 2$, $H = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \end{pmatrix}$.

8.14. $g(x) = 1 + x$, $h(x) = 1 + x + x^2 + x^3 + x^4$, $H = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \end{pmatrix}$.

8.15. Код не является циклическим.

8.16. $h(x) = 1 + x^2 + x^3$, $G = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}$, $H = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}$, $d = 3$.

8.17. $G = \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \end{pmatrix}$, $g(x) = 1 + 2x^2$.

8.18. $g(x) = x^{10} + x^8 + x^5 + x^4 + x^2 + 1$,

$$G = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \end{pmatrix},$$

$h(x) = x^5 + x^3 + x + 1$,

$$H = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

Глава 9

9.1. $p^k(1-p)^{n-k}$.

9.2. $p^k(1-p)^{n-k}C_n^k$, где C_n^k — биномиальные коэффициенты.

9.3. а) $\mathbf{P}\{X=x_1\}=\gamma=\dfrac{2^C\beta+\beta-1}{(\beta-\alpha)(1+2^C)}$, $\mathbf{P}\{X=x_2\}=1-\gamma$,

$$C^*=-\left(\alpha\gamma\log_2(\alpha(1+2^C))+(1-\alpha)\gamma\log_2\frac{(1-\alpha)(1+2^C)}{2^C}+\right.$$

$$\left.+(1-\gamma)\beta\log_2(\beta(1+2^C))+(1-\gamma)(1-\beta)\log_2\frac{(1-\beta)(1+2^C)}{2^C}\right),$$

$$C=\frac{\beta\log_2\beta+(1-\beta)\log_2(1-\beta)-\alpha\log_2\alpha-(1-\alpha)\log_2(1-\alpha)}{\alpha-\beta};$$

б) $\mathbf{P}\{X=x_1\}=3/5$, $\mathbf{P}\{X=x_2\}=2/5$, $C^*=-0{,}7-0{,}4\log 0{,}4-0{,}1\log 0{,}1\approx$ $\approx$ 0,160964; в) $C^*=0$ для любого распределения входного сигнала; г) $\mathbf{P}\{X=x_1\}=0{,}5$, $\mathbf{P}\{X=x_2\}=0$, $\mathbf{P}\{X=X_3\}=0{,}5$, $C^*=1$; д) $\mathbf{P}\{X=x_1\}=$ $=$ 0,5, $\mathbf{P}\{X=x_2\}=0$, $\mathbf{P}\{X=X_3\}=0{,}5$, $C^*=0{,}75$; е) $\mathbf{P}\{X=x_1\}=0{,}5$, $\mathbf{P}\{X=x_2\}=0$, $\mathbf{P}\{X=X_3\}=0{,}5$, $C^*=0{,}75$; ж) $\mathbf{P}\{X=x_1\}=\mathbf{P}\{X=X_3\}=$ $=\dfrac{2^a-1}{2^{a+1}+1}\approx 0{,}363561$, $\mathbf{P}\{X=x_2\}=\dfrac{3}{2^{a+1}+1}\approx 0{,}272878$, $a=3(-2+\dfrac{7}{4}\log_2 3)$, $C^*\approx 0{,}326305$; з) $\mathbf{P}\{X=x_1\}=\mathbf{P}\{X=X_3\}=1/2$, $\mathbf{P}\{X=x_2\}=0$, $C^*=2/3$; и) $\mathbf{P}\{X=x_1\}\approx 0{,}240701$, $\mathbf{P}\{X=x_2\}\approx 0{,}364079$, $\mathbf{P}\{X=X_3\}\approx 0{,}395220$, $C^*\approx 0{,}131951$; к) $\mathbf{P}\{X=x_1\}=0{,}5$, $\mathbf{P}\{X=x_2\}=0$, $\mathbf{P}\{X=X_3\}=0{,}5$, $C^*=\log_2 3+\dfrac{1}{2}\log_2\dfrac{1}{2}+\dfrac{1}{3}\log_2\dfrac{1}{3}+\dfrac{1}{6}\log_2\dfrac{1}{6}$; л) $\mathbf{P}\{X=x_1\}=\mathbf{P}\{X=x_2\}=$ $=\mathbf{P}\{X=X_3\}=\dfrac{1}{3}$, $C^*=\dfrac{2}{3}(\log_2\dfrac{1}{3}-\log_2\dfrac{2}{9})\approx 0{,}389975$; м) $\mathbf{P}\{X=x_1\}=\ldots=$ $=\mathbf{P}\{X=X_n\}=\frac{1}{n}$, $C^*=\log_2 n$; н) $C^*=0$ для любого распределения входного сигнала;

