А. А. ФЕЛЬДБАУМ
А. Д. ДУДЫКИН
А. П. МАНОВЦЕВ
Н. Н. МИРОЛЮБОВ

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ СВЯЗИ И УПРАВЛЕНИЯ

ФМ

ГЛАВА X

ЭЛЕМЕНТЫ ТЕОРИИ ИНФОРМАЦИИ

§ 1. Общие понятия и количественная мера информации

Общие сведения. Понятие «информация» может быть истолковано как некоторая совокупность сведений, определяющих меру наших знаний о тех или иных событиях, явлениях или фактах. Это определение или, точнее, пояснение лишь немного дополняет имеющееся у каждого интуитивное представление об этом понятии и ничего не дает для построения количественной теории информации, которая могла бы использоваться при решении инженерных задач. Для такой теории необходимо ввести «операционное» определение этого понятия, в основе которого лежит указание на способ измерения.

Легко убедиться в том, что введение количественной меры информации является весьма трудной задачей. Действительно, информация может быть чрезвычайно разнообразной: мы можем получить извещение о приезде знакомых или родственников, о получении премии или болезни друга; можем услышать по радио или прочесть в газетах о тех или иных событиях; узнать о новом открытии или изобретении и т. п. Различная информация будет вызывать у нас различные эмоции и будет представлять различную ценность. Иногда краткое, содержащее лишь несколько слов, извещение может иметь для нас неизмеримо большее значение, чем текст, состоящий из многих слов и страниц. Из двух книг равного объема мы можем извлечь совершенно различную информацию. Два радиосигнала, имеющие одинаковую мощность и длительность, могут переносить различную по содержанию и объему информацию. Очевидно, что количественная мера информации не должна противоречить нашим интуитивным представлениям, должна охватывать то общее, что присуще всему многообразию различной информации, и, главное, эта мера должна быть полезной для теории и практики построения различных систем передачи и преобразования информации.

Первая попытка определения количественной меры информации была предпринята Р. В. Хартли в 1928 г. *), однако данное им определение оказалось недостаточно универсальным.

В 1933 г. В. А. Котельников доказал теорему (см. гл. II), которая приобрела фундаментальное значение в теории передачи непрерывных сообщений.

Математический аппарат современной теории информации, в его основных чертах, впервые был разработан в 1947 г. А. Н. Колмогоровым и изложен в его докладе «Статистическая теория колебаний с непрерывным спектром» на общем собрании Академии наук СССР.

Основные соотношения, определяющие количественную меру информации, и основные теоремы теории информации были сформулированы К. Шэнноном и опубликованы в 1949 г. **). В последующих трудах советских ученых А. Н. Колмогорова, А. Я. Хинчина, А. А. Харкевича, А. М. Яглома, Р. Л. Добродушина, М. С. Пинскера, А. Н. Железнова и др., а также в работах ряда иностранных ученых (А. Файнстейна, С. Голдмана, Л. Бриллюэна, Л. Заде и др.) теория информации получила свое дальнейшее развитие.

Кратко характеризуя содержание теории информации, можно сказать, что в ней рассматриваются:

общая количественная мера информации, не зависящая от природы объектов;

общие законы передачи, обработки и хранения информации. Эти законы имеют характер законов сохранения и устанавливают ряд важнейших предельных соотношений, как-то: максимальную скорость передачи информации, наибольшую возможную эффективность кодирования сигналов и др.;

общие зависимости, определяющие влияние тех или иных факторов на преобразование, скорость передачи, потерю и возможность хранения информации.

Количественная мера информации при равновозможных событиях. Обратим, прежде всего, внимание на то обстоятельство, что всякая информация добывается нами в результате того или иного опыта (действия). Таким опытом может быть получение телеграммы или письма, чтение книги и т. п.; измерение тем или иным инструментом (прибором); простое визуальное наблюдение события или наблюдение с помощью тех или иных средств; прием и обработка радиосигналов и т. д.

Чем можно охарактеризовать факт получения информации? Можно сказать, что во всех случаях до опыта на интересующий нас вопрос или на вопрос, который нам могут задать о некоторой ситуации,

*) Р. Хартли, Передача информации. Теория информации и ее приложения. Сборник переводов под редакцией А. А. Харкевича, Физматгиз, Москва, 1959.

**) Теория передачи электрических сигналов при наличии помех. Сборник переводов под ред. А. Н. Железнова, ИЛ, Москва, 1953.

мы не можем ответить однозначно. В лучшем случае мы можем высказать лишь некоторую совокупность различных предположений (альтернатив). Таким образом, до опыта имеет место бо́льшая или меньшая неопределенность в интересующей нас ситуации или в исходе тех или иных событий. После опыта, т. е. после получения информации, ситуация становится более определенной и на поставленный вопрос мы можем ответить либо однозначно, либо, по крайней мере, количество возможных различных ответов уменьшится и, следовательно, уменьшится и существовавшая ранее неопределенность. Рассмотрим два примера.

П р и м е р 1. Перед наблюдателем поставлены две урны. Известно, что в одну из них (урна A) положены белые шары, а в другую (урна $Б$) — черные. Внешне урны ничем не отличаются, поэтому на вопрос, которая из них урна A и которая $Б$, наблюдатель, не имея никаких дополнительных сведений, может дать два равновозможных ответа.

Для получения необходимой информации достаточно вынуть шар из какой-либо урны. После такого опыта ситуация становится полностью определенной, а ответ — однозначным.

П р и м е р 2. Известно, что интересующий нас размер детали может находиться в пределах 120—180 *мм*. Поэтому на вопрос о том, каков этот размер с точностью до миллиметра, можно высказать 61 предположение: 120, 121, 122, ..., 180 *мм*. Пусть в результате измерения, производимого с точностью до ± 5 *мм*, установлено, что размер равен 130 *см*. Тогда после такого опыта на интересующий нас вопрос можно высказать уже 11 предположений: 125, 126, ..., 135 *мм*.

Таким образом, после получения информации хотя и остается некоторая неопределенность, но эта неопределенность становится меньшей, чем раньше.

Следовательно, для оценки количества получаемой информации необходимо найти меру неопределенности той или иной ситуации. Уменьшение неопределенности в результате опыта может быть принято за наиболее общую меру количества получаемой информации. В этом смысле говорят, что информация обратна неопределенности.

Меру количества информации, получаемой в результате того или иного опыта, можно было бы установить как функцию отношения числа равновозможных ответов до опыта (n) и после опыта (n_e), т. е. как функцию отношения $\frac{n}{n_e}$.

Интуиция подсказывает, что количество получаемой в результате опыта информации должно быть тем больше, чем больше это отношение. Так, например, если мы сравним два случая, когда число равновозможных ответов до опыта $n=2$ и $n=100$, то ясно, что во втором случае ситуация является более неопределенной, и если в результате опыта эта неопределенность полностью снимается

($n_e=1$), то следует считать, что количество полученной информации при $n=100$ больше.

Вместо понятия «равновозможные события» (или ответы) удобнее пользоваться понятием о «вероятности этих событий».

Если рассматриваемые события (ответы) равновозможны, то априорная (доопытная) вероятность события x_i равна, очевидно, $p(x_i)=\frac{1}{n}$, а апостериорная (после опыта или после сообщения) вероятность $p_e(x_i)=\frac{1}{n_e}$. В таком случае количество информации, которое мы получаем о событии x_i в результате опыта, должно быть функцией отношения

$$\frac{n}{n_e}=\frac{p_e(x_i)}{p(x_i)}.$$

Ниже будет показано, что наиболее удобной является логарифмическая функция, и, следовательно, количество информации, получаемое о событии x_i в результате опыта, может быть принято равным

$$I(i)=K\log_a\frac{p_e(x_i)}{p(x_i)}.$$

Выбор коэффициента K и основания логарифмов (a) не имеет принципиального значения, ибо он определяет лишь масштаб или единицу количества информации. Обычно выбирают логарифм с основанием два ($a=2$) и $K=1$, тогда

$$\boxed{I(i)=\log_2\frac{p_e(x_i)}{p(x_i)}\ \text{дв. ед.}^{*)}} \tag{10.1}$$

Единица количества информации при таком выборе a и K называется *двоичной*.

Для рассмотренного выше примера 1 в качестве события можно указать, например, ответ: урной A является урна, стоящая справа или слева; очевидно, что априорная вероятность дать правильный ответ $p(A)=\frac{1}{2}$, а апостериорная (после того как вынут шар) $p_e(A)=1$.

Из (10.1) следует, что получаемое при вытаскивании шара количество информации равно $I(A)=1$ дв. ед.

Таким образом, получение одной двоичной единицы количества информации соответствует тому, что мы узнаем, какое из двух равновозможных событий имеет место или какая из двух равновозможных гипотез правильна. Такими двумя событиями (гипотезами) могут быть также ответы «да» или «нет» на какой-либо вопрос. Если эти ответы равновозможны (равновероятны), то, получив один из них, мы тем самым получаем одну двоичную единицу информации.

*) При практическом использовании формулы (10.1) следует иметь в виду, что $\log_2 x \approx 3{,}32\log_{10} x$.

Для примера 2 в качестве события можно рассматривать ответ: размер детали равен l миллиметров. Очевидно, что в данном случае $p(l)=\frac{1}{61}$, а $p_c(l)=\frac{1}{11}$; тогда согласно (10.1) количество информации, получаемое в результате измерения, равно

$$I(l)=\log_2\frac{61}{11}\approx 3{,}32\log_{10}\frac{61}{11}$$

и, следовательно,

$$I(l)\approx 1{,}84 \text{ дв. ед.}$$

Если события (ответы) равновозможны и если к тому же $p_c(l)=1$, т. е. после опыта ситуация полностью определена, формула (10.1) может быть представлена в виде

$$I(l)=\log_2 n. \qquad (10.2)$$

Такая мера количества информации была предложена в 1928 г. Р. Хартли. Из изложенного и дальнейших рассуждений будет видно, что (10.2) относится лишь к весьма частному случаю. Р. Хартли правильно поставил вопрос о необходимости измерения количества информации и исключения психологических факторов при установлении меры, однако он не смог найти наиболее общего подхода к решению этой задачи, что было значительно позднее сделано К. Шэнноном.

Количественная мера информации при конечном ансамбле независимых событий. Формула (10.1) устанавливает непосредственную связь между количеством информации, получаемой о некотором событии в результате опыта, и изменением вероятности этого события до и после опыта. Эта связь может быть обобщена и на случай, когда имеется некоторое конечное множество независимых событий с разными априорными вероятностями.

Рассмотрим некоторое конечное множество X событий $x_1, x_2, \ldots, x_n$. Такими событиями может быть, например, изменение положения отдельных механизмов на каком-либо объекте: открытие и закрытие клапанов, срабатывание и отпускание реле и т. п.

Допустим, что данные события независимы и несовместны, а априорные вероятности их соответственно равны $p(x_1), p(x_2), \ldots, p(x_n)$, причем $p(x_1)+p(x_2)+\ldots+p(x_n)=1$. Последнее означает, что в течение некоторого наблюдаемого отрезка времени всегда происходит одно из этих событий.

Множество с известным распределением вероятностей его элементов принято называть *ансамблем*.

Рассматриваемый ансамбль событий может быть описан конечной схемой вида

$$X=\begin{pmatrix} x_1 & x_2 & \ldots & x_n \\ p(x_1) & p(x_2) & \ldots & p(x_n)\end{pmatrix}. \qquad (10.3)$$

Ансамбль рассматривается как некоторая модель физической системы, которая может находиться в n различных состояниях или в которой может происходить n различных событий. Пока что мы рассматриваем случай, когда эти состояния или события независимы и несовместны.

Используя формулу (10.1), можно сказать, что достоверное сообщение о том, что из всех событий x происходит событие x_i, несет в себе количество информации, равное

$$I(x_i)=\log\frac{1}{p(x_i)} \text{ дв. ед.}^{*)}. \qquad (10.4)$$

Следовательно, сообщение о событии несет тем большее количество информации, чем меньше априорная вероятность этого события. Это положение хорошо согласуется с интуитивным представлением об информации. Так, например, сообщение о том, что летом дни длиннее ночи, для любого человека с минимальным жизненным опытом или образованием не несет никакой информации, ибо этот факт априори (до этого сообщения) уже ему известен. Сообщение о том, что Ваша жена или знакомая родила трех близнецов, несет значительно большее количество информации, чем сообщение о том, что она родила одного сына или дочь, ибо статистикой установлено, что тройня бывает в среднем одна на 8000 случаев.

Формула (10.4) указывает, что в конечном ансамбле X сообщения о разных событиях в общем случае несут разное количество информации. При решении большинства задач, связанных с построением систем передачи и преобразования информации, необходимо знать среднее количество информации, приходящееся на одно сообщение. Последнее в соответствии с правилами теории вероятности может быть определено как математическое ожидание величины $I(x_i)$, т. е.

$$H(X)=M[I(x_i)]$$

и, следовательно,

$$H(X)=-\sum_{i=1}^{n}p(x_i)\log p(x_i)\,\frac{\text{дв. ед.}}{\text{сообщ.}}. \qquad (10.5)$$

В данном случае через $H(X)$ обозначено среднее количество информации, приходящееся на одно достоверное сообщение о событии x при передаче большого числа таких сообщений. По аналогии с соответствующим соотношением в термодинамике величина $H(X)$ называется *энтропией* конечного ансамбля X.

Следует сразу же отметить, что общность формул, определяющих значение энтропии в термодинамике и теории информации, имеет глубокий физический смысл, связанный с оценкой неопределенности ситуации (см. ниже).

*) Ниже, если не оговаривается противное, знаком log обозначается логарифм при основании два.

В качестве примера определения $H(X)$ рассмотрим ансамбль двух событий x_1 и x_2 с априорными вероятностями $p(x_1)=p$ и $p(x_2)=q=1-p$. В этом случае согласно (10.5)

$$H(X) = -p\log p - (1-p)\log(1-p)\ \frac{\text{дв. ед.}}{\text{сообщ.}}.$$

На рис. 10.1 приведен график зависимости величины $H(X)$ от p. Из этого графика видно, что энтропия $H(X)$ равна нулю при $p=0$ и $p=1$ и имеем максимальное значение $H(X)=1$ при $p=\frac{1}{2}$. Эти результаты нетрудно объяснить: действительно, при $p=0$ всегда происходит лишь событие x_2 и потому сообщения о нем не несут никакой информации. Аналогично при $p=1$ всегда происходит лишь событие x_1. Если $p \ll q$, то событие x_2 происходит часто, а событие x_1 — редко. Сообщения о событии x_2 в этом случае несут очень малое количество информации, а сообщения о событии x_1 хотя и несут большое количество информации, но число этих сообщений очень мало, и, таким образом, количество информации, получаемое в среднем на одно сообщение, также оказывается малым.

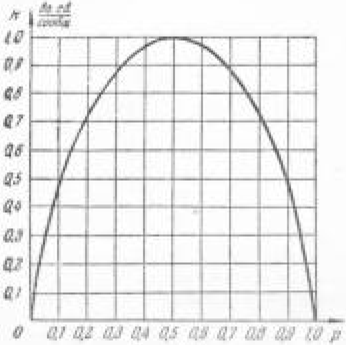

Рис. 10.1.

Аналогичная картина имеет место и при $p \gg q$. Если же $p=q=\frac{1}{2}$, то вероятности событий x_1 и x_2 одинаковы; сообщение о любом из них несет 1 дв. ед. информации, и энтропия $H(X)$ оказывается максимальной.

Определение количественной меры общей неопределенности ситуации. Для того чтобы более строго доказать (10.5), нужно ввести количественную меру неопределенности исхода конечного ансамбля событий. Введя такую меру, как уже указывалось выше, мы сможем оценивать количество получаемой информации через изменение неопределенности, наступающее в результате того или иного опыта и, в частности, в результате передачи и приема сообщения.

Количественная оценка неопределенности исхода событий, удовлетворяющая некоторым общим требованиям, была построена К. Шэнноном, затем более строго — А. Я. Хинчиным *) и, наконец,

*) А. Я. Хинчин, Понятие энтропии в теории информации, Успехи матем. наук, 1953, вып. 3, стр. 3—20.

при наименьшем числе исходных предпосылок — А. Файнстейном *). Задача ставилась следующим образом: необходимо с точностью до постоянного множителя найти функцию $H(p_1, p_2, \ldots, p_n)$, значение которой являлось бы мерой неопределенности исхода событий в ансамбле X. При этом

$$p_i = p(x_i). \tag{10.6}$$

К отыскиваемой функции А. Файнстейн предъявляет всего лишь три требования:

1. $H(p, 1-p)$ должна быть непрерывной функцией от p на интервале $0 \leqslant p \leqslant 1$.

2. $H(p_1, p_2, \ldots, p_n)$ должна быть симметричной функцией своих аргументов. Например,

$$H(p_1, p_2) = H(p_2, p_1).$$

3. Если $p_n = q_1 + q_2 > 0$, то

$$H(p_1, p_2, \ldots, p_{n-1}, q_1, q_2) = H(p_1, p_2, \ldots, p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right).$$

Необходимость требований 1 и 2 достаточно очевидна. Требование 1 означает, что при весьма малом изменении распределения вероятностей отдельных событий неопределенность ситуаций также должна изменяться незначительно. Например, для двух ансамблей

$$\begin{pmatrix} x_1 & x_2 \\ 0{,}1 & 0{,}9 \end{pmatrix} \text{ и } \begin{pmatrix} x_1 & x_2 \\ 0{,}101 & 0{,}899 \end{pmatrix}$$

разница в неопределенности исхода событий должна быть малой. Существо требования 2 сводится к тому, что неопределенность исхода событий в ансамбле не должна зависеть от порядка их нумерации.

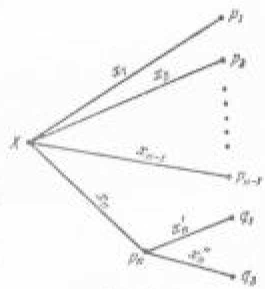

Рис. 10.2.

Требование 3 менее очевидно, оно иллюстрируется диаграммой, показанной на рис. 10.2. В данном случае событие x_n состоит из двух несовместных событий x_n' и x_n''. Вероятность события x_n равна p_n, а вероятности событий x_n' и x_n'' соответственно равны q_1 и q_2. Очевидно, что величины $\frac{q_1}{p_n}$ и $\frac{q_2}{p_n}$ определяют условные вероятности событий x_n' и x_n'', т. е. вероятности исхода x_n' и x_n'' при условии,

*) А. Файнстейн, Основы теории информации, ИЛ, Москва, 1960.

что x_n имеет место. Требование 3 говорит о том, что неопределенность исхода ансамбля всех событий $x_1, x_2, \ldots, x_{n-1}, x_n', x_n''$ должна равняться взвешенной сумме неопределенности исхода событий $x_1, x_2, \ldots, x_{n-1}, x_n$ и неопределенности исхода событий x_n', x_n''.

Веса слагаемых равны вероятности рассматриваемой совокупности событий. Вероятность того, что произойдет одно из событий $x_1, x_2, \ldots, x_n$, равна единице, а неопределенность исхода в этом случае равняется $H(p_1, p_2, \ldots, p_n)$; вероятность того, что произойдет одно из событий x_n' или x_n'', равна $p_n = q_1 + q_2$, а неопределенность исхода, если событие x_n имеет место, равна

$$H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right).$$

Сформулированные три условия однозначно определяют искомую функцию $H(p_1, p_2, \ldots, p_n)$, которая может быть записана в виде

$$H(p_1, p_2, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log_a p_i.$$

Если выбрать $a = 2$ и учесть (10.6), то видно, что последнее равенство совпадает с (10.5).

Вывод соотношения (10.5), определяющего меру общей неопределенности ситуации.

Теорема. *Три требования, сформулированных в предыдущем пункте, определяют функцию $H(p_1, p_2, \ldots, p_n)$ с точностью до постоянного множителя.*

Вывод проведем согласно с А. Файнстейном с небольшими изменениями и пояснениями.

Лемма 1. *Имеет место равенство $H(1, 0) = 0$.*

Доказательство. Используя сформулированное выше требование 3, положим

$$p_1 = \frac{1}{2}, \quad q_1 = \frac{1}{2}, \quad q_2 = 0, \quad p_2 = q_1 + q_2 = \frac{1}{2},$$

тогда

$$H\left(\frac{1}{2}, \frac{1}{2}, 0\right) = H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2} H(1, 0). \tag{1}$$

В силу требований 2 и 3, полагая

$$p_1 = 0, \quad q_1 = \frac{1}{2}, \quad q_2 = \frac{1}{2} \quad \text{и} \quad p_2 = 1,$$

имеем

$$H\left(\frac{1}{2}, \frac{1}{2}, 0\right) = H\left(0, \frac{1}{2}, \frac{1}{2}\right) = H(0, 1) + H\left(\frac{1}{2}, \frac{1}{2}\right). \tag{2}$$

Поскольку $H(1, 0) = H(0, 1)$, то одновременное выполнение (1) и (2) возможно лишь при $H(1, 0) = 0$.

Лемма 2. *Имеет место равенство $H(p_1, \ldots, p_n, 0) = H(p_1, \ldots, p_n)$.*

Доказательство. Полагая, что $p_n > 0$ и рассматривая p_n и 0 как q_1 и q_2, согласно условию 3, имеем

$$H(p_1, \ldots, p_n, 0) = H(p_1, \ldots, p_n) + p_n H(1, 0).$$

Учитывая лемму 1, получаем искомое равенство.

Лемма 3.

$$H(p_1, \ldots, p_{n-1}, q_1, \ldots, q_m) = H(p_1, \ldots, p_n) + p_n H\left(\frac{q_1}{p_n}, \ldots, \frac{q_m}{p_n}\right),$$

где $p_n = q_1 + \ldots + q_m > 0$.

Доказательство. Для $m = 2$ данное равенство совпадает с требованием 3. Применим индукцию по m. Положим, что для некоторого m равенство справедливо. Тогда, дважды используя требование 3, имеем

$$H(p_1, \ldots, p_{n-1}, q_1, \underbrace{q_2, \ldots, q_{m+1}}_{m \text{ членов}}) =$$
$$= H(p_1, \ldots, p_{n-1}, q_1, p') + p' H\left(\frac{q_2}{p'}, \ldots, \frac{q_{m+1}}{p'}\right) =$$
$$= H(p_1, \ldots, p_{n-1}, p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{p'}{p_n}\right) + p' H\left(\frac{q_2}{p'}, \ldots, \frac{q_{m+1}}{p'}\right),$$

где $p' = q_2 + \ldots + q_{m+1}$, $p_n = q_1 + p'$.

Опять используя справедливость утверждения для m, напишем

$$H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}, \ldots, \frac{q_{m+1}}{p_n}\right) = H\left(\frac{q_1}{p_n}, \frac{p'}{p_n}\right) + \frac{p'}{p_n} H\left(\frac{q_2}{p'}, \ldots, \frac{q_{m+1}}{p'}\right).$$

Умножая обе части данного равенства на p_n и подставляя его в предыдущее равенство, получаем

$$H(p_1, \ldots, p_{n-1}, q_1, q_2, \ldots, q_{m+1}) = H(p_1, \ldots, p_{n-1}, p_n) +$$
$$+ p_n \cdot H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}, \ldots, \frac{q_{m+1}}{p_n}\right).$$

Последнее соотношение распространяет утверждение леммы для $m + 1$, что и требовалось доказать.

Лемма 4.

$$H(q_{11}, q_{12}, \ldots, q_{1m_1}; q_{21}, q_{22}, \ldots, q_{2m_2}; \ldots; q_{n1}, q_{n2}, \ldots, q_{nm_n}) =$$
$$= H(p_1, p_2, \ldots, p_n) + \sum_{i=1}^{n} p_i H\left(\frac{q_{i1}}{p_i}, \frac{q_{i2}}{p_i}, \ldots, \frac{q_{im_i}}{p_i}\right),$$

где $p_i = q_{i1} + q_{i2} + \ldots + q_{im_i}$, $i = 1, 2, \ldots, n$.

Доказательство. Используя лемму 3, получаем

$$H(q_{11}, q_{12}, \ldots, q_{1m_1}; \ldots; q_{n1}, q_{n2}, \ldots, q_{nm_n}) =$$
$$= H(q_{11}, q_{12}, \ldots, q_{1m_1}; \ldots; q_{n-1,1}, q_{n-1,2}, \ldots, q_{n-1,m_{n-1}}; p_n) +$$
$$+ p_n H\left(\frac{q_{n1}}{p_n}, \frac{q_{n2}}{p_n}, \ldots, \frac{q_{nm_n}}{p_n}\right).$$

Переставляя переменные таким образом, чтобы p_n оказалось крайним слева, мы продолжим сведение, пока после n шагов не получим желаемого результата.

Введем теперь в рассмотрение функцию

$$F(n)=H\left(\frac{1}{n},\ \ldots,\ \frac{1}{n}\right)$$

при $n \geqslant 2$, ибо $F(1)=H(1)=H(1,\ 0)=0$.

Лемма 5. $F(mn)=F(m)+F(n)$.

Доказательство. В равенстве леммы 4 примем $m_1=m_2=\ldots$ $\ldots=m_n=m,\ q_{ij}=\frac{1}{mn}$; тогда

$$p_i=m\frac{1}{mn}=\frac{1}{n},\quad \frac{q_{ij}}{p_i}=\frac{\frac{1}{mn}}{\frac{1}{n}}=\frac{1}{m}$$

и, следовательно,

$$H\left(\frac{1}{mn},\ \frac{1}{mn},\ \ldots,\ \frac{1}{mn}\right)=H\left(\frac{1}{n},\ \frac{1}{n},\ \ldots,\ \frac{1}{n}\right)+n\frac{1}{n}H\left(\frac{1}{m},\ \frac{1}{m},\ \ldots,\ \frac{1}{m}\right),$$

что и доказывает утверждение данной леммы.

Следствие леммы 5.

$$F(n^m)=mF(n).$$

Лемма 6. $F(n)=\log_a n$, *где a — произвольное число.*

Доказательство. Обозначим

$$F(n)=u \tag{3}$$

и введем в рассмотрение обратную функцию

$$\varphi[F(n)]=n. \tag{4}$$

Тогда из следствия леммы 5 имеем:

$$F(n^m)=mF(n)=mu$$

и, следовательно,

$$\varphi[F(n^m)]=\varphi(mu),$$

но согласно (3) и (4)

$$\varphi[F(n^m)]=n^m=[\varphi(u)]^m.$$

Таким образом, в итоге получим

$$\varphi(mu)=[\varphi(u)]^m. \tag{5}$$

Будем решать функциональное уравнение (5) для функции $\varphi(u)$, положив $u=1$, тогда

$$\varphi(m)=[\varphi(1)]^m.$$

Если обозначить $\varphi(1)=a$, то $\varphi(m)=a^m$ и, следовательно, для целых u также $\varphi(u)=a^u$.

Положим теперь в (5) $u=\frac{1}{m}$; тогда $a=\left[\varphi\left(\frac{1}{m}\right)\right]^m$ или $\varphi\left(\frac{1}{m}\right)=a^{\frac{1}{m}}$.

Далее, полагая в (5) $u=\frac{1}{n}$, имеем

$$\varphi\left(\frac{m}{n}\right)=\left[\varphi\left(\frac{1}{n}\right)\right]^m,$$

но из предыдущего равенства видно, что $\varphi\left(\frac{1}{n}\right)=a^{\frac{1}{n}}$ и, следовательно,

$$\varphi\left(\frac{m}{n}\right)=a^{\frac{m}{n}}.$$

Таким образом, для любого рационального значения u имеем

$$\varphi(u)=a^u.$$

В силу непрерывности функций $\varphi(u)$ для всех значений u справедливо следующее равенство:

$$\varphi(u)=a^u.$$

Учитывая (3), получаем $\varphi[F(n)]=a^{F(n)}$ или $n=a^{F(n)}$, откуда *)

$$F(n)=\log_a n.$$

Теперь мы имеем все необходимое для завершения доказательства теоремы.

