Федеральное агентство по образованию

Государственное образовательное учреждение
высшего профессионального образования

Владимирский государственный университет

С. В. ФЕДОРОВ

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ТЕОРИИ ИНФОРМАЦИИ

Учебное пособие

Владимир 2010

УДК 519.72(075.8)
ББК 22.18я73
Ф 33

Рецензенты:
Доктор физико-математических наук, профессор
зав. кафедрой теоретической физики
Владимирского государственного гуманитарного университета
В. Г. Рау

Доктор технических наук, профессор кафедры
радиотехники и радиосистем
Владимирского государственного университета
А. П. Галкин

Печатается по решению редакционного совета
Владимирского государственного университета

Федоров, С. В.
Ф 33 Математические основы теории информации : учеб. пособие / С. В. Федоров; Владим. гос. ун-т. – Владимир : Изд-во Владим. гос. ун-та, 2010. − 72 с.
ISBN 978-5-9984-0060-5

Подготовлено в соответствии с программой курса «Математические основы теории информации» специальностей 210301, 210302. Рассмотрены вопросы кодирования дискретных источников, взаимная информация и ее свойства, кодирование в дискретных и непрерывных каналах, кодирование источников с заданным критерием качества.

Предназначено для студентов 2-го курса очной и заочной форм обучения специальностей 210301 – радиофизика и электроника, 210302 – радиотехника.

Может быть полезно всем студентам, изучающим вопросы кодирования и передачи информации.

Табл. 2. Ил. 1. Библиогр.: 8 назв.

УДК 519.72(075.8)
ББК 22.18я73

ISBN 978-5-9984-0060-5

Предисловие

Теория информации **-** ветвь статистической теории связи. Основы заложены в классических трудах Н. Винера, А.Н. Колмогорова, И.А. Котельникова, К. Шеннона.

Основное содержание теории информации - исследование методов кодирования для экономного представления сообщений различных источников и надежной передачи сообщений по каналам связи с шумом.

В основе теории информации лежат статистическое описание (модель) источников сообщений и каналов связи, а также измерение на этой основе количества информации между сообщениями по Шеннону (количество информации определяется только вероятностными свойствами сообщений).

Предмет теории информации - теоремы, устанавливающие предельные возможности различных методов обработки и передачи сообщений (зависят только от статистических свойств источника и канала).

Три типичные задачи теории информации

1. Задан источник сообщений. Требуется найти наименьшее количество символов, необходимое для указания последовательности сообщений. Если задан критерий качества восстановления сообщений, то ошибка не должна превосходить заданную величину.

2. Задан канал связи. Требуется найти наибольшую возможную скорость передачи информации по этому каналу с произвольно малой ошибкой.

3. Заданы источник, канал и критерий качества. Необходимо найти минимально достижимую ошибку при передаче сообщений.

В результате решения этих задач зачастую не удается найти конкретный метод обработки сообщений, но произвести анализ систем передачи сообщений можно всегда, т.е. на основе теории информации можно оценить информационные параметры конкретных систем и сравнить их с теоретически достижимыми.

Теория информации возникла в 1948 г. из задач радиосвязи и телеграфии. Основы науки заложены К. Шенноном в двух статьях − “Математическая теория связи” и “Связь при наличии шума”.

До сих пор теория информации носит достаточно сложный математический характер. Для хорошего освоения материала необходимы знания теории вероятности, элементов теории случайных процессов, основы матричной алгебры и функционального анализа.

КОДИРОВАНИЕ ДИСКРЕТНЫХ ИСТОЧНИКОВ

Дискретные источники (ДИ) - наиболее простой объект теории информации. Для изучения ДИ требуется наименьшее количество определений и вспомогательных результатов.

Задача кодирования ДИ часто встречается на практике и называется задачей сжатия данных: как наиболее экономно представить с помощью кодовых символов последовательность дискретных сообщений.

1. Дискретные ансамбли и источники.

Пусть $X = \{x_1, x_2, ..., x_M\}$- множество, состоящее из M элементов. X, Y, ... - обозначения множеств; x, y, ... - элементы множеств, нижний индекс − номер элемента в множестве. Природа элементов несущественна, но мы будем называть их сообщениями. На конечном множестве X задано распределение вероятностей $p(x)$, если каждому $x_i \in X$ сопоставлено $p(x_i)$, причем

$$p(x_i) \geq 0,\ i = 1,2,...,M,\ \sum_{i=1}^{M} p(x_i) = 1.$$

Пусть A − подмножество X, $A \subseteq X$, тогда число $\Pr(A) = \sum_{x_i \in A} p(x_i)$ − вероятность того, что при случайном выборе сообщений из X будет выбрано сообщение, принадлежащее A (вероятность множества A).

Определение. Конечное множество X вместе с заданным на нем распределением $p(x)$называется дискретным вероятностным ансамблем (дискретным ансамблем) и обозначается $\{X, p(x)\}$ (иногда просто X).

Пусть X и Y − конечные множества, тогда множество, элементы которого представляют собой все возможные упорядоченные пары $(x_i, y_j), x_i \in X, y_j \in Y, i = 1,...M; j = 1,...,N$, называется произведением множеств X и Y – XY, XY и YX - различные множества. Если $X = Y$, то $XY = X^2$, по аналогии определяются множества $X_1 X_2 ... X_n$ и X^n.

Пусть задан ансамбль $\{XY, p(x,y)\}$, тогда $p_1(x_i) = \sum_{y_j \in Y} p(x_i, y_j)$ и $p_2(y_j) = \sum_{x_i \in X} p(x_i, y_j)$, т.е. из совместно заданных X и Y мы можем получить задание X и Y по отдельности.

Если для вероятностей, заданных на XY, справедливо $p(x_i, y_j) = p(x_i)p(y_j)$ для всех $x_i \in X$ и $y_j \in Y$, то X и Y называются статистически независимыми ансамблями.

Пусть задан ансамбль $\{XY, p(x,y)\}$ и для $x_i \in X$ $p_1(x_i) \neq 0$, тогда число $p(y_j | x_i) = \dfrac{p(x_i, y_j)}{p_1(x_i)}$ называется условной вероятностью сообщения y_j при условии, что x_i известно (условная вероятность y_j относительно x_i). Распределение $p(y|x_i)$ называется условным распределением на множестве Y относительно сообщения x_i. Таким образом, задание ансамбля $\{XY, p(x,y)\}$ определяет также и условные ансамбли $\{X, p(x|y)\}$, $p_2(y) \neq 0$ и $\{Y, p(y|x)\}, p_1(x) \neq 0$.

Пусть A - произвольное подмножество элементов из X такое, что $\Pr_1(A) = \sum\limits_{x \in A} p_1(x) \neq 0$, тогда число $p(y_j | A) = \dfrac{1}{\Pr_1(A)} \sum\limits_{x_i \in A} p(x_i y_j)$ называется условной вероятностью сообщения y_j на множестве Y относительно множества A.

Если выбрать $A=X$, то $p(y_j | A) = p_2(y_j)$, т.е. условное распределение относительно X - это безусловное распределение на множестве Y. Таким образом определены $\{X, p(x|B)\}, \Pr(B) \neq 0, B \subseteq Y$ и $\{Y, p(y|A)\}, A \subseteq X$, $\Pr_1(A) \neq 0$.

Рассмотрим вероятностный ансамбль $\{X_1, X_2, ..., X_n, p(x^{(1)}, ..., x^{(n)})\}$. Пусть

$$p_1(x^{(1)}) = \sum_{X_2} ... \sum_{X_n} p(x^{(1)}, ..., x^{(n)}), \text{ и } p_n(x^{(n)}) = \sum_{X_1} ... \sum_{X_{n-1}} p(x^{(1)}, ..., x^{(n)}) \quad (1.1)$$

безусловные распределения на множествах $X_1, X_2, ..., X_n$.

Условие статистической независимости ансамблей $p(x^{(1)}, ...x^{(n)}) = p_1(x^{(1)}) p_2(x^{(2)}) ... p_n(x^{(n)})$ для любых $x^{(1)} \in X_1, ..., x^{(n)} \in X_n$.

Можно ввести (задать) всевозможные совокупности по $m \leq n$ ансамблей

$$\{X_{i_1}, ..., X_{i_m}, p(x^{(i_1)}, ..., x^{(i_m)})\}, \quad (1.2)$$

где $p(x^{(i_1)}, ..., x^{(i_m)}) = \sum\limits_{X_{j_1}} ... \sum\limits_{X_{j_{n-m}}} p(x^{(1)}, ..., x^{(n)})$, суммирование производится по всем X_j, если j не содержится среди чисел $i_1, ..., i_m$.

Можно также получить различные условные ансамбли, вводя соответствующие условные распределения.

Дискретный источник - источник, выдающий в каждый определенный момент времени одно из сообщений дискретного множества X. В некоторых случаях в разные моменты времени могут использоваться разные множества.

Для полного задания источника необходимо дать вероятностное описание процесса появления сообщений на выходе источника, т. е. определить вероятности появления любых сочетаний сообщений.

Определение. Пусть U_X − дискретный источник, выбирающий сообщения из множества X. Источник U_X задан, если для любых n = 1, 2, ... и любых i = 0, ±1, ±2, ... задано семейство распределений вероятностей $\left\{p(x^{(i+1)},...,x^{(i+n)})\right\}, x^{(i_1)} \in X, j = i+1,...,i+n,$ удовлетворяющих условию ***согласованности***, состоящему в том, что распределение вероятностей $p(x^{(i_1)},...,x^{(i_m)})$ для любого набора позиций $i_1,...,i_m$ определено однозначным образом.

Распределение вероятности $p(x^{(i_1)},...,x^{(i_m)})$ на последовательностях $p(x^{(i_1)},...,x^{(i_n)})$ может быть получено многими способами, например:

$$p(x^{(2)}, x^{(3)}) = \sum_{X_1} p(x^{(1)}, x^{(2)}, x^{(3)}),$$

$$p(x^{(2)}, x^{(3)}) = \sum_{X_4} p(x^{(2)}, x^{(3)}, x^{(4)}).$$

Условие согласованности обеспечивает совпадение всех этих распределений, поэтому что определение источника совпадает с определением случайного процесса, порождаемого этим источником.

Всякое n-мерное распределение вероятности в общем случае является функцией $2n$ переменных $p(x^{(i_1)},...,x^{(i_n)}; i_1,...,i_n)$, т.е. вероятность некоторого отрезка сообщений зависит как от самого отрезка, так и от его расположения на оси времени.

Однако важный класс источников обладает однородностью во времени, или ***стационарностью***. Условие стационарности дискретного процесса

$$p(x^{(i_1+j)},...,x^{(i_n+j)}; i_1 + j,...,i_n + j) = p(x^{(i_1)},...,x^{(i_n)}; i_1,...,i_n).$$

Определение. ДИ называется источником без памяти, если для любых n = 1, 2, ..., любых i = 0, ±1, ±2, ... и любых последовательностей $(x^{(i+1)},...,x^{(i+n)})$, $x^{(i)} \in X$ имеет место равенство (в общем случае)

$$p(x^{(i+1)},...,x^{(i+n)}) = \prod_{j=1}^{n} p_{i+j}(x^{(i+j)}).$$

В случае стационарного ДИ без памяти $p(x^{(i+1)},...,x^{(i+n)}) = \prod_{i=1}^{n} p(x^{(i+j)})$.

Стационарный источник без памяти иногда называют ***постоянным***.

2. Случайные величины. Закон больших чисел

Пусть $\{X, p(x)\}$ - дискретный ансамбль, $\varphi(x)$ − функция, определенная на $X=\{x_1, ..., x_M\}$ и принимающая значения на числовой оси. x_i может иметь произвольную природу, но $\varphi(x_i)$ - уже числа.

Всякая действительная функция $\varphi(x_i)$, заданная на произвольном дискретном ансамбле, порождает действительную дискретную случайную величину (СВ).

Пусть X − числовое множество, тогда число $\mathrm{M}X = \sum_{x\in X} xp(x)$ − математическое ожидание СВ x. Число $\mu_k = \mathrm{M}(X - \mathrm{M}X)^k = \sum_{x\in X} (x - \mathrm{M}X)^k p(x)$ - k-й центральный момент СВ x. При этом μ_2 - дисперсия x, равная σ_X^2. Если $Y = \varphi(X)$ - СВ, определенная на $\{X,p(x)\}$, то:

$$\mathrm{M}\,Y = \sum_{y\in Y} yp_2(y) = \sum_{y\in Y} y \sum_{x:\varphi(x)=y} p_1(x) = \sum_{x\in X} \varphi(x)p_1(x). \qquad (2.1)$$

Основное свойство математического ожидания

$$\mathrm{M}[\alpha_1 X_1 + ... + \alpha_n X_n] = \alpha_1 \mathrm{M}X_1 + ... + \alpha_n \mathrm{M}X_2 ,$$

где $\alpha_1,...,\alpha_n$ − неслучайные числа.

Доказательство: рассмотрим $\{X_1,..., X_n, p(x^{(1)},..., x^{(i)})\}$ и положим $Y = \alpha_1 X_1 + ... + \alpha_n X_n$, тогда из (2.1)

$$\mathrm{M}Y = \sum_{X_1...X_n} (\alpha_1 x^{(1)} + ... + \alpha_n x^{(n)}) p(x^{(1)},...,x^{(n)}) =$$

$$= \alpha_1 \sum_{X_1...X_n} x^{(1)} p(x^{(1)},...,x^{(n)}) + ... + \alpha_n \sum_{X_1...X_n} x^{(n)} p(x^{(1)},...,x^{(n)}) =$$

$$= \alpha_1 \sum_{X_1} x^{(1)} p_1(x^{(1)}) + ... + \alpha_n \sum_{X_n} x^{(n)} p_n(x^{(n)}) \alpha_1 MX_1 + ... + \alpha_n MX_n .$$

Предпоследнее равенство справедливо в связи с (1.1).

Пусть совместно заданы X и Y, $p(x, y)$ − распределение вероятностей на множестве всех пар (x,y) - XY.

X и Y - статистически независимы, если $p(x,y)=p_1(x)p_2(y)$ для всех x и y. Число

$$\mathrm{K}_{XY}=\mathrm{M}(X-\mathrm{M}X)(Y-\mathrm{M}Y)=\sum_{XY}(x-\mathrm{M}X)(y-\mathrm{M}Y)p(x,y),$$

где $\mathrm{M}X=\sum_{X,Y}xp(x,y)$, $\mathrm{M}Y=\sum_{X,Y}yp(x,y)$, называется корреляционным моментом СВ x и y. Если x и y независимы, то $\mathrm{K}_{XY}=0$ (в общем случае не наоборот). Действительно,

$$\mathrm{K}_{XY}=\sum_X\sum_Y(x-\mathrm{M}X)(y-\mathrm{M}Y)p_1(x)p_2(y)=(\sum_X(x-\mathrm{M}X)p_1(x))\times$$

$$\times(\sum_Y(y-\mathrm{M}Y)p_2(y))=0.$$

Пусть $\mathrm{M}X=0$ и σ_X^2 - дисперсия СВ X, тогда для произвольного положительного ε

$$\sigma_X^2=\sum_X x^2p(x)\geq\sum_{|x|\geq\varepsilon}x^2p(x)\geq\varepsilon^2\sum_{|x|\geq\varepsilon}p(x)=\varepsilon^2\mathrm{Pr}(|x|\geq\varepsilon),$$

отсюда следует неравенство Чебышева:

$$Pr(\,|x|\,\geq\varepsilon\,)\leq\sigma_X^2\Big/\varepsilon^2.$$

Пусть $X_1,...,X_2$ − независимые СВ, имеющие одинаковые распределения вероятностей $p(x)$. Они соответствуют n независимым экспериментам, проводимым в i-й момент времени (от $i=1$ до $i=n$). Пусть Y - среднее арифметическое этих СВ:

$$y=\frac{1}{n}\sum_{i=1}^{n}x^{(i)},\ \text{тогда}\ \mathrm{M}Y=\frac{1}{n}\sum_{i=1}^{n}\mathrm{M}X_i=m_X,$$

где m_X − математическое ожидание $X_1,...,X_n$, имеющих одинаковые распределения.

$$\sigma_Y^2=\mathrm{M}(Y-\mathrm{M}Y)^2=\mathrm{M}[\frac{1}{n}\sum_{i=1}^{n}(x^{(i)}-m_X)]^2=$$

$$=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{M}(X_i-m_X)(X_j-m_X)=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{K}_{X_iX_j}=\frac{1}{n}\sigma_X^2;\ \sigma_Y^2=\frac{1}{n}\sigma_X^2,$$

так как $\mathrm{K}_{X_iX_j}=0$ при $i\neq j$ (независимость), а $\mathrm{K}_{X_iX_j}=\sigma_X^2$ (одинаковые распределения). Применим неравенство Чебышева к СВ Y:

$$\mathrm{Pr}(|\,Y-\mathrm{M}Y|\geq\varepsilon)\leq\sigma_Y^2\Big/\varepsilon^2,\ \ \mathrm{Pr}(|\frac{1}{n}\sum_{i=1}^{n}x_i-m_X|\geq\varepsilon)\leq\frac{\sigma_X^2}{n\varepsilon^2}.\qquad(2.2)$$

Теорема 2.1 (Закон больших чисел в форме Чебышева). Пусть $X_1,...,X_n$ - независимые одинаково распределенные дискретные СВ, имеющие конечные математические ожидания и дисперсию, тогда для любых положительных ε и δ найдется такое *N*, зависящее от ε и δ, что для всех $n > N$ вероятность того, что среднее арифметическое СВ $X_1,...,X_n$ будет отличаться от математического ожидания m_X каждой из СВ на величину, не меньшую ε, не превосходит δ:

$$\Pr(|\frac{1}{n}\sum_{i=1}^{n} x^{(i)} - m_X| \geq \varepsilon) \leq \delta .$$

Доказательство очевидно с учетом соотношений (2.2).

3. Количество информации в сообщении. Энтропия

Определение. Количеством собственной информации (или собственной информацией) в сообщении $x_i \in X$ называется число $I(x_i)$, определяемое как

$$I(x_i) = -\log p(x_i),\ i=1,2,\ ...,\ L.$$

Наиболее часто употребимы логарифмы по основанию 2 и натуральные логарифмы. Как правило, мы будем использовать двоичный логарифм [бит].

Основные свойства количества информации

1. Собственная информация неотрицательна. Она равна 0, только если вероятность сообщения равна 1.

2. Рассмотрим ансамбль $\{XY,p(x,y)\}$. Для каждой пары сообщений x_i и y_j собственная информация $I(x_i, y_j) = -\log p(x_i, y_j)$. Если сообщения независимы, то $I(x_i, y_j) = -\log p_1(x_i) - \log p_2(y_j) = I(x_i) + I(y_j)$ - свойство аддитивности информации.

Количество информации на ансамбле $\{X, p(x)\}$ является действительной функцией и, следовательно, представляет собой случайную величину со значениями $I(x_1), ..., I(x_L)$.

Определение. Математическое ожидание $H(X)$ случайной величины $I(x)$, определенной на $\{X, p(x)\}$, называется энтропией этого ансамбля:

$$H(X) = MI(x) = \sum_{x\in X} I(x)p(x) = -\sum_{x\in X} p(x)\log p(x). \qquad (3.1)$$

По определению собственная информация принимает значение ∞ при $p(x_i) = 0$, но энтропия любого дискретного ансамбля конечна, так как

$$\lim z \log z = \lim \frac{1/z}{1/z^2} \log e = 0 .$$

Свойства энтропии

1. Энтропия всякого дискретного ансамбля неотрицательна $H(X) \geq 0$. Равенство нулю возможно только при некотором $x_i \in X$, для которого $p(x_i) = 1$, а вероятности всех остальных сообщений равны нулю. Неотрицательность $H(X)$ следует из неотрицательности $I(x_i)$ для всех x_i.

2. Пусть L – число сообщений в ансамбле X, тогда

$$H(X) \leq \log L, \qquad (3.2)$$

равенство возможно только при одинаковых вероятностях всех сообщений (x_i) множества X. Доказательство основано на неравенстве

$$\ln x \leq x - 1. \qquad (3.3)$$

Рассмотрим выражение

$$H(X) - \log L = -\sum_X p(x) \log p(x) - \log L \sum_X p(x) =$$

$$= -\sum_X p(x)[\log p(x) + \log L] = \log e \sum_X p(x) \ln \frac{1}{Lp(x)} .$$

Используем неравенство (3.3)

$$H(X) - \log L \leq \log e \sum_X p(x)[\frac{1}{Lp(x)} - 1] = \log e[\sum_X \frac{p(x)}{Lp(x)} - \sum_X p(x)] = 0 ,$$

отсюда получаем неравенство (3.2). Равенство в (3.2) возможно только при $\frac{1}{Lp(x)} = 1$ для всех $x \in X$, т. е. при $p(x) = 1/L$ - равномерном распределении. $H(X) = 1/L$ - при равновероятных сообщениях.

