А.Г.Дьячков

Теория информации

§6. Прямая теорема Шеннона для дискретного канала без памяти

Содержание

6.1 Определения, обозначения и формулировка теоремы

Будем рассматривать *дискретный канал без памяти* (ДКБП) с матрицей *переходных* вероятностей

$$\mathbf{P} = \|P(j|k)\|, \quad k = \overline{0, K-1}, \quad j = \overline{0, J-1},$$

имеющий $K \geq 2$ *входных* и $J \geq 2$ *выходных* символов, которые отождествим с целыми числами от 0 до $K-1$ и от 0 до $J-1$, соответственно. Число $P(j|k)$ означает вероятность получить на выходе канала связи символ j, если на вход канала был подан символ k. Если же по каналу передано слово $\mathbf{x} = (x_1, \dots, x_N)$, $x_i = \overline{0, K-1}$, длины N, то вероятность получить на выходе слово $\mathbf{y} = (y_1, \dots, y_N)$, $y_j = \overline{0, J-1}$, определяется соотношением

$$P_N(\mathbf{y}|\mathbf{x}) := \prod_{n=1}^{N} P(y_n|x_n). \tag{1}$$

Пусть $X = (\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(M))$, —код объема M. Кодовое слово длины N

$$\mathbf{x}(m) = (x_1(m), x_2(m), \dots, x_N(m)), \quad x_n(m) = \overline{0, K-1}, \quad n = \overline{1, N},$$

посылается по каналу, если передаваемое сообщение $\xi = m$, $m = \overline{1, M}$. При этом, согласно (1), вероятность получить на выходе канала связи слово $\mathbf{y} = (y_1, y_2, \dots, y_N)$ есть

$$P_N(\mathbf{y}|\mathbf{x}(m)) = \prod_{n=1}^{N} P(y_n|x_n(m)). \tag{2}$$

Как и в **§2**, будем рассматривать (N, R) –*блоковые коды*, т.е. коды X объема

$$M := \lceil \exp\{RN\} \rceil, \tag{3}$$

где параметр $R > 0$, называемый *скоростью* кода не зависит от длины блока N.

Описание *декодирования*, т.е. вычисления *сообщения адресата* $\bar{\xi}$ по выходному сигналу $\mathbf{y}$ с помощью *декодирующей* функции

$$D(\mathbf{y}) = m, \textit{если}\ \mathbf{y} \in D_m, \quad m = \overline{1, M+1},$$

$$(0, J-1)^N = D_1 + D_2 + \cdots + D_M + D_{M+1},$$

почти дословно повторяет соответствующее описание для двоичного случая из **§2**. Напомним, что при $m = \overline{1, M}$ *условная вероятность ошибки* кода X

$$\mathcal{P}_m(X) := \Pr\{\xi \neq \tilde{\xi} \mid \xi = m\} = \sum_{\mathbf{y} \in \bar{D}_m} P_N(\mathbf{y}|\mathbf{x}(m)), \tag{4}$$

а *средняя вероятность ошибки* кода X

$$\mathcal{P}(X) := \Pr\{\xi \neq \tilde{\xi}\} = \frac{1}{M} \sum_{m=1}^{M} \mathcal{P}_m(X). \tag{5}$$

Введем также

$$\mathcal{P}^*(X) := \max_{m=\overline{1,M}} \mathcal{P}_m(X) - \tag{6}$$

максимальную вероятность ошибки кода X .

Отметим, что с помощью рассуждений, почти дословно повторяющих рассуждения из **§2**, можно показать, что для заданного кода X декодирование по *максимуму правдоподобия* (МП–декодирование), когда

$$D_m = D_m^{(\text{МП})} := \{\mathbf{y} \,:\, \forall\, m' \neq m, P_N(\mathbf{y}|\mathbf{x}(m')) \leq P_N(\mathbf{y}|\mathbf{x}(m))\}, \tag{7}$$

максимизирует среднюю *вероятность правильного решения*

$$1 - \mathcal{P}(X) = \frac{1}{M} \sum_{m=1}^{M} \sum_{\mathbf{y} \in D_m} P_N(\mathbf{y}|\mathbf{x}(m)), \tag{8}$$

т.е. *минимизирует* среднюю вероятность ошибки $\mathcal{P}(X)$.

Для заданных N , $R > 0$ и матрицы $\mathbf{P} = \|P(j|k)\|$ переходных вероятностей ДКБП (1) введем среднюю вероятность ошибки *оптимального (наилучшего)* (N,R) –кода

$$\mathcal{E}_N(\mathbf{P}, R) := \min_{X,D} \mathcal{P}(X),$$

где минимум средней вероятности ошибки берется по всевозможным (N,R) –кодам и декодирующим функциям. Аналогично определяется максимальная вероятность ошибки оптимального (наилучшего) (N,R) –кода

$$\mathcal{E}_N^*(\mathbf{P}, R) := \min_{X,D} \mathcal{P}^*(X).$$

Очевидно, что $\mathcal{E}_N^*(\mathbf{P}, R) \geq \mathcal{E}_N(\mathbf{P}, R)$.

Пусть $\mathbf{Q} = (Q(0), \ldots, Q(K-1))$ — распределение вероятностей на выходных символах канала (1), а $\tilde{\mathbf{Q}} = (\tilde{Q}(0), \ldots, \tilde{Q}(J-1))$, где

$$\tilde{Q}(j) = \sum_{k=0}^{K-1} Q(k) P(j|k), \quad j = \overline{0, J-1}, \tag{9}$$

соответствующее распределение вероятностей на выходных символах канала (1). Для рассматриваемого ДКБП с фиксированной переходной матрицей $\mathbf{P} = \|P(j|k\|$, $k = \overline{0, K-1}$, $j = \overline{0, J-1}$, введем *информационную функцию* канала

$$\mathbf{J}(\mathbf{Q}, \mathbf{P}) := \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} Q(k) P(j|k) \log \frac{P(j|k)}{\tilde{Q}(j)}, \tag{10}$$

определенную на множестве вероятностных распределений $\mathbf{Q}$. Число

$$C = C(\mathbf{P}) := \max_{\mathbf{Q}} \mathbf{J}(\mathbf{Q}, \mathbf{P}) \tag{11}$$

называется *пропускной способностью* канала без памяти на единицу времени. В **§5** была доказана *обратная теорема Шеннона*, которая утверждала, что

$$\liminf_{N\to\infty} \mathcal{E}_N(\mathbf{P}, R) \geq 1 - \frac{C}{R},$$

а потому при $R > C$ величина $\mathcal{E}_N(\mathbf{P}, R) \not\to 0$ при $N \to \infty$. Аналогичный результат, очевидно, справедлив и для $\mathcal{E}_N^*(\mathbf{P}, R)$.

Замечание. В разделе **6.4** будет показано, что информационная функция (10) является выпуклой вверх функцией аргумента $\mathbf{Q}$. Поэтому, в силу известной теоремы Куна–Таккера, необходимые условия экстремума в (11), вычисляемые по теореме Лагранжа