

волны) — до 30 ГГц, оптический (инфракрасный диапазон) — 0.15–400 ТГц, оптический (видимый свет) — 400–700 ТГц, оптический (ультрафиолетовый диапазон) — 0.7–1.75 ПГц.



Рис. 3

Типичные современные каналы: телеграфный и телефонный. Перспективные, внедряемые ныне: оптоволоконный (терабоды) и цифровой телефонный (ISDN, Integrated Services Digital Networks) — 57–128 Кбод.

В реальных оптоволоконных системах скорость гораздо ниже теоретических пределов (редко превосходит 1–10 Гбод).

Наиболее широко пока используются телефонные линии связи. Здесь достигнута скорость более 50 Кбод!

6. Способы измерения информации

Понятие количества информации естественно возникает, например, в следующих типовых случаях:

1. Равенство вещественных переменных $a = b$, включает в себе информацию о том, что a равно b . Про равенство $a^2 = b^2$ можно сказать, что оно несет меньшую информацию, чем первое, т.к. из первого следует второе, но не наоборот. Равенство $a^3 = b^3$ несет в себе информацию по объему такую же, как и первое;

2. Пусть происходят некоторые измерения с некоторой погрешностью. Тогда чем больше будет проведено измерений, тем больше информации об измеряемой сущности будет получено;

3. М.о. некоторой сл.в. содержит в себе информацию о самой сл.в. Для сл.в., распределенной по нормальному закону, с известной дисперсией знание м.о. дает полную информацию о сл.в.;

4. Рассмотрим схему передачи информации. Пусть передатчик описывается сл.в. X , тогда из-за помех в канале связи на приемник будет приходить сл.в. $Y = X + Z$, где Z — это сл.в., описывающая помехи. В этой схеме можно говорить о количестве информации, содержащейся в сл.в. Y , относительно X . Чем ниже уровень помех (дисперсия Z мала),

тем больше информации можно получить из Y . При отсутствии помех Y содержит в себе всю информацию об X .

В 1865 г. немецкий физик Рудольф Клаузиус ввел в статистическую физику понятие энтропии или меры уравниваемости системы.

В 1921 г. основатель большей части математической статистики, англичанин Роналд Фишер впервые ввел термин “информация” в математику, но полученные им формулы носят очень специальный характер.

В 1948 г. Клод Шеннон в своих работах по теории связи выписывает формулы для вычисления количества информации и энтропии. Термин “энтропия” используется Шенноном по совету патриарха компьютерной эры фон Неймана, отметившего, что полученные Шенноном для теории связи формулы для ее расчета совпали с соответствующими формулами статистической физики, а также то, что “точно никто не знает” что же такое энтропия.

► Упражнение 4

Какое из соотношений несет в себе больше информации $x = 5$ или $x > 3$?

7. Вероятностный подход к измерению дискретной и непрерывной информации

В основе теории информации лежит предложенный Шенноном способ измерения количества информации, содержащейся в одной сл.в. относительно другой сл.в. Этот способ приводит к выражению количества информации числом.

Для д.с.в. X и Y , заданных законами распределения $P(X = X_i) = p_i$, $P(Y = Y_j) = q_j$ и совместным распределением $P(X = X_i, Y = Y_j) = p_{ij}$, количество информации, содержащейся в X относительно Y , равно

$$I(X, Y) = \sum_{i,j} p_{ij} \log_2 \frac{p_{ij}}{p_i q_j}.$$

Для непрерывных сл.в. X и Y , заданных плотностями распределения вероятностей $p_X(t_1)$, $p_Y(t_2)$ и $p_{XY}(t_1, t_2)$, аналогичная формула имеет вид

$$I(X, Y) = \iint_{\mathbb{R}^2} p_{XY}(t_1, t_2) \log_2 \frac{p_{XY}(t_1, t_2)}{p_X(t_1)p_Y(t_2)} dt_1 dt_2.$$

Очевидно, что

$$P(X = X_i, X = X_j) = \begin{cases} 0, & \text{при } i \neq j \\ P(X = X_i), & \text{при } i = j \end{cases}$$