3. Пусть X и Y - статистически независимые ансамбли, тогда для каждой пары $x_i \in X$ и $y_j \in Y$ выполняется свойство аддитивности. Беря математическое ожидание от обеих частей формулы, его выражающей, получим

$$H(X, Y) = MI(x_i y_j,) = M[I(x_i) + I(y_j)] = H(X) + H(Y), \qquad (3.4)$$

свойство аддитивности энтропии.

4. Условная информация. Условная энтропия

Пусть $\{XY, p(x,y)\}$ - пара совместно заданных ансамблей. На каждом из множеств могут быть определены различные условные распределения. Зафиксируем некоторое сообщение $y \in Y$, $p(y) \neq 0$ и рассмотрим условное

распределение $p(x \mid y)$ на X. Для каждого $x \in i$ определена собственная информация

$$\mathrm{I}(x \mid y) = -\log p(x \mid y),$$

которая называется условной собственной информацией сообщения x при фиксированном сообщении y. $\mathrm{I}(x \mid y)$ - случайная величина на ансамбле $\{\mathrm{X}, p(x \mid y)\}$, ее математическое ожидание

$$\mathrm{H}(X \mid y) = -\sum_X p(x|y) \log p(x|y)$$

называется условной энтропией ансамбля X относительно $y \in Y$. Условная энтропия также - СВ на ансамбле $\{Y, p(y)\}$.

Определение. Математическое ожидание $\mathrm{H}(X \mid Y)$ случайной величины $\mathrm{H}(X \mid y)$, определенной на ансамбле $\{Y, p(y)\}$, называется условной энтропией ансамбля X относительно ансамбля Y:

$$\mathrm{H}(X \mid Y) = \mathrm{MH}(X|y) = \sum_Y p(y) H(X \mid y) = -\sum_Y \sum_X p(x,y) \log p(x \mid y).$$

Все эти понятия определены для $y \in Y$, $p(y) \neq 0$.

Свойства энтропии и условной энтропии

1. Условная энтропия не превосходит энтропию того же ансамбля

$$\mathrm{H}(X \mid Y) \leq \mathrm{H}(X). \qquad (4.1)$$

Равенство осуществляется только при статистической независимости X и Y.

Доказательство: с помощью неравенства (3.3):

$$\mathrm{H}(X \mid Y) - \mathrm{H}(X) = \sum_X \sum_Y p(x,y) \log p(x \mid y) - \sum_X [-p(x) \log p(x)] =$$

$$= -\sum_X \sum_Y p(x,y)[\log p(x \mid y) - \log p(x)] = \sum_{XY} p(x,y) \log \frac{p(x)}{p(x \mid y)} \leq$$

$$\leq \log e \sum_{XY} p(x,y)\left[\frac{p(x)}{p(x|y)} - 1\right] = \log e\left[\sum_{XY} p(x)p(y) - \sum_{XY} p(x,y)\right] = 0.$$

Равенство выполняется только при независимых x и y.

2. Математическое ожидание собственной информации пары (x, y)

$$\mathrm{H}(X,Y) = \sum_{XY} I(x,y) p(x,y) = -\sum_{XY} p(x,y) \log p(x,y)$$

называется энтропией ансамбля XY. Используя $p(x, y) = p(y)p(x \mid y)$, получим

$$\mathrm{H}(X,Y) = -\sum_{XY} p(x,y) \log p(x|y) - \sum_{XY} p(x,y) \log p(y) = \mathrm{H}(X|Y) + \mathrm{H}(Y),$$

аналогично можно получить $\mathrm{H}(X,Y) = \mathrm{H}(Y|X) + \mathrm{H}(X)$ − свойства аддитивности энтропии. В случае независимых X и Y получаем (3.4).

3. Задан ансамбль $\{X, p(x)\}$, на нем определено отображение $\varphi(x)$ множества X во множество Y. Это отображение определяет ансамбль $\{Y, p(y)\}$, для которого $p(y) = \sum_{x:\varphi(x)=y} p(x)$. Пусть $H(X)$ и $H(Y)$ - энтропии ансамблей X и Y, тогда $H(Y) \le H(X)$. Знак равенства возможен только в случае обратимости $\varphi(x)$, т. е. каждому y соответствует только один $x \in X$. Действительно, так как любое сообщение x однозначно определяет соответствующее значение $y = \varphi(x)$, то $H(Y|X) = 0$,

$$H(Y) \le H(Y) + H(X|Y) = H(Y|X) + H(X).$$

При произвольных отображениях X в Y энтропия не возрастает. Она сохраняется, если Y однозначно определяет X.

4. Пусть $\{XYZ, p(x,y, z)\}$ - три совместно заданных ансамбля и $I(x|y,z) = -\log p(x|y, z)$ - условная собственная информация x при фиксированных сообщениях y и z, где $p(x|y, z) = \frac{p(x,y,z)}{\sum_X p(x,y,z)}$. Число

$$H(X|YZ) = MI(x|y, z) = -\sum_{XYZ} p(x,y,z)\log p(x|y,z)$$

называется условной энтропией X относительно пары ансамблей Y и Z. Имеет место неравенство $H(X|YZ) \le H(X|Y)$. Доказательство аналогично доказательству неравенства (4.1). Равенство возможно только при $p(x|y, z) = p(x|y)$, т. е. при статистической независимости X и Z при данном y.

Обобщение на случай n совместно заданных ансамблей

$$H(X_i|X_{i-1}...X_{i-s}) \le H(X_i|X_{i-1}...X_{i-m}) \qquad (4.2)$$

для любых s и m, $1 \le m \le s \le i$.

Для любой последовательности $(x^{(1)}, ...) \in X_1, ..., X_n$ ее вероятность

$$p(x^{(1)}, ..., x^{(n)}) = p(x^{(1)})\, p(x^{(2)}|x^{(1)}) ... p(x^{(n)}|x^{(n-1)}, ..., x^{(1)}),$$

где $$p(x^{(i)}|x^{(i-1)},...,x^{(1)}) = \frac{p(x^{(1)},...,x^{(i)})}{p(x^{(1)},...,x^{(i-1)})}$$

и $$p(x^{(1)}, ..., x^{(i)}) = \sum_{X_{i+1}...X_n} p(x^{(1)},...,x^{(n)}).$$

Тогда

$$I(x^{(1)}, ..., x^{(n)}) = I(x^{(1)}) + I(x^{(2)}|x^{(1)}) + ... + I(x^{(n)}|x^{(n-1)},...,x^{(1)}), \qquad (4.3)$$

где все члены, кроме первого, - условная собственная информация сообщения $x^{(i)}$. Вычисляем математическое ожидание обеих частей (4.3):

$$\mathrm{H}(X_1 \ldots X_n) = \mathrm{H}(X_1) + \mathrm{H}(X_2 | X_1) + \ldots + \mathrm{H}(X_n | X_{n-1}, \ldots, X_1) =$$

$$= \sum_{i=1}^{n} \mathrm{H}(X_i | X_{i-1}, \ldots, X_1) \ . \qquad (4.4)$$

Используя (4.2), можно записать:

$$\mathrm{H}(X_1 \ldots X_n) \le \sum_{i=1}^{n} \mathrm{H}(X_i) .$$

5. Энтропия на сообщение стационарного дискретного источника

Рассмотрим стационарный ДИ, выбирающий сообщения из множества X. X_i - ансамбль сообщений в i-й момент времени. Для стационарного источника все n-мерные распределения вероятности не зависят от сдвига во времени, следовательно, все величины, зависящие только от этих распределений, тоже не зависят от сдвига во времени, например $\mathrm{H}(X^n)$, $X^n = X_{i+1}, \ldots, X_{i+n}$ не зависят от i.

Пусть X^{n-1}, $X^n = X^{n-1} X_n$ - два ансамбля длины $n-1$ и n и

$$\mathrm{H}(X_n | X^{n-1}) = -\sum_{X^n} p(x^{(1)}, \ldots x^{(n)}) \log p(x^{(n)} | x^{(1)}, \ldots, x^{(n-1)})$$

условная энтропия ансамбля X_n относительно ансамбля $X^{(n-1)}$.

Теорема 5.1. Для всякого стационарного ДИ последовательность $\mathrm{H}(X_n | X^{n-1})$, $n = 1,2,\ldots$ имеет предел $\lim\limits_{n \to \infty} \mathrm{H}(X_n | X^{(n-1)}) = \mathrm{H}(X | X^{\infty})$.

Покажем, что последовательность

$$\mathrm{H}(X_1), \mathrm{H}(X_2 | X_1), \ldots, \mathrm{H}(X_n | X_{n-1} \ldots X_1) \qquad (5.1)$$

не возрастает.

Непосредственно применить (4.2) нельзя, так как последовательность (5.1) - последовательность энтропий различных ансамблей (X_1, X_2, ...), а не одного, как в (4.2). Однако в случае стационарного источника для всех i, j, n $\mathrm{H}(X_j | X_{j-1} \ldots X_{j-i}) = \mathrm{H}(X_n | X_{n-1} \ldots X_{n-i})$, поскольку и правая, и левая части определяются $i+1$- мерными распределениями вероятностей. Применяя теперь неравенство (4.2), получаем, что последовательность (5.1) не возрастает.

С другой стороны, все члены этой последовательности ограничены снизу (0), поэтому последовательность имеет предел. Обозначим его $\mathrm{H}(X|X^{\infty})$ и теорема доказана.

Рассмотрим последовательность

$$\mathrm{H}_{n}(X)=\frac{1}{n}H(X^{n}),\ n=1,2,\ldots\ , \qquad (5.2)$$

Если она имеет предел при $n\to\infty$, то он (предел) представляет собой среднее количество информации, порождаемое источником в единицу времени, и называется энтропией стационарного источника на сообщение.

Теорема 5.2. Для всякого стационарного ДИ последовательность (5.2) имеет предел, причем

$$\lim_{\mathrm{n}\to\infty}\mathrm{H}_{n}(X)=\mathrm{H}(X\mid X^{\infty}).$$

Покажем, что последовательность (5.2) не возрастает. Рассмотрим $\mathrm{H}(X^{n+1})$, где $X^{n+1}=X^{n}X_{n+1}$ - ансамбль последовательностей длины $n+1$. Используя свойство аддитивности энтропии, получим:

$$\mathrm{H}(X^{n+1})=\mathrm{H}(X^{n})+\mathrm{H}(X_{n+1}\mid X^{n})=\mathrm{H}(X^{n})+\mathrm{H}(X_{n}\mid X^{n})\le$$
$$\le\mathrm{H}(X^{n})+\mathrm{H}(X_{n}\mid X^{n-1})\ . \qquad (5.3)$$

Используя (4.4) и (4.2), можно получить:

$$\mathrm{H}(X^{n})=\sum_{i=1}^{n}\mathrm{H}(X_{i}\mid X^{i-1})\ge n\mathrm{H}(X_{n}\mid X^{n-1})\ .$$

Заменяя $\mathrm{H}(X_{n}|X^{n-1})$ на $\frac{1}{n}\mathrm{H}(X^{n})$ в (5.3), мы только усиливаем неравенство, поэтому $\mathrm{H}(X^{n+1})\le\frac{n+1}{n}\mathrm{H}(X^{n})$. Разделив обе части на $n+1$, получим $\mathrm{H}_{n+1}(X)\le\mathrm{H}_{n}(X)$ - последовательность энтропий не возрастает, а так как она ограничена снизу (0), то имеет предел.

Можно записать

$$\mathrm{H}_{n}(X)=\frac{1}{n}\sum_{i=1}^{n}\mathrm{H}(X_{i}\mid X^{i-1})=\frac{1}{n}\sum_{i=1}^{k}\mathrm{H}(X_{i}\mid X^{i-1})+\frac{1}{n}\sum_{i=k+1}^{n}\mathrm{H}(X_{i}\mid X^{i-1})\le$$
$$\le\frac{k}{n}\mathrm{H}(X)+\frac{n-k}{n}\mathrm{H}(X_{k+1}\mid X^{k}),\ \ k\le n.$$

Выберем k так, чтобы для заданного $\varepsilon>0$ выполнялось

$$\mathrm{H}(X_{k+1}|X^{k})-\mathrm{H}(X\mid X^{\infty})\le\frac{\varepsilon}{2}.$$

Вследствие теоремы 5.1 это всегда можно сделать (H $(X_{k+1}|X^k)$ стремится к H($X|X^\infty$), не возрастая при $n \to \infty$).

По выбранному k определим N так, чтобы при всех $n > N$:

$$\frac{k}{n}\mathrm{H}(X) \le \frac{\varepsilon}{2}.$$

Тогда для любого $\varepsilon > 0$ всегда найдется N такое, что для всех $n > N$

$$\mathrm{H}_n(X) \le \mathrm{H}(X \mid X^\infty) + \varepsilon \ . \qquad (5.4)$$

С другой стороны,

$$\mathrm{H}_n(X) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{H}(X_i \mid X^{i-1}) \ge \mathrm{H}(X_n \mid X^{n-1}) \ge \mathrm{H}(X \mid X^\infty). \qquad (5.5)$$

Так как ε - произвольное число больше нуля, то из (5.4) и (5.5) следует, что H($X|X^\infty$) является пределом $\mathrm{H}_n(X)$ при $n \to \infty$.

6. Постановка задачи кодирования равномерными кодами

Обозначим через A множество, состоящее из D элементов (D >1): $A = \{a_1,\ldots,a_D\}$, и назовем его алфавитом кода источника. Элементы A – ***кодовые символы***. Последовательность кодовых символов – ***кодовое слово***. Любое семейство кодовых слов – ***код над алфавитом.***

Определение. Код называется равномерным, если все его слова имеют одинаковую длину m, это число называется **длиной кода**. Если хотя бы два кодовых слова имеют различные длины, то код называется неравномерным.

Количество слов равномерного кода длины m – не больше D^m, неравномерного кода с максимальной длиной слова m – $D(D^m - 1)/(D-1)$.

Определение. Кодированием сообщений ансамбля X посредством кода называется отображение (необязательно взаимно однозначное) множества сообщений во множество кодовых слов.

Из всего множества сообщений выделяют два подмножества: однозначно кодируемых и декодируемых блоков и подмножество неоднозначно кодируемых и декодируемых блоков (“ошибка”).

При кодировании последовательность сообщений на выходе источника разбивается на блоки длиной n и кодер сопоставляет каждому блоку соответствующее слово длиной m; m определяется числом кодовых слов $M \le D^m$, где D – объем кодового алфавита.

Описанный способ кодирования называется равномерным кодированием. Ошибкой способа является появление неоднозначно кодируемого блока.

При равномерном кодировании количество D-ичных кодовых символов, приходящихся на одно сообщение, равно $\frac{1}{n}\frac{\log M}{\log D}$.

Определение. Число $R = \frac{\log M}{n}$ называется скоростью равномерного кодирования источника посредством кода с M кодовыми словами при разбиении последовательности сообщений на блоки длиной n (обычно – количество двоичных символов на сообщение).

Все коды, имеющие одинаковое число кодовых слов и однозначно кодирующие одно и то же множество сообщений, имеют одинаковые скорость и вероятность ошибки.

Наименьшая скорость кодирования является характеристикой источника сообщений и называется скоростью создания информации.

Определение. Скоростью создания информации ДИ при равномерном кодировании называется наименьшее число H такое, что для любого $R > H$ и любого сколь угодно малого $\delta > 0$ найдется n (длина кодовых сообщений) и равномерный код со скоростью кодирования R, для которого вероятность неправильного декодирования не превосходит δ.

Поиски H связаны с доказательством двух утверждений.

1. Прямая теорема кодирования для любого $R > H$ и произвольного δ найдутся n и код со скоростью R, кодирующий отрезки сообщений длины n, для которого вероятность ошибки не больше δ.

2. Обратная теорема кодирования для любого $R < H$ найдется зависящее от R $\delta > 0$ такое, что для всех n и для всех равномерных кодов со скоростью R вероятность ошибки при декодировании больше, чем это δ.

7. Теорема о высоковероятных множествах. Источник без памяти

Для ДИ без памяти и любой последовательности сообщений $x = (x^{(1)}, ..., x^{(n)}) \in X^n$ на его выходе:

$$p(\boldsymbol{x}) = \prod_{1}^{n} p(x^{(i)}).$$

Пусть H(X) - энтропия ансамбля $\{X, p(x)\}$ и ε - некоторое положительное число. Определим подмножество $T(\varepsilon)$ множества X^n как

$$T_n(\varepsilon) = \{\boldsymbol{x}: \mathrm{H}(X) - \varepsilon \le \frac{1}{n}\mathrm{I}(\boldsymbol{x}) \le \mathrm{H}(X) + \varepsilon\}, \qquad (7.1)$$

где $\mathrm{I}(\boldsymbol{x}) = -\log p(\boldsymbol{x})$ - собственная информация последовательности $\boldsymbol{x} \in X^n$.

Теорема. Для любых положительных чисел ε и δ найдется такое N, что для всех $n > N$ выполняется следующие неравенства:

$$\Pr(\boldsymbol{x} \in T_n(\varepsilon)) \ge 1 - \delta; \qquad (7.2)$$

$$(1-\delta)2^{n[\mathrm{H}(X)-\varepsilon]} \le |\mathrm{T}_n(\varepsilon)| \le 2^{n[\mathrm{H}(X)+\varepsilon]}, \qquad (7.3)$$

где $|T_n(\varepsilon)|$ - число элементов в множестве $T_n(\varepsilon)$.

В силу отсутствия памяти

$$\mathrm{I}(\boldsymbol{x}) = \sum_{i=1}^{n} \mathrm{I}(x^{(i)}), \qquad (7.4)$$

где $\mathrm{I}(x^{(i)})$ - независимые одинаково распределенные СВ, принимающие ограниченные значения. Поэтому можно применить закон больших чисел:

$$\Pr(|\frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(x^{(i)}) - \mathrm{H}(X)| > \varepsilon) \le \delta, \qquad (7.5)$$

$$\text{или} \quad \Pr(|\frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(x^{(i)}) - \mathrm{H}(X)| \le \varepsilon) \ge 1 - \delta, \qquad (7.6)$$

но (7.1) - (7.6) и есть доказательство справедливости (7.2).

Из определения множества $T_n(\varepsilon)$ следует, что для всякой последовательности $\boldsymbol{x} \in T_n(\varepsilon)$

$$2^{-n[\mathrm{H}(X)+\varepsilon]} \le p(\boldsymbol{x}) \le 2^{-n[\mathrm{H}(X)-\varepsilon]}. \qquad (7.7)$$

Тогда $$1 \ge \sum_{\mathrm{T}_n(\varepsilon)} p(\boldsymbol{x}) \ge |\mathrm{T}_n(\varepsilon)|\, 2^{-n[\mathrm{H}(X)+\varepsilon)]},$$

откуда следует справедливость правого неравенства в (7.3).

Суммируя теперь все элементы правого неравенства (7.7), получим:

$$1 - \delta \le \sum_{T_n(\varepsilon)} p(\boldsymbol{x}) \le |T_n(\varepsilon)| 2^{-n[\mathrm{H}(X)-\varepsilon]},$$

что эквивалентно левой части неравенства (7.3). Теорема доказана.

Последовательности сообщений на выходе ДИ без памяти можно разделить на множества $T_n(\varepsilon)$ и $\overline{T_n}(\varepsilon)$(дополнительное до X^n). Вероятности последовательностей из $T_n(\varepsilon)$ весьма близки друг к другу (7.7), а суммарная вероятность близка к 1 (7.2). Но доля $T_n(\varepsilon)$ может быть очень мала в X^n (число элементов - L):

$$|X^n| = L^n = 2^{n\log L}; \alpha = \frac{|\mathrm{T}_n(\varepsilon)|}{|X^n|} \le 2^{-n[\log L - \mathrm{H}(X) - \varepsilon]}.$$

Если $\log L - \mathrm{H}(X) - \varepsilon = a > 0$, ($\mathrm{H}(X) < \log L$), то ε достаточно мало, и α при увеличении n убывает к 0 как 2^{-an}.

Эти свойства множества $T_n(\varepsilon)$ позволяют называть его множеством типичных последовательностей, или высоковероятным множеством ДИ без памяти.

8. Скорость создания информации источником без памяти при равномерном кодировании

Скорость создания информации ДИ без памяти при равномерном кодировании равна его энтропии; чтобы это утверждать, необходимо доказать прямую и обратную теоремы кодирования.

Предположим, ДИ без памяти выдает сообщения из множества X с распределением вероятностей $p(x)$, $p(x) \neq 0$, $x \in X$ и $\mathrm{H}(X)$ - энтропия ансамбля $\{X, p(\mathrm{x})\}$.

Теорема 8.1 (прямая теорема кодирования). Пусть $\mathrm{R} > \mathrm{H}(X)$, тогда для любого положительного δ существует код со скоростью R, который кодирует ДИ без памяти с вероятностью ошибки, не превышающей δ.