Пусть r и s — целые числа, причем $r<s$. Положим в равенстве леммы 4

$$q_{ij}=\frac{1}{s},\quad m_1=r,\quad m_2=s-r,$$

тогда

$$p_1=\frac{r}{s},\quad p_2=\frac{s-r}{s},$$

и данное равенство запишется в виде

$$H\left(\frac{1}{s},\ \ldots,\ \frac{1}{s}\right)=H\left(\frac{r}{s},\ \frac{s-r}{s}\right)+\frac{r}{s}H\left(\frac{1}{r},\ \ldots,\ \frac{1}{r}\right)+$$
$$+\frac{s-r}{s}H\left(\frac{1}{s-r},\ \ldots,\ \frac{1}{s-r}\right).$$

Обозначая $\frac{r}{s}=p$, из последнего соотношения имеем

$$H(p,\ 1-p)=F(s)-pF(r)-(1-p)F(s-r)=$$
$$=\log_a s-p\log_a r-(1-p)\log_a(s-r).$$

Используя очевидное равенство $\log_a s=p\log_a s+(1-p)\log_a s$, получим

$$H(p,\ 1-p)=-p\log_a p-(1-p)\log_a(1-p).$$

Ввиду непрерывности функции $H(p,\ 1-p)$ данный результат распространяется также и на все иррациональные p.

Используя условие 3 для случая $p_2=q_1+q_2$, имеем

$$H(p_1,\ q_1,\ q_2)=H(p_1,\ p_2)+p_2H\left(\frac{q_1}{p_2},\ \frac{q_2}{p_2}\right)=$$
$$=-p_1\log_a p_1-p_2\log_a p_2-q_1\log_a\frac{q_1}{p_2}-q_2\log_a\frac{q_2}{p_2}=$$
$$=-p_1\log_a p_1-q_1\log_a q_1-q_2\log_a q_2.$$

*) Приведенное доказательство леммы 6 выполнено Р. С. Гутер.

Изменяя обозначения, получим

$$H(p_1, p_2, p_3) = -p_1 \log_a p_1 - p_2 \log_a p_2 - p_3 \log_a p_3.$$

Аналогично индукцией по n можно доказать, что

$$H(p_1, p_2, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log_a p_i.$$

Основные свойства энтропии. Сформулируем некоторые основные свойства энтропии как меры неопределенности исхода ансамбля событий.

1) $H = 0$ лишь в том случае, когда все вероятности $p(x_i)$, кроме одной, равны нулю, а эта единственная вероятность равна единице. Следовательно, $H = 0$ только в случае полной определенности исхода опыта, а в остальных случаях $H > 0$.

Последнее вытекает из того, что вероятность события p_i может лежать лишь в интервале между нулем и единицей, т. е. $0 \leqslant p_i \leqslant 1$ и, следовательно, $-p_i \log p_i \geqslant 0$.

2) При заданном n H максимальна и равна

$$H(X) = \log n \tag{10.7}$$

лишь тогда, когда все события равновероятны, т. е.

$$p(x_1) = p(x_2) = \ldots = p(x_n) = \frac{1}{n}.$$

Интуитивно ясно, что такое состояние соответствует наибольшей неопределенности. Данное свойство энтропии при $n = 2$ иллюстрируется рис. 10.1.

В общем случае это свойство энтропии может быть доказано следующим образом.

Обозначим $p(x_i) = p_i$ и представим (10.5) в виде

$$H = -\sum_{i=1}^{n} p_i \log p_i. \tag{1}$$

Кроме того, должно соблюдаться очевидное условие

$$\sum_{i=1}^{n} p_i - 1 = 0. \tag{2}$$

Найдем значения p_i, при которых энтропия H имеет максимальное значение.

Согласно правилу отыскания относительного максимума функции нескольких переменных

$$\frac{\partial}{\partial p_i}\left[H + \lambda\left(\sum_{i=1}^{n} p_i - 1\right)\right] = 0,$$

где λ — множитель Лагранжа.

Подставляя в последнее равенство значение H из (1) и выполняя дифференцирование, получаем

$$-\frac{1}{\ln 2}(\ln p_i + 1) + \lambda = 0 \quad \text{или} \quad p_i = e^{-\lambda_1}, \tag{3}$$

где $\lambda_1 = 1 - \lambda \ln 2$.

Из (2) и (3) следует, что $ne^{-\lambda_1} = 1$, откуда $\lambda_1 = \ln n$. Подставляя это значение λ_1 в (3), получим

$$p_i = \frac{1}{n},$$

что соответствует равной вероятности событий (сообщений).

Нетрудно убедиться в том, что найденное экстремальное значение p_i соответствует максимуму *).

3) Возьмем два ансамбля событий: X, которое описывается схемой (10.3) и Y со схемой

$$Y = \begin{pmatrix} y_1 & y_2 & \cdots & y_m \\ p(y_1) & p(y_2) & \cdots & p(y_m) \end{pmatrix}. \tag{10.8}$$

Будем теперь рассматривать совместные события x_i и y_j. Всевозможные пары (x, y) могут рассматриваться как элементы нового объединенного ансамбля $X \otimes Y$.

Обозначим через $p(x_i, y_j)$ вероятность совместного появления событий x_i и y_j.

Схема объединенного ансамбля может быть записана в виде

$X \otimes Y =$

$y_j \backslash x_i$	x_1	x_2	$\ldots$	x_n
y_1	$p(x_1, y_1)$	$p(x_2, y_1)$	$\ldots$	$p(x_n, y_1)$
y_2	$p(x_1, y_2)$	$p(x_2, y_2)$	$\ldots$	$p(x_n, y_2)$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$
y_m	$p(x_1, y_m)$	$p(x_2, y_m)$	$\ldots$	$p(x_n, y_m)$

Объединенный ансамбль описывает (характеризует) более сложную физическую систему, состоящую из двух подсистем X и Y, или такую систему, в которой состояния или события могут быть разделены на две характерные группы X и Y. В частности, события x могут порождать (быть причиной) событий y, или наоборот.

Объединенный ансамбль $X \otimes Y$ может рассматриваться как некий новый ансамбль, в котором возможны nm различных событий (x, y) с заданным распределением вероятностей $p(x, y)$.

*) Приведенное доказательство выполнено А. А. Красовским.

Энтропия такого ансамбля, т. е. энтропия исхода совместных событий (x, y), согласно (10.5), равна

$$H(X, Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log p(x_i, y_j). \qquad (10.9)$$

Если $H(X)$ определяется (10.5), а

$$H(Y) = -\sum_{j=1}^{m} p(y_j)\log p(y_j), \qquad (10.10)$$

то можно показать, что

$$\boxed{H(X, Y) \leqslant H(X) + H(Y).} \qquad (10.11)$$

При этом знак равенства имеет место тогда и только тогда, когда

$$p(x_i, y_j) = p(x_i)\,p(y_j) \text{ при всех } i, j,$$

т. е. лишь в случае, когда события x и y независимы.

Таким образом, при объединении независимых ансамблей (подсистем) их энтропии складываются.

Для получения (10.11) предварительно докажем, что если

$$\sum_{i=1}^{s} q_i = \sum_{i=1}^{s} p_i = 1, \qquad (1)$$

то

$$-\sum_{i=1}^{s} q_i \log q_i \leqslant -\sum_{i=1}^{s} q_i \log p_i.$$

При этом равенство выполняется лишь в случае, когда $p_i = q_i$.

Для доказательства (2) найдем условия, при которых функция

$$f(p_1, p_2, \ldots, p_s) = -\sum_{i=1}^{s} q_i \log p_i$$

имеет минимальное значение.

Используя способ множителей Лагранжа, получим

$$\frac{\partial}{\partial p_i}\left[f + \lambda\left(\sum_{i=1}^{s} p_i - 1\right)\right] = 0$$

или

$$-\frac{q_i}{p_i \ln 2} + \lambda = 0,$$

откуда

$$p_i = \frac{q_i}{\lambda \ln 2}.$$

Подставляя это значение p_i в (1), находим $\lambda \ln 2 = 1$ и, следовательно, $p_i = q_i$. По обычным правилам можно убедиться в том, что это значение p_i соответствует минимуму f, откуда непосредственно следует (2).

Для доказательства (10.11) учтем, что

$$p(x_i) = \sum_{j=1}^{m} p(x_i, y_j) \quad \text{и} \quad p(y_j) = \sum_{i=1}^{n} p(x_i, y_j),$$

и перепишем равенства (10.5) и (10.10) в форме

$$H(X) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log p(x_i),$$

$$H(Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log p(y_j).$$

Отсюда

$$H(X) + H(Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log p(x_i)\,p(y_j). \qquad (3)$$

Так как

$$\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i)\,p(y_j) = 1,$$

то для правых частей (10.9) и (3) можно использовать соотношение (2) *); тогда

$$-\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log p(x_i, y_j) \leqslant -\sum p(x_i, y_j)\log p(x_i)\,p(y_j),$$

откуда непосредственно следует (10.11).

4. Известно **), что вероятность совместного события равна

$$p(x_i, y_j) = p(x_i)\,p(y_j \mid x_i), \qquad (10.12)$$

где $p(y_j \mid x_i)$ — вероятность события y_j при условии, что произошло событие x_i (условная вероятность y_j). Если учесть также, что

$$\sum_{j=1}^{m} p(x_i)\,p(y_j \mid x_i) = p(x_i),$$

то равенство (10.9) можно представить в виде

$$H(X, Y) = -\sum_{i=1}^{n} p(x_i)\log p(x_i) - \sum_{i=1}^{n} p(x_i)\sum_{j=1}^{m} p(y_j \mid x_i)\log p(y_j \mid x_i)$$

или

$$H(X, Y) = H(X) + H(Y \mid X), \qquad (10.13)$$

где

$$H(Y \mid X) = -\sum_{i=1}^{n} p(x_i)\sum_{j=1}^{m} p(y_j \mid x_i)\log p(y_j \mid x_i). \qquad (10.14)$$

$H(Y \mid X)$ мы будем называть условной энтропией ансамбля Y.

*) Изменив индексацию, двойные суммы можно записать как одинарные, при этом $q_l = p(x_i, y_j)$, $p_l = p(x_i)\,p(y_j)$, $s = mn$.

**) См., например, Б. Р. Левин, Теория случайных процессов и ее применение в радиотехнике, «Советское радио», Москва, 1960.

Из структуры (10.14) следует, что $H(Y|X)$ равно математическому ожиданию величин

$$H(Y|x_i) = -\sum_{j=1} p(y_j|x_i)\log p(y_j|x_i).$$

Последние являются энтропией событий y в случае, если известно, что имеет место событие x_i. Величина $H(Y|x_i)$ может быть названа частной условной энтропией ансамбля Y. Следовательно, условная энтропия $H(Y|X)$ равна среднему значению частных условных энтропий и характеризует среднюю неопределенность исхода событий y при известных событиях x.

Легко видеть, что условная энтропия $H(Y|X)$, так же как энтропия $H(X)$ или $H(Y)$, — величина положительная, т. е. $H(Y|X) \geqslant 0$.

Из (10.13) и (10.11) следует, что

$$H(Y) \geqslant H(Y|X).$$

Равенство выполняется лишь в том случае, когда события x и y независимы.

Последнее соотношение показывает, что неопределенность исхода событий y никогда не может возрастать из-за того, что нам известен исход событий x. Она не изменяется, если события y и x независимы, и уменьшается в противном случае.

Из (10.9) следует, что

$$H(Y, X) = H(X, Y), \tag{10.15}$$

следовательно, соотношение (10.13) может быть еще записано в виде

$$\boxed{H(X, Y) = H(Y) + H(X|Y).} \tag{10.16}$$

Описанные свойства энтропии хорошо согласуются с различными требованиями, которые следует предъявить к количественной мере неопределенности исхода событий. Хотя таковые требования формулируются на основе интуитивных соображений, однако эти соображения являются, по сути, обобщением большого жизненного опыта. Последующее построение количественной теории и сравнение ее выводов с результатами экспериментов позволяют дать окончательное заключение о правомерности и целесообразности принимаемых для вывода основных соотношений.

Определение количества информации при неполной достоверности результатов опыта. Мы уже указывали, что информация о том или ином событии или факте добывается всегда в результате того или иного опыта, причем после опыта ситуация оказывается не всегда полностью определенной, и мы не можем с полной достоверностью утверждать, какое именно событие имело место. Определим количество информации, приходящееся в среднем на один опыт, когда полная достоверность его исхода отсутствует.

Допустим, что интересующие нас события или факты составляют ансамбль X, описываемый (10.3), а результаты опытов (например, получение сообщений), на основе которых мы выносим суждение об исходе событий x, составляют ансамбль Y, описываемый (10.8). Обозначим через $p(x_i|y_j)$ вероятность того, что при известном нам исходе опыта y_j имело место событие x_i. Если, например,

$$p(x_i|y_j) = \begin{cases} 1 & \text{при } j = i, \\ 0 & \text{при } j \neq i, \end{cases} \tag{10.17}$$

то в результате опыта ситуация полностью определена, и мы можем с полной достоверностью утверждать, какое событие x_i имело место. Так как неопределенность исхода событий x до опыта равна $H(X)$, а после опыта неопределенность отсутствует, то в этом случае

$$I(Y, X) = H(X)\ \frac{\text{дв. ед.}}{\text{опыт}}, \tag{10.18}$$

где $I(Y, X)$ — количество информации, содержащееся в среднем в опытах y относительно событий x.

В более общем случае, когда $0 < p(x_i|y_j) < 1$ для различных j, после опыта остается неопределенность, которая в среднем может характеризоваться условной энтропией $H(X|Y)$.

Количество информации, получаемое в среднем на один опыт, теперь может быть найдено как изменение неопределенности в результате опыта, и, следовательно,

$$\boxed{I(Y, X) = H(X) - H(X|Y)\ \frac{\text{дв. ед.}}{\text{опыт}},} \tag{10.19}$$

где, согласно (10.14),

$$H(X|Y) = -\sum_{j=1}^{m} p(y_j)\sum_{i=1}^{n} p(x_i|y_j)\log p(x_i|y_j)\ \frac{\text{дв. ед.}}{\text{опыт}}. \tag{10.20}$$

Легко видеть, что в частном случае, когда выполняется (10.17), условная энтропия $H(X|Y) = 0$ и (10.19) совпадает с (10.18).

Интересно отметить, что формула (10.19) может быть получена также из (10.1), которая для рассматриваемого случая перепишется в виде

$$I(y_j, x_i) = \log\frac{p(x_i|y_j)}{p(x_i)}.$$

При этом $I(y_j, x_i)$ — количество информации, содержащееся в опыте y_j относительно события x_i.

Для усреднения необходимо рассмотреть всевозможные пары (y, x); тогда

$$I(Y, X) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(y_j, x_i)\log\frac{p(x_i|y_j)}{p(x_i)}. \tag{10.21a}$$

Если учесть, что

$$p(y_j, x_i) = p(x_i)\, p(y_j \mid x_i) = p(y_j)\, p(x_i \mid y_j),$$

то после элементарных преобразований получим (10.19). Соотношение (10.19) является фундаментальным и дает ответ на вопрос о количестве информации, содержащемся в среднем в опытах y относительно случайных событий x.

В дальнейшем мы будем иметь дело всегда лишь с количеством информации, приходящимся в среднем на опыт или сообщение, и потому указание об усреднении будем опускать. Подставляя в (10.19) значение $H(X \mid Y)$ из (10.16), получим симметричную форму

$$I(Y, X) = H(Y) + H(X) - H(X, Y). \tag{10.21б}$$

Если учесть также (10.15), то из (10.21б) следует, что

$$I(Y, X) = I(X, Y). \tag{10.22}$$

Последнее соотношение означает, что количество информации, содержащееся в опытах y относительно случайных событий x, равно количеству информации, содержащейся в случайных событиях x относительно исхода опытов y. Соотношения (10.22) и (10.19) позволяют написать еще одну формулу, которая может использоваться для вычисления количества информации

$$I(Y, X) = H(Y) - H(Y \mid X). \tag{10.23}$$

Из (10.19) или (10.23) следует: 1) информация, содержащаяся в опытах y относительно событий x, не превосходит энтропии событий — $H(X)$, т. е.

$$I_{\max}(Y, X) = H(X).$$

Этот максимум достигается лишь при выполнении (10.17), т. е. если исход опыта достоверно определяет событие.

2) $I(X, X) = H(X)$, т. е. энтропия может быть истолкована как информация, содержащаяся в событиях (опытах) относительно самих себя. Из этого также непосредственно вытекает, что энтропия событий есть наибольшее количество информации об этих событиях, которое можно получить из опытов.

3) $I(Y, X) = 0$, если $p(x_i, y_j) = p(x_i)\, p(y_j)$ при всех значениях i и j. Это означает, что информация равна нулю, если исход опытов не зависит от исхода событий. Последний вывод хорошо согласуется с интуитивным представлением о возможности получения информации о событиях на основании некоторых опытов. При вычислении количества информации по формулам (10.19) или (10.23) необходимо

иметь в виду основные соотношения теории вероятностей:

$$\left.\begin{aligned}
&\sum_i p(x_i) = 1,\\
&\sum_j p(y_j) = 1,\\
&\sum_i p(y_j \mid x_i)\, p(x_i) = p(y_j),\\
&\sum_j p(x_i \mid y_j)\, p(y_j) = p(x_i),\\
&\sum_{i,j} p(x_i)\, p(y_j \mid x_i) = \sum_{i,j} p(y_j)\, p(x_i \mid y_j) = 1.
\end{aligned}\right| \tag{10.24}$$

В качестве примера рассмотрим случай, когда

$$p(x_1) = p(x_2) = \frac{1}{2},$$
$$p(y_1 \mid x_1) = p(y_2 \mid x_2) = 0{,}9,$$
$$p(y_2 \mid x_1) = p(y_1 \mid x_2) = 0{,}1.$$

Для вычисления $I(Y, X)$ воспользуемся (10.19), (10.20) и (10.5). Из (10.5) получаем

$$H(X) = -p(x_1)\log p(x_1) - p(x_2)\log p(x_2) = \\ = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1\ \frac{\text{дв. ед.}}{\text{опыт}}.$$

Для вычисления $p(y_1)$ и $p(y_2)$ воспользуемся (10.24). В данном случае

$$p(y_1) = p(y_1 \mid x_1)\, p(x_1) + p(y_1 \mid x_2)\, p(x_2) = 0{,}9\cdot\frac{1}{2} + 0{,}1\cdot\frac{1}{2} = \frac{1}{2},$$
$$p(y_2) = 1 - p(y_1) = 1 - \frac{1}{2} = \frac{1}{2}.$$

Условные вероятности $p(x_i \mid y_j)$ найдем из равенства

$$p(x_i)\, p(y_j \mid x_i) = p(y_j)\, p(x_i \mid y_j)$$

и, следовательно,

$$p(x_i \mid y_j) = \frac{p(x_i)\, p(y_j \mid x_i)}{p(y_j)}.$$

Для нашей задачи имеем

$$p(x_i) = p(y_j) \quad \text{при} \quad i, j = 1, 2$$

и, следовательно,

$$p(x_i \mid y_j) = p(y_j \mid x_i)$$

или

$$p(x_1 \mid y_1) = p(y_1 \mid x_1) = 0{,}9,$$
$$p(x_2 \mid y_1) = p(y_1 \mid x_2) = 0{,}1,$$
$$p(x_2 \mid y_2) = p(y_2 \mid x_2) = 0{,}9,$$
$$p(x_1 \mid y_2) = p(y_2 \mid x_1) = 0{,}1.$$

Из (10.20) получаем

$$H(X|Y) = -p(y_1)[p(x_1|y_1)\log p(x_1|y_1) + p(x_2|y_1)\log p(x_2|y_1)] - p(y_2)[p(x_1|y_2)\log p(x_1|y_2) + p(x_2|y_2)\log p(x_2|y_2)]$$

или

$$H(X|Y) = -\frac{1}{2}(0{,}9\log_2 0{,}9 + 0{,}1\log_2 0{,}1) - \frac{1}{2}(0{,}1\log_2 0{,}1 + 0{,}9\log_2 0{,}9),$$

$$H(X|Y) = 0{,}468\,\frac{\text{дв. ед.}}{\text{опыт}}.$$

В итоге согласно (10.19) имеем

$$I(Y|X) = 1 - 0{,}468 = 0{,}532\,\frac{\text{дв. ед.}}{\text{опыт}}.$$

Если бы в рассматриваемом случае двух событий опыт был бы достоверным, т. е. выполнялись условия

$$p(y_2|x_1) = p(y_1|x_2) = 0,$$

то очевидно, что

$$I(Y|X) = H(X) = 1\,\frac{\text{дв. ед.}}{\text{опыт}}.$$

Таким образом, недостоверность опыта, когда вероятность ошибочного исхода составляет 0,1, приводит к уменьшению количества информации, получаемой из опыта, почти в 2 раза.

Общие замечания о введенной количественной мере информации. Фундаментальные соотношения (10.19), (10.21б) и (10.23) определяют количественную меру информации, содержащейся в одном случайном событии (опыте) — Y относительно другого случайного события — X. При этом мы имели в виду такие события, у которых число возможных исходов (опытов) конечно. События с бесконечно большим числом исходов рассматриваются ниже (см. § 5).

Из структуры указанных формул видно, что в основу измерения количества информации кладутся лишь вероятностные (статистические) характеристики событий. При этом совершенно игнорируется смысловое содержание информации или, как говорят, ее семантика, а также ценность информации для ее получателя, возможные последствия, вызываемые получением данной информации, и другие обстоятельства.

Так, например, отца, получившего сообщение о том, что его жена родила тройню, могут волновать не редкость (малая вероятность) этого события, а такие вопросы, как необходимость расширения

квартиры, дополнительных закупок вещей, и пр. Теория информации все это игнорирует и количество информации, содержащееся в данном сообщении, определяет лишь на основе статистики событий.

Приведем еще один пример. Допустим, что организована лотерея, где вероятности выигрыша различных предметов одинаковы. Тогда сообщения о выигрыше авторучки или автомобиля с точки зрения рассматриваемой теории содержат одинаковое количество информации. Ясно, что для получателя этих сообщений будет не безразлично, какое из них он получит.

Такой подход к определению количественной меры информации в современной теории имеет как сильные, так и слабые стороны.

Достоинствами введенной статистической меры количества информации являются:

1) общность (универсальность) этой меры, ибо она применима к информации любого вида и содержания;

2) объективность и независимость от психологических факторов, ибо статистические показатели событий объективны и устанавливаются на основе эксперимента;

3) для анализа работы и выбора оптимальных характеристик большинства систем передачи и преобразования информации эта мера оказывается достаточной и наиболее рациональной.

В то же время то обстоятельство, что современная теория информации принимает во внимание лишь статистические характеристики рассматриваемых событий, ограничивает область ее применения.

Для случаев, когда должны приниматься во внимание семантика или ценность информации, эта теория неприменима. В последнее время делаются попытки расширения области применения теории информации и введения, например, объективных показателей ценности информации.

Теория информации, разработанная уже в настоящее время, использующая статистический подход к определению количественной меры информации, позволяет анализировать и устанавливать общие законы для громадного числа явлений в технике и живой природе.

Рассмотрим два типа важнейших обменов, происходящих в природе, технике и обществе: энергетический и информационный. Энергетический обмен связан с преобразованием одного вида энергии в другой, с выполнением определенной работы рабочими органами машин и т. п.

Для изучения энергетических обменов необходимо было найти общую меру энергии для различных ее видов, законы преобразования энергии, факторы, определяющие коэффициент полезного действия, и т. п. Все эти общие закономерности были установлены еще в XVIII—XIX веках.

Информационные обмены хотя и связаны с энергетическими процессами, однако последние играют в них вспомогательную роль. Важными для информационных обменов являются количество получаемых или перерабатываемых сведений (информации), их достоверность и потери при различных преобразованиях.

Общие законы для информационных обменов устанавливаются современной теорией информации, что непосредственно указывает на ее большое значение.

§ 2. Информационные характеристики источников дискретных сообщений

Понятие об эргодическом источнике сообщений. Источником сообщений может быть объект, состояние которого определяется некоторым физическим процессом, протекающим во времени или в пространстве по случайному (заранее не известному нам) закону.

Сообщения об этом процессе могут непосредственно восприниматься или фиксироваться (регистрироваться) наблюдателем или автоматической аппаратурой, либо передаваться на расстояние по тому или иному каналу.

Источниками сообщений, для которых характерно изменение их состояния во времени, могут быть: человек, произносящий речь или играющий на музыкальном инструменте, прибор, измеряющий температуру, влажность, давление или какой-либо другой физический параметр, и т. п.

Примерами источников сообщений с пространственным распределением носителя информации являются книги, картины, запись на магнитную пленку и др. Очень часто имеет место сочетание распределений в пространстве и во времени: действие на сцене, в театре, кинокартина и. п.

При передаче и преобразовании информации, как правило, происходит преобразование пространственного распределения во временное, в силу чего в дальнейшем мы будем рассматривать лишь источники сообщений, характеризуемые случайным процессом, развивающимся во времени. Однако все полученные ниже выводы могут быть непосредственно применены и к источникам с пространственным распределением носителя информации. В настоящем параграфе рассматриваются характеристики источников дискретных сообщений.

Такими дискретными сообщениями могут быть: буквы или простые цифры, совокупности букв — слова или фразы, имеющие определенное смысловое содержание, типовые команды или распоряжения, извещения о возможных дискретных состояниях объектов или событиях, результаты измерений, выражаемые числами, и т. д.

Будем полагать, что число различных сообщений конечно, и обозначать их символами $x_1, x_2, \ldots, x_n$. Заметим, что в данном случае различными символами могут обозначаться как элементарные сообщения типа «да», «нет» или «включить», «отключить», цифры 0, 1, ..., так и более сложные, например, стандартные тексты, числа с заданным числом разрядов и т. п.

Важно лишь, чтобы каждое сообщение, независимо от сложности его смыслового содержания, было вполне определено.

Источник дискретных сообщений вырабатывает некоторую последовательность символов x_i, причем порядок следования этих символов случаен и характеризуется некоторой совокупностью вероятностей.

Математическая модель такой системы называется *стохастическим процессом*.

Нас будет интересовать вопрос о том, какое в среднем количество информации создается таким источником на один символ или в единицу времени. Для ответа на этот вопрос необходимо выяснить, какие вероятностные (статистические) показатели могут охарактеризовать рассматриваемый источник. Нетрудно убедиться в том, что одних вероятностей появления отдельных символов, т. е. схемы (10.3), в данном случае для описания источника недостаточно. Допустим, например, что передается последовательность из символов x_1, x_2, x_3 и x_4 с вероятностями $p(x_1)=0,5$; $p(x_2)=0,25$; $p(x_3)=p(x_4)=0,125$. При этом символ x_4 всегда передается после символа x_3. Легко видеть, что хотя вероятности передачи символов x_3 и x_4 одинаковы, однако символ x_4 не несет никакой информации, ибо, получив символ x_3, мы уже достоверно знаем, что следующим будет символ x_4. Следовательно, необходимы более глубокие и детальные вероятностные характеристики источника и, в частности, нужно учитывать зависимость вероятности передачи данного символа x_i от того, какие символы были переданы ранее. Вероятностные или, как их еще называют, корреляционные связи между символами могут распространяться на большие или меньшие группы символов. Допустим, например, что передаваемая последовательность символов есть ..., x_g, x_h, x_i, x_j, ...; если символы независимы, то условная вероятность передачи символа x_j равна

$$p(x_j \mid x_i, x_h, x_g, \ldots) = p(x_j)$$

для всех g, h, i, j; если имеется корреляционная связь только между двумя соседними символами, то

$$p(x_j \mid x_i, x_h, x_g, \ldots) = p(x_j \mid x_i),$$

т. е. вероятность передачи символа x_j зависит лишь от того, каков был предшествующий символ x_i, и не зависит от символов, передававшихся ранее; в случае, когда такая связь имеется между тремя соседними символами,

$$p(x_j \mid x_i, x_h, x_g, \ldots) = p(x_j \mid x_i, x_h).$$

Аналогичные соотношения могут быть написаны при распространении корреляционных связей на группы из большего числа символов.

У большинства источников, встречающихся на практике, корреляционные связи распространяются на конечное число предшествующих символов; такие источники называются *эргодическими*. В эргодическом источнике для символов $x_g, \ldots, x_s$, отстоящих достаточно далеко друг от друга, корреляционная связь отсутствует, т. е. появление x_s не зависит от того, каково было x_g, и, следовательно, $p(x_s \mid x_g) = p(x_s)$ для всех g.

