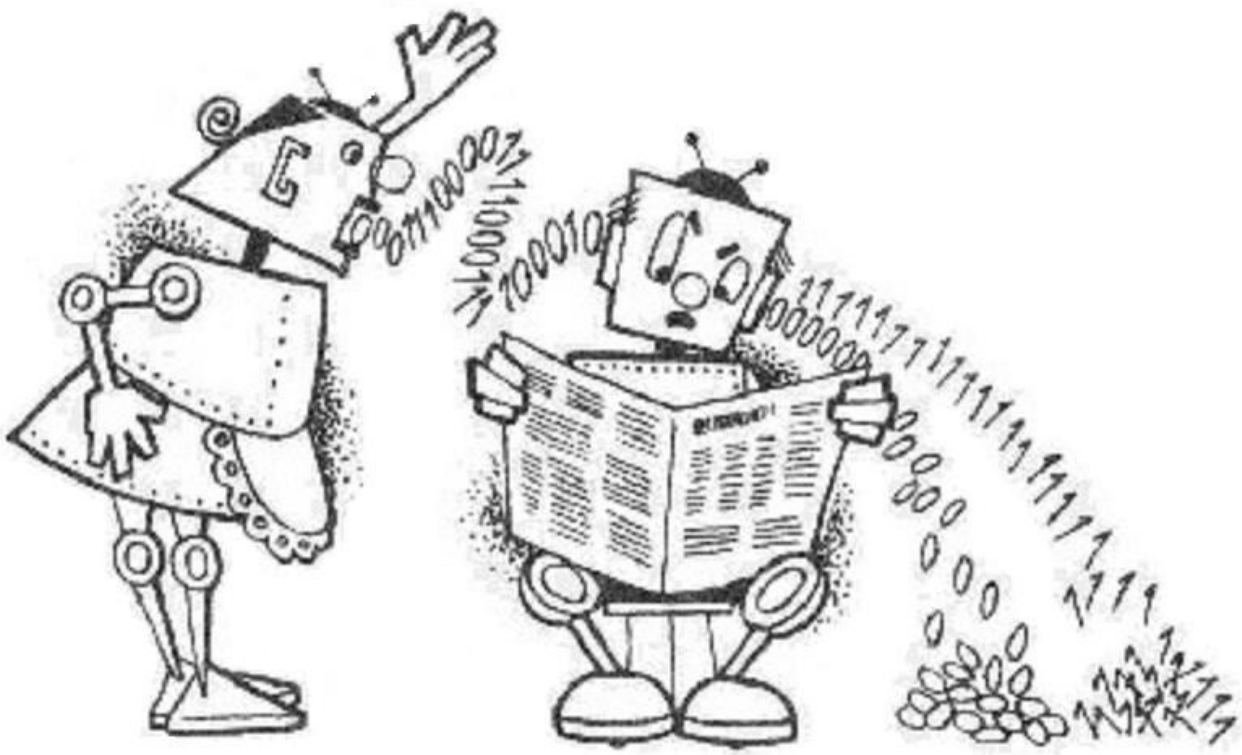
УНИВЕРСИТЕТ ИТМО

Блинова И.В., Попов И.Ю.

Теория информации



Санкт-Петербург
2018

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

УНИВЕРСИТЕТ ИТМО

Блинова И.В., Попов И.Ю.

Теория информации

РЕКОМЕНДОВАНО К ИСПОЛЬЗОВАНИЮ В УНИВЕРСИТЕТЕ ИТМО по направлению подготовки 01.03.02 в качестве учебно-методического пособия для реализации основных профессиональных образовательных программ высшего образования бакалавриата.

УНИВЕРСИТЕТ ИТМО

Санкт-Петербург

2018

Блинова И.В., Попов И.Ю. Теория информации. Учебное пособие. – СПб: Университет ИТМО, 2018. – 84 с.

Рецензенты:
д.ф.-м.н., профессор В.Д. Лукьянов, зав. учебного центра ОАО «Авангард»;
к.ф.-м.н., доцент кафедры ВМ Университета ИТМО А.И. Трифанов.

Предлагаемое пособие предназначено для студентов академического бакалавриата. В пособии разобраны следующие вопросы: энтропия как мера степени неопределенности, измерение информации, энтропия и информация для непрерывных систем, приложение теории информации к задачам передачи сообщений, передача сообщений при наличии помех, коды, обнаруживающие и исправляющие ошибки, семантическая информация.

УНИВЕРСИТЕТ ИТМО

Университет ИТМО – ведущий вуз России в области информационных и фотонных технологий, один из немногих российских вузов, получивших в 2009 году статус национального исследовательского университета. С 2013 года Университет ИТМО – участник программы повышения конкурентоспособности российских университетов среди ведущих мировых научно-образовательных центров, известной как проект «5 в 100». Цель Университета ИТМО – становление исследовательского университета мирового уровня, предпринимательского по типу, ориентированного на интернационализацию всех направлений деятельности.

© Университет ИТМО, 2018

© Блинова И.В., Попов И.Ю., 2018

Содержание

Предмет теории информации

Теория информации наука, которая изучает количественные закономерности, связанные с получением, передачей, обработкой и хранением информации.

Теория информации тесно связана с математическими науками, в частности с теорией вероятностей, которая является для нее математическим фундаментом. В свою очередь, теория информации является математическим фундаментом для теории связи. Часто теория информации рассматривается как раздел теории вероятностей или как раздел теории связи. Таким образом, предмет «Теория информации» весьма узок, т.к. зажат между чистой математикой и прикладными разделами теории связи.

Получение, передача, обработка и хранение информации является основой любой управляющей системы. Принцип работы управляющей системы заключается в том, что движение и действие больших масс или передача и преобразование больших количеств энергии направляется и контролируется при помощи небольших количеств энергии, несущих информацию.

Любая информация, для того, чтобы быть переданной должна быть соответствующим образом «закодирована», т.е. переведена на язык специальных символов и сигналов.

Одной из задач теории информации является отыскание наиболее экономных методов кодирования, позволяющих передать заданную информацию с помощью минимального количества символов. Эта задача решается как при отсутствии, так и при наличии помех в линии связи.

Другая задача теории информации: источник информации непрерывно передает информацию по линии связи приемнику. Какова должна быть пропускная способность линии связи, чтобы передать всю поступающую информацию.

Ряд задач теории информации относится к определению объема запоминающих устройств, предназначенных для хранения информации.

Для решения таких задач нужно научиться измерять количественно объем информации, пропускную способность линии связи и их чувствительность к помехам.

Итак, теория информации представляет собой математическую теорию, посвященную измерению количества информации, преобразованию информации, передаче информации по линиям связи, изучению методов построения различных кодов, и пр.

Глава 1.
Энтропия как мера степени неопределенности

1. Определение энтропии

Любое сообщение в теории информации является сведением о некоторой физической системе. Если состояние системы известно заранее, то нет смысла передавать сообщение.

В качестве объекта, о котором передается информация, будем рассматривать физическую систему X, которая случайным образом может оказаться в том или ином состоянии, т.е. систему, которой присуща какая-то степень неопределенности.

Сведения, полученные о системе, будут тем ценнее, чем больше была неопределенность системы до получения сведений. Степень неопределенности физической системы определяется числом ее возможных состояний и вероятностями состояний.

Рассмотрим систему X, которая может принимать конечное множество состояний $x_1, x_2, ...x_n$ с вероятностями $p_1, p_2, ...p_n$ $(\sum p_i = 1)$, где

$$p_i = P(X \sim x_i)$$

или

x_i	x_1	x_2	...	x_n
p_i	p_1	p_2	...	p_n

В качестве меры неопределенности системы применяется специальная характеристика, называемая *энтропией*.

Энтропией называется величина, вычисляемая по формуле:

$$H(X) = -\sum_{i=1}^{n} p_i \log_a p_i$$

Областью допустимых значений величины p является отрезок $[0, 1]$, доопределим значение в этой формуле при $p = 0$ по непрерывности значением 0: $\lim\limits_{p \to 0+0} p \log p = 0$. Основание логарифма можно взять любым $a > 1$. Выбор основания равносилен выбору единицы измерения энтропии. Если за основание выбрать число 10, то говорят о «десятичных единицах» энтропии (дитах). На практике в качестве основания удобнее использовать число 2. При выполнении вычислений будем считать $a = 2$. В этом случае

за единицу измерения энтропии принимается энтропия простейшей системы, которая имеет два равновозможных состояния, а сама энтропия измеряется в «двоичных единицах» или битах (bit).

x_i	x_1	x_2
p_i	$\frac{1}{2}$	$\frac{1}{2}$

Здесь $H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = 1$ бит.

Бит – это очень маленькая единица, поэтому часто используется величина в 8 раз большая – байт (byte). Байт обозначают латинской буквой B (1 B=8 bit). Для бита и для байта существуют производные от них единицы, образуемые при помощи приставок кило (K), мега (M), гига (G), тера (T), пета (P) и пр. Но для битов и байтов они означают не степень 10, а степень 2: кило – $2^{10} = 1024 \approx 10^3$, мега (M) $2^{20} =\approx 10^6$, гига $2^{30} \approx 10^9$, тера $2^{40} =\approx 10^{12}$, пета $2^{50} \approx 10^{15}$.

Рассмотрим систему, которая имеет n равновероятных состояний:

x_i	x_1	x_2	...	x_n
p_i	$\frac{1}{n}$	$\frac{1}{n}$	...	$\frac{1}{n}$

Здесь имеем: $H(X) = -n\frac{1}{n}\log\frac{1}{n} = \log n$. Следовательно, энтропия системы с равновозможными состояниями равна логарифму числа состояний.

2. Свойства энтропии

1. Если состояние системы в точности известно заранее, то ее энтропия равна нулю.

$\triangle$ В этом случае все вероятности $p_1, p_2, ...p_n$ в формуле для энтропии обращаются в ноль, кроме одной, которая равна 1.

Слагаемое $p_k \log p_k = 0$, т.к. $\log 1 = 0$. Остальные слагаемые обращаются в ноль, т.к.

$$\lim_{p\to 0} p\log p = 0.\square$$

2. Энтропия системы с конечным множеством состояний достигает максимума, когда все состояния равновероятны.

$\triangle$ $H(X) = -\sum_{i=1}^{n} p_i \log p_i$ – энтропия системы. Рассмотрим функцию

$$H(p_1, ...p_n) = -p_1\log p_1 - ... - p_n \log p_n.$$

Найдем условный экстремум этой функции при условии $p_1 + ... + p_n = 1$.

Чтобы найти условный экстремум функции $H(p_1, ...p_n)$, надо определить обычный экстремум фунции

$$L(p_1, ...p_n, \lambda) = -p_1\log p_1 - ... - p_n\log p_n + \lambda(p_1 + ... + p_n - 1)$$

(метод неопределенных множителей Лагранжа).

Необходимое условие экстремума: $L'_{p_i} = 0$, $(i = 1, ...n)$ и $L'_\lambda = 0$.

Дифференцируем функцию $L(p_1, ...p_n, \lambda)$ и приравниваем производные нулю. Получим систему уравнений:

$$\begin{cases} \log p_1 = \lambda - \log e \\ ... \\ \log p_n = \lambda - \log e \\ p_1 + .. + p_n = 1 \end{cases}$$

Система имеет единственное решение $p_1 = ... = p_n = \frac{1}{n}$.

Достаточное условие экстремума:

$d^2L = -\frac{1}{p_1}dp_1^2 - ... - \frac{1}{p_n}dp_n^2 < 0$. Значит функция $H(p_1, ...p_n)$ имеет условный максимум в точке $(\frac{1}{n}, ...\frac{1}{n})$.

Максимальная энтропия системы $H_{max}(X) = \log n$.

Следовательно, максимальное значение энтропии системы с конечным множеством состояний равно логарифму числа состояний и достигается, когда все состояния равновероятны. □

Введем специальную функцию:

$$\eta(p) = -p \log p$$

Тогда формула для энтропии примет вид:

$$H(X) = \sum_{i=1}^{n} \eta(p_i)$$

Представим формулу для энтропии в виде математического ожидания. Рассмотрим $\log P(X)$ как случайную величину. Когда система X принимает значения $x_1, ..., x_n$, случайная величина $\log P(X)$ принимает значения $\log p(x_1), ... \log p(x_n)$ с вероятностями $p(x_1), ...p(x_n)$:

$\log p(x_i)$	$\log p(x_1)$	...	$\log p(x_n)$
$p(x_i)$	$p(x_1)$	...	$p(x_n)$

Тогда, $M[-\log P(X)] = -p(x_1) \log p(x_1) - ... - p(x_n) \log p(x_n)$.

Следовательно, $H(X) = M[-\log P(X)]$.

Пример 1. Найти энтропию системы X, вероятности состояний которой заданы законом распределения:

x_i	x_1	x_2	x_3	x_4	x_5	x_6
p_i	0,2	0,3	0,1	0,05	0,15	0,2

Решение. Воспользовавшись Приложением 1, получим
$H(X) = \sum_{i=1}^{6} \eta(p_i) = 0,4644 + 0,5211 + 0,3322 + 0,2161 + 0,4105 + 0,4644 = 2,4087$ бит.

Пример 2. Определить максимально возможную энтропию технического устройства, состоящего из четырех элементов (устройство выходит из строя при отказе любого из элементов).

Решение. Рассмотрим техническое устройство как систему X, которая может находиться в $2^4 = 16$ состояниях. Энтропия системы достигает максимума, когда все состояния равновероятны и равна $H(X) = \log 16 = 4$ бит.

Пример 3. Написать функцию $H(p_1, p_2)$, определяющую энтропию системы с тремя состояниями. Чему равно наибольшее значение этой функции?

Решение. Рассмотрим систему X с тремя состояниями x_1, x_2, x_3. Вероятности состояний:

x_i	x_1	x_2	x_3
p_i	p_1	p_2	$1 - p_1 - p_2$

Энтропия такой системы равна:

$$H(p_1, p_2) = -p_1 \log p_1 - p_2 \log p_2 - (1 - p_1 - p_2) \log(1 - p_1 - p_2).$$

Энтропия системы максимальна, если состояния системы равновероятны:

x_i	x_1	x_2	x_3
p_i	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

и равна логарифму числа состояний. Следовательно, функция $H(p_1, p_2)$ достигает максимума в точке $\left(\frac{1}{3}, \frac{1}{3}\right)$ и наибольшее значение функции равно $H_{max}(p_1, p_2) = \log 3$.

Пример 4. Имеются две урны. В первой – 5 красных, 7 белых шаров. Во второй – 8 красных и 6 белых шаров. Из каждой урны берут по два шара. Исход какого из двух опытов следует считать более неопределенным?

Решение. Рассмотрим первый опыт, извлечение двух шаров из первой урны, как систему X с тремя исходами:

исходы x_i	x_1 (2 кр.ш.)	x_2 (1 кр.ш.+1 бел.ш.)	x_3 (2 бел.ш.)
вероятности исходов	$\frac{5}{12} \cdot \frac{4}{11} = \frac{20}{132}$	$\frac{5}{12} \cdot \frac{7}{11} + \frac{7}{12} \cdot \frac{5}{11} = \frac{70}{132}$	$\frac{7}{12} \cdot \frac{6}{11} = \frac{42}{132}$

Найдем энтропию системы X:
$H(X) = \eta\left(\frac{20}{132}\right) + \eta\left(\frac{70}{132}\right) + \eta\left(\frac{42}{132}\right) = 0,4118 + 0,4854 + 0,5256 = 1,4228$ бит.

Рассмотрим второй опыт, извлечение двух шаров из второй урны, как систему Y с тремя исходами:

исходы y_j	y_1 (2 кр.ш.)	y_2 (1 кр.ш.+1 бел.ш.)	y_3 (2 бел.ш.)
вероятности исходов	$\frac{8}{14}\cdot\frac{7}{13}=\frac{56}{182}$	$\frac{8}{14}\cdot\frac{6}{13}+\frac{6}{14}\cdot\frac{8}{13}=\frac{96}{182}$	$\frac{6}{14}\cdot\frac{5}{13}=\frac{30}{182}$

Найдем энтропию системы Y:

$H(Y) = \eta\left(\frac{56}{182}\right)+\eta\left(\frac{96}{182}\right)+\eta\left(\frac{30}{182}\right) = 0,5230+0,4870+0,4277 = 1,4377$ бит.

Т.к. $H(X) < H(Y)$, то исход второго опыта является более неопределенным.

Пример 5. Из наблюдений за погодой известно, что для пункта N вероятность того, что 30 мая будет дождь равна 0,15, а вероятность того, что дождя не будет 0,85. Для этого же пункта вероятность того, что 30 октября будет дождь равна 0,8, будет снег – 0,1, осадков не будет – 0,1. В какой из перечисленных дней погоду следует считать более неопределенной?

Решение. Рассмотрим опыт по выяснению погоды 30 мая как систему X с тремя состояниями:

дождь	снег	отсутствие осадков
0,15	0	0,85

$$H(X)=\eta(0,85)+\eta(0,15)=0,1993+0,4105=0,6098 \text{ бит}$$

Рассмотрим опыт по выяснению погоды 30 октября как систему Y_1 с тремя состояниями:

дождь	снег	отсутствие осадков
0,8	0,1	0,1

Имеем

$$H(Y_1)=\eta(0,8)+\eta(0,1)+\eta(0,1)=0,2575+0,3322+0,3322=0,9219 \text{ бит.}$$

Можно рассматривать опыт по выяснению погоды 30 октября и как систему Y_2 с двумя состояниями:

наличие осадков	отсутствие осадков
0,9	0,1

Получим

$$H(Y_2)=\eta(0,9)+\eta(0,1)=0,1368+0,3322=0,4690 \text{ бит.}$$

Если из всех характеристик погоды интересоваться лишь наличием осадков, тогда $H(X) > H(Y_2)$, следовательно, погоду 30 мая следует считать более неопределенной. Если, кроме наличия осадков, интересоваться еще и характером осадков, тогда $H(X) < H(Y_1)$ и погоду 30 октября следует считать более неопределенной.

3. Энтропия сложной системы

Рассмотрим сложную систему, полученную объединением двух или более простых систем.

Под объединением двух систем X и Y с возможными состояниями $x_1, ...x_n$ и $y_1, ...y_m$ понимается сложная система (X, Y), состояния которой (x_i, y_j) представляют собой все всевозможные комбинации состояний x_i, y_j $(i = 1, ...n,\ j = 1...m)$ систем X и Y.

Пусть p_{ij} – вероятность того, что система (X, Y) будет в состоянии (x_i, y_j):

$$p_{ij} = P((X \sim x_i)(Y \sim y_j)).$$

(x_i, y_j)	x_1	x_2	...	x_n
y_1	p_{11}	p_{21}	...	p_{n1}
y_2	p_{12}	p_{22}	...	p_{n2}
...	...	...	...	...
y_m	p_{1m}	p_{2m}	...	p_{nm}

Энтропия сложной системы равна сумме произведений вероятностей всех возможных ее состояний на их логарифмы с обратным знаком:

$$H(X, Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \log p_{ij}$$

или

$$H(X, Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} \eta(p_{ij}).$$

Энтропия сложной системы в форме математического ожидания:

$$H(X, Y) = M[-\log P(X, Y)].$$

Рассмотрим случай, когда системы X и Y независимы, т.е. принимают свои состояния независимо одна от другой. Найдем энтропию такой системы.

По теореме умножения вероятностей для независимых систем: $P(X, Y) = P(X) \cdot P(Y)$. Тогда,

$$\log P(X, Y) = \log P(X) + \log P(Y).$$

Так как $H(X,Y) = M[-\log P(X,Y)]$, то $H(X,Y) = M[-\log P(X) - \log P(Y)] = M[-\log P(X)] - M[-\log P(Y)]$. Следовательно,

$$H(X,Y) = H(X) + H(Y).$$

Итак, при объединении независимых систем их энтропии складываются (теорема сложения энтропий).

Для произвольного числа независимых систем $X_1, X_2, ...X_n$:

$$H(X_1, X_2, ...X_n) = \sum_{i=1}^{n} H(X_i).$$

Если системы зависимы, то энтропия сложной системы меньше, чем сумма энтропий ее составных частей. Покажем это на следующих примерах.

Пример 6. Вероятности состояний независимых систем X и Y заданы таблицами:

x_i	x_1	x_2
p_i	0,6	0,4

y_j	y_1	y_2
p_j	0,5	0,5

Найти энтропию объединенной системы (X,Y).

Решение. Т.к. системы X и Y независимы, то $H(X,Y) = H(X) + H(Y)$. Тогда,

$H(X) = \eta(0,6) + \eta(0,4) = 0,4422 + 0,5288 = 0,971$ бит,

$H(Y) = \eta(0,5) + \eta(0,5) = 0,5 + 0,5 = 1$ бит,

$H(X,Y) = 0,971 + 1 \approx 1,971$ бит.

Пример 7. Системы X и Y зависимы. Вероятности их состояний распределены как в примере 6. Найти энтропию объединенной системы (X,Y), если совместное распределение вероятностей описывается таблицей:

(x_i, y_j)	x_1	x_2
y_1	0,4	0,1
y_2	0,2	0,3

Решение.

$H(X,Y) = \eta(0,4) + \eta(0,1) + \eta(0,2) + \eta(0,3) = 0,5288 + 0,3322 + 0,4644 + 0,5211 = 1,8465$ бит.

4. Условная энтропия

Пусть имеются две зависимые системы X и Y. Пусть система X приняла состояние x_i.

$p(y_j/x_i)$ – условная вероятность того, что система Y примет состояние y_j при условии, что система X находится в состоянии x_i:

$$p(y_j/x_i) = P(Y \sim y_j/X \sim x_i).$$

Тогда,

$$H(Y/x_i) = -\sum_{j=1}^{m} p(y_j/x_i) \log p(y_j/x_i) \tag{1}$$

или

$$H(Y/x_i) = \sum_{j=1}^{m} \eta(p(y_j/x_i))$$

где $H(Y/x_i)$– *частная условная энтропия* системы Y при условии, что система X находится в состоянии x_i.

Частная условная энтропия зависит от того, какое состояние x_i приняла система X.

Определим *среднюю* (*полную*) условную энтропию системы Y с учетом того, что система X может принимать разные состояния:

$$H(Y/X) = \sum_{i=1}^{n} p_i H(Y/x_i). \tag{2}$$

Используя выражение (1), получим

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

Внесем p_i под знак второй суммы, получим

$$H(Y/X) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p_i p(y_j/x_i) \log p(y_j/x_i) \tag{3}$$

или

$$H(Y/X) = \sum_{i=1}^{n} \sum_{j=1}^{m} p_i \eta(p(y_j/x_i)).$$

С другой стороны, по теореме умножения вероятностей $p_{ij} = p_i \cdot p(y_j/x_i)$. Используя выражение (3), получим

$$H(Y/X) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \log p(y_j/x_i). \tag{4}$$

Придадим выражению (4) форму математического ожидания:

$$H(Y/X) = M[-\log P(Y/X)].$$

Величина $H(Y/X)$ характеризует степень неопределенности системы Y, остающуюся после того, как состояние системы X полностью определилось.

$H(Y/X)$ – *полная условная энтропия системы* Y *относительно системы* X.

Аналогично определим полную условную энтропию системы X относительно системы Y. Пусть $r_j = P(Y \sim y_j)$, тогда

$$H(X/Y) = \sum_{j=1}^{m} r_j H(X/y_j) \tag{5}$$

или

$$H(X/Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} r_j \eta(p(x_i/y_j)).$$

Пример 8. Сложная система (X, Y), задана таблицей:

(x_i, y_j)	x_1	x_2	x_3	x_4
y_1	0,5	0,3	0	0,01
y_2	0,1	0,01	0,06	0,02

Найти полные условные энтропии $H(Y/X)$ и $H(X/Y)$, и частные энтропии $H(Y/x_i)$, $i = 1, 2, 3, 4$ и $H(X/y_j)$, $j = 1, 2$.