Согласно теореме о высоковероятных множествах для любых ε и $\delta > 0$ найдется N такое, что при $n > N$ вероятность появления последовательности сообщений $\boldsymbol{x}$, не принадлежащей высоковероятному множеству $\mathrm{T}_n(\varepsilon)$, не превышает δ. Если это множество и принять в качестве множества однозначно кодирующих последовательностей, то вероятность ошибки декодирования не превышает δ и количество слов (кодовых) не меньше $|\mathrm{T}_n(\varepsilon)|$. Это условие удовлетворяется, если

$$2^{n\mathrm{R}} \ge 2^{n[\mathrm{H}(X)+\varepsilon]},$$

откуда следует, что $\mathrm{R} \ge \mathrm{H}(X) + \varepsilon$. Теорема доказана.

Доказано существование не одного, а целой последовательности кодов, кодирующих отрезки сообщений длины $n = N+1, N+2, \ldots$ и имеющих скорость $\mathrm{R} = \mathrm{H}(X) + \varepsilon$ с ошибкой не больше δ.

Теорема 8.2 (обратная теорема кодирования). Пусть R < H(X), тогда существует зависящее от R число $\delta > 0$ такое, что для каждого кода ДИ без памяти со скоростью R, вероятность ошибки $P_{en} \geq \delta$ и $\lim\limits_{n\to\infty} P_{en} = 1$.

Пусть $\varepsilon = H(X) - R > 0$ и $T_n(\varepsilon/2)$, $n = 1, 2, \ldots$ - последовательность высоковероятных множеств. Обозначим T_n - последовательность произвольных множеств $T_n \subseteq X^n$ и $|T_n| = 2^{nR} = 2^{n[H(X)-\varepsilon]}$. Для каждого n множество T_n будем рассматривать как множество однозначно кодируемых последовательностей сообщений, поэтому

$$1 - P_{en} = \Pr(T_n) = \Pr(T_n \cap \overline{T_n}(\varepsilon/2)) + \Pr(T_n \cap T_n(\varepsilon/2) \leq$$
$$\leq \Pr(\overline{T_n}(\varepsilon/2)) + \Pr(T_n \cap T_n(\varepsilon/2)), \qquad (8.1)$$

где $\overline{T_n}(\varepsilon/2)$ - дополнение множества $T_n(\varepsilon/2)$ до множества X^n. Из теоремы п. 7 следует, что

$$\lim_{n\to\infty} \Pr(\overline{T_n}(\varepsilon/2)) = 0, \ \varepsilon > 0. \qquad (8.2)$$

Для всякой последовательности $\boldsymbol{x} \in T_n(\varepsilon/2) \cap T_n$ имеем $p(\boldsymbol{x}) \leq 2^{-n[H(X)-\varepsilon/2]}$, так как $\boldsymbol{x} \in T_n(\varepsilon/2)$, $|T_n \cap T_n(\varepsilon/2)| \leq |T_n| = 2^{nR}$, следовательно:

$$\Pr(T_n \cap T_n(\varepsilon/2)) = \sum_{\boldsymbol{x} \in T_n \cap T_n(\varepsilon/2)} p(\boldsymbol{x}) \leq 2^{nR} 2^{-n(H(X)-\varepsilon/2)} = 2^{-n\varepsilon/2}. \qquad (8.3)$$

Из (8.2), (8.3) и (8.1) следует второе утверждение теоремы.

Докажем первое утверждение теоремы. При R < H(X) множество $\overline{T}_n$ не пусто для каждого n; $p(x) \neq 0$ для всех $x \in X$ и каждая последовательность $\boldsymbol{x} \in X^n$ имеет ненулевую вероятность, следовательно, для каждого n есть $\delta(n) > 0$, такое, что $P_{en} \geq \delta(n)$. В силу того, что $P_{en} \to 1$, найдется N такое, что для всех $n > N$ будет $P_{en} \geq 1/2$. Полагая $\delta = \min\{\delta(1), \delta(2), \ldots, 1/2\}$, получим, что для любого n и T_n

$$P_{en} \geq \delta > 0,$$

что и завершает доказательство теоремы.

Итак, скорость создания информации при равномерном кодировании ДИ без памяти равна его энтропии. Энтропия источника есть наименьшее количество двоичных символов на сообщение на выходе наилучшего дво-

ичного кодера для этого источника при условии, что сообщения источника могут быть восстановлены по выходу кодера сколь угодно точно.

9. Эргодические дискретные источники

Пусть U_X − стационарный источник, выбирающий сообщения из множества X, $\ldots x^{(-1)}, x^{(0)} x^{(1)}, \ldots$ − последовательности сообщений на его выходе. Пусть $\varphi(x_1,\ldots,x_k)$ - функция, определенная на X^k и отображающая отрезки сообщений длины k в числовую ось. Пусть $z^{(i)} = \varphi(x^{(i+1)},\ldots x^{(i+k)})$, i=1,2, ...− последовательность случайных величин, стационарных и поэтому одинаково распределенных; m_z - математическое ожидание $z^{(i)}$.

Определение. Стационарный ДИ называется эргодическим, если для любого k любой действительной функции $\varphi(x_1,\ldots,x_k), \mathrm{M}\varphi(\bullet) < \infty,$ определенной на X^k, любых положительных ε и δ найдется такое N, что для всех $n > N$:

$$\Pr(|\frac{1}{n}\sum_{i=1}^{n} z^{(i)} - m_z| \geq \varepsilon) \leq \delta. \qquad (9.1)$$

Теорема 9.1. Всякий стационарный ДИ без памяти является эргодическим.

Зафиксируем k и некоторую $\varphi(x_1,\ldots,x_k),$ заданную на X^k, положим $n = kl$, $z^{(i)} = \varphi(x^{(i)},\ldots,x^{(i+k-1)}),$ $i = 1,\ldots,n,$ и рассмотрим сумму

$$\sum_{i=0}^{n-1} z^{(i)} = \sum_{j=1}^{k}\sum_{s=1}^{l} z^{((s-1)k+j)} = \sum_{j=1}^{k} w_j, \qquad (9.2)$$

где $w_j = \sum_{s=1}^{l} z^{((s-1)k+j)}$; $z^{(i)}$ не являются независимыми, однако $z^{((s-1)k+j)}$, S = 1, 2, ..., l независимы, так как не имеют общих аргументов, поэтому w_j - сумма независимых одинаково распределенных СВ. Используя (9.2),

$$\Pr(|\frac{1}{n}\sum_{i=1}^{n} z^{(i)} - m_z| \geq \varepsilon) = \Pr(|\frac{1}{n}\sum_{j=1}^{k} w_j - m_z| \geq \varepsilon) =$$

$$= \Pr(|\frac{1}{k}\sum_{j=1}^{k} (\frac{1}{l} w_j - m_z)| \geq \varepsilon) \leq \Pr(|\frac{1}{l} w_j - m_z| \geq \varepsilon \text{ хотя бы при одном } j),$$

w_j одинаково распределены, поэтому $\Pr(|\frac{1}{l}w_j - m_z| \geq \varepsilon)$ одинаковы для всех j и

$$\Pr(|\frac{1}{l}w_j - m_z| \geq \varepsilon) = \Pr(|\frac{1}{l}\sum_{s=1}^{l} z^{((s-1)k+j)} - m_z| \geq \varepsilon). \qquad (9.3)$$

Теперь можно применить закон больших чисел. При достаточно больших l (при фиксированном k) и достаточно больших $n = kl$ правая часть (9.3) не превышает $\delta > 0$. Теорема доказана.

Стационарность и независимость - достаточные условия эргодичности. Но, возвращаясь к методу доказательства, видим, что достаточно стационарности и независимости через некоторый интервал $i \geq N$.

Рассмотрим неравенство (источник эргодический)

$$\Pr(|\frac{1}{n}\mathrm{I}(\boldsymbol{x}) - \mathrm{H}(X \mid X^{\infty})| \geq \varepsilon) \leq \delta \ . \qquad (9.4)$$

Оно не совпадает с (9.1)

$$\frac{1}{n}\mathrm{I}(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(x^{(i)} \mid x^{(i-1)},...,x^{(1)}),$$

все слагаемые имеют разные математические ожидания и ни одно из них не совпадает с $\mathrm{H}(X|X^{\infty})$.

Лемма Мак-Миллана. Для любого эргодического ДИ, любых ε и $\delta > 0$ найдется N такое, что для всех $n > N$ выполняется неравенство (9.4) (без доказательства).

Из леммы можно легко вывести теорему о высоковероятных множествах, а также прямую и обратную теоремы кодирования при равномерном кодировании эргодического ДИ.

Теорема 9.2 (о высоковероятных множествах). Пусть U_X - эргодический ДИ и $\{X^n, p(\boldsymbol{x})\}$, $n = 1, 2, ...$ - ансамбли последовательностей длины n на его выходе. Пусть $\mathrm{H}(X|X^{\infty})$ - энтропия на сообщение и ε - некоторое положительное число. Пусть $T_n(\varepsilon)$ - подмножество X^n, определенное как:

$$T_n(\varepsilon) = \{\boldsymbol{x} : \mathrm{H}(X \mid X^{\infty}) - \varepsilon \leq \frac{1}{n}\mathrm{I}(\boldsymbol{x}) \leq \mathrm{H}(X \mid X^{\infty}) + \varepsilon\},$$

где $\mathrm{I}(\boldsymbol{x}) = -\log p(\boldsymbol{x})$ - собственная информация последовательности $\boldsymbol{x} \in X^n$.

Тогда для любых $\varepsilon > 0$ и $\delta > 0$ найдется такое N, что для всех $n > N$ выполняются неравенства

$$\Pr(\boldsymbol{x} \in T_n(\varepsilon)) \geq 1 - \delta,$$

$$(1-\delta)2^{n(\mathrm{H}(X|X^{\infty})-\varepsilon)} \le |\,\mathrm{T}_n(\varepsilon)\,| \le 2^{n(\mathrm{H}(X|X^{\infty})+\varepsilon)}.$$

Доказательство полностью повторяет доказательство теоремы п. 7.

Теорема 9.3 (прямая теорема кодирования). Пусть R > H(X |X^{∞}), тогда для любого δ > 0 существует код со скоростью R, который кодирует эргодический ДИ с вероятностью ошибки меньшей δ.

Теорема 9.4 (обратная теорема кодирования). Пусть R < H(X | X^{∞}), тогда существует зависящее от R такое δ > 0, что для каждого кода со скоростью R, кодирующего эргодический ДИ, вероятность ошибки превосходит δ. И для любой последовательности кодов со скоростью R

$$\lim_{n\to\infty} P_{en} = 1,$$

где P_{en} – вероятность ошибки для кода, кодирующего отрезки сообщений длиной n.

Доказательства теорем 9.3 и 9.4 практически не отличаются от доказательств соответствующих теорем для ДИ без памяти (теоремы 8.1 и 8.2).

Из теорем 9.3 и 9.4 следует, что скорость создания информации эргодическим ДИ при равномерном кодировании равна его энтропии на сообщение H(X | X^{∞}), как и для постоянных источников, - наименьшее количество двоичных символов, приходящихся на одно сообщение на выходе наилучшего двоичного кодера, при условии, что сообщения могут быть восстановлены на выходе кодера сколь угодно точно.

10. Постановка задачи неравномерного кодирования

Если закодировать часто встречающееся сообщение коротким кодовым словом, а редко встречающееся - длинным, то длина кодового слова становится случайной величиной и ее среднюю величину можно уменьшить по сравнению с равномерным кодом $\bar{m}(X) = \sum_{x;\in \mathrm{X}} m_i p(x_i)$, где m_i - длина кодового слова для сообщения x_i на $\{X, p(x)\}$.

Определение. Число $\mathrm{R} = \dfrac{\bar{m}\left(X^n\right)\log D}{n}$ называется средней скоростью неравномерного кодирования посредством D-ичного кода при разбиении последовательности сообщений на блоки длины n.

R зависит от выбора n и множества кодовых слов. Коды, в которых ни одно слово не является началом другого, называются префиксными.

Коды, в которых любая последовательность кодовых слов допускает однозначное разбиение на кодовые слова, называются кодами со свойством однозначного декодирования.

Определение. Скоростью создания информации ДИ при неравномерном кодировании называется наименьшее число H такое, что для любого R > H найдется *n* (длина кодовых сообщений) и неравномерный код со средней скоростью кодирования R, который допускает однозначное декодирование.

Для доказательства того, что H есть скорость создания информации, необходимо доказать прямую и обратную теоремы кодирования. Но прежде сформулируем необходимое условие однозначной декодируемости кода.

Теорема. Предположим, что однозначно декодируемый код состоит из M слов, длины которых $m_1, ..., m_M$ и кодовый алфавит содержит D символов. Тогда

$$\sum_{i=1}^{M} D^{-m_i} \le 1. \qquad (10.1)$$

Пусть L - произвольное положительное число, тогда

$$\left(\sum_{i=1}^{M} D^{-m_i} \right)^L = \sum_{i_1=1}^{M} \dots \sum_{i_L=1}^{M} D^{-\left(m_{i_1} + \dots + m_{i_L}\right)}. \qquad (10.2)$$

Каждое слагаемое в (10.2) - возможная последовательность из L кодовых слов, а сумма $m_{i1} + ... + m_{iL}$ - их соответствующие длины.

Допустим, A_j - число последовательностей с длиной j, тогда

$$\left(\sum_{i=1}^{M} D^{-m_i} \right)^L = \sum_{j=1}^{Lm} A_j D^{-j}, \qquad (10.3)$$

где m - $\max(m_1,...,m_M)$. Код однозначно декодируем, если при любом L и любом j имеется единственная последовательность кодовых символов длины j, образованная L кодовыми словами. Так как D^j - максимальное количество различных последовательностей длины j, то $A_j \le D^j$, подставляя в (10.3), получаем:

$$\left(\sum_{i=1}^{M} D^{-m_i} \right)^L \le Lm, \text{ или}$$

$$\sum_{i=1}^{M} D^{-m_i} \le (Lm)\frac{1}{L} = 2^{1/L \cdot \log mL}.$$

Переходя к пределу при $L \to \infty$, получим утверждение теоремы.

11. Кодовые деревья. Неравенство Крафта

Префиксные коды хорошо описываются специальными графами - кодовыми деревьями.

D-ичное дерево - граф (система узлов и связывающих их ребер), не имеющий замкнутых путей (петель), в котором из каждого узла выходит не более *D* ребер и в каждый узел (кроме корня) входит одно ребро. Каждому ребру сопоставляют один кодовый символ, из одного узла выходят ребра, соответствующие разным символам.

Ярус порядка *i* - узлы дерева, отстоящие от корня на *i* ребер;

Порядок узла - номер его яруса;

Порядок дерева - максимальный из порядков его узлов;

Концевой узел - узел, из которого не выходит ни одного ребра.

Код является префиксным, если кодовые слова соответствуют только концевым узлам.

Теорема (неравенство Крафта). Для существования префиксного кода в алфавите объема *D* с длинами кодовых слов $m_1, ..., m_M$ необходимо и достаточно, чтобы

$$\sum_{i=1}^{M} D^{-m_i} \le 1. \qquad (11.1)$$

Необходимость. Предположим, существует кодовое дерево с концевыми узлами порядков $m_1, ...,m_M$. Максимальное количество узлов на ярусе j - D^j. Пусть $m = \max\{m_1, ...,m_M\}$. Рассмотрим концевой узел порядка m_i, отстоящий от яруса m на m-m_i ребер и, следовательно, исключающий из яруса m D^{m-m_i} возможных узлов.

Так как количество исключаемых узлов не может быть больше общего количества узлов яруса m, то

$$\sum_{i=1}^{M} D^{m-m_i} \le D^m .$$

Разделим неравенство на D^m и получим исходное (11.1).

Достаточность. Необходимо доказать, что при выполнении (11.1) дерево с концевыми узлами порядков $m_1, ...,m_M$ может быть построено. Предположим, что среди $\{m_1, ...,m_M\}$ число S встречается ровно α_S раз (S =1, 2, ..., m), тогда

$$\sum_{i=1}^{M} D^{-m_i} = \sum_{S=1}^{m} \alpha_S D^{-S} \le 1; \qquad (11.2)$$

$$\sum_{S=1}^{i-1} \alpha_S D^{-S} + \alpha_i D^{-i} + \sum_{S=i+1}^{m} \alpha_S D^{-S} \le 1;$$

$$\alpha_i \le D^i - \sum_{S=1}^{i-1} \alpha_S D^{i-S} - \sum_{S=i+1}^{m} \alpha_S D^{i-S} \le D^i - \sum_{S=1}^{i-1} \alpha_S D^{i-S}. \quad (11.3)$$

Применим метод полной индукции. Дерево, содержащее α_1 концевых узлов порядка 1, может быть построено из (11.2) $\alpha_1 D^{-1} \le 1$ и $\alpha_1 \le D$ - максимально возможное количество концевых узлов порядка 1.

Предположим, что дерево с α_S концевыми узлами порядка S = 1, 2, ..., i-1 может быть построено. Докажем, что к этому дереву можно добавить еще α_i концевых узлов порядка i.

Если верно предположение индукции, то из яруса порядка i исключаются $\sum_{S=1}^{i-1} \alpha_S \cdot D^{i-S}$ возможных концевых узлов; остается $D^i - \sum_{S=1}^{i-1} \alpha_S \cdot D^{i-S}$ свободных узлов в ярусе i, но из (11.3) следует, что количество свободных узлов превосходит α_i, т.е. то, которое необходимо добавить. Следовательно, к дереву α_S с концевыми узлами порядка S =1, 2, ..., i-1 могут быть добавлены α_i концевых узлов порядка i.

12. Неравномерное кодирование стационарных источников

Вначале некоторые вспомогательные утверждения. Пусть $\{X, p(x)\}$ – произвольный дискретный ансамбль с энтропией H(X), $\overline{m}(X) = \sum_{i=1}^{M} m_i p(x_i)$ - средняя длина слов D-ичного кода, сопоставляющихся сообщениям ансамбля X.

Теорема 12.1. Для любого кода со свойством однозначного декодирования

$$\overline{m}(X) \ge \frac{\mathrm{H}(X)}{\log D};$$

$$\overline{m}(X)\log D = \sum_{i=1}^{M} m_i \cdot p(x_i)\log D = \sum_{i=1}^{M} p(x_i)\log D^{m_i};$$

$$\mathrm{H}(X) - \overline{m}(X)\log D = -\sum_{i=1}^{M} p(x_i)\log p(x_i) + \sum_{i=1}^{M} p(x_i)\log D^{-m_i} = \quad (12.1)$$

$$= \sum_{i=1}^{M} p(x_i)\log\frac{D^{-m_i}}{p(x_i)} \le \log e \sum_{i=1}^{M} p(x_i)\left(\frac{D^{-m_i}}{p(x_i)} - 1\right) \le 0.$$

Последнее неравенство - следствие (10.1). Теорема доказана.

Неравенства (12.1) превращаются в равенства при

$$p(x_i) = D^{-m_i},\ i = 1,2,...,M\ . \qquad (12.2)$$

Таким образом, если выполняется (12.2), то существует D-ичное дерево с концевыми узлами порядков m_1, ..., m_M и соответствующий D-ичный код будет иметь среднюю длину $\overline{m}(X) = \dfrac{\mathrm{H}(X)}{\log D}$, т.е. равную нижней границе.

Коды, для которых средняя длина кодовых слов (и соответственно скорость кодирования) равна наименьшему значению, называются оптимальными.

Теорема 12.2. Существует D-ичный код со свойством однозначного декодирования, для которого $\overline{m}(X) < \dfrac{\mathrm{H}(X)}{\log D} + 1$.

Пусть m_i' - наименьшее целое число такое, что $m_i' \geq \dfrac{\mathrm{I}(x_i)}{\log D}$, i = 1, 2, ..., M.

Очевидно, $\dfrac{\mathrm{I}(x_i)}{\log D} \leq m_i' < \dfrac{\mathrm{I}(x_i)}{\log D} + 1$. Поскольку $\sum\limits_{i=1}^{M} D^{-m_i'} \leq \sum\limits_{i=1}^{M} D^{-\frac{\mathrm{I}(x_i)}{\log D}} =$

$= \sum\limits_{i=1}^{M} p(x_i) = 1$, то, по неравенству Крафта, существует дерево с концевыми узлами порядков $m_1', ..., m_M'$. Соответствующий код будет иметь среднюю длину

$$\overline{m}(X) = \sum_{i=1}^{M} m_i' p(x_i) < \frac{\mathrm{H}(X)}{\log D} + 1\ .$$

Теорема доказана.