о) для $0\leqslant p\leqslant 1/3$:

$$\mathbf{P}\{X=x_1\}=\frac{1-3p}{3(1-2p)},\mathbf{P}\{X=x_2\}=\frac{1-3p+3p^2}{3(1-2p)(1-p)},$$

$$\mathbf{P}\{X=X_3\}=\frac{1}{3(1-p)},\ C^*(p)=\log_2 3+p\log_2 p+(1-p)\log_2(1-p);$$

для $1/3<p\leqslant p^*$, где p^* решение уравнения $1-p=1+p\log_2 p+(1-p)\log_2(1-p)$:

$$\mathbf{P}\{X=x_1\}=0,\ \mathbf{P}\{X=x_2\}=\frac{1}{2},\ \mathbf{P}\{X=X_3\}=\frac{1}{2},C^*(p)=1-p;$$

для $p^*<p\leqslant 1$:

$$\mathbf{P}\{X=x_1\}=\mathbf{P}\{X=x_2\}=\frac{1}{2},\ \mathbf{P}\{X=X_3\}=0,$$

$$C^*(p)=1+p\log_2 p+(1-p)\log_2(1-p);$$

график зависимости C^* от p изображен на рис. 1П;

п) $\mathbf{P}\{X=x_1\}=\mathbf{P}\{X=x_2\}=\dfrac{2^{-h(p)}}{1+2^{1-h(p)}}$, $\mathbf{P}\{X=x_3\}=\dfrac{1}{1+2^{1-h(p)}}$,

$$C^*(p)=-\left(\frac{2^{1-h(p)}}{1+2^{1-h(p)}}\log_2\frac{2^{-h(p)}}{1+2^{1-h(p)}}+\frac{1}{1+2^{1-h(p)}}\log_2\frac{1}{1+2^{1-h(p)}}\right)+$$

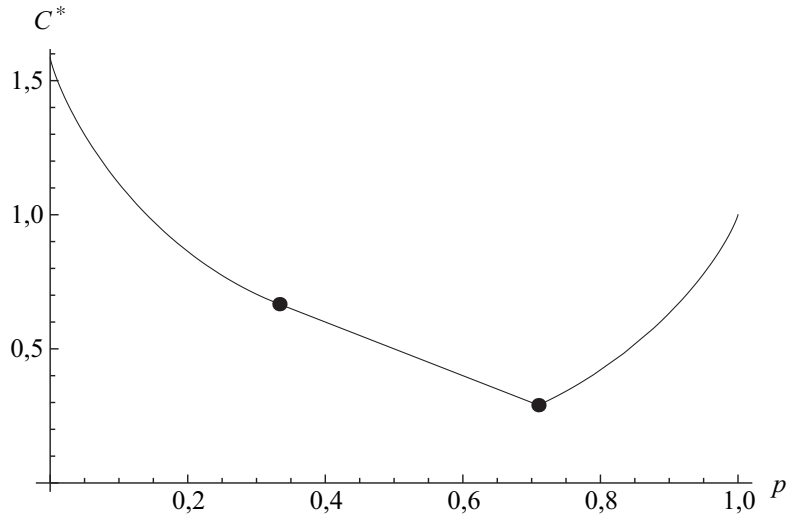


Рис. 1П. График зависимости пропускной способности от параметра p

$$+\frac{2^{1-h(p)}p}{1+2^{1-h(p)}}\log_2 p+\frac{2^{1-h(p)}(1-p)}{1+2^{1-h(p)}}\log_2(1-p),$$

$$h(p)=-p\log_2 p-(1-p)\log_2(1-p);$$

график зависимости C^* от p изображен на рис. 2П.