Решение. Напишем таблицы вероятностей для систем X и Y:

x_i	x_1	x_2	x_3	x_3
p_i	0,6	0,31	0,06	0,03

y_j	y_1	y_2
r_j	0,81	0,19

1. Т.к. $p_{ij} = p_i \cdot p(y_j/x_i)$, то $p(y_j/x_i) = \frac{p_{ij}}{p_i}$. Найдем условные вероятности $p(y_j/x_i)$:

$$p(y_1/x_1) = \tfrac{0,5}{0,6} \quad p(y_1/x_2) = \tfrac{0,3}{0,31} \quad p(y_1/x_3) = 0 \quad p(y_1/x_4) = \tfrac{0,01}{0,03}$$
$$p(y_2/x_1) = \tfrac{0,1}{0,6} \quad p(y_2/x_2) = \tfrac{0,01}{0,31} \quad p(y_2/x_3) = 1 \quad p(y_2/x_4) = \tfrac{0,02}{0,03}$$

$H(Y/x_1) = \sum_{j=1}^{2} \eta(p(y_j/x_1)) = \eta(p(y_1/x_1)) + \eta(p(y_2/x_1)) = \eta\left(\frac{0,5}{0,6}\right) + \eta\left(\frac{0,1}{0,6}\right) = \eta(0,833) + \eta(0,166) = 0,2196 + 0,4301 = 0,6497$

$H(Y/x_2) = \eta\left(\frac{0,3}{0,31}\right) + \eta\left(\frac{0,01}{0,31}\right) = \eta(0,967) + \eta(0,032) = 0,0468 + 0,1589 = 0,2057$

$H(Y/x_3) = 0$

$H(Y/x_4) = \eta\left(\frac{0,01}{0,03}\right) + \eta\left(\frac{0,02}{0,03}\right) = \eta(0,333) + \eta(0,666) = 0,5283 + 0,3905 = 0,9188$

$H(Y/X) = \sum_{i=1}^{4} p_i H(Y/x_i) = p_1 H(Y/x_1) + p_2 H(Y/x_2) + p_3 H(Y/x_3) + p_4 H(Y/x_4) = 0,6 \cdot 0,6497 + 0,31 \cdot 0,2057 + 0,03 \cdot 0,9188 = 0,4811$ бит.

2. Т.к. $p_{ij} = r_j \cdot p(x_i/y_j)$, то $p(x_i/y_j) = \frac{p_{ij}}{r_j}$. Получим таблицу условных вероятностей $p(x_i/y_j)$:

$$p(x_1/y_1) = \tfrac{0,5}{0,81} \quad p(x_2/y_1) = \tfrac{0,3}{0,81} \quad p(x_3/y_1) = 0 \quad p(x_4/y_1) = \tfrac{0,01}{0,81}$$
$$p(x_1/y_2) = \tfrac{0,1}{0,19} \quad p(x_2/y_2) = \tfrac{0,01}{0,19} \quad p(x_3/y_2) = \tfrac{0,06}{0,19} \quad p(x_4/y_2) = \tfrac{0,02}{0,19}$$

$H(X/y_1) = \sum_{i=1}^{4} \eta(p(x_i/y_1)) = \eta(p(x_1/y_1)) + \eta(p(x_2/y_1)) + \eta(p(x_3/y_1)) + \eta(p(x_4/y_1)) = \eta\left(\tfrac{0,5}{0,81}\right) + \eta\left(\tfrac{0,3}{0,81}\right) + \eta\left(\tfrac{0,01}{0,81}\right) = \eta(0,617) + \eta(0,370) + \eta(0,012) = 0,4298 + 0,5307 + 0,0766 = 1,0371$

$H(X/y_2) = \eta\left(\tfrac{0,1}{0,19}\right) + \eta\left(\tfrac{0,01}{0,19}\right) + \eta\left(\tfrac{0,06}{0,19}\right) + \eta\left(\tfrac{0,02}{0,19}\right) = \eta(0,526) + \eta(0,052) + \eta(0,315) + \eta(0,105) = 0,4875 + 0,2218 + 0,5250 + 0,3414 = 1,5757$

$H(X/Y) = \sum_{j=1}^{2} r_j H(X/y_j) = r_1 H(X/y_1) + r_2 H(X/y_2) = 0,81 \cdot 1,0371 + 0,19 \cdot 1,5757 = 1,1394$ бит.

Пример 9. Опыт X состоит в извлечение двух шаров из урны, содержащей 16 красных и 7 белых шаров. Опыт Y - в извлечение из той же урны еще одного шара. Чему равна энтропия $H(Y)$ опыта Y и условная энтропия $H(Y/X)$ этого опыта?

Решение. Рассмотрим опыт X, как систему с тремя исходами. Вероятности исходов:

исходы x_i	x_1	x_2	x_3
	2 кр.ш.	1 кр.ш.+1 бел.ш.	2 бел.ш.
вероятности исходов p_i	$\frac{16}{23} \cdot \frac{15}{22} = \frac{240}{506}$	$\frac{16}{23} \cdot \frac{7}{22} + \frac{7}{23} \cdot \frac{16}{22} = \frac{224}{506}$	$\frac{7}{23} \cdot \frac{6}{22} = \frac{42}{506}$

Пока ничего не известно об опыте X, опыт Y следует рассматривать как извлечение одного шара из урны, содержащей первоначальный набор шаров (16 красных и 7 белых). Рассмотрим опыт Y, как систему с двумя исходами. Вероятности исходов:

исходы y_j	y_1	y_2
	1 кр.ш.	1 бел.ш.
вероятности исходов r_j	$\frac{16}{23}$	$\frac{7}{23}$

$H(Y) = \eta\left(\tfrac{16}{23}\right) + \eta\left(\tfrac{7}{23}\right) \approx 0,8870$

После того, как результат опыта X известен, вероятности исходов опыта Y будут иметь другие значения:

$$p(y_1/x_1) = \tfrac{14}{21} \quad p(y_1/x_2) = \tfrac{15}{21} \quad p(y_1/x_3) = \tfrac{16}{21}$$
$$p(y_2/x_1) = \tfrac{7}{21} \quad p(y_2/x_2) = \tfrac{6}{21} \quad p(y_2/x_3) = \tfrac{5}{21}$$

Найдем условные энтропии опыта Y относительно каждого исхода x_1, x_2, x_3 опыта X:

$H(Y/x_1) = \eta\left(\frac{14}{21}\right) + \eta\left(\frac{7}{21}\right) = \eta(0,666) + \eta(0,333) = 0,3905 + 0,5283 = 0,9188$

$H(Y/x_2) = \eta\left(\frac{15}{21}\right) + \eta\left(\frac{6}{21}\right) = \eta(0,714) + \eta(0,285) = 0,3470 + 0,5161 = 0,8631$

$H(Y/x_3) = \eta\left(\frac{16}{21}\right) + \eta\left(\frac{5}{21}\right) = \eta(0,761) + \eta(0,238) = 0,2999 + 0,4949 = 0,7948$

Тогда, $H(Y/X) = p_1 \cdot H(Y/x_1) + p_2 \cdot H(Y/x_2) + p_3 \cdot H(Y/x_3) = \frac{240}{506} \cdot 0,9188 + \frac{224}{506} \cdot 0,8631 + \frac{42}{506} \cdot 0,7948 = 0,8838$ бит.

5. Объединение зависимых систем

Найдем энтропию объединенной системы через энтропию ее составных частей.

Теорема. Если две системы X и Y объединить в одну, то энтропия объединенной системы равна энтропии одной из ее составных частей плюс условная энтропия второй части относительно первой:

$$H(X,Y) = H(X) + H(Y/X).$$

$\triangle$ Напишем $H(X,Y)$ в форме математического ожидания:

$$H(X,Y) = M[-\log P(X,Y)].$$

По теореме умножения вероятностей

$$P(X,Y) = P(X)P(Y/X).$$

Тогда,

$$\log P(X,Y) = \log P(X) + \log P(Y/X).$$

Имеем

$$M[-\log P(X,Y)] = M[-\log P(X)] + M[-\log P(Y/X)],$$

тогда $H(X,Y) = M[-\log P(X,Y)]$, $H(X) = M[-\log P(X)]$ и $H(Y/X) = M[-\log P(Y/X)]$, и

$$H(X,Y) = H(X) + H(Y/X). \quad \square$$

Аналогично можно показать, что $H(X,Y) = H(Y) + H(X/Y)$.

В частном случае, когда X и Y независимы, $H(Y/X) = H(Y)$, и, следовательно,

$$H(X,Y) = H(X) + H(Y).$$

В общем случае будем иметь

$$H(X,Y) \leqslant H(X) + H(Y).$$

Неравенство следует из того, что полная условная энтропия $H(Y/X)$ не может превосходить безусловную:

$$H(Y/X) \leqslant H(Y),$$

т.е. степень неопределенности системы не может увеличиваться от того, что состояние какой-то другой системы стало известным.

Из неравенства $H(X,Y) \leqslant H(X)+H(Y)$ следует, что энтропия сложной системы достигает максимума в крайнем случае, когда ее составные части независимы.

Другой крайний случай, когда состояние одной из систем (X) полностью определяет состояние другой (Y). В этом случае $H(Y/X) = 0$, и $H(X,Y) = H(X)$.

Если состояние каждой из систем X и Y однозначно определяет состояние другой, т.е. системы эквивалентны, то $H(X,Y) = H(X) = H(Y)$.

Обобщим теорему об энтропии сложной системы:

$H(X_1, X_2, X_3 ... X_n) = H(X_1) + H(X_2/X_1) + H(X_3/X_1, X_2) + ... + H(X_n/X_1, X_2, ... X_{n-1})$,

т.е. энтропия каждой последующей системы вычисляется при условии, что состояние всех предыдущих известно.

Пример 10. В условиях примера **9** найти энтропию $H(X)$ опыта X и условную энтропию $H(X/Y)$ этого опыта.

Решение. Имеем $H(X,Y) = H(X) + H(Y/X)$ и $H(X,Y) = H(Y) + H(X/Y)$, то $H(X)+H(Y/X) = H(Y)+H(X/Y)$. Следовательно, получим

$$H(X/Y) = H(X) + H(Y/X) - H(Y).$$

Вычисляя, получим $H(X) = \eta(0,474)+\eta(0,442)+\eta(0,083) = 0,5105+ 0,5206 + 0,2980 = 1,3291$ бит.

$H(X/Y) = H(X) - H(Y) + H(Y/X) = 1,3291 - 0,8870 + 0,8838 = 1,3259$ бит.

Глава 2.
Измерение информации

1. Определение информации

Выше определили *энтропию*, как меру неопределенности состояния физической системы. После получения сведений неопределенность может быть уменьшена. Чем больше получено сведений, тем больше будет информация о системе, менее неопределенным будет состояние системы. *Поэтому количество информации о системе измеряют уменьшением энтропии этой системы.*

Рассмотри систему X. Пусть в результате наблюдений над системой состояние системы становится полностью известным. Энтропия системы до наблюдений $H(X)$. После получения сведений состояние системы полностью определилось и энтропия системы стала равна нулю. При этом информация, получаемая в результате выяснения состояния системы X, равна уменьшению энтропии:

$$I(X) = H(X) - 0,$$

т.е.

$$I(X) = H(X) = -\sum_{i=1}^{n} p_i \log p_i.$$

Тогда, величина $I(X)$ — среднее значение случайной величины $-\log p_i$ ($i = 1, 2, ...n$):

$-\log p_i$	$-\log p_1$	$-\log p_2$	...	$-\log p_n$
p_1	p_1	p_2	...	p_n

При этом каждое отдельное слагаемое $-\log p_i$ рассматривается, как *частная информация*, получаемая от отдельного сообщения, состоящего в том, что система X находится в состоянии x_i. Введем величину:

$$I(x_i) = -\log p_i.$$

Тогда информация $I(X)$ представляется как средняя, или полная, информация, получаемая от всех возможных отдельных сообщений с учетом их вероятностей.

2. Свойства информации

1. Полная и частная информация не могут быть отрицательными:

$$I(X) \geqslant 0, \quad I(x_i) \geqslant 0.$$

2. Если все всевозможные состояния системы одинаково вероятны ($p_1 = p_2 = ...p_n = \frac{1}{n}$), то частная информация $I(x_i)$ от каждого отдельного сообщения,

$$I(x_i) = -\log p_i = \log n,$$

равна средней (полной) информации

$$I(X) = -n\frac{1}{n}\log\frac{1}{n} = \log n.$$

3. Если состояния системы обладают различными вероятностями, то информации, получаемые от разных сообщений неодинаковы. Наибольшую информацию несут сообщения о тех событиях которые изначально были наименее вероятны.

Рассмотрим систему X, вероятности состояний которой связаны с законом распределения:

x_i	x_1	x_2
p_i	0,99	0,01

Частная информация $I(x_1)$, получаемая от сообщения, что система X приняла состояние x_1, равна:

$$I(x_1) = -\log 0,99 \approx 0,0144 \text{ бит},$$

а частная информация в случае $X \sim x_2$:

$$I(x_2) = -\log 0,01 \approx 6,6438 \text{ бит}.$$

Пример 11. Лыжник съезжает с годы без падения с вероятностью $0,95$. Какое количество информации мы получим, узнав, что лыжник упал на склоне?

Решение. Рассмотрим лыжника, как систему X с двумя состояниями:

x_i	x_1	x_2
p_i	0,95	0,05

Сообщение «лыжник упал» несет информацию, равную:
$I(x_2) = -\log 0.05 = 4,3219$ бит.

Пример 12. Игрок наудачу бросает два игральных кубика. Какое количество информации при этом получает игрок?

Решение. Рассмотрим два кубика, как систему X с $6 \cdot 6 = 36$ равновероятными состояниями. Тогда,

$I(X) = \log 36 = 5,1699$ бит.

Пример 13. Из колоды в 36 карт наудачу берут три. Найти частную информацию из сообщения: «выпали шестерка, семерка, туз».

Решение. Вероятность события «выпали шестерка, семерка, туз» равна $p_i = 6 \cdot \frac{4 \cdot 4 \cdot 4}{36 \cdot 35 \cdot 34}$. Частная информация в этом случае равна:

$I(x_i) = -\log p_i = 6,8017$ бит.

Пример 14. Вероятность попадания в цель стрелка при одном выстреле 0,15. Стрелок производит по цели k независимых выстрелов. После чего, поступает сообщение, поражена цель или нет. Если цель поражена, стрельба прекращается. При каком k количество информации, содержащееся в сообщении, будет наибольшим?

Решение. Физическая система X — состояние цели после k-ого выстрела. Система X имеет два состояния:

x_1	x_2
цель поражена	цель не поражена
$p_1 = 1 - (1 - 0,15)^k$	$p_2 = (1 - 0,15)^k$

Информация, содержащаяся в сообщении о состоянии цели будет максимальной, если оба состояния x_1 и x_2 равновероятны: $p_1 = p_2$.
$1 - (1 - 0,15)^k = (1 - 0,15)^k \quad \Rightarrow 2 \cdot (1 - 0,15)^k = 1 \quad \Rightarrow \log(1 - 0,15)^k = \log \frac{1}{2} \quad \Rightarrow \quad k \log(1 - 0,15) = -1 \quad \Rightarrow \quad k = \frac{-1}{\log(1-0,15)} = 4$ выстрела.

3. Информация об одной системе, содержащаяся в другой системе

В предыдущих примерах предполагалась, что наблюдение ведется непосредственно за самой системой X. На практике часто система X оказывается недоступной. При этом выясняется состояние другой системы Y, связанной с X. Например, вместо непосредственного наблюдения за космическим кораблем ведется наблюдение за системой сигналов, передаваемых его аппаратурой.

Различия между интересующей нас системой X и поддающейся наблюдению системой Y бывают двух типов:

1. различия за счет того, что некоторые состояния системы X не находят отражение в системе Y. Например, за счет округления численных данных.

2. различия за счет ошибок: неточностей измерения параметров системы X и ошибок при передаче сообщений. Например, искажение сигнала в следствии помех.

В случае, когда интересующая система X и наблюдаемая система Y различны, то количество информации о системе X дающее наблюдение системы Y определяется как уменьшение энтропии системы X в результате получения сведений о состоянии системы Y:

$$I(Y,X) = H(X) - H(X/Y). \tag{1}$$

Величина $I(Y,X)$ — средняя, или полная, информация о системе X, содержащаяся в системе Y.

Теорема. Докажем, что $I(Y,X) = I(X,Y)$, т.е. из двух систем каждая содержит относительно другой одну и ту же полную информацию.

$\triangle$ Запишем энтропию объединенной системы двумя равносильными формулами:

$$H(X,Y) = H(X) + H(Y/X) \quad \text{и} \quad H(X,Y) = H(Y) + H(X/Y).$$

Следовательно,

$$H(X) + H(Y/X) = H(Y) + H(X/Y)$$

$$H(X) - H(X/Y) = H(Y) - H(Y/X)$$

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

$I(X,Y)$ — полная взаимная информация, содержащаяся в системах X и Y (Рис. 1).

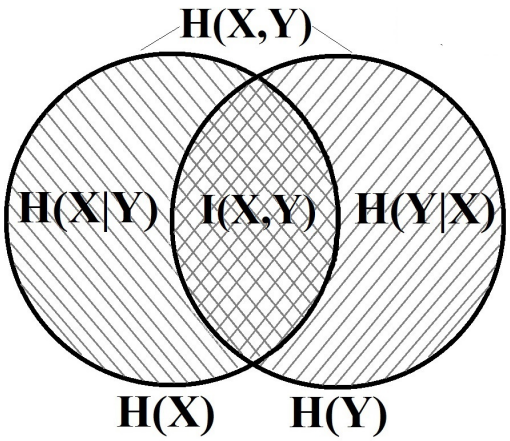


Рис.1. Полная взаимная информация, содержащаяся в системах X и Y.

Посмотрим, во что обращается информация в крайних случаях полной зависимости и полной независимости систем.

1. Если X и Y независимы, то $H(Y/X) = H(Y)$ и $I(X,Y) = 0$, т.е. полная взаимная информация, содержащаяся в независимых системах равна нулю (нельзя получить сведения о системе, наблюдая вместо нее другую, никак с ней не связанную).

2. Если состояние системы X полностью определяет состояние системы Y и наоборот, т.е. системы эквивалентны, то $H(X) = H(Y)$ и $H(X/Y) = H(Y/X) = 0$. Следовательно,

$$I(X,Y) = I(X) = I(Y) = H(X) = H(Y).$$

3. Рассмотрим случай, когда между системами X и Y наблюдается строгая односторонняя зависимость: состояние одной системы Y полностью определяет состояние другой X (подчиненной системы), но не наоборот.

X — подчиненная система, тогда $H(X/Y) = 0$ и $I(X,Y) = H(X)$, т.е. полная взаимная информация, содержащаяся в системах, из которых одна является подчиненной, равна энтропии подчиненной системы.

Получим выражение для информации $I(X,Y)$ через энтропию объединенной системы $H(X,Y)$ и энтропию ее составных частей $H(X)$ и $H(Y)$. Т.к.

$$H(X,Y) = H(Y) + H(X/Y), \quad \text{то} \quad H(X/Y) = H(X,Y) - H(Y).$$

Тогда,

$$I(X,Y) = H(X) - H(X/Y) = H(X) + H(Y) - H(X,Y). \tag{2}$$

Значит, полная взаимная информация, содержащаяся в двух системах, равна сумме энтропий этих систем минус энтропия объединенной системы.

Получим выражение для информации $I(X,Y)$ через вероятности состояний систем. Имеем равенство

$$I(X,Y) = H(X) + H(Y) - H(X,Y),$$

где

$$H(X) = M[-\log P(X)], \quad H(Y) = M[-\log P(Y)],$$

$$H(X,Y) = M[-\log P(X,Y)],$$

тогда

$$I(X,Y) = M[-\log P(X) - \log P(Y) + \log P(X,Y)],$$

или

$$I(X,Y) = M\left[\log \frac{P(X,Y)}{P(X)P(Y)}\right].$$

Запишем последнюю формулу в виде:

$$I(X,Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}p_{ij}\log\frac{p_{ij}}{p_i r_j}, \tag{3}$$

где $p_{ij}=P((X\sim x_i)(Y\sim y_j))$, $p_i=P(X\sim x_i)$, $r_j=P(Y\sim y_j)$.

Пример 15. Вероятности состояний зависимых систем X и Y заданы таблицей:

(x_i,y_j)	x_1	x_2
y_1	0,45	0,15
y_2	0,05	0,35

Найти полную взаимную информацию, содержащуюся в этих системах.

Решение. Найдем вероятности состояний каждой из систем:

x_i	x_1	x_2
p_i	0,5	0,5

y_j	y_1	y_2
r_j	0,6	0,4

$H(X)=1$,

$H(Y)=\eta(0,6)+\eta(0,4)=0,4422+0,5288=0,971$

$H(X,Y)=\eta(0,45)+\eta(0,15)+\eta(0,05)+\eta(0,35)=0,5184+0,4105+0,2161+0,5301=1,6751$

$I(X,Y)=H(X)+H(Y)-H(X,Y)=0,2959$ бит.

Пример 16. Вероятности состояний независимых систем X и Y заданы таблицами:

x_i	$x_1=0$	$x_2=1$
p_i	0,4	0,6

y_j	$y_1=-1$	$y_2=1$
r_j	0,8	0,2

Система Z связана с системами X и Y соотношением $Z=X-(Y-X)^2$. Найти полную информацию $I(Z,Y)$ о системе Y, содержащуюся в системе Z.

Решение. Найдем состояния системы Z. Если $x_1=0$ и $y_1=-1$, то $z=-1$. Если $x_1=0$ и $y_2=1$, то $z=-1$. Если $x_2=1$ и $y_1=-1$, то $z=-3$. Если $x_2=1$ и $y_2=1$, то $z=1$.

Найдем вероятности состояний системы Z:

z_k	$z_1=-3$	$z_2=-1$	$z_3=1$
q_k	$0,6\cdot0,8=0,48$	$0,4\cdot0,8+0,4\cdot0,2=0,4$	$0,6\cdot0,2=0,12$

Найдем вероятности состояний объединенной системы (Y,Z).

(y_j,z_k)	z_1	z_2	z_3
y_1	0,48	0,32	0
y_2	0	0,08	0,12

$H(Y) = \eta(0,8) + \eta(0,2) = 0,2575 + 0,4644 = 0,7219$

$H(Z) = \eta(0,48)+\eta(0,4)+\eta(0,12) = 0,5083+0,5288+0,3671 = 1,4042$

$H(Y,Z) = \eta(0,48) + \eta(0,32) + \eta(0,08) + \eta(0,12) = 0,5083 + 0,5260 + 0,2915 + 0,3671 = 1,6929$

$I(Z,Y) = H(Z) + H(Y) - H(Y,Z) = 0,4332$ бит.

Пример 17. Опыт X состоит в извлечении 21 шара из урны, содержащей 17 белых и 7 черных шаров (всего 24 шара). Опыт Y состоит в извлечении из той же урны еще одного шара. Чему равна информация $I(X,Y)$ об опыте Y, содержащаяся в опыте X?

Решение. Рассмотрим опыт X, как систему с четырьмя исходами. Вероятности исходов:

x_1 (осталось 3 б.ш.)	x_2 (осталось 2б.ш. и 1 ч.ш.)	x_3 (осталось 1б.ш. и 2ч.ш.)	x_4 (осталось 2 б.ш.)
$p_1 = \frac{17\cdot16\cdot15}{24\cdot23\cdot22} = \frac{4080}{21144}$	$p_2 = 3 \cdot \frac{17\cdot16\cdot7}{24\cdot23\cdot22} = \frac{5712}{21144}$	$p_3 = 3 \cdot \frac{17\cdot7\cdot6}{24\cdot23\cdot22} = \frac{2142}{21144}$	$p_4 = \frac{7\cdot6\cdot5}{24\cdot23\cdot22} = \frac{210}{21144}$

Пока ничего не известно об опыте X, опыт Y следует рассматривать как извлечение одного шара из урны, содержащей первоначальный набор шаров (17 белых и 7 черных). Рассмотрим опыт Y, как систему с двумя исходами. Вероятности исходов:

исходы y_j	y_1 1 б.ш.	y_2 1 ч.ш.
вероятности исходов r_j	$\frac{17}{24}$	$\frac{7}{24}$

$H(Y) = \eta\left(\frac{17}{24}\right) + \eta\left(\frac{7}{24}\right) = 0,3527 + 0,5182 = 0,8709$

После того, как результат опыта X известен, вероятности исходов опыта Y будут иметь другие значения:

$p(y_1/x_1) = 1 \qquad p(y_1/x_2) = \frac{2}{3} \qquad p(y_1/x_3) = \frac{1}{3} \qquad p(y_1/x_4) = 0$

$p(y_2/x_1) = 0 \qquad p(y_2/x_2) = \frac{1}{3} \qquad p(y_2/x_3) = \frac{2}{3} \qquad p(y_2/x_4) = 1$

Найдем условные энтропии опыта Y относительно каждого исхода x_1, x_2, x_3, x_4 опыта X:

$H(Y/x_1) = H(Y/x_4) = 0$

$H(Y/x_2) = H(Y/x_3) = \eta\left(\frac{2}{3}\right) + \eta\left(\frac{1}{3}\right) = \eta(0,6666) + \eta(0,3333) = 0,3905 + 0,5283 = 0,9188$

Тогда, $H(Y/X) = p_1 \cdot H(Y/x_1) + p_2 \cdot H(Y/x_2) + p_3 \cdot H(Y/x_3) + p_4 \cdot H(Y/x_4) = \frac{5712}{12144} \cdot 0,9188 + \frac{2142}{12144} \cdot 0,9188 = 0,5942$ бит

$I(X,Y) = H(Y) - H(Y/X) = 0,8709 - 0,5942 = 0,2767$ бит.