Обобщим результат на случай кодирования последовательностей. Пусть $\left\{X^n, p(\boldsymbol{x})\right\}$ - произвольный дискретный ансамбль последовательностей сообщений; n - длина последовательностей; $\mathrm{H}(X^n)$ - энтропия ансамбля. Тогда для любого D-ичного кода, однозначно кодирующего последовательности из X^n, среднее количество $\overline{m}$ символов на сообщение

$$\bar{m} = \frac{\bar{m}\left(X^n\right)}{n} \geq \frac{\mathrm{H}\left(X^n\right)}{n \log D}. \qquad (12.3)$$

Существует код, для которого $\bar{m} < \dfrac{\mathrm{H}\left(X^n\right)}{n \log D} + \dfrac{1}{n}$. (12.4)

Рассмотрим стационарный источник U_X, выбирающий сообщения из множества X, с энтропией на сообщение $\mathrm{H}\left(X \middle| X^\infty\right)$. С помощью D-ичного кода каждой последовательности сообщений $\boldsymbol{x} = \left(x^{(1)},\ldots,x^{(n)}\right) \in X^n$ ставится в соответствие кодовое слово длины $m(\boldsymbol{x})$. Средняя длина слова $\bar{m}\left(X^n\right) = \sum\limits_{X^n} p(\boldsymbol{x}) m(\boldsymbol{x})$, среднее число кодовых символов на сообщение $\bar{m} = \dfrac{\bar{m}\left(X^n\right)}{n} = \dfrac{1}{n} \sum\limits_{X^n} p(\boldsymbol{x}) m(\boldsymbol{x})$. Средняя скорость кодирования

$$\mathrm{R} = \bar{m} \log D = \frac{\log D}{n} \sum_{X^n} p(\boldsymbol{x}) m(\boldsymbol{x}). \qquad (12.5)$$

Теорема 12.3 (прямая теорема кодирования). Для любого кода, кодирующего источник U_X однозначно, средняя скорость кодирования

$$\mathrm{R} \geq \mathrm{H}\left(X \middle| X^\infty\right).$$

Из (12.5), (12.3) и теорем 5.1 и 5.2 следует, что для всех n = 1, 2,... выполняются неравенства

$$\mathrm{R} \geq \frac{1}{n}\mathrm{H}\left(X^n\right) \geq \mathrm{H}\left(X \middle| X^{n-1}\right) \geq \mathrm{H}\left(X \middle| X^\infty\right).$$

Теорема 12.4 (обратная теорема кодирования). Пусть ε - произвольное положительное число, D - число элементов кодового алфавита. Существует однозначно декодируемый D-ичный код, кодирующий источник U_X, для которого $\mathrm{R} < \mathrm{H}\left(X \middle| X^\infty\right) + \varepsilon$.

Согласно теореме 5.2 для любого $\varepsilon_1 > 0$ найдется такое $N(\varepsilon_1)$, что для всех $n > N(\varepsilon_1)$:

$$\frac{1}{n}\mathrm{H}\left(X^n\right) < \mathrm{H}\left(X \middle| X^\infty\right) + \varepsilon_1.$$

Отсюда и из (12.4) вытекает, что для произвольного целого $D > 0$ существует однозначно декодируемый D-ичный код со средним количеством символов на сообщение

$$\bar{m} < \frac{\mathrm{H}\left(X^n\right)}{n\log D} + \frac{1}{n} < \frac{\mathrm{H}\left(X\middle|X^\infty\right)}{\log D} + \frac{\varepsilon_1}{\log D} + \frac{1}{n}$$

и, следовательно, скоростью кодирования

$$R < \mathrm{H}\left(X\middle|X^\infty\right) + \varepsilon_1 + \frac{\log D}{n}.$$

Полагая $\varepsilon_1 = \varepsilon/2$, получим, что для всех $n > \frac{\log D}{\varepsilon - \varepsilon_1} = \frac{2\log D}{\varepsilon}$ имеет место $\varepsilon_1 + \frac{\log D}{n} \le \varepsilon$.

Утверждение теоремы справедливо для всех n, больших, чем максимальное из $N(\varepsilon/2)$ и $2\log D/\varepsilon$. Теорема доказана.

Из теорем 12.3 и 12.4 следует, что средняя скорость создания информации таким источником равна энтропии на сообщении $\mathrm{H}(X|X^\infty)$, т.е. той же величине, что и при равномерном кодировании. Минимальное количество символов на сообщение может быть сделано сколь угодно близким к $\mathrm{H}(X|X^\infty)$ при обоих методах кодирования.

13. Оптимальные неравномерные коды

Оптимальным называется код, средняя длина кодовых слов которого равна минимально возможной.

Рассмотрим методы построения оптимальных кодов. Возьмем простейший случай: $p(x_i) = D^{-m_i}$, i =1, 2, ..., M. В этом случае существует оптимальный D-ичный однозначно декодируемый код с $\bar{m}(x) = \frac{\mathrm{H}(X)}{\log D}$ (см. теорему 12.1). Сообщению x_i сопоставляется слово длины $m_i = \frac{-\log p(x_i)}{\log D}$.

Всякое дерево с набором концевых вершин m_1, ..., m_M в этом случае дает оптимальный код.

Рассмотрим метод построения такого дерева (Шеннон - Фано).

Делим множество сообщений на D подмножеств с одинаковыми суммарными вероятностями и каждому подмножеству присваиваем первый кодовый символ α_i, i = 1, 2, ..., D, A = { α_1, α_2, ..., α_D } - кодовый алфавит.

С каждым из полученных подмножеств поступаем аналогичным образом, присваивая последующие кодовые символы, и так, пока в каждом из подмножеств не останется по одному сообщению (табл. 13.1).

Таблица 13.1

x	$p(x_i)$	1 шаг	2 шаг	3 шаг	Кодовое слово
x_1	1/2	I			0
x_2	1/4	II	I		10
x_3	1/8		II	I	110
x_4	1/8			II	111

Рассмотрим построение оптимального префиксного кода ($D = 2$) в случае произвольных $p(x_i)$; $p(x_1) \geq p(x_2) \geq ... \geq p(x_M)$.

Лемма 13.1. В оптимальном коде слово, соответствующее наименее вероятному сообщению, имеет наибольшую длину.

Предположим $m_i > m_M$ для некоторого $i<M$. Поменяем местами m_i и m_M кодовые слова, тогда:

$$\bar{m}' = \bar{m} - p(x_i)m_i - p(x_M)m_M + p(x_i)m_M + p(x_M)m_i =$$

$$= \bar{m} - (m_i - m_M)(p(x_i) - p(x_M)) < \bar{m},$$

что противоречит оптимальности полученного кода.

Лемма 13.2. В оптимальном двоичном префиксном коде два наименее вероятных сообщения кодируются словами одинаковой длины, одно из которых оканчивается нулем, другое - единицей. Последний символ кода служит отличительной особенностью от другого кодового слова такой же длины; если такого слова нет, то код можно укоротить, уменьшив среднюю длину кодовых слов, следовательно, в оптимальном коде существует два самых длинных кодовых слова, отличающихся в самом последнем символе.

Предположим, что одно из них стоит не на $(M-1)$-м месте, а на i-м $(i < M - 1)$, тогда $m_i = m_M > m_{M-1}$, следовательно, есть возможность уменьшить $\bar{m}$, поменяв местами i-е и $M-1$-е слова, что противоречит оптимальности исходного кода.

Если в X' $M - 1$ сообщений и существует префиксный декодируемый код для него, то для X, состоящего из M сообщений, код можно получить добавлением к кодирующему слову для x_{M-1} “0”, а для x_M - “1”.

Лемма 13.3. Если оптимален однозначно декодируемый префиксный код для X', то оптимален полученный из него префиксный код для X. Обозначим: $\bar{m}'$ - средняя длина кодовых слов, $p(x_{M-1}) + p(x_M) = p(x'_{M-1})$.

$$\bar{m} = \sum_{i=1}^{M} m_i p(x_i) = \sum_{i=1}^{M-2} m_i p(x_i) + m_{M-1} p(x_{M-1}) + m_M p(x_M) =$$

$$= \sum_{i=1}^{M-1} m'_i p(x'_i) - m'_{M-1}\left[p(x_{M-1}) + p(x_M)\right] + m_{M-1} p(x_{M-1}) + m_M p(x_M) =$$

$$= \bar{m}' + p(x_{M-1})(m_{M-1} - m'_{M-1}) + p(x_M)(m_M - m'_{M-1}),$$

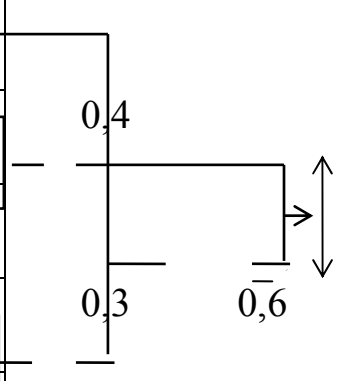
но $\begin{cases} m_i = m'_i, i = 1,2,\ldots, M-2 \\ m_M = m_{M-1} = m'_{M-1} + 1 \end{cases}$ и $\bar{m} = \bar{m}' + p(x'_{M-1})$,

$\bar{m}$ и $\bar{m}$ отличаются на константу $p(x'_{M-1})$, не зависящую от выбора кодовых слов, поэтому $\bar{m}$ - минимальна.

Таким образом, объединяя наименее вероятные сообщения, последовательно можно прийти к ансамблю, содержащему два сообщения “0” и “1”, провести обратную процедуру, присваивая каждый раз новые кодовые символы. Этот метод построения оптимального префиксного кода называется методом Хаффмена (табл. 13.2).

Т а б л и ц а 13.2

x_i	$p(x_i)$	Кодовое слово
x_1	0,3	01
x_2	0,25	11
x_3	0,15	10
x_4	0,15	001
x_5	0,15	000



Вопросы для самопроверки

1. В чем различие в определений дискретных ансамблей и источников?
2. Каков механизм перехода от случайных событий к случайным величинам?

3. Каковы условия выполнения закона больших чисел?

4. От чего зависит количество информации в сообщении?

5. Сформулируйте свойства энтропии и условной энтропии.

6. Для каких источников и почему применимо понятие «энтропия на сообщение»?

7. Опишите процесс равномерного кодирования дискретного источника.

8. Что такое – высоковероятные множества?

9. Сформулируйте понятие «скорость создания информации» для дискретного источника без памяти.

10. Какие источники называются эргодическими?

11. Какая проблема возникает при переходе от равномерного кодирования к неравномерному?

12. Дайте определение префиксного кода.

13. В чем отличие префиксного кода от однозначно декодируемого кода?

14. Какое свойство кода определяется с помощью неравенства Крафта?

15. Для чего служат кодовые деревья?

16. Перечислите признаки оптимального кода.

ВЗАИМНАЯ ИНФОРМАЦИЯ И ЕЕ СВОЙСТВА

В этом разделе определяется информация между сообщениями и ансамблями и изучаются их свойства.

Трудность - отсутствие понятия собственной информации для непрерывных ансамблей, которая преодолевается с помощью введения функции плотности вероятности (ФПВ).

При совместном рассмотрении дискретных и непрерывных ансамблей используется техника обобщенных функций, с помощью которой вводится понятие ФПВ дискретных случайных величин.

14. Количество информации между дискретными ансамблями

Определение. Количеством информации в сообщении $x \in X$ о сообщении $y \in Y$ называется величина

$$\mathrm{I}(x;y) = \mathrm{I}(y) - \mathrm{I}(y|x) = \log \frac{p(y|x)}{p(y)}.$$

Может принимать различные по величине и знаку значения, но могут возникать неопределенности типа 0/0, либо если условная вероятность не

определена. Решение*:* доопределение или устранение сообщений, вероятности которых равны 0.

Так как $p(x,y) = p(x|y)p(y) = p(y|x)p(x)$, то $\mathrm{I}(x;y) = \mathrm{I}(x) - \mathrm{I}(x|y) = \mathrm{I}(y;x)$ и $\mathrm{I}(y;x)$ - количество взаимной информации между сообщениями x и y (взаимная информация).

В симметричной форме $\mathrm{I}(x;y) = \log\dfrac{p(x,y)}{p(x)p(y)}$ для трех ансамблей X, Y и Z справедливы соотношения

$$\mathrm{I}(y;z \mid x) = \mathrm{I}(y \mid x) - \mathrm{I}(y \mid xz) = \log\frac{p(y \mid xz)}{p(y \mid x)};$$

$$\mathrm{I}((x,y);z) = \mathrm{I}(x,y) - \mathrm{I}((x,y) \mid z) = \log\frac{p(x,y \mid z)}{p(x,y)}.$$

На основе свойства аддитивности собственной информации

$$\mathrm{I}((x,y);\mathrm{z}) = \mathrm{I}(x) + \mathrm{I}(y \mid x) - \mathrm{I}(x \mid z) - \mathrm{I}(y \mid xz) =$$

$$= \mathrm{I}(x;z) + \mathrm{I}(y;z;x) = \mathrm{I}(y;z) + \mathrm{I}(x;z \mid y),$$

свойство аддитивности взаимной информации.

Взаимная информация - случайная величина на ансамбле и обладает различными числовыми характеристиками.

Определение. Математическое ожидание случайной величины $\mathrm{I}(x;y)$ на ансамбле $\{XY, p(x,y)\}$ называется средним количеством взаимной информации (средней взаимной информацией) между ансамблями $\{X, p(x)\}$ и $\{Y, p(y)\}$ и обозначается

$$\mathbf{I}(X;Y) = \mathrm{MI}(x;y) = \sum_{XY} p(x,y)\log\frac{p(x|y)}{p(x)}.$$

Определение. Математическое ожидание случайной величины $\mathrm{I}(x;y)$ на ансамбле $\{X, p(x|y)\}$, $p(y) \neq 0$ называется средней взаимной информацией между ансамблем X и сообщением $y \in Y$ и обозначается

$$\mathbf{I}(X;y) = \mathrm{M}_y\mathrm{I}(x;y) = \sum_{X} p(x|y)\log\frac{p(x|y)}{p(x)}.$$

Аналогично $\mathbf{I}(x;Y) = \mathrm{M}_x\mathrm{I}(x;y) = \sum_{Y} p(y \mid x)\log\dfrac{p(y \mid x)}{p(y)}$.

Средняя взаимная информация не определена для сообщений, вероятность которых равна 0; ее можно доопределить, но не зависимо от способа доопределения:

$$\mathrm{M}\mathbf{I}(X;y)=\sum_{Y}p(y)\mathbf{I}(X;y)=\mathbf{I}(X;Y);$$
$$\mathrm{M}\mathbf{I}(x;Y)=\sum_{X}p(x)\mathbf{I}(x;Y)=\mathbf{I}(X;Y), \qquad (14.1)$$

т.е. среднюю взаимную информацию можно определить двояко (как в определении и как в соотношениях (14.1)).

Математическое ожидание собственной информации - энтропия ансамбля, поэтому по аналогии

$$\mathbf{I}(X;Y)=\mathrm{H}(X)-\mathrm{H}(X|Y)=\mathrm{H}(Y)-\mathrm{H}(Y|X). \qquad (14.2)$$

Свойства средней взаимной информации

Теорема 14.1. Средняя взаимная информация между сообщением, вероятность которого отлична от нуля, и ансамблем, а также средняя взаимная информация между двумя ансамблями неотрицательна.

Покажем, что $\mathbf{I}(X;y)\ge 0$, тогда второе утверждение будет следовать из (14.1). Рассмотрим

$$-\mathbf{I}(X;y)=\sum_{X}p(x|y)\log\frac{p(x)}{p(x|y)}\le\log e\sum_{X}p(x|y)\left[\frac{p(x)}{p(x|y)}-1\right]=0.$$

$\mathbf{I}(X;y)=\mathbf{I}(X;Y)=0$ только при статистической независимости всех x и y. Теорема доказана.

Рассмотрим ансамбль троек $\{XYZ,\ p(x,y,z)\}$.

Определение. Математическое ожидание случайной величины $\mathrm{I}(x;y|z)$ на условном ансамбле $\{XY,\ p(x,\ y|z)\}$ называется средней взаимной информацией между ансамблями X и Y относительно сообщения z из ансамбля Z и обозначается

$$\mathbf{I}(X;Y\,|\,z)=\mathrm{M}_{z}\mathrm{I}(x;y\,|\,z)=\sum_{XY}p(x,y\,|\,z)\log\frac{p(x\,|\,yz)}{p(x\,|\,z)},$$

случайная величина на ансамбле $\{Z,\ p(z)\}$.

Определение. Математическое ожидание случайной величины $\mathrm{I}(X;Y|z)$ на ансамбле $\{Z,\ p(z)\}$ называется средней взаимной информацией между ансамблями X и Y относительно Z и обозначается

$$\mathbf{I}(X;Y|Z)=\mathrm{M}\mathbf{I}(X;Y|z)=\sum_{XYZ}p(x,y,z)\log\frac{p(x|yz)}{p(x|z)},$$

Можно определить также

$$\mathbf{I}(XY;Z)=\mathrm{MI}((x,y);z)=\sum_{XYZ}p(x,y,z)\log\frac{p(xy\,|\,z)}{p(x,y)}.$$

Из свойства аддитивности

$$\mathbf{I}(XY;Z)=\mathbf{I}(X;Z)+\mathbf{I}(Y;Z|X)=\mathbf{I}(Y;Z)+\mathbf{I}(X;Z|Y), \qquad (14.3)$$

а из (14.2)

$$\mathbf{I}(XY;Z) = \mathrm{H}(XY) - \mathrm{H}(XY|Z) = \mathrm{H}(Z) - \mathrm{H}(Z|XY).$$

Для доказательства неувеличения средней взаимной информации при преобразованиях введем некоторое преобразование $\varphi(\bullet)$, отображающее элементы множества X на элементы множества Z (в общем случае неоднозначно). Предположим, что задан ансамбль $\{XY, p(x, y)\}$ со средней взаимной информацией $\mathbf{I}(X;Y)$, тогда преобразование $\varphi(\bullet)$ определяет ансамбль $\{ZY, p(z, y)\}$, для которого

$$p(z,y) = \sum_{x:\varphi(x)=z} p(x,y).$$

Теорема 14.2. Для любого отображения $Z = \varphi(X)$ ансамбля X в ансамбль Z

$$\mathbf{I}(X;Y) \geq \mathbf{I}(Z;Y). \qquad (14.4)$$

Равенство соблюдается при обратимости отображения (взаимной однозначности).

Рассмотрим множество XYZ. Так как при выбранном x сообщение z однозначно определено и не зависит от y, то $p(z|xy)=p(z|x)$ или $p(x, y, z) = p(x, y)p(z|x)$, откуда следует, что

$$\mathrm{I}(y;z|x) = \log\frac{p(z|xy)}{p(z|x)} = 0 = \mathbf{I}(Y;Z|X).$$

Учитывая (14.3),

$$\mathbf{I}(XZ;Y) = \mathbf{I}(X;Y) + \mathbf{I}(Y;Z|X) = \mathbf{I}(X;Y).$$

С другой стороны, в силу неотрицательности средней взаимной информации $\mathbf{I}(X;Y|Z)$

$$\mathrm{I}(XZ;Y) = \mathrm{I}(X;Y) + \mathrm{I}(X;Y \mid Z) \geq \mathrm{I}(Z;Y),$$

что и доказывает (14.4).

Равенство в (14.4) возможно только в случае $\mathbf{I}(X;Y|Z) = 0$, что справедливо при $p(x|yz) = p(x|z)$ для всех $(x\ ,\ y,\ z) \in XYZ$, т.е. при статистической независимости x и y при выбранном z. Это условие всегда выполняется, если x и z однозначно определяют друг друга, т.е. отображение $\varphi(\bullet)$ обратимо.

Утверждение теоремы справедливо не только для детерминированного отображения X в Z, но и при произвольных случайных отображениях, при условии независимости z и y при выбранном x.

Физическое толкование: никакая обработка сигнала (сообщения) не добавляет информации о нем.

15. Непрерывные ансамбли и источники. Обобщение понятия количества информации

Непрерывные источники, порождающие непрерывные ансамбли, важный класс источников сообщений (например, речь).

Определим вероятностные характеристики непрерывных источников (ансамблей).

$F(x)$ - функция распределения случайной величины x

$$F(x) = \Pr(-\infty < x' \le x).$$

Если существует $f(x)$ такая, что для всех x на числовой оси

$$F(x) = \int_{-\infty}^{x} f(x')dx',$$

то она называется функцией плотности вероятности (ФПВ), или просто плотностью вероятности случайной величины X.

Вероятность появления СВ в любом интервале определяется как

$$\Pr(a,b) = F(b) - F(a) = \int_{a}^{b} f(x)dx.$$

$F(x)$ - неотрицательна и монотонно не убывает, $F(-\infty) = 0$, $F(\infty) = 1$. ФПВ неотрицательна и ее интеграл в бесконечных пределах равен единице - условие нормировки.

$F(x)$ - универсальная функция и применима к дискретной СВ - в этом случае $F(x)$ - ступенчатая функция с конечным числом ступеней (ФПВ в обычном смысле не существует). Если для $F(x)$ в каждой точке может быть определена производная, то распределение соответствует непрерывной СВ. Существуют смешанные типы распределения: $F(x)$ непрерывна (справа) везде, за исключением конечного числа точек (ступеньки); и, наконец, $F(x)$ непрерывна в каждой точке (справа), но ФПВ всюду либо на каком-то интервале не существует.

Определение. Непрерывным ансамблем, задаваемым плотностью вероятности $f(x)$, будем называть пару $\{X, f(x)\}$, где X - числовая ось и распределение вероятностей задается ФПВ $f(x)$.