Эргодические последовательности обладают свойствами, аналогичными свойствам эргодических функций (см. гл. VIII). Любая достаточно длинная (с большим числом символов) эргодическая последовательность с вероятностью, как угодно близкой к единице, будет *типичной*. Последнее означает, что частота передачи любого символа (например, x_i)*) в этой последовательности отличается от вероятности передачи этого символа — $p(x_i)$ на сколь угодно малую величину, частота передачи x_j после x_i мало отличается от условной вероятности $p(x_j \mid x_i)$ и. п. Таким образом, достаточно длинная эргодическая последовательность, являющаяся частью всей последовательности символов, вырабатываемой источником сообщения, с вероятностью, близкой к единице, характеризует вероятности отдельных символов и корреляционные связи между ними, присущие источнику. Данное свойство эргодических источников является весьма важным (см. ниже). В качестве примера эргодических сообщений можно привести язык. Почти в любой книге на данном языке (если это только не весьма специализированная литература, как, например, толковый или энциклопедический словарь или труд, насыщенный специальными терминами) частота отдельных букв и сочетаний разных букв одинакова, хотя смысловое содержание книг различно. Это обстоятельство позволяет применить математический аппарат при изучении структуры языков и имеет большое значение для построения систем связи, машин для печати, перевода и др.

Энтропия эргодического источника. Полученное ранее соотношение (10.5) не может быть использовано для вычисления энтропии эргодического источника, поскольку оно справедливо лишь для независимых сообщений и, следовательно, не учитывает корреляционных связей. Учет корреляционных связей значительно упрощается для эргодического источника сообщений. Для такого источника может быть найдено конечное число характерных состояний — $S_1, S_2, \ldots$ таких, что условная вероятность появления очередного символа зависит лишь от того, в каком из этих состояний находится источник. Вырабатывая очередной символ, источник переходит из одного состояния в другое либо возвращается в исходное состояние**).

Рассмотрим частные случаи. Если корреляционные связи в последовательностях, вырабатываемых некоторым источником, отсутствуют, то у источника имеется лишь одно характерное состояние S_1. Вероятность появления символа x_i в момент, когда система находится в этом состоянии, равна $p(x_i)$; выработав символ x_i, источник возвращается в то же состояние S_1. На рис. 10.3 приведена диаграмма перехода для источника, имеющего одно характерное состояние S_1 и вырабатывающего три различных символа — x_1, x_2 и x_3. Состояние источника показано на графике точкой, линии со стрелками характеризуют процесс генерации символов, надписи у линий указывают вероятность этого процесса, когда состояние источника нам известно.

*) Частота x_i в данном случае определяется как отношение числа символов x_i к общему числу символов в рассматриваемой последовательности.

**) Стохастический процесс такого типа в математике называется цепью Маркова.

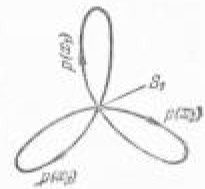

Рис. 10.3.

Когда корреляционные связи имеют место лишь между двумя соседними символами, вероятность появления символа x_i зависит лишь от того, какой символ был выработан до этого. Источник, генерирующий n разных символов — $x_1, x_2, \ldots, x_n$, в этом случае может иметь n характерных состояний: S_1 — после появления символа x_1, S_2 — после появления символа x_2 и т. д. Пример такого источника для случая $n = 3$ приведен на диаграмме рис. 10.4. Для описания такого источника необходимо задать распределение вероятностей $p(x_i)$ и вероятностей переходов $p(x_i \mid x_j)$ для всех i, j. Вместо этого могут быть заданы вероятности всех возможных пар символов — $p(x_i, x_j)$.

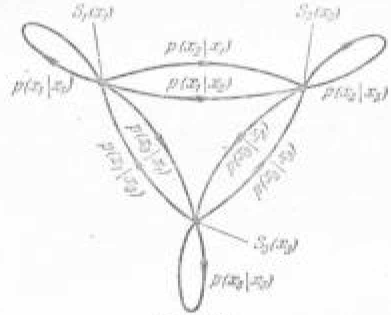

Рис. 10.4.

Если известны $p(x_i, x_j)$, то $p(x_i)$ и $p(x_i \mid x_j)$ могут быть найдены по известным формулам

$$p(x_i) = \sum_j p(x_i, x_j), \qquad p(x_i \mid x_j) = \frac{p(x_i, x_j)}{p(x_j)}.$$

Если коррелятивные связи имеются только между тремя символами, то вероятность появления символа x_l зависит от того, какие два символа были выработаны перед этим, следовательно, число характерных состояний источника определяется числом различных пар x_l, x_j. Для описания такого источника должны быть заданы вероятности появления отдельных символов $p(x_l)$ и вероятности переходов $p(x_l|x_j, x_k)$, либо вероятности всех возможных групп, состоящих из трех символов — $p(x_l, x_j, x_k)$.

Энтропию источника сообщений мы будем вычислять в предположении, что он работает длительное время, и всякий раз, когда мы ждем появления очередного символа, нам известно, какие символы были выработаны ранее, и, следовательно, нам известно, в каком характерном состоянии находится источник. Очевидно, что при этом достаточно помнить лишь те символы, которые коррелятивно связаны с ожидаемым. При таком условии энтропия источника сообщений должна вычисляться уже как условная энтропия по формуле, аналогичной (10.14).

Допустим, что рассматриваемый источник сообщений $x_1, \ldots, x_n$ имеет характерные состояния $S_1, S_2, \ldots, S_n$, при этом $p(S_l|S_k)$ есть вероятность того, что источник, находясь в состоянии S_k, перейдет в состояние S_l при появлении очередного символа; тогда согласно (10.14)

$$H(X) = -\sum_{k} p(S_k) \sum_{l|k} p(S_l|S_k) \log p(S_l|S_k) \; \frac{\text{дв. ед.}}{\text{символ}}, \tag{10.25}$$

где $p(S_k)$ — вероятность состояния S_k. Знаки у сумм означают: k — суммирование по всем возможным состояниям, $l|k$ — по всем возможным переходам из состояния S_k в S_l.

Соотношение (10.25) может быть также записано в виде

$$H(X) = \sum_{k} p(S_k) H(S_k) \; \frac{\text{дв. ед.}}{\text{символ}}, \tag{10.26}$$

где

$$H(S_k) = -\sum_{l|k} p(S_l|S_k) \log p(S_l|S_k) \; \frac{\text{дв. ед.}}{\text{символ}}. \tag{10.27}$$

$H(S_k)$ есть энтропия источника в состоянии S_k и, следовательно, энтропия $H(X)$ есть среднее значение (математическое ожидание) энтропий всех характерных состояний источника.

В случае, когда символы источника независимы, имеется лишь одно состояние S_1, вероятность которого $p(S_1) = 1$. При появлении символа x_l источник вновь возвращается в состояние S_1, и при этом $p(S_1|S_1) = p(x_l)$ и, следовательно,

$$H(X) = H(S_1) = -\sum_{l=1}^{n} p(x_l) \log p(x_l),$$

что полностью совпадает с (10.5).

Если коррелятивные связи имеются лишь между двумя соседними символами, то $p(S_k) = p(x_k)$ и $p(S_l|S_k) = p(x_l|x_k)$. Из (10.25) тогда получим

$$H(X) = -\sum_{k=1}^{n} p(x_k) \sum_{l=1}^{n} p(x_l|x_k) \log p(x_l|x_k) \tag{10.28}$$

или, так как порядок суммирования значения не имеет, то

$$H(X) = -\sum_{k=1}^{n} \sum_{l=1}^{n} p(x_k, x_l) \log p(x_l|x_k) \; \frac{\text{дв. ед.}}{\text{символ}}. \tag{10.29}$$

В случае, когда коррелятивные связи имеются между тремя символами, характерные состояния определяются передачей двух символов, и их удобно нумеровать двумя индексами; так, если генерируются $x_k x_j$, то источник переходит в состояние S_{kj}. Тогда

$$p(S_{kj}) = p(x_k, x_j) \text{ и } p(S_{jl}|S_{kj}) = p(x_l|x_k, x_j).$$

Из (10.25) получаем

$$H(X) = -\sum_{k=1}^{n} \sum_{j=1}^{n} p(x_k, x_j) \sum_{l=1}^{n} p(x_l|x_k, x_j) \log p(x_l|x_k, x_j)$$

или

$$H(X) = -\sum_{k=1}^{n} \sum_{j=1}^{n} \sum_{l=1}^{n} p(x_k, x_j, x_l) \log p(x_l|x_k, x_j) \; \frac{\text{дв. ед.}}{\text{символ}}.$$

Аналогичные соотношения получаются и в случаях, когда коррелятивные связи распространяются на большее число символов.

Рассмотрим численные примеры на определение энтропии источника.

П р и м е р 3. Источник вырабатывает четыре символа: x_1, x_2, x_3, x_4 с вероятностями $p(x_1) = p(x_2) = p(x_3) = p(x_4) = \frac{1}{4}$.

Коррелятивные связи между различными символами отсутствуют. Непосредственно используя (10.7), получим

$$H(X) = 2 \; \frac{\text{дв. ед.}}{\text{символ}}.$$

П р и м е р 4. Вероятности появления символов источника равны

$$p(x_1) = \frac{1}{2}, \quad p(x_2) = \frac{1}{4}, \quad p(x_3) = p(x_4) = \frac{1}{8}.$$

Коррелятивные связи между символами отсутствуют.

Для этого случая, используя (10.5), находим

$$H(X) = \frac{1}{2} \log_2 2 + \frac{1}{4} \log_2 4 + 2 \, \frac{1}{8} \log_2 8$$

или

$$H(X) = \frac{7}{4} = 1{,}75 \; \frac{\text{дв. ед.}}{\text{символ}}.$$

Пример 5. Вероятности появления символов источника такие же, как и в примере 4, но между двумя соседними символами имеют место коррелятивные связи, которые описываются таблицей 10.1. Во 2-м столбце этой таблицы приведены вероятности различных сочетаний из двух символов, что дает полное описание статистических свойств данного источника. Из нее могут быть получены вероятности появления отдельных символов $p(x_i)$ и вероятности переходов $p(x_j|x_i)$, приведенные в 3-м столбце.

Таблица 10.1

x_ix_j	$p(x_i, x_j)$	$p(x_j\|x_i)$	x_ix_j	$p(x_i, x_j)$	$p(x_j\|x_i)$
x_1x_1	13/32	13/16	x_3x_1	0	0
x_1x_2	3/32	3/16	x_3x_2	0	0
x_1x_3	0	0	x_3x_3	0	0
x_1x_4	0	0	x_3x_4	1/8	1
x_2x_1	1/32	1/8	x_4x_1	1/16	1/2
x_2x_2	1/8	1/2	x_4x_2	1/32	1/4
x_2x_3	3/32	3/8	x_4x_3	1/32	1/4
x_2x_4	0	0	x_4x_4	0	0

Из таблицы 10.1 видно, что за символом x_3 в данном источнике всегда следует символ x_4, а после символа x_1 генерируется либо тот же символ x_1, либо символ x_2. Так как вероятности некоторых пар символов равны нулю, то всего в рассматриваемом источнике имеются девять характерных состояний.

Используя (10.28), непосредственно получим

$$H(X) = 0{,}886 \ \frac{\text{дв. ед.}}{\text{символ}},$$

Избыточность источника сообщений. Вспомним теперь, что энтропия характеризует количество информации, приходящееся в среднем на одно сообщение. Рассмотренные примеры 3, 4 и 5 показывают, что при одинаковом количестве различных символов (сообщений) количество информации, приходящееся на одно сообщение, может быть различным в зависимости от статистических характеристик источника. Энтропия источника максимальна и равна $H_{\max} = \log n$, если символы вырабатываются с равными вероятностями; если же это не так и некоторые символы повторяются часто, а другие редко, то энтропия источника уменьшается, а при появлении дополнительных коррелятивных связей между символами энтропия становится еще меньшей. Это положение хорошо согласуется с интуитивным представлением о количестве информации, вырабатываемой тем или иным

источником. Так, например, если из предшествовавшего опыта свойства лектора или докладчика известны настолько хорошо, что слушатели с высокой степенью достоверности знают, о чем он будет говорить, то количество информации, сообщаемой таким лектором, будет очень малым, несмотря на большое число произнесенных слов.

Для того чтобы выяснить, насколько хорошо в источнике сообщений используются разные символы (а источник будет тем лучше, чем больше информации он будет вырабатывать), вводится параметр, называемый *избыточностью* и равный

$$R = \frac{H_{\max} - H(X)}{H_{\max}}. \tag{10.30}$$

При этом $H_{\max} = \log n$ есть максимальная энтропия или наибольшее количество информации, которое может приходиться на один символ источника при данном числе n используемых символов.

Из (10.30) видно, что при $R = 0$ энтропия источника $H(X) = H_{\max}$, т. е. источник генерирует максимальное количество информации на символ. Если $R = 1$, то $H(X) = 0$ и, следовательно, информация, вырабатываемая источником, равна нулю. В общем случае $0 \leqslant R \leqslant 1$. Чем меньше избыточность R, тем рациональнее работает источник, тем большее количество информации он вырабатывает.

Следует, однако, иметь в виду, что не всегда нужно стремиться к тому, чтобы $R = 0$. Некоторая избыточность бывает полезной для обеспечения надежности передачи, регистрации и других преобразований информации. Известно, например, что лектора, который не повторяет или не разъясняет более подробно, на примерах, отдельные положения, слушать и конспектировать значительно труднее, чем лектора, который в разумной мере пользуется этими приемами.

Если не различать буквы «е» и «ё», а также мягкий и твердый знаки, то в русском алфавите всего 31 буква, к ним нужно добавить еще пробел между словами, так что всего получается 32 символа. Если бы все символы были равновероятны, то энтропия такого языка была бы равна

$$H_0 = \log_2 32 = 5 \ \frac{\text{дв. ед.}}{\text{символ}}.$$

В действительности, однако, вероятности различных символов различны; так, например, вероятность буквы «о» равна приблизительно 0,09, а буквы «ф» — 0,002. Кроме того, между символами имеют место значительные коррелятивные связи.

Проведенные исследования *), дают следующие значения энтропии: при учете разной вероятности отдельных символов

$$H_1 = 4{,}39 \ \frac{\text{дв. ед.}}{\text{символ}},$$

*) Д. С. Лебедев, В. А. Гармаш, О возможности увеличения скорости передачи телеграфных сообщений, Электросвязь, 1958, № 1, стр. 68—69.

при учете коррелятивных связей между двумя символами

$$H_2 = 3{,}41\ \frac{\text{дв. ед.}}{\text{символ}},$$

при учете коррелятивных связей между тремя символами

$$H_3 = 3\ \frac{\text{дв. ед.}}{\text{символ}}.$$

Таким образом, можно утверждать, что избыточность русского языка

$$R > \frac{5-3}{5} = 0{,}4.$$

Анализ английского языка *) с учетом коррелятивных связей, распространяющихся на восемь соседних букв, показал, что избыточность его $R > 0{,}5$. Если учесть коррелятивные связи, распространяющиеся на достаточно большое число букв, то можно, по-видимому, убедиться, что избыточность русского и других европейских языков более 50%. Наличие этой избыточности позволяет легко исправлять отдельные ошибки или восстанавливать пропуски букв и даже слогов без искажения текста.

Для примеров, рассмотренных выше, получаем:

в примере 3 $R = 0$,

в примере 4 $R = \frac{2-1{,}75}{2} = 0{,}125$,

в примере 5 $R = \frac{2-0{,}886}{2} \approx 0{,}56$.

Связь между энтропией и числом различных последовательностей сообщений. Рассмотрим последовательности сообщений эргодического источника, содержащие M символов. Кратко будем говорить: последовательности C длины M.

Для таких последовательностей можно доказать следующую теорему, имеющую фундаментальное значение в теории информации.

Теорема 1. Сколь бы ни были малы два числа $\varepsilon > 0$ и $\eta > 0$, при достаточно большом M все последовательности могут быть разбиты на две группы, обладающие следующими свойствами:

1) Вероятность $p(C)$ любой последовательности первой группы (типичные последовательности) удовлетворяет неравенству

$$\left| \frac{\log \frac{1}{p(C)}}{M} - H \right| < \eta, \tag{10.31}$$

где H — энтропия источника, определяемая по (10.25).

*) А. А. Харкевич, Очерки общей теории связи, Гос. изд. техн.-теор. литер., 1955.

2) Сумма вероятностей последовательностей второй группы (нетипичные последовательности) меньше ε.

Доказательство этой теоремы приведено ниже.

Из данной теоремы следует, что для всех типичных последовательностей

$$2^{-M(H+\eta)} < p(C) < 2^{-M(H-\eta)}.$$

Для достаточно длинных последовательностей с весьма малой погрешностью можно принять

$$p(C) \approx 2^{-MH}. \tag{10.32}$$

Последнее означает, что вероятности всех типичных, достаточно длинных последовательностей одинаковы и, если учесть свойство 2), то число таких последовательностей равно

$$N_t \approx \frac{1}{p(C)} \quad \text{или} \quad N_t \approx 2^{HM}. \tag{10.33а}$$

Точнее это соотношение может быть записано в виде

$$2^{(H-\eta)M} < N_t < 2^{(H+\eta)M}, \tag{10.33б}$$

где η может быть как угодно мало, причем если $M \to \infty$, то $\eta \to 0$.

Нетипичные последовательности будут появляться (вырабатываться источником) весьма редко, поэтому с этими последовательностями в ряде случаев можно не считаться.

В случае, если все n символов источника независимы и равновероятны, то энтропия его определяется выражением (10.7). Тогда из (10.33а) имеем

$$N_t = 2^{M \log_2 n} \quad \text{или} \quad N_t = n^M. \tag{10.34}$$

Легко видеть, что (10.34) определяет число всех возможных последовательностей длины M, содержащих n различных символов.

В общем случае, когда вероятности символов не равны и между ними имеют место коррелятивные связи, $H < \log_2 n$ и, следовательно, $N_t < n^M$.

Так как в (10.33а) H стоит в показателе степени, то может быть, что $N_t \ll n^M$. Уменьшение N_t объясняется тем, что многие возможные последовательности оказываются весьма маловероятными и становятся нетипичными.

Доказательство теоремы 1. Последовательность C длины M образуется в результате поочередного перехода источника из одного характерного состояния в другое.

Рассмотрим переход из состояния S_k в состояние S_l по множестве последовательностей C.

Поскольку источник — эргодический, то при достаточно большом M все множество последовательностей может быть разбито на две группы: типичные и нетипичные последовательности. Число M всегда может быть выбрано настолько большим, что сумма вероятностей всех нетипичных последовательностей меньше ε. Учитывая, что для типичной последовательности частота событий может быть как угодно близка к их вероятности, можно утверждать, что в каждой типичной последовательности источник пребывает в состоянии S_k приблизительно $Mp(S_k)$ раз, а число переходов из состояния S_k в состояние S_l приблизительно равно $Mp(S_k)\,p(S_l\,|\,S_k)$.

Точнее можно написать, что число таких переходов равно

$$M\,[p(S_k)\,p(S_l\,|\,S_k)\pm\delta],$$

где δ с увеличением M может быть сделано как угодно малым.

Вероятность того, что в рассматриваемой последовательности имеет место M переходов из состояния S_k в S_l, равна

$$p(S_l\,|\,S_k)^{M\,[p(S_k)\,p(S_l|S_k)\pm\delta]}.$$

Вероятность появления конкретной последовательности C определяется как вероятность всех возможных переходов и, следовательно,

$$p(C)=\prod_{l,\,k}p(S_l\,|\,S_k)^{M\,[p(S_k)\,p(S_l|S_k)\pm\delta]},$$

где индексы l, k у символа произведения $\prod$ показывают, что произведения берутся по всем возможным состояниям S_k и переходам из состояния S_k в S_l.

Из последнего равенства получаем

$$\frac{\log p(C)}{M}=\sum_{l,\,k}[p(S_k)\,p(S_l\,|\,S_k)\pm\delta]\log p(S_l\,|\,S_k)$$

или

$$\frac{\log\dfrac{1}{p(C)}}{M}=-\sum_{k}\sum_{l\,|\,k}p(S_k)\,p(S_l\,|\,S_k)\log p(S_l\,|\,S_k)\pm\delta\sum_{k}\sum_{l\,|\,k}\log p(S_l\,|\,S_k).$$

Если учесть, что первая сумма в правой части последнего равенства совпадает с (10.25), а вторая вследствие произвольной малости δ всегда может быть меньше η, то непосредственно получим (10.31).

Поток информации источника сообщений. Соотношение (10.25) позволяет определить количество информации, переносимое (в среднем) одним символом источника. При работе источника сообщений на его выходе отдельные символы появляются через некоторые интервалы времени; в этом смысле мы можем говорить о *длительности отдельных символов*, и, следовательно, может быть поставлен вопрос о количестве информации, вырабатываемой источником в единицу времени.

Для ответа на этот вопрос вычислим среднюю длительность символа.

Обозначим через

$\tau_{x_i}(S_l\,|\,S_k)$ — длительность символа x_i, переводящего источник из состояния S_k в состояние S_l,

$p_{x_i}(S_l\,|\,S_k)$ — вероятность того, что источник, находясь в состоянии S_k, будет переведен в состояние S_l символом x_i; тогда средняя длительность символа

$$\bar{\tau}_{и}=\sum_{k}p(S_k)\sum_{l\,|\,k}\sum_{i}p_{x_i}(S_l\,|\,S_k)\,\tau_{x_i}(S_l\,|\,S_k). \qquad (10.35)$$

Интересующая нас величина равна

$$\overline{H}(X)=\frac{H(X)}{\bar{\tau}_{и}}\,\frac{\text{дв. ед.}}{\text{сек}}. \qquad (10.36)$$

Энтропия источника, приходящаяся на единицу времени, может быть названа скоростью создания сообщений *) или же *потоком информации*. Двоичная единица информации часто обозначается через *бит* (английское bit, сокращение от «binary digit» — двоичная единица); поток информации в этом случае выражается в $\frac{\text{бит}}{\text{сек}}$.

Очевидно, что поток информации зависит от количества различных символов, вырабатываемых источником, их длительности и статистических характеристик источника.

Если длительности всех символов равны, т. е. $\tau_{x_i}(S_l\,|\,S_k)=\tau$ для всех x_i, S_l и S_k, то, как и следует ожидать, (10.35) дает $\bar{\tau}_{и}=\tau$, и тогда

$$\overline{H}(X)=\frac{1}{\tau}H(X)\,\frac{\text{дв. ед.}}{\text{сек}}. \qquad (10.37)$$

В этом случае поток информации максимальный, если энтропия источника на символ максимальна. Если же символы имеют разную длительность, то такой простой связи уже нет. Для увеличения потока информации необходимо по возможности уменьшить среднюю длительность символов — $\bar{\tau}_{и}$. С этой целью, например, необходимо, чтобы длительность тех символов, вероятность появления которых больше (встречаются чаще), была меньше, чем для символов, вероятность появления которых относительно невелика. Таким образом, для получения большого потока информации на выходе источника необходимо не только обеспечить по возможности большую энтропию на символ, но и правильно выбрать длительность разных символов.

§ 3. Дискретные каналы без шумов

Функции, выполняемые дискретным каналом передачи информации. В соответствии с изложенным ранее (см. гл. II), функциональная схема дискретного канала передачи информации (информационного канала) может быть представлена в виде, показанном на рис. 10.5.

На вход такого канала подаются дискретные сообщения (x). Для передачи на расстояние воздействия любого источника сообщения, как правило, преобразуются в первичные электрические сигналы. Поэтому в дальнейшем будем считать, что сообщения (x) поступают в форме электрических сигналов. Последние с помощью кодирующего устройства преобразуются в кодированные сигналы y.

*) Термин, используемый К. Шэнноном.

Как известно (см. гл. II), для кодирования используется некоторый алфавит элементарных сигналов (символов) — $y_1, y_2, \ldots, y_m$, и существо кодирования сводится к представлению отдельных сообщений (x_i) или последовательностей сообщений (C_l) некоторыми определенными комбинациями символов используемого алфавита.

Кодированные сигналы передаются по каналу связи и на его приемном конце восстанавливаются в точно такой же или в иной форме (z). Декодирующее устройство преобразует кодированные сигналы (z) в сообщения (w) в форме, наиболее приемлемой для выходных (конечных) преобразователей. Сообщения (w) формируются в виде электрических сигналов в одной или нескольких выходных цепях информационного канала.

Рис. 10.5.

Формирование сигналов в различных выходных цепях эквивалентно формированию последовательности разных сигналов в одной цепи. В последующем для простоты мы будем иметь в виду схему с одной выходной цепью. Очевидно также, что с информационной точки зрения физическая реализация (форма или вид модуляции) кодированных сигналов на передающей (y) и на приемной (z) сторонах, а также выходных сигналов (w) значения не имеет. Важным является лишь установление определенного соответствия между y и x, между z и y и, наконец, между w и z.

Не нарушая общности рассуждений, при отсутствии шумов (помех) можно принять $z = y$. Способ кодирования должен быть таким, чтобы в случае отсутствия шумов на приемной стороне по полученным кодированным сигналам можно было однозначно восстановить вид переданных сообщений, т. е. чтобы $w = x$. Это накладывает некоторые ограничения на допустимые комбинации символов кода. Так, например, если мы закодируем символы $x_1 \to 1$, $x_2 \to 0$ и $x_3 \to 01$ *), то при получении сигнала 01 на приемной стороне мы не будем знать, что было передано: сообщение x_3 или два сообщения x_2x_1. Следовательно, необходимо либо отказаться от такого способа кодирования, либо заранее условиться, что комбинацию x_2x_1 передавать нельзя (она запрещена). На возможную последовательность

*) Различные символы сигнала y обозначены разными цифрами.

символов кода могут накладываться и другие ограничения. Так, в случае, когда для разделения разных сообщений передается специальный разделительный символ y_p (например, пауза между импульсами), причем аппаратура построена таким образом, что при передаче подряд двух (y_py_p) или более таких символов она воспринимаются как один, то такие комбинации следует считать запрещенными.

Совокупность всех подобных запретов, обусловленных принятым способом кодирования и построением аппаратуры, мы будем относить к *фиксированным ограничениям*, накладываемым на информационный канал.

Понятие о пропускной способности информационного канала. Обозначим через x_T последовательность сообщений, вырабатываемых источником за время T. Соответствующую ей последовательность принятых сообщений обозначим через w_T. В канале без шумов x_T однозначно определяет w_T, а в канале с шумами (см. § 4) при передаче x_T на приемной стороне могут образовываться различные последовательности w_T.

Пусть $I(W_T, X_T)$ определяет количество информации, содержащееся в последовательностях сообщений w_T на выходе канала о последовательностях x_T на его входе. Очевидно, что $I(W_T, X_T)$ зависит от статистических характеристик источника сообщений и характеристик помех, действующих в канале, а также от интервала времени T.

Рассмотрим предел

$$\lim_{T\to\infty} \frac{I(W_T, X_T)}{T}. \qquad (10.38)$$

При передаче сообщений эргодического источника при $T\to\infty$ с вероятностью, как угодно близкой к единице, последовательность сообщений источника (x_T) будет типичной. При соблюдении некоторых условий (например, эргодический характер помех, действующих в канале) последовательность выходных сообщений (w_T) также будет типичной и потому следует ожидать, что вычисляемый предел может являться некоторой характеристикой работы информационного канала, указывающей среднее количество информации, получаемое на выходе канала за единицу времени, т. е. *скорость передачи информации* *). Обозначим эту скорость через $\bar{I}(W, X)$, тогда

$$\bar{I}(W, X) = \lim_{T\to\infty} \frac{I(W_T, X_T)}{T}. \qquad (10.39)$$

Очевидно, что возможные значения скорости передачи информации $\bar{I}(W, X)$ зависят от свойств (статистических характеристик) источника сообщений, метода кодирования сообщений и свойств канала

*) Эта же величина может называться *потоком информации на выходе канала*.

связи. В частности, при одном и том же способе кодирования длительность символов передаваемых сигналов может быть различной и зависит от вида модуляции и ширины полосы пропускания канала связи. С изменением длительности символов меняется и скорость передачи информации.