и, следовательно,

$$I(X, X) = \sum_i p_i \log_2 \frac{p_i}{p_i p_i} = - \sum_i p_i \log_2 p_i.$$

Энтропия д.с.в. X в теории информации определяется формулой

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

Свойства меры информации и энтропии:

- 1) $I(X, Y) \geq 0$, $I(X, Y) = 0 \Leftrightarrow X$ и Y независимы;
- 2) $I(X, Y) = I(Y, X)$;
- 3) $HX = 0 \Leftrightarrow X$ — константа;
- 4) $I(X, Y) = HX + HY - H(X, Y)$, где $H(X, Y) = - \sum_{i,j} p_{ij} \log_2 p_{ij}$;
- 5) $I(X, Y) \leq I(X, X)$. Если $I(X, Y) = I(X, X)$, то X — функция от Y .

1) Логарифмированием из очевидного для всех x неравенства $e^{x-1} \geq x$ (равенство устанавливается только при $x = 1$) получается неравенство $x - 1 \geq \ln x$ или $\frac{x-1}{\ln 2} \geq \log_2 x$.

$$\begin{aligned} -I(X, Y) &= \sum_{i,j} p_{ij} \log_2 \frac{p_i q_j}{p_{ij}} \leq \sum_{i,j} p_{ij} \frac{\frac{p_i q_j}{p_{ij}} - 1}{\ln 2} = \\ &= \sum_{i,j} \frac{p_i q_j - p_{ij}}{\ln 2} = \frac{\sum_i p_i \sum_j q_j - \sum_{i,j} p_{ij}}{\ln 2} = \frac{1 - 1}{\ln 2} = 0, \end{aligned}$$

т.е. $I(X, Y) = 0$ только при $p_{ij} = p_i q_j$ для всех i и j , т.е. при независимости X и Y . Если X и Y независимы, то $p_{ij} = p_i q_j$ и, следовательно, аргументы логарифмов равны 1 и, следовательно, сами логарифмы равны 0, что означает, что $I(X, Y) = 0$;

2) Следует из симметричности формул относительно аргументов;

3) Если $HX = 0$, то все члены суммы, определяющей HX , должны быть нули, что возможно только тогда и только тогда, когда X — константа;

4) Из четырех очевидных соотношений

$$\sum_j p_{ij} = p_i, \quad \sum_i p_{ij} = q_j,$$

$$HX = - \sum_i p_i \log_2 p_i = - \sum_{i,j} p_{ij} \log_2 p_i,$$

$$HY = - \sum_j q_j \log_2 q_j = - \sum_{i,j} p_{ij} \log_2 q_j$$

получается

$$HX + HY - H(X, Y) = \sum_{i,j} p_{ij} (\log_2 p_{ij} - \log_2 q_j - \log_2 p_i) = I(X, Y);$$

5) Нужно доказать $I(X, Y) = HX + HY - H(X, Y) \leq HX$ или $HY - H(X, Y) \leq 0$.

$$HY - H(X, Y) = - \sum_{i,j} p_{ij} \log_2 q_j + \sum_{i,j} p_{ij} \log_2 p_{ij} = \sum_{i,j} p_{ij} \log_2 (p_{ij}/q_j),$$

но $p_{ij} = P(X = X_i, Y = Y_j) \leq q_j = P(Y = Y_j)$, а значит аргументы у всех логарифмов не больше 1 и, следовательно, значения логарифмов не больше 0, а это и значит, что вся сумма не больше 0.

Если $HX = I(X, X) = I(X, Y)$, то для каждого i p_{ij} равно либо q_j , либо 0. Но из $p_{ij} = P(X = X_i, Y = Y_j) = P(X = X_i/Y = Y_j)P(Y = Y_j) \in \{q_j, 0\}$ следует $P(X = X_i/Y = Y_j) \in \{0, 1\}$, что возможно только в случае, когда X — функция от Y .

При независимости сл. в. X и Y одна из них ничем не описывает другую, что и отражается в том, что для таких сл. в. $I(X, Y) = 0$.

Рассмотрим пример измерения количества информации при подбрасывании двух игральных костей.