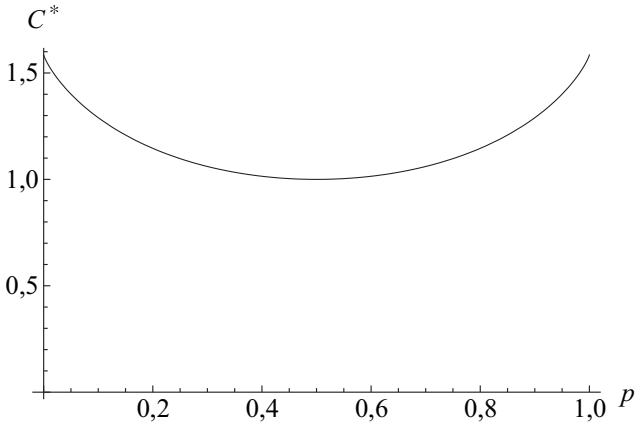


Рис. 2П. График зависимости пропускной способности от параметра p

9.4. $\mathbf{H}\{X\}=3/2, \mathbf{H}\{Y\}=1, \mathbf{H}\{Z\}=1, \mathbf{H}\{Y,Z\}=2, \mathbf{I}\{X,Y\}=0, \mathbf{I}\{X,Z\}=0, \mathbf{I}\{X,(Y,Z)\}=1.$

9.5. $\mathbf{I}\{X,Y\}=(3/4)\log_2 3-1$, $\mathbf{I}\{Y,Z\}=1+\frac{2}{3}\log_2\frac{2}{3}+\frac{1}{3}\log_1\frac{2}{3}$, $\mathbf{I}\{X,Z\}=$ $=1+\frac{7}{12}\log_2\frac{7}{12}+\frac{5}{12}\log_1\frac{5}{12}$.

9.6. $C^*(p) = 1 - h\left(\sum\limits_{k=0}^{\lfloor n/2\rfloor} C_n^{2k}(1-p)^{2k}p^{n-2k}\right)$, где C_n^k — биномиальные коэффициенты; $\lim_{n\to\infty} C^*(p) = \begin{cases} 0, & 0 < p < 1, \\ 1, & p = 0 \text{ или } p = 1. \end{cases}$

9.7. $\mathsf{P}\{X=0\} = \ldots = \mathsf{P}\{X=4\} = 0{,}2$, $C^* = \log_2 5 - 1$.

9.8. $\mathsf{P}\{X=0\} = \ldots = \mathsf{P}\{X=10\} = 1/11$, $C^* = \log_2 11 - \log_2 3$.

9.9. $C^*(a) = \begin{cases} 1/2, & a = 1, \\ 1, & a \neq 1. \end{cases}$

9.10. $C^*(a) = -\left(\dfrac{1}{1+2^{\frac{h(p)}{1-p}}}\log_2\dfrac{1}{1+2^{\frac{h(p)}{1-p}}} + \dfrac{2^{\frac{h(p)}{1-p}}}{1+2^{\frac{h(p)}{1-p}}}\log_2\dfrac{2^{\frac{h(p)}{1-p}}}{1+2^{\frac{h(p)}{1-p}}}\right) + \dfrac{p}{(1-p)\left(1+2^{\frac{h(p)}{1-p}}\right)}\log_2 p + \dfrac{1}{1+2^{\frac{h(p)}{1-p}}}\log_2(1-p)$ (рис. 3П).

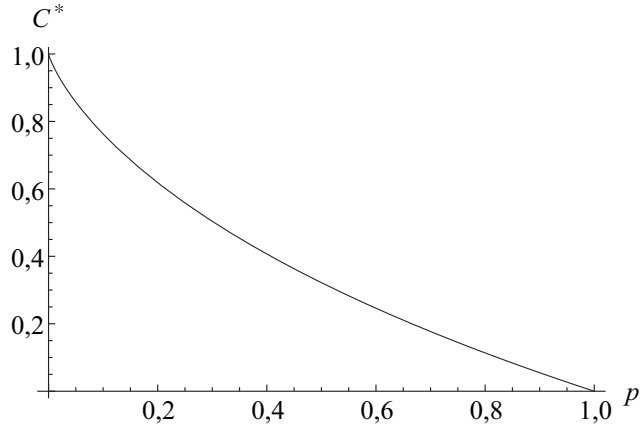


Рис. 3П. График зависимости пропускной способности от параметра p

9.11. $\lambda_1 = \lambda_2 = \lambda = 1 - (1-p)^2(1+2p)$.