Пара ансамблей (совместно заданных) - действительная плоскость XY, $F(x,y)$ - совместная функция распределения на множестве XY, а

$$F_1(x) = F(x,\infty),\ F_2(y) = F(\infty, y).$$

Определение. Пусть распределение вероятностей на X, Y и XY задают ФПВ $f_1(x), f_2(y), f(x, y)$, определяемые соотношениями:

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)dx,dy;\ f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy;\ f_2(y) = \int_{-\infty}^{\infty} f(x,y)dx\,.$$

В этом случае $\{XY, f(x, y)\}$ есть система двух совместно заданных непрерывных ансамблей $\{X, f_1(x)\}$ и $\{Y, f_2(y)\}$. Аналогично с дискретными ансамблями вводятся условные непрерывные ансамбли $\{X, f(x \mid y)\}$ и $\{Y, f(y \mid x)\}$, где $f(x|y) = f(x, y)/f_2(y)$.

Аналогично вводятся n-мерные непрерывные ансамбли и понятие статистической независимости:

$$f(x^{(1)}, ..., x^{(n)}) = f_1(x^{(1)}) \ldots f_n(x^{(n)}).$$

Все числовые характеристики непрерывных СВ определяются так же, как и в случае дискретных с заменой вероятностей на ФПВ, а сумм - на интегралы. Для непрерывных СВ справедливы неравенства Чебышева и закон больших чисел.

Переход от непрерывного ансамбля к дискретному: разбиваем Y на непересекающиеся подмножества $\{y_1, ..., y_N\}$, где событие y_i соответствует попаданию непрерывной СВ в подмножество B_i и $p(y_i) = \int_{B_i} f(y)dy$.

Этот процесс называется процессом дискретизации. При совместном задании непрерывного X и дискретизированного ансамбля Y возникает семейство условных вероятностей и условных плотностей вероятности

$$f(x \mid y_i) = f(x \mid B_i) = \frac{1}{p_2(y_i)}\int_{B_i} f(x,y)dy,\ \ p_2(y) \neq 0,$$

$$\text{откуда } f_1(x) = \sum_{i=1}^{N} f(x \mid y_i)\,p_2(y_i)$$

$$\text{и } \ p(y_i \mid x) = \int_{B_i} \frac{f(x,y)}{f_1(x)}dy = \frac{f(x \mid y_i)\,p_2(y_i)}{f_1(x)},\quad f_1(x) \neq 0\,.$$

Таким образом, ансамбль $XY_д$ можно задавать двумя эквивалентными способами: $f_1(x)p(y_i|x)$ либо $f(x|y_i)p_2(y_i)$. Обе функции совпадают и определяют функцию распределения на множестве $XY_д$:

$$F(x,y)=\begin{cases}\int\limits_{-\infty}^{x}\sum\limits_{y_i\le y} f_1(x)p(y_i\mid x)dx;\\ \int\limits_{-\infty}^{x}\sum\limits_{y_i\le y} f(x\mid y_i)p_2(y_i)dx.\end{cases}$$

Совместное рассмотрение дискретных и непрерывных ансамблей можно проводить с помощью функций распределения, но в терминах ФПВ. Это проще, поэтому введем понятие ФПВ для дискретных множеств с помощью обобщенной функции - дельта-функции Дирака, определяемой следующим формальным равенством:

$$\int\limits_{\Delta}\delta(x)\varphi(x)dx=\begin{cases}\varphi(0),\ 0\in\Delta\\ 0,\ 0\notin\Delta\end{cases},$$

где $\varphi(x)$ - произвольная непрерывная функция; Δ - произвольный интервал на числовой оси.

Очевидно, что

$$\int\limits_{\Delta}\delta(x-x_0)\varphi(x)dx=\varphi(x_0)\ \text{ и }\int\limits_{\Delta}\delta(x-x_0)dx=1,$$

если $x_0\in\Delta$.

Дельта-функцию можно умножать на число, складывать с другими дельта-функциями и с обычными интегрируемыми функциями.

Формальное выражение для ФПВ дискретного множества X

$$f^*(x)=\sum_{i=1}^{M}p_i\delta(x-x_i),$$

так как $F(x)=\sum\limits_{x_i\le x}p(x_i)=\int\limits_{-\infty}^{x}f^*(x')dx'$.

Для произведения дискретных множеств XY

$$f^*(x,y)=\sum_{i,j}p(x_i,y_j)\delta(x-x_i)\delta(y-y_j).$$

Отношение выражений, содержащих дельта-функции, вообще говоря, не определено, и поэтому условные ФПВ не являются ни обычными, ни обобщенными ФПВ. Однако если положить

$$\frac{\sum_{j=1}^{N} a_j \delta(y - y_j)}{\sum_{j=1}^{N} b_j \delta(y - y_j)} = \frac{a_i}{b_i}, \quad y = y_i, \quad i = 1,..., N$$

для всех выражений такого вида с $b_i \neq 0$, то

$$f^*(x \mid y) = \frac{\sum_{i,j} p(x_i, y_j)\delta(x - x_i)\delta(y - y_j)}{\sum_{i,j} p(x_i, y_j)\delta(y - y_j)} = \sum_{i,j}^{N} \frac{p(x_i, y_j)}{p_2(y_j)}\delta(x - x_i),$$

где $p_2(y_j) = \sum_i p(x_i, y_j)$, и функция $f^*(x \mid y)$ становится обобщенной ФПВ для всех $y = y_j$. Для остальных y эта функция не определена. Кроме того,

$$\frac{f^*(x, y)}{f_1^*(x) f_2^*(y)} = \frac{p(x_i, y_j)}{p_1(x_i) p_2(y_j)},$$

где $p_1(x_i) = \sum_j p(x_i, y_j)$.

В случае непрерывных ансамблей вероятность каждого “сообщения” равна нулю и, следовательно, собственная информация бесконечна. Однако взаимная информация, как правило, ограничена.

Определение. Количеством взаимной информации между сообщениями $x \in X$ и $y \in Y$ непрерывного ансамбля $\{XY, f(x,y)\}$ называется величина

$$\mathrm{I}(x; y) = \log \frac{f(x, y)}{f(x) f(y)}, \tag{15.1}$$

определенная для всех пар (x,y) таких, что $f(x) \neq 0, f(y) \neq 0$. Для остальных пар сообщений количество взаимной информации в случае необходимости доопределяется произвольным образом.

Выражение (15.1) аналогично выражению, определяющему взаимную информацию между сообщениями дискретных ансамблей.

Для непрерывного ансамбля $\{XYZ, f(x,y,z)\}$ условная информация между x и y при фиксированном z

$$\mathrm{I}(x; y \mid z) = \log \frac{f(y \mid xz)}{f(y \mid z)} = \log \frac{f(x, y \mid z)}{f(x \mid z) f(y \mid z)};$$

информация между парой сообщений (x,y) и z

$$\mathrm{I}((x, y); z) = \log \frac{f(x, y \mid z)}{f(x, y)}.$$

Математическое ожидание СВ I(x; y)

$$\mathbf{I}(X;Y) = \mathrm{MI}(x;y) = \iint\limits_{XY} f(x,y)\log\frac{f(x,y)}{f(x)f(y)}dxdy \qquad (15.2)$$

называется средней взаимной информацией между непрерывными ансамблями X и Y. Математическое ожидание

$$\mathbf{I}(X;Y \mid z) = \mathrm{M_Z I}(x;y \mid z) = \iint\limits_{XY} f(x,y \mid z)\log\frac{f(xy \mid z)}{f(x \mid z)f(y \mid z)}dxdy,$$

средняя взаимная информация между непрерывными ансамблями X и Y относительно сообщения z ансамбля Z. Математическое ожидание

$$\mathbf{I}(X;Y \mid Z) = \mathrm{MI}(x;y \mid z) = \iiint\limits_{XYZ} f(x,y,z)\log\frac{f(x,y \mid z)}{f(x \mid z)f(y \mid z)}dxdydz -$$

средняя взаимная информация между непрерывными ансамблями X и Y относительно непрерывного ансамбля Z. Математическое ожидание

$$\mathbf{I}(XY;Z) = \mathrm{MI}((xy);z) = \iiint\limits_{XYZ} f(x,y,z)\log\frac{f(x,y \mid z)}{f(x,y)}dxdydz \ -$$

средняя взаимная информация между непрерывным ансамблем XY и непрерывным ансамблем Z.

Приведенные формулы могут быть использованы и в случае, когда часть ансамблей являются непрерывными, а часть - дискретными. В этом случае ФПВ для дискретных ансамблей рассматриваются как обобщенные либо используются дискретные функции распределения, но тогда соответствующие интегралы должны быть заменены на суммы.

Средняя взаимная информация между непрерывными или непрерывными и дискретными ансамблями обладает многими из свойств, ранее сформулированными для дискретных ансамблей, так теоремы 14.1 и 14.2 остаются справедливыми.

Определение непрерывного источника (с дискретным временем) опирается на определение последовательности непрерывных ансамблей. Все остальное, что было сказано о задании дискретных источников (в том числе стационарность источника без памяти), переносится без сущест-

венных изменений на случай непрерывных источников. Единственное отличие - замена вероятности на ФПВ, а сумм - на интегралы.

16. Относительная энтропия и ее свойства

Все свойства средней взаимной информации для дискретных и непрерывных ансамблей являются общими. Отличие - отсутствие собственной информации на непрерывных ансамблях и, как следствие, отсутствие определения энтропии.

Однако можно ввести некоторые аналоги энтропии в непрерывном случае. Рассмотрим среднюю взаимную информацию между непрерывными ансамблями X и Y:

$$\mathbf{I}(X;Y) = \iint\limits_{XY} f(x,y)\log\frac{f(x\,|\,y)}{f(x)}dxdy = \iint\limits_{XY} f(x,y)\frac{f(y\,|\,x)}{f(y)}dxdy\,.$$

Обозначим

$$\mathrm{H}_0(X) = -\int\limits_X f(x)\log f(x)dx; \quad \mathrm{H}_0(Y) = -\int\limits_Y f(y)\log f(y)dy,$$
$$\mathrm{H}_0(X\,|\,Y) = -\iint\limits_{XY} f(x,y)\log f(x\,|\,y)dxdy,$$
$$\mathrm{H}_0(Y\,|\,X) = -\iint\limits_{XY} f(x,y)\log f(y\,|\,x)dxdy,$$

используя (15.1), получим $\mathbf{I}(X;Y) = \mathrm{H}_0(X) - \mathrm{H}_0(X\,|\,Y) = \mathrm{H}_0(Y) - \mathrm{H}_0(Y\,|\,X)$.

Величины $\mathrm{H}_0(\bullet)$, при условии существования соответствующих интегралов, называются относительными (дифференциальными) энтропиями непрерывных ансамблей.

В случае непрерывных ансамблей относительная энтропия, в отличие от дискретного случая, может принимать различные по знаку значения.

Величина

$$\mathrm{H}_0(XY) = -\iint\limits_{XY} f(x,y)\log f(x,y)dxdy$$

называется относительной энтропией ансамбля XY. Представляя $f(x,y)$ в виде произведения условной и безусловной ФПВ, получим:

$$\mathrm{H}_0(XY) = \mathrm{H}_0(X) + \mathrm{H}_0(Y\,|\,X) = \mathrm{H}_0(Y) + \mathrm{H}_0(X\,|\,Y),$$

свойство аддитивности относительной энтропии.

Используя равенство $f(\boldsymbol{x}) = f(x^{(1)})f(x^{(2)} \mid x^{(1)})...f(x^{(n)} \mid x^{(1)}...x^{(n-1)})$, получим:

$$\mathrm{H}_0(X_1...X_n) = -\int f(\boldsymbol{x}) \log f(\boldsymbol{x}) d\boldsymbol{x} =$$
$$= \mathrm{H}_0(X_1) + \mathrm{H}_0(X_2 \mid X_1) + ... + \mathrm{H}_0(X_n \mid X_1...X_{n-1}).$$

С помощью неравенства логарифма легко доказать, что $\mathrm{H}_0(Y|X) \le \mathrm{H}_0(Y)$ и равенство выполняется только в случае статистической независимости X и Y.

Относительная энтропия определяется ФПВ на ансамбле, и естественным является вопрос о том, для каких распределений она больше. Но вопрос не имеет смысла без дополнительных ограничивающих предположений.

Пусть Q - класс ФПВ на числовой оси и для каждой $f(x) \in Q$ выполняется условие

$$\int_{-\infty}^{\infty} x^2 f(x) dx \le c^2,$$

условие ограничения средней мощности СВ.

Теорема 16.1. Для любой ФПВ $f(x) \in Q$ выполняется неравенство

$$\mathrm{H}_0(X) \le \frac{1}{2} \log 2\pi e c^2, \tag{16.1}$$

где $\mathrm{H}_0(X)$ - относительная энтропия ансамбля $\{X, f(x)\}$. Равенство имеет место в случае, когда

$$f(x) = \frac{1}{c\sqrt{2\pi}} \exp\left[-\frac{x^2}{2c^2} \right],$$

распределение является гауссовским с нулевым средним и дисперсией c^2.

Пусть $f(x)$ - произвольная функция из Q, тогда

$$-\int_{-\infty}^{\infty} f(x) \log \frac{1}{c\sqrt{2\pi}} \exp\left[-\frac{x^2}{2c^2} \right] dx =$$
$$= \frac{1}{2} \log 2\pi c^2 + \frac{\log e}{2c^2} \int_{-\infty}^{\infty} x^2 f(x) dx \le \frac{1}{2} \log 2\pi e c^2, \tag{16.2}$$

так как интеграл второго слагаемого меньше c^2. Используя это вспомогательное неравенство, получим

$$\mathrm{H}_0(X)-\frac{1}{2}\log 2\pi ec^2 \le -\int_{-\infty}^{\infty} f(x)\log f(x)dx+\int_{-\infty}^{\infty} f(x)\log\frac{1}{c\sqrt{2\pi}}\exp\left[-\frac{x^2}{2c^2}\right]dx=$$

$$=\int_{-\infty}^{\infty} f(x)\log\frac{\exp\left[-\frac{x^2}{2c^2}\right]}{\mathrm{f}(x)c\sqrt{2\pi}}dx.$$

Применяя неравенство логарифма, получим:

$$\mathrm{H}_0(X)-\frac{1}{2}\log 2\pi ec^2 \le \log e\int_{-\infty}^{\infty} f(x)\{\frac{\exp[-\frac{x^2}{2c^2}]}{f(x)c\sqrt{2\pi}}-1\}dx=$$

$$=\log e\{\frac{1}{c\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp[-\frac{x^2}{2c^2}]dx-\int_{-\infty}^{\infty} f(x)dx\}=0,$$

так как каждое из выражений в фигурных скобках равно единице. Таким образом, получено неравенство (16.1). Равенство, учитывая (16.2), соблюдается только в случае, если $f(x)$ - гауссовская ФПВ. Условие равенства нулю математического ожидания несущественно, поэтому его можно опустить.

Рассмотрим относительную энтропию системы случайных величин X_1, ..., X_n с совместной ФПВ $f(x^{(1)}, ..., x^{(n)})$ и m_1, ..., m_n - математические ожидания этих величин. Обозначим через K_{ij} корреляционный момент СВ X_i и X_j.

$\mathrm{K}_{ij}=\mathrm{M}(X_i-m_i)(X_j-m_j)=\int(x^{(i)}-m_i)(x^{(j)}-m_j)f(x^{(1)},...,x^{(n)})dx^{(1)}...dx^{(n)}$. Матрица $\mathbf{K}=[\mathrm{K}_{ij}]i\ ;\ j=1, ..., n$ называется корреляционной матрицей системы СВ X_1, ..., X_n.

Относительная энтропия системы определяется соотношением

$$\mathrm{H}_0(X_1,...,X_n)=\int f(\boldsymbol{x})\log f(\boldsymbol{x})d\boldsymbol{x}.$$

Введем необходимые дополнительные определения. Обозначим $\mathbf{K}^{-1}$ матрицу, обратную $\mathbf{K}$, т.е. такую, что $\mathbf{K}^{-1}\mathbf{K}=\mathbf{I}_n$, где $\mathbf{I}_n$ - единичная матрица порядка n. Обратная матрица всегда существует, если $\det\mathbf{K}\neq 0$. Так

как $\mathrm{K}_{ij} = \mathrm{K}_{ji}$, то матрицы $\mathbf{K}$ и $\mathbf{K}^{-1}$ являются симметрическими. Из определения обратной матрицы и матричного умножения следует, что

$$\sum_{j=1}^{n} \mathrm{K}^{*}_{ij}\mathrm{K}_{ji'} = \begin{cases} 1 & \text{при } i = i', \\ 0 & \text{при } i \neq i', \end{cases}$$

где K^{*}_{ij} - элементы матрицы $\mathbf{K}^{-1}$, а отсюда и из симметричности $\mathbf{K}$ следует

$$\sum_{i,j=1}^{n} \mathrm{K}^{*}_{ij}\mathrm{K}_{ij} = n. \qquad (16.3)$$

Функция

$$f_G(x^{(1)},...,x^{(n)}) = (2\pi)^{-n/2}(\det \mathbf{K})^{-1/2} \exp[-\frac{1}{2}\sum_{i,j}(x^{(i)} - m_i)(x^{(j)} - m_j)\mathrm{K}^{*}_{ij}]$$

называется ФПВ n-мерного невырожденного гауссовского распределения вероятности, а СВ $X_1,...,X_n$ - совместно гауссовскими. В этом случае каждая из СВ системы также имеет гауссовское распределение вероятностей.

Теорема 16.2. Пусть Q_n - класс ФПВ n случайных величин с $m,...,m_n$ и заданной корреляционной матрицей $\mathbf{K}$, $\det \mathbf{K} \neq 0$. Для любой функции $f(\boldsymbol{x}) \in Q_n$ выполняется неравенство

$$\mathrm{H}_0(X_1,...,X_n) \leq \frac{n}{2}\log 2\pi e + \frac{1}{2}\log\det \mathbf{K},$$

равенство имеет место в том случае, когда $f(\boldsymbol{x})$ есть ФПВ n-мерного гауссовского распределения вероятностей, т.е. когда $f(\boldsymbol{x}) = f_{\mathrm{G}}(\boldsymbol{x})$.

Получим вначале вспомогательное равенство. Пусть $f(\boldsymbol{x})$ - произвольная функция из Q_n, тогда

$$F = -\int f(\boldsymbol{x})\log\{(2\pi)^{\frac{-\mathrm{n}}{2}}(\det \mathbf{K})^{-\frac{1}{2}}\exp[-\frac{1}{2}\sum_{ij}(x^{(i)} - m_i)(x^{(j)} - m_j)\mathrm{K}^{*}_{ij}]\}d\boldsymbol{x} =$$

$$= \frac{n}{2}\log 2\pi + \frac{1}{2}\log\det \mathbf{K} + \frac{\log e}{2}\int f(\boldsymbol{x})\sum_{i,j}(x^{(i)} - m_i)(x^{(i)} - m_j)\mathrm{K}^{*}_{ij}d\boldsymbol{x}.$$

Очевидно, что

$$\int f(\boldsymbol{x})\sum_{i,j}(x^{(i)} - m_i)(x^{(j)} - m_j)\mathrm{K}^{*}_{ij}d\boldsymbol{x} =$$

$$= \sum_{i,j}\mathrm{K}^{*}_{ij}\int f(\boldsymbol{x})(x^{(i)} - m_i)(x^{j} - m_j)d\boldsymbol{x} = \sum_{i,j}\mathrm{K}^{*}_{ij}\mathrm{K}_{ij} = n \qquad \text{(см. (16.3))},$$

откуда $F = \frac{n}{2}\log 2\pi e + \frac{1}{2}\log\det \mathbf{K}.$

Используя это равенство и неравенство для логарифма, получим:

$$\mathrm{H}_0(X_1,\ldots X_n) - \mathrm{F} = -\int \mathrm{f}(\boldsymbol{x})\,\log f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} + \int f(\boldsymbol{x})\,\log f_{\mathrm{G}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} =$$

$$= \int f(\boldsymbol{x})\,\log\frac{f_G(\boldsymbol{x})}{f(\boldsymbol{x})}d\boldsymbol{x} \le \int f(\boldsymbol{x})[\frac{f_G(\boldsymbol{x})}{f(\boldsymbol{x})} - 1]d\boldsymbol{x} = 0\ ,$$

что и доказывает теорему.

Как и в одномерном случае, относительная энтропия системы СВ не зависит от математического ожидания, поэтому упоминание о них в теореме 16.2 можно опустить.