Пусть заданы д. с. в. X_1 , X_2 и Y . X_1 и X_2 — количества очков, выпавших соответственно на 1-й и 2-й игральной кости, а $Y = X_1 + X_2$. Найти $I(Y, X_1)$, $I(X_1, X_1)$, $I(Y, Y)$.

Законы распределения вероятностей для д. с. в. X_1 и X_2 совпадают, т. к. кости одинаковые и без изъянов.

$$\begin{array}{c|cccccc} X_1 & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline p & & 1/6 & & & & \end{array}, \text{ т. е. при } j = 1 \dots 6 \quad q_j = P(X_1 = j) = 1/6.$$

Закон распределения вероятностей для д. с. в. Y ,

$$P(Y = i) = P(X_1 + X_2 = i), \quad i = 2 \dots 12,$$

вследствие того, что X_1 , X_2 — независимы и поэтому

$$P(X_1 = n, X_2 = m) = P(X_1 = n)P(X_2 = m),$$

будет

$$p_i = P(X_1 + X_2 = i) = \sum_{\substack{n+m=i \\ 1 \leq n, m \leq 6}} P(X_1 = n)P(X_2 = m) = \sum_{\substack{n+m=i \\ 1 \leq n, m \leq 6}} 1/36.$$

Таблицы, определяющие Y :

$X_2 \setminus X_1$	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12,

$Y = X_1 + X_2$	2	3	4	5	6	7	8	9	10	11	12
p	$1/36$	$2/36$	$3/36$	$4/36$	$5/36$	$6/36$	$5/36$	$4/36$	$3/36$	$2/36$	$1/36$,

т.е. при $i = 2...12$, $p_i = P(Y = i) = (6 - |7 - i|)/36$.

Закон совместного распределения вероятностей д.с.в. X_1 и Y будет

$$p_{ij} = P(Y = i, X_1 = j) = P(Y = i/X_1 = j)P(X_1 = j),$$

например,

$$\begin{aligned} P(Y = 2, X_1 = 1) &= P(Y = 2/X_1 = 1)P(X_1 = 1) = \\ &= P(X_2 = 1)P(X_1 = 1) = 1/36. \end{aligned}$$

В общем случае получится

$$p_{ij} = P(Y = i, X_1 = j) = \begin{cases} 1/36, & \text{при } 1 \leq i - j \leq 6, \\ 0, & \text{иначе.} \end{cases}$$

$X_1 \setminus Y$	2	3	4	5	6	7	8	9	10	11	12
1	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	0	0	0	0	0
2	0	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	0	0	0	0
3	0	0	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	0	0	0
4	0	0	0	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	0	0
5	0	0	0	0	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	0
6	0	0	0	0	0	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$

Тогда

$$\begin{aligned} I(Y, X_1) &= \sum_{j=1}^6 \sum_{1 \leq i-j \leq 6} p_{ij} \log_2 \frac{p_{ij}}{p_i q_j} = \\ &= \frac{1}{36} \sum_{j=1}^6 \sum_{1 \leq i-j \leq 6} \log_2 \frac{1}{6p_i} = \\ &= \frac{1}{36} \left(\sum_{i=2}^7 \log_2 \frac{1}{6p_i} + \sum_{i=3}^8 \log_2 \frac{1}{6p_i} + \dots + \sum_{i=6}^{11} \log_2 \frac{1}{6p_i} + \sum_{i=7}^{12} \log_2 \frac{1}{6p_i} \right) = \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{36}((\log_2 \frac{6}{1} + \log_2 \frac{6}{2} + \dots + \log_2 \frac{6}{6}) + \dots + (\log_2 \frac{6}{6} + \log_2 \frac{6}{5} + \dots + \log_2 \frac{6}{1})) = \\
&= \frac{1}{36}(\underline{2 \log_2 6} + 4 \log_2 3 + 6 \log_2 2 + 8 \log_2 \frac{3}{2} + 10 \log_2 \frac{6}{5} + \underline{6 \log_2 1}) = \\
&= (2 + 2 \log_2 3 + 4 \log_2 3 + 6 + 8 \log_2 3 - 8 + 10 \log_2 3 + 10 - 10 \log_2 5)/36 = \\
&= (10 + 24 \log_2 3 - 10 \log_2 5)/36 \approx 0.69 \text{ бит/символ.}
\end{aligned}$$