Пример 18. Из наблюдений за погодой известно, что 30 мая вероятность дождя 0,55; а вероятность того, что дождя не будет 0,45. Определенный метод прогноза на 30 мая оказывается правильным в 0,8 случаях, если предсказывается дождь, и в 0,65 случаях, если предсказывается отсутствие дождя. Какое количество информации дает прогноз о реальной погоде 30 мая.

Решение. Рассмотрим *определение* погоды (до прогноза) 30 мая как систему Y с двумя состояниями:

y_1 будет дождь	y_2 дождя не будет
0,55	0,45

$$H(Y) = \eta(0,55) + \eta(0,45) = 0,4744 + 0,5184 = 0,9928 \text{ бит.}$$

Рассмотрим *предсказание* погоды (определенным методом прогноза) 30 мая как систему X с двумя состояниями: x_1 — предсказание дождя, x_2 — предсказание отсутствия дождя. Вероятности $p(x_1)$ и $p(x_2)$ пока не известны.

После того, как прогноз сделан, вероятности исходов опыта Y имеют другие значения:

$p(y_1/x_1) = 0,8$	$p(y_1/x_2) = 0,35$
$p(y_2/x_1) = 0,2$	$p(y_2/x_2) = 0,65$

$$H(Y/x_1) = \eta(0,8) + \eta(0,2) = 0,2575 + 0,4644 = 0,7219,$$

$$H(Y/x_2) = \eta(0,35) + \eta(0,65) = 0,5301 + 0,4040 = 0,9341.$$

$I(X,Y)$ — полезность прогноза, т.е. информация о том, насколько предсказание погоды уменьшает неопределенность о реальной погоде 30 мая.

$I(X,Y) = H(Y) - H(Y/X)$, где $H(Y/X) = p(x_1) \cdot H(Y/x_1) + p(x_2) \cdot H(Y/x_2)$.

Найдем вероятности $p(x_1)$ и $p(x_2)$ ($p(x_2) = 1 - p(x_1)$). Вероятность того, что 30 мая будет дождь:

$$p(y_1) = p(x_1) \cdot p(y_1/x_1) + p(x_2) \cdot p(y_1/x_2).$$

Отсюда $0,55 = p(x_1) \cdot 0,8 + p(x_2) \cdot 0,35$. Следовательно, $p(x_1) = \frac{4}{9}$, $p(x_2) = \frac{5}{9}$.

Тогда,

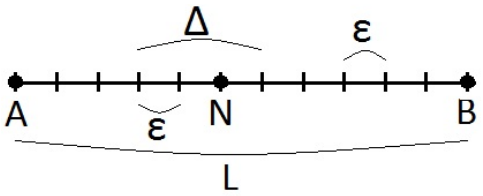
$$H(Y/X) = \frac{4}{9} \cdot 0,7219 + \frac{5}{9} \cdot 0,9341 = 0,8397 \text{ бит},$$

$$I(X,Y) = H(Y) - H(Y/X) = 0,9928 - 0,8397 = 0,1531 \text{ бит.}$$

Пример 19. Опыт Y состоит в определении положения точки N относительно которой заранее известно, что она расположена на отрезке AB длины L. Опыт X — в измерении длины отрезка AN с помощью измерительного прибора, дающего значение длины с точностью до Δ (например, при помощи линейки, на которой нанесена шкала с делениями длины Δ). Чему равна информация $I(X,Y)$, содержащаяся в результате измерения, относительно истинного положения точки N?

Решение. Опыт Y может иметь бесконечно много исходов (точка N может совпадать с любой точкой отрезка AB). Опыту Y нельзя приписать никакой конечной энтропии. Но, информация $I(X,Y) = H(Y) - H(Y/X)$ имеет определенное конечное значение.

Разобьем отрезок AB на маленькие отрезки длины ε. Выберем ε так, чтобы на отрезке AB и на отрезке Δ уложилось целое число таких малых отрезков, т.е., чтобы $\frac{L}{\varepsilon}$ и $\frac{\Delta}{\varepsilon}$ были целыми числами.



Будем искать положение точки N с точностью до ε. До измерения известно, что точка N располагается где-то на отрезке AB.

Опыт Y_ε — определение положения точки N с точностью до ε имеет $\frac{L}{\varepsilon}$ равновероятных исходов. Значит,

$$H(Y_\varepsilon) = \log \frac{L}{\varepsilon}$$

После того как произвели опыт X (измерили AN), выяснили, что N помещается внутри меньшего интервала длины Δ. При известном исходе X, опыт Y_ε будет иметь $\frac{\Delta}{\varepsilon}$ равновероятных исходов. Следовательно,

$$H(Y_\varepsilon/X) = \log \frac{\Delta}{\varepsilon}$$

Отсюда,

$$I(X,Y_\varepsilon) = H(Y_\varepsilon) - H(Y_\varepsilon/X) = \log \frac{L}{\varepsilon} - \log \frac{\Delta}{\varepsilon} = \log \frac{L}{\Delta}.$$

При неограниченном уменьшении ε обе энтропии будут неограниченно возрастать, но информация $I(X,Y_\varepsilon)$ при этом не меняется. Поэтому, информация $I(X,Y)$ относительно истинного положения точки N, которую можно определить как предел $\lim_{\varepsilon\to\infty} I(X,Y_\varepsilon)$, также равна $\log \frac{L}{\Delta}$.

4. Частная информация о системе

Выше рассмотрели *полную* информацию о системе X, содержащуюся в сообщении о том, в каком состоянии находится система Y.

I. Найдем *частную* информацию о системе X, содержащуюся в отдельном сообщении, указывающем, что система Y находится в конкретном состоянии y_j. Обозначим эту информацию $I(y_j, X)$. При этом должно быть выполнено:

$$I(Y, X) = \sum_{j=1}^{m} r_j I(y_j, X). \tag{4}$$

Имеем $p_{ij} = r_j p(x_i/y_j)$, тогда

$I(X, Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \log \frac{p_{ij}}{p_i r_j} = \sum_{i=1}^{n} \sum_{j=1}^{m} r_j p(x_i/y_j) \log \frac{r_j p(x_i/y_j)}{p_i r_j} = \sum_{j=1}^{m} r_j \sum_{i=1}^{n} p(x_i/y_j) \log \frac{p(x_i/y_j)}{p_i}$.

Отсюда, сравнив с (4), получим

$$I(y_j, X) = \sum_{i=1}^{n} p(x_i/y_j) \log \frac{p(x_i/y_j)}{p_i}.$$

Примем это выражение за определение частной информации о системе, содержащейся в сообщении о событии.

Докажем, что частная информация $I(y_j, X)$ не может быть отрицательна.

$\triangle$ Пусть $\frac{p(x_i/y_j)}{p_i} = q_{ij}$, тогда $\log \frac{p(x_i/y_j)}{p_i} = \log q_{ij}$.

При любом $x > 0$ выполнено $\ln x \leqslant x - 1$, имеем при $x = \frac{1}{q_{ij}}$

$$\ln \frac{1}{q_{ij}} \leqslant \frac{1}{q_{ij}} - 1 \quad \Rightarrow \quad -\ln q_{ij} \leqslant \frac{1}{q_{ij}} - 1 \quad \Rightarrow \quad \ln q_{ij} \geqslant 1 - \frac{1}{q_{ij}}.$$

$$\log q_{ij} = \frac{\ln q_{ij}}{\ln 2} \geqslant \frac{1}{\ln 2} \left(1 - \frac{1}{q_{ij}}\right)$$

Т.к. $\quad I(y_j, X) = \sum_{i=1}^{n} p(x_i/y_j) \log \frac{p(x_i/y_j)}{p_i} \quad$ и $\quad q_{ij} = \frac{p(x_i/y_j)}{p_i}, \quad$ то

$$I(y_j, X) = \sum_{i=1}^{n} p(x_i/y_j) \log q_{ij} \geqslant \sum_{i=1}^{n} p(x_i/y_j) \frac{1}{\ln 2} \left(1 - \frac{1}{q_{ij}}\right) =$$

$$= \frac{1}{\ln 2} \sum_{i=1}^{n} p(x_i/y_j) \left(1 - \frac{p_i}{p(x_i/y_j)}\right) = \frac{1}{\ln 2} \left(\sum_{i=1}^{n} p(x_i/y_j) - \sum_{i=1}^{n} p_i\right)$$

С учетом равенств $\sum_{i=1}^{n} p(x_i/y_j) = 1$ и $\sum_{i=1}^{n} p_i = 1$, получим $I(y_j, X) \geqslant 0$.

Доказали, что частная информация о системе X, заключенная в сообщении о состоянии y_j системы Y, не может быть отрицательной. $\square$

Значит, неотрицательна и полная взаимная информация, как математическое ожидание неотрицательной случайной величины: $I(X,Y) > 0$.

Имеем $I(X,Y) = H(Y) - H(Y/X)$, тогда $H(Y) - H(Y/X) \geqslant 0$ и, следовательно, $H(Y) \geqslant H(Y/X)$. Полная условная энтропия системы не превосходит ее безусловной энтропии.

Преобразуем формулу для частной информации, введя вместо условных $p(x_i/y_j)$ вероятностей безусловные. Т.к. $p(x_i/y_j) = \frac{p_{ij}}{r_j}$, то

$$I(y_j, X) = \sum_{i=1}^{n} \frac{p_{ij}}{r_j} \log \frac{p_{ij}}{p_i r_j}.$$

Таким образом, определили частную информацию о системе X, содержащуюся в сообщении «система Y находится в состоянии y_j».

II. Определим частную информацию о событии $X \sim x_i$, содержащуюся в событии $Y \sim y_j$, т.е. получим информацию «от события к событию». Введем информацию «от события к событию» следующим образом:

$$I(y_j, x_i) = \log \frac{p(x_i/y_j)}{p_i}. \tag{5}$$

Частная информация о событии, получаемая в результате сообщения о другом событии, равна логарифму отношения вероятности первого сообщения после сообщения о другом событии к его же вероятности до сообщения.

Из формулы (5) видно, что если вероятность события $X \sim x_i$ в результате сообщения $Y \sim y_j$ увеличивается, т.е.

$$p(x_i/y_j) > p_i,$$

то информация $I(y_j, x_i)$ положительна. В противном случае она отрицательна.

В частном случае, когда появление события $Y \sim y_j$ полностью исключает возможность появления события $X \sim x_i$ (т.е. когда эти события несовместны), то

$$I(y_j, x_i) = -\infty.$$

Т.к. $I(y_j, x_i) = \log \frac{p(x_i/y_j)}{p_i} = \log \frac{p_{ij}}{p_i r_j}$, то частная информация симметрична относительно x_i и y_j. Следовательно,

$$I(x_i, y_j) = I(y_j, x_i).$$

Таким образом, ввели *три вида информации*:

1. $I(Y,X)$ — полная информация о системе X, содержащаяся в системе Y.

2. $I(y_j,X)$ — частная информация о системе X, содержащаяся в событии (сообщении) $Y\sim y_j$.

3. $I(y_j,x_i)$ — частная информация о событии $X\sim x_i$, содержащаяся в событии $Y\sim y_j$.

Пример 20. Вероятности состояний зависимых систем X и Y заданы таблицей:

(x_i,y_j)	x_1	x_2	x_3
y_1	0,08	0,15	0,4
y_2	0,1	0,25	0,02

1. Найти частную информацию $I(y_2,X)$ о системе X, содержащуюся в сообщении: «система Y находится в состоянии y_2».

2. Найти частную информацию $I(y_1,x_2)$ о событии «система X находится в состоянии x_2», содержащуюся в сообщении: «система Y находится в состоянии y_1».

3. Найти полную информацию $I(Y,X)$ о системе X, содержащуюся в системе Y.

Решение. Напишем таблицы вероятностей для систем X и Y:

x_i	x_1	x_2	x_3
p_i	0,18	0,4	0,42

y_j	y_1	y_2
r_j	0,63	0,37

1. $I(y_2,X)=\sum_{i=1}^{3}\frac{p_{i2}}{r_2}\log\frac{p_{i2}}{p_ir_2}=\frac{p_{12}}{r_2}\log\frac{p_{12}}{p_1r_2}+\frac{p_{22}}{r_2}\log\frac{p_{22}}{p_2r_2}+\frac{p_{32}}{r_2}\log\frac{p_{32}}{p_3r_2}=\frac{0,1}{0,37}\log\frac{0,1}{0,18\cdot0,37}+\frac{0,25}{0,37}\log\frac{0,25}{0,4\cdot0,37}+\frac{0,02}{0,37}\log\frac{0,02}{0,42\cdot0,37}=0,1584+0,5110-0,1598=0,5096$ бит.

2. $I(y_1,x_2)=\log\frac{p_{21}}{p_2r_1}=\log\frac{0,15}{0,4\cdot0,63}=-0,7484$ бит.

3. $I(Y,X)=H(X)+H(Y)-H(X,Y)=1,4997+0,9506-2,1759=0,2744$ бит.

Пример 21. Опыт X состоит в извлечении трех шаров из урны, содержащей 16 красных и 6 белых шаров (всего 22 шара). Опыт Y в извлечении из той же урны еще двух шаров.

1. Чему равна частная информация $I(x_2,Y)$ об опыте Y, содержащаяся в сообщении: «при первом изъятии достали два красных и один белый шар»?

2. Чему равна частная информации $I(x_4,y_1)$ о событии: «при втором изъятии достали два красных шара», содержащаяся в сообщении: «при первом изъятии достали три белых шара»?

3. Чему равна полная информация об опыте Y, содержащаяся в опыте X?

Решение. Рассмотрим опыт X, как систему с четырьмя исходами. Вероятности исходов:

x_1 (3 кр.ш.)	x_2 (2 кр.ш. и 1 б.ш.)	x_3 (1кр.ш. и 2б.ш.)	x_4 (3 б.ш.)
$p_1 = \frac{16\cdot15\cdot14}{22\cdot21\cdot20} = \frac{3360}{9240}$	$p_2 = 3\cdot\frac{16\cdot15\cdot6}{22\cdot21\cdot20} = \frac{4320}{9240}$	$p_3 = 3\cdot\frac{6\cdot5\cdot16}{22\cdot21\cdot20} = \frac{1440}{9240}$	$p_4 = \frac{6\cdot5\cdot4}{22\cdot21\cdot20} = \frac{120}{9240}$

Пока ничего не известно об опыте X, опыт Y следует рассматривать как извлечение двух шаров из урны, содержащей первоначальный набор шаров (16 красных и 6 белых). Рассмотрим опыт Y, как систему с тремя исходами. Вероятности исходов:

исходы y_j	y_1 (2 кр.ш.)	y_2 (1 кр.ш. и 1 б.ш.)	y_3 (2 б.ш.)
вероятности исходов r_j	$\frac{16\cdot15}{22\cdot21} = \frac{240}{462}$	$2\cdot\frac{16\cdot6}{22\cdot21} = \frac{192}{462}$	$\frac{6\cdot5}{22\cdot21} = \frac{30}{462}$

После того, как результат опыта X известен, вероятности исходов опыта Y будут иметь другие значения:

$p(y_1/x_1) = \frac{156}{342}$ $p(y_1/x_2) = \frac{182}{342}$ $p(y_1/x_3) = \frac{210}{342}$ $p(y_1/x_4) = \frac{240}{342}$
$p(y_2/x_1) = \frac{156}{342}$ $p(y_2/x_2) = \frac{140}{342}$ $p(y_2/x_3) = \frac{120}{342}$ $p(y_2/x_4) = \frac{96}{342}$
$p(y_3/x_1) = \frac{30}{342}$ $p(y_3/x_2) = \frac{20}{342}$ $p(y_3/x_3) = \frac{12}{342}$ $p(y_3/x_4) = \frac{6}{342}$

1. $I(x_2, Y) = \sum_{j=1}^{3} = p(y_j/x_2)\log\frac{p(y_j/x_2)}{r_j} = \frac{182}{342}\log\frac{182\cdot462}{342\cdot240} + \frac{140}{342}\log\frac{140\cdot462}{342\cdot192} + \frac{20}{342}\log\frac{20\cdot462}{342\cdot30} = 0,0185 - 0,0089 - 0,0088 = 0,0008$ бит

2. $I(x_4, y_1) = \log\frac{p(y_1/x_4)}{r_1} = \log\frac{240\cdot462}{342\cdot240} = 0,4338$ бит

3. $H(Y) = \eta\left(\frac{240}{462}\right) + \eta\left(\frac{192}{462}\right) + \eta\left(\frac{30}{462}\right) = 1,2715$

$H(Y/x_1) = \eta\left(\frac{156}{342}\right) + \eta\left(\frac{156}{342}\right) + \eta\left(\frac{30}{342}\right) = 1,3397$

$H(Y/x_2) = \eta\left(\frac{182}{342}\right) + \eta\left(\frac{140}{342}\right) + \eta\left(\frac{20}{342}\right) = 1,2502$

$H(Y/x_3) = \eta\left(\frac{210}{342}\right) + \eta\left(\frac{120}{342}\right) + \eta\left(\frac{12}{342}\right) = 1,1315$

$H(Y/x_4) = \eta\left(\frac{240}{342}\right) + \eta\left(\frac{96}{342}\right) + \eta\left(\frac{6}{342}\right) = 0,9734$

$H(Y/X) = p_1\cdot H(Y/x_1) + p_2\cdot H(Y/x_2) + p_3\cdot H(Y/x_3) + p_4\cdot H(Y/x_4) = \frac{3360}{9240}\cdot1,3397 + \frac{4320}{9240}\cdot1,2502 + \frac{1440}{9240}\cdot1,1315 + \frac{120}{9240}\cdot0,9734 = 1,2606$

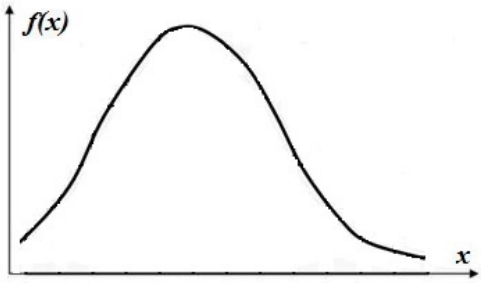
$I(X, Y) = H(Y) - H(Y/X) = 1,2715 - 1,2606 = 0,0109$ бит.

Глава 3.
Энтропия и информация для непрерывных систем

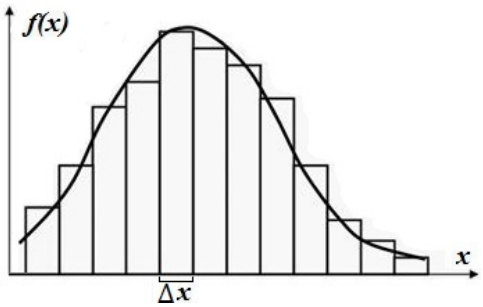
1. Энтропия для непрерывных систем

До сих пор рассматривались системы, состояния которых $x_1, ...x_n$ можно перечислить. На практике встречаются системы, состояния которых непрерывно переходят одно в другое. Для таких систем распределение вероятностей характеризуется плотностью. Такие системы будем называть непрерывными системами.

Рассмотрим систему X, определяемую непрерывной случайной величиной X с плотностью распределения $f(x)$.



Установим некоторый отрезок Δx, в пределах которого состояния системы X будем считать неразличимыми, т.е. сведем непрерывную систему к дискретной. Это равносильно замене непрерывной кривой на ступенчатую. При этом каждый участок Δx заменяется одной точкой-представителем x_i.



Площади прямоугольников равны вероятностям попадания в соответствующие разряды: $f(x_i)\Delta x$.

Определим приближенно энтропию системы X, рассматриваемой с точностью до Δx:

$H_{\Delta x}(X) = -\sum_i f(x_i)\Delta x \log[f(x_i)\Delta x] = -\sum_i f(x_i)\Delta x(\log[f(x_i)] + \log \Delta x) = -\sum_i f(x_i)\Delta x \log[f(x_i)] - \sum_i f(x_i)\Delta x \log \Delta x =$
$= -\sum_i (f(x_i)\log[f(x_i)])\Delta x - \log \Delta x \sum_i f(x_i)\Delta x.$

При достаточно малом Δx:

$$\sum_i (f(x_i)\log[f(x_i)])\Delta x \approx \int_{-\infty}^{+\infty} f(x)\log f(x)dx$$

$$\sum_i f(x_i)\Delta x \approx \int_{-\infty}^{+\infty} f(x)dx = 1.$$

Тогда, формула для энтропии принимает вид

$$H(X) = -\int_{-\infty}^{+\infty} f(x)\log f(x)dx - \log \Delta x.$$

Здесь при $\Delta x \to 0$ выполнено $\log \Delta x \to -\infty$, т.е., чем точнее мы хотим задать состояние системы X, тем большую степень неопределенности необходимо устранить. При неограниченном уменьшении Δx эта неопределенность растет тоже неограниченно.

Величина $H^*(X) = -\int_{-\infty}^{+\infty} f(x)\log f(x)dx$ называется «приведенной энтропией» непрерывной системы X. Тогда

$$H(X) = H^*(X) - \log \Delta x.$$

Получим выражение для энтропии в форме математического ожидания. Перепишем выражение для энтропии в виде:

$H(X) = -\int_{-\infty}^{+\infty} f(x)\log(f(x))\,dx - \log(\Delta x)\int_{-\infty}^{+\infty} f(x)dx =$
$= -\int_{-\infty}^{+\infty} f(x)\log(f(x))\,dx - \int_{-\infty}^{+\infty} f(x)\log(\Delta x)\,dx =$
$= -\int_{-\infty}^{+\infty} f(x)\log(f(x)\Delta x)\,dx$

Это выражение — математическое ожидание функции $-\log(f(x)\Delta x)$ от случайной величины X с плотностью $f(x)$:

$$H(X) = M[-\log(f(X)\Delta x)]$$

2. Условная энтропия для непрерывных систем

Рассмотрим две непрерывные системы X и Y, в общем случае зависимые.

$f(x,y)$ — плотность распределения для состояний объединенной системы (X,Y).

$f_1(x)$, $f_2(y)$ — плотности распределения систем X и Y соответственно.

$f(y/x)$, $f(x/y)$ — условные плотности распределения.

Определим частную условную энтропию $H(Y/x)$, т.е. энтропию системы Y при условии, что система X приняла определенное состояние x:

$$H(Y/x) = -\int_{-\infty}^{+\infty} f(y/x)\log f(y/x)dy - \log\Delta y.$$

Чтобы получить выражение для полной (средней) условной энтропии $H(Y/X)$, надо осреднить частную условную энтропию $H(Y/x)$ по всем состояниям x с учетом их вероятностей, характеризуемых плотностью $f_1(x)$:

$$H(Y/X) = -\iint_{-\infty}^{+\infty} f_1(x)f(y/x)\log f(y/x)dxdy - \log\Delta y.$$

Полная условная энтропия в форме математического ожидания:

$H(Y/X) = M[-\log f(Y/X)] - \log\Delta y$ или $H(Y/X) = M[-\log(f(Y/X)\Delta y)]$

3. Энтропия объединенной непрерывной системы

Рассмотрим объединенную систему (X, Y). Пусть Δx и Δy — «участки нечувствительности» для систем X и Y соответственно. Для объединенной системы «участок нечувствительности» — элементарный прямоугольник $\Delta x\Delta y$. Тогда энтропия объединенной системы (X, Y):

$$H(X, Y) = M[-\log(f(X, Y)\Delta x\Delta y)]$$

Пусть $f(x, y) = f_1(x)f(y/x)$, тогда $f(X, Y) = f_1(X)f(Y/X)$. Следовательно,

$$H(X, Y) = M[-\log f_1(X) - \log f(Y/X) - \log\Delta x - \log\Delta y] =$$
$$= M[-\log(f_1(X)\Delta x)] + M[-\log(f(Y/X)\Delta y)]$$

Имеем, $H(X, Y) = H(X) + H(Y/X)$. Таким образом, для непрерывных систем остается в силе правило сложения энтропий.