Положим, что нам известна некоторая совокупность фиксированных ограничений, накладываемых на информационный канал. К фиксированным ограничениям мы будем относить: параметры канала связи и, в частности, длительность передаваемых символов сигнала, используемый код, методы декодирования сигналов и накладываемые в связи с этим запреты и т. п.

Назовем *пропускной способностью* информационного канала (C) максимальное значение (точнее, верхнюю грань) скорости передачи информации при заданных фиксированных ограничениях, т. е.

$$C = \operatorname*{Sup}_{\substack{\alpha_1 \in B_1 \\ \cdots \\ \alpha_i \in B_i \\ \cdots}} [\bar{I}(W, X)] \frac{\text{дв. ед.}}{\text{сек}}. \tag{10.40}$$

В данном равенстве обозначение Sup указывает, что вычисляется верхняя грань, а запись $\alpha_i \in B_i$ говорит о том, что параметр α_i удовлетворяет заданному фиксированному ограничению, т. е. лежит в некоторой области B_i.

Способ отыскания верхней грани зависит от того, какая совокупность фиксированных ограничений задана.

Если информационный канал определен полностью, то верхняя грань должна отыскиваться по статистическим характеристикам источника сообщений, т. е. отыскиваются распределение вероятностей сообщений (x) и коррелятивные связи между ними, при которых скорость передачи информации будет наибольшей.

Таким образом, *пропускная способность есть характеристика канала* и не зависит от фактической скорости передачи информации от данного источника.

Если совокупность заданных фиксированных ограничений не полностью определяет канал, а имеется возможность изменения тех или иных параметров, например, длительности символов, способа декодирования и пр., то последние выбираются из условия получения наибольшей скорости передачи информации. Во всех случаях отыскиваются оптимальные статистические характеристики источника, при которых скорость передачи информации максимальна.

Аналогично тому как это было сделано для информационного канала, можно определить пропускную способность канала связи

$$C_c = \operatorname*{Sup}_{\substack{\alpha_1 \in B_1 \\ \cdots \\ \alpha_i \in B_i \\ \cdots}} [\bar{I}(Z, Y)] \frac{\text{дв. ед.}}{\text{сек}}, \tag{10.41}$$

где

$$\bar{I}(Z, Y) = \lim_{T \to \infty} \frac{I(Z_T, Y_T)}{T}. \tag{10.42}$$

При этом z_T и y_T — сигналы длительностью T соответственно на выходе и входе канала связи.

Канал связи является частью информационного канала (см. рис. 10.5). Рассматривая канал связи в отдельности и определяя его пропускную способность, мы можем не накладывать никаких фиксированных ограничений на способ кодирования, вид модуляции и структуру сигналов на его входе. Тем самым создаются возможности использования оптимальных сигналов, при которых скорость передачи информации будет предельно большой. Построение реального информационного канала неизбежно связано с введением тех или иных дополнительных фиксированных ограничений, что приводит к недоиспользованию пропускной способности канала связи. Таким образом, обычно

$$C_c > C.$$

Если пропускную способность канала связи определять с учетом всех фиксированных ограничений, накладываемых работой кодирующих и декодирующих устройств, то мы получим *)

$$C = C_c.$$

Это положение следует также из того обстоятельства, что кодирующее и декодирующее устройства в лучшем случае могут быть построены таким образом, чтобы в них не происходила потеря информации. При этом вся информация, которая извлекается из сигнала на выходе канала связи, будет передана к получателю сообщения. Очевидно, что обратного неравенства ($C > C_c$) быть не может.

Если количество фиксируемых ограничений достаточно велико, так что их совокупность полностью определяет реальный канал, то пропускную способность часто называют *реальной*. Если же фиксируются лишь самые общие показатели, как-то: полоса частот передаваемого сигнала, средняя или пиковая мощность передающего устройства, характер и уровень помех, то пропускную способность иногда называют *идеальной* или *предельной*.

Ниже мы увидим, что пропускная способность является важнейшей характеристикой каналов.

Пропускная способность дискретных каналов при отсутствии шумов. Ранее уже указывалось, что в схеме рис. 10.5 при отсутствии шумов, не нарушая общности, можно считать, что $z = y$, а $w = x$ и, следовательно, для сообщений, передаваемых за время T,

*) В этом случае, определяя, например, $I(Z_T, Y_T)$, мы учитываем, каким способом сигнал Z_T будет реально декодироваться.

$Z_T = y_T$ и $w_T = x_T$. Учитывая свойства количественной меры информации, изученные в § 1, получим

$$I(W_T, X_T) = I(X_T, X_T) = H(X_T)$$

и

$$I(Z_T, Y_T) = I(Y_T, Y_T) = H(Y_T).$$

В таком случае из (10.39) и (10.40) получим

$$C = \sup\left\{\lim_{T\to\infty} \frac{H(X_T)}{T}\right\}\ ^{*)}, \tag{10.43}$$

а из (10.41) и (10.42)

$$C_\varepsilon = \sup\left\{\lim_{T\to\infty} \frac{H(Y_T)}{T}\right\}. \tag{10.44}$$

Обозначим через $N(T)$ число всех возможных последовательностей сообщений длительностью T. Из свойств энтропии, рассмотренных в § 1, следует, что $H(X_T)$ будет максимальна, если всевозможные последовательности сообщений равновероятны. Это максимальное значение может быть определено по (10.7) и равно $\log N(T)$. Из (10.43) тогда следует, что для заданного информационного канала

$$C = \lim_{T\to\infty} \frac{\log N(T)}{T}. \tag{10.45}$$

Рассматривая один лишь канал связи, аналогично получим

$$C_\varepsilon = \lim_{T\to\infty} \frac{\log N_\varepsilon(T)}{T}, \tag{10.46}$$

где $N_\varepsilon(T)$ — число всех возможных последовательностей кодированных сигналов длительностью T.

Соотношения (10.45) и (10.46) обычно используют в качестве определения пропускной способности дискретного канала без шумов.

Вычисление пропускной способности дискретных каналов. Найдем пропускную способность некоторых дискретных каналов.

Канал А. Для передачи сообщений используется код с основанием a (т. е. с a различными символами), длительность всех символов кода одинакова и равна τ. Другие фиксированные ограничения отсутствуют.

Для вычисления C_ε рассмотрим последовательность из M символов. Длительность такой последовательности равна $T = M\tau$. При

*) В дальнейшем мы будем опускать указание о том, что верхняя грань ищется при заданных фиксированных ограничениях.

$T\to\infty$ число символов в одной последовательности $M\to\infty$. Очевидно, что всего можно образовать a^M последовательностей длиной в M символов, следовательно, $N_\varepsilon(M\tau) = a^M$ и из (10.46) получим

$$C_\varepsilon = \lim_{M\to\infty} \frac{\log a^M}{M\tau}$$

или

$$C_\varepsilon = \frac{1}{\tau}\log a. \tag{10.47}$$

Нужно отметить, что если рассматривать выход кодирующего устройства как источник сообщений, то согласно (10.7) $\log a$ есть энтропия этого источника, т. е. $\log a = H(Y)$, у которого корреляционные связи между символами отсутствуют и вероятности передачи различных символов одинаковы. Отсюда следует и весьма важное обратное утверждение: *для того чтобы в рассматриваемом канале скорость передачи была максимальной, необходимо, чтобы вероятности передачи различных символов были одинаковыми.*

При использовании двоичного кода $a = 2$ из (10.47) получаем, что

$$C_\varepsilon = \frac{1}{\tau}\ \frac{\text{дв. ед.}}{\text{сек}}. \tag{10.48}$$

Для дискретных каналов принято обозначать $\frac{1}{\tau} = V$, где V называют скоростью передачи, которая выражается в бодах, если τ измеряется в секундах. Таким образом, $C_\varepsilon = V$, т. е. пропускная способность двоичного канала А, выражаемая в двоичных единицах в секунду, равна скорости передачи в бодах.

Канал Б. Символы кода (или отдельные сообщения) имеют одинаковую длительность, равную τ. На допустимую последовательность символов накладываются некоторые фиксированные запреты.

При вычислении C_ε будем рассматривать кодирующее устройство как источник информации и используем результаты § 2.

Рассмотрим сигнал y_T длительностью $T = M\tau$. Энтропия этого сигнала

$$H(Y_T) = MH(Y),$$

где $H(Y)$ — энтропия источника, вычисляемая с учетом наложенных фиксированных запретов.

Очевидно, что скорость передачи информации в таком канале

$$\bar{I} = \frac{H(Y_T)}{T} = \frac{H(Y)}{\tau}, \tag{10.49а}$$

а пропускная способность

$$C_\varepsilon = \frac{H(Y)_{\max}}{\tau}, \tag{10.49б}$$

где $H(Y)_{\max}$ — максимально возможное значение энтропии кодированного сигнала с учетом наложенных запретов.

Канал В. Для передачи сообщений используется код с основанием a, длительности символов кода различны и равны τ_1, τ_2, ..., τ_a. Фиксированные ограничения на допустимую последовательность передачи различных символов не накладываются.

Анализ этого случая *) показывает, что на достаточно большом интервале времени T число возможных последовательностей равно

$$N_c(T) = Ar_1^{\frac{T}{\tau_0}}, \tag{10.50а}$$

где A — некоторая постоянная, τ_0 — произвольный отрезок времени, который обычно удобно выбирать равным минимальной длительности символа.

Значение r_1 находится как наибольший, действительный, положительный корень уравнения

$$1 - r^{-\frac{\tau_1}{\tau_0}} - r^{-\frac{\tau_2}{\tau_0}} - \ldots - r^{-\frac{\tau_a}{\tau_0}} = 0.$$

Подставляя значение для $N_c(T)$ в (10.46) и переходя к пределу, получим

$$C_c = \frac{1}{\tau_0} \log r_1. \tag{10.50б}$$

Можно показать, что для получения максимальной скорости передачи по каналу, у которого длительности символов не равны, вероятности передачи разных символов также должны быть неравными. Так, в частности, для рассматриваемого случая вероятность передачи символа y_j должна быть равной

$$p(y_j) = r_1^{-\frac{\tau_j}{\tau_0}}. \tag{10.51}$$

Коррелятивные связи между символами должны при этом отсутствовать.

В литературе *) рассматриваются также способы вычисления пропускной способности дискретных каналов и в более общем случае, когда на допустимую последовательность передачи символов с различной длительностью накладываются те или иные запреты. Интересно отметить, что и в этом случае формулы (10.50а) и (10.50б) остаются справедливыми, но только величина r_1 должна вычисляться из более общего соотношения.

Введение тех или иных запретов в допустимой последовательности символов, естественно, приводит к появлению коррелятивных связей в сигнале с оптимальной статистической структурой.

Примеры вычисления пропускной способности дискретных каналов и скорости передачи информации. Если кодирующее и декодирующее устройства информационного канала удовлетворяют рассмотренным фиксированным ограничениям для каналов А и Б, то

*) См., например, С. Голдман, Теория информации, ИЛ, Москва, 1957.

его пропускная способность $C = C_c$ и может вычисляться по соответствующим формулам. Следует, однако, иметь в виду, что фактическая скорость передачи информации далеко не всегда должна быть равной пропускной способности канала.

Рассмотрим два простейших примера.

Пример 6. Источник сообщений такой же, как и в примере 3 § 2. В канале типа А используется двоичный код. Сообщения кодируются в соответствии с таблицей 10.2. Длительность символов $\tau = 1,25$ *мсек*.

Таблица 10.2

Сообщение	Код	Сообщение	Код
x_1	00	x_3	10
x_2	01	x_4	11

Согласно (10.48) пропускная способность данного информационного канала равна

$$C = \frac{1}{\tau} = \frac{1}{1,25 \cdot 10^{-3}} = 800 \, \frac{\text{дв. ед.}}{\text{сек}}.$$

Нетрудно убедиться в том, что в этом случае фактическая скорость передачи информации будет равна пропускной способности, т. е. $\bar{I} = C$. Действительно, так как передаваемые сообщения равновероятны, то в длинной последовательности сообщений числа нулей и единиц в сигнале будут одинаковыми. Это видно хотя бы из того, что кодовая группа 00 будет передана столько же раз, что и кодовая группа 11, а кодовая группа 01 — столько же раз, как и группа 10.

Следовательно, вероятности передачи символов 0 и 1 равны, т. е. $p(0) = p(1)$, а ранее уже было найдено, что это является условием обеспечения максимальной скорости передачи информации в канале А.

Пример 7. Источник сообщений тот же, что и в примере 4 § 2. Остальные условия такие же, как и в примере 6.

Так как в данном случае вероятности передачи разных сообщений различны и, в частности,

$$p(x_1) > p(x_2) > p(x_3) = p(x_4),$$

то если время передачи T достаточно велико, число нулей в сигнале будет больше числа единиц, т. е. $p(0) > p(1)$, следовательно, сигнал не будет удовлетворять оптимальным статистическим характеристикам канала А.

Для определения фактической скорости передачи информации используем (10.39) и учтем, что в канале без помех

$$I(W_T,\ X_T) = H(X_T).$$

Рассмотрим последовательность из M сообщений и обозначим $\bar{\tau}_c$ среднюю длительность сигнала, обеспечивающего передачу одного сообщения. Тогда время передачи сообщений будет $T = M\bar{\tau}_c$.

Если энтропия источника сообщений равна $H(X)$, то информация, содержащаяся в M сообщениях, равна

$$H(X_T) = MH(X).$$

Учитывая последние два равенства, из (10.39) получаем

$$\bar{I} = \frac{H(X)}{\bar{\tau}_c}. \qquad (10.52)$$

В данном примере $\bar{\tau}_c = 2\tau$, и, следовательно,

$$\bar{I} = \frac{H(X)}{2\tau}.$$

В примере 4 для рассматриваемого источника было найдено, что $H(X) = \frac{7}{4}\ \frac{\text{дв. ед.}}{\text{сообщ.}}$, тогда $\bar{I} = \frac{7}{8\tau}$ или $= \frac{7}{8\cdot 1{,}25\cdot 10^{-3}} = 700\frac{\text{дв. ед.}}{\text{сек}}$, т. е. $\bar{I} = \frac{7}{8}C = 0{,}875\ C$.

Таким образом, отход от оптимальных статистических характеристик сигналов приводит к тому, что фактическая скорость передачи информации становится меньше пропускной способности канала. Этот вывод естественным образом следует из самого понятия пропускной способности канала.

Эффективное кодирование. Рассмотренные примеры 6 и 7 показывают, что фактическая скорость передачи информации может быть максимальной (равной пропускной способности), если статистические характеристики источника сообщений определенным образом согласованы со свойствами информационного канала. Для каждого источника сообщений это согласование может быть достигнуто специальным выбором способа кодирования и декодирования сообщений. Такое кодирование сообщений, при котором достигается наилучшее использование пропускной способности канала связи (т. е. наибольшая скорость передачи информации), называется *эффективным*. Очевидно, что эффективное кодирование должно обеспечивать:

1) при заданной статистике источника сообщений формирование кодированных сигналов с оптимальными статистическими характеристиками, при которых достигается наибольшая скорость передачи информации;

2) возможность декодирования сигналов на приемной стороне, т. е. разделения сигналов отдельных сообщений или последовательностей сообщений, опознание этих сигналов и восстановление переданных сообщений.

Способы эффективного кодирования зависят от вида и свойств информационного канала. В настоящее время разработано большое количество различных способов эффективного кодирования *), однако практическое применение пока находят лишь немногие из них. Рассмотрим примеры эффективного кодирования для двоичного канала А.

Выше мы уже видели, что в канале А скорость передачи информации максимальна, если вероятности передачи различных символов равны. Следовательно, в двоичном канале эффективный код должен быть таким, чтобы при заданной статистике источника сообщений вероятности передачи символов 0 и 1 в канале связи были бы равны. Одним из способов, который позволяет приблизиться к этому требованию, является использование кода Шэннона — Фано.

Построение кода Шэннона — Фано для источника сообщений примера 4 § 2 иллюстрируется таблицей 10.3.

Таблица 10.3

Сообщение	Вероятность	Номер деления на подгруппы				Символы кода: позиции				Длительность сигнала
		1	2	3		1	2	3		
x_1	0,5	I				0				τ
x_2	0,25	II	I			1	0			2τ
x_3	0,125	II	II	I		1	1	0		3τ
x_4	0,125	II	II	II		1	1	1		3τ

Для составления такой таблицы поступают следующим образом: все сообщения выписываются в порядке убывания вероятности их передачи. Затем производится ряд последовательных делений сообщений на подгруппы I и II. Всякий раз деление производится из условия, чтобы суммы вероятностей передачи сообщений подгрупп I и II были бы по возможности равными. В данном примере это условие удается выполнить точно; так, при первом делении в подгруппу I входит сообщение x_1, вероятность его передачи 0,5; в подгруппу II входят сообщения x_2, x_3 и x_4, сумма вероятностей их передачи тоже 0,5. При втором делении в подгруппу I вошло сообщение x_2 с вероятностью передачи 0,25, а в подгруппу II — сообщения x_3 и x_4,

*) См., например, А. А. Харкевич, Очерки общей теории связи, Гостехиздат, Москва, 1955.

сумма вероятностей передачи которых тоже 0,25. Деления на подгруппы продолжают до тех пор, пока в каждой подгруппе не окажется по одному сообщению.

Номер подгруппы, в которую попадает данное сообщение при каждом делении, определяет символ на соответствующей позиции записи кода этого сообщения. В рассматриваемой таблице принадлежность к подгруппе I обозначается символом 0, а к подгруппе II — символом 1. Так, в частности, первое деление дало на первой позиции записи кода для сообщения x_1 символ 0, и для остальных сообщений — символ 1.

Как видно, полученный код является неравномерным, т. е. сигналы разных сообщений могут иметь различное число символов, а следовательно, и разную длительность.

Нетрудно убедиться, что построенный таким образом код удовлетворяет указанным ранее требованиям к эффективному коду. Действительно, любой символ в передаваемом сигнале будет принадлежать некоторой позиции одного из сигналов, однако из таблицы 10.3 легко видеть, что для любой позиции всей совокупности кодированных сигналов вероятности передачи 0 и 1 одинаковы, следовательно, и в передаваемом по каналу связи сигнале вероятности передачи символов 0 и 1 одинаковы, т. е. выполняется первое требование. Построенные кодовые группы имеют ту особенность, что начальная часть сигнала каждого сообщения не воспроизводит сигнала какого-либо другого сообщения. Следовательно, при поочередной передаче различных сообщений их сигналы могут быть разделены на приемной стороне без введения каких-либо дополнительных разделительных сигналов или специальных символов. В этом легко убедиться, выписав любую последовательность кодированных сигналов. Таким образом, выполняется и второе требование.

Средняя длительность сигнала, затрачиваемая на передачу одного сообщения, в рассматриваемом примере равна

$$\bar{\tau}_c = \tau \cdot 0,5 + 2\tau \cdot 0,25 + 3\tau \cdot 0,125 + 3\tau \cdot 0,125 = 1,75\tau.$$

Используя (10.52), определим скорость передачи информации

$$\bar{I} = \frac{H(X)}{\bar{\tau}_c} = \frac{H(X)}{1,75\tau};$$

учитывая, что для используемого источника сообщений

$$H(X) = 1,75 \ \frac{\text{дв. ед.}}{\text{сообщ.}} \quad \text{(см. пример 4)},$$

получаем

$$\bar{I} = \frac{1}{\tau} = C,$$

т. е. в данном случае применение кода Шэннона — Фано позволило полностью согласовать статистические характеристики источника со

свойствами канала. Конечно, это объясняется тем, что значения вероятностей передачи различных сообщений в примере выбраны такими*), что условия деления на подгруппы всегда удается выполнить точно. Проследим еще, что дают различные способы кодирования, на следующем примере.

Пример 8. Источник сообщений описывается схемой

$$X = \begin{pmatrix} x_1 & x_2 & x_3 \\ 0,2 & 0,7 & 0,1 \end{pmatrix}.$$

Коррелятивные связи между сообщениями отсутствуют. Передача информации производится по двоичному каналу А. Длительность символов кодированного сигнала равна τ. Определим скорость передачи информации при использовании равномерного кода и кода Шэннона — Фано.

Используя (10.5), найдем

$$H(X) = -(0,2\log_2 0,2 + 0,7\log_2 0,7 + 0,1\log_2 0,1)$$

или

$$H(X) = 1,16 \ \frac{\text{дв. ед.}}{\text{сек}}.$$

Согласно (10.52) скорость передачи информации равна

$$\bar{I} = \frac{1,16}{\bar{\tau}_c} \ \frac{\text{дв. ед.}}{\text{сек}}.$$

При равномерном коде для передачи каждого сообщения придется израсходовать два символа — таблица 10.4.

Таблица 10.4.

Сообщение	Код
x_1	00
x_2	01
x_3	10

Очевидно, что в этом случае $\bar{\tau}_c = 2\tau$ и, следовательно,

$$\bar{I} = \frac{1,16}{2\tau} = 0,58\,C \ \frac{\text{дв. ед.}}{\text{сек}},$$

где $C = \frac{1}{\tau}$ — пропускная способность данного канала.

*) Пример заимствован у Шэннона.

Для построения кода Шэннона — Фано составлена таблица 10.5.

Таблица 10.5

Сообщения	Вероятность	Номер деления на подгруппы			Символы кода			Длительность сигнала
		1	2		1	2		
x_2	0,7	I			0			τ
x_1	0,2	II	I		1	0		2τ
x_3	0,1		II		1	1		2τ

Средняя длительность сигнала при этом будет

$$\overline{\tau}_c = \tau\cdot 0,7 + 2\tau\cdot 0,2 + 2\tau\cdot 0,1 = 1,3\tau.$$

Скорость передачи информации в этом случае равна

$$\overline{I} = \frac{1,16}{1,3\tau} = 0,89C.$$

Как видим, переход к более эффективному коду позволяет увеличить скорость передачи информации в $\frac{0,89}{0,58} = 1,53$ раза.

Эффективность кодирования может быть еще увеличена, если от кодирования одиночных сообщений перейти к кодированию последовательностей (групп) сообщений.

Таблица 10.6

Сообщение	Вероятность	Деление на подгруппы						Код	Длительность сигнала
x_2x_2	0,49	I						0	τ
x_1x_2	0,14		I	I				1 0 0	3τ
x_2x_1	0,14			II				1 0 1	3τ
x_2x_3	0,07			I	I			1 1 0 0	4τ
x_3x_2	0,07	II			II			1 1 0 1	4τ
x_1x_1	0,04		II		I			1 1 1 0	4τ
x_1x_3	0,02			II		I		1 1 1 1 0	5τ
x_3x_1	0,02				II	II	I	1 1 1 1 1 0	6τ
x_3x_3	0,01						II	1 1 1 1 1 1	6τ

В таблице 10.6 дано построение кода Шэннона — Фано для случая передачи всех возможных пар сообщений.

Средняя длительность сигнала, обеспечивающего передачу одного сообщения, в этом случае будет

$$\overline{\tau}_c = \frac{1}{2}(\tau\cdot 0,49 + 3\tau\cdot 0,28 + 4\tau\cdot 0,18 + 5\tau\cdot 0,02 + 6\tau\cdot 0,03) = 1,165\tau.$$

Соответствующая скорость передачи сообщений равна

$$\overline{I} = \frac{1,16}{1,165\tau} = 0,995C.$$

Если кодировать последовательности по три и более сообщений, то скорость передачи информации может еще больше приблизиться к пропускной способности.

Кодирование последовательностей сообщений особенно эффективно при наличии коррелятивных связей между ними. Легко видеть, что кодирование отдельных сообщений кодом Шэннона — Фано коррелятивных связей не учитывает. При кодировании же последовательностей сообщений вероятности этих последовательностей будут вычисляться с учетом коррелятивных связей. Кроме того, с удлинением кодируемых сообщений коррелятивные связи между ними ослабевают.

Заметим еще, что описанный метод построения эффективного кода Шэннона — Фано может быть распространен и на коды с основанием больше двух.

Основная теорема Шэннона для дискретного канала без шумов. Основная теорема Шэннона для дискретного канала без шумов дает ответ на вопрос о том, в какой мере скорость передачи информации может быть приближена к пропускной способности информационного канала. Она может быть сформулирована в следующем виде.

Теорема 2. Если поток информации, вырабатываемой источником, равен

$$\overline{H}(X) = C_c - \varepsilon, \tag{10.53}$$

где ε может быть как угодно малым, то всегда можно найти такой способ кодирования, который обеспечит передачу всех сообщений, вырабатываемых источником, причем скорость передачи информации будет равна

$$\overline{I} = C_c - \varepsilon.$$

Обратное утверждение заключается в том, что невозможно обеспечить длительную передачу всех сообщений источника, у которого

$$\overline{H}(X) > C_c.$$

Доказательство этой теоремы приводится ниже, здесь же мы приведем лишь некоторые рассуждения, которые позволят лучше уяснить ее суть.

Если $\bar{\tau}_н$ — средняя длительность одного сообщения, то при достаточно большом T возможна передача $M = \frac{T}{\bar{\tau}_н}$ различных сообщений.

Из (10.33а) следует, что число типичных последовательностей x_T сообщений длительностью T равно

$$N_{x_T} \approx 2^{\frac{T}{\bar{\tau}_н} H} = 2^{\bar{H}T}.$$

В то же время на основании (10.46) можно утверждать, что число различных последовательностей кодированных сигналов y_T длительностью T при большом T равно $N_c(T) \approx 2^{C_c T}$ или, учитывая (10.53),

$$N_c(T) \approx 2^{(\bar{H}+\alpha)T} > N_{x_T}.$$

Последнее означает, что эти последовательности сигналов обеспечивают кодирование всех типичных последовательностей сообщений при скорости передачи информации, близкой к C_c, ибо для каждой типичной последовательности сообщений x_T может быть выбран некоторый сигнал y_T и остается еще небольшой резерв сигналов длительностью T. Что касается нетипичных последовательностей сообщений, то суммарная вероятность их очень мала, и на скорость передачи информации они влияния не оказывают. Эти последовательности могут кодироваться сигналами с большей длительностью (с большим числом символов).

Нетрудно убедиться в справедливости обратного утверждения теоремы. Если $\bar{H}(X) > C$, то число различных последовательностей сигналов оказывается недостаточным для кодирования типичных последовательностей сообщений.

Существенно обратить внимание на то обстоятельство, что для приближения скорости передачи информации к пропускной способности, в общем случае*) требуется кодировать последовательность сообщений с большой длительностью — T. Кодирование длинных последовательностей сообщений вызывает:

значительное усложнение кодирующих и декодирующих устройств,

задержку во времени передачи. Величина этой задержки может достигать $2T$.

В заключение приведем другие математические формулировки основной теоремы Шэннона.

Учитывая, что

$$\bar{H}(X) = \frac{H(X)}{\bar{\tau}_н} = V_н H(X),$$

*) В частных случаях, как мы видели на примерах, скорость передачи сообщений может достигать или быть близкой к пропускной способности канала при кодировании отдельных сообщений или коротких последовательностей.

где $V_н = \frac{1}{\bar{\tau}_н}$ — среднее количество сообщений, передаваемых в единицу времени, представим (10.53) в виде

$$V_н = \frac{C_c}{H(X)} - \varepsilon_н, \qquad (10.54)$$

где $\varepsilon_н = \frac{\varepsilon}{H(X)}$ может быть сколь угодно малым.

При использовании канала A

$$\bar{\tau}_н = \bar{m}_c \tau, \qquad (10.55)$$

где $\bar{m}_c$ — среднее количество символов в кодированном сигнале, расходуемое на передачу одного сообщения, τ — как и ранее, длительность символа в сигнале. Из (10.55), (10.47) и (10.53) для этого типа каналов получаем

$$\frac{H(X)}{\bar{m}_c \tau} = \frac{1}{\tau} \log a - \varepsilon$$

или *)

$$\bar{m}_c = \frac{H(X)}{\log a} + \varepsilon_m, \qquad (10.56)$$

где ε_m может быть как угодно малой величиной.

Для двоичных кодов $a = 2$, и тогда

$$\bar{m}_{c,2} = H(X) + \varepsilon_m,$$

т. е. при двоичном кодировании среднее число символов в передаваемом сигнале, приходящееся на одно сообщение, может принимать значения, как угодно близкие к энтропии источника сообщений.