$$I(X_1, X_1) = I(X_2, X_2) = -\sum_{j=1}^6 q_j \log_2 q_j = \log_2 6 = 1 + \log_2 3 \approx 2.58 \text{ бит/сим.}$$

$$\begin{aligned}
I(Y, Y) &= -\sum_{i=2}^{12} p_i \log_2 p_i = \\
&= \frac{1}{36}(2 \log_2 36 + 4 \log_2 18 + 6 \log_2 12 + 8 \log_2 9 + 10 \log_2 \frac{36}{5} + 6 \log_2 6) = \\
&= (4 + 4 \log_2 3 + 4 + 8 \log_2 3 + 12 + 6 \log_2 3 + 16 \log_2 3 + 20 + 20 \log_2 3 - 10 \log_2 5 + \\
&\quad + 6 + 6 \log_2 3)/36 = (46 + 60 \log_2 3 - 10 \log_2 5)/36 \approx 3.27 \text{ бит/сим.}
\end{aligned}$$

Здесь $0 < I(Y, X_1) = I(Y, X_2) < I(X_1, X_1) = I(X_2, X_2) < I(Y, Y)$, что соответствует свойствам информации.

Подчеркнутый член $\frac{1}{36} 2 \log_2 6 = I(X_1, X_1)/18$ в расчете $I(X_1, Y)$ соответствует информации о двух случаях из 36, когда $Y = 2$ и $Y = 12$, которые однозначно определяют X_1 . Шесть случаев, когда $Y = 7$, не несут никакой информации об X_1 , что соответствует подчеркнутому члену $6 \log_2 1 = 0$.

Расчеты можно проводить, используя 4-е свойство информации, через энтропию.

$$H(Y, X_1) = -\sum_{i,j} p_{ij} \log_2 p_{ij} = \log_2 36 = 2(1 + \log_2 3) = 2H X_1 \approx 5.17 \text{ бит/сим.}$$

$$I(Y, X_1) = H X_1 + H Y - H(X_1, Y) = H Y - H X_1 \approx 3.27 - 2.58 = 0.69 \text{ бит/сим.}$$

Расчет количества информации с использованием 4-го свойства, а не определения, обычно требует меньше вычислений.

Рассмотрим более простой пример. Пусть д.с.в. X равна количеству очков, выпавших на игральной кости, а д.с.в. Y равна 0, если выпавшее количество очков нечетно, и 1, если выпавшее количество очков четно. Найти $I(X, Y)$ и $I(Y, Y)$.

Составим законы распределения вероятностей д.с.в. X и Y .

X	1	2	3	4	5	6
p	$1/6$					

Y	0	1
p	$1/2$	

Таким образом, при $i = 1 \dots 6$ $p_i = P(X = i) = 1/6$ и, соответственно, при $j = 0 \dots 1$ $q_j = P(Y = j) = 1/2$.

Составим также закон совместного распределения вероятностей этих д.с.в.

X	1	3	5	2	4	6	1	3	5	2	4	6
Y	0	0	0	1	1	1	1	1	1	0	0	0
p	1/6						0					

Таким образом, $p_{ij} = P(X = i, Y = j) = \begin{cases} 0, & \text{если } i + j \text{ — чётно,} \\ 1/6, & \text{иначе.} \end{cases}$

$$I(X, Y) = \sum_{i,j} p_{ij} \log_2 \frac{p_{ij}}{p_i q_j} = 6 \frac{1}{6} \log_2 2 = 1 \text{ бит/сим.}$$

$$I(Y, Y) = - \sum_{j=0}^1 q_j \log_2 q_j = 2 \frac{1}{2} \log_2 2 = 1 \text{ бит/сим.}$$

Точное количество выпавших очков даёт точную информацию о чётности, т.е. 1 бит. Из $I(X, Y) = I(Y, Y) = 1$ бит/сим и 3-го свойства информации следует, что информация об X полностью определяет Y , но не наоборот, т.к. $I(X, Y) \neq I(X, X) = 1 + \log_2 3 \approx 2.58$ бит/сим. Действительно, Y функционально зависит от X , а X от Y функционально не зависит.