4. Информация для непрерывных систем

Количество информации вычисляется как разность двух энтропий. При этом неограниченно возрастающие слагаемые в формулах для энтропий взаимно уничтожаются. Все виды информации, связанные с непрерывными величинами, оказываются не зависящими от «участка нечувствительности».

$$H(Y) = M[-\log(f_2(Y)\Delta y)] \quad \text{и} \quad H(Y/X) = M[-\log(f(Y/X)\Delta y)]$$

$I(X, Y) = H(Y) - H(Y/X) = M[-\log(f_2(Y)\Delta y)] - M[-\log(f(Y/X)\Delta y)] =$
$M[-\log(f_2(Y)\Delta y) + \log(f(Y/X)\Delta y)] = M\left[\log\frac{f(Y/X)\Delta y}{f_2(Y)\Delta y}\right] = M\left[\log\frac{f(X,Y)}{f_1(X)\cdot f_2(Y)}\right]$

Отсюда, выражение для полной взаимной информации, содержащейся в двух непрерывных системах X и Y:

$$I(X,Y) = \int\limits_{-\infty}^{+\infty}\!\!\!\int f(x,y)\log\frac{f(x,y)}{f_1(x)f_2(y)}dxdy.$$

Пример 22. Найти энтропию непрерывной системы X, все состояния которой на отрезке $[a,b]$ одинаково вероятны.

Решение. Состояния системы X на отрезке $[a,b]$ одинаково вероятны. Для плотности вероятности имеем:

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{если } a \leqslant x \leqslant b; \\ 0, & \text{если } x < a, x > b. \end{cases}$$

$H(X) = -\int_{-\infty}^{+\infty} f(x)\log f(x)dx - \log\Delta x = -\int_a^b \frac{1}{b-a}\log\frac{1}{b-a}dx - \log\Delta x = -\frac{1}{b-a}\log\frac{1}{b-a}x\Big|_a^b - \log\Delta x = -\log\frac{1}{b-a} - \log\Delta x = \log\frac{b-a}{\Delta x}$.

Пусть $a = 3$ и $b = 6$, тогда $H(X) = \log\frac{3}{\Delta x}$.

Если $\Delta x = 1$, то $H(X) = \log\frac{3}{1} = 1,584$ бит

Если $\Delta x = 0,1$, то $H(X) = \log\frac{3}{0,1} = 2,131$ бит

Если $\Delta x = 0,01$, то $H(X) = \log\frac{3}{0,01} = 8,228$ бит

Пример 23. Найти энтропию непрерывной системы X, состояния которой распределены по нормальному закону с математическим ожиданием m_x и среднеквадратичным отклонением σ_x.

Решение. Состояния системы распределены по нормальному закону, плотность вероятности задается формулой:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_x}e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}.$$

$$H^*(X) = M[-\log f(X)] = M\left[-\log\left(\frac{1}{\sqrt{2\pi}\sigma_x}e^{-\frac{(X-m_x)^2}{2\sigma_x^2}}\right)\right] =$$

$$= M\left[-\log\left(\frac{1}{\sqrt{2\pi}\sigma_x}\right) - \log\left(e^{-\frac{(X-m_x)^2}{2\sigma_x^2}}\right)\right] =$$

$$= M\left[\log\left(\sqrt{2\pi}\sigma_x\right) + \frac{(X-m_x)^2}{2\sigma_x^2}\log e\right] = \log\left(\sqrt{2\pi}\sigma_x\right) + \frac{\log e}{2\sigma_x^2}M\left[(X-m_x)^2\right]$$

Т.к. $M\left[(X-m_x)^2\right] = D[X] = \sigma_x^2$, то

$$H^*(X) = \log\left(\sqrt{2\pi}\sigma_x\right) + \frac{\log e}{2\sigma_x^2}\sigma_x^2 = \log\left(\sqrt{2\pi e}\sigma_x\right)$$

$$H(X) = H^*(X) - \log \Delta x = \log\left(\sqrt{2\pi e}\sigma_x\right) - \log \Delta x = \log \frac{\sqrt{2\pi e}\sigma_x}{\Delta x}$$

При $\sigma_x = 1$ имеем $H(X) = \log \frac{\sqrt{2\pi e}}{\Delta x}$

Если $\Delta x = 0,1$, то $H(X) = \log \frac{\sqrt{2\pi e}}{1} = 2,046$ бит

Если $\Delta x = 0,01$, то $H(X) = \log \frac{\sqrt{2\pi e}}{0,1} = 5,368$ бит

Если $\Delta x = 0,001$, то $H(X) = \log \frac{\sqrt{2\pi e}}{0,01} = 8,689$ бит

Энтропия случайной величины не зависит от ее математического ожидания.

Пример 24. Состояние летательного аппарата характеризуется двумя случайными величинами: высотой полета H и модулем скорости V. Высота летательного аппарата распределена с равномерной плотностью на участке (h_1, h_2), скорость V — по нормальному закону с м.о. v_0 и с.к.о. σ_v. Величины H и V независимы. Найти энтропию объединенной системы.

Решение.

$$H(H) = \log \frac{h_2 - h_1}{\Delta h}, \qquad H(V) = \log \frac{\sqrt{2\pi e}\sigma_v}{\Delta v}$$

где Δh и Δv — «участки нечувствительности» при определении высоты и скорости соответственно.

Т.к. величины H и V независимы, то

$$H(H,V) = H(H)+H(V) = \log \frac{h_2 - h_1}{\Delta h} + \log \frac{\sqrt{2\pi e}\sigma_v}{\Delta v} = \log\left[\frac{h_2 - h_1}{\Delta h} \cdot \frac{\sqrt{2\pi e}\sigma_v}{\Delta v}\right]$$

Каждый множитель под знаком логарифма показывает, сколько «участков нечувствительности» укладывается на некотором отрезке, характерном для случайной величины. В случае с равномерной плотностью этот участок представляет собой участок возможных значений случайной величины. В случае нормального распределения этот участок равен $\sqrt{2\pi e}\sigma$, где σ - с.к.о.

Пример 25. Случайная величина X распределена по нормальному закону ($m_x = 0$, σ_x). Производится измерение случайной величины X. Ошибка измеренная Z также распределена по нормальному закону ($m_z = 0$, σ_z). X и Z независимые. Случайная величина $Y = X + Z$ — результат измерения. Какое количество информации о величине X содержит величина Y?

Решение. Воспользуемся формулой $I(X,Y) = M\left[\log \frac{f(X,Y)}{f_1(X)\cdot f_2(Y)}\right]$.

Преобразуем $\log \frac{f(x,y)}{f_1(x)\cdot f_2(y)} = \log \frac{f_1(x)\cdot f(y/x)}{f_1(x)\cdot f_2(y)} = \log \frac{f(y/x)}{f_2(y)}$

Т.к. $\sigma_y = \sqrt{D(Y)} = \sqrt{D(X+Z)} = \sqrt{D(X)+D(Z)} = \sqrt{\sigma_x^2+\sigma_z^2}$, то

$$f_2(y) = f_2(x+z) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{y^2}{2\sigma_y^2}} = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_x^2+\sigma_z^2}} e^{-\frac{y^2}{2(\sigma_x^2+\sigma_z^2)}}$$

После того, как о состоянии величины X все станет известно, состояние величины Y $(Y = X + Z)$ будет зависеть только от величины Z. Тогда,

$$f(y/x) = \frac{1}{\sqrt{2\pi}\sigma_z} e^{-\frac{(y-x)^2}{2\sigma_z^2}} = \frac{1}{\sqrt{2\pi}\sigma_z} e^{-\frac{z^2}{2\sigma_z^2}}$$

$$\log \frac{f(x,y)}{f_1(x)\cdot f_2(y)} = \log \frac{f(y/x)}{f_2(y)} = \log \left[\frac{\frac{1}{\sqrt{2\pi}\sigma_z} e^{-\frac{z^2}{2\sigma_z^2}}}{\frac{1}{\sqrt{2\pi}\sqrt{\sigma_x^2+\sigma_z^2}} e^{-\frac{y^2}{2(\sigma_x^2+\sigma_z^2)}}} \right] =$$

$$= \log \left[\frac{\sqrt{2\pi}\sqrt{\sigma_x^2+\sigma_z^2}}{\sqrt{2\pi}\sigma_z} e^{\frac{y^2}{2(\sigma_x^2+\sigma_z^2)} - \frac{z^2}{2\sigma_z^2}} \right] = \log \frac{\sqrt{\sigma_x^2+\sigma_z^2}}{\sigma_z} + \frac{1}{2\ln 2}\left(\frac{y^2}{\sigma_x^2+\sigma_z^2} - \frac{z^2}{\sigma_z^2} \right)$$

$$I(X,Y) = M\left[\log \frac{f(X,Y)}{f_1(X)\cdot f_2(Y)} \right] = \log \frac{\sqrt{\sigma_x^2+\sigma_z^2}}{\sigma_z} + \frac{1}{2\ln 2}\left(\frac{M[Y^2]}{\sigma_x^2+\sigma_z^2} - \frac{M[Z^2]}{\sigma_z^2} \right)$$

Т.к. $m_x = 0$ и $m_z = 0$, то $m_y = 0$, тогда
$M[Y^2] = M[(Y-m_y)^2] = D[Y] = \sigma_y^2 = \sigma_x^2 + \sigma_z^2$
$M[Z^2] = M[(Z-m_z)^2] = D[Z] = \sigma_z^2$

$$I(X,Y) = \log \frac{\sqrt{\sigma_x^2+\sigma_z^2}}{\sigma_z}$$

Если $\sigma_x = 5$, $\sigma_z = 2$, то $I(X,Y) = \log \frac{\sqrt{29}}{2} = 1,42$ бит.

Глава 4.
Приложение теории информации к задачам передачи сообщений

1. Виды информации

Информация может быть двух видов: дискретная (цифровая) или непрерывная (аналоговая). Дискретная информация характеризуется последовательными точными значениями некоторой величины, а непрерывная — непрерывным процессом изменения некоторой величины.

Дискретная информация удобнее для обработки человеком, но непрерывная информация чаще встречается в практической работе, поэтому необходимо уметь переводить непрерывную информацию в дискретную и наоборот.

При переводе непрерывной информации в дискретную важна так называемая *частота дискретизации* ν, определяющая период ($T = \frac{1}{\nu}$) между измерениями значений непрерывной величины (Рис. 2).

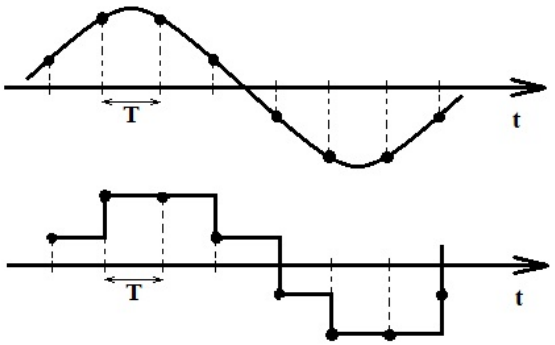


Рис.2. Перевод непрерывной информации в дискретную

Чем выше частота дискретизации, тем точнее происходит перевод непрерывной информации в дискретную. Но с ростом этой частоты растет размер дискретных данных, получаемых при таком переводе, и, следовательно, сложность их обработки, передачи и хранения. Однако, для повышения точности дискретизации необязательно безграничное увеличение ее частоты. Эту частоту разумно увеличивать только до предела, определяемого теоремой о выборках (теорема Котельникова).

Согласно этой теореме, аналоговый сигнал может быть сколь угодно точно дискретизирован, если не содержит частот, выше половины частоты дискретизации.

При преобразовании дискретной информации в непрерывную, определяющей является скорость этого преобразования: чем она выше, тем более высокочастотный аналоговый сигнал будет получен. Но чем более высокие частоты встречаются в полученном сигнале, тем сложнее с ним работать.

Устройства для преобразования непрерывной информации в дискретную называются АЦП (аналого-цифровой преобразователь), а устройства для преобразования дискретной информации в непрерывную — ЦАП (цифроаналоговый преобразователь).

2. Основные определения

При передаче сообщений по линиям связи всегда приходится пользоваться тем или иным *кодом*, т.е. представлением сообщения в виде ряда сигналов.

Кодирование — отображение состояния одной физической системы с помощью состояния некоторой другой. Рассмотрим наиболее простой случай кодирования — случай, когда обе системы X и Y (отображаемая и отображающая) имеют конечное число возможных состояний.

Пусть имеется некоторая система X (например, буква русского алфавита), которая может случайным образом принимать одно из состояний $x_1, x_2, ...x_n$ с вероятностями $p_1, p_2, ...p_n$.

Хотим закодировать ее при помощи другой системы Y, возможные состояния которой $y_1, y_2, ...y_m$.

Если $m < n$, то нельзя каждое состояние системы X закодировать с помощью одного состояния системы Y. В этом случае одно состояние системы X приходиться отображать с помощью определенной комбинации состояний системы Y.

Выбор таких комбинаций называется кодированием.

Коды различаются по числу *элементарных сигналов*, из которых формируются комбинации (т.е. по числу возможных состояний системы Y). Код с двумя элементарными сигналами (0 и 1) называется *двоичным* (т.е. $m = 2$). Будем рассматривать случай побуквенного кодирования, т.е. случай, когда каждая буква сообщения передается по отдельности. В этом случае каждой из n букв исходного алфавита сопоставляется *кодовое обозначение*, т.е. последовательность из двух элементарных сигналов (0 и 1). Для задания кода надо перечислить n таких последовательностей из цифр 0 и 1.

Одно и то же сообщение можно закодировать различными способами. Будем считать кодирование тем более выгодным, чем меньше элементарных сигналов приходится затратить на передачу сообщения.

После того, как сообщение передано, его надо однозначно *декодировать*. Для этого необходимо знать, где кончается кодовое обозначение одной буквы и начинается другое. Для этой цели можно ввести специальный разделительный знак («запятую») или использовать кодовые обозначения фиксированной длины (*равномерные коды*). Посмотрим, как можно составить неравномерные коды, не содержащие запятые.

Неравномерный код может быть однозначно декодирован, если никакое кодовое обозначение не совпадает с началом какого-либо другого более длинного кодового обозначения.

Например, если 011 - кодовое обозначение какой-нибудь буквы, то не может быть уже букв с кодовым обозначением 01 или 01100.

Коды, удовлетворяющие указанному условию, называют *мгновенно декодируемыми* (префиксными). Будем рассматривать только такие коды.

3. Экономность кода. Наилучший равномерный код

Будем измерять выгодность (экономность) данного двоичного кода при помощи максимального числа элементарных сигналов, необходимых для передачи одной буквы.

Наилучший равномерный код

Число a цифр в десятичной записи числа b определяется неравенством:

$$10^{a-1} \leqslant b < 10^a$$

т.е. числа в промежутке от $10^{2-1} = 10$ до $10^2 - 1 = 99$ будут двузначными.

Аналогично, в двоичной системе счисления число a «цифр» в записи числа b определяется неравенством:

$$2^{a-1} \leqslant b < 2^a$$

Если $a = 2$, то это будут числа от $2^1 = 2$ до $2^2 - 1 = 3$. Т.е. **10**, **11**.

Если $a = 3$, то это будут числа от $2^2 = 4$ до $2^3 - 1 = 7$. Т.е. **100**, **101**, **110**, **111** — трехзначные.

Если выписать первые b целых чисел, начиная с 0 (т.е. 0,1,2,...$b-1$), то окажется, что двоичная запись всех этих чисел содержит не более a знаков. Добавим в начало двоичной записи всех менее, чем a-значных чисел некоторое число нулей, тогда получим равномерный двоичный код для b-буквенного алфавита с минимальной возможной длиной кодовых обозначений.

Пример 26. Пусть имеется алфавит из 10 букв. Построить равномерный двоичный код с минимальной длиной кодовых обозначений.

Решение. Числа от 0 до 9 в двоичной системе счисления: 0, 01, 10, 11, 100, 101, 110, 111, 1000, 1001.

Дополним первые 8 чисел нулями до четырех знаков. Получим 0000, 0001, 0010, 0011, 0100, 0101, 0111, 1000, 1001.

На изображение одной буквы тратится не менее $\log n = \log 10 = 3,32$ элементарных сигнала, а именно **4**.

Пример 27. Закодировать равномерным двоичным кодом буквы русской азбуки: а,б,в,...э,ю,я,_ (пробел). Всего 32 буквы (ь=ъ, е=ё).

Решение. Не меняя порядок букв, занумеруем их, приписав каждой букве числа от 0 до 31. Затем переведем нумерацию в двоичную систему и дополним, где надо в начале нули. Получим: а∼00000; б∼00001; в∼00010; ... я∼11110; _(пробел)∼11111.

При таком кодировании, на изображение одной буквы тратится $\log n = \log 32 = 5$ элементарных сигналов.

С другой стороны, код будет самым экономным, когда каждый элементарный сигнал будет передавать максимальную информацию.

Рассмотрим элементарный сигнал, как физическую систему с двумя возможными состояниями: 0 и 1.

0	1
p_1	p_2

Информация, которую дает этот сигнал равна энтропии этой системы и максимальна в случае, когда оба состояния равновероятны. Тогда элементарный сигнал будет передавать 1 бит информации.

Следовательно, для оптимального кодирования необходимо, чтобы элементарные сигналы в закодированном сообщении встречались в среднем одинаково часто. Этой цели удается достигнуть, используя методы кодирования Шеннона-Фано и Хафмена.

4. Коды Шеннона-Фано и Хафмена

При составлении кода Шеннона-Фано и кода Хафмена основное значение играет *вероятностное* среднее значение числа элементарных сигналов, приходящихся на одну букву сообщения. Рассмотрим сообщение, записанное при помощи n «букв», частоты появления которых на *любом* месте сообщения характеризуются вероятностями $p_1, p_2, ...p_n$ $(\sum p_i = 1)$. При этом, вероятность p_i появления i-й буквы на любом месте сообщения не зависит от того, какие буквы стояли на всех предыдущих местах.

Среднее число двоичных элементарных сигналов, приходящихся в закодированном сообщении на одну букву исходного сообщения, не может

быть меньше величины $H = -p_1 \log p_1 - ... - p_n \log p_n$, где H — энтропия опыта, состоящего в распознавании одной буквы текста (т.е. энтропия одной буквы).

Следовательно, при любом методе кодирования для записи длинного сообщения из K букв требуется не меньше, чем KH двоичных знаков. Это следует из того, что информация, содержащаяся в отрывке текста, содержащем K букв, равна KH (в случае, если отдельные буквы независимые).

С другой стороны, информация, содержащаяся в одном элементарном сигнале (0 или 1), не может превосходить одного бита. Если вероятности $p_1, ...p_n$ не все равны между собой, то $H < \log n$.

Следовательно, если учитывать частоты появления букв в тексте, то можно построить код более экономный, чем наилучший равномерный код, требующий не менее $K \log n$ двоичных знаков для записи текста из K букв.

4.1. Код Шеннона-Фано

Расположим все кодируемые буквы в один столбик в порядке убывания вероятностей. Разделим их на две приблизительно равновероятные группы. Для первой группы символов на первом месте комбинации ставим 0, для второй группы — 1. Далее каждая группа снова делится на две приблизительно равновероятные подгруппы. Для символов первой подгруппы на первом месте ставится 0, для второй — 1, и т.д.

Основной принцип этого метода кодирования заключается в том, что каждая цифра кодового обозначения принимает оба возможных для нее значения (0 и 1) по возможности с одинаковой вероятностью. Никакое кодовое обозначение при таком методе не может оказаться началом более длинного кодового обозначения. Существенно, также, что буквам, имеющим большую вероятность, соответствуют более короткие кодовые обозначения.

Пример 28. Имеется алфавит, содержащий 9 букв. Частоты появления букв равны 0,3; 0,2; 0,15; 0,1; 0,1; 0,05; 0,05; 0,03; 0,02. Все буквы независимы. Составить кодовые обозначения по методу Шеннона-Фано.

Решение.

Среднее число элементарных сигналов, приходящихся на одну букву сообщения, равно

№ буквы	1	2	3	4	5	6	7	8	9
Вероятность	0,3	0,2	0,15	0,1	0,1	0,05	0,05	0,03	0,02
Количество сигналов	2	2	3	3	3	4	5	6	6

$0,3 \cdot 2 + 0,2 \cdot 2 + 0,15 \cdot 3 + 0,1 \cdot 3 + 0,1 \cdot 3 + 0,05 \cdot 4 + 0,05 \cdot 5 + 0,03 \cdot 6 + 0,02 \cdot 6 = 2.8$

При этом энтропия одной буквы передаваемого сообщения равна:

$$H = 0.5211+0.4644+0.4105+2\cdot 0.3322+2\cdot 0.2161+0.1518+0.1129 = 2.7573$$

№	Вероятность	Разбиение на подгруппы	Кодовые обозначения
1	0,3	I I	**11**
2	0,2	II	**10**
3	0,15	I I I	**011**
4	0,1	II	**010**
5	0,1	I II	**001**
6	0,05	I II I	**0001**
7	0,05	II II I	**00001**
8	0,03	II II I I	**000001**
9	0,02	II	**000000**

Следовательно, некоторые кодовые обозначения могут иметь весьма значительную длину, но среднее значение длины такого обозначения оказывается лишь немногим больше H.

Замечание. В рассмотренном выше примере наилучший равномерный код для 10-буквенного алфавита состоит из 4-значных кодовых обозначений ($2^{4-1} < 10 < 2^4$).

Пример 29. Применить метод Шеннона-Фано к примеру кодирования букв русского алфавита. См. пример 27.

Решение. В русском языке буквы «а», «о», «е» встречаются чаще, а буквы «щ», «ф» реже.

В таблице 1 приведены частоты букв русского алфавита и кодовые обозначения, построенные методом Шеннона-Фано.

Средняя информация, содержащаяся в одной букве передаваемого текста равна энтропии одной буквы:

$$H = -\sum_{i=1}^{32} p_i \log p_i = \eta(0,145) + \eta(0,095) + ... + \eta(0,002) \approx 4,42\text{бит}$$

Среднее число элементарных сигналов, приходящихся на одну букву: $n_{\text{ср}} = 3 \cdot 0,145 + 3 \cdot 0,095 + 4 \cdot 0,074 + ... + 9 \cdot 0,002 = 4,45\text{бит}$

Средняя информация, приходящаяся на один элементарный символ равна: $I_{1c} = \frac{H}{n_{\text{ср}}} = \frac{4,42}{4,45} \approx 0,994\text{бит}$

Следовательно, информация на один символ близка к своему верхнему пределу 1, а выбранный код близок к оптимальному.

При использовании наилучшего равномерного кода информация на один элементарный сигнал равна $I_{1c} = \frac{4,42}{5} \approx 0,884\text{бит}$

Таблица 1. Частоты букв русского алфавита; кодовые обозначения, построенные методом Шеннона-Фано

Буквы	Частоты букв	Кодовые обозначения
пробел	0,145	000
о	0,095	001
е	0,074	0100
а	0,064	0101
и	0,064	0110
т	0,056	0111
н	0,056	1000
с	0,047	1001
р	0,041	10100
в	0,039	10101
л	0,036	10110
к	0,029	10111
м	0,026	11000
д	0,026	110010
п	0,024	110011
у	0,021	110100
я	0,019	110110
ы	0,016	110111
з	0,015	111000
ь,ъ	0,015	111001
б	0,015	111010
г	0,014	111011
ч	0,013	111100
й	0,010	1111010
х	0,009	1111011
ж	0,008	1111100
ю	0,007	1111101
ш	0,006	11111100
ц	0,004	11111101
щ	0,003	11111110
э	0,003	111111110
ф	0,002	111111111

4.2. Код Хафмена

Рассмотрим алфавит X, содержащий буквы $x_1, ...x_n$, вероятности появления которых в алфавите $p_1, ...p_n$. Расположим буквы в порядке убывания вероятностей.