Доказательство теоремы 2. Справедливость прямого утверждения теоремы 2 докажем в формулировке (10.54).

Для краткости энтропию источника сообщений обозначим через

$$H = H(X). \qquad (1)$$

Из (10.46) следует, что если T достаточно велико, то

$$N_c(T) > 2^{(C_c - \theta)T}, \qquad (2)$$

где θ может быть как угодно мало.

Разобьем все возможные последовательности из M сообщений на две группы

$$N_x = N_{xI} + N_{xII}, \qquad (3)$$

где N_{xI} — типичные последовательности, N_{xII} — нетипичные последовательности.

Из (10.33б) видно, что если M велико, то

$$N_{xI} < 2^{(H+\eta)M}, \qquad (4)$$

где η может быть как угодно малой величиной.

Если источник может вырабатывать n различных сообщений, то из них может быть образовано всего n^M различных последовательностей из M сооб-

*) Строгое доказательство данного соотношения дано у А. Файнстейна.

щений. Следовательно,

$$N_{x11} < n^M = 2^{MH_m}, \tag{4a}$$

где $H_m = \log_2 n$. Выберем отрезок времени

$$T_1 = \left(\frac{H}{C_c} + \lambda\right) M \tag{5}$$

и параметр

$$\lambda = \frac{\eta + \theta \dfrac{H}{C_c}}{C_c - \theta}. \tag{6}$$

Из (6) ясно, что если η и θ малы, то λ также мало. При таком выборе λ $(C_c - \theta)\, T_1 = (H + \eta)\, M$, и, следовательно, согласно (2) и (4),

$$N_{c1} > 2^{(C_c - \theta) T_1} = 2^{(H + \eta) M} > N_{x1}. \tag{7}$$

Последнее неравенство означает, что число различных последовательностей сигналов (N_{c1}) длительностью T_1 таково, что ими можно закодировать все типичные последовательности сообщений (N_{x1}) и останется еще некоторое количество последовательностей сигналов, которые могут быть, например, использованы для фиксирования начала и конца передачи нетипичных последовательностей сообщений.

Для передачи нетипичных последовательностей сообщений (N_{x11}) возьмем отрезок времени, равный

$$T_{11} = \left(\frac{H_m}{C_c} + \varphi\right) M. \tag{8}$$

Выберем

$$\varphi = \frac{\theta \dfrac{H_m}{C_c}}{C_c - \theta}, \tag{9}$$

тогда $(C_c - \theta)\, T_{11} = MH_m$; из (9) видно, что если θ мало, то φ также мало.

Из (2), (4а) и последнего равенства следует, что

$$N_{c11} > 2^{(C_c - \theta) T_{11}} = 2^{MH_m} > N_{x11}. \tag{10}$$

Данное неравенство свидетельствует, что число различных последовательностей сигналов длительностью T_{11} позволяет с избытком закодировать все нетипичные последовательности сообщений.

Так как $H_m > H$, то, очевидно, и $T_{11} > T_1$, т. е. полагаем, что нетипичные последовательности сообщений кодируются сигналами с большей длительностью (с большим числом символов).

Из теоремы 1 (см. стр. 776) известно, что если M достаточно велико, то суммарная вероятность передачи всех нетипичных последовательностей сообщений может быть как угодно малой. Примем, что она меньше δ; тогда очевидно, что среднее время передачи последовательности из M сообщений удовлетворяет неравенству $\overline{T} < (1 - \delta)\, T_1 + \delta T_{11}$. Подставляя в это неравенство значения T_1 и T_{11} из (5) и (8), получим

$$V_n = \frac{M}{\overline{T}} > \frac{1}{(1 - \delta)\left(\dfrac{H}{C_c} + \lambda\right) + \delta\left(\dfrac{H_m}{C_c} + \varphi\right)}.$$

Если учесть, что при $M \to \infty$ величины δ, λ и φ стремятся к нулю, то из последнего неравенства следует (10.54).

При этом ε_n будет малой величиной, стремящейся к нулю, когда δ, λ и φ стремятся к нулю.

§ 4. Дискретные каналы с шумами

Особенности работы дискретных каналов с шумами. Воздействие различного рода помех и собственных шумов (в общем случае будем говорить о шумах) на всякий реальный канал связи приводит к искажению передаваемых сигналов, в результате чего, получив сигнал на приемной стороне, мы не можем с полной достоверностью утверждать, какое сообщение было передано, ибо искажение сигнала шумами может привести к тому, что при передаче сообщения x_i на приемной стороне будет зарегистрировано некоторое другое сообщение x_j, где j не совпадает с i.

Если нам известны априорные статистические характеристики источника сообщений, а также характеристики помех, воздействующих на канал, то, получив сообщение, мы можем лишь узнать новое апостериорное распределение вероятности передачи различных сообщений (см. гл. XI), с помощью которого можно вычислить вероятность правильного решения о том, какое из возможных сообщений было передано. Таким образом, работа информационного канала с математической точки зрения, в конечном счете, сводится к изменению на приемной стороне распределения вероятностей передачи различных сообщений.

Ясно, что на практике мы не можем довольствоваться лишь вычислением апостериорного распределения, а аппаратура всегда строится таким образом, что в ней с тем или иным риском (вероятностью ошибки) принимается решение о том, какое сообщение было передано.

В схеме рис. 10.5 в случае воздействия помех нет однозначного соответствия между сигналами на входе канала связи (y) и сигналами на его выходе (z). При передаче сигнала y_i выходные сигналы (z) могут принимать различные значения в зависимости от того, каков был шум в момент приема. В общем случае сигнал z может принимать непрерывное (бесконечно большое число) множество различных значений. Принятие решения соответствует некоторому разделению всего множества (z) на области $Z_1, Z_2, \ldots, Z_m$, так что если принятый сигнал z принадлежит области Z_j, т. е. $z \in Z_j$, то принимается решение о том, что был передан сигнал y_j.

В основу работы решающей схемы могут быть положены те или иные критерии (см. гл. XI); так, в частности, при использовании критерия В. А. Котельникова *) минимизируется среднее значение вероятности принятия ошибочного решения. Применение этого критерия на практике часто затрудняется тем, что он требует знания априорного распределения вероятности передачи различных сообщений. Кроме того, в системах, оптимальных по этому критерию,

*) В литературе используется также название «критерий идеального наблюдателя».

может оказаться, что маловероятные сообщения в основном принимаются неправильно, так как это почти не влияет на среднюю вероятность ошибки.

Основные соотношения для дискретного канала с шумами в теории информации выведены для решающих схем, в которых равномерно ограничена вероятность ошибки при опознании любого переданного сигнала y_j. В этом случае для вероятности правильного принятия решения $p(z_j|y_j)$, т. е. условной вероятности того, что при передаче сигнала y_j будет принято решение z_j, справедливо соотношение

$$p(z_j|y_j) \geqslant q_p.$$

При этом q_p — вероятность правильного решения, гарантируемая для всех сообщений, т. е. для всех j.

Пропускная способность. Соотношение (10.41), определяющее пропускную способность дискретного канала связи, является общим, оно справедливо и для дискретных каналов с шумами. Разница между такими каналами и рассмотренными выше каналами без шумов состоит лишь в способе вычисления количества информации, содержащейся в последовательности выходных сигналов Z_T, о входных сигналах Y_T, т. е. $I(Z_T, Y_T)$.

Для вычисления $I(Z_T, Y_T)$ можно использовать (10.19) или (10.23). Из этих соотношений получаем

$$I(Z_T, Y_T) = H(Z_T) - H(Z_T|Y_T) = H(Y_T) - H(Y_T|Z_T). \tag{10.57}$$

В дальнейшем мы будем полагать, что шумы, действующие в канале связи, имеют эргодический характер. Это значит, что, например, при длительной многократной передаче сигнала y_j сигналы z на выходе канала с вероятностью, как угодно близкой к единице, образуют типичную последовательность. То же самое справедливо и при передаче эргодической последовательности различных сигналов y.

При таком условии выход канала связи может рассматриваться как эргодический источник, и к нему применены результаты, полученные в § 2.

Для последовательности длительностью T, содержащей M сигналов такого источника, имеем

$$H(Z_T) = MH(Z), \tag{10.58}$$

где $H(Z)$ — энтропия выходного сигнала или, точнее, энтропия выхода канала связи, рассматриваемого как эргодический источник.

Величина $H(Z)$ может быть подсчитана по формуле, аналогичной (10.25),

$$H(Z) = -\sum_k \sum_{l|k} p(Q_k)\, p(Q_l|Q_k) \log p(Q_l|Q_k). \tag{10.59}$$

При этом Q_l и Q_k обозначены характерные состояния выхода канала связи.

Такое же соотношение получим и для вычисления условной энтропии

$$H(Z_T|Y_T) = MH(Z|Y), \tag{10.60}$$

где $H(Z|Y)$ — энтропия выходного сигнала канала связи при известных входных сигналах.

Повторяя рассуждения, приведенные при выводе (10.25), получим

$$H(Z|Y) = \sum_j p(y_j)\, H(Z|y_j), \tag{10.61}$$

где

$$H(Z|y_j) = -\sum_k \sum_{l|k} p(Q_k)\, p(Q_l|Q_k, y_j) \log p(Q_l|Q_k, y_j). \tag{10.62}$$

При этом $p(Q_l|Q_k, y_j)$ — условная вероятность перехода выхода канала связи из состояния Q_k в состояние Q_l при передаче сигнала y_j.

Из (10.57), (10.58) и (10.60) следует, что

$$I(Z_T, Y_T) = MH(Z) - MH(Z|Y).$$

При определении скорости передачи информации по (10.42) учтем, что $\lim\limits_{T\to\infty}\left(\frac{T}{M}\right) = \bar{\tau}_c$; при этом, как и ранее, $\bar{\tau}_c$ — средняя длительность сигнала одного сообщения. Тогда получим

$$\bar{I}(Z, Y) = \bar{H}(Z) - \bar{H}(Z|Y), \tag{10.63}$$

где

$$\left.\begin{aligned} \bar{H}(Z) &= \frac{H(Z)}{\bar{\tau}_c} \\ \text{и} \quad \bar{H}(Z|Y) &= \frac{H(Z|Y)}{\bar{\tau}_c}. \end{aligned}\right\} \tag{10.64}$$

Повторяя рассуждения, аналогично найдем

$$\bar{I}(Z, Y) = \bar{H}(Y) - \bar{H}(Y|Z). \tag{10.65}$$

В последнем равенстве $\bar{H}(Y)$ — поток информации на выходе кодирующего устройства, $\bar{H}(Y|Z)$ характеризует потерю информации, обусловленную действием помех.

Из найденных соотношений и (10.41) следует, что пропускная способность канала связи при наличии помех может быть определена из условия

$$C_c = \operatorname{Sup}\{\bar{H}(Z) - \bar{H}(Z|Y)\} \tag{10.66}$$

или

$$C_c = \operatorname{Sup}\{\bar{H}(Y) - \bar{H}(Y|Z)\}. \tag{10.67}$$

Оба определения равноправны и дают одно и то же значение C_ε. Использование того или иного определения диктуется удобством анализа. При отыскании оптимальных статистических характеристик передаваемых сигналов (y) необходимо иметь в виду следующее:

Характерные состояния выхода канала связи (Q_0, Q_1) могут определяться двумя обстоятельствами:

а) наличием фиксированных ограничений, т. е. запретов, накладываемых на допустимую последовательность передачи различных сигналов, и

б) корреляционными связями между символами, вызываемыми действием шумов.

Каналы, у которых на каждый передаваемый сигнал (символ) шум воздействует независимо от того, какие сигналы передавались ранее, называются каналами без памяти. В этих каналах шумы не вызывают дополнительных корреляционных связей между сигналами. В настоящее время основные выводы теории информации получены применительно к каналам без памяти.

Выход канала связи без памяти при отсутствии фиксированных ограничений на допустимый порядок передачи различных сигналов может иметь одно характерное состояние, и тогда энтропии $H(Z)$ и $H(Y)$ могут определяться по формуле (10.5), а условные энтропии $H(Z|Y)$ и $H(Y|Z)$ — по формуле (10.14).

Примеры определения пропускной способности дискретных каналов с шумами. В качестве примера рассмотрим двоичный канал связи без памяти, по которому передаются два различных сигнала (символа) y_0 и y_1. Какие-либо запреты (фиксированные ограничения) на допустимую последовательность передачи сигналов отсутствуют. Длительности сигналов y_0 и y_1 одинаковы и равны τ (канал А) и, следовательно, $\bar{\tau} = \tau$.

На приемной стороне решающая схема, в соответствии с используемым методом приема, разделяет принимаемые сигналы на две области Z_0 и Z_1, так что если сигнал попадет в область Z_0, то считается (принимается решение), что был передан сигнал y_0, а если — в Z_1, то — y_1.

Для краткости записи обозначим априорные вероятности передачи сигналов $p(y_0) = p_0$ и $p(y_1) = p_1 = 1 - p_0$, а вероятности ошибочного решения $p(Z_0|y_1) = p_{\text{I}}$ и $p(Z_1|y_0) = p_{\text{II}}$.

В общем случае полагаем, что вероятности ошибок первого (p_{I}) и второго (p_{II}) видов могут быть неравными.

Очевидно, что при данных обозначениях вероятности правильного решения будут $p(Z_0|y_0) = 1 - p_{\text{II}}$ и $p(Z_1|y_1) = 1 - p_{\text{I}}$.

Используя правила теории вероятностей, можно сразу написать значения априорных вероятностей различных решений на приемной стороне

$$\left.\begin{aligned} p(Z_0) &= p_0(1 - p_{\text{II}}) + (1 - p_0)\, p_{\text{I}}, \\ p(Z_1) &= (1 - p_0)(1 - p_{\text{I}}) + p_0 p_{\text{II}}. \end{aligned}\right\} \qquad (10.68)$$

Применяя (10.5) и (10.64), получаем

$$\bar{H}(Z) = -\frac{1}{\tau}\left[p(Z_0)\log p(Z_0) + p(Z_1)\log p(Z_1)\right]. \qquad (10.69)$$

Применяя (10.14), непосредственно имеем

$$H(Z|Y) = -p(y_0)\left[p(Z_0|y_0)\log p(Z_0|y_0) + p(Z_1|y_0)\log p(Z_1|y_0)\right] + \\ + p(y_1)\left[p(Z_0|y_1)\log p(Z_0|y_1) + p(Z_1|y_1)\log p(Z_1|y_1)\right]$$

или в принятых обозначениях вероятностей, учитывая (10.68),

$$\bar{H}(Z|Y) = -\frac{1}{\tau}\left[p_0(1 - p_{\text{II}})\log(1 - p_{\text{II}}) + p_0 p_{\text{II}}\log p_{\text{II}} + \\ + (1 - p_0)\, p_{\text{I}}\log p_{\text{I}} + (1 - p_0)(1 - p_{\text{I}})\log(1 - p_{\text{I}})\right]. \qquad (10.70)$$

Подставляя найденные значения $\bar{H}(Z)$ и $\bar{H}(Z|Y)$ в (10.63), получим соотношение, определяющее скорость передачи информации — $\bar{I}$, из которого может быть найдено оптимальное значение априорной вероятности ($p_{0\,\text{opt}}$) передачи сигнала y_0, при котором $\bar{I}$ принимает максимальное значение. Так как больше никаких других параметров канала менять мы не можем, то, зная $p_{0\,\text{opt}}$, найдем C_ε. Рассмотрим два крайних случая.

Пример 9. Симметричный канал. В этом канале $p_{\text{I}} = p_{\text{II}} = p_{\text{ош}}$. Из (10.70) находим

$$\bar{H}(Z|Y) = -\frac{1}{\tau}\left[(1 - p_{\text{ош}})\log(1 - p_{\text{ош}}) + p_{\text{ош}}\log p_{\text{ош}}\right].$$

Как видно, $\bar{H}(Z|Y)$ не зависит от p_0. В таком случае, скорость передачи информации будет максимальной, когда величина $\bar{H}(Z)$ максимальна. Последнее справедливо, если $p(Z_0) = p(Z_1) = 0{,}5$. При этом из (10.69) $\bar{H}(Z)_{\max} = \frac{1}{\tau}$, и тогда

$$C_\varepsilon = \frac{1}{\tau}\left[1 + (1 - p_{\text{ош}})\log(1 - p_{\text{ош}}) + p_{\text{ош}}\log p_{\text{ош}}\right]. \qquad (10.71)$$

Из (10.68) находим, что при равенстве $p(Z_0)$ и $p(Z_1)$ $p_{0\,\text{opt}} = \frac{1}{2}$, и значит, и $p_{1\,\text{opt}} = \frac{1}{2}$.

Таким образом, условие обеспечения максимальной скорости передачи информации в симметричном канале А с шумами такое же, как и в канале без шумов. Однако из сравнения (10.71) с (10.48) следует, что ошибки в решениях о принадлежности принятых сигналов приводят к уменьшению пропускной способности канала.

На рис. 10.6 приведен график зависимости $C_\varepsilon\tau$ от $p_{\text{ош}}$. Как видно из этого графика, при $p_{\text{ош}} = \frac{1}{2}$ пропускная способность канала $C_\varepsilon = 0$. Этот результат станет очевидным, если учесть, что если

$p_{\text{ош}}=\frac{1}{2}$, то при передаче каждого из сигналов y_0 и y_1 на приемной стороне с равной вероятностью может быть принято решение Z_0 и Z_1. Эти решения будут иметь такую же ценность, как если бы передача сигналов прекратилась и на приемной стороне мы бы принимали решения по результатам бросания монеты (цифра или герб). Интересно также отметить, что если $p_{\text{ош}}>\frac{1}{2}$, то с увеличением $p_{\text{ош}}$ пропускная способность возрастает. Этот на первый взгляд парадоксальный вывод становится очевидным, если принять во внимание, что мы всегда можем изменить правило распознавания сигналов на обратное, т. е. считать, что решение Z_1 соответствует передаче y_0, а решение Z_0 — передаче y_1. Тогда вероятность ошибочного решения станет равной $1-p_{\text{ош}}$.

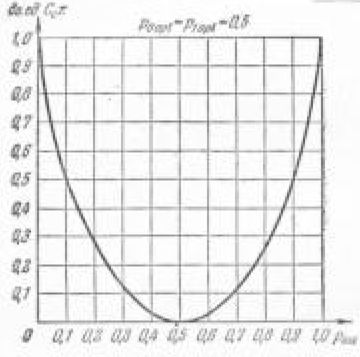

Рис. 10.6.

Пример 10. В канале возможны лишь ошибки одного вида (крайняя несимметрия). Допустим, что $p_1=p(Z_0|y_1)=0$, $p_{\text{II}}=p(Z_1|y_0)>0$, т. е. ошибочные решения имеют место лишь при передаче сигналов y_0.

Например, импульсы всегда регистрируются правильно, а паузы в полезном сигнале вследствие воздействия шума иногда регистрируются на приемной стороне как импульсы.

Из (10.63), (10.69), (10.68) и (10.70) для этого случая получаем

$$\bar{I}=-\frac{1}{\tau}\left[p_0\log p_0-p_0p_{\text{II}}\log p_0p_{\text{II}}+(1-p_0+p_0p_{\text{II}})\log(1-p_0+p_0p_{\text{II}})\right].$$

Исследуя это соотношение на максимум, находим

$$p_{0\,\text{opt}}=\frac{1}{1+p_{\text{II}}^{\frac{p_{\text{II}}}{1-p_{\text{II}}}}-p_{\text{II}}}.$$

Пропускная способность этого канала определяется следующим выражением:

$$C_c=-\frac{1}{\tau}\left[p_{0\,\text{opt}}\log p_{0\,\text{opt}}-p_{0\,\text{opt}}p_{\text{II}}\log(p_{0\,\text{opt}}p_{\text{II}})+\right.$$
$$\left.+(1-p_{0\,\text{opt}}+p_{0\,\text{opt}}p_{\text{II}})\log(1-p_{0\,\text{opt}}+p_{0\,\text{opt}}p_{\text{II}})\right]. \quad (10.72)$$

На рис. 10.7 приведены графики $C_c\tau$, $p_{0\,\text{opt}}$ и $p_{1\,\text{opt}}=1-p_{0\,\text{opt}}$ как функции от вероятности ошибки p_{II}. Как видим, в данном случае с увеличением p_{II} от 0 до 1 пропускная способность падает от $\frac{1}{\tau}$ до нуля. При $p_{\text{II}}=\frac{1}{2}$

$$C_c=\frac{0{,}318}{\tau}\ \frac{\text{дв. ед.}}{\text{сек}}.$$

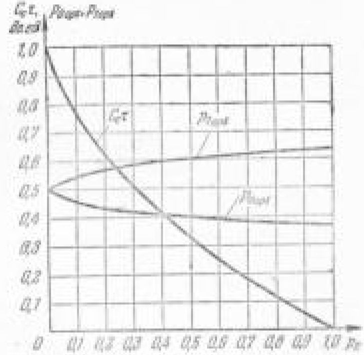

Рис. 10.7.

С увеличением p_{II} оптимальные статистические характеристики меняются: $p_{0\,\text{opt}}$ падает, а $p_{1\,\text{opt}}$ растет, т. е. сигналы y_0 должны передаваться реже, поскольку они чаще искажаются, а сигналы y_1 должны передаваться чаще. Изменение $p_{0\,\text{opt}}$ и $p_{1\,\text{opt}}$, однако, не очень велико, при $p_{\text{II}}\to 1$ $p_{0\,\text{opt}}\to\frac{1}{e}\approx 0{,}368$ и соответственно $p_{1\,\text{opt}}\to 1-\frac{1}{e}\approx 0{,}632$.

Средняя вероятность ошибки для рассматриваемого крайне несимметричного канала равна $p_{\text{ош}}=p_{0\,\text{opt}}p_{\text{II}}$, а для симметричного канала (пример 9) $p_1=p_{\text{II}}=p_{\text{ош}}$.

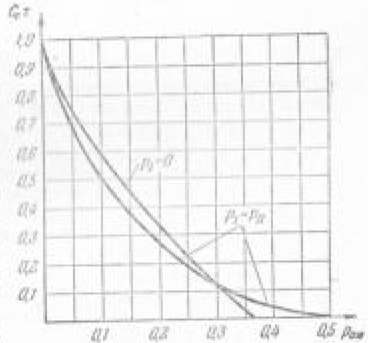

Рис. 10.8.

На рис. 10.8 представлены графики зависимости пропускной способности канала связи (точнее, $C_c\tau$) от средней вероятности ошибки ($p_{\text{ош}}$) для симметричного и крайне несимметричного каналов. Из этих графиков следует, что если $p_{\text{ош}}\leqslant 0{,}3$, то одним и тем же значением $p_{\text{ош}}$ соответствуют приблизительно одинаковые пропускные способности симметричного и несимметричного каналов.

Основная теорема Шэннона для дискретного канала с шумами. Для дискретного канала с шумами К. Шэнноном доказана следующая важная теорема.

Теорема 3. Если поток информации $\bar{H}(X)$, вырабатываемый источником,

$$\bar{H}(X) = C_c - \varepsilon, \tag{10.73}$$

где ε как угодно мало, то существует способ кодирования, при котором все сообщения, вырабатываемые источником, смогут быть переданными, а вероятность ошибочного опознания любого переданного сообщения может быть как угодно малой, т. е.

$$p_p < \eta, \tag{10.74}$$

где p_p — вероятность неправильного опознания любого переданного сообщения, η — как угодно малая величина ($\eta > 0$).

Обратное утверждение рассматриваемой теоремы состоит в том, что если $\bar{H}(X) > C_c$, то не существует способа кодирования, обеспечивающего передачу всех сообщений с малой вероятностью ошибки.

Доказательство сформулированной теоремы приведено ниже.

Из этого доказательства следует, что в общем случае выполнение (10.73) может быть достигнуто с помощью кодирования достаточно длинных последовательностей сообщений. Если M_x — число сообщений в одной кодируемой последовательности C_x, то оказывается, что чем меньше ε в (10.73) и η в (10.74), тем большим должно быть M_x, т. е. тем длиннее должна быть кодируемая последовательность. В пределе при $\varepsilon \to 0$ и $\eta \to 0$ необходимо, чтобы $M_x \to \infty$.

Кодирование длинных последовательностей сообщений связано с неблагоприятными обстоятельствами, отмеченными при изучении теоремы 2 для дискретного канала без шумов.

Следует, однако, иметь в виду, что теорема 3 для дискретных каналов с шумами, так же как и теорема 2 для каналов без шумов, не утверждает, что кодирование длинных последовательностей сообщений является единственным методом эффективного кодирования. Смысл этих теорем состоит в утверждении существования эффективных методов кодирования и в установлении количественных пределов максимально возможной скорости передачи информации. В связи с этим важными являются не только прямые, но и обратные утверждения этих теорем. Из доказательства теорем 2 и 3 вытекает лишь, что кодированием достаточно длинных последовательностей сообщений всегда можно как угодно близко подойти к максимально возможной скорости передачи сообщений (при минимальной вероятности ошибки для каналов с шумами). Последнее, однако, не означает, что не могут существовать другие способы эффективного кодирования. Наоборот, на ряде частных примеров можно показать, что такие способы существуют *).

*) Таким хотя и специально подобранным примером для канала без шумов является код в таблице 10.3.

К сожалению, в настоящее время не найдено еще общих способов построения эффективных кодов для каналов с шумами, удовлетворяющих различным требованиям практики. Постепенно, однако, такие способы выявляются. Весьма интересным и важным является утверждение теоремы 3 о том, что в канале с шумами при сколь угодно малой ненадежности передачи сообщений ($\eta \to 0$) скорость передачи информации может быть как угодно близкой к C_c. Ранее господствовало мнение, основанное на интуитивных соображениях, что при этих требованиях скорость передачи информации должна неограниченно уменьшаться.

Фундаментальное значение теоремы 3 состоит в том, что она позволяет, зная предельные (теоретические) значения скорости передачи информации C_c, оценить эффективность используемых методов кодирования и модуляции сигналов в заданной системе передачи информации.

Доказательство теоремы 3. Справедливость утверждения теоремы 3 мы докажем, применяя кодирование достаточно длинных последовательностей C_x, состоящих из M_x сообщений.

Аналогично тому как это было сделано при доказательстве теоремы 2, можно показать, что средняя скорость передачи информации определяется лишь типичными последовательностями; что же касается нетипичных последовательностей, то при кодировании их необходимо лишь заботиться о том, чтобы обеспечить малую вероятность неправильного их распознавания на приемной стороне, а эффективность кодирования этих последовательностей значения не имеет, ибо они передаются очень редко и влияние их на среднюю скорость передачи информации с увеличением M_x может быть весьма малым.

Допустим, что для кодирования типичной последовательности сообщений C_x используется последовательность C_y, состоящая из M_y сигналов (символов).

Согласно (10.33а), если M_y достаточно велико, то число типичных последовательностей C_y сигналов, образуемых на передающей стороне, равно

$$N_1 \approx 2^{H(Y)\,M_y}, \tag{1}$$

а на приемной стороне

$$N_2 \approx 2^{H(Z)\,M_y}. \tag{2}$$

Рассуждая так же, как при выводе (10.33а), можно показать, что каждая передаваемая последовательность сигналов C_y вследствие искажений, претерпеваемых ею из-за воздействия шумов, порождает на приемной стороне в среднем

$$N_3 \approx 2^{H(Z|Y)\,M_y} \tag{3}$$

типичных последовательностей сигналов C_z (рис. 10.9). В действительности, конечно, число различных последовательностей сигналов на приемной стороне, порождаемых одной передаваемой последовательностью, будет больше N_3, однако последовательности, не учитываемые (3), нетипичны и суммарная вероятность их может быть при большом M_y как угодно малой.

Для получения малой вероятности неправильного опознания принятой последовательности сигналов необходимо, чтобы при передаче некоторой последовательности C_{yi} все порождаемые ею на приемной стороне типичные последовательности C_z попадали в одну соответствующую область решающей схемы — C_{zi}. Казалось бы, что число таких областей может быть взято равным $\frac{N_2}{N_3}$, однако это не так, ибо в таком случае области решений будут близко располагаться друг от друга и вероятность неправильного решения может быть большой. Для отдаления областей различных решений друг от друга возьмем число их равным

$$N_p = \frac{N_2}{N_3} e^{-\alpha M_y}, \tag{4}$$

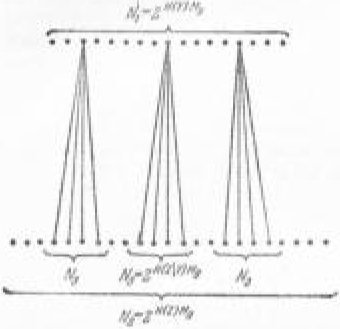

Рис. 10.9.