Расчёты через энтропию будут следующими

$$H(X, Y) = - \sum_{i,j} p_{ij} \log_2 p_{ij} = \log_2 6 = 1 + \log_2 3 = HX,$$

$$I(X, Y) = HX + HY - HX = HY = 1 \text{ бит/сим.}$$

► Упражнение 5

Найти энтропию д.с.в. X , заданной распределением

X	1	2	3	4	5	6	7	8
p	0.1	0.2	0.1	0.05	0.1	0.05	0.3	0.1.

► Упражнение 6

Значения д.с.в. X_1 и X_2 определяются подбрасыванием двух идеальных монет, а д.с.в. Y равна сумме количества “гербов”, выпавших при подбрасывании этих монет. Сколько информации об X_1 содержится в Y ?

► Упражнение 7

Сколько информации об X_1 содержится в д.с.в. $Z = (X_1 + 1)^2 - X_2$, где независимые д.с.в. X_1 и X_2 могут с равной вероятностью принимать значение либо 0, либо 1? Найти HX_1 и HZ . Каков характер зависимости между X_1 и Z ?

► Упражнение 8

Д.с.в. X_1, X_2 — зависимы и распределены также как и соответствующие д.с.в. из предыдущей задачи. Найти $I(X_1, X_2)$, если совместное распределение вероятностей X_1 и X_2 описывается законом

X_1	0	0	1	1
X_2	0	1	0	1
p	1/3	1/6	1/6	1/3.

► Упражнение 9

Д.с.в. X_1 и X_2 определяются подбрасыванием двух идеальных тетраэдров, грани которых помечены числами от 1 до 4. Д.с.в. Y равна сумме чисел, выпавших при подбрасывании этих тетраэдров, т.е. $Y = X_1 + X_2$. Вычислить $I(X_1, Y)$, HX_1 и HY .

► Упражнение 10

Подсчитать сколько информации об X_1 содержится в д.с.в. $Z = X_1 * X_2$, а также HZ . Д.с.в. X_1 и X_2 берутся из предыдущего упражнения.

► Упражнение 11

Д.с.в. X_1 может принимать три значения $-1, 0$ и 1 с равными вероятностями. Д.с.в. X_2 с равными вероятностями может принимать значения $0, 1$ и 2 . X_1 и X_2 — независимы. $Y = X_1^2 + X_2$. Найти $I(X_1, Y)$, $I(X_2, Y)$, HX_1 , HX_2 , HY .

► Упражнение 12

Найти энтропии д.с.в. X , Y , Z и количество информации, содержащейся в $Z = X + Y$ относительно Y . X и Y — независимы и задаются распределениями

X	0	1	3	4
p	1/8	1/8	1/4	1/2

Y	-2	2
p	3/8	5/8

8. Смысл энтропии Шеннона

Энтропия д.с.в. — это минимум среднего количества бит, которое нужно передавать по каналу связи о текущем значении данной д.с.в.

Рассмотрим пример (скачки). В заезде участвуют 4 лошади с равными шансами на победу, т.е. вероятность победы каждой лошади равна $1/4$. Введем д.с.в. X , равную номеру победившей лошади. Здесь $HX = 2$. После каждого заезда по каналам связи достаточно будет передавать два бита информации о номере победившей лошади. Кодировать номер лошади следующим образом: 1—00, 2—01, 3—10, 4—11. Если ввести функцию $L(X)$, которая возвращает длину сообщения, кодирующего заданное значение X , то м.о. $ML(X)$ — это средняя длина сообщения, кодирующего X . Можно формально определить L через две функции $L(X) = \text{len}(\text{code}(X))$, где $\text{code}(X)$ каждому значению X ставит в соответствие некоторый битовый код, причем, взаимно однозначно, а len возвращает длину в битах для любого конкретного кода. В этом примере $ML(X) = HX$.

Пусть теперь д.с.в. X имеет следующее распределение

$$P(X = 1) = \frac{3}{4}, P(X = 2) = \frac{1}{8}, P(X = 3) = P(X = 4) = \frac{1}{16},$$