В алфавите X две наименее вероятные буквы x_{n-1}, x_n объединим в одну, получим алфавит X_1, содержащий буквы $x_1, x_2, ...x_{n-2}, y$. Алфавит X_1 получен из X с помощью однократного сжатия.

Расположим буквы алфавита X_1 в порядке убывания вероятностей и подвергнем его сжатию. Получим алфавит X_2. Алфавит X_2 получен из X с помощью двукратного сжатия.

После $n-2$ кратного сжатия придем к алфавиту X_{n-2}, содержащему две буквы.

Припишем двум буквам последнего алфавита X_{n-2} кодовые обозначения 1 и 0.

Далее, если кодовые обозначения уже приписаны всем буквам алфавита X_j, то буквам предыдущего алфавита X_{j-1}, сохранившимся и в алфавите X_j, припишем те же кодовые обозначения, которые они имели в алфавите X_{j-1}. А двум буквам x' и x'' алфавите X_j, слившимся в одну букву y алфавита X_{j-1}, припишем обозначения, получающиеся из кодового обозначения буквы y, добавлением цифр 1 и 0 в конце.

Пример 30. Имеется алфавит, содержащий 9 букв. Частоты появления букв равны 0,3; 0,2; 0,15; 0,1; 0,1; 0,05; 0,05; 0,03; 0,02. Все буквы независимы. Составить кодовые обозначения по методу Хафмена.

Решение. Расположим буквы исходного алфавита в порядке убывания вероятностей и подвергнем его сжатию. Получим алфавиты X_1,...X_7.

№	X	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	0,3	0,3	0,3	0,3	0,3	0,3	0,4	0,6
2	0,2	0,2	0,2	0,2	0,2	0,3	0,3	0,4
3	0,15	0,15	0,15	0,15	0,2	0,2	0,3	
4	0,1	0,1	0,1	0,15	0,15	0,2		
5	0,1	0,1	0,1	0,1	0,15			
6	0,05	0,05	0,1	0,1				
7	0,05	0,05	0,05					
8	0,03	0,05						
9	0,02							

Припишем двум буквам алфавита X_7 кодовые обозначения 1 и 0. Переходя от X_7 к X, получим кодовые обозначения для исходного алфавита.

№	X		X_1		X_2		X_3		X_4		X_5		X_6		X_7	
1	0,3	11	0,3	11	0,3	11	0,3	11	0,3	11	0,3	11	0,4	0	0,6	1
2	0,2	01	0,2	01	0,2	01	0,2	01	0,2	01	0,3	10	0,3	11	0,4	0
3	0,15	101	0,15	101	0,15	101	0,15	101	0,2	00	0,2	01	0,3	10		
4	0,1	001	0,1	001	0,1	001	0,15	100	0,15	101	0,2	00				
5	0,1	000	0,1	000	0,1	000	0,1	001	0,15	100						
6	0,05	1000	0,05	1000	0,1	1001	0,1	000								
7	0,05	10011	0,05	10011	0,05	1000										
8	0,03	100101	0,05	10010												
9	0,02	100100														

Среднее число элементарных сигналов, приходящихся на одну букву $n_{cp} = 2.8$. Энтропия $H = 2,7573$.

Кодирование по методу Хафмена, так же, как и по методу Шеннона-Фано не является однозначно определенной процедурой.

5. Блоковые коды

Кодирование «по буквам» не является самым экономичным. При передаче длинных сообщений более выгодно использовать блоковые коды. Так, в русском языке между соседними буквами осмысленного текста есть зависимость. Например, после мягкого знака не может быть «и» или «а», а после шипящих «я» или «ю» и т.д. Можно объединять в блоки такие сочетания, как «ает», «тся», «ние» и т.п.

Т.к. при объединении зависимых систем $H(X,Y) \leqslant H(X) + H(Y)$, то кодирование сразу крупных блоков позволяет добиться того, что среднее число элементарных сигналов, приходящихся на одну букву сообщения будет сколь угодно близким к энтропии одной буквы.

Пример 31. Алфавит состоит из двух букв: p(X)=0,8, P(Y)=0,2. Составить кодовые обозначения методом Шеннона-Фано для: (1) побуквенного кодирования алфавита, (2) кодирования блоков из двух букв, (3) кодирования блоков из трех букв.

Решение. (1)Применим метод Шеннона-Фано к побуквенному кодированию исходного алфавита:

	Вероятность	Кодовые обозначения
X	0,8	**1**
Y	0,2	**0**

Энтропия одной буквы: $H = \eta(0,8) + \eta(0,2) = 0,7219$.

Среднее число элементарных сигналов: $n_{cp1} = 0,8 \cdot 1 + 0,2 \cdot 1 = 1$.

(2) Применим метод Шеннона-Фано к кодирования всевозможных двубуквенных комбинаций:

№	Вероятность	Кодовые обозначения
XX	0,64	**1**
XY	0,16	**01**
YX	0,16	**001**
YY	0,04	**000**

Среднее число элементарных сигналов, приходящееся на одну двубуквенную комбинацию: $n_{ср2} = 0,64 \cdot 1 + 0,16 \cdot 2 + 0,16 \cdot 3 + 0,04 \cdot 3 = 1,56$. Среднее число элементарных сигналов, приходящееся на одну букву: $n_{ср} = \frac{n_{ср2}}{2} = 0,78$.

Число 0,78 более приближено к энтропии одной буквы $H = 0,7219$ по сравнению с числом 1 в случае побуквенного кодирования.

(3) Применим метод Шеннона-Фано к кодирования всевозможных трехбуквенных комбинаций:

№	Вероятность	Кодовые обозначения
XXX	0,512	**1**
XXY	0,128	**011**
XYX	0,128	**010**
YXX	0,128	**001**
YYX	0,032	**00011**
YXY	0,032	**00010**
XYY	0,032	**00001**
YYY	0,008	**00000**

Среднее число элементарных сигналов, приходящееся на одну трехбуквенную комбинацию: $n_{ср3} = 2,184$. Среднее число элементарных сигналов, приходящееся на одну букву: $n_{ср} = \frac{n_{ср3}}{3} = 0,7280$.

Число 0,7280 еще более приближено к энтропии одной буквы $H = 0,7219$.

6. Обобщение для k-ичных кодов

Пусть для составления кода используется k элементарных сигналов. Тогда для составления кода Шеннона-Фано надо разбивать группы символов не на две, а на k частей.

При составлении кода Хафмена надо использовать операцию сжатия алфавита, при которой каждый раз сливаются не две, а k букв исходного

алфавита, имеющих наименьшие вероятности. При этом, чтобы в результате сжатий получить алфавит из k букв, необходимо, чтобы число n букв исходного алфавита удовлетворяло условию: $n = k+z(k-1)$, где z — целое.

При любом методе кодирования, использующем k-ичный код, среднее число элементарных сигналов, приходящихся на одну букву сообщения, никогда не может быть меньше отношения $\frac{H}{\log k}$, где H — энтропия одной буквы сообщения, $\log k$ — максимальная информация, которую содержит один элементарный сигнал. Однако, оно может быть сделано сколь угодно близким к этой величине, если кодировать достаточно длинные блоки.

Пример 32. Алфавит состоит из восьми букв, вероятности которых 0,3;0,2;0,15; 0,12;0,1;0,05;0,05;0,03. Составить троичный алгоритм Хафмена.

Решение. Для выполнения условия $n = k+z(k-1)$ при $k = 3$ необходимо, чтобы число букв исходного алфавита было равно $n = 9$. Добавим еще одну букву, вероятность которой 0.

№	X		X_1		X_2		X_3	
1	0,3	1	0,3	1	0,3	1	0,47	0
2	0,2	00	0,2	00	0,23	2	0,3	1
3	0,15	01	0,15	01	0,2	00	0,23	2
4	0,12	02	0,12	02	0,15	01		
5	0,1	20	0,1	20	0,12	02		
6	0,05	22	0,08	21				
7	0,05	210	0,05	22				
8	0,03	211						
9	0	–						

Среднее число элементарных сигналов $n_{cp} = 1,78$ не будет меньше отношения $\frac{H}{\log k} = \frac{0{,}5211+0{,}4644+0{,}4105+0{,}3671+0{,}3322+0{,}2161\cdot 2+0{,}1518}{\log 3} = \frac{2{,}6793}{1{,}5849} = 1,6904$.

7. Словарно-ориентированные методы кодирования. Методы Лемпела-Зива

Преимущество словарных алгоритмов состоит в том, что они позволяют кодировать последовательности символов разной длины. Первым был опубликован алгоритм LZ77 (1977 г). После он был многократно модифицирован. Многие модификации получали название LZX, где X — первая буква имени автора модификации. Популярность алгоритмов LZ объясняется их простотой при высокой эффективности сжатия.

Ниже рассмотрим два алгоритма: LZ77, LZ78.

7.1. Алгоритм LZ77

Алгоритм LZ77 использует «скользящее» по сообщению окно, разделенное на две неравные части. Первая, большая по размеру, включает уже просмотренную часть сообщения. Вторая, намного меньшая, является буфером, содержащим еще незакодированные символы входного потока. Алгоритм пытается найти в словаре (большей части окна) фрагмент, совпадающий с содержимым буфера.

Алгоритм LZ77 выдает коды, состоящие из трех элементов: (1)смещение в словаре относительно его начала строки, совпадающего с началом содержимого буфера; (2)длина этой строки; (3)первый символ буфера, следующий за совпадением.

Пример 33. Закодировать по алгоритму LZ77 строку «ЗЕЛЕНАЯ_ЗЕЛЕНЬ_ЗЕЛЕНЕЕТ». Размер буфера 7 символов, а словаря — 9 символов.

Решение.

Словарь (9)									Буфер (7)							Код
1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	
.	.	.	.	.	.	.	.	.	З	Е	Л	Е	Н	А	Я	$\langle 0,0,\text{З}\rangle$
.	.	.	.	.	.	.	.	З	Е	Л	Е	Н	А	Я	_	$\langle 0,0,\text{Е}\rangle$
.	.	.	.	.	.	.	З	Е	Л	Е	Н	А	Я	_	З	$\langle 0,0,\text{Л}\rangle$
.	.	.	.	.	.	З	**Е**	Л	**Е**	Н	А	Я	_	З	Е	$\langle 7,1,\text{Н}\rangle$
.	.	.	.	З	Е	Л	Е	Н	А	Я	_	З	Е	Л	Е	$\langle 0,0,\text{А}\rangle$
.	.	.	З	Е	Л	Е	Н	А	Я	_	З	Е	Л	Е	Н	$\langle 0,0,\text{Я}\rangle$
.	.	З	Е	Л	Е	Н	А	Я	_	З	Е	Л	Е	Н	Ь	$\langle 0,0,\text{\_}\rangle$
.	**З**	**Е**	**Л**	**Е**	**Н**	А	Я	_	**З**	**Е**	**Л**	**Е**	**Н**	Ь	_	$\langle 1,5,\text{Ь}\rangle$
А	Я	_	**З**	**Е**	**Л**	**Е**	**Н**	Ь	_	**З**	**Е**	**Л**	**Е**	**Н**	Е	$\langle 2,6,\text{Е}\rangle$
Н	Ь	_	З	**Е**	Л	Е	Н	Е	**Е**	Т	.	.	.	.	.	$\langle 4,1,\text{Т}\rangle$

Длина кода вычисляется следующим образом: длина строки не может быть больше размера буфера +1, а смещение не может быть больше размера словаря. Следовательно, длина двоичного кода смещения будет округленным в большую сторону log(размер словаря), а длина двоичного кода для длины строки будет округленным в большую сторону log(размер буфера +1). А символ кодируется 8 битами (например, ASCII+).

Длина полученного кода: $10\cdot(\log(9)+\log(7+1)+8)\approx 10\cdot(4+3+8)$, т.е. 150 бит.

Длина исходной стоки 23 символа, т.е. $23\cdot 8=184$ бит.

Пример 34. Длина словаря 9 символов. Распаковать сообщение, сжатое LZ77: ⟨0, 0, З⟩, ⟨0, 0, Е⟩, ⟨0, 0, Л⟩, ⟨7, 1, Н⟩, ⟨0, 0, А⟩, ⟨0, 0, Я⟩, ⟨0, 0, _⟩, ⟨1, 5, Ь⟩, ⟨2, 6, Е⟩, ⟨4, 1, Т⟩.

Решение.

Входной код	Печать	Словарь (9)								
		1	2	3	4	5	6	7	8	9
⟨0, 0, З⟩	З	.	.	.	.	.	.	.	.	З
⟨0, 0, Е⟩	Е	.	.	.	.	.	.	.	З	Е
⟨0, 0, Л⟩	Л	.	.	.	.	.	.	З	**Е**	Л
⟨7, 1, Н⟩	ЕН	.	.	.	.	З	Е	Л	Е	Н
⟨0, 0, А⟩	А	.	.	.	З	Е	Л	Е	Н	А
⟨0, 0, Я⟩	Я	.	.	З	Е	Л	Е	Н	А	Я
⟨0, 0, _⟩	_	.	**З**	**Е**	**Л**	**Е**	**Н**	А	Я	_
⟨1, 5, Ь⟩	ЗЕЛЕНЬ	А	Я	_	**З**	**Е**	**Л**	**Е**	**Н**	Ь
⟨2, 6, Е⟩	_ЗЕЛЕНЕ	Н	Ь	_	З	**Е**	Л	Е	Н	Е
⟨4, 1, Т⟩	ЕТ	_	З	Е	Л	Е	Н	Е	Е	Т

7.2. Алгоритм LZ78

Алгоритм LZ77 не позволяет кодировать строки, отстоящие друг от друга на расстояние большее длины словаря, а длина строки, которую можно закодировать ограничена размером буфера. Если увеличивать размеры словаря и буфера, то это снизит эффективность кодирования.

LZ78 не использует «скользящее» окно, он хранит словарь из уже просмотренных фраз. При старте алгоритма этот словарь содержит одну пустую строку длины ноль. Алгоритм считывает символы сообщения до тех пор, пока накапливаемая строка входит целиком в одну из фраз словаря. Как только эта строка перестанет соответствовать хотя бы одной фразе словаря, алгоритм генерирует код, состоящий из индекса строки в словаре, которая до последнего введенного символа содержала входную строку, и символа, нарушившего совпадение. Затем в словарь добавляется введенная строка. Если словарь уже заполнен, то из него предварительно удаляют менее всех используемую в сравнениях фразу.

Длина полученного двоичного кода будет округленным в большую сторону log(размер словаря) +8 (8 битами кодируются символы, например, ASCII+).

Пример 35. Закодировать по алгоритму LZ78 строку «ЗЕЛЕНАЯ_ЗЕЛЕНЬ_ЗЕЛЕНЕЕТ», используя словарь длиной 16 фраз.

Решение.

Входная фраза (в словарь)	Код	Позиция словаря
«»		0
З	$\langle 0, \text{З}\rangle$	1
Е	$\langle 0, \text{Е}\rangle$	2
Л	$\langle 0, \text{Л}\rangle$	3
ЕН	$\langle 2, \text{Н}\rangle$	4
А	$\langle 0, \text{А}\rangle$	5
Я	$\langle 0, \text{Я}\rangle$	6
_	$\langle 0, \_\rangle$	7
ЗЕ	$\langle 1, \text{Е}\rangle$	8
ЛЕ	$\langle 3, \text{Е}\rangle$	9
Н	$\langle 0, \text{Н}\rangle$	10
Ь	$\langle 0, \text{Ь}\rangle$	11
_З	$\langle 7, \text{З}\rangle$	12
ЕЛ	$\langle 2, \text{Л}\rangle$	13
ЕНЕ	$\langle 4, \text{Е}\rangle$	14
ЕТ	$\langle 2, \text{Т}\rangle$	15

Указатель на любую фразу такого словаря — это число от 0 до 15, для его кодирования достаточно $\log(16) = 4$ бит.

Длина полученного кода равна: $15 \cdot (4 + 8) = 180$ бит.

Пример 36. Длина словаря 16 фраз. Распаковать сообщение, сжатое LZ78: $\langle 0, \text{З}\rangle$, $\langle 0, \text{Е}\rangle$, $\langle 0, \text{Л}\rangle$, $\langle 2, \text{Н}\rangle$, $\langle 0, \text{А}\rangle$, $\langle 0, \text{Я}\rangle$, $\langle 0, \_\rangle$, $\langle 1, \text{Е}\rangle$, $\langle 3, \text{Е}\rangle$, $\langle 0, \text{Н}\rangle$, $\langle 0, \text{Ь}\rangle$, $\langle 7, \text{З}\rangle$, $\langle 2, \text{Л}\rangle$, $\langle 4, \text{Е}\rangle$, $\langle 2, \text{Т}\rangle$.

Решение.

Входной код	Печать (словарь)	Позиция словаря
«»		0
⟨0, З⟩	З	1
⟨0, Е⟩	Е	2
⟨0, Л⟩	Л	3
⟨2, Н⟩	ЕН	4
⟨0, А⟩	А	5
⟨0, Я⟩	Я	6
⟨0, _⟩	_	7
⟨1, Е⟩	ЗЕ	8
⟨3, Е⟩	ЛЕ	9
⟨0, Н⟩	Н	10
⟨0, Ь⟩	Ь	11
⟨7, З⟩	_З	12
⟨2, Л⟩	ЕЛ	13
⟨4, Е⟩	ЕНЕ	14
⟨2, Т⟩	ЕТ	15

8. Сжатие информации с потерями

Выше рассмотренные алгоритмы кодирования информации обеспечивали возможность полного восстановления исходных данных. Но иногда для повышения степени сжатия можно отбрасывать часть исходной информации, т.е. производить сжатие с потерями.

Сжатие с потерями используется в основном для трех видов данных: полноцветная графика, звук и видеоинформация.

Основная идея сжатия графической информации с потерями заключается в следующем. Каждая точка в картинке характеризуется тремя равноважными атрибутами: яркостью, цветом и насыщенностью. Но глаз человека воспринимает эти атрибуты не как равные. Глаз воспринимает полностью только информацию о яркости в гораздо меньшей степени о цвете и насыщенности, что позволят отбрасывать часть информации о двух последних атрибутах без потери качества изображения.

Сжатие видеоинформации основано на том, что при переходе от одного кадра к другому на экране почти ничего не меняется. Таким образом, сжатая видеоинформация представляет собой запись некоторых базовых

кадров и последовательности изменений в них. При этом часть информации может отбрасываться. Сжатую таким образом информацию можно дальше сжимать и другими методами.

9. Общая схема передачи сообщений по линии связи. Пропускная способность линии связи

Общая схема передачи сообщений представлена на рисунке. Сообщения X_1 и Y_1 на входе и на выходе записываются на одном и том же языке и могут отличаться лишь в результате искажений. Сигналы X и Y на входе и на выходе представляют собой последовательности элементарных сигналов.

Если по линии связи за единицу времени можно передать L элементарных сигналов, принимающих k различных значений, то скорость передачи сообщений по такой линии не может быть больше, чем

$$v = \frac{L}{\frac{H}{\log k}} = \frac{L \log k}{H} \text{букв в единицу времени}$$

Величина $C = L \log k$ (числитель) зависит от самой линии связи. Величина H (знаменатель) характеризует передаваемое сообщение.

Величина C — пропускная способность линии связи (емкость линии связи), показывает наибольшее количество единиц информации, которое можно передать по линии в единицу времени (т.к. один элементарный сигнал может содержать самое большее $\log k$ единиц информации).

Если по линии связи передается два элементарных сигнала ($k = 2$), то пропускная способность линии связи измеряется в количестве переданных за одну секунду бит или в бодах (baud): 1 бод=1 бит/сек.

Пропускную способность канала связи без помех можно приблизительно вычислить, зная максимальную частоту волновых процессов, допустимых в этом канале. Можно считать, что максимальная скорость передачи данных может быть не меньше, чем эта частота. Например, при предельной частоте, равной 1000 Гц, можно обеспечить скорость передачи данных не меньше 1 Кбод. Примеры каналов связи и связанных с ними предельных частот: телеграф — 140 Гц, телефон — 3.1 КГц, короткие

волны (10-100 м)– 3-30 МГц, УКВ(1-10 м) — 30-300 МГц, спутник (сантиметровые волны) — до 30 ГГц, оптический (инфракрасный диапазон) — 0.15–400 ТГц, оптический (видимый свет) — 400–700 ТГц, оптический (ультрафиолетовый диапозон) — 0.7—1.75 ПГц.

Глава 5.
Передача сообщений при наличии помех

На практике процесс передачи информации почти всегда сопровождается помехами. Наличие помех приводит к потере информации. Чтобы на приемном конце линии точно восстановить передаваемое сообщение, нужно принимать меры. Например, ввести «избыточность» в передаваемое сообщение (передавать сообщение несколько раз или передавать сообщение «по буквам», когда вместо каждой буквы передается знакомое слово или имя).

Рассмотрим простейшую схему дискретной линии связи, т.е. предполагаем, что по линии передается лишь конечное число различных «элементарных сигналов» постоянной длительности. Наличие помех в такой линии приводит к путанице, т.е. элементарный сигнал одного типа может быть ошибочно принят за сигнал другого типа.

1. Математическое описание линии связи с помехами

Пусть линия связи использует n элементарных сигналов $x_1, x_2, ...x_n$. Из-за наличия помех сигнал x_i $(i = 1, 2, ...n)$ может быть иногда принят на приемном конце за какой-то другой x_j $(x_i \neq x_j)$. Т.е., пусть

$p(x_1/x_1)$ – вероятность того, что передавая сигнал x_1, на приемном конце получен правильный сигнал x_1.

$p(x_2/x_1), p(x_3/x_1), ...p(x_n/x_1)$ – вероятность, что сигнал x_1 будет на приемном конце расшифрован как $x_2, x_3, ...x_n$.

$p(x_1/x_2), p(x_2/x_2), ...p(x_n/x_2)$ – вероятность получения на приемном конце сигналов $x_1, x_2, ...x_n$, если на самом деле передавался сигнал x_2.

Получим таблицу вероятностей, которые статистически характеризуют помехи:

$$\begin{array}{cccc} p(x_1/x_1) & p(x_2/x_1) & ... & p(x_n/x_1) \\ p(x_1/x_2) & p(x_2/x_2) & ... & p(x_n/x_2) \\ ... & ... & ... & ... \\ p(x_1/x_n) & p(x_2/x_n) & ... & p(x_n/x_n) \end{array}$$

Таким образом, чтобы описать линию связи с помехами, надо задать (1) n – число различных элементарных сигналов, (2)L – скорость передачи элементарных сигналов, (3) n^2 неотрицательных чисел $p(x_i/x_j)$, удовлетворяющих n условиям $p(x_1/x_i) + p(x_2/x_i) + ... + p(x_n/x_i) = 1$ $(i = 1, 2, ...n)$, характеризующих влияние помех.

Пусть теперь на приемном конце линии помехи так исказили передаваемый сигнал, что его нельзя отождествить ни с одним из n используемых

сигналов x_i. Тогда будем считать, что на приемном конце линии получено не n, а m каких-то других сигналов (где $n > m$ или $n < m$ или $n = m$). $y_1, y_2, ...y_m$ – полученные сигналы, все или частично отличающиеся от передаваемых $x_1, x_2, ...x_n$.

Тогда помехи статистически характеризуются nm неотрицательными числами:

$$\begin{array}{cccc} p(y_1/x_1) & p(y_2/x_1) & ... & p(y_m/x_1) \\ p(y_1/x_2) & p(y_2/x_2) & ... & p(y_m/x_2) \\ ... & ... & ... & ... \\ p(y_1/x_n) & p(y_2/x_n) & ... & p(y_m/x_n) \end{array}$$

где $p(y_j/x_i)$ – вероятность того, что на приемном конце принят сигнал y_j, если на самом деле передан x_i.

В этом случае линия характеризуется числами n, m, L и nm числами $p(y_j/x_i)$ $(p(y_1/x_i) + p(y_2/x_i) + ... + p(y_m/x_i) = 1)$.

Пример 37. Линия связи использует четыре различных элементарных сигнала. Вероятность безошибочной передачи сигнала равна 0,73. В случае ошибки вероятность стирания равна 0,003, а если стирания не произошло, то каждый из сигналов с равной вероятностью воспринимается как любой из трех, отличных от него. Построить таблицу вероятностей приема.