9.12. $\lambda^{(a)} = 0{,}028$, $\lambda^{(b)} = 0{,}792$, $\lambda^{(a)}(p) \leqslant \lambda^{(b)}(p)$, при $p \leqslant 1/2$.

9.13. а) $A_1 = \{(000),(001),(010),(100)\}$, $A_2 = \{(011),(101),(110),(111)\}$, $\lambda \approx 0{,}06075$, $0{,}105687 \approx \mathsf{H}\{X^3|Y^3\} < h(\lambda) \approx 0{,}330415$; б) $A_1 = \{(000),(001),(010),(100)\}$, $A_2 = \{(011),(101),(110),(111)\}$, $\lambda \approx 0{,}15625$, $0{,}249364 \approx \mathsf{H}\{X^3|Y^3\} < h(\lambda) \approx 0{,}625662$; в) $A_1 = \{(000)\}$, $A_2 = \{(011),(101),(110),(111),(001),(010),(100)\}$, $\lambda \approx 0{,}288576$, $0{,}38181 \approx \mathsf{H}\{X^3|Y^3\} < h(\lambda) \approx 0{,}866875$; г) $A_1 = \emptyset$, $A_2 = \{(011),(101),(110),(111),(001),(010),(100),(000)\}$, $\lambda = 0{,}2$, $0{,}329897 \approx \mathsf{H}\{X^3|Y^3\} < h(\lambda) \approx 0{,}721928$.

9.14. а) $A_1 = \{(0000),(0001),(0010),(0100),(0101),(0110),(1000),(1001),(1010)\}$, $A_2 = \{(0011),(0111),(1011)\}$, $A_3 = \{(1100),(1101),(1110)\}$, $A_4 = \{(1111)\}$; б) $A_1 = \{(0000),(000e),(00e0),(00ee),(0e00),(0e0e),(0ee0),(0eee),(e000),(e00e),(e0e0),(e0ee),(ee00),(ee0e),(eee0),(eeee)\}$; $A_2 = \{(0011),(001e),$

(00e1), (oe11), (0e1e), (0ee1), (e011), (e01e), (e0e1), (ee11), (ee1e), (eee1)}; A_3 ==={(1100), (110e), (11e0), (11ee), (1e00),(1e0e), (1ee0), (1eee), (e100), (e10e), (e1e0), (e1ee) }; $A_4 = \{(1111), (111e), (11e1), (1e11), (1e1e), (1ee1), (e111), (e11e), (e1e1)\}$.

9.15. $A_1 = \{(00000), (00001), (00010), (00011), (00100), (00110), (01000), (10000), (10001), (10100)\}$, $A_2 = \{(00101), (01001), (01011), (01100), (01101), (01110), (01111), (11001), (11100), (11101) \}$, $A_3 = \{(00111), (10011), (10101), (10110), (10111), (11111) \}$; $A_4 = \{(01010), (10010), (11000), (11010), (11011), (11110)\}$.

9.16. $\mathcal{D}_L : A_1 = \{(00000), (00001), (00010), (00011), (00100), (00110), (01000), (01001), (01100), (10000)\}$, $A_2 = \{(00111), (01011), (01101), (01110), (01111), (11111) \}$, $A_3 = \{(00101), (10001), (10111), (10100), (10101), (10110), (10111), (11001), (11100), (11101) \}$; $A_4 = \{(01010), (10010), (11000), (11010), (11011), (11110)\}$.

$\mathcal{D}_{AP}$: $A_1 = \{(00000), (00001), (00010), (00011), (00100), (00101), (00110), (01000), (01001), (01010), (01100), (10000), (10001), (10010), (10011), (10100), (10110), (11000), (11001), (11100)\}$, $A_2 = \{(00111), (01011), (01101), (01110), (01111), (11111) \}$, $A_3 = \{(10101), (10111), (11101)\}$; $A_4 = \{(11010), (11011), (11110)\}$.

9.17. $\lambda_1 = 0{,}5$, $\lambda_2 = 0{,}7$, $\lambda_3 = 1$, $\lambda = 11/15$, $1{,}30096 \approx \mathbf{H}\{X|Y\} < h(\lambda) + \lambda \log_2 2 = h(\lambda) + \lambda \approx 1{,}56997$.