Вопросы для самопроверки

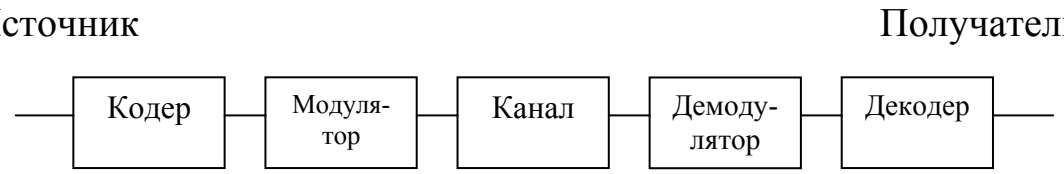
1. Что определяется с помощью понятия «взаимная информация»?
2. При каком условии сохраняется количество средней взаимной информации при функциональном отображении одного ансамбля в другой?
3. Какая сложность возникает в связи с оценкой количества информации при переходе от дискретных источников к непрерывным?
4. Сформулируйте понятие «относительная энтропия».
5. При каком распределении непрерывного сообщения относительная энтропия принимает максимальное значение?
6. Перечислите свойства относительной энтропии.

КОДИРОВАНИЕ В ДИСКРЕТНЫХ КАНАЛАХ

В любом канале передачи сообщений присутствует шум, что приводит к ненулевой вероятности появления ошибок. Основная задача, которую мы будем решать, - определение по статистической модели канала величины наибольшей скорости, при которой возможна передача сообщений с произвольно малой вероятностью ошибки.

17. Классификация каналов связи

В понятие “канал связи“ могут быть включены не только сам канал, но и все устройства, работающие в системе связи между “источником“ и “получателем“ (рисунок).



Структурная схема канала связи

Что касается источника и получателя, то будем полагать, что они, имея собственные кодер и декодер, выбраны достаточно хорошо, и что символы, появляющиеся на выходе кодера источника, независимы и равновероятны.

Устройство, сопоставляющее символу или группе символов на выходе кодера соответствующий входной сигнал канала, называется модулятором. Устройство, выполняющее обратное преобразование, называется демодулятором. В зависимости от построения пар “кодер - декодер“ и “модулятор - демодулятор“ зависит эффективность уменьшения влияния шумов. Будем предполагать, что “модулятор – демодулятор” и, следовательно, система используемых сигналов выбраны, а задача состоит в выборе пары “кодер – декодер”.

Определение. Канал называется дискретным по входу (выходу), если множество входных (выходных) сигналов конечно.

Иногда дискретным называется канал со счетным множеством сигналов.

Определение. Канал называется непрерывным по входу (выходу), если множество входных (выходных) сигналов несчетно.

Обычно множество входных сигналов обозначается X с элементами x, а выходных - Y с элементами y.

Определение. Канал называется каналом с дискретным временем, если сигналы на его входе и выходе представляют собой конечные или бесконечные последовательности с элементами из алфавитов X и Y соответственно. Дискретный по входу и выходу канал с дискретным временем будем называть дискретным каналом (ДК).

Определение. Канал называется каналом с непрерывным временем, если сигналы на его входе и выходе представляют собой действительные функции времени. Непрерывный по входу и выходу канал с непрерывным временем будем называть непрерывным каналом.

Будем говорить, что ДК задан, если для любых целых чисел n и j и любых последовательностей

$$(x^{(j)}, x^{(j+1)}, ..., x^{(n+j-1)}), \quad (y^{(j)}, y^{(j+1)}, ..., y^{(n+j-1)})$$

с элементами из дискретных множеств X и Y заданы условные (переходные) вероятности $p(y^{(j)}, y^{(j+1)}, ..., y^{(n+j-1)} \mid x^{(j)}, x^{(j+1)}, ..., x^{(n+j-1)})$.

Определение. ДК называется каналом без памяти, если для любых n и j, а также для любых последовательностей $(x^{(j)}, ..., x^{(n+j-1)})$ и $(y^{(j)}, ..., y^{(n+j-1)})$ имеют место равенства

$$p(y^{(j)}, ..., y^{(n+j-1)} \mid x^{(j)}, ..., x^{(n+j-1)}) = \prod_{i=j}^{n+j-1} p_i(y^{(i)} \mid x^{(i)}),$$

где $p_i(y \mid x)$ - вероятность для момента времени i получения на выходе канала сигнала y, если на входе был сигнал x .

Определение. ДК без памяти удовлетворяет условию стационарности, если для любых $i, j, x \in X, \ y \in Y$

$$p_j(y \mid x) = p_i(y \mid x).$$

Будем предполагать, что переходные вероятности удовлетворяют следующим условиям согласованности:

$$\sum_{Y^{n-k+1} \ldots Y^n} p(y^{(1)}, ..., y^{(n)} \mid x^{(1)}, ..., x^{(n)}) = p(y^{(1)}, ..., y^{(k)} \mid x^{(1)}, ..., x^{(k)}),$$

$$n = 1, 2, \ldots, \ k = 1, \ldots, n-1.$$

Каналы, удовлетворяющие условиям согласованности, называются каналами без предвосхищения, т.е. вероятность появления выходного сигнала в некоторый момент времени не зависит от сигналов, которые появятся на входе канала в последующие моменты времени.

Распределение вероятностей на входе канала не является его параметром, поскольку определяется входными устройствами, но не самим каналом.

18. Постановка задачи кодирования в дискретном канале

Определение. Кодом длиной n и объемом M для канала называется множество из M пар $\{\boldsymbol{u}_1, A_1; \boldsymbol{u}_2, A_2; \ldots; \boldsymbol{u}_M, A_M\}$, где $\boldsymbol{u}_i \in X^n$, $i = = 1, 2, ..., M$ - последовательности длины n, образованные входными сигналами канала и называемые кодовыми словами ($\boldsymbol{u}_i \neq \boldsymbol{u}_j$ при $i \neq j$) и $A_i \subseteq Y^n$, i = 1, 2, ...,

M - решающие области, образованные выходными последовательностями канала, причем при $i \neq j$ множества A_i и A_j не пересекаются.

Задание кода предполагает задание и правила, по которому приемник принимает решение о переданном кодовом слове (если на выходе канала $\boldsymbol{y} \in A_i$, то кодировалось сообщение $\boldsymbol{u}_i$).

Определение. Скоростью кода (скоростью передачи) называется величина

$$\mathrm{R} = \frac{1}{n}\log M,$$

максимальное количество информации, которое можно передать с помощью одного символа ($\log M$ - максимальное количество информации в одном кодовом слове).

Для источника n - длина отрезка кодируемых сообщений, для канала n - длина кодового слова. Код длины n со скоростью R и объемом $M = 2^{n\mathrm{R}}$ будем обозначать G(n, R). Обозначим λ_i - вероятность ошибки декодирования при условии передачи слова $\boldsymbol{u}_i$:

$$\lambda_i = \sum_{\mathrm{y} \in \overline{A_i}} p(y \mid \boldsymbol{u}_i),$$

где $\overline{A_i}$ - дополнение множества A_i до Y^n.

В качестве меры надежности передачи с помощью кода G(n, R) можно использовать две величины: максимальную вероятность ошибки $\Lambda = \max\{\lambda_1, ..., \lambda_{\mathrm{M}}\}$ и среднюю вероятность ошибки:

$$\lambda = \sum_{i=1}^{M} \lambda_i p(\boldsymbol{u}_i) \quad . \tag{18.1}$$

Так как $p(\boldsymbol{u}_i)$ характеризует источник, а не канал, то среднюю вероятность ошибки можно определить как

$$\lambda = \frac{1}{M} \sum_{i=1}^{M} \lambda_i. \tag{18.2}$$

Выражения (18.1) и (18.2) совпадают при оптимальном кодировании источника.

Определение. Пропускной способностью канала с дискретным временем называется такое максимальное число C, что для любого сколь угодно малого $\delta > 0$ и для любого $\mathrm{R} < \mathrm{C}$ существует такой код G(n, R), что максимальная вероятность ошибки удовлетворяет неравенству $\Lambda < \delta$.

Для определения числа C как пропускной способности канала необходимо доказать прямую и обратную теоремы кодирования.

Лемма. Пусть существует код объема M с вероятностью ошибки λ, определенной в (18.2), тогда существует код объема $M/2$, максимальная вероятность ошибки которого $\Lambda \le 2\lambda$.

Предположим, что кодовые слова $\boldsymbol{u}_1$, ..., $\boldsymbol{u}_M$ упорядочены по невозрастанию ошибки ($\lambda_i \ge \lambda_j$ при $i < j$). Тогда

$$\lambda = \frac{1}{M}\sum_{i=1}^{M}\lambda_i \ge \frac{1}{M}\sum_{i=1}^{M/2}\lambda_i \ge \frac{1}{2}\lambda_{M/2}$$

и для всех $j > M/2$, $\lambda_j \le \lambda_{M/2} \le 2\lambda$, т. е. код объема $M/2$, образованный словами $\boldsymbol{u}_{M/2+1},...,\boldsymbol{u}_M$, имеет максимальную вероятность ошибки, не превосходящую 2λ.

19. Неравенство Фано

С помощью неравенства Фано доказываются обратные теоремы кодирования для различных каналов.

Пусть задан дискретный ансамбль $\{UW,p(u,w)\}$, $U = \{u_1,u_2,...,u_M\}$, $W = \{w_1,w_2,...,w_L\}$. Событие E - появление пары (u_i,w_j), $i \ne j$ - будем называть “ошибкой“. Положим

$$\lambda_j = \Pr(\mathrm{E}\,|\,w_j) = \sum_{i:i\ne j} p(u_i\,|\,w_j) = 1 - p(u_j\,|\,w_j), \qquad (19.1)$$

условная вероятность ошибки при фиксированном $w_j \in W$;

$$\lambda = \Pr(\mathrm{E}) = \sum_{i,j:i\ne j} p(u_i w_j) = \sum_{j=1}^{L}\lambda_j p(w_j),$$

средняя вероятность ошибки.

На множестве $E = \{\mathrm{E},\overline{\mathrm{E}}\}$ для каждого $w_j \in W$ определено условное распределение вероятностей $\{\lambda_j, 1-\lambda_j\}$. Это распределение совместно с безусловным распределением $p(w_j)$ задает ансамбль EW, для которого

$$\mathrm{H}(E|w_j) = -\lambda_j \log\lambda j - (1-\lambda_j)\log(1-\lambda_j) = h(\lambda_j), \qquad (19.2)$$

где $\mathrm{H}(E|W) = \sum_{j=1}^{L}\mathrm{H}(E|w_j)\,p(w_j)$.

Безусловное распределение вероятностей на E есть $\{\lambda, 1-\lambda\}$, при этом $\mathrm{H}(\mathrm{E}) = h(\lambda)$.

Теорема (**Неравенство Фано**). Для любого дискретного ансамбля $\{UW, p(u, w)\}$, $|U| = M$, справедливо неравенство

$$\mathrm{H}(U|W) \le h(\lambda) + \lambda \log M. \tag{19.3}$$

Рассмотрим условную энтропию $\mathrm{H}(U|w_j)$. При $j \le M$ имеем

$$\mathrm{H}(U|w_j) = -\sum_{i=1}^{M} p(u_i \mid w_j)\log p(u_i \mid w_j) =$$

$$= -p(u_j|w_j)\log p(u_j|w_j) - [1-p(u_j|w_j)]\log[1-p(u_j|w_j)] - \tag{19.4}$$

$$-\sum_{i:i=j} p(u_i \mid w_j)\ \log p(u_i \mid w_j) +$$

$$+[1-p(u_j|w_j)]\log[1-p(u_j|w_j)] = \mathrm{H}(E|w_j) - \lambda_j \sum_{i\ne j} \frac{p(u_i|w_j)}{\lambda_j}\log\frac{p(u_i|w_j)}{\lambda_j},$$

где последнее равенство следует из (19.1) и (19.2) так же, как и

$$\sum_{i\ne j} \frac{p(u_i \mid w_j)}{\lambda_j} = 1, \tag{19.5}$$

поэтому второе слагаемое в (19.4) есть умноженная на λ_j энтропия ансамбля, состоящего из $M-1$ сообщений, вероятности которых указаны как слагаемые в сумме (19.5). Если эту энтропию оценить сверху величиной $\log M$, то

$$\mathrm{H}(U \mid w_j) \le \mathrm{H}(E \mid w_j) + \lambda_j \log M = h(\lambda_j) + \lambda_j \log M. \tag{19.6}$$

Если $L > M$, то при $j > M$

$$\mathrm{H}(U \mid w_j) = -\sum_{i=1}^{M} p(u_i \mid w_j)\log p(u_i \mid w_j) \le \log M, \tag{19.7}$$

и всегда происходит ошибка, поэтому $\lambda_j = 1$ и $\mathrm{H}(E \mid w_j) = h(\lambda_j) = 0$. Следовательно, из (19.7) вытекает, что (19.6) имеет место при всех $j = 1, ..., L$.

Усредним (19.6) по ансамблю W. В результате получим, что

$$\mathrm{H}(U \mid W) \le \mathrm{H}(E|W) + \lambda \log M,$$

но $\mathrm{H}(E|W) \le \mathrm{H}(E) = h(\lambda)$, поэтому неравенство (19.3) справедливо.

Если W - множество решений, U - множество кодовых слов, то $\mathrm{H}(U|W)$ характеризует количество информации, потерянное в канале из-за шума, - ненадежность передачи с помощью кода G(n, R). Неравенство Фано (19.3) устанавливает связь между ненадежностью передачи и средней вероятностью ошибки декодирования для кода G(n, R).

20. Общая обратная теорема кодирования для дискретного канала

Пусть задан канал (ансамбль) $\{X^n Y^n, p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})\}$ и $\mathbf{I}(X^n;Y^n)$ - средняя взаимная информация между последовательностями на входе и выходе канала

$$\mathbf{I}(X^n;Y^n)=\sum_{X^n}\sum_{Y^n} p(\boldsymbol{x})p(\boldsymbol{y}\mid\boldsymbol{x})\log\frac{p(\boldsymbol{y}|\boldsymbol{x})}{p(\boldsymbol{y})},$$

где $p(\boldsymbol{y})=\sum_{X^n} p(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x})$. Обозначим через C^* максимальное значение средней взаимной информации между входом и выходом канала:

$$\mathrm{C}^*=\sup_{n,p(x)}\frac{1}{n}\mathbf{I}(X^n;Y^n)$$

и будем называть его информационной емкостью ДК.

Теорема (обратная теорема кодирования). Пусть C^* - информационная емкость ДК и $\mathrm{R}=\mathrm{C}^*+\varepsilon$, где ε - произвольное положительное число. Тогда существует положительное число δ, зависящее от R, такое, что для всякого кода G(n,R) $\lambda\ge\delta$.

Зафиксируем некоторое n и рассмотрим код G(n, R) с $M=2^{nR}$ кодовыми словами $\{\boldsymbol{u}_1,...,\boldsymbol{u}_M\}$. Зададим распределение вероятности на X^n:

$$p(\boldsymbol{x})=\begin{cases}1/M & \text{для } \boldsymbol{x}\in\{\boldsymbol{u}_1,...,\boldsymbol{u}_{\mathrm{M}}\},\\ 0 & \boldsymbol{x}\notin(\boldsymbol{u}_1,...,\boldsymbol{u}_{\mathrm{M}}).\end{cases} \qquad (20.1)$$

Пусть $\mathbf{I}(X^n;Y^n)$ - средняя взаимная информация между входом и выходом канала. $\mathbf{I}(X^n;Y^n)=\mathbf{I}(U;Y^n)$, где U - ансамбль слов рассматриваемого кода . Из определения информационной емкости следует, что

$$n\mathrm{C}^*\ge\mathbf{I}(U;Y^n). \qquad (20.2)$$

Пусть W - отображение ансамбля Y^n сообщений на выходе канала в решения (ансамбль решений). Отображение задается посредством набора решающих областей $A_1,\dots,A_M$:

$$w=\begin{cases} w_i & \boldsymbol{y}\in A_i, i=1,\dots M, \\ w_{M+1} & \text{если} \quad \boldsymbol{y}\notin \bigcup A_i. \end{cases}$$

В результате преобразования информация не возрастает, поэтому

$$\mathbf{I}(U;Y^n)\geq \mathbf{I}(U;W).$$

Так как $\mathrm{H}(U\,|W) = \mathrm{H}(U) - \mathbf{I}(U;W)$ и $\mathrm{H}(U) = \log M$ (согласно (20.1)), то, используя (20.2), получим

$$\mathrm{H}(U\,|W) = \log M - \mathbf{I}(U;W)\geq \log M - n\mathrm{C}^*$$

или

$$\mathrm{H}(U\,|W)\geq n(\mathrm{R}-\mathrm{C}^*)=n\varepsilon. \qquad (20.3)$$

Воспользуемся неравенством Фано. Обозначим через λ_{0n} наименьший корень уравнения

$$h(\lambda)+\lambda\log M = n\varepsilon.$$

Тогда из неравенства Фано и (20.3) следует, что средняя вероятность ошибки λ для кода $\mathrm{G}(n, \mathrm{R})$ удовлетворяет неравенству $\lambda\geq\lambda_{0n}$. Из свойств функции $\varphi(\lambda)=h(\lambda)+\lambda\log M$ следует, что при $M\geq 1$ число $\lambda_{0n}>0$ при всех n и $\lambda_{0n}\geq\lambda_{01}>0$. Полагая $\lambda_{01}=\delta$, получим, что $\lambda\geq\delta$ для любого кода.

21. Информационная емкость дискретного канала без памяти

Теорема. Информационная емкость C^* ДК без памяти определяется соотношением

$$\mathrm{C}^* = \max_{\{p(x)\}} \mathbf{I}(X;Y),$$

где максимум разыскивается по всем распределениям $p(x)$ на X.

Пусть $p(\boldsymbol{x})$ - произвольное распределение вероятностей на входе канала. Средняя взаимная информация между входом и выходом канала

$$\mathbf{I}(X^n;Y^n)=\mathrm{H}(Y^n)-\mathrm{H}(Y^n\,|\,X^n)\;=$$

$$=\mathrm{H}(Y^n)+\sum_{X^n}\sum_{Y^n}p(\boldsymbol{x})\,p(\boldsymbol{y}\,|\,\boldsymbol{x})\log\prod_{i=1}^{n}p(y^{(i)}\,|\,x^{(i)})=$$

$$=\mathrm{H}(Y^n)+\sum_{i=1}^{n}\sum_{X^n}\sum_{Y^n}p(\boldsymbol{x})\,p(\boldsymbol{y}|\boldsymbol{x})\log p(y^{(i)}\,|\,x^{(i)})= \qquad (21.1)$$

$$= \mathrm{H}(Y^n) + \sum_{i=1}^{n} \sum_{X_i} \sum_{Y_i} p(x^{(i)}) p(y^{(i)} \mid x^{(i)}) \log p(y^{(i)} \mid x^{(i)}) =$$

$$= \mathrm{H}(Y^n) - \sum_{i=1}^{n} \mathrm{H}(Y_i \mid X_i) \le \sum_{i=1}^{n} \mathrm{H}(Y_i) - \sum_{i=1}^{n} \mathrm{H}(Y_i \mid X_i) = \sum_{i=1}^{n} \mathbf{I}(X_i; Y_i).$$

Знак равенства в неравенстве (21.1) справедлив в случае статистической независимости $Y_1, ..., Y_n$, т.е. когда $p(\boldsymbol{y}) = \prod_{i=1}^{n} p_i(y^{(i)})$ для всех $\boldsymbol{y} \in Y^n$. Условие выполняется, если выбрать $p(\boldsymbol{x}) = \prod_{i=1}^{n} p_i(x^{(i)})$.

Для произвольного входного распределения из (21.1) имеем

$$\mathbf{I}(X^n; Y^n) \le \sum_{i=1}^{n} \max_{\{p_i(x^{(i)})\}} \mathbf{I}(X_i; Y_i) = n \max_{\{p(x)\}} \mathbf{I}(X; Y), \qquad (21.2)$$

где последнее равенство выполняется в силу стационарности канала.

Итак, при статистической независимости входных сигналов и одинаковом их распределении в соотношении (21.2) достигается знак равенства, отсюда

$$\sup_{n\{p(x)\}} \frac{1}{n} \mathbf{I}(X^n; Y^n) = \max_{\{p(x)\}} \mathbf{I}(X; Y).$$

Знак sup заменен на max, так как множество всех распределений $\{p(x)\}$ на конечном множестве X замкнуто.

22. Симметричные дискретные каналы без памяти

Одномерные условные вероятности, которые описывают ДК без памяти, удобно записывать в виде $L \times N$ матрицы переходных вероятностей канала:

$$\begin{bmatrix} p(y_1|x_1) \ldots . p(y_N|x_1) \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ p(y_1|x_L) \ldots . p(y_N|x_L) \end{bmatrix}, \qquad (22.1)$$

в которой сумма элементов любой строки равна единице.

Определение. ДК называется симметричным по входу, если все строки матрицы переходных вероятностей образованы перестановками элементов первой строки.

Определение. ДК называется симметричным по выходу, если все столбцы матрицы переходных вероятностей образованы перестановками элементов первого столбца.

Определение. ДК называется симметричным, если он симметричен по входу и выходу - и строки, и столбцы матрицы переходных вероятностей образованы перестановками одного и того же набора чисел.