Решение. Если стирания не произошло, то каждый из сигналов на приемном конце линии будет принят как любой из трех, отличных от него с вероятностью $\frac{1-0,73-0,03}{3} = 0,08$. Тогда,

$$\begin{array}{ccccc} 0,73 & 0,08 & 0,08 & 0,08 & 0,003 \\ 0,08 & 0,73 & 0,08 & 0,08 & 0,003 \\ 0,08 & 0,08 & 0,73 & 0,08 & 0,003 \\ 0,08 & 0,08 & 0,08 & 0,73 & 0,003 \end{array}$$

2. Пропускная способность линии связи с помехами

Рассмотрим сложную систему, состоящую из источника информации X, линии связи и приемника Y.

Источник информации представляет собой физическую систему X с n возможными состояниями $x_1, x_2, ...x_n$ с вероятностями $p_1, p_2, ...p_n$. Будем

рассматривать эти состояния, как элементарные сигналы, которые может передавать источник X по линии связи к приемнику Y.

Количество информации на один сигнал, которое дает источник, равно энтропии на один сигнал:

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i$$

При наличии ошибок количество информации, содержащееся в системе Y относительно X равно:

$$I(Y, X) = H(X) - H(X/Y)$$

Будем рассматривать $H(X/Y)$ как потерю информации на один элементарный сигнал из-за помех.

Определим пропускную способность линии связи с помехами, т.е. максимальное количество информации, которое способна передавать линия связи в единицу времени.

Пусть линия связи способна передавать L элементарных сигналов. В отсутствии помех пропускная способность линии была бы равна: $C = L \log n$.

При наличии помех пропускная способность линии определяется как

$$C = L \max I(Y, X)$$

где $\max I(Y, X)$ – максимальная информация, на один сигнал, которую может передать линия при наличии помех.

Вычисление этой величины является сложной задачей. Она зависит от того, как и с какими вероятностям искажаются сигналы, происходит ли их перепутывание и т.д. Но для простейших случаев можно рассчитать пропускную способность линии связи.

1. Линия связи использует n элементарных сигналов $x_1, x_2, ...x_n$. Сигналы на приемном конце линии $y_1, y_2, ...y_m$ совпадают с сигналами $x_1, x_2, ...x_n$ ($n = m$) и $p(y_j/x_i) = 1$ при $i = j$. Следовательно, $p(y_j/x_i) = 0$ при $i \neq j$. Т.е. всегда принимается тот же самый сигнал, который был передан (помехи не препятствуют или отсутствуют). Таблица вероятностей помех:

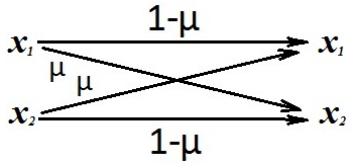
$$\begin{pmatrix} 1 & 0 & ... & 0 \\ 0 & 1 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 1 \end{pmatrix}$$

Здесь $H(X/Y) = 0$. Откуда, $\max I(Y/X) = \max H(X) = \log n$. Это наибольшее значение, когда все состояния передаваемого сигнала равновероятны ($p(x_1) = ... = p(x_n) = \frac{1}{n}$). В этом случае $C = L \log n$.

2. Линия связи может передавать два элементарных сигнала x_1 и x_2 в количестве L сигналов в единицу времени. В процессе передачи каждый сигнал с вероятностью μ может быть заменен на противоположный. Такую линию называют двоичной симметричной линией связи. Найти пропускную способность линии связи.

Таблица вероятностей помех:

$$\begin{pmatrix} 1-\mu & \mu \\ \mu & 1-\mu \end{pmatrix}$$



Определим максимальную информацию на один элементарный сигнал, которую может передавать линия связи. Пусть источник производит сигналы x_1 и x_2 с вероятностями p и $1-p$:

$X:$	x_1	x_2
	p	$1-p$

Энтропия источника $H(X) = -p \log p - (1-p) \log(1-p)$.

Информация на один элементарный сигнал: $I(X,Y) = H(Y) - H(Y/X)$.

Найдем $H(Y/X)$. Известно, что $H(Y/X) = pH(Y/x_1) + (1-p)H(Y/x_2)$.

Вычислим $H(Y/x_1)$. Предположим, что передан сигнал x_1. Найдем условные вероятности того, что при этом система Y находится в состоянии y_1 (т.е. получен сигнал x_1) и в состоянии y_2 (т.е. получен сигнал x_2).

$p(y_1/x_1) = 1-\mu$ – вероятность того, что сигнал не перепутан,

$p(y_2/x_1) = \mu$ – вероятность того, что сигнал перепутан.

Следовательно, $H(Y/x_1) = -\sum_{i=1}^{2} p(y_i/x_1) \log p(y_i/x_1) = -(1-\mu) \log(1-\mu) - \mu \log \mu$.

Пусть передан сигнал x_2. Вычислим $H(Y/x_2)$.

$p(y_1/x_2) = \mu$ (сигнал перепутан), $p(y_2/x_2) = 1-\mu$ (не перепутан).

$H(Y/x_2) = -\mu \log \mu - (1-\mu) \log(1-\mu)$

Следовательно, $H(Y/x_1) = H(Y/x_2) = -\mu \log \mu - (1-\mu) \log(1-\mu)$.

$H(Y/X) = pH(Y/x_1) + (1-p)H(Y/x_2) = H(Y/x_1)(1-p+p) = H(Y/x_1)$

$H(Y/X) = -\mu \log \mu - (1-\mu) \log(1-\mu) = \eta(\mu) + \eta(1-\mu)$

Условная энтропия зависит только от вероятности ошибки μ.

Найдем полную информацию, передаваемую одним сигналом:

$I(Y/X) = H(Y) - H(Y/X)$.

Пусть r – вероятность того, что на выходе появится символ x_1, т.е. система Y будет находиться в состоянии y_1.

$Y:$	y_1	y_2
	r	$1-r$

$H(Y) = -r\log r - (1-r)\log(1-r) = \eta(r) + \eta(1-r)$,

$I(Y/X) = H(Y) - H(Y/X) = (\eta(r) + \eta(1-r)) - (\eta(\mu) + \eta(1-\mu))$.

Информация $I(Y,X)$ максимальна, когда выражение $(\eta(r)+\eta(1-r))$ максимально. Это достигается, при $r = \frac{1}{2}$, т.е. когда на приемнике оба сигнала равновероятны.

Следовательно, $\max I(Y/X) = 1 - (\eta(\mu) + \eta(1-\mu))$.

Значит, пропускная способность линии связи:

$$C = L\left[1 - (\eta(\mu) + \eta(1-\mu))\right].$$

Пример 38. Определить пропускную способность линии связи с помехами, передающую 70 сигналов 0 и 1 в единицу времени, причем каждый из сигналов заменяется противоположным с вероятностью 0,05.

Решение. $C = L(1-(\eta(\mu)+\eta(1-\mu))) = 70(1-(0,2161)+0,0703) = 49,952 \approx 50$ бит в единицу времени.

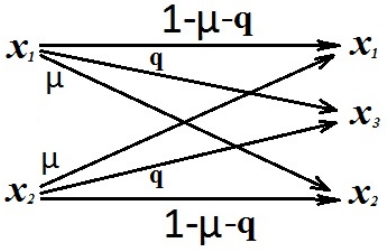
3. Линия связи может передавать два элементарных сигнала x_1 и x_2 в количестве L сигналов в единицу времени. В процессе передачи каждый сигнал с вероятностью μ может быть заменен на противоположный или он так искажен, что его невозможно узнать. В этом последнем случае будем считать, что принят какой-то новый сигнал x_3 (т.е. сигнал стерся или не допускает расшифровки). При этом вероятности стирания обоих сигналов x_1 и x_2 равны q. Такую линию называют двоичной симметричной линией со стиранием. Найти пропускную способность этой линии связи.

Таблица вероятностей помех:

$$\begin{pmatrix} 1-\mu-q & \mu & q \\ \mu & 1-\mu-q & q \end{pmatrix}$$

Определим максимальную информацию на один элементарный сигнал, которую может передавать линия связи. Пусть источник производит сигналы x_1 и x_2 с вероятностями p и $1-p$, соответственно:

$X:$	x_1	x_2
	p	$1-p$



Энтропия источника $H(X) = -p\log p - (1-p)\log(1-p)$.

Информация на один элементарный сигнал: $I(X,Y) = H(Y) - H(Y/X)$.

Найдем $H(Y/X)$. Известно, что $H(Y/X) = pH(Y/x_1) + (1-p)H(Y/x_2)$.

Вычислим $H(Y/x_1)$. Предположим, что передан сигнал x_1. Найдем условные вероятности того, что при этом система Y находится в состоянии y_1 (т.е. получен сигнал x_1), в состоянии y_2 (т.е. получен сигнал x_2) и в состоянии y_3 (т.е. сигнал или стерся, или не может быть расшифрован).

$p(y_1/x_1) = 1 - \mu - q$ – вероятность того, что сигнал не искажен,

$p(y_2/x_1) = \mu$ – вероятность того, что сигнал перепутан.

$p(y_3/x_1) = q$ – вероятность того, что произошло стирание сигнала.

Следовательно, $H(Y/x_1) = -\sum_{i=1}^{3} p(y_i/x_1)\log p(y_i/x_1) = -(1-\mu-q)\log(1-\mu-q) - \mu\log\mu - q\log q$.

Пусть передан сигнал x_2. Вычислим $H(Y/x_2)$.

$p(y_1/x_2) = \mu$ (сигнал перепутан), $p(y_2/x_2) = 1 - \mu - q$ (не искажен), $p(y_3/x_2) = q$ (стирание сигнала).

$H(Y/x_2) = -(1-\mu-q)\log(1-\mu-q) - \mu\log\mu - q\log q$

Следовательно, $H(Y/x_1) = H(Y/x_2) = -(1-\mu-q)\log(1-\mu-q) - \mu\log\mu - q\log q$.

$H(Y/X) = pH(Y/x_1) + (1-p)H(Y/x_2) = H(Y/x_1)(1-p+p) = H(Y/x_1)$

$H(Y/X) = -(1-\mu-q)\log(1-\mu-q) - \mu\log\mu - q\log q = \eta(1-\mu-q) + \eta(\mu) + \eta(q)$

Найдем полную информацию, передаваемую одним сигналом:

$I(Y/X) = H(Y) - H(Y/X)$.

Пусть r_1 – вероятность того, что на выходе появится сигнал x_1, т.е. система Y будет находиться в состоянии y_1; r_2 – вероятность того, что на выходе появится сигнал x_2. Вероятность того, что произойдет стирание сигнала известна и равна q.

$Y:$	y_1	y_2	y_3
	r_1	r_2	q

$H(Y) = -r_1\log r_1 - r_2\log r_2 - q\log q = \eta(r_1) + \eta(r_2) + \eta(q)$, где $r_1 + r_2 + q = 1$

$I(Y/X) = H(Y) - H(Y/X) = [\eta(r_1) + \eta(r_2) + \eta(q)] - [\eta(1-\mu-q) + \eta(\mu) + \eta(q)]$.

Информация $I(Y,X)$ будет максимальна, когда выражение $\eta(r_1) + \eta(r_2) + \eta(q)$ максимально. Это достигается, когда энтропия системы Y принимает наибольшее значение. В данном случае вероятность стирания при любых вероятностях $p(x_1)$ и $p(x_2)$ будет равна числу q, которое характеризует линию связи. Следовательно, q фиксировано, а r_1 и r_2 – это вероятности появления на приемном конце линии сигналов x_1 и x_2 соответственно, которые зависят от $p(x_1)$ и $p(x_2)$.

Рассмотрим функцию $H(r_1, r_2) = -r_1 \log r_1 - r_2 \log r_2 - q \log q$. Найдем условный экстремум этой функции при условии $r_1 + r_2 + q = 1$.

Чтобы найти условный экстремум функции $H(r_1, r_2)$, надо найти обычный экстремум функции

$$F(r_1, r_2, \lambda) = -r_1 \log r_1 - r_2 \log r_2 - q \log q + \lambda(r_1 + r_2 + q - 1)$$

(метод неопределенных множителей Лагранжа).

Необходимое условие экстремума: $\frac{\partial F}{\partial r_1} = 0$, $\frac{\partial F}{\partial r_2} = 0$ и $\frac{\partial F}{\partial \lambda} = 0$.

Дифференцируем функцию $F(r_1, r_2, \lambda)$ и приравниваем производную нулю. Получим:

$$\begin{cases} \log r_1 = \lambda - \log e \\ \log r_2 = \lambda - \log e \\ r_1 + r_2 = 1 - q \end{cases}$$

Система имеет решение $r_1 = r_2 = \frac{1-q}{2}$.

Достаточное условие экстремума: $d^2L = -\frac{1}{r_1} dr_1^2 - \frac{1}{r_2} dr_2^2 < 0$. Значит функция $H(r_1, r_2)$ имеет условный максимум в точке $(\frac{1-q}{2}, \frac{1-q}{2})$.

Максимальная энтропия системы $H_{max}(Y) = -\frac{1-q}{2} \log \frac{1-q}{2} - \frac{1-q}{2} \log \frac{1-q}{2} - q \log q = -(1-q) \log \frac{1-q}{2} - q \log q$.

Следовательно, $\max I(Y/X) = -(1-q) \log \frac{1-q}{2} - q \log q - [-(1-\mu-q) \log(1-\mu-q) - \mu \log \mu - q \log q] = -(1-q)[\log(1-q) - \log 2] + (1-\mu-q) \log(1-\mu-q) + \mu \log \mu = (1-q)[1 - \log(1-q)] + (1-\mu-q) \log(1-\mu-q) + \mu \log \mu$
Окончательно имеем:

$$\max I(Y/X) = (1-q)[1 - \log(1-q)] - \eta(1-\mu-q) - \eta(\mu).$$

Значит, пропускная способность двоичной симметричной линии связи со стиранием:

$$C = L\left[(1-q)[1 - \log(1-q)] - \eta(1-\mu-q) - \eta(\mu)\right].$$

Пример 39. Определить пропускную способность линии связи с помехами, передающую 70 сигналов 0 и 1 в единицу времени, причем каждый

из сигналов заменяется противоположным с вероятностью 0,05, кроме того, в процессе передачи сигналы могут искажаться так, что их невозможно распознать. Вероятность искажения 0,07.

Решение.

$C = L\,[(1-q)[1-\log(1-q)] - \eta(1-\mu-q) - \eta(\mu)] =$

$= 70\,[(1-0,07)[1-\log(1-0,07)] - \eta(1-0,05-0,07) - \eta(0,05)] =$

$= 45,42$ бит в единицу времени.

Теорема Шеннона. Пусть имеется источник информации X, энтропия которого в единицу времени (X) и линия связи с пропускной способностью C. Тогда, если $H(X) > C$, то при любом кодировании передача сообщений без задержек и искажений невозможна. Если же $H(X) < C$, то всегда можно закодировать достаточно длинное сообщение так, чтобы оно было передано без задержек и искажений с вероятностью, сколь угодно близкой к единице.

Пример 40. Имеется источник информации с энтропией $H(X) =$ 90 бит в единицу времени и две линии связи, каждая из которых способна передавать 55 двоичных знаков 0 или 1 в единицу времени. Каждый двоичный знак заменяется противоположным с вероятностью $\mu = 0.05$. Достаточна ли пропускная способность этих линий связи для передачи информации без задержек и искажений?

Решение. $L = 55$. Пропускная способность одной линии связи $C = 55(1 - (\eta(0,05) + \eta(1-0,05))) = 55(1-0,2161-0,0703) = 39,248$ бит.

Т.к. линии связи две, то максимальное количество информации, которое может быть передано $2 \cdot 39,248 = 78,496$ бит.

Т.к. $78,496 < 90$, то пропускной способности таких линий связи недостаточно для обеспечения передачи информации.

Глава 6.
Коды, обнаруживающие и исправляющие ошибки

1. Избыточность кодовых обозначений

Пусть по линии связи можно передавать два элементарных сигнала (0 и 1) и эти же два сигнала могут быть приняты на приемном конце линии. Будем рассматривать равномерные коды длины N. Тогда кодовые обозначения представляют собой 2^N различных последовательностей $x_0x_1...x_{N-1}$, где все x_i $(i = 0, ...N-1)$ принимает значения 0 или 1. Рассмотрим крайние случаи.

Если все 2^N различных последовательностей принять за кодовые обозначения, то скорость передачи информации будет наибольшей

$$L \text{ бит/ед.вр. или } \frac{L}{H} \text{ букв/ед.вр.}$$

Но при этом на приемном конце линии не будет возможности определить, имелись ли ошибки при передаче.

Ограничимся теперь меньшим числом кодовых обозначений. Тогда возникнет «избыточность кода», которую можно использовать для передачи сведений об ошибках, вносимых линией связи.

Самый простой пример – передавать по линии связи две цепочки длины N (00..0, 11..1). На приемном конце линии расшифровывать принятую цепочку как 00..0, если она содержит больше нулей, и как 11..1 в противном случае. Такой метод обеспечит малую вероятность ошибки, но скорость передачи будет очень низка.

Наибольший интерес представляют коды, обеспечивающие хорошую скорость передачи и одновременно позволяющие исправить многие ошибки в передаваемых сообщениях.

2. Прием проверки на четность для обнаружения одиночной ошибки

Рассмотрим пример использования избыточности кодовых обозначений для обнаружения *одиночной* ошибки. При этом будем исключать возможность двух и более ошибок.

Сопоставим кодовые обозначения всевозможным цепочкам $x_0x_1...x_{N-2}$ из $N-1$ чисел 0 и 1 (таких цепочек будет 2^{N-1}), а N-ую цифру x_{N-1} будем выбирать так, чтобы сумма $x_0+x_1+...+x_{N-1}$ была четной. После передачи

на приемной конце линии получена цепочка $x'_0x'_1...x'_{N-1}$. В случае одиночной ошибки среди N элементарных сигналов сумма $x'_0 + x'_1 + ... + x'_{N-1}$ будет нечетной.

Этот прием позволяет обнаружить нечетное число ошибок, при этом четное число ошибок не будет обнаружено. Однако, в случае, если вероятность появления более одной ошибки среди N сигналов мала, вышеописанный метод представляет ценность. В случае обнаружения ошибки, сообщение можно либо игнорировать, либо запросить повторно. Скорость передачи при таком методе кодирования остается большой:

$$\frac{N-1}{N}L \text{ бит/ед.вр. или } \frac{N-1}{N}\frac{L}{H} \text{ букв/ед.вр.}$$

Прием проверки на четность можно применить несколько раз. Часто это позволяет не только обнаружить, но и исправить ошибку.

Пусть $N = 3$, при этом используется два кодовых обозначения. x_0 – информационный сигнал, x_1, x_2 – контрольные сигналы. Подберем x_1 и x_2 так, чтобы обе суммы $x_0 + x_1$ и $x_0 + x_2$ были четные. Тогда получим два кодовых обозначения 000 и 111.

Проверив на приемном конце линии суммы $x'_0 + x'_1$ и $x'_0 + x'_2$, можно точно установить, какая цепочка $x'_0x'_1x'_2$ была передана.

Если обе суммы $x'_0 + x'_1$ и $x'_0 + x'_2$ четные, то ошибок при передаче не было.

Если нечетной будет одна из сумм, значит, ошибочно принят входящий в эту сумму контрольный сигнал x_1 или x_2.

Если обе суммы нечетные, значит неверно принят информационный сигнал x_0.

При таком методе кодирования будут исправлены все одиночные ошибки. Скорость передачи при этом будет

$$\frac{L}{3} \text{ бит/ед.вр.}$$

Пусть теперь $N = 7$. x_0, x_1, x_2, x_3 – информационные сигналы. Тогда число кодовых обозначений будет $2^4 = 16$. Первые четыре сигнала в кодовых обозначениях:

0000	0100	1000	1100
0001	0101	1001	1101
0010	0110	1010	1110
0011	0111	1011	1111

x_4, x_5, x_6 – контрольные сигналы. Подберем их так, чтобы были четными суммы: $s_1 = x_0+x_1+x_2+x_4$, $s_2 = x_0+x_1+x_3+x_5$, $s_3 = x_0+x_2+x_3+x_6$. Получим:

0000000	0100110	1000111	1100001
0001011	0101101	1001100	1101010
0010101	0110011	1010010	1110100
0011110	0111000	1011001	1111111

На приемном конце линии получим цепочку сигналов $x'_0x'_1x'_2x'_3x'_4x'_5x'_6$. Если один элементарный сигнал принят неправильно, то хотя бы одна из сумм $s'_1 = x'_0 + x'_1 + x'_2 + x'_4$, $s'_2 = x'_0 + x'_1 + x'_3 + x'_5$, $s'_3 = x'_0 + x'_2 + x'_3 + x'_6$ окажется нечетной.

Если все суммы четные, то сообщение принято без ошибок.

Нечетность одной из сумм s'_1, s'_2, s'_3 указывает на ошибочно принятый входящий в эту сумму один из трех контрольных сигналов $(x'_4x'_5x'_6)$.

Нечетность двух сумм говорит о неверно принятом того из трех сигналов x'_1, x'_2, x'_3, который входит в обе эти суммы.

Если все три суммы нечетные, то неверно принят сигнал x'_0, который входит во все эти суммы.

Пусть на приемном конце линии получено сообщение 0001101. Суммы $s'_1 = x'_0+x'_1+x'_2+x'_4$ и $s'_2 = x'_0+x'_1+x'_3+x'_5$- нечетные, а сумма s'_3 – четная. Значит, неверно принят входящий в суммы s'_1, s'_2 сигнал x'_1. На самом деле передавался 0101101.

Использование таких кодовых обозначений обеспечивает скорость передачи

$$\frac{4L}{7} \text{ бит/ед.вр.} = \frac{4L}{7H} \text{ букв/ед.вр.}$$

и позволяет исправить все одиночные ошибки в блоках из семи элементарных сигналов.

Рассмотрим пример практического использования вышеописанного метода кодирования при $N = 7$.

Пример 41. Пусть имеем двоичную симметричную линию связи. Вероятность перепутывания сигналов при передаче равна 0,01.

Пропускная способность такой линии связи равна $C = L[1-[\eta(0,01)+\eta(0,99)]] = 0,92L$ бит/ед.вр..

Т.е. существует код, позволяющий передавать $0,92$ бит/ед.вр., при этом вероятность ошибки при декодировании равна 0,01.

Воспользуемся вышеописанным кодом с $N = 7$. При этом будем передавать информацию со скоростью $0,58L$ бит/ед.вр., т.е. меньшей предельной скорости в **1,6** раз.

Но, вероятность ошибки при декодировании будет равна вероятности, что из семи переданных элементарных сигналов приняты с ошибкой два или больше.

$$p = 1 - [p_7(0) + p_7(1)] = 1 - [(0,99)^7 + C_7^1(0,01)^1(0,99)^6] = 0,002$$

Вероятность ошибки при приеме одного элементарного сигнала уменьшилась с 0,01 до 0,002, т.е. в **пять** раз.

В общем случае кодовых обозначений длины N число K контрольных сигналов кода, исправляющие все одиночные ошибки, должно удовлетворять неравенству

$$\log(N+1) \leqslant K < \log(N+1) + 1 \text{ или } 2^{K-1} - 1 < N \leqslant 2^K - 1$$

Число информационных сигналов $M = N - K$.

Код, использующий кодовые обозначения длины N, состоящий из $M = N - K$ информационных сигналов и K контрольных сигналов, используемых для проверок на четность, будем называть (N, M)-кодом. Скорость передачи информации для такого кода будет равна $\frac{M}{N}L$ бит/ед.вр.

Т.к. $K < \log(N+1) + 1$, то при большом K,N $(K < N)$ число K будет гораздо меньше, чем N. Поэтому скорость передачи при большом N близка к максимальной скорости L бит/ед.вр. Однако, очень большое N выбирать невыгодно, т.к. при этом сильно увеличивается вероятность наличия больше одной ошибки в блоке из N сигналов. На практике приходится выбирать не слишком большое и не слишком малое значение N.

3. Прием проверки на четность для обнаружения одной или двух ошибок

Рассмотрим код, исправляющий одну или две ошибки для случая $N = 5$. Исключим возможность искажения больше, чем двух сигналов из пяти.

Здесь потребуется 4 проверки на четность. Значит из пяти сигналов четыре должны быть контрольными.

x_0 – информационный сигнал. Контрольные сигналы x_1, x_2, x_3, x_4 можно подобрать из условия:

$$s_1 = x_0 + x_1, s_2 = x_0 + x_2, s_3 = x_0 + x_3, s_4 = x_0 + x_4.$$

Тогда четность всех рассматриваемых сумм на приемном конце линии означает отсутствие ошибок.