9.18. $A_1 = \{x_1, x_2\}$, $A_2 = \emptyset$, $A_3 = \{x_3\}$, $\lambda_1 = 1/6$, $\lambda_2 = 1$, $\lambda_3 = 1/2$, $\lambda = 11/24$, $1{,}38172 \approx \mathbf{H}\{X|Y\} < h(\lambda) + \lambda \log_2 2 = h(\lambda) + \lambda \approx 1{,}45332$.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Анализ биологических последовательностей / Р. Дурбин [и др.]. М., 2006.

2. *Андерсон Дж.* А. Дискретная математика и комбинаторика : пер. с англ. М., 2004.

3. *Блейхут Р.* Теория и практика кодов, контролирующих ошибки : пер. с англ. М., 1986.

4. *Быховский М. А.* Пионеры информационного века. История развития теории связи. М., 2006.

5. *Верн Ж.* Жангада. Минск, 1992.

6. *Городилова А. А., Токарева Н. Н., Шушуев Г. И.* Криптография и криптоанализ : сб. задач. Новосибирск, 2014.

7. *Дуб Дж.* Вероятностные процессы. М., 1956.

8. *Духин А. А.* Теория информации. М., 2007.

9. *Кейперс Л., Нидеррайтер Г.* Равномерное распределение последовательностей. М., 1985.

10. *Кемени Дж., Снелл Дж.* Конечные цепи Маркова. М., 1970.

11. *Кнут Д.* Искусство программирования : в 3 т. М., 2000. Т. 2 : Получисленные алгоритмы.

12. *Колесник В. Д., Полтырев Г. Ш.* Курс теории информации. М., 1982.

13. *Королюк В. С.* Справочник по теории вероятностей и математической статистике. М., 1985.

14. Криптология : учебник / Ю. С. Харин [и др.]. Минск, 2013.

15. Лекции по теории графов / В. А. Емеличев [и др.]. М., 1990.

16. *Лидл Р., Нидеррайтер Г.* Конечные поля : в 2 т. М., 1988. Т. 1, 2.

17. *Лидовский В. В.* Теория информации : учеб. пособие. М., 2003.

18. *Максимов Ю. И.* О цепях Маркова, связанных с двоичными регистрами сдвига со случайными элементами // Тр. по дискретной математике. М., 1997. Т. 1. С. 203–220.

19. *Марков А. А.* Введение в теорию кодирования. М., 1982.

20. *Никитин Г. И.* Сверточные коды : учеб. пособие. СПб., 2001.

21. Основы криптографии / А. П. Алферов [и др.]. М., 2001.

22. *Орлов В. А., Филиппов Л. И.* Теория информации в упражнениях и задачах. М., 1976.

23. *Сачков В. Н.* Введение в комбинаторные методы дискретной математики. М., 2004.

24. *Стратонович Р. Л.* Теория информации. М., 1975.

25. *Тарасенко Ф. П.* Введение в курс теории информации. Томск, 1973.

26. *Харин Ю. С.* Вероятностно-статистический анализ цепей Маркова высокого порядка // Вестн. БГУ. Сер. 1. 2006. № 3. С. 80–86.

27. *Харин Ю. С.* Цепи Маркова с *r*-частичными связями и их статистическое оценивание // Докл. НАН Беларуси. 2004. Т. 48, № 1. С. 40–44.

28. *Харин Ю. С., Берник В. И., Матвеев Г. В.* Математические основы криптологии. М., 1999.

29. *Харин Ю. С., Гурин А. С.* Статистические оценки параметров векторных авторегрессионных временных рядов при наличии пропущенных значений и их асимптотические свойства // Докл. НАН Беларуси. 2006. Т. 50. № 1. С. 18–24.

30. *Харин Ю. С., Зуев Н. М., Жук Е. Е.* Теория вероятностей, математическая и прикладная статистика. Минск, 2011.

31. *Харин Ю. С., Петлицкий А. И.* Цепь Маркова *s*-го порядка с *r* частичными связями и их статистическое оценивание // Дискретная математика. 2007. Т. 12, вып. 2. С. 109–130.

32. *Харин Ю. С., Степанова М. Д.* Практикум на ЭВМ по математической статистике. Минск, 1987.

33. *Чечета С. И.* Введение в дискретную теорию информации и кодирование : учеб. пособие. М., 2011.