Свойство 1. Если канал симметричен по входу, то условная энтропия H(Y | X) не зависит от распределения вероятностей на входе и равна

$$\mathrm{H}(Y \mid X) = -\sum_{j=1}^{N} p_j \log p_j ,$$

где p_j - элементы первой строки матрицы (22.1).

$$\mathrm{H}(Y \mid X) = \mathrm{MH}(Y \mid x) = -\sum_{i=1}^{L} p(x_i) \sum_{j=1}^{N} p_j \log p_j = -\sum_{j=1}^{N} p_j \log p_j \ ,$$

так как $\sum_{j=1}^{N} p_j \log p_j$ не зависит от $\underline{x}$, а $\sum_{i=1}^{L} p(x_i) = 1$.

Свойство 2. Если канал симметричен по выходу и распределение вероятностей на его входных сигналах равномерное, то равномерным является и распределение вероятностей на его выходных сигналах.

Действительно, пусть $p(x_i) = 1/L$, тогда

$$p(y_j) = \sum_{i=0}^{L} p(x_i) p(y_j \mid x_i) = \frac{1}{L} \sum_{i=1}^{L} q_i , \qquad 22.2)$$

где $q_1, \ldots, q_L$ - элементы первого столбца матрицы (22.1). Правая часть (22.2) не зависит от *j,* поэтому выходные сигналы канала равновероятны.

Найдем информационную емкость C^* симметричного ДК без памяти:

$$\mathrm{C}^* = \max_{\{p(x)\}} \mathbf{I}(X;Y) = \max_{\{p(x)\}} \left[\mathrm{H}(Y) - \mathrm{H}(Y \mid X)\right].$$

$\mathrm{H}(Y \mid X)$ не зависит от распределения на входе (свойство 1), поэтому нужно максимизировать H(*Y*), а для дискретных ансамблей энтропия максимальна для равномерных распределений. По свойству 2 входные сигналы также должны иметь равномерное распределение, тогда

$$\mathrm{H}(Y) = \log N, \ \mathrm{C}^* = \log N + \sum_{j=1}^{N} p_j \log p_j .$$

Если канал не симметричен по выходу, то может не существовать распределения на входе, при котором выходные сигналы равновероятны, тогда

$$\mathrm{C}^* \le \log N + \sum_{j=1}^{N} p_j \log p_j .$$

23. Стационарные каналы с аддитивным по модулю *L*-шумом

Суммой $x + y$ по модулю L называется такое число z, $0 \le z \le L-1$, что $x+y-z$ нацело делится на L. Это записывается как $z = (x+y) \bmod L$.

Сумма двух последовательностей $(\boldsymbol{x}+\boldsymbol{y}) \bmod L$ определяется как их покомпонентная сумма по модулю L. Если $z = (x+y) \bmod L$, то $z^{(i)} = (x^{(i)} + y^{(i)}) \bmod L$.

Определение. Стационарным ДК с аддитивным по модулю L шумом (AL-каналом) называется канал, переходные вероятности которого определяются соотношением $p(y|x) = q(z)$, где z - корень уравнения $z + \boldsymbol{x} = \boldsymbol{y} \bmod L$ и $q(z)$ - распределение, задаваемое для каждого n, n= 1, 2, ... стационарным источником U_Z, называемого в этом случае источником шума. При этом, как множество X на входе, так и множество Y на выходе канала представляют собой L чисел $\{0,1,..., L-1\}$.

Матрица, составленная из переходных вероятностей AL-канала, удовлетворяет определению симметричного канала, следовательно,

$$\max_{\{p(x)\}} \mathbf{I}(X^n; Y^n) = n \log L + \sum_{Z^n} q(\mathrm{z}) \log q(\mathrm{z}) = n \log L - \mathrm{H}(Z^n),$$

$$\max \frac{1}{n} \mathbf{I}(X^n; Y^n) = \log L - \mathrm{H}_n(Z), \qquad (23.1)$$

где $\mathrm{H}_n(Z) = \frac{1}{n}\mathrm{H}(Z^n)$.

Из свойств стационарного источника (см. п. 5) известно, что $\mathrm{H}_{n+1}(Z) \le \mathrm{H}_n(Z)$ и $\lim\limits_{n\to\infty} \mathrm{H}_n(Z) = \mathrm{H}(Z \mid Z^\infty)$, где $\mathrm{H}(Z|Z^\infty)$ - энтропия на сообщение стационарного источника U_Z. Поэтому правая часть (23.1) не убывает с ростом n, и информационная емкость AL-канала

$$\mathrm{C}^* = \sup_{n,\{p(x)\}} \frac{1}{n} \mathbf{I}(X^n, Y^n) = \lim_{n\to\infty} (\log L - \mathrm{H}_n(Z)) = \log L - \mathrm{H}(Z \mid Z^\infty). \qquad (23.2)$$

В предположении эргодичности источника шума в AL-канале докажем обратную теорему кодирования.

Теорема. Пусть C^* - информационная емкость AL-канала с эргодическим шумом. Тогда для любого $R > C^*$ и для любой последовательности кодов $\{G(n,R)\}$, $n = 1, 2, \ldots$ $\lim\limits_{n\to\infty} \Lambda_n = 1$, где Λ_n - максимальная вероятность ошибки для кода $G(n,R)$.

Положим $R = C^* + \varepsilon$ $(\varepsilon > 0)$ и рассмотрим произвольный код $G(n, R) = \{\boldsymbol{u}_1, A_1; \ldots; \boldsymbol{u}_M A_M\}, M = 2^{nR}$. Предположим, что решающая область A_1 имеет наименьший объем. Тогда

$$|A_1| \le \frac{2^{n \log L}}{2^{nR}} = 2^{n(\log L - R)}, \qquad (23.3)$$

где числитель - общее число выходных последовательностей канала, а знаменатель - количество решающих областей. Из (23.2) и (23.3) следует, что

$$|A_1| \le 2^{n[H(Z|Z^\infty)-\varepsilon]}. \qquad (23.4)$$

Определим множество B_1 следующим образом:

$$B_1 = \{\boldsymbol{z} : (\boldsymbol{u}_1 + \boldsymbol{z}) \bmod L \in A_1\},$$

т.е. правильное декодирование происходит при передаче слова $\boldsymbol{u}_1$ и шумовой последовательности, принадлежащей области B_1. Пусть λ_{1n} - вероятность ошибки при передаче $\boldsymbol{u}_1$, тогда

$$\Lambda \ge \lambda_{1n} = \Pr(\boldsymbol{z} \in \overline{B_1} \mid \boldsymbol{u}_1) = 1 - \Pr(\boldsymbol{z} \in B_1),$$

причем последнее равенство следует из независимости источника шума и передаваемых сообщений.

Так как $|B_1| = |A_1|$ и не превосходит правой части неравенства (23.4), то из обратной теоремы для равномерного кодирования эргодического источника (теорема 9.4) следует, что для любого множества B_1 с таким числом элементов вероятность события $\{\boldsymbol{z} \in B_1\}$ стремится к нулю при стремлении n к бесконечности. Это означает, что λ_{1n} стремится к единице. Теорема доказана.

В основе доказанной теоремы лежит сопоставление объемов решающей области и высоковероятного множества эргодического источника. Чтобы вероятность ошибки могла быть сделана сколь угодно малой, необходимо (но не достаточно), чтобы каждая решающая область имела объем не меньший, чем объем высоковероятного множества.

24. Неравенство Файнстейна

Неравенство Файнстейна − одно из важнейших теоретико-информационных неравенств, с помощью которого можно доказать прямые теоремы кодирования для различных каналов связи.

Рассмотрим ДК с переходными вероятностями $p(\boldsymbol{y}|\boldsymbol{x})$, $\boldsymbol{x} \in X^n, \boldsymbol{y} \in Y^n$. I($\boldsymbol{x}$; $\boldsymbol{y}$) - информация между $\boldsymbol{x}$ и $\boldsymbol{y}$, вычисляемая как

$$I(\boldsymbol{x};\boldsymbol{y}) = \log \frac{p(\boldsymbol{y}|\boldsymbol{x})}{p(\boldsymbol{y})},$$

где $p(\boldsymbol{y}) = \sum_{X^n} p(\boldsymbol{y}|\boldsymbol{x}) p(\boldsymbol{x})$.

Обозначим через S некоторое подмножество множества X^n, а через V_τ - множество таких пар ($\boldsymbol{x}$; $\boldsymbol{y}$), что

$$V_\tau = \{(\boldsymbol{x},\boldsymbol{y}) : \mathrm{I}(\boldsymbol{x};\boldsymbol{y}) > n\tau\},$$

где τ - некоторое положительное число.

Теорема (**неравенство Файнстейна**). Пусть τ - произвольное положительное число, S - произвольное подмножество множества X^n и $p(\boldsymbol{x})$ - произвольное распределение вероятностей на X^n, тогда для каждого n, (n = 1, 2, ...) существует код G(n, R), каждое слово которого принадлежит S, а максимальная вероятность Λ_n ошибки декодирования удовлетворяет неравенству

$$\Lambda_n \le \frac{1}{\Pr(S)} [2^{-n(\tau - R)} + 1 - \Pr(V_\tau)], \qquad (24.1)$$

где $\Pr(S) = \sum_{\mathrm{x} \in S} p(\boldsymbol{x})$ и $\Pr(V_\tau) = \sum_{(\boldsymbol{x},\boldsymbol{y}) \in V_\tau} p(\boldsymbol{y}|\boldsymbol{x}) p(\boldsymbol{x})$. Доказательство опускаем.

Для доказательства прямой теоремы кодирования с использованием (24.1) необходимо установить: можно ли подобрать τ и входное распределение так, чтобы оба слагаемых в правой части неравенства убывали к нулю при возрастании n. Эта задача сводится к исследованию поведения случайной величины $\frac{1}{n}\mathrm{I}(\boldsymbol{x};\boldsymbol{y})$.

Предположим, что S - X^n, тогда $\Pr(S) = 1$. Положим R= $\mathrm{C}^* - \varepsilon$, $\varepsilon > 0$ и

$$\tau = \mathrm{C}^* - \frac{\varepsilon}{2}, \qquad (24.2)$$

где C^* - информационная емкость дискретного по времени канала, определяемая следующим соотношением:

$$C^* = \sup \frac{1}{n} \mathbf{I}(X^n; Y^n).$$

Тогда, как это следует из (24.1), первое слагаемое равно $2^{-n\varepsilon/2}$ и стремится к нулю при возрастании n, второе слагаемое определяется поведением величины $\Pr(\overline{V_\tau})$, которую можно представить следующим образом:

$$\Pr(\overline{V_\tau}) = 1 - \Pr(V_\tau) = \Pr(\frac{1}{n} \mathrm{I}(\boldsymbol{x};\boldsymbol{y}) \le C^* - \frac{\varepsilon}{2}) =$$

$$= \Pr\{\frac{1}{n} \mathrm{I}(\boldsymbol{x};\boldsymbol{y}) \le \frac{1}{n} \mathbf{I}(X^n; Y^n) - \frac{\varepsilon}{2} + (C^* - \frac{1}{n} \mathbf{I}(X^n; Y^n))\}.$$

Предположим, что при каждом n выбирается такое входное распределение $p(\boldsymbol{x})$, при котором достигается максимум средней взаимной информации. Тогда найдется такое n, что

$$C^* - \frac{1}{n} \mathbf{I}(X^n; Y^n) \le \frac{\varepsilon}{4} \qquad (24.3)$$

и

$$\Pr(\overline{V_\tau}) \le \Pr\{\frac{1}{n} \mathrm{I}(\boldsymbol{x};\boldsymbol{y}) \le \frac{1}{n} \mathbf{I}(X^n; Y^n) - \frac{\varepsilon}{4}\}. \qquad (24.4)$$

Для моделей многих каналов связи верхняя грань C^* достигается при $n \to \infty$, поэтому (24.3) выполняется для всех достаточно больших n, и вопрос о поведении вероятности ошибки сводится к исследованию правой части неравенства (24.4), т. е. величины $\frac{1}{n} \mathrm{I}(\boldsymbol{x};\mathbf{y})$.

25. Прямая теорема кодирования для канала без памяти

Для произвольного ДК без памяти информационная емкость

$$C^* = \max_{\{p(x)\}} \mathbf{I}(X;Y), \qquad (25.1)$$

и при любом распределении вероятностей на входе

$$\frac{1}{n} \mathrm{I}(\boldsymbol{x};\boldsymbol{y}) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}(x^{(i)}; y^{(i)}), \quad \boldsymbol{x} \in X^n, \boldsymbol{y} \in Y^n, \qquad (25.2)$$

где

$$\mathrm{I}(x^{(i)}; y^{(i)}) = \log \frac{p(y^{(i)} \mid x^{(i)})}{p(y^{(i)})}, \quad i = 1,2,...,n,$$

независимые одинаково распределенные случайные величины с ограниченной дисперсией.

Теорема. Пусть C^* - информационная емкость ДК без памяти. При любом $R < C^*$ и любом положительном δ существует код $G(n, R)$, максимальная вероятность ошибки которого удовлетворяет неравенству $\Lambda_n \le \delta$.

Пусть $\varepsilon = C^* - R$. Выберем τ в соответствии с (24.2), а распределение $p(\boldsymbol{x})$ таким, которое максимизирует среднюю взаимную информацию в (25.1). По теореме п. 24 при $S = X^n$ существует код $G(n, R)$, максимальная вероятность ошибки декодирования которого удовлетворяет неравенству

$$\Lambda_n \le 2^{-n(\tau - R)} + \Pr(\overline{V_\tau}) = 2^{-n\varepsilon/2} + \Pr(\overline{V_\tau}), \qquad (25.3)$$

где $\Pr(\overline{V_\tau})$ удовлетворяет неравенству (24.4).

Поскольку справедливо (25.2), то $\mathbf{I}(X^n; Y^n) = n\mathbf{I}(X; Y)$ и

$$\Pr(\overline{V_\tau}) \le \Pr\{\frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(x^{(i)}; y^{(i)}) - \mathbf{I}(X; Y) \le -\frac{\varepsilon}{4}\} \le$$

$$\le \Pr\{|\frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(x^{(i)}; y^{(i)}) - \mathbf{I}(X; Y)| \ge \frac{\varepsilon}{4}\}. \qquad (25.4)$$

В силу ранее перечисленных свойств величины $\mathrm{I}(x^{(i)}; y^{(i)})$ и того, что $\mathrm{MI}(x^{(i)}; y^{(i)}) = \mathbf{I}(X; Y),$ по закону больших чисел правая часть (25.4) стремится к нулю при увеличении n. Таким образом, оба слагаемых в (25.3) стремятся к нулю. Это доказывает существование кода $G(n, R)$ и такого n, при котором максимальная вероятность ошибки меньше любого заданного наперед δ. Теорема доказана.

Этот результат совместно с обратной теоремой кодирования позволяет сформулировать следующее утверждение.

Пусть C - пропускная способность ДК без памяти, т. е. такое число, что для каждого $R < C$ существует код $G(n, R)$ с максимальной вероятностью ошибки, меньшей наперед заданного произвольного положительного числа, и что для любого $R > C$ не существует кода с таким свойством. Пусть C^*- информационная емкость, определяемая выражением (25.1). Тогда $C = C^*$.

26. Прямая теорема кодирования для стационарных каналов с аддитивным эргодическим шумом

Теорема. Пусть C^* − информационная емкость стационарного ДК с аддитивным по модулю L-шумом. Пусть, кроме того, источник шума в таком канале - эргодический. Тогда при любом $R < C^*$ и любом положительном δ существует код $G(n, R)$, максимальная вероятность ошибки которого удовлетворяет неравенству $\Lambda \le \delta$.

Пусть $\varepsilon = C^* - R$. Выберем τ в соответствии с (24.2). Равномерное распределение $p(\boldsymbol{x})$ максимизирует среднюю взаимную информацию на сообщение. По теореме п. 24 при $S = X^n$ существует код, максимальная вероятность ошибки которого в рассматриваемом канале удовлетворяет неравенству

$$\Lambda_n < 2^{-n(\tau - R)} + \Pr(\overline{V_\tau}) = 2^{-n\varepsilon/2} + \Pr(\overline{V_\tau}), \qquad (26.1)$$

где

$$\Pr(\overline{V_\tau}) = \Pr(\mathrm{I}(\boldsymbol{x};\boldsymbol{y}) \le n\tau) = \Pr\left(\frac{1}{n}\mathrm{I}(\boldsymbol{x};\boldsymbol{y}) \le C^* - \frac{\varepsilon}{2}\right) =$$

$$= \Pr\left(\frac{1}{n}\log q(\boldsymbol{z}) \le -\mathrm{H}(Z \mid Z^\infty) - \frac{\varepsilon}{2}\right) \le \Pr\left(\left|\frac{1}{n}\mathrm{I}(\boldsymbol{z}) - \mathrm{H}(Z \mid Z^\infty)\right| \ge \frac{\varepsilon}{2}\right).$$

Из леммы Мак-Миллана следует, что, выбирая n достаточно большим, можно сделать $\Pr(\overline{V_\tau})$ меньше, чем δ. Вместе с (26.1) это завершает доказательство теоремы.

Полученный результат совместно с обратной теоремой кодирования позволяет сформулировать следующее утверждение: пусть C - пропускная способность AL-канала с эргодическим шумом и пусть C^* - его информационная емкость, тогда $C = C^*$.

27. Декодирование для кодов с заданным множеством кодовых слов

Рассмотрим некоторый код $\{X^n Y^n, p(\boldsymbol{y}|\boldsymbol{x})\}$ и обозначим через $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_M$, $\boldsymbol{u}_i \in X^n$ его кодовые слова.

Предположим, что набор кодовых слов фиксирован и требуется указать наилучший выбор решающих областей $A_1, \ldots, A_M \subseteq Y^n$, при котором средняя или максимальная вероятность ошибки минимизируется.

Пусть $\{p(\boldsymbol{u}_1), \ldots, p(\boldsymbol{u}_M)\}$ - распределение вероятностей на множестве кодовых слов. Тогда апостериорные вероятности кодовых слов

$$p(\boldsymbol{u}_i \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{u}_i)\,p(\boldsymbol{u}_i)}{p(\boldsymbol{y})},$$

где $p(\boldsymbol{y}) = \sum_{i=1}^{M} p(\boldsymbol{y}|\boldsymbol{u}_i)\,p(\boldsymbol{u}_i)$ - безусловное распределение вероятностей на выходе канала. Обозначим $\lambda(\boldsymbol{y})$ - вероятность ошибки при условии, что на выходе канала последовательность $\boldsymbol{y}$. Тогда средняя вероятность ошибки будет равна

$$\lambda = \sum_{Y^n} \lambda(\boldsymbol{y})\,p(\boldsymbol{y}) = \sum_{i=1}^{M} \sum_{A_i} \lambda(\boldsymbol{y})\,p(\boldsymbol{y}). \qquad (27.1)$$

Так как $\lambda(\boldsymbol{y}) = 1 - p(\boldsymbol{u}_i \mid \boldsymbol{y}),\ \boldsymbol{y} \in A_i,$ то совместно с (27.1) это дает

$$\lambda = 1 - \sum_{i=1}^{M} \sum_{A_i} p(\boldsymbol{y})\,p(\boldsymbol{u}_i \mid \boldsymbol{y}).$$

Таким образом, при данном наборе кодовых слов средняя вероятность ошибки минимальна, если

$$A_i = \{\boldsymbol{y} : p(\boldsymbol{u}_i \mid \boldsymbol{y}) \geq p(\boldsymbol{u}_j \mid \boldsymbol{y}) \quad \text{для всех } j = 1, 2, ..., \mathrm{M}\}. \qquad (27.2)$$

Правило декодирования, определяемое разбиением на решающие области (27.2), выбирает при каждом $y \in Y^n$ то кодовое слово, которое имеет наибольшую апостериорную вероятность $p(\boldsymbol{u}_i \mid \boldsymbol{y})$. Это правило называется декодированием по максимуму апостериорной вероятности (МАВ-декодирование). Оно минимизирует среднюю вероятность ошибки.

МАВ-декодирование зависит от апостериорного распределения вероятностей на множестве кодовых слов. Декодирование по максимуму правдоподобия (МП-декодирование) по построению не зависит от апостериорного распределения . При МП-декодировании

$$A_i = \{\boldsymbol{y} : p(\boldsymbol{y}|\boldsymbol{u}_i) > p(\boldsymbol{y}|\boldsymbol{u}_j) \quad \text{для всех } j = 1, 2, ...,M\}. \qquad (27.3)$$

Если все кодовые слова равновероятны, то разбиения на решающие области (2.72) и (27.3) совпадают.

Согласно определению

$$\mathrm{I}(\boldsymbol{x}; \boldsymbol{y}) = \log \frac{p(\boldsymbol{y}|\boldsymbol{x})}{p(\boldsymbol{y})},$$

поэтому МП-декодирование эквивалентно выбору такого кодового слова, которое при данной выходной последовательности $\boldsymbol{y}$ максимизирует величину взаимной информации между ним и $\boldsymbol{y}$.

Вопросы для самопроверки

1. Перечислите типы дискретных каналов связи.