Нечетность одной суммы s_i говорит об ошибке в соответствующем сигнале x_i.

Нечетность двух сумм s_i и s_j означает, что с ошибкой переданы сигналы x_i и x_j.

Нечетность трех сумм, например, всех, кроме s_i указывает на невернопринятые сигналы x_i и x_0.

Если все суммы нечетные, то с ошибкой передан один сигнал x_0.

В общем случае кодов, исправляющих одну или две ошибки в блоках из произвольного числа N сигналов, число K контрольных сигналов не может быть меньше, чем

$$\log(C_N^2 + C_N^1 + 1) = \log \frac{N^2 + N + 1}{2}.$$

Однако, на вопрос как именно подбирать контрольные сигналы ответить непросто. Часто эта задача сводится к перебору огромного числа возможностей.

Еще в более общем случае кодов, позволяющих обнаружить и исправить в цепочке из N сигналов любое число ошибок, не превосходящего заданного n, число контрольных вопросов K не может быть меньше, чем

$$\log(C_N^n + C_N^{n-1} + ... + 1).$$

Однако, это неравенство не дает ответа, как именно нужно выбирать необходимые проверки на четность. Кроме того, для некоторых K, удовлетворяющих этому неравенству, невозможно построить нужный код.

4. Матричное кодирование

Рассмотрим все арифметические действия с числами 0 и 1, принятые в 2-арифметике. Эти действия могут иметь два возможных результата 0, символизирующий, что результат четный и 1, означающий нечетность результата:

$$0 + 0 = 0, 0 + 1 = 1, 1 + 0 = 1, 1 + 1 = 0$$

$$0 \cdot 0 = 0, 0 \cdot 1 = 0, 1 \cdot 0 = 0, 1 \cdot 1 = 1$$

Кроме того, т.к. $1 + 1 = 0$, то $1 = -1$.

Общий (M,N)-код с проверками на четность задается $K = N - M$ равенствами вида

$$\begin{gathered} x_M = a_{M,0}x_0 + a_{M,1}x_1 + ... + a_{M,M-1}x_{M-1} \\ x_{M+1} = a_{M+1,0}x_0 + a_{M+1,1}x_1 + ... + a_{M+1,M-1}x_{M-1} \\ ..... \\ x_{N-1} = a_{N-1,0}x_0 + a_{N-1,1}x_1 + ... + a_{N-1,M-1}x_{M-1} \end{gathered} \tag{1}$$

Все коэффициенты $a_{M,0}$, $a_{M,1}$,..., $a_{N-1,M-1}$ - числа 0 или 1.

Для данного (M,N)-кода $x_0x_1...x_{M-1}$ - информационные сигналы. Контрольные сигналы x_i $(i = M, M+1, ...N-1)$ удовлетворяют равенствам (1).

Для задания кода надо указать все коэффициенты $a_{i,j}$. При этом удобно сначала перенести все левые части $x_M, xM+1, ..., x_{N-1}$ вправо, а затем записать все коэффициенты в получившихся равенствах в виде таблицы из K строк и N столбцов:

$$\mathbf{A} = \begin{pmatrix} a_{M,0} & a_{M,1} & ... & a_{M,M-1} & 1 & 0 & ... & 0 \\ a_{M+1,0} & a_{M+1,1} & ... & a_{M+1,M-1} & 0 & 1 & ... & 0 \\ ... & ... & ... & ... & ... & ... & ... & ... \\ a_{N-1,0} & a_{N-1,1} & ... & a_{N-1,M-1} & 0 & 0 & ... & 1 \end{pmatrix} \tag{2}$$

Общий (N,M)-код с проверками на четность задается $(K \times M)$-матрицей из нулей и единиц.

Информационные сигналы $x_0x_1...x_{M-1}$ здесь могут быть любыми, а контрольные сигналы определяются по информационным с помощью равенств (1). Общее число кодовых обозначений равно 2^M.

Замечание. В более общем случае код с проверками на четность определяют как совокупность цепочек $x_0x_1...x_{N-1}$ таких, что числа $x_0, x_1, ...,$ x_{N-1} удовлетворяют K соотношениям:

$$\begin{cases} a_{M,0}x_0 + a_{M,1}x_1 + ... + a_{M,N-1}x_{N-1} = 0 \\ a_{M+1,0}x_0 + a_{M+1,1}x_1 + ... + a_{M+1,N-1}x_{N-1} = 0 \\ ..... \\ a_{N-1,0}x_0 + a_{N-1,1}x_1 + ... + a_{N-1,N-1}x_{N-1} = 0 \end{cases} \tag{3}$$

Матрица (3) будет уже произвольной матрицей $(K \times N)$, состоящей из нулей и единиц.

Коды, задаваемые равенствами вида (1) и матрицей (2), называют *систематическими* кодами с проверками на четность. Можно показать, что произвольный код с проверками на четность всегда может быть записан как систематический код с K контрольных сигналов. Дальше будем рассматривать только систематические коды.

Систематический код задается матрицей A (2), называемой *проверочной* матрицей кода. Пусть $\mathbf{x} = (x_0x_1...x_{N-1})$ – одно из кодовых обозначений кода, то равенство (1) удобно изображать в виде равенства:

$$\mathbf{Ax} = 0.$$

Пусть по линии связи передавалось кодовое обозначение $\mathbf{x} = (x_0, x_1, ..., x_{N-1})$. В результате искажений на приемном конце линии будет принята цепочка $\mathbf{x'} = (x'_0, x'_1, ..., x'_{N-1})$, отличная от той, которая передавалась. Подставим цепочку $\mathbf{x'}$ в левые части равенства (3), т.е. выполним Ax'. Получаемая при этом цепочка $Ax' = \mathbf{s} = (s_M, s_{M+1}, ...s_{N-1})$ уже на будет нулевой.

Пусть $\mathbf{e} = (e_0, e_1, ..., e_{N-1}) = (x'_0 - x_0, x'_1 - x_1, ..., x'_{N-1} - x_{N-1})$ – блок ошибок, содержащий единицы на местах, соответствующих сигналам x_i, искаженным при передаче, а нули на всех остальных местах.

Тогда, $\mathbf{Ae} = A(\mathbf{x'} - \mathbf{x}) = A\mathbf{x'} - A\mathbf{x} = A\mathbf{x'}$, т.к. $\mathbf{Ax} = 0$.

Следовательно,

$$\mathbf{Ae} = \mathbf{s} \tag{4}$$

Существует много цепочек $e = (e_0, e_1, ..., e_{N-1})$, удовлетворяющих K равенствам (4). Поэтому, нельзя однозначно восстановить блок ошибок, а следовательно, переданную цепочку $a = a' + e$. Задача нахождения нужного блока e может считаться решенной лишь для некоторых частных случаев со специальной структурой проверочной матрицы A.

Пусть проверочная матрица $\mathbf{A}$ содержит $K = N - M$ строк. Обозначим столбики матрицы $\mathbf{A}$ через $a_0, a_1, ..., a_{N-1}$. В случае систематического кода K столбцов $a_M, ..., a_{N-1}$ все будут содержать по одной единице и $N - M - 1$ нулей. Тогда матрицу $\mathbf{A}$ можно представить в виде:

$$\mathbf{B} = (\mathbf{a}_0, \mathbf{a}_1, ...\mathbf{a}_{N-1})$$

Пусть $e = (e_0, e_1, ...e_{N-1})$ – блок ошибок, содержащий единицы на местах тех элементарных сигналов передаваемого кодового обозначения, которые исказились при передаче. Перепишем равенство (4) в виде:

$$e_0 a_0 + e_1 a_1 + ... + e_{N-1} a_{N-1} = s$$

Следовательно, блок s, равен сумме столбцов проверочной матрицы A, отвечающих сигналам, искаженным при передаче. В частности, одиночным ошибкам соответствуют блоки s, совпадающие со столбцами a_i проверочной матрицы A. Отсутствию ошибок отвечает блок $s = 0$ из одних нулей. Поэтому для того, чтобы код с проверками на четность позволил различить и случай отсутствия ошибок, и все случаи одиночных ошибок, надо, чтобы все столбцы проверочной матрицы A были различны и ни один из них не был нулевым.

Общее число возможных различных K-значных блоков a равно числу целых чисел, записываемых в двоичной системе при помощи не более, чем K цифр, т.е. равно 2^K. Т.к. нулевой блок при этом исключается, то число

различных столбцов будет равно $2^K - 1$. Следовательно, *код с проверками на четность, исправляющий все одиночные ошибки и содержащий K контрольных сигналов, должен состоять из кодовых обозначений, длина которых не превосходит* $2^K - 1$.

Для задания такого кода надо в качестве столбцов матрицы A взять все числа от 0 до $2^K - 1$, записанные в двоичной системе и перечисленные в возрастающем порядке. Получаемый при этом код будет систематическим, но только контрольными сигналами здесь будут не последние K сигналов, а какие-то сигналы с другими номерами.

Пусть $N = 15$, $K = 4$. Тогда соответствующую проверочную матрицу (4×15) можно записать в виде:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix} \tag{5}$$

При такой матрице роль контрольных сигналов будут играть первый, второй, четвертый и восьмой сигналы. Остальные 11 сигналов будут информационные.

Блок s будет нулевым в случае отсутствия ошибок. В случае одиночной ошибки он будет равен соответствующему столбику A, т.е. задавать двоичную запись номера того сигнала, который исказился при передаче.

Коды, исправляющие одиночные ошибки в блоках из $N < 2^K - 1$ легко получить, вычеркнув из соответствующей проверочной матрицы A некоторое число лишних столбцов, которое можно выбрать произвольно из числа тех, которые содержат не меньше, чем две единицы.

Код можно улучшить, если добавить к каждому кодовому обозначению дополнительный $(K + 1)$-й контрольный сигнал x_N, позволяющий обнаружить, но не исправить также и все двойные ошибки. Для этого надо выбрать этот добавочный сигнал, так, чтобы он давал четное число в сумме со всеми остальными сигналами: $x_0 + x_1 + ... + x_N = 0$.

Это соответствует добавлению к матрице A сначала добавочного последнего столбца из одних нулей, а затем добавочной строки из $N + 1$ единиц. Тогда, при отсутствии ошибок, блок s будет состоять из одних нулей. В случае одной ошибки первые K элементов блока s будут представлять собой запись двоичного числа в пределах от 0 до $2^K - 1$, а последний будет равен 1. В случае двух ошибок последний элемент блока s будет равен 0, а остальные будут содержать хотя бы одну единицу.

В случае, если при передаче возникла не одна, а две ошибки, то блок s будет равен сумме соответствующих столбцов матрицы A. Для того, чтобы можно было обнаружить ошибку, необходимо, чтобы суммы любых двух столбцов отличалась от всех столбцов и от всех прочих их попарных сумм. Матрицу, удовлетворяющую этим условиям можно попытаться построить с помощью перебора. Однако, решение этой задачи даже для небольших N представляет сложность.

Пример 42. Рассмотрим код с $N = 7$, $K = 3$. Пусть проверочная матрица (3×7) имеет вид:

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

Т.к. $M = 4$, то можно составить $2^4 = 16$ кодовых обозначений, несущих информацию:

0000	0100	1000	1100
0001	0101	1001	1101
0010	0110	1010	1110
0011	0111	1011	1111

С помощью матрицы A каждой цепочке информационных сигналов надо приписать контрольные сигналы, применяя правило (1):

0000000	0100111	1000011	1100100
0001101	0101010	1001110	1101001
0010110	0110001	1010101	1110010
0011011	0111100	1011000	1111111

Пусть передана цепочка 0011011, а на приемном конце линии принята цепочка 00**0**1011. Для проверки правильности передачи умножим матрицу A на принятое кодовое обозначение. Получим $s = (1, 1, 0)$. Т.к. 110–элементы третьего столбика проверочной матрицы A, то третий сигнал передан с ошибкой.

Пусть передана цепочка 1010101, а на приемном конце линии принята цепочка 1010**1**1. Снова умножим матрицу A на принятое кодовое обозначение. Получим $s = (0, 1, 0)$. Значит, неверно принят шестой сигнал.

5. Алгебраическое кодирование

Сопоставим кодовому обозначению $a = (a_0, a_1, ..., a_{N-1})$ многочлен степени не выше $N-1$:

$$a(x) = a_0 + a_1 x + a_2 x^2 + ... + a_{N-1} x^{N-1},$$

относительно неизвестного x с коэффициентами 0 и 1. Будем рассматривать код, как совокупность кодовых многочленов $a(x)$. При этом должно быть выполнено два условия: (1) сумма любых двух многочленов, принадлежащих совокупности, должна принадлежать совокупности кодовых многочленов $a(x)$, (2) нулевой многочлен тоже обязательно принадлежит этой совокупности.

Совокупностью многочленов, удовлетворяющих указанным двум условиям, будет, например, совокупность многочленов $a(x)$ степени не выше некоторого $N-1$, делящихся без остатка на какой-либо фиксированный многочлен $g(x) = g_0 + g_1 x + ... + g_K x^K$ степени $K < N-1$, т.е. представимых в виде

$$a(x) = c(x)g(x),$$

где $c(x)$ – произвольный многочлен, степень которого не превосходит $N - K - 1$.

Тогда, совокупности многочленов соответствует определенный код с проверками на четность. Такой код называется *кодом, порожденным многочленом* $g(x)$. Многочлен $g(x)$ называется *порождающим* многочленом кода.

Пусть $a(x)$ – кодовый многочлен:
$a(x) = a_0 + a_1 x + ... + a_{K-1} x^{K-1} + a_K x^K + a_{K+1} x^{K+1} + ... + a_{N-1} x^{N-1}$

Здесь последние $M = N-K$ коэффициентов $a_K, a_{K+1}, ..., a_{N-1}$ можно выбрать произвольным образом, а первые K коэффициентов $a_0, a_1, ..., a_{K-1}$ будут однозначно определять условиями делимости $a(x)$ на $g(x)$. А именно, многочлен $a_0 + a_1 x + ... + a_{K-1} x^{K-1}$ должен равняться остатку от деления $a_K x^K + a_{K+1} x^{K+1} + ... + a_{N-1} x^{N-1}$ на $g(x)$.

Следовательно, последние $N-K$ сигналов $a_K, a_{K+1}, ..., a_{N-1}$ – информационные сигналы, а первые K сигналов $a_0, a_1, ..., a_{K-1}$ – контрольные. Общее число кодовых слов – 2^{N-K}.

Пусть на приемном конце линии связи принят блок $a' = (a'_0, a'_1, ..., a'_{N-1})$. Соответствующий ему многочлен $a'(x) = a'_0 + a'_1 x + ... a'_{N-1} x^{N-1}$ отличается от переданного $a(x) = a_0 + a_1 x + ... a_{N-1} x^{N-1}$ на многочлен ошибок $e(x) = e_0 + e_1 x + ... e_{N-1} x^{N-1}$, где $e_i = a'_i - a_i$. Из-за наличия добавленного многочлена $e(x)$, многочлен $a'(x)$ уже не будет делиться без остатка на $g(x)$.

Ненулевой остаток $r(x)$ от деления $a'(x)$ на $g(x)$ свидетельствует о наличии искажений и содержит всю информацию об ошибках.

Пример 43.

Пусть $N = 7$, $M = 4$. a_3, a_4, a_5, a_6 – информационные сигналы. Тогда число кодовых обозначений будет $2^{7-3} = 16$. Последние четыре сигнала в кодовых обозначениях:

0000	0100	1000	1100
0001	0101	1001	1101
0010	0110	1010	1110
0011	0111	1011	1111

a_0, a_1, a_2 – контрольные сигналы. Подберем их так, чтобы было выполнено условие: многочлен $a_0 + a_1x + a_2x^2$ должен равняться остатку от деления многочлена $a_3x^3 + a_4x^4 + a_5x^5 + a_6x^6$ на $g(x)$.

$g(x) = x^3 + x + 1$ – порождающий многочлен.

1. Составим кодовое обозначение, в котором **0101** – информационные сигналы. Найдем контрольные сигналы.

$x^6 + x^4$ – многочлен $a_3x^3 + a_4x^4 + a_5x^5 + a_6x^6$, соответствующий информационным сигналам **0101**.

$$\begin{array}{l|l} x^6 + x^4 & x^3 + x + 1 \\ \cline{2-2} x^6 + x^4 + x^3 & x^3 + 1 \\ \cline{1-1} \quad x^3 & \\ \quad x^3 + x + 1 & \\ \cline{1-1} \quad\quad x + 1 & \end{array}$$

Остатку $x + 1$ соответствуют сигналы 110.

Кодовое обозначение - 110 0101.

Действительно, при делении $\frac{x^6+x^4+x+1}{x^3+x+1} = x^3 + 1$ получаем нулевой остаток.

Пусть кодовое обозначение 110 0101 передали с ошибкой, тогда получим:

(1) Получено кодовое обозначение: **0**10 0101

$$\begin{array}{l|l} x^6 + x^4 + x & x^3 + x + 1 \\ \cline{2-2} x^6 + x^4 + x^3 & x^3 + 1 \\ \cline{1-1} \quad x^3 + x & \\ \quad x^3 + x + 1 & \\ \cline{1-1} \quad\quad \mathbf{1} & \end{array}$$

(2)Получено кодовое обозначение: 1**0**0 0101

$$\begin{array}{l|l} x^6+x^4+1 & x^3+x+1 \\ \cline{2-2} x^6+x^4+x^3 & x^3+1 \\ \cline{1-1} \quad x^3+1 & \\ \quad x^3+x+1 & \\ \cline{1-1} \mathbf{x} & \end{array}$$

(3)Получено кодовое обозначение: 11**1** 0101

$$\begin{array}{l|l} x^6+x^4+x^2+x+1 & x^3+x+1 \\ \cline{2-2} x^6+x^4+x^3 & x^3+1 \\ \cline{1-1} \quad x^3+x^2+x+1 & \\ \quad x^3+x+1 & \\ \cline{1-1} \mathbf{x^2} & \end{array}$$

(4)Получено кодовое обозначение: 110 **1**101

$$\begin{array}{l|l} x^6+x^4+x^3+x+1 & x^3+x+1 \\ \cline{2-2} x^6+x^4+x^3 & x^3 \\ \cline{1-1} \mathbf{x+1} & \end{array}$$

(5)Получено кодовое обозначение: 110 0**0**01

$$\begin{array}{l|l} x^6+x+1 & x^3+x+1 \\ \cline{2-2} x^6+x^4+x^3 & x^3+x+1 \\ \cline{1-1} \quad x^4+x^3+x+1 & \\ \quad x^4+x^2+x & \\ \cline{1-1} \quad x^3+x^2+x & \\ \quad x^3+x+1 & \\ \cline{1-1} \mathbf{x^2+x} & \end{array}$$

(6)Получено кодовое обозначение: 110 01**1**1

$$\begin{array}{l|l} x^6+x^5+x^4+x+1 & x^3+x+1 \\ \cline{2-2} x^6+x^4+x^3 & x^3+x^2 \\ \cline{1-1} \quad x^5+x^3+x+1 & \\ \quad x^5+x^3+x^2 & \\ \cline{1-1} \mathbf{x^2+x+1} & \end{array}$$

(7)Получено кодовое обозначение: 110 010**0**

$$\begin{array}{l|l} x^4+x+1 & x^3+x+1 \\ \cline{2-2} x^4+x^2+x & x \\ \cline{1-1} \mathbf{x^2+1} & \end{array}$$

2. Составим кодовое обозначение, в котором **0100** – информационные сигналы. Найдем контрольные сигналы.

x^4 – многочлен $a_3x^3 + a_4x^4 + a_5x^5 + a_6x^6$, соответствующий информационным сигналам **0100**.

$$\begin{array}{l|l} x^4 & x^3+x+1 \\ \hline x^4+x^2+x & x \\ \hline x^2+x & \end{array}$$

Остатку $x^2 + x$ соответствуют сигналы 011.

Кодовое обозначение - 011 0100.

Пусть кодовое обозначение 011 0100 передали с ошибкой, тогда получим:

(1)Получено кодовое обозначение: **1**11 0100

$$\begin{array}{l|l} x^4+x^2+x+1 & x^3+x+1 \\ \hline x^4+x^2+x & x \\ \hline \mathbf{1} & \end{array}$$

(2)Получено кодовое обозначение: 0**0**1 0100

$$\begin{array}{l|l} x^4+x^2 & x^3+x+1 \\ \hline x^4+x^2+x & x \\ \hline \mathbf{x} & \end{array}$$

(3)Получено кодовое обозначение: 01**0** 0100

$$\begin{array}{l|l} x^4+x & x^3+x+1 \\ \hline x^4+x^2+x & x \\ \hline \mathbf{x^2} & \end{array}$$

(4)Получено кодовое обозначение: 011 **1**100

$$\begin{array}{l|l} x^4+x^3+x^2+x & x^3+x+1 \\ \hline x^4+x^2+x & x+1 \\ \hline x^3 & \\ \hline x^3+x+1 & \\ \hline \mathbf{x+1} & \end{array}$$

(5)Получено кодовое обозначение: 011 0**0**00

$$\begin{array}{l|l} x^2+x & x^3+x+1 \\ \hline & \end{array}$$

В остатке $\mathbf{x^2+x}$

(6)Получено кодовое обозначение: 011 01**1**0

$$\begin{array}{l|l} x^5+x^4+x^2+x & x^3+x+1 \\ \cline{2-2} x^5+x^3+x^2 & x^2+x+1 \\ \cline{1-1} x^4+x^3+x & \\ x^4+x^2+x & \\ \cline{1-1} x^3+x^2 & \\ x^3+x+1 & \\ \cline{1-1} \mathbf{x^2+x+1} & \end{array}$$

(7)Получено кодовое обозначение: 011 010**1**

$$\begin{array}{l|l} x^6+x^4+x^2+x & x^3+x+1 \\ \cline{2-2} x^6+x^4+x^3 & x^3+1 \\ \cline{1-1} x^3+x^2+x & \\ x^3+x+1 & \\ \cline{1-1} \mathbf{x^2+1} & \end{array}$$

Кодовые обозначения с контрольными сигналами:

000 0000	011 0100	110 1000	101 1100
101 0001	110 0101	011 1001	000 1101
111 0010	100 0110	001 1010	010 1110
010 0011	001 0111	100 1011	111 1111

Остатки при делении, указывающие на позицию, в которой произошла ошибка:

Место ошибки	1	2	3	4	5	6	7
Остоток $r(x)$	1	x	x^2	$x+1$	x^2+x	x^2+x+1	x^2+1

6. Циклические коды

В алгебраической теории кодирования большое внимание уделяется специальным классам кодов, порожденных многочленами, называемыми *циклическими кодами.*

Код называется циклическим, если для каждого его кодового обозначения $a=(a_0,a_1,a_2,...,a_{N-1})$ блок $(a_{N-1},a_0,a_1,...,a_{N-2})$, полученный из a с помощью циклического сдвига, также является кодовым обозначением. Следовательно, и блок $(a_{N-i},a_{N-i+1},...,a_{N-i-1})$, полученный из a с помощью i-кратного циклического сдвига, также будет кодовым обозначением.

Все циклические коды порождены многочленами. $g(x)$ – порождающий многочлен. Пусть

$$a(x)=a_0+a_1x+a_2x^2...+a_{N-1}x^{N-1}$$

тогда $a_1(x)=a_{N-1}+a_0x+a_1x^2...+a_{N-2}x^{N-1}$.

Т.к. $x \cdot a(x) = x \cdot (a_0 + a_1 x + a_2 x^2 ... + a_{N-1} x^{N-1}) = a_0 x + a_1 x^2 + a_2 x^3 ... + a_{N-2} x^{N-1} + a_{N-1} x^N$ тогда $x \cdot a(x) - a_{N-1} x^N + a_{N-1} = a_{N-1} + a_0 x + a_1 x^2 + a_2 x^3 ... + a_{N-2} x^{N-1} = a_1(x)$.

Получим

$$a_1(x) = x \cdot a(x) - a_{N-1}(x^N - 1).$$

$a_1(x)$ будет кодовым обозначением, если делится на $g(x)$ без остатка. Т.к. $a(x)$ – кодовое обозначение, то $a(x)$ делится без остатка на $g(x)$. Значит, $a_1(x)$ будет делиться без остатка на $g(x)$, только, если $g(x)$ – делитель многочлена $x^N - 1$.