34. *Шеннон К.* Работы по теории информации и кибернетики. М., 1963.

35. *Ширяев А. Н.* Вероятность. М., 1980.

36. *Шоломов Л. А.* Основы теории дискретных логических и вычислительных устройств. М., 1980.

37. Эконометрическое моделирование / Ю. С. Харин [и др.]. Минск, 2003.

38. *Bonachela J. A., Hinrichsen H., Munoz M. A.* Entropy estimates of small data sets // J. of physics A: mathematical and theoretical. 2008. Vol. 41, № 20. P. 85–108.

39. *Borda M.* Fundamentals in information theory and coding. B., 2011.

40. *Buhlmann P., Wyner A.* Variable length Markov chains // The annals of statistics. 1999. Vol. 27, № 2. P. 480–513.

41. *Cover T. M., Thomas J. A.* Elements of information theory. N. Y., 1991.

42. *Jacobs P. A., Lewis P. A. W.* Discrete time series generated by mixtures // J. of the royal statistical society. Ser. B. 1978. Vol. 40, № 1. P. 94–105.

43. *MacKay D. I. C.* Information theory, inference, and learning algorithms. Cambridge, 2004.

44. *Raftery A. E.* A model for high-order Markov chains // J. of the royal statistical society. Ser. B. 1985. Vol. 47, № 3. P. 528–539.

45. *Yeung R. W.* Information theory and network coding. N. Y., 2008.

СОДЕРЖАНИЕ

Глава 3. ПРЕОБРАЗОВАНИЯ СЛУЧАЙНЫХ ПРОЦЕССОВ. ЭРГОДИЧНОСТЬ СЛУЧАЙНЫХ ПРОЦЕССОВ

Глава 4. СТАЦИОНАРНЫЕ ИСТОЧНИКИ СООБЩЕНИЙ И ИХ ЭНТРОПИЙНЫЕ СВОЙСТВА

Глава 5. МАРКОВСКИЕ ИСТОЧНИКИ СООБЩЕНИЙ И ИХ СВОЙСТВА

Глава 6. ШЕННОНОВСКИЙ ПОДХОД К ШИФРОВАНИЮ

Глава 7. ОПТИМАЛЬНОЕ КОДИРОВАНИЕ

Глава 8. ЛИНЕЙНЫЕ КОДЫ, ИСПРАВЛЯЮЩИЕ ОШИБКИ

Глава 9. ДИСКРЕТНЫЕ (БЕЗ ПАМЯТИ) КАНАЛЫ ПЕРЕДАЧИ ИНФОРМАЦИИ И СВЯЗАННЫЕ С НИМИ ТЕОРЕМЫ КОДИРОВАНИЯ

Учебное издание

Харин Юрий Семенович
Бодягин Игорь Александрович
Вечерко Егор Валентинович

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ТЕОРИИ ИНФОРМАЦИИ

Учебное пособие

Редактор *Н. Ф. Акулич*
Художник обложки *Т. Ю. Таран*
Технический редактор *Л. В. Жаборовская*
Компьютерная верстка *И. А. Бодягина*
Корректор *Е. В. Гордейко*

Подписано в печать 25.04.2018. Формат 70×100/16. Бумага офсетная.
Печать офсетная. Усл. печ. л. 24,51. Уч.-изд. л. 25,11.
Тираж 100 экз. Заказ 187

Белорусский государственный университет.
Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий № 1/270 от 03.04.2014.
Пр. Независимости, 4, 220030, Минск.

Республиканское унитарное предприятие
«Издательский центр Белорусского государственного университета».
Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий № 2/63 от 19.03.2014.
Ул. Красноармейская, 6, 220030, Минск.

Харин, Ю. С.

Х20 Математические основы теории информации : учеб. пособие / Ю. С. Харин, И. А. Бодягин, Е. В. Вечерко. – Минск : БГУ, 2018. – 302 с.

ISBN 978-985-566-525-1.

Изложены математические основы теории информации. Рассмотрены решения типовых задач. Представлены задания для тестов, задачи и упражнения, ответы на которые приведены в приложениях.

Для студентов учреждений высшего образования, обучающихся по специальностям «Компьютерная безопасность», «Прикладная криптография».

УДК 519.72(075.8)
ББК 22.18я73-1