2. В чем различие при формулировании понятий «скорости кодирования» для дискретных источников «скорости кода» для дискретных каналов?

3. Какие типы ошибок применяются для оценки качества передачи сообщений?

4. Дайте определение пропускной способности канала.

5. Что устанавливает неравенство Фано?

6. Как определяется информационная емкость дискретного канала?

7. Сформулируйте свойства симметричных дискретных каналов без памяти.

8. Что такое AL-канал?

9. Взаимосвязь каких параметров кодирования в каналах устанавливает неравенство Файнстейна?

10. Сформулируйте прямую теорему кодирования для стационарных дискретных каналов с аддитивным эргодическим шумом.

11. При каком условии минимизируется средняя вероятность ошибки декодирования сообщений с заданным множеством кодовых слов?

КОДИРОВАНИЕ В НЕПРЕРЫВНЫХ КАНАЛАХ

Все результаты, полученные ранее для ДК, легко переносятся на непрерывные каналы (НК) с дискретным временем. Такое перенесение достигается заменой переходных вероятностей, задающих ДК, на условные функции плотности вероятностей, задающие НК с дискретным временем, а также сумм на интегралы. Специфическим вопросам кодирования в НК с дискретным временем, связанным с учетом ограничения мощности передатчика, по существу и посвящены пункты этого раздела.

28. Канал с дискретным временем. Обратная теорема кодирования

НК с дискретным временем считается заданным, если для любого $n=1, 2, \ldots$ и любых последовательностей $\boldsymbol{x} = \in X^n$, $\boldsymbol{y} = \left(y^{(1)}, \ldots, y^{(n)} \right) \in Y^n$, где X и Y - числовые оси, заданы n-мерные функции плотности вероятности $f(\boldsymbol{y} \mid \boldsymbol{x})$, описывающие передачу последовательностей длины n в таком канале.

НК без памяти обладает свойством $f(\boldsymbol{y}\,|\,\boldsymbol{x}) = \prod_{i=1}^{n} f_i\left(y^{(i)}\,|\,x^{(i)}\right)$, а в соответствующем стационарном канале - $f_i(y\,|\,x) = f_j(y\,|\,x)$.

Основное различие ДК и НК заключается в том, что средняя взаимная информация в НК может быть сделана какой угодно большой, в то время как в ДК максимум средней взаимной информации на сообщение по всем распределениям вероятностей на входе давал пропускную способность канала. Величина средней взаимной информации в НК с аддитивным шумом прямо связана с отношением сигнал/шум, вот почему введение ограничения (физически обоснованного) на мощность передатчика дает возможность найти ее реальное значение.

Определение. Пусть $\boldsymbol{u}_1,...,\boldsymbol{u}_m$ - последовательности длины n (кодовые слова), образованные входными сигналами канала, и $A_1,...,A_m$ - непересекающиеся подмножества (решающие области), образованные выходными сигналами канала. Кодом для НК с непрерывным временем, удовлетворяющим ограничению P на среднюю мощность, будем называть множество пар $\{\boldsymbol{u}_1, A_1; ...; \boldsymbol{u}_m, A_m\}$ таких, что

$$\frac{1}{n}\sum_{j=1}^{n}\left(u_i^{(j)}\right)^2 \le P \qquad (28.1)$$

где $i = 1, ..., m$ для каждого кодового слова $\boldsymbol{u}_i$. Число $R = \log M/n$ называется скоростью, а число n - длиной кода. Вероятность ошибки при декодировании $\boldsymbol{u}_i$

$$\lambda_i = 1 - \int_{A_i} f(\boldsymbol{y}\,|\,\boldsymbol{u})\,d\boldsymbol{y}.$$

Определение. Пропускной способностью НК с дискретным временем при ограничении P на среднюю мощность входных сигналов называется максимальное число C такое, что для любого сколь угодно малого положительного δ и любого $R < C$ существует код G(n,R), все слова которого удовлетворяют ограничению (28.1), а максимальная вероятность ошибки удовлетворяет неравенству $\Lambda \le \delta$.

Обозначим через $\Phi_n(P)$ множество всех ФПВ на X^n таких, что

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{M}\, X_i^2 \le P, \qquad (28.2)$$

где $\mathrm{M}\, X_i^2 = \int\limits_{X^n} \boldsymbol{x}^2 f(\boldsymbol{x})\, d\boldsymbol{x} = \int\limits_{X} \left(x^{(i)}\right)^2 f_i\left(x^{(i)}\right) dx^{(i)}$,

и $f_i\left(x^{(i)}\right) = \int \ldots \int f(\boldsymbol{x})\, dx^{(1)} \ldots dx^{(i-1)} dx^{(i+1)} \ldots dx^{(n)}$.

Определение. Информационной емкостью НК с дискретным временем при ограничении P на среднюю мощность входных сигналов называется число C*, определяемое следующим соотношением:

$$\mathrm{C}^* = \sup_{n, \Phi_n(P)} \frac{1}{n} \mathbf{I}\left(X^n; Y^n\right). \tag{28.3}$$

Теорема (обратная теорема кодирования). Пусть C*- информационная емкость указанного выше канала. Пусть ε - произвольное положительное число и $\mathrm{R} = \mathrm{C}^* + \varepsilon$. Тогда найдется такое положительное число δ, зависящее от R, что для всякого кода G(n, R), удовлетворяющего ограничению P на среднюю мощность, средняя вероятность $\lambda \geq \delta$.

Зафиксируем n и рассмотрим некоторый код G(n, R) при $\mathrm{R} = \mathrm{C}^* + \varepsilon$, все слова которого удовлетворяют условию (28.1). Обозначим $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ слова этого кода и получим:

$$f(\boldsymbol{x}) = \frac{1}{M} \sum_{i=1}^{M} \delta(\boldsymbol{x} - \boldsymbol{u}_i), \tag{28.4}$$

где $\delta(x)$ − дельта-функция Дирака.

Функция $f(\boldsymbol{x})$ - обобщенная ФПВ, приписывающая одинаковые вероятности $1/M$ всем кодовым словам. Функция $f(\boldsymbol{x})$ принадлежит $\Phi_n(P)$, так как неравенство (28.1) можно представить в следующей векторной форме:

$$\frac{1}{n}\left(\boldsymbol{u}_i \boldsymbol{u}_i^T\right) \leq P,$$

где T - символ транспонирования вектора $\boldsymbol{u}_i$.

Неравенство (28.2) представим следующим образом:

$$\frac{1}{n} \int\limits_{X^n} \left(\boldsymbol{x}\boldsymbol{x}^T\right) f(\boldsymbol{x})\, d\mathrm{x} = \frac{1}{n} \frac{1}{M} \sum_{i=1}^{M} \int\limits_{X^n} \left(\boldsymbol{x}\boldsymbol{x}^T\right) \delta(\boldsymbol{x} - \boldsymbol{u}_i)\, d\mathrm{x} = \frac{1}{n} \frac{1}{M} \sum_{i=1}^{M} \left(\boldsymbol{u}_i \boldsymbol{u}_i^T\right) \leq P,$$

т. е. для ФПВ (28.4) неравенство (28.2) выполняется.

Из определения информационной емкости следует, что для ФПВ (28.4) выполняются неравенства:

$$nC^* \geq \mathbf{I}\left(X^n; Y^n\right) = \mathbf{I}\left(U; Y^n\right) \geq \mathbf{I}\left(U; W\right),$$

где U - ансамбль кодовых слов с равномерным распределением вероятностей, W - ансамбль решений. Последнее неравенство - следствие невозрастания средней взаимной информации при преобразованиях. Доказательство теоремы завершается применением неравенства Фано и рассуждений, приведенных при доказательстве обратной теоремы кодирования для ДК.

29. Канал без памяти с дискретным временем

Общим результатом, который позволяет получить прямые теоремы кодирования не только в случае каналов без памяти, является неравенство Файнстейна. Приведем его формулировку для НК с дискретным временем.

Теорема 29.1 (**неравенство Файнстейна**). Пусть задан произвольный НК с дискретным временем. Пусть τ - произвольное положительное число, S − произвольное подмножество множества X^n входных сигналов, $f(\boldsymbol{x})$ − произвольная ФПВ на X^n. Тогда для любых значений n и R существует код G(n, R), каждое слово которого принадлежит S, а максимальная вероятность ошибки удовлетворяет неравенству

$$\Lambda_n \leq \frac{1}{\Pr(S)}\left[2^{-n(\tau - R)} + \Pr\left(\bar{V}_\tau\right)\right],$$

$$V_\tau = \left\{(\boldsymbol{x}, \boldsymbol{y}) : \mathrm{I}(\boldsymbol{x}; \boldsymbol{y}) = \log\frac{f(\boldsymbol{y} \mid \boldsymbol{x})}{f(\boldsymbol{y})} > n\tau\right\},$$

где

$$\Pr\left(\bar{V}_\tau\right) = 1 - \int_{V_\tau} f(\boldsymbol{x}) f(\boldsymbol{y} \mid \boldsymbol{x})\, d\boldsymbol{x} d\boldsymbol{y}, \quad \Pr(S) = \int_S f(\boldsymbol{x})\, d\boldsymbol{x}.$$

В случае НК без памяти, как и для ДК без памяти, информационную емкость канала можно искать не по (28.3), а полагая $n = 1$.

Зададим НК без памяти, т.е. определим условные ФПВ

$$f(\boldsymbol{y} \mid \boldsymbol{x}) = \prod_{i=1}^{n} f\left(y^{(i)} \mid x^{(i)}\right) \tag{29.1}$$

и множество ФПВ на X:

$$\Phi(P) = \left\{f(x) : \int_{-\infty}^{\infty} x^2 dx \leq P\right\}. \tag{29.2}$$

Теорема 29.2. Информационная емкость НК без памяти с ограничением P на среднюю мощность входных сигналов определяется соотношением

$$C^* = \max_{\Phi(P)} \mathbf{I}(X;Y).$$

Пусть $f(x)$ принадлежит $\Phi_n(P)$ и удовлетворяет ограничению (28.1). Тогда из (29.1), как и в случае с ДК, следует:

$$\mathbf{I}\left(X^n;Y^n\right) \le \sum_{i=1}^{n} \mathbf{I}_i(X;Y), \tag{29.3}$$

где $$\mathbf{I}_i(X;Y) = \int_X \int_Y f_i\left(x^{(i)}\right) f\left(y^{(i)} \mid x^{(i)}\right) \log \frac{f\left(y^{(i)} \mid x^{(i)}\right)}{f_i\left(y^{(i)}\right)} dx^{(i)} dy^{(i)}.$$

Равенство в (29.3) выполняется при статистической независимости выходных сигналов. Это достигается статистической независимостью сигналов на входе, т.е.

$$f(\boldsymbol{x}) = \prod_{i=1}^{n} f_i\left(x^{(i)}\right).$$

Средняя взаимная информация является выпуклой вверх функцией относительно входных распределений. Поэтому

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{I}_i(X;Y) \le \mathbf{I}_0(X;Y), \tag{29.4}$$

где $$\mathbf{I}_0(X;Y) = \int_X \int_Y f_0(x) f(y \mid x) \log \frac{f(y \mid x)}{f_0(y)} dx dy,$$

$$f_0(y) = \int_X f_0(x) f(y \mid x)\, dx; f_0(x) = \frac{1}{n}\sum_{i=1}^{n} f_i(x). \tag{29.5}$$

ФПВ $f_0(x)$ принадлежит множеству $\Phi(P)$, так как из (29.5) следует, что

$$P \ge \frac{1}{n}\sum_{i=1}^{n} \mathrm{M}X^2 = \frac{1}{n}\sum_{i=1}^{n} \int_X f_i(x) x^2 dx = \int_X x^2 f_0(x)\, dx.$$

В неравенстве (29.4) равенство достигается при $f_i(x) = f_0(x)$. Таким образом,

$$\max_{f(\boldsymbol{x})\in\Phi_n(P)} \frac{1}{n}\mathbf{I}\left(X^n;Y^n\right) \le \max_{f(\boldsymbol{x})\in\Phi_n(P)} \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}_i(X;Y) \le \max_{f(\boldsymbol{x})\in\Phi_n(P)} \mathbf{I}(X;Y). \tag{29.6}$$

Легко найти функцию $f(\boldsymbol{x})$, на которой достигается максимум в (29.6). Положим, $f(\boldsymbol{x}) = \prod_{i=1}^{n} f_0\left(x^{(i)}\right)$. Выходные сигналы при таком выборе вход-

ного распределения вероятностей статистически независимы. Следовательно, имеет место первое неравенство в (29.6). Кроме того, $f_i(\boldsymbol{x}) = f_0(\boldsymbol{x})$ и, следовательно, имеет место и второе неравенство в (29.6). Отсюда

$$\mathrm{C}^* = \sup_{n, f(\boldsymbol{x}) \in \Phi_n(P)} \frac{1}{n} \mathbf{I}\left(X^n; Y^n\right) = \max_{f(\boldsymbol{x}) \in \Phi_n(P)} \mathbf{I}(X;Y).$$

Теорема доказана.

Рассмотрим НК без памяти с дискретным временем и аддитивным гауссовским шумом. Предположим, что любая выходная последовательность $\boldsymbol{Y}$ может быть записана в виде

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{Z} = (X_1 + Z_1, ..., X_n + Z_n) \quad , \tag{29.7}$$

причем случайные последовательности $\boldsymbol{X}$, $\boldsymbol{Y}$ статистически независимы, а Z_i - независимые случайные гауссовские величины:

$$f_Z(\boldsymbol{z}) = \prod_{i=1}^{n} f_Z\left(z^{(i)}\right) \text{ , где } f_Z\left(z^{(i)}\right) = \frac{1}{\sigma_Z \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_Z^2} \left(z^{(i)}\right)^2 \right].$$

Число σ_Z^2 называется мощностью шума. Из (29.7) и статистической независимости $\boldsymbol{X}$, $\boldsymbol{Y}$ следует, что

$$f(\boldsymbol{y} \mid \boldsymbol{x}) = f_Z(\boldsymbol{y} - \boldsymbol{x}) = \prod_{i=1}^{n} f_Z\left(y^{(i)} - x^{(i)}\right).$$

Теорема 29.3. Информационная емкость НК без памяти с дискретным временем, с аддитивным гауссовским шумом мощностью σ_Z^2 и ограничением P на среднюю мощность входных сигналов определяется соотношением

$$\mathrm{C}^* = \frac{1}{2} \log\left(1 + \frac{P}{\sigma_Z^2}\right). \tag{29.8}$$

Для рассматриваемого сигнала

$$\mathbf{I}(X;Y) = H_0(Y) - H_0(Y \mid X) = H_0(Y) - H_0(Z) \le$$
$$\le \frac{1}{2} \log 2\pi e(P + \sigma_Z^2) - \frac{1}{2} \log 2\pi e \sigma_Z^2 = \frac{1}{2} \log\left(1 + \frac{P}{\sigma_Z^2}\right).$$

Равенство достигается, когда Y является гауссовской случайной величиной с дисперсией $P + \sigma_Z^2$, что справедливо, когда X является гауссовской величиной, имеет нулевое математическое ожидание и дисперсию P. Таким образом,

$$C^* = \max_{\Phi(P)} \mathbf{I}(X;Y) = \frac{1}{2}\log\left(1 + \frac{P}{\sigma_Z^2}\right),$$

где максимум достигается выбором гауссовской ФПВ $f(x)$ с параметрами $MX = 0;\ M X^2 = 0$.

Теорема 29.4 (прямая теорема кодирования). Пусть C* − информационная емкость НК без памяти с дискретным временем и аддитивным гауссовским шумом мощностью σ_Z^2 при ограничении P на среднюю мощность входных сигналов, определяемая (29.8). При любом $R < C^*$ и любом положительном δ существует код $G(n, R)$, удовлетворяющий ограничению P на среднюю мощность кодовых слов, максимальная вероятность ошибки декодирования которого удовлетворяет неравенству $\Lambda_n \le \delta$. Доказательство опускаем.

Следствием обратной и прямой теорем кодирования для НК без памяти с дискретным временем, аддитивным гауссовским шумом и ограничением на среднюю мощность входных сигналов является утверждение, выражаемое в виде $C = C^*$.

Вопросы для самопроверки

1. Дайте определение кода для непрерывного канала с непрерывным временем.

2. Дайте определение пропускной способности и информационной емкости непрерывного канала с дискретным временем.

3.Сформулируйте теорему о неравенстве Файнстейна в случае непрерывного канала без памяти с дискретным временем.

4. Как определяется информационная емкость непрерывного канала без памяти с ограничением на среднюю мощность входных сигналов?

ЗАКЛЮЧЕНИЕ

В пособии содержатся основные математические сведения, необходимые для ознакомления с теорией информации, их, конечно, явно недостаточно для глубокого понимания математических основ теории информации. Студентам, желающим более детально познакомиться с этими основами, можно рекомендовать кроме литературы, приведенной в библиографическом списке, следующие книги: Б.В. Гнеденко «Курс теории вероятностей», В. Феллер «Введение в теорию вероятностей и ее приложения», Ю.В. Похоров, Ю.А. Розанов «Теория вероятностей. Основные понятия, предельные теоремы, случайные процессы». По книге Б.З. Вулиха «Введение в функциональный анализ» можно изучить основные понятия функционального анализа.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Галагер, Р.* Теория информации и надежная связь /Р. Галагер. − М.: Сов. радио, 1974. – 376 с.

2. *Игнатов, В.А.* Теория информации и передача сигналов: учебник / В.А. Игнатов. − М.: Радио и связь, 1991. – 452 с.

3. *Кузьмин, И.В.* Основы теории информации и кодирования: учебник / И.В. Кузьмин, В.А. Кедрус. − Киев: Выща шк., 1986. – 324 с.

4. *Колесник, В.Д.* Курс теории информации: учеб. пособие / В.Д. Колесник, Г.Ш. Полтырев. − М.: Наука, 1982. – 416 с.

5. *Мурылев, А.В.* Теория информации: конспект лекций / А.В. Мурылев. − Саратов: Изд-во Сарат. гос. ун-та, 1977. – 254 с.

6. *Файнстейн, А.* Основы теории информации / А. Файнстейн. − М.: Иностр. лит., 1960. – 342 с.

7. *Фано, Р.* Передача информации. Статистическая теория связи / Р. Фано. М.: − Мир, 1965. – 420 с.

8. *Цымбал, В.П.* Теория информации и кодирования / В.П. Цымбал. − Киев: Наук. думка, 1977. – 356 с.

СПИСОК СИМВОЛОВ И ОБОЗНАЧЕНИЙ

ДИ – дискретный источник
ДК – дискретный канал
МАВ-декодирование – декодирование по максимуму апостериорной вероятности
МП-декодирование – декодирование по максимуму правдоподобия
НК – непрерывный канал
СВ – случайная величина
ФПВ – функция плотности вероятности
AL-канал – стационарный канал с аддитивным по модулю L-шумом
C – пропускная способность канала
C* – информационная емкость канала
$F(x)$ – функция распределения вероятности случайной величины x
$f(x)$ – функция распределения плотности вероятности случайной величины x
$H(X)$ – энтропия ансамбля X
$H(X \mid X^{\infty})$ – энтропия на сообщение
$H_0(X)$ – относительная (дифференциальная) энтропия ансамбля X
$I(x_i)$ – количество собственной информации в сообщении x_i
$I(x; y)$ – количество взаимной информации между сообщениями x и y
$\mathbf{I}(X; Y)$ – среднее количество взаимной информации между ансамблями X и Y
K_{XY} – корреляционный момент между случайными величинами x и y
$\mathbf{K} = [K_{ij}]$ – корреляционная матрица случайных величин
MX – оператор математического ожидания на множестве X
m_X – математическое ожидание случайной величины x
$\overline{m}(X)$ – средняя длина кодовых слов при неравномерном кодировании
$Pr(A)$ – вероятность множества A
P_{en} – вероятность ошибки кодирования
$p(x_i)$ – вероятность сообщения x_i
$p(x|y)$ – условная вероятность x при известном y
R – скорость кодирования
$T_n(\varepsilon)$ – высоковероятное множество
$\left|T_n(\varepsilon)\right|$ – количество элементов, принадлежащих высоковероятному множеству
U_X – источник, выбирающий сообщения из множества X
X, Y, A – множества
x, y, a – элементы множеств X, Y, A
$\{X, p(x)\}$ – вероятностный ансамбль X
$\delta(x)$ – дельта-функция Дирака
Λ – максимальная вероятность ошибки декодирования
λ – средняя вероятность ошибки декодирования
λ_i – вероятность ошибки декодирования
μ_k – k-й центральный момент случайной величины
σ_X^2 – дисперсия случайной величины x

ОГЛАВЛЕНИЕ

Учебное издание

ФЕДОРОВ Сергей Владимирович

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ТЕОРИИ ИНФОРМАЦИИ

Учебное пособие

Подписано в печать 31.05.10.

Формат 60x84/16. Усл. печ. л. 4,18. Тираж 100 экз.

Заказ

Издательство

Владимирского государственного университета

600000, Владимир, ул. Горького, 87.