Следовательно, код, порожденный многочленом $g(x)$, будет циклическим только, если $g(x)$ – делитель многочлена $x^N - 1$.

Для определения степени порождающего полинома можно воспользоваться выражением $K \geqslant \log(N + 1)$, где N - длина строящегося циклического кода.

Степень полинома K	Порождающий полином $g(x)$
2	$x^2 + x + 1$
3	$x^3 + x + 1$
4	$x^4 + x + 1$
5	$x^5 + x^2 + 1$, $x^5 + x^4 + x^3 + x^2 + 1$, $x^5 + x^4 + x^2 + x + 1$
6	$x^6 + x + 1$, $x^6 + x^5 + x^2 + x + 1$
7	$x^7 + x^3 + 1$, $x^7 + x^3 + x^2 + x + 1$, $x^7 + x^4 + x^3 + x^2 + 1$
8	$x^8 + x^7 + x^6 + x^5 + x^2 + x + 1$, $x^8 + x^4 + x^3 + x^2 + 1$, $x^8 + x^6 + x^5 + x + 1$

Если $g(x)$ – делитель $x^N - 1$, то $x^N - 1 = g(x) \cdot h(x)$.

Т.к. $a(x) = g(x) \cdot c(x)$, то $a(x) \cdot h(x) = g(x)c(x)h(x) = c(x)(x^N - 1)$.

Значит, кодовые многочлены циклического кода – это такие многочлены (степени не выше $N - 1$), для которых $a(x)h(x)$ делится без остатка на $(x^N - 1)$.

Такое свойство многочленов $a(x)$ позволяет найти ошибки при передаче. Если $a'(x) = a(x) + e(x)$, где $e(x) \neq 0$, то $a'(x)h(x)$ уже не будет делиться на $x^N - 1$ без остатка. Вся информация об ошибках будет содержаться в остатке от деления.

При декодировании большую роль играет многочлен $h(x)$, который называется *проверочным многочленом циклического кода.* На приемном конце линии многочлен $a'(x)$ следует умножить на проверочный многочлен

$h(x)$, тогда остаток от деления этого произведения на x^N-1 определит расшифровку этого сообщения.

Пример 44. Составить циклический код, если $N = 7$.

Решение. Если $N = 7$, то $K = 3$, тогда проверочный многочлен: $(x^7 - 1 = (x^3 + x + 1)(x^4 + x^2 + x + 1))$.

Возьмем кодовое обозначение из предыдущего примера: 1110010. Тогда должны быть кодовыми обозначениями 0111001, 1011100, 0101110, 0010111, 1001011, 1100101.

Соответствующие им кодовые многочлены: $x^5 + x^2 + x + 1 = (x^3 + x + 1)(x^2 + 1)$, $x^6 + x^3 + x^2 + x = (x^3 + x + 1)(x^3 + x)$, $x^4 + x^3 + x^2 + 1 = (x^3+x+1)(x+1)$, $x^5+x^4+x^3+x = (x^3+x+1)(x^2+x)$, $x^6+x^5+x^4+x^2 = (x^3 + x + 1)(x^3 + x^2)$, $x^6 + x^5 + x^3 + 1 = (x^3 + x + 1)(x^3 + x^2 + x + 1)$, $x^6 + x^4 + x + 1 = (x^3 + x + 1)(x^3 + 1)$. Все они делятся на $g(x) = x^3 + x + 1$ без остатка.

При изучении кодов, позволяющих исправлять ошибки при передаче, важную роль играет *расстояние Хэмминга* $|b-a|$ между двумя цепочками $b = (b_0, b_1, ...b_{N-1})$ и $a = (a_0, a_1, ...a_{N-1})$, которое равно числу цифр a_i, таких, что $b_i \neq a_i$.

Пример 45. Расстояние Хэмминга между кодовыми обозначениями 111 0010 и 100 0110 равно $d = 3$.

Расстояние Хэмминга равно числу искажений отдельных передаваемых сигналов, приводящим к тому, что переданная цепочка a воспринимается на приемном конце линии как цепочка b. Чем больше будет расстояние Хэмминга между отдельными кодовыми обозначениями, тем меньше будет вероятность перепутать эти обозначения на приемном конце, т.е. тем лучше будет используемый код. Поэтому важной характеристикой кода является отвечающее ему *кодовое расстояние* $D = min|a^{(i)} - a^{(j)}|$ – расстояние Хэмминга между самыми близкими различными кодовыми обозначениями данного кода.

Пример 46. Для кода 111 0010, 001 1010, 000 1101, 011 0100 найти кодовое расстояние.

Решение.

	111 0010	001 1010	000 1101	011 0100
111 0010	0	3	7	3
001 1010		0	4	4
000 1101			0	4
011 0100				0

Кодовое расстояние равно $D = 3$.

В случае кода, позволяющего исправить любое, не превосходящее n число ошибок, он не должен содержать двух таких кодовых обозначений $a^{(i)} = (a_0^{(i)}, a_1^{(i)}, ..., a_{N-1}^{(i)})$ и $a^{(j)} = (a_0^{(j)}, a_1^{(j)}, ..., a_{N-1}^{(j)})$, что изменив какие-то n или менее цифр первого из них и какие-то n или менее цифр второго, мы получим одну и ту же цепочку b. Т.е., приняв на приемном конце линии эту цепочку b, мы не сможем выяснить, было ли передано обозначение $a^{(i)}$ или $a^{(j)}$. Следовательно, все расстояния должны быть больше $2n$. Значит, кодовое расстояние кода $D \geqslant 2n + 1$.

Итак, если кодовое расстояние $D \geqslant 2n + 1$, то код позволяет исправить любое, не превосходящее n, число ошибок и позволяет обнаружить (но не исправить) наличие не менее, чем $n + 1$ ошибок.

Приложение 1. Таблица величин

$$\eta(p) = -p \log p$$

p	0	1	2	3	4	5	6	7	8	9
0,00	–	0,0100	0,0179	0,0251	0,0319	0,0382	0,0443	0,0501	0,0557	0,0612
0,01	0,0664	0,0716	0,0766	0,0815	0,0862	0,0909	0,0955	0,0999	0,1043	0,1086
0,02	0,1129	0,1170	0,1211	0,1252	0,1291	0,0330	0,1369	0,1407	0,1444	0,1481
0,03	0,1518	0,1554	0,1589	0,1624	0,1659	0,1693	0,1727	0,1760	0,1793	0,1825
0,04	0,1858	0,1889	0,1921	0,1952	0,1983	0,2013	0,2043	0,2073	0,2103	0,2132
0,05	0,2161	0,2190	0,2218	0,2246	0,2274	0,2301	0,2329	0,2356	0,2383	0,2409
0,06	0,2435	0,2461	0,2487	0,2513	0,2538	0,2563	0,2588	0,2613	0,2637	0,2661
0,07	0,2686	0,2709	0,2733	0,2756	0,2780	0,2803	0,2826	0,2848	0,2871	0,2893
0,08	0,2915	0,2937	0,2959	0,2980	0,3002	0,3023	0,3044	0,3065	0,3086	0,3106
0,09	0,3127	0,3147	0,3167	0,3187	0,3207	0,3226	0,3246	0,3265	0,3284	0,3303
0,10	0,3322	0,3341	0,3359	0,3378	0,3398	0,3414	0,3432	0,3450	0,3468	0,3485
0,11	0,3503	0,3520	0,3537	0,3555	0,3571	0,3588	0,3605	0,3622	0,3638	0,3654
0,12	0,3671	0,3687	0,3703	0,3719	0,3734	0,3750	0,3766	0,3781	0,3796	0,3811
0,13	0,3826	0,3841	0,3856	0,3871	0,3886	0,3900	0,3915	0,3929	0,3943	0,3957
0,14	0,3971	0,3985	0,3999	0,4012	0,4026	0,4040	0,4053	0,4066	0,4079	0,4092
0,15	0,4105	0,4118	0,4131	0,4144	0,4156	0,4169	0,4181	0,4194	0,4206	0,4218
0,16	0,4230	0,4242	0,4254	0,4266	0,4277	0,4289	0,4301	0,4312	0,4323	0,4335
0,17	0,4346	0,4357	0,4368	0,4379	0,4390	0,4400	0,4411	0,4422	0,4432	0,4443
0,18	0,4453	0,4463	0,4474	0,4484	0,4494	0,4504	0,4514	0,4523	0,4533	0,4543
0,19	0,4552	0,4562	0,4571	0,4581	0,4590	0,4599	0,4608	0,4617	0,4626	0,4635
0,20	0,4644	0,4653	0,4661	0,4670	0,4678	0,4687	0,4695	0,4704	0,4712	0,4720
0,21	0,4728	0,4736	0,4744	0,4752	0,4760	0,4768	0,4776	0,4783	0,4791	0,4798
0,22	0,4806	0,4813	0,4820	0,4828	0,4835	0,4842	0,4849	0,4856	0,4863	0,4870
0,23	0,4877	0,4883	0,4890	0,4897	0,4903	0,4010	0,4916	0,4923	0,4929	0,4935
0,24	0,4941	0,4947	0,4954	0,4960	0,4966	0,4971	0,4977	0,4983	0,4989	0,4994
0,25	0,5000	0,5006	0,5011	0,5016	0,5022	0,5027	0,5032	0,5038	0,5043	0,5048

p	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0,26	0,5053	0,5058	0,5063	0,5068	0,5072	0,5077	0,5082	0,5087	0,5091	0,5096
0,27	0,5100	0,5105	0,5109	0,5113	0,5118	0,5122	0,5126	0,5130	0,5134	0,5138
0,28	0,5142	0,5146	0,5150	0,5154	0,5158	0,5161	0,5165	0,5169	0,5172	0,5176
0,29	0,5179	0,5182	0,5186	0,5189	0,5192	0,5196	0,5199	0,5202	0,5205	0,5208
0,30	0,5211	0,5214	0,5217	0,5220	0,5222	0,5225	0,5228	0,5230	0,5233	0,5235
0,31	0,5238	0,5240	0,5243	0,5245	0,5247	0,5250	0,5252	0,5254	0,5256	0,5258
0,32	0,5260	0,5262	0,5264	0,5266	0,5268	0,5270	0,5272	0,5273	0,5275	0,5277
0,33	0,5278	0,5280	0,5281	0,5283	0,5284	0,5286	0,5287	0,5288	0,5289	0,5290
0,34	0,5292	0,5293	0,5294	0,5295	0,5296	0,5297	0,5298	0,5299	0,5299	0,5300
0,35	0,5301	0,5302	0,5302	0,5303	0,5304	0,5304	0,5305	0,5305	0,5305	0,5306
0,36	0,5306	0,5306	0,5307	0,5307	0,5307	0,5307	0,5307	0,5307	0,5307	0,5307
0,37	0,5307	0,5307	0,5307	0,5307	0,5307	0,5306	0,5306	0,5306	0,5305	0,5305
0,38	0,5304	0,5304	0,5303	0,5303	0,5302	0,5302	0,5301	0,5300	0,5300	0,5299
0,39	0,5298	0,5297	0,5296	0,5295	0,5294	0,5293	0,5292	0,5291	0,5290	0,5289
0,40	0,5288	0,5286	0,5285	0,5284	0,5283	0,5281	0,5280	0,5278	0,5277	0,5275
0,41	0,5274	0,5272	0,5271	0,5269	0,5267	0,5266	0,5264	0,5262	0,5260	0,5258
0,42	0,5256	0,5255	0,5253	0,5251	0,5249	0,5246	0,5244	0,5242	0,5240	0,5238
0,43	0,5236	0,5233	0,5231	0,5229	0,5226	0,5224	0,5222	0,5219	0,5217	0,5214
0,44	0,5211	0,5209	0,5206	0,5204	0,5201	0,5198	0,5195	0,5193	0,5190	0,5187
0,45	0,5184	0,5181	0,5187	0,5175	0,5172	0,5169	0,5166	0,5163	0,5160	0,5157
0,46	0,5153	0,5150	0,5147	0,5144	0,5140	0,5137	0,5133	0,5130	0,5127	0,5123
0,47	0,5120	0,5116	0,5112	0,5109	0,5105	0,5102	0,5098	0,5094	0,5090	0,5087
0,48	0,5083	0,5079	0,5075	0,5071	0,5067	0,5063	0,5059	0,5055	0,5051	0,5047
0,49	0,5043	0,5039	0,5034	0,5030	0,5026	0,5022	0,5017	0,5013	0,5009	0,5004
0,50	0,5000	0,4996	0,4991	0,4987	0,4982	0,4978	0,4973	0,4968	0,4964	0,4959
0,51	0,4954	0,4950	0,4945	0,4940	0,4935	0,4930	0,4926	0,4921	0,4916	0,4911
0,52	0,4906	0,4901	0,4896	0,4891	0,4886	0,4880	0,4875	0,4870	0,4865	0,4860
0,53	0,4854	0,4849	0,4844	0,4839	0,4833	0,4828	0,4822	0,4817	0,4811	0,4806
0,54	0,4800	0,4795	0,4789	0,4784	0,4778	0,4772	0,4767	0,4761	0,4755	0,4750
0,55	0,4744	0,4738	0,4732	0,4726	0,4720	0,4714	0,4708	0,4702	0,4797	0,4691
0,56	0,4684	0,4678	0,4672	0,4666	0,4660	0,4654	0,4648	0,4641	0,4635	0,4629
0,57	0,4623	0,4616	0,4610	0,4603	0,4597	0,4591	0,4584	0,4578	0,4571	0,4565
0,58	0,4558	0,4551	0,4545	0,4538	0,4532	0,4525	0,4518	0,4512	0,4505	0,4498
0,59	0,4491	0,4484	0,4477	0,4471	0,4464	0,4457	0,4450	0,4443	0,4436	0,4429
0,60	0,4422	0,4415	0,4408	0,4401	0,4393	0,4386	0,4379	0,4372	0,4365	0,4357
0,61	0,4350	0,4343	0,4335	0,4328	0,4321	0,4313	0,4306	0,4298	0,4291	0,4383
0,62	0,4276	0,4268	0,4261	0,4253	0,4246	0,4238	0,4230	0,4223	0,4215	0,4207
0,63	0,4199	0,4192	0,4184	0,4176	0,4168	0,4160	0,4153	0,4145	0,4137	0,4129
0,64	0,4121	0,4113	0,4105	0,4097	0,4089	0,4080	0,4072	0,4064	0,4056	0,4048
0,65	0,4040	0,4032	0,4023	0,4015	0,4007	0,3998	0,3990	0,3982	0,3973	0,3965

p	0	1	2	3	4	5	6	7	8	9
0,66	0,3957	0,3948	0,3940	0,3931	0,3922	0,3914	0,3905	0,3897	0,3888	0,3880
0,67	0,3871	0,3862	0,3954	0,3845	0,3836	0,3828	0,3819	0,3810	0,3801	0,3792
0,68	0,3784	0,3775	0,3766	0,3757	0,3748	0,3739	0,3730	0,3721	0,3712	0,3703
0,69	0,3694	0,3685	0,3676	0,3666	0,3657	0,3648	0,3639	0,3630	0,3621	0,3611
0,70	0,3602	0,3593	0,3583	0,3574	0,3565	0,3555	0,3546	0,3536	0,3527	0,3518
0,71	0,3508	0,3499	0,3489	0,3480	0,3470	0,3461	0,3451	0,3441	0,3432	0,3422
0,72	0,3412	0,3403	0,3393	0,3383	0,3373	0,3364	0,3354	0,3344	0,3334	0,3324
0,73	0,3314	0,3304	0,3295	0,3285	0,3275	0,3265	0,3255	0,3245	0,3235	0,3225
0,74	0,3215	0,3204	0,3194	0,3184	0,3174	0,3164	0,3154	0,3144	0,3133	0,3123
0,75	0,3113	0,3103	0,3092	0,3082	0,3071	0,3061	0,3051	0,3040	0,3030	0,3019
0,76	0,3009	0,2999	0,2988	0,2978	0,2967	0,2956	0,2946	0,2935	0,2925	0,2914
0,77	0,2903	0,2893	0,2882	0,2871	0,2861	0,2850	0,2839	0,2828	0,2818	0,2807
0,78	0,2796	0,2785	0,2774	0,2763	0,2853	0,2741	0,2731	0,2720	0,2709	0,2698
0,79	0,2687	0,2676	0,2664	0,2653	0,2642	0,2631	0,2620	0,2609	0,2598	0,2587
0,80	0,2575	0,2564	0,2553	0,2542	0,2431	0,2519	0,2508	0,2497	0,2485	0,2474
0,81	0,2462	0,2451	0,2440	0,2428	0,2417	0,2405	0,2394	0,2382	0,2371	0,2359
0,82	0,2348	0,2336	0,2324	0,2313	0,2301	0,2290	0,2278	0,2268	0,2255	0,2243
0,83	0,2231	0,2220	0,2208	0,2196	0,2184	0,2172	0,2160	0,2149	0,2137	0,2125
0,84	0,2113	0,2101	0,2089	0,2077	0,2065	0,2053	0,2041	0,2029	0,2017	0,2005
0,85	0,1993	0,1981	0,1969	0,1957	0,1944	0,1932	0,1920	0,1908	0,1896	0,1884
0,86	0,1871	0,1859	0,1847	0,1834	0,1822	0,1810	0,1797	0,1785	0,1773	0,1760
0,87	0,1748	0,1735	0,1723	0,1711	0,1698	0,1686	0,1673	0,1661	0,1648	0,1635
0,88	0,1623	0,1610	0,1598	0,1585	0,1572	0,1560	0,1547	0,1534	0,1522	0,1509
0,89	0,1496	0,1484	0,1471	0,1458	0,1445	0,1432	0,1419	0,1407	0,1394	0,1381
0,90	0,1368	0,1355	0,1342	0,1329	0,1316	0,1303	0,1290	0,1277	0,1264	0,1251
0,91	0,1238	0,1225	0,1212	0,1199	0,1186	0,1173	0,1159	0,1146	0,1133	0,1120
0,92	0,1107	0,1094	0,1080	0,1067	0,1054	0,1040	0,1027	0,1014	0,1000	0,1987
0,93	0,0974	0,0960	0,0947	0,0933	0,0920	0,0907	0,0893	0,0880	0,0868	0,0853
0,94	0,0839	0,0826	0,0812	0,0798	0,0785	0,0771	0,0758	0,0744	0,0730	0,0717
0,95	0,0703	0,0689	0,0676	0,0662	0,0648	0,0634	0,0621	0,0607	0,0593	0,0579
0,96	0,0565	0,0552	0,0538	0,0524	0,0510	0,0496	0,0482	0,0468	0,0454	0,0440
0,97	0,0426	0,0412	0,0398	0,0384	0,0370	0,0356	0,0342	0,0328	0,0314	0,0300
0,98	0,0286	0,0271	0,0257	0,0243	0,0230	0,0214	0,0201	0,0186	0,0172	0,0158
0,99	0,0140	0,0129	0,0115	0,0101	0,0086	0,0072	0,0058	0,0043	0,0029	0,0014

Приложение 2.
Семантическая информация

Смысл сообщений не имеет никакого отношения к теории информации, целиком построенной на положениях теории вероятностей. Но в 50-х годах двадцатого века Бар-Хиллелом и Карнапом была предложена теория семантической информации. Семантическая информация трактовалась авторами, как синоним смыслового содержания.

Рассматривается предложение s. $p(s)$ – логическая вероятность предложения s. Предложены две основные меры семантической информации. Первая из них $cont(s) = 1 - p(s)$. Вторая $inf(s) = \log \frac{1}{1 - cont(s) = \log \frac{1}{p(s)} = -\log p(s)}$

Примером одной из таких мер является функция $inf(s) = -\log p(s)$, где s – это предложение, смысловое содержание которого измеряется, $p(s)$ – вероятность истинности s. Некоторые свойства этих функций-мер:

1. если $s_1 \to s_2$ (из s_1 следует s_2) – истино, то $inf(s_1) \geqslant inf(s_2)$;

2. $inf(s) \geqslant 0$, $cont(s) \geqslant 0$;

3. если s – истино, то $inf(s) = 0$;

4. Для двух логически независимых предложений $inf(s_1 \wedge s_2) = inf(s_1) + inf(s_2)$, но $cont(s_1 \wedge s_2) < cont(s_1) + cont(s_2)$, где $\wedge$ – знак логической связки "И".

Значение этой функции-меры inf больше для предложений, исключающих большее количество возможностей.

Пример 1. Из s_1 – "$a < 8$"и s_2 – "$a = 3$"следует, что $s_2 \to s_1$, т.е. $inf(s_2) \geqslant inf(s_1)$. Действительно, s_2 исключает больше возможностей, чем s_1.

Пример 2. Известно, что высказываение s_1 истино на 50%, а высказывание s_2 истино на 25%. Найти $inf(s)$ и $cont(s)$ предложений s_1 и s_2.

$inf(s_1) = -\log(\frac{1}{2}) = 1$, $inf(s_2) = -\log(\frac{1}{4}) = 2$,

$cont(s_1) = 1 - \frac{1}{2} = \frac{1}{2}$, $cont(s_2) = 1 - \frac{1}{4} = \frac{3}{4}$.

Список литературы

[1] А.М. Яглом, И.М. Яглом. Вероятность и информация. М.: Наука, 1973.

[2] В.В. Лидовский. Теория информации: Учебное пособие. М.: Компания Спутник +, 2004. 112 с.

[3] Е.С. Вентцель. Теория вероятностей. М.: Высшая школа, 2001г. 575с. Глава 18. Основные понятия теории информации. С.468-514.

[4] Р.Л. Стратонович. Теория информации. М.: Сов. радио, 1975, 424 с.

[5] Б.Д. Кудряшов. Теория информации. СПб.: Питер, 2009. 320с.

[6] В.П. Цымбал. Задачник по теории информации и кодированию. Ленанд, 2014, 280 с.

[7] О.С. Литвинская, Н.И. Чернышев. Основы теории передачи информации: учебное пособие. М.:КНОРУС, 2017. 168 с.

[8] В.А. Орлов, Л.И. Филиппов. Теория информации в упражнениях и задачах. Учебное пособие для втузов. М.: Высшая школа, 1976, 136 с.

[9] Л. Бриллюэн. Наука и теория информации. М.: Государственное издательство физико-математической литературы, 1960, 720 с.

[10] Р. Галлагер. Теория информации и надежная связь. М.: Сов. радио, 1974, 304с.

[11] А.Н. Колмогоров. Теория информации и теория алгоритмов. М.: Наука, 1987.

[12] https://ru.wikipedia.org/wiki/Семантическая информация

УНИВЕРСИТЕТ ИТМО

Миссия университета – генерация передовых знаний, внедрение инновационных разработок и подготовка элитных кадров, способных действовать в условиях быстро меняющегося мира и обеспечивать опережающее развитие науки, технологий и других областей для содействия решению актуальных задач.

КАФЕДРА ВЫСШЕЙ МАТЕМАТИКИ

Кафедра высшей математики – крупнейшая в Санкт-Петербургском национальном исследовательском университете информационных технологий, механики и оптики. С момента основания на ней работали такие выдающиеся ученые, как И.П. Натансон, В.А. Тартаковский, В.Н. Попов, И.А. Молотков, А.Г. Аленицын, В.В. Жук и другие. Научные интересы сотрудников покрывают практически все разделы математики. На кафедре сложилась мощная научная школа по математическому моделированию сложных физических систем. В последнее время активно развивается направление, связанное с нанофизикой и нанотехнологиями, квантовым компьютером и квантовыми коммуникациями. Сотрудники кафедры активно участвуют в международных научных конференциях, работают в рамках Российских и международных научных проектов. Сложилось тесное научное сотрудничество с Санкт-Петербургским государственным университетом, Петербургским отделением Математического института имени В.А. Стеклова РАН, лабораторией физикохимии наносистем Института химии силикатов РАН и другими научными центрами как в России, так и за рубежом: университетами Марселя и Тулона (Франция), Ювяскиля (Финляндия), Гумбольдтовским университетом Берлина (Германия).

Блинова Ирина Владимировна
Попов Игорь Юрьевич

Теория информации

Учебное пособие

В авторской редакции
Редакционно-издательский отдел Университета ИТМО
Зав. РИО Н.Ф. Гусарова
Подписано к печати
Заказ №
Тираж
Отпечатано на ризографе