где α — некоторая малая величина. N_p может быть выбрано равным числу передаваемых типичных последовательностей сообщений C_x. Коль скоро это так, то число различных типичных последовательностей сигналов на приемной стороне будет равно

$$N_{z1} = N_p N_3 = N_2 \cdot 2^{-\alpha M_y}$$

и, следовательно,

$$\frac{N_{z1}}{N_2} = 2^{-\alpha M_y}. \tag{5}$$

Как бы ни было мало α, при достаточно большом M_y величина $2^{-\alpha M_y}$ будет достаточно малой и, следовательно, при выборе числа N_p, согласно (4), на приемной стороне из всех возможных типичных последовательностей сигналов N_2 используется лишь как угодно малая (при большом M_y) часть их N_{z1}; в таком случае используемые сигналы могут быть выбраны так, чтобы области принятия различных решений были достаточно далеко разнесены друг от друга, и тогда вероятность принятия неправильного решения будет приемлемо малой.

Данное положение более наглядно поясняется с точки зрения теории корректирующих кодов.

Из (2) следует, что с увеличением M_y число различных последовательностей сигналов может неограниченно возрастать; в то же время, согласно (5), из всего этого множества последовательностей используется как угодно малая часть. В таком случае всегда может быть построен код, исправляющий любое заданное число ошибок и, следовательно, обеспечивающий как угодно малую вероятность неправильного опознания принятой последовательности.

Из (4), (2) и (3) следует, что

$$N_p \approx 2^{[H(Z) - H(Z|Y) - \alpha] M_y}. \tag{6}$$

Пусть

$$T = M_y \bar{\tau}_c, \tag{7}$$

где $\bar{\tau}_c$ — средняя длительность сигнала одного сообщения, а T — время передачи последовательности сигналов. Тогда (6) преобразуется к виду

$$N_p \approx 2^{[\bar{H}(Z) - \bar{H}(Z|Y) - \varepsilon] T},$$

где $\varepsilon = \frac{\alpha}{\bar{\tau}_c}$ — сколь угодно малая величина.

Используя (10.33а) для источника сообщений на входе канала, получим

$$N_{x1} \approx 2^{H(X) M_x} = 2^{\bar{H}(X) T}. \tag{9}$$

Приравнивая $N_{x1} = N_p$, из (9), (8) и сказанного ранее получаем, что все типичные последовательности сообщений могут быть переданы с как угодно малой вероятностью неправильного распознания, если только

$$\bar{H}(X) = \bar{H}(Z) - \bar{H}(Z|Y) - \varepsilon.$$

Это соотношение и (10.65) свидетельствуют о справедливости прямого утверждения теоремы 3.

Обратное утверждение вытекает из того соображения, что если $\bar{H}(X) > C_c$ и, следовательно,

$$\bar{H}(X) > \bar{H}(Z) - \bar{H}(Z|Y),$$

то $N_p < N_{x1}$. Последнее означает, что не все типичные последовательности сообщений могут быть закодированы таким образом, чтобы они опознавались на приемной стороне с приемлемо малой ошибкой.

§ 5. Информационные характеристики источников непрерывных сообщений

Понятие об источниках непрерывных сообщений. Для источника непрерывных сообщений характерным является непрерывное изменение во времени или в пространстве физического параметра, значение которого измеряется или передается на расстояние для получения информации (сведений) о том или ином явлении, факте, процессе и т. п. Так, например, непрерывное изменение во времени давления на мембрану микрофона, создаваемого говорящим человеком, позволяет передать на расстояние информацию о содержании его речи, интонации и тембре голоса и т. п.

При исследовании различных процессов или явлений производят измерение большого количества разных величин: температуры, влажности, давления, вязкости, освещенности, перемещений, скорости движения, механического напряжения, электрического напряжения, силы тока, частоты и т. п.

Результаты таких измерений либо непосредственно регистрируются (фиксируются), либо передаются на расстояние.

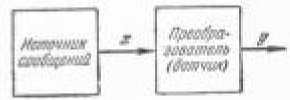

Рис. 10.10.

Для регистрации или передачи на расстояние параметра x, величина которого непрерывно изменяется, как правило, с помощью специального устройства (датчика) производится преобразование этого параметра в электрический сигнал y. Общая схема такого преобразования показана на рис. 10.10. Смысл этого преобразования состоит в установлении некоторого соответствия между y и x.

Таким соответствием может быть функциональная связь $y=f(x)$. Однако всякий реальный преобразователь вносит случайные погрешности, нарушающие функциональную связь, и интересующее нас соответствие между y и x может характеризоваться лишь условным распределением $W(y|x)$ или зависимостью математического ожидания случайной величины Y от x, т. е. $M[Y]=f(x)$, или дисперсией Y при известном x, т. е. $D[Y|x]$.

Примерами преобразователей могут быть микрофон, фотоэлемент, термоэлемент, потенциометрический датчик и т. п.

Для распространения теории информации на случай источников непрерывных сообщений мы вначале определим количество информации, содержащееся в одном замере Y случайной величины X, а затем рассмотрим методику определения количества информации, содержащегося в случайном сигнале $Y(t)$ на выходе преобразователя, о случайной функции $X(t)$, которая описывает изменение параметра X во времени.

Количество информации, содержащееся в одном замере непрерывной случайной величины. Для конкретности будем полагать, что в схеме рис. 10.10 преобразователь представляет собой измерительный прибор, показание которого y рассматривается как результат измерения случайной величины. Такая конкретизация не повлияет на общность соотношений, которые будут получены ниже. Определим, какое количество информации о величине X содержится в величине Y.

Легко убедиться в том, что если бы измерение было абсолютно точным, т. е. не содержало погрешности, то количество информации должно было быть бесконечно большим. Действительно, будем выражать результат измерения y числом в двоичной системе счисления. Если x иррационально, а это всегда может быть, поскольку x непрерывно изменяется в некотором интервале, то для того, чтобы y точно описывало x, потребуется двоичное число с бесконечно большим числом разрядов, что соответствует бесконечно большому числу двоичных единиц информации. Другими словами, если x — непрерывная случайная величина, которая может принимать всевозможные значения в некотором интервале $x=x_{\min} \div x_{\max}$, то число этих значений бесконечно велико, и точное определение значения x при измерении соответствует получению бесконечно большого количества информации. Это обстоятельство не позволяет понятие об энтропии дискретных случайных величин непосредственно распространить на непрерывные величины, ибо для дискретных величин $H(X)=I(X, X)$ (см. § 1), а для непрерывных величин $I(X, X)$ бесконечно велико.

В действительности, однако, измерение (определение) любой случайной величины может быть выполнено лишь с некоторой степенью точности, и в лучшем случае мы можем определить лишь функцию распределения получаемой погрешности. Наличие погрешности ведет

к тому, что количество информации, содержащееся в случайной величине Y, о величине X становится конечным. Убедимся в этом на простом примере.

Обозначим через

$$L_x = x_{\max} - x_{\min}, \tag{10.74'}$$

где L_x — интервал (шкала) изменения параметра x. Допустим, что измерение x ведется с точностью ε, так что показания y измерительного прибора могут иметь лишь $m=\frac{L_x}{\varepsilon}$ различных значений, которые мы будем обозначать через y_*. Очевидно, что информация, содержащаяся в одном измерении, будет максимальна, если все результаты равновероятны; тогда

$$I(Y, X)_{\max} = H(Y_*) = \log_2 m$$

или

$$I(Y, X)_{\max} = \log_2 \frac{L_x}{\varepsilon}. \tag{10.75}$$

Из последнего соотношения следует, что при конечном, хотя и малом ε величина $I(Y, X)_{\max}$ также конечна, если же $\varepsilon \to 0$, то $I(Y, X)_{\max} \to \infty$.

Заметим еще, что если x измеряется с некоторой погрешностью, то для его выражения числом в двоичной системе потребуется лишь конечное число разрядов.

Сказанное дает основание распространить соотношения, определяющие $I(Y, X)$ для дискретных случайных величин, на непрерывные величины. Положим, что нам известны:

$W(x)$ — априорная функция распределения (плотность вероятности) случайной величины X.

$W(y|x)$ — условная функция распределения случайной величины Y при известном x.

Последняя функция характеризует погрешность преобразования. Так, например, допустим, что

$$y = x + \delta, \tag{10.76}$$

где δ — погрешность измерения, являющаяся случайной величиной, независимой от x. Тогда очевидно, что

$$W(y|x) = w_\Delta(\delta),$$

где $W_\Delta(\delta)$ — плотность распределения вероятности погрешностей измерения.

Так как $\delta = y - x$, то соответственно получим

$$W(y|x) = W_\Delta(y-x). \tag{10.77}$$

На рис. 10.11 приведены примеры графиков плотности распределения вероятностей $W_\Delta(\delta)$ и $W(y|x)$.

Зная $W(x)$ и $W(y|x)$, можно найти плотность совместного распределения вероятностей

$$W(x, y) = W(x)\,W(y|x) \tag{10.78}$$

и плотность распределения вероятностей случайной величины Y

$$W(y)=\int\limits_{L_x} W(x,\ y)\,dx, \tag{10.79}$$

где знак L_x указывает на то, что интегрирование ведется по всем возможным значениям x.

Найдем также плотность условного распределения вероятностей

$$W(x\,|\,y)=\frac{W(x,\ y)}{W(y)}. \tag{10.80}$$

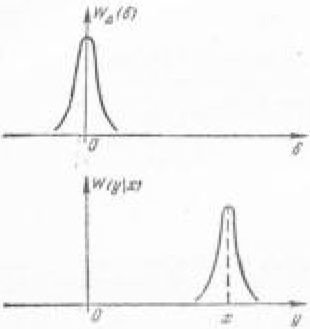

Рис. 10.11.

Совокупность перечисленных распределений вероятностей полностью описывает вероятностные зависимости между величинами x и y.

Для того чтобы перейти от изучения дискретных величин к исследованию непрерывных, мы можем, как это обычно делается, непрерывную величину в первом приближении представить как дискретную. Так, в частности, разбив интервал L_x на конечное число малых интервалов Δx и перенумеровав их, можно совокупность всех значений x внутри i-го интервала представлять одним средним значением x_i, априорная вероятность которого будет $p(x_i)=W(x_i)\,\Delta x$. Поступая аналогично с величиной y и используя (10.21a), получим

$$I(Y,\ X)\approx\sum_i\sum_j W(x_i,\ y_j)\,\Delta x\,\Delta y\log\frac{W(x_i\,|\,y_j)\,\Delta x}{W(x_i)\,\Delta x}.$$

В дроби, стоящей под знаком логарифма, умножим числитель и знаменатель на $W(y_j)$, тогда

$$I(Y,\ X)\approx\sum_i\sum_j W(x_i,\ y_j)\,\Delta x\,\Delta y\log\frac{W(x_i,\ y_j)}{W(x_i)\,W(y_j)}.$$

Переходя к пределу при $\Delta x\to 0$ и $\Delta y\to 0$, получим

$$\boxed{I(Y,\ X)=\int\limits_{L_x}\int\limits_{L_y} W(x,\ y)\log\frac{W(x,\ y)}{W(x)\,W(y)}\,dx\,dy.} \tag{10.81}$$

Полученное соотношение (10.81) является весьма важным. Оно дает ответ на вопрос о количестве информации, содержащейся в непрерывной случайной величине Y, о непрерывной величине X. В этой форме оно выведено К. Шэнноном. В работах академика А. Н. Колмогорова*) даются более общие соотношения, применимые к широкому классу случайных объектов. Значительное развитие теории информации для непрерывных сообщений дано в работе М. С. Пинскера**).

Отметим наиболее важные свойства соотношения (10.81).

1. $I(X,\ Y)=I(Y,\ X)$. Это следует из того, что

$$W(x,\ y)=W(y,\ x).$$

2. $I(Y,\ X)\geqslant 0$, причем $I(Y,\ X)=0$, если X и Y независимые случайные величины.

В последнем случае $W(x,\ y)=W(x)\,W(y)$ и выражение, стоящее под знаком логарифма, становится равным единице.

3. Значение $I(Y,\ X)$ не зависит от способа отсчета величин x и y или, как говорят, от выбора системы координат.

Так, в частности, если мы вместо y будем выражать результат измерения новой величиной $y_1=f(y)$, где $f(y)$ — некоторая регулярная функция от y, то можно доказать***), что

$$I(Y_1,\ X)=I(Y,\ X).$$

Последнее равенство означает, например, что результат измерения длины при данной погрешности измерения содержит одно и то же количество информации независимо от того, будем ли мы выражать его в метрах, сантиметрах или в других единицах.

В общем случае, если от величин y и x перейти к новым величинам y_1 и x_1, функционально связанным с первыми так, что $y_1=f_1(x,\ y)$ и $x_1=f_2(x,\ y)$, то

$$I(Y_1,\ X_1)=I(Y,\ X).$$

Указанные свойства соотношения (10.81) согласуются с общими требованиями к количественной мере информации, рассмотренными в § 1.

Энтропия непрерывных случайных величин. Если учесть, что $W(x,\ y)=W(x)\,W(y\,|\,x)=W(y)\,W(x\,|\,y)$, то (10.81) может быть преобразовано к виду

$$I(Y,\ X)=H(X)-H(X\,|\,Y), \tag{10.82}$$

где

$$H(X)=-\int\limits_{L_x} W(x)\log W(x)\,dx, \tag{10.83}$$

$$H(X\,|\,Y)=-\int\limits_{L_x}\int\limits_{L_y} W(y)\,W(x\,|\,y)\log W(x\,|\,y)\,dx\,dy. \tag{10.84}$$

*) А. Н. Колмогоров, Теория передачи информации, Изд. АН СССР, Москва, 1956.

**) М. С. Пинскер, Информация и информационная устойчивость случайных величин и процессов, Изд. АН СССР, Москва, 1960.

***) См., например, С. Голдман, Теория информации, ИЛ, Москва, 1957.

Соотношение (10.83) по структуре сходно с (10.5), а (10.84) — с (10.20). В силу этого $H(X)$ и $H(X|Y)$ могут быть названы соответственно энтропией и условной энтропией непрерывной случайной величины X.

Рассмотрим совокупность двух случайных величин X_1, X_2. Это может понадобиться в том случае, когда случайная величина X зависит от двух случайных параметров. Например, текущее значение гармонического колебания определяется его амплитудой (X_1) и полной фазой (X_2). Такая совокупность случайных величин может рассматриваться как вектор X с двумя координатами (X_1, X_2).

Пусть $W(x_1, x_2)$ есть совместная плотность распределения вероятностей X_1 и X_2. Тогда, распространяя формулу (10.83) на этот случай, получим

$$H(X_1, X_2) = -\int\limits_{L_{x_1}}\int\limits_{L_{x_2}} W(x_1, x_2)\log W(x_1, x_2)\,dx_1\,dx_2. \tag{10.85}$$

Очевидно, что и для n-мерной случайной величины $(X_1, \ldots, X_n)$ справедливо

$$H(X) = H(X_1, \ldots, X_n) =$$
$$= -\int\limits_{L_{x_1}}\ldots\int\limits_{L_{x_n}} W(x_1, \ldots, x_n)\log W(x_1, \ldots, x_n)\,dx_1 \ldots dx_n. \tag{10.86}$$

Нетрудно убедиться в том, что энтропии, определяемая соотношениями (10.83) — (10.86), обладает многими свойствами энтропии дискретных случайных величин.

1. Если X ограничено некоторым объемом V в своем пространстве ($V = L_{x_1}L_{x_2}\ldots L_{x_n}$), то $H(X)$ максимальна и равна $\log V$, если совместное распределение $W(x_1, \ldots, x_n)$ постоянно и равно $\frac{1}{V}$ в этом объеме.

2. При любых X_1 и X_2

$$H(X_1, X_2) \leqslant H(X_1) + H(X_2), \tag{10.87}$$

причем равенство выполняется только тогда, когда X_1 и X_2 независимы.

3. $$H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y) \tag{10.88}$$

и

$$H(Y|X) \leqslant H(Y).$$

Несмотря на общность свойств энтропии дискретных и непрерывных случайных величин, между ними имеется и существенное отличие. Как уже указывалось выше, для дискретных случайных величин $H(X) = I(X, X)$. Для непрерывных величин это равенство не имеет места, в чем нетрудно убедиться из сопоставления (10.83) и (10.81).

Таким образом, для дискретных случайных величин энтропия определяется как количество информации, получаемое при полной достоверности опыта, *а для непрерывных случайных величин энтропия может быть определена лишь как некоторая величина, с помощью которой можно определить количество информации по (10.82).* Следует также иметь в виду, что значение энтропии зависит от выбора системы координат (способа отсчета x и y). Эта зависимость такова, что при вычислении разности в (10.82) соответствующие слагаемые взаимно компенсируются так, что $I(Y, X)$ от выбора координат не зависит *).

Примеры вычисления энтропии непрерывных случайных величин и среднего количества информации в одном замере. Пример 11. Случайная величина x распределена равномерно, так что

$$W(x) = \begin{cases} \dfrac{1}{L_x} \text{ при } x_{\min} \leqslant x \leqslant x_{\max}, \\ 0 \text{ вне этого интервала,} \end{cases} \tag{10.89}$$

$$L(x) = x_{\max} - x_{\min}.$$

Вычислим энтропию $H(X)$. Из (10.83) получим

$$H(X) = -\int\limits_{x_{\min}}^{x_{\max}} \frac{1}{L_x}\log\frac{1}{L_x}\,dx,$$

откуда

$$H(X) = \log L_x. \tag{10.90}$$

Пример 12. Случайная величина x распределена по нормальному закону

$$W(x) = \frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{x^2}{2\sigma^2}}. \tag{10.91}$$

Используя (10.83), получаем

$$H(X) = -\int\limits_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{x^2}{2\sigma^2}}\log\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{x^2}{2\sigma^2}}\,dx.$$

Вычисление этого интеграла дает

$$H(X) = \log\left(\sigma\sqrt{2\pi e}\right). \tag{10.92}$$

В приведенных примерах вычислялись энтропии известных функций распределения. Используя методы вариационного исчисления,

*) Для непрерывного случая энтропия может быть истолкована как мера неопределенности в выбранной системе координат.

можно решать и задачи обратного типа, например найти функцию распределения $W(x)$, обеспечивающую максимальное значение энтропии $H(X)$ при некоторых заданных ограничениях. Если таким ограничением является конечная величина L_x, так что, например, $W(x) \geqslant 0$ только при $x_{\min} \leqslant x \leqslant x_{\max}$ и $W(x) = 0$ вне этого интервала, то наибольшее значение $H(X)$ дает равномерное распределение (10.89) *).

Если заданное ограничение может быть представлено в виде

$$\int_{-\infty}^{\infty} x^2 W(x)\,dx = \sigma^2,$$

то наибольшее значение $H(X)$ достигается при нормальном распределении (10.91).

Таким образом, нормальное распределение дает максимальное значение энтропии, если интервал изменения случайной величины не ограничен, а ограничено среднеквадратичное значение этой величины.

Пример 13. Определим количество информации, содержащееся в одном замере случайной величины x, если $W(x)$ описывается (10.89), а результат измерения описывается (10.76), т. е. если $y = x + \delta$. При этом погрешность измерения δ не зависит от x и распределена по нормальному закону

$$W_\Delta(\delta) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{\delta^2}{2\sigma^2}}. \tag{10.93}$$

При вычислении воспользуемся соотношением

$$I(Y, X) = H(Y) - H(Y|X). \tag{10.94}$$

Справедливость последнего вытекает из (10.82) и свойства $I(Y, X) = I(X, Y)$.

Из (10.77) и (10.93) имеем

$$W(y|x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(y-x)^2}{2\sigma^2}}. \tag{10.95}$$

Заменив в (10.84) x на y, а y на x и учитывая (10.89) и (10.95), находим

$$H(Y|X) = -\int_{-\infty}^{\infty}\int_{x_{\min}}^{x_{\max}} \frac{1}{L_x}\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(y-x)^2}{2\sigma^2}} \log \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(y-x)^2}{2\sigma^2}}\,dy\,dx.$$

*) В общем виде это сформулировано при указании свойства 1 энтропии $H(X)$.

Вычисление этого интеграла дает

$$H(Y|X) = \log \sigma \sqrt{2\pi e}. \tag{10.96}$$

Для определения $H(Y)$ необходимо найти плотность распределения $W(y)$. Так как y является суммой двух случайных независимых величин x и δ с известными законами распределения, то $W(y)$ является сверткой распределений слагаемых *), т. е.

$$W(y) = \int_{-\infty}^{\infty} W(x)\, W_\Delta(y - x)\,dx.$$

Проинтегрировав, получим

$$W(y) = \frac{1}{L_x}\left[F\left(\frac{y - x_{\min}}{\sigma}\right) - F\left(\frac{y - x_{\max}}{\sigma}\right)\right], \tag{10.97}$$

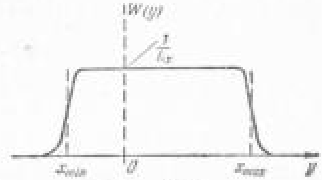

Рис. 10.12.

где $F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}}\,dt$ — табулированная функция Лапласа.

График функции $W(y)$ приведен на рис. 10.12.

Из свойств функции Лапласа легко убедиться, что с большой степенью точности

$$W(y) \approx \begin{cases} \frac{1}{L_x} & \text{при} \quad x_{\min} + 3\sigma < y < x_{\max} - 3\sigma, \\ 0 & \text{при} \quad y < x_{\min} - 3\sigma \quad \text{и} \quad y > x_{\max} + 3\sigma. \end{cases}$$

Таким образом, полученное распределение отличается от равномерного лишь вблизи значений $y = x_{\min}$ и $y = x_{\max}$ на участках протяженностью 6σ. Если погрешность измерения невелика, то $6\sigma \ll L_x$.

Для равномерного распределения $W(y)$, используя (10.90), получим $H(Y) = \log L_x$. Учет небольшого отличия от равномерного распределения в следующем приближении дает

$$H(Y) \approx \log_2 L_x + 2{,}606 \frac{\sigma}{L_x}. \tag{10.98}$$

Из (10.94), (10.96) и (10.98) получаем

$$I(Y, X) \approx -\log_2 \frac{\sigma}{L_x}\sqrt{2\pi e} + 2{,}606 \frac{\sigma}{L_x}\ \frac{\text{дв. ед.}}{\text{замер}}.$$

*) См., например, Б. Р. Левин. Теория случайных процессов и ее применение в радиотехнике, «Советское радио», Москва, 1960.

Количество информации о непрерывной случайной величине при заданных требованиях к верности воспроизведения. Соотношение (10.81) определяет количество информации, содержащееся в случайной величине Y (например, показании измерительного прибора), о случайной величине X, коль скоро нам известны характеристики вероятностных связей между Y и X (например, функция распределения погрешности преобразования).

Однако возможна, а в ряде случаев и необходима другая постановка задачи.

Допустим, что по случайной величине Y нужно судить о величине X. В дальнейшем будем говорить, что *Y воспроизводит X*. Так как по y нельзя определить точное значение x, то мы можем предъявить некоторые требования к допустимым значениям ошибок, или, будем говорить, требования к *верности воспроизведения*.

В дальнейшем необходимо найти количество информации, содержащееся в случайной величине Y, о случайной величине X при заданных требованиях к верности воспроизведения.

Рассмотрим вначале, как можно задать такие требования, или, короче, какими могут быть *критерии верности воспроизведения*.

В общем случае такой критерий может требовать, чтобы плотность совместного распределения вероятностей $W(x, y)$ принадлежала к некоторому классу функций $\mathcal{W}$ или

$$W(x, y) \in \mathcal{W}.$$

В частных случаях при сопоставлении значений x и y нас могут интересовать величины $\rho(x, y) = |y - x|$ или $\rho(x, y) = (y - x)^2$, или какая-нибудь другая функция $\rho(x, y)$, имеющая природу «расстояния». В таком случае наиболее удобным критерием будет среднее значение функции $\rho(x, y)$ на множестве возможных значений x и y, т. е.

$$g = \int\limits_{L_x} \int\limits_{L_y} W(x, y) \rho(x, y) \, dx \, dy. \tag{10.98'}$$

При использовании критерия (10.98') требование к верности воспроизведения может быть задано в форме

$$g \leqslant \varepsilon, \tag{10.99}$$

где ε — приемлемо малая величина.

Так как $W(x, y) = W(x) W(y|x)$, то при выбранной функции $\rho(x, y)$ и известном распределении $W(x)$ мы можем добиваться выполнения заданного условия (10.99), варьируя в (10.98') функцию условного распределения $W(y|x)$. По-видимому, условию (10.98') будет удовлетворять не одна, а некоторое множество функций $W(y|x)$ и, следовательно, $W(x, y)$. Подставляя эти функции в (10.81), можно найти соответствующие им значения $I(Y, X)$. Очевидно, что наиболее выгодной будет функция $W(y|x)$, при которой $I(Y, X)$ имеет наименьшее значение, ибо выбор такой $W(y|x)$ позволяет при получении минимального количества информации выполнить заданные требования к верности воспроизведения. *Наименьшее значение $I(Y, X)$, при котором удовлетворяются заданные требования к верности воспроизведения, А. Н. Колмогоров назвал ε-энтропией.* Согласно определению ε-энтропия равна

$$H_\varepsilon(X) = \inf_{g \leqslant \varepsilon} \{I(Y, X)\}. \tag{10.100}$$

Знаки Inf и $g \leqslant \varepsilon$ указывают, что отыскивается нижняя грань (наименьшее значение) при условии выполнения заданного требования к верности воспроизведения. Эта грань отыскивается по множеству функций $W(y|x)$.

Из изложенного следует, что ε-энтропия равна минимальному количеству информации, которое должно содержать сообщение о случайной непрерывной величине X (например, результат измерения) при условии выполнения заданных требований к верности воспроизведения.

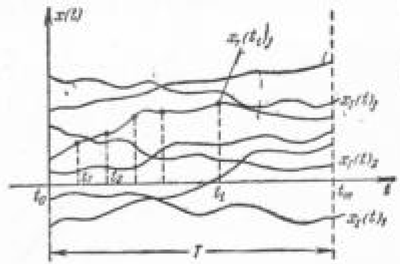

Рис. 10.13.

Представление непрерывного сообщения многомерным вектором. Рассмотрим реализацию $x(t)$ случайной функции X на некотором временном интервале T. На рис. 10.13 показаны реализации $x_T(t)_1, x_T(t)_2, \ldots$ Для простоты записи начало рассматриваемого интервала T будем совмещать с началом отсчета времени ($t = 0$). Такое совмещение нисколько не снижает общности всех последующих рассуждений.

В дальнейшем реализацию случайной функции на некотором интервале времени $T x_T(t)$ будем называть непрерывным сообщением или, короче, сообщением.

Совокупность всех возможных сообщений образует множество. Множество, для которого известны статистические закономерности, мы, как и ранее, будем называть ансамблем. Решение задачи об определении количества информации возможно лишь для величин или функций с известными статистическими свойствами, и потому ниже мы будем говорить лишь об ансамблях.

Любая реализация $x(t)=x_T(t)$, случайной функции на рассматриваемом интервале времени T может характеризоваться некоторой совокупностью параметров или, как говорят, функция $x(t)$ на отрезке времени T может быть представлена точкой или вектором в многомерном пространстве. В качестве координат такого пространства обычно используется совокупность (набор) коэффициентов разложения функции $x_T(t)$ в ряд по некоторой полной системе регулярных функций.

При использовании ряда Маклорена такими функциями являются $1,\ t,\ t^2,\ \ldots$ В результате разложения имеем

$$x_T(t)=a_0+a_1t+a_2t^2+\ldots+a_lt^l+\ldots$$

В общем случае ряд бесконечен, однако с некоторой допустимой погрешностью можно ограничиться конечным числом членов, и тогда

$$x_T(t)\approx\sum_{l=0}^{m-1}a_lt^l.$$

Коэффициенты $a_0,\ a_1,\ \ldots,\ a_{m-1}$ могут рассматриваться как координаты $x_T(t)$.

Если функцию $x_T(t)$ периодически продолжить вне рассматриваемого интервала T, то можно использовать тригонометрический ряд Фурье, дающий разложение $x_T(t)$ по системе гармонических функций:

$$x_T(t)\approx a_0+\sum_{l=1}^{\frac{m-1}{2}}(a_l\cos l\Omega t+b_l\sin l\Omega t),$$

где $\Omega=\frac{2\pi}{T}$.

При таком разложении координатами являются величины

$$a_0,\ a_1,\ b_1,\ a_2,\ b_2,\ \ldots,\ a_{\frac{m-1}{2}},\ b_{\frac{m-1}{2}}.$$

При использовании ряда Котельникова разложение производится по системе функций отсчетов

$$Sa\frac{\pi}{T_0}(t-lT_0),\quad \text{где}\quad l=0,\ 1,\ 2,\ \ldots$$

При этом $Sa\,u=\frac{\sin u}{u}$, а T_0 — некоторый период опроса.

$$x_T(t)\approx\sum_{l=0}^{m-1}x_T(lT_0)\,Sa\frac{\pi}{T_0}(t-lT_0), \qquad (10.101)$$

где

$$m=\frac{T}{T_0}+1. \qquad (10.102)$$

При использовании ряда Котельникова координатами является совокупность значений функции $x_T(t)$ в точках опроса:

$$x_T(0),\ x_T(T_0),\ x_T(2T_0),\ \ldots,\ x_T[(m-1)T_0].$$

Можно использовать и другие способы разложения функций $x_T(t)$ и, следовательно, другие способы задания координат. Важно, чтобы разложение было применимо к любому сообщению $x_T(t)$. Обозначим совокупность координат, характеризующих данное сообщение, через $x_1,\ x_2,\ \ldots,\ x_m$. В общем случае может использоваться любая приемлемая система координат и потому $x_1,\ x_2,\ \ldots$ могут быть коэффициентами того или иного разложения функции $x_T(t)$ в ряд и не являться значениями этой функции в точках опроса.

Для различных сообщений эти координаты принимают различные значения. В таком случае ансамбль случайных функций $x_T(t)$ (ансамбль сообщений) может характеризоваться совокупностью случайных величин $X_1,\ X_2,\ \ldots,\ X_m$. Если координаты $x_1,\ x_2,\ \ldots,\ x_m$ могут изменяться непрерывно, то статистическое описание ансамбля сообщений дается совместной плотностью вероятностей $W(x_1,\ x_2,\ \ldots,\ x_m)$. Последняя зависит от свойств источника сообщений.

Обратим еще раз внимание на то обстоятельство, что все разложения дают приближенное представление функции $x_T(t)$ на рассматриваемом интервале T. Обозначим через $x_T^0(t)$ сумму m первых членов ряда для этой функции. Для оценки величины погрешности используются те или иные функции оценки, например функция

$$\rho^2(x_T,\ x_T^0)=\frac{1}{T}\int_0^T[x_T(t)-x_T^0(t)]^2\,dt \qquad (10.103)$$

дает оценку по среднеквадратичному отклонению, а функция

$$\rho_1(x_T,\ x_T^0)=\frac{1}{T}\int_0^T|x_T(t)-x_T^0(t)|\,dt \qquad (10.104)$$

— оценку среднего значения абсолютной ошибки. Возможны и другие функции оценок*).

Если в последние соотношения подставить $x_T^0(t)$ в виде суммы m первых членов используемого ряда и выполнить интегрирование, то величина $\rho(x_T,\ x_T^0)$ будет определяться значением координат, т. е. $\rho(x_T,\ x_T^0)=\rho(x_1,\ x_2,\ \ldots,\ x_m)$.

Погрешность представления ансамбля сообщений в принятой системе координат может быть оценена по максимальному значению

*) Функция $\rho(x_T,\ x_T^0)$ имеет природу «расстояния», а способ ее задания определяет метрику в функциональном пространанстве.

$\rho_{\max}(x_T, x_T^0)$ или же (что более характерно) по среднему значению

$$g_k = \int\limits_{L_{x_1}} \int\limits_{L_{x_2}} \cdots \int\limits_{L_{x_m}} \rho(x_1, x_2, \ldots, x_m) W(x_1, x_2, \ldots, x_m) \times \\ \times dx_1 dx_2 \ldots dx_m. \qquad (10.105)$$

Приемлемость используемой системы координат оценивается в зависимости от выполнения условий

$$\rho_{\max}(x_T, x_T^0) < \eta_1 \qquad (10.106)$$

или

$$g_k < \eta_1, \qquad (10.107)$$

где η_1 — некоторая малая величина, определяемая условиями задачи.

Погрешность, определяемую по (10.106) или (10.107), мы будем называть *погрешностью системы координат*.

Заметим, что чем меньше должно быть η_1, тем больше членов ряда должно быть использовано при описании $x_T(t)$. Последнее означает увеличение числа координат m.

Количество информации, содержащееся в воспроизведении непрерывного сообщения. Допустим, что нами выбрана некоторая приемлемая m-мерная система координат. Тогда можно говорить о том, что ансамбль всех возможных сообщений $x_T(t)$ представляется ансамблем m-мерных векторов $(x_1, x_2, \ldots, x_m)$.

Информация о таком ансамбле конечна, если его сообщения воспроизводятся с некоторой погрешностью.

Допустим, что функция $y_T(t)$ является воспроизведением сообщения $x_T(t)$. Примерами такого воспроизведения служат показания измерительного регистрирующего прибора (осциллограмма, запись на магнитной пленке и т. п.), напряжение на выходе телеметрического датчика и др.

Воспроизведение $y_T(t)$, так же как и сообщение $x_T(t)$, может быть представлено вектором $(y_1, y_2, \ldots, y_m)$ в многомерном пространстве функций, где $y_1, y_2, \ldots, y_m$ — координаты $y_T(t)$. Точное значение функции, представляемое вектором $(y_1, \ldots, y_m)$, обозначим $y_T^0(t)$. Погрешность воспроизведения сообщения $x_T(t)$ функцией $y_T(t)$ разделим на две части: погрешность системы координат и погрешность преобразования.

Для определения погрешности преобразования используем функцию оценки $\rho(x_T^0, y_T^0)$, которую определим аналогично (10.103) или (10.104) либо другим подобным способом. Представляя функции $x_T^0(t)$ и $y_T^0(t)$ в выбранной системе координат и выполнив интегрирование, найдем, что величина $\rho(x_T^0, y_T^0)$ определяется значениями соответствующих координат, т. е. $\rho(x_T^0, y_T^0) = \rho(x_1, \ldots, x_m; y_1, \ldots, y_m)$.

Погрешность преобразования для ансамбля сообщений может быть оценена по (10.98′), если только учесть, что x_T и y_T — многомерные

величины. Приняв во внимание последнее обстоятельство, получим

$$g_n = \int\limits_{L_{x_1}} \cdots \int\limits_{L_{x_m}} \int\limits_{L_{y_1}} \cdots \int\limits_{L_{y_m}} W(x_1, \ldots, x_m; y_1, \ldots, y_m) \rho(x_1, \ldots \\ \ldots, x_m; y_1, \ldots, y_m) dx_1 \ldots dx_m dy_1 \ldots dy_m. \qquad (10.108)$$

Потребуем, чтобы

$$g_n < \eta_2, \qquad (10.109)$$

где η_2 — некоторая малая величина.

Будем полагать, что погрешности выбранной системы координат пренебрежимо малы, т. е. $x_T(t) = x_T^0(t)$ и $y_T(t) = y_T^0(t)$, так что суммарная погрешность (g) воспроизведения сообщения $x_T(t)$ функцией $y_T(t)$ определяется погрешностью преобразования (g_n)

$$g = g_n. \qquad (10.110)$$

Повторив рассуждения, приведенные при выводе (10.81) для случая, когда X и Y — многомерные величины, определим количество информации, содержащееся в воспроизведении Y_T сообщений X_T, в виде

$$I(Y_T, X_T) = \int\limits_{L_{x_1}} \cdots \int\limits_{L_{x_m}} \int\limits_{L_{y_1}} \cdots \int\limits_{L_{y_m}} W(x_1, \ldots, x_m; y_1, \ldots, y_m) \times \\ \times \log \frac{W(x_1, \ldots, x_m; y_1, \ldots, y_m)}{W(x_1, \ldots, x_m) W(y_1, \ldots, y_m)} dx_1 \ldots dx_m dy_1 \ldots dy_m. \qquad (10.111)$$

Совместная плотность распределения вероятностей $W(x_1, \ldots, x_m; y_1, \ldots, y_m)$ может быть представлена в виде

$$W(x_1, \ldots, x_m; y_1, \ldots, y_m) = \\ = W(x_1, \ldots, x_m) W(y_1, \ldots, y_m | x_1, \ldots, x_m),$$

где $W(y_1, \ldots, y_m | x_1, \ldots, x_m)$ — условная функция распределения координат $y_1, \ldots, y_m$ при известных координатах $x_1, \ldots, x_m$. Эта функция определяется свойствами преобразователя сообщения $x_T(t)$ в воспроизведение $y_T(t)$.

Функция распределения координат воспроизведения может быть определена из соотношения

$$W(y_1, \ldots, y_m) = \int\limits_{L_{x_1}} \cdots \int\limits_{L_{x_m}} W(x_1, \ldots, x_m; y_1, \ldots, y_m) dx_1 \ldots dx_m. \qquad (10.112)$$

Равенство (10.111) приводится также к виду

$$I(Y_T, X_T) = H(X_T) - H(X_T | Y_T). \qquad (10.113)$$

При этом энтропия $H(X_T)$ определяется по (10.86), а условная энтропия $H(X/Y_T)$ — по формуле, аналогичной (10.84), с той лишь разницей, что вместо одномерных должны быть записаны многомерные распределения, а вместо однократных интегралов — многократные. Так как основные свойства энтропии для многомерного случая остаются такими же, как и для одномерного, то, кроме (10.113), количество информации может определяться из соотношения

$$I(Y_T, X_T) = H(Y_T) - H(Y_T \mid X_T). \qquad (10.114)$$

Таким образом, в многомерном случае, как и в одномерном, количество информации может быть подсчитано при условии, что известна функция распределения ансамбля сообщений $W(x_1, \ldots, x_m)$ и условная функция распределения $W(y_1, \ldots, y_m \mid x_1, \ldots, x_m)$.

Задаваясь последней, можно оценить по (10.108) погрешность преобразования g_n и количество информации $I(Y_T, X_T)$.

Так же как и для одномерных случайных величин, ε-энтропию источника непрерывных сообщений определим из условия

$$H_\varepsilon(X_T) = \inf_{g_n < \eta} [I(Y_T, X_T)]. \qquad (10.115)$$

При этом нижняя грань отыскивается по всем возможным условным функциям распределения $W(y_1, \ldots, y_m \mid x_1, \ldots, x_m)$.

ε-энтропия в рассматриваемом случае имеет такое же толкование, как и в одномерном.

Модели сигналов. При выборе наиболее приемлемой системы координат используют тот или иной способ представления функций $x_T(t)$ или, как говорят, используют различные модели сигналов.

Выбор модели соответствует замене функции $x(t)$ функцией, для которой погрешность системы координат пренебрежимо мала. Кроме того, для облегчения расчетов желательно, чтобы используемые координаты были независимыми.

Рассмотрим две модели сигналов.

а) Модель, в которой принимается, что спектр функции $x(t)$ ограничен и максимальная частота в спектре равна F_m. Как известно, в этом случае (см. гл. II), если период опроса выбрать равным $T_0 = \frac{1}{2F_m}$, бесконечный ряд Котельникова точно описывает функции $x(t)$, т. е.

$$x(t) = \sum_{l=-\infty}^{\infty} x\left(\frac{l}{2F_m}\right) Sa\, 2\pi F_m \left(t - \frac{l}{2F_m}\right). \qquad (10.116)$$

Существенной особенностью функции с ограниченным спектром является то, что она отлична от нуля на всей временной оси $(-\infty < t < \infty)$, за исключением, быть может, счетного бесконечного множества моментов времени. Более того, если бы можно было точно определить значение этой функции на некотором конечном

интервале времени, то теоретически можно было бы рассчитать все ее координаты $x\left(\frac{l}{2F_m}\right)$ и, следовательно, предсказать значения функции $x(t)$ на все последующие моменты времени. В этом смысле говорят, что функция с ограниченным спектром детерминирована.

Реальные сигналы (функции $x(t)$), как правило, имеют конечную длительность и не детерминированы, ибо в противном случае они не несли бы никакой информации (все можно было бы предсказать наперед). В действительности, однако, мы никогда не можем определить значения $x(t)$ со сколь угодно высокой точностью. Такое определение возможно лишь с некоторой погрешностью. При наличии такой погрешности свойства детерминированности функции $x(t)$ теряются.

Кроме того, функция с ограниченным спектром вне некоторого интервала времени T может иметь величину, много меньшую уровня шумов или разрешающей способности (чувствительности) измерительных устройств. Поэтому всякую реальную функцию $x(t)$, ограниченную конечным интервалом времени, мы можем продолжить в сторону больших и меньших значений времени и с некоторой погрешностью аппроксимировать функцией с ограниченным спектром. Эти соображения делают возможным использование функции с ограниченным спектром в качестве модели сигналов.

Если принять, что вне некоторого интервала T значения функции с ограниченным спектром пренебрежимо малы, то из (10.116) следует, что функция $x_T(t)$ на этом интервале определяется значением

$$m = \frac{T}{T_0} + 1 \approx 2F_m T \qquad (10.117)$$

координат. Указанные отличия сигналов с ограниченным спектром от реальных приводят к ряду существенных затруднений при попытке построения теории, использующей эти сигналы в качестве модели для реальных сигналов. Так, в частности, весьма трудным оказывается выбор величины F_m, от которой, согласно (10.117), зависит число координат, описывающих сигнал.

б) Модель, в которой считается, что интервал корреляции в сигналах ограничен. Эта модель была предложена Н. А. Железновым*). Здесь сигнал рассматривается как нестационарный стохастический процесс с конечной длительностью T. Энергетический спектр таких сигналов сплошной и отличен от нуля на всех частотах, за исключением, быть может, счетного бесконечного множества частот. Интервал корреляции τ_0 принимается ограниченным, причем $\tau_{0\,max} \ll T$.

Следует отметить, что и эта модель также является некоторой идеализацией, ибо в обычных физических системах с накопителями

*) Н. А. Железнов, Некоторые вопросы теории информационных электрических систем, Изд. ЛКВВИА им. А. Ф. Можайского, Ленинград, 1960.

энергии (например, в фильтрах) переходные процессы теоретически продолжаются бесконечно долго и, следовательно, ограничивая интервал корреляции, мы тем самым отбрасываем влияние предшествовавшей части сигнала, удаленной от рассматриваемой на время более $\tau_{0\,max}$. Достоинством модели Н. А. Железнова является то, что она лишена внутренних противоречий, удовлетворяет требованиям к носителю информации и полнее отражает реальную картину.

Из нестационарных сигналов может быть выделен большой класс квазистационарных, которые удовлетворяют условию стационарности на интервале T, за исключением малой области у краев, и равны нулю вне этого интервала. Основные выводы теории стационарных сигналов с ограниченным спектром остаются справедливыми и для квазистационарных сигналов.

§ 6. Непрерывные каналы с шумами

Общие сведения о непрерывных каналах передачи информации. Непрерывные каналы обеспечивают передачу непрерывных сообщений. Функциональная схема непрерывных каналов сходна со схемой, показанной на рис. 10.5 для дискретных каналов, и отличается от последней тем, что в ней вместо кодирующих и декодирующих устройств может использоваться более широкий класс различных преобразователей. Во всех случаях непрерывное сообщение $x(t)$ преобразуется вначале в электрический сигнал $y(t)$, который затем уже, после ряда преобразований, передается на расстояние, так что обычно сигнал $z(t)$, получаемый на приемной стороне, отличается от сигнала $y(t)$ лишь наличием некоторых искажений, обусловленных действием шумов и несовершенством аппаратуры.

Для обеспечения передачи сигналов на расстояние используются различные методы преобразования и модуляции. Так, например, в случае использования радиоканала связи применяют:

а) непосредственную модуляцию одного из параметров несущих колебаний (амплитуды, частоты или фазы) по закону $y(t)$;

б) предварительную модуляцию поднесущих колебаний (гармонические или импульсные поднесущие колебания), которые в свою очередь определяют закон модуляции несущих колебаний (см. гл. II);

в) квантование сигналов $y(t)$ по дискретным уровням и передачу в точках опроса номеров ближайших квантованных уровней.

Если в случае а) имеет место как бы непрерывная передача значения $y(t)$, то в случае б) при использовании импульсных поднесущих колебаний производится передача значения $y(t)$ лишь в определенные моменты времени (точки опроса), соответствующие моментам формирования импульсов. В случае в) для передачи номеров ближайших квантованных уровней в точках опроса используют те же методы, что и в дискретных каналах.

Во всех случаях аппаратурой на приемной стороне производятся соответствующие обратные преобразования, в результате чего формируется сигнал $z(t)$, несущий информацию о сигнале $y(t)$, а следовательно, и о сообщении $x(t)$.

Пропускная способность непрерывных каналов. Независимо от конкретного характера преобразований сигналов, происходящих в непрерывном канале связи, мы можем рассматривать этот канал как некоторый преобразователь (рис. 10.14), устанавливающий соответствие между сигналами на выходе $z(t)$ и на входе $y(t)$. В результате воздействия шумов (помех) и других случайных факторов соответствие между $z(t)$ и $y(t)$ носит вероятностный характер. Эти

Рис. 10.14.

соображения позволяют для вычисления количества информации, содержащегося в случайном сигнале Z_T, о случайном сигнале Y_T применить соотношения, рассмотренные в § 5. Так, в частности, для вычисления $I(Z_T, Y_T)$ можно использовать (10.114), тогда

$$I(Z_T, Y_T) = H(Z_T) - H(Z_T \mid Y_T). \tag{10.118}$$

Для непрерывных каналов с шумами также может быть использовано понятие о пропускной способности согласно (10.41). Как и для дискретных каналов понятие о пропускной способности непрерывных каналов приобретает смысл, если ансамбль передаваемых непрерывных сообщений обладает эргодическими свойствами.

Рассмотрим случай, когда входные сигналы $y(t)$ и выходные сигналы $z(t)$ являются стационарными, эргодическими и стационарно связанными функциями времени. Возьмем отрезки этих функций на временном интервале T, полагая, что вне этого интервала $y(t)$ и $z(t)$ невелики, так что функции $y_T(t)$ и $z_T(t)$ с малой погрешностью определяются m координатами, в качестве которых служат их значения y_i и z_i в точках опроса. Будем полагать также, что значения каждой функции в точках опроса независимы (не коррелированы).

Определим $H(Z_T)$ и $H(Z_T \mid Y_T)$. Согласно (10.86)

$$H(Z_T) = -\int\limits_{L_{z_1}} \dots \int\limits_{L_{z_m}} W(z_1, \dots, z_m) \log W(z_1, \dots, z_m)\, dz_1 \dots dz_m. \tag{10.119}$$

Так как $z_1, z_2, \dots, z_m$ — независимые случайные величины, то

$$W(z_1, z_2, \dots, z_m) = W(z_1)\, W(z_2) \dots W(z_m).$$

Выполняя интегрирование, находим

$$H(\mathbf{Z}_T)=\sum_{l=1}^{m} H(Z_l), \tag{10.120}$$

где

$$H(Z_l)=-\int\limits_{L_{z_l}} W(z_l)\log W(z_l)\,dz_l \tag{10.121}$$

— энтропия l-го опроса воспроизведения.

Для стационарных и квазистационарных функций $W(z_l)$ для всех l одинаковы. Тогда $H(Z_1)=H(Z_2)=\ldots=H(Z_m)=H(Z)$ и (10.120) можно записать в виде

$$H(\mathbf{Z}_T)=mH(Z). \tag{10.122}$$

Используя (10.84) и учитывая многомерность $\mathbf{Z}_T$ и $\mathbf{Y}_T$, определим условную энтропию

$$H(\mathbf{Z}_T\,|\,\mathbf{Y}_T)=-\int\limits_{L_{y_1}}\ldots\int\limits_{L_{y_m}}\int\limits_{L_{z_1}}\ldots\int\limits_{L_{z_m}} W(y_1,\ldots,y_m;\,z_1,\ldots,z_m)\times$$
$$\times\log W(z_1,\ldots,z_m\,|\,y_1,\ldots,y_m)\,dy_1\ldots dy_m\,dz_1\ldots dz_m.$$

Положим, что значение z_l зависит лишь от значения y_l и не зависит от y_j, если $j\neq l$; так как, кроме того, y_l и y_j — независимые случайные величины, то в таком случае

$$W(y_1,\ldots,y_m;\,z_1,\ldots,z_m)=W(y_1,z_1)\,W(y_2,z_2)\ldots W(y_m,z_m)$$

и

$$W(z_1,\ldots,z_m\,|\,y_1,\ldots,y_m)=W(z_1\,|\,y_1)\,W(z_2\,|\,y_2)\ldots W(z_m\,|\,y_m).$$

Учитывая эти соотношения и производя интегрирование, получим

$$H(\mathbf{Z}_T\,|\,\mathbf{Y}_T)=\sum_{l=1}^{m} H(Z_l\,|\,Y_l),$$

где

$$H(Z_l\,|\,Y_l)=-\int\limits_{L_{y_l}}\int\limits_{L_{z_l}} W(y_l,z_l)\log W(z_l\,|\,y_l)\,dy_l\,dz_l. \tag{10.123}$$

Для стационарно связанных процессов $W(y_l,z_l)$ и $W(z_l\,|\,y_l)$ для всех l одинаковы, тогда

$$H(Z_1\,|\,Y_1)=H(Z_2\,|\,Y_2)=\ldots=H(Z_m\,|\,Y_m)=H(Z\,|\,Y)$$

и, следовательно,

$$H(\mathbf{Z}_T\,|\,\mathbf{Y}_T)=mH(Z\,|\,Y). \tag{10.124}$$

Из (10.118), (10.122) и (10.124) получаем

$$I(\mathbf{Z}_T\,|\,\mathbf{Y}_T)=mH(Z)-mH(Z\,|\,Y).$$

Разделив это равенство на величину рассматриваемого интервала времени T, имеем

$$\bar{I}(\mathbf{Z}_T\,|\,\mathbf{Y}_T)=\bar{H}(Z)-\bar{H}(Z\,|\,Y). \tag{10.125}$$

При этом

$$\bar{H}(Z)=F_0H(Z),$$
$$\bar{H}(Z\,|\,Y)=F_0H(Z\,|\,Y),$$

где $F_0=\frac{m}{T}=\frac{1}{T_0}$ — частота опроса.

Соотношение (10.125) определяет поток информации, получаемый на выходе канала связи, при сделанных допущениях о характере передаваемых сигналов и выборе частоты опроса.

Из (10.123) и (10.121) следует, что этот поток зависит от характеристик шумов и других случайных факторов, действующих в канале и определяющих функцию $W(z_l\,|\,y_l)$, и от статистики передаваемых сигналов ($W(y)$), поскольку

$$W(z_l)=\int\limits_{L_{y_l}} W(z_l\,|\,y_l)\,W(y_l)\,dy_l.$$

Если $W(z_l\,|\,y_l)$ определяется свойствами канала, то, варьируя $W(y)$, можно найти такую функцию распределения входных сигналов, при которой поток получаемой информации будет наибольшим. Эти соображения позволяют написать выражение для пропускной способности непрерывного канала связи в виде соотношения

$$C_c=\operatorname{Sup}\,[\bar{H}(Z)-\bar{H}(Z\,|\,Y)], \tag{10.126}$$

которое совпадает с (10.66). Общие свойства функций, определяющих энтропию, позволяют заключить, что для непрерывных каналов справедливо также и (10.67).

Все сказанное ранее об учете фиксированных ограничений при отыскании верхней грани по (10.66) и (10.67) относится и к непрерывному каналу.

Пример определения пропускной способности непрерывного канала. Рассмотрим случай, когда все случайные искажения выходных сигналов могут рассматриваться как аддитивный шум $n(t)$, т. е.

$$z(t)=y(t)+n(t). \tag{10.127}$$

Допустим также, что этот шум — стационарный, имеет нормальный закон распределения в полосе частот $0\div F_m$, средняя мощность его равна $P_N=\sigma_N^2$.

Определим пропускную способность канала связи, полагая, что полоса пропускания его ограничена пределами 0 и F_m, а средняя мощность полезного сигнала также ограничена и равна P_S.

Для заданных условий, когда полоса пропускания канала связи ограничена, спектр функций $n(t)$, $y(t)$ и $z(t)$ также ограничен частотой F_m. При представлении таких функций рядом В. А. Котельникова T_0 необходимо выбрать равным $T_0 = \frac{1}{2F_m}$; тогда, согласно (10.125),

$$\bar{I} = 2F_m[H(Z) - H(Z \mid Y)]. \tag{10.128}$$

При таком выборе периода опроса значения $z(t)$ в точках опроса полностью определяют сигнал с ограниченным спектром и, следовательно, несут всю информацию о нем. Если учесть, что реальный сигнал имеет неограниченный спектр, то значения сигнала в точках опроса описывают его с верностью, определяемой отбрасыванием несущественной части спектра реального сигнала.

Из (10.127) следует, что

$$W(z_i \mid y_i) = W_N(n_i) = W_N(z_i - y_i), \tag{10.129}$$

где $W_N(n_i)$ — функция распределения шума.

Функцию совместного распределения $W(y_i, z_i)$ представим в виде

$$W(y_i, z_i) = W(y_i)\,W(z_i \mid y_i) = W(y_i)\,W_N(z_i - y_i). \tag{10.130}$$

Подставляя в (10.123) значения функций распределения из (10.129) и (10.130) и производя замену переменных, получим

$$H(Z \mid Y) = H(N), \tag{10.131}$$

где $H(N)$ — энтропия одного опроса шума, равная

$$H(N) = -\int\limits_{L_{n_i}} W_N(n_i) \log W_N(n_i)\, dn_i. \tag{10.132}$$

Для определения пропускной способности используем (10.126) и учтем (10.128) и (10.131); тогда

$$C_c = \operatorname{Sup}\,[2F_m[H(Z) - H(N)]].$$

В данном случае $H(N)$ будет энтропией нормального распределения, для вычисления которой можно использовать (10.92), т. е.

$$H(N) = \log \sigma_N \sqrt{2\pi e}\,,$$

следовательно,

$$C_c = \operatorname{Sup}\,\{2F_m[H(Z) - \log \sigma_N \sqrt{2\pi e}]\}. \tag{10.133}$$

Если F_m и σ_N заданы, то выражение, стоящее в квадратных скобках, будет наибольшим, если энтропия выходных сигналов $H(Z)$ будет максимальна. Так как средние мощности шума $n(t)$ и полезных сигналов $y(t)$ ограничены, то и средняя мощность выходных сигналов $z(t)$ также ограничена. При таком условии $H(Z)$ будет

наибольшей, если Z_i имеет нормальный закон распределения. Так как N_i имеет нормальный закон распределения, то, следовательно, и Y_i (см. (10.127)) должно иметь нормальный закон распределения. Последнее означает, что характер полезных сигналов в данном случае должен быть таким же, как и характер шума. Если Y_i и N_i независимы, то $\sigma_Z^2 = \sigma_Y^2 + \sigma_N^2$ и, следовательно,

$$H(Z) = \log \sqrt{(\sigma_Y^2 + \sigma_N^2)\,2\pi e}\,.$$

Подставляя это значение $H(Z)$ в (10.133) и учитывая, что $\sigma_Y^2 = P_S$, и $\sigma_N^2 = P_N$, получим

$$C_c = F_m \log\left(1 + \frac{P_S}{P_N}\right). \tag{10.134а}$$

При выводе (10.134а) мы определяли информацию, содержащуюся в сигнале на выходе канала связи относительно сигналов на его входе. Однако такие же рассуждения могут быть проведены и для радиосигналов, т. е. мы можем определять количество информации, содержащейся в радиосигнале, действующем на входе радиоприемного устройства, о сигналах, действующих на выходе радиопередающего устройства. Можно показать (см. С. Голдман), что и радиосигнал, спектр которого ограничен полосой Δf_c, а длительность равна T, полностью определяется $2T\,\Delta f_c$ координатами.

В рамках тех же допущений, какие были сделаны при выводе (10.134а), для радиотракта канала связи получим соотношение, идентичное (10.134а),

$$C_c = \Delta f_c \log\left(1 + \frac{P_{S\,\text{вх}}}{P_{N\,\text{вх}}}\right), \tag{10.134б}$$

где $P_{S\,\text{вх}}$ и $P_{N\,\text{вх}}$ — соответственно средние мощности полезного сигнала и шума на входе радиоприемного устройства.

При использовании формул (10.134а) и (10.134б) нужно всегда иметь в виду те ограничения, которые накладывались на канал при их выводе.

Если при анализе конкретной радиолинии вычисление по (10.134а) дает величину C_c меньшую, чем по (10.134б), то это означает, что обработка сигнала в приемном устройстве ведется не оптимальным способом.

Основная теорема Шэннона для непрерывных каналов. Для непрерывных каналов справедлива следующая теорема, аналогичная теоремам для дискретных каналов.

Т е о р е м а 4. Если ε-энтропия $\overline{H}_\varepsilon(X)$ источника непрерывных сообщений, определяющая количество информации, вырабатываемое источником в единицу времени *), при заданной оценке g верности

*) Поток информации непрерывного источника.

воспроизведения, равна $\overline{H}_\varepsilon(X) = C_\varepsilon - \alpha$, где α как угодно мало, то существует метод передачи, при котором все сообщения, вырабатываемые источником, могут быть переданы, а верность воспроизведения при этом как угодно близка к g.

Обратное утверждение этой теоремы говорит о том, что такая передача невозможна, если

$$\overline{H}_\varepsilon(X) > C_\varepsilon.$$

Эта теорема доказывается так же, как и теорема 3, с помощью предельного перехода от дискретного случая к непрерывному.

В непрерывном случае, так же как и в дискретном, для того, чтобы реальная скорость передачи информации была близка к пропускной способности канала связи, необходимо выполнить ряд преобразований сигналов с целью согласования вероятностных характеристик входных сигналов $x(t)$ с оптимальными вероятностными характеристиками сигналов $y(t)$, соответствующих наибольшей скорости передачи информации при заданной верности воспроизведения. В общем случае, чем больше должна быть верность воспроизведения при скорости передачи информации, близкой к C_ε, тем все более длинные сигналы $x_T(t)$ должны кодироваться и декодироваться.

Теорема 4 позволяет находить предельно достижимую эффективность непрерывных каналов и с этой точки зрения оценивать методы передачи, используемые в реальных каналах.

Следует иметь в виду, что вычисления C_ε и ε-энтропии для заданных условий работы канала часто весьма затруднительны.

В заключение отметим, что в настоящее время продолжается весьма интенсивная работа многих ученых по исследованию различных проблем теории информации и по совершенствованию систем обработки и передачи информации в свете тех возможностей, которые открываются теорией информации.

Большое значение теория информации имеет также в раскрытии механизма информационных обменов в живой природе и в связи с этим для решения многих важнейших задач кибернетики.

§ 7. Примеры

1. Колода состоит из 32 карт от семерок и старше. Игрок 1 вытаскивает любую карту. Игрок 2 должен угадать, какая карта вытащена, задавая вопросы, на которые игрок 1 дает ответ «да» или «нет». Определить минимальное число вопросов, которое гарантирует отгадывание вытащенной карты.

Решение. Число разных карт $n = 32$. Так как любая карта может быть вытащена с равной вероятностью, то согласно (10.7) энтропия системы

$$H(X) = \log 32 = 5\,\frac{\text{дв. ед.}}{\text{событ.}}$$

Такое же число двоичных единиц должно содержать сообщение об этом событии, т. е.

$$I(Y, X) = H(X) = 5\,\frac{\text{дв. ед.}}{\text{сообщ.}}$$

Ответ «да» или «нет» может содержать одну двоичную единицу информации. Следовательно, достаточно в среднем задать пять вопросов, ответы на которые должны позволить отгадать карту.

В общем случае энтропия дает среднее количество информации, приходящееся на один опыт при производстве большого числа опытов. Однако в данном случае, когда события равновероятны, среднее количество информации, приходящееся на сообщение, равно количеству информации, содержащемуся в единичном сообщении, которое определяется по (10.4). Поэтому может быть найдена система пяти вопросов, которая при каждом единичном опыте гарантирует угадывание. Очевидно, что система вопросов должна быть такова, чтобы ответы «да» и «нет» были равновероятны, ибо только в этом случае они всякий раз будут нести 1 дв. ед. информации. Нетрудно сообразить, какова может быть такая система вопросов.

Допустим, например, вытащена пиковая девятка. Тогда

Вопрос 1: Масть красная?

Ответ: Нет.

(Вывод: вытащена трефа или пика.)

Вопрос 2: Масть трефа?

Ответ: Нет.

(Вывод: вытащена пика.)

Вопрос 3: Старше десятки?

Ответ: Нет.

(Вывод: вытащена семерка или восьмерка, или девятка, или десятка.)

Вопрос 4 Старше восьмерки?

Ответ: Да.

(Вывод: вытащена девятка или десятка.)

Вопрос 5: Девятка?

Ответ: Да.

(Вывод: пиковая девятка.)

2. На контролируемом объекте имеются три клапана, каждый из которых может находиться в одном из двух положений («закрыт» или «открыт»). С объекта передаются сообщения об изменении положения клапанов. Наблюдением в течение длительного отрезка времени установлено, что из 100 переданных сообщений 70 относится к 1-му клапану, 20 — ко 2-му и 10 — к 3-му. Клапаны работают независимо друг от друга.

Определить количество информации, содержащееся в одном сообщении.

Решение. Вычисление может быть произведено по (10.5).

Обозначим:

x_{13} — сообщение о том, что 1-й клапан закрылся;

x_{10} — сообщение о том, что 1-й клапан открылся.

Аналогично для 2-го и 3-го клапанов введем обозначения x_{23}, x_{20}, x_{33}, x_{30}. Для каждого клапана число сообщений об его открытии и закрытии будет одинаковым и, следовательно, вероятности этих сообщений равны между собой, т. е. $p(x_{13}) = p(x_{10})$, $p(x_{23}) = p(x_{20})$ и $p(x_{33}) = p(x_{30})$.

Используя статистику наблюдений, непосредственно получим

	x_{13}	x_{10}	x_{23}	x_{20}	x_{33}	x_{30}
$p(x)$	0,35	0,35	0,1	0,1	0,05	0,05

Применяя (10.5), имеем

$$H(X) = -2\cdot 0{,}35 \log 0{,}35 - 2\cdot 0{,}1 \log 0{,}1 - 2\cdot 0{,}05 \log 0{,}05$$

или

$$H(X) = 2{,}156\ \frac{\text{дв. ед.}}{\text{сообщ.}}$$

Примечание. Для того чтобы упростить расчеты, соотношение (10.5) представим в виде

$$H(X) = \sum_{i=1}^{n} \eta(p_i), \tag{1п}$$

где $\eta(p) = -p \log_2 p$ — функция, таблица 1п которой приведена ниже *).

Таблица 1п

Значения функции $\eta(p) = -p \log_2 p$

p	$\eta(p)$	Δ	p	$\eta(p)$	Δ	p	$\eta(p)$	Δ	p	$\eta(p)$	Δ
0	0	664	0,25	0,5000	53	0,50	0,5000	—46	0,75	0,3113	—104
0,01	0,0664	464	0,26	5053	47	0,51	4954	—48	0,76	3009	—106
0,02	1128	390	0,27	5100	42	0,52	4906	—52	0,77	2903	—107
0,03	1518	340	0,28	5142	37	0,53	4854	—54	0,78	2796	—109
0,04	1858	303	0,29	5179	32	0,54	4800	—56	0,79	2687	—112
0,05	2161	274	0,30	5211	27	0,55	4744	—59	0,80	2575	—113
0,06	2435	251	0,31	5238	22	0,56	4685	—62	0,81	2462	—114
0,07	2686	229	0,32	5260	18	0,57	4623	—65	0,82	2348	—117
0,08	2915	211	0,33	5278	14	0,58	4558	—67	0,83	2231	—119
0,09	3126	196	0,34	5292	9	0,59	4491	—69	0,84	2112	—120
0,10	3322	181	0,35	5301	5	0,60	4422	—72	0,85	1992	—121
0,11	3503	168	0,36	5306	1	0,61	4350	—74	0,86	1871	—123
0,12	3671	155	0,37	5307	—2	0,62	4276	—77	0,87	1748	—125
0,13	3826	145	0,38	5305	—7	0,63	4199	—78	0,88	1623	—127
0,14	3971	134	0,39	5298	—10	0,64	4121	—81	0,89	1496	—128
0,15	4105	125	0,40	5288	—14	0,65	4040	—83	0,90	1368	—130
0,16	4230	116	0,41	5274	—18	0,66	3957	—86	0,91	1238	—131
0,17	4346	107	0,42	5256	—20	0,67	3871	—87	0,92	0,1107	—134
0,18	4453	99	0,43	5236	—24	0,68	3784	—90	0,93	0,0974	—135
0,19	4552	92	0,44	5210	—26	0,69	3694	—92	0,94	839	—136
0,20	4644	84	0,45	5184	—29	0,70	3602	—94	0,95	703	—138
0,21	4728	78	0,46	5153	—33	0,71	3508	—96	0,96	565	—139
0,22	4806	71	0,47	5120	—37	0,72	3412	—98	0,97	426	—140
0,23	4877	67	0,48	5083	—40	0,73	3314	—99	0,98	286	—142
0,24	4941	59	0,49	5043	—43	0,74	3215	—102	0,99	144	—144
0,25	0,5000		0,50	0,5000		0,75	0,3113		1,00	0	

3. В объекте, аналогичном описанному в п. 2, производится периодический контроль состояния клапанов. Количество сообщений, передаваемых о состоянии каждого клапана, одинаково. Наблюдением установлено, что в среднем клапан 1 «закрыт» 98%, клапан 2 — 80%, а клапан 3 — 0,6% всего времени. В остальное время клапаны открыты.

*) Таблица заимствована из книги: Е. С. Вентцель, Теория вероятностей, Физматгиз, 1962.

Определить количество информации, содержащейся в одном сообщении.

Решение. Обозначим через x_1 событие, состоящее в том, что передается извещение о состоянии клапана 1, т. е. извещение клапан 1 «закрыт» — $x_{1з}$ или «открыт» — $x_{2о}$. Аналогично для клапана 2 рассматриваем события x_2 и $x_{2з}$, $x_{2о}$, а для клапана 3 — x_3 и $x_{3з}$, $x_{3о}$.

Так как все клапаны контролируются одинаково часто, то вероятности событий x_1, x_2 и x_3 равны, т. е.

$$p(x_1) = p(x_2) = p(x_3) = \frac{1}{3}.$$

Состояния «закрыт» и «открыт» несовместны и, следовательно, $p(x_1) = p(x_{1з}) + p(x_{1о})$, а также $p(x_2) = p(x_{2з}) + p(x_{2о})$ и $p(x_3) = p(x_{3з}) + p(x_{3о})$.

Из статистики наблюдений следует, что для клапана 1

$$p(x_{1з}) = 0{,}98\,p(x_1) = \frac{0{,}98}{3} \approx 0{,}326,$$

тогда

$$p(x_{1о}) = p(x_1) - p(x_{1з}) = \frac{1}{3} - 0{,}326,$$

т. е.

$$p(x_{1о}) \approx 0{,}007.$$

Аналогично вычисляются вероятности передачи других сообщений. Результаты расчета вероятностей сообщений и функций $\eta(p) = -p \log p$ сведем в таблицу.

x	$x_{1з}$	$x_{1о}$	$x_{2з}$	$x_{2о}$	$x_{3з}$	$x_{3о}$
$p(x)$	0,326	0,007	0,269	0,065	0,002	0,331
$\eta(p)$	0,527	0,05	0,509	0,256	0,0179	0,528

В итоге получаем

$$H(X) = \sum \eta(p) = 1{,}888\ \frac{\text{дв. ед.}}{\text{сообщ.}}$$

Обратим внимание на то, что сообщения, которые передаются редко, $x_{1о}$ и $x_{3з}$, вносят очень малый вклад в значение $H(X)$.

4. Определить избыточность источников информации для условий задач пп. 2 и 3.

Решение. Расчет может быть произведен по (10.30).

В обоих случаях число различных сообщений $n = 6$ и, следовательно, по (10.7)

$$H_{max} = \log n = \log 6 = 2{,}585.$$

Тогда для условия п. 1

$$R = \frac{2{,}585 - 2{,}156}{2{,}585} \approx 0{,}167,$$

а для условия п. 2

$$R = \frac{2{,}585 - 1{,}888}{2{,}585} \approx 0{,}272.$$

5. Ансамбли событий X и Y объединены. Вероятности совместных событий (x, y) приведены в следующей таблице:

$p(x_i y_j)$

y_j \ x_i	x_1	x_2	x_3
y_1	0,1	0,2	0,3
y_2	0,25	0	0,15

Определить:
а) энтропию ансамблей X и Y,
б) энтропию объединенного ансамбля $X \otimes Y$,
в) условные энтропии ансамблей.

Решение. Используя (10.24), определим вероятности событий в ансамблях X и Y

$$p(x_1) = p(x_1, y_1) + p(x_1, y_2) = 0{,}1 + 0{,}25 = 0{,}35.$$

Аналогично получим

$$p(x_2) = 0{,}2; \quad p(x_3) = 0{,}3 + 0{,}15 = 0{,}45;$$
$$p(y_1) = 0{,}1 + 0{,}2 + 0{,}3 = 0{,}6; \quad p(y_2) = 0{,}25 + 0{,}15 = 0{,}4.$$

Зная вероятности событий по (10.5), получим

$$H(X) = -0{,}35 \log 0{,}35 - 0{,}2 \log 0{,}2 - 0{,}45 \log 0{,}45 = 1{,}512 \ \frac{\text{дв. ед.}}{\text{сообщ.}}.$$

Аналогично

$$H(Y) = -0{,}6 \log 0{,}6 - 0{,}4 \log 0{,}4 = 0{,}971 \ \frac{\text{дв. ед.}}{\text{сообщ.}}.$$

Для определения энтропии объединенного ансамбля используем (10.9); тогда для заданных значений $p(x_i, y_j)$ имеем

$$H(X, Y) = -0{,}1 \log 0{,}1 - 0{,}25 \log 0{,}25 - 0{,}2 \log 0{,}2 - 0{,}3 \log 0{,}3 - 0{,}15 \log 0{,}15$$

или

$$H(X, Y) = 2{,}228 \ \frac{\text{дв. ед.}}{\text{сообщ.}}.$$

Условные энтропии $H(X \mid Y)$ и $H(Y \mid X)$ могут вычисляться различными способами, для заданных условий удобнее воспользоваться (10.13) и (10.16).

Применяя (10.16), получаем

$$H(X \mid Y) = H(X, Y) - H(Y) = 2{,}228 - 0{,}971 = 1{,}257 \ \frac{\text{дв. ед.}}{\text{сообщ.}},$$

а из (10.13)

$$H(Y \mid X) = H(X, Y) - H(X) = 2{,}228 - 1{,}512 = 0{,}716 \ \frac{\text{дв. ед.}}{\text{сообщ.}}.$$

6. Для условий задачи п. 5 определить количество информации, содержащееся в событиях y относительно событий x.

Решение. Применим (10.19) или (10.23).

Используя результаты п. 5 и (10.19), имеем

$$I(Y, X) = H(X) - H(X \mid Y) = 1{,}512 - 1{,}257 = 0{,}255 \ \frac{\text{дв. ед.}}{\text{сообщ.}}.$$

Как видно, $I(Y, X) \ll H(X)$, что говорит о сравнительно слабой связи между событиями y и x. Последнее проявляется также в том, что энтропия объединенного ансамбля $H(X, Y) = 2{,}226 \ \frac{\text{дв. ед.}}{\text{сообщ.}}$ не сильно отличается от суммы энтропий ансамблей

$$H(X) + H(Y) = 1{,}512 + 0{,}971 = 2{,}483 \ \frac{\text{дв. ед.}}{\text{сообщ.}}.$$

7. В информационном канале используется алфавит с четырьмя различными символами. Длительности всех символов одинаковы и равны $\tau = 1$ *мсек*. Фиксированные ограничения на допустимую последовательность символов не накладываются (канал типа А).

Определить пропускную способность канала при отсутствии шумов.

Решение. Для расчета используем (10.47). В данном случае $a = 4$ и $C_t = \dfrac{\log 4}{10^{-3}} = 2000 \ \dfrac{\text{дв. ед.}}{\text{сообщ.}}$.

8. В информационном канале используется сменно-качественный код, при котором запрещается передача подряд двух одинаковых символов. Алфавит кода состоит из четырех различных символов. Вероятности передачи всех разрешенных пар символов одинаковы. Длительности всех символов также одинаковы и равны $\tau = 1$ *мсек* (канал типа Б).

Определить скорость передачи информации.

Решение. Ответ на поставленный вопрос дает (10.49а).

Обозначим символы кода y_1, y_2, y_3 и y_4. Из характера наложенного ограничения следует, что корреляционные связи имеют место лишь между двумя символами и, следовательно, энтропия сигнала может быть определена по (10.29). Обозначим пары поочередно передаваемых символов (y_i, y_j). Так как комбинации (y_i, y_j) при $i = j$ запрещены, то вероятности таких пар $p(y_i, y_j) = 0$. Всего разрешенных пар будет $3 \times 4 = 12$. Согласно условию, вероятности передачи равны тогда

$$p(y_i, y_j) = \frac{1}{12} \quad \text{при} \quad i \neq j.$$

Для большей наглядности все возможные значения $p(y_i, y_j)$ сведем в следующую таблицу.

$p(y_i, y_j)$

y_j \ y_i	y_1	y_2	y_3	y_4
y_1	0	$^1/_{12}$	$^1/_{12}$	$^1/_{12}$
y_2	$^1/_{12}$	0	$^1/_{12}$	$^1/_{12}$
y_3	$^1/_{12}$	$^1/_{12}$	0	$^1/_{12}$
y_4	$^1/_{12}$	$^1/_{12}$	$^1/_{12}$	0

Вероятность передачи любого символа может быть теперь определена по обычной формуле

$$p(y_i) = \sum_j p(y_i, y_j) = 3 \frac{1}{12} = \frac{1}{4}.$$

Как видно, вероятности передачи разных символов также равны. Определим условную вероятность $p(y_j \mid y_i)$ при $j \neq i$.

Очевидно, что

$$p(y_j \mid y_i) = \frac{p(y_i, y_j)}{p(y_i)} = \frac{^1/_{12}}{^1/_4}$$

или $p(y_j \mid y_i) = \frac{1}{3}$ для всех i и j, если только $i \neq j$. Из заданного условия и составленной таблицы также следует, что $p(y_j \mid y_i) = 0$ при $j = i$.

Используя теперь (10.29), найдем

$$H(Y) = -12 \frac{1}{12} \log \frac{1}{3} = 1{,}585 \frac{\text{дв. ед.}}{\text{символ}}.$$

Подставляя это значение $H(Y)$ в (10.49а), имеем

$$\bar{I} = \frac{H(Y)}{\tau} = \frac{1{,}585}{10^{-3}}$$

или

$$\bar{I} = 1585 \frac{\text{дв. ед.}}{\text{сек}}.$$

9. Доказать, что в условиях задачи п. 8 найденная скорость передачи информации является максимальной и равна пропускной способности канала.

Доказательство. Используем (10.28) и запишем это отношение для рассматриваемого случая в виде

$$H(Y) = -\sum_{i=1}^{4} p(y_i) \sum_{j=1}^{4} p(y_j \mid y_i) \log p(y_j \mid y_i).$$

Из свойства 2 энтропии (см. § 1) следует, что при данном y_i сумма

$$-\sum_{j=1}^{4} p(y_j \mid y_i) \log p(y_j \mid y_i)$$

имеет максимальное значение, если условные вероятности $p(y_j \mid y_i)$ всех разрешенных переходов $(j \neq i)$ равны между собой. Так как всего разрешенных переходов три, то необходимо, чтобы

$$p(y_j \mid y_i) = \frac{1}{3}.$$

Определим $p(y_1)$ из очевидного равенства

$$p(y_1) = p(y_2)\,p(y_1 \mid y_2) + p(y_3)\,p(y_1 \mid y_3) + \\ + p(y_4)\,p(y_1 \mid y_4) = \frac{1}{3}\,[p(y_2) + p(y_3) + p(y_4)]$$

или $p(y_1) = \frac{1}{3}[1 - p(y_1)]$, откуда $p(y_1) = \frac{1}{4}$; аналогичные значения получим для y_2, y_3 и y_4. Найденные значения вероятностей, при которых $H(Y)$ имеет максимальное значение, равны соответствующим величинам задачи п. 8, и, следовательно, рассмотренная скорость передачи информации максимальна и, согласно (10.49б), равна пропускной способности канала.

Найденное условие равной вероятности разрешенных переходов может быть распространено и на общий случай фиксированных ограничений для канала типа Б.

10. По информационному каналу передается шесть различных сообщений. Энтропия источника $H(X) = 1{,}888 \frac{\text{дв. ед.}}{\text{сообщ.}}$ (см. п. 3). В канале используется простой двоичный код.

Определить избыточность в передаваемом сигнале.

Решение. При простом двоичном коде число разрядов кодовой группы должно быть $m \geq \log M$, где M — число передаваемых сообщений, в данном случае $M = 6$, и тогда $m \geq \log 6 = 2{,}585$ и, следовательно, $m = 3$.

Кодовая таблица может, например, быть

x_j	x_1	x_2	x_3	x_4	x_5	x_6
Код	000	001	010	011	100	101

Если фиксированные ограничения на допустимую последовательность передачи символов не накладываются, то максимальная энтропия кодированного сигнала, отнесенная к одному символу двоичного кода,

$$H_{m\,\text{симв}} = 1 \frac{\text{дв. ед.}}{\text{символ}},$$

а для одного сообщения

$$H_{m\,\text{сообщ}} = m H_{m\,\text{симв}} = 3 \frac{\text{дв. ед.}}{\text{сообщ.}}$$

Используя (10.30), находим

$$R = \frac{3 - 1{,}888}{3} = 0{,}368.$$

Как видно, избыточность достаточно большая. При передаче фактически используется лишь $(1 - R)\,100 \approx 63\%$ пропускной способности канала.

11. В информационном канале для передачи сообщений используются три различных символа с длительностями $\tau_1 = \tau_2 = 10$ *мсек* и $\tau_3 = 20$ *мксек*. Фиксированные ограничения на допустимую последовательность передачи символов не накладываются (канал типа В).

Определить пропускную способность канала.

Решение. Для ответа на вопрос используем (10.50б).

Обозначим $\tau_1 = \tau_2 = \tau_0$, тогда $\tau_3 = 2\tau_0$.

Для определения величины r_1 составим уравнение

$$1 - r^{-\frac{\tau_1}{\tau_0}} - r^{-\frac{\tau_2}{\tau_0}} - r^{-\frac{\tau_3}{\tau_0}} = 0, \quad \text{или} \quad 1 - 2r^{-1} - r^{-2} = 0.$$

Запишем его в виде

$$\left(\frac{1}{r}\right)^2 + 2\left(\frac{1}{r}\right) - 1 = 0.$$

Корни этого уравнения будут $\frac{1}{r} = -1 \pm \sqrt{1+1}$; так как нам нужен наибольший положительный корень r_1, то

$$\frac{1}{r_1} = \sqrt{2} - 1 \approx 0{,}414$$

и

$$r_1 = \frac{1}{0{,}414} \approx 2{,}418.$$

Из (10.50б) находим

$$C_c = \frac{1}{10 \cdot 10^{-3}} \log 2{,}418 = 127 \frac{\text{дв. ед.}}{\textit{сек}}.$$

12. Определить оптимальные статистические характеристики кодированного сигнала для условия задачи п. 11.

Решение. Используя (10.51), находим вероятности передачи отдельных символов, при которых скорость передачи информации может быть равной пропускной способности канала

$$p(y_1) = r_1^{-\frac{\tau_1}{\tau_0}} = r_1^{-1} = 0{,}414,$$

$$p(y_2) = r_1^{-\frac{\tau_2}{\tau_0}} = r_1^{-1} = 0{,}414,$$

$$p(y_3) = r_1^{-\frac{\tau_3}{\tau_0}} = r_1^{-2} = 0{,}172.$$

13. Определить объем информации, содержащейся в изображении при условии, что последнее разлагается на 500 строк по 500 элементов в каждой строке. Яркость каждого элемента передается восемью квантованными уровнями. Различные градации яркости равновероятны, а яркости разных элементов не коррелированы.

Решение. Так как все элементы независимы и равноправны, то энтропия изображения

$$H(X_{\text{из}}) = MH(X_{\text{эл}}) \frac{\text{дв. ед.}}{\text{изобр.}},$$

где M — число элементов разложения,

$$M = 500 \times 500 = 250\,000,$$

$H(X_{\text{эл}})$ — энтропия одного элемента.

Так как различные градации яркости не коррелированы и равновероятны, то согласно (10.7)

$$H(X_{\text{эл}}) = \log_2 8 = 3.$$

В итоге получаем

$$H(X_{\text{из}}) = 250\,000 \times 3 = 750\,000 \frac{\text{дв. ед.}}{\text{изобр.}}.$$

14. Изображение п. 13 должно быть передано по радиолинии, на входе которой действует белый гауссовский шум с удельной мощностью $N_0 = 8 \cdot 10^{-18} \frac{\textit{вт}}{\textit{кгц}}$. Ширина полосы пропускания приемного устройства выбрана равной $\Delta f_c = 1000$ *гц*. Время передачи 1 час.

Определить минимально возможное значение мощности полезного сигнала на входе радиоприемного устройства.

Решение. Определим вначале скорость передачи информации. Так как время передачи изображения $T_{\text{из}} = 3600$ *сек*, то скорость передачи информации должна быть

$$\bar{I} = \frac{H(X_{\text{из}})}{T_{\text{из}}} = \frac{750\,000}{3600} = 208{,}1 \frac{\text{дв. ед.}}{\textit{сек}}.$$

Положим, что радиолиния идеальна и скорость передачи информации равна пропускной способности, определяемой по (10.134б), т. е. $\bar{I} = C_c$, и, следовательно,

$$208{,}1 = 1000 \log \left(1 + \frac{P_{S_{\text{вх}}}}{P_{N_{\text{вх}}}}\right),$$

откуда получаем

$$\frac{P_{S_{\text{вх}}}}{P_{N_{\text{вх}}}} \approx 0{,}155.$$

Так как

$$P_{N_{\text{вх}}} = N_0 \Delta f_c = 8 \cdot 10^{-18} \cdot 1 = 8 \cdot 10^{-18} \textit{ вт},$$

то

$$P_{S_{\text{вх}}} = P_{N_{\text{вх}}} \cdot 0{,}155 = 8 \cdot 10^{-18} \cdot 0{,}155$$

или

$$P_{S_{\text{вх}}} = 1{,}32 \cdot 10^{-18} \textit{ вт}.$$

Для реальной линии потребуется, конечно, обеспечить значительно большую мощность на входе приемного устройства. Найденное значение является нижним теоретическим пределом.

15. Изображение, переданное по радиолинии, записывается на магнитную ленту. По ширине магнитной ленты записываются 30 дв. ед. информации (10 элементов), плотность записи по длине $10 \frac{\text{дв. ед.}}{\textit{мм}}$.

Определить длину магнитной ленты для записи одного изображения.

Решение. Так как общий объем информации $H(X_{\text{из}}) = 750\,000 \frac{\text{дв. ед.}}{\text{изобр.}}$, то требуемая длина магнитной ленты для записи одного изображения будет

$$l = \frac{750\,000}{30 \cdot 10} = 2500 \textit{ мм}.$$

При расчете не учитываем необходимости записи некоторых служебных сигналов (сигналы синхронизации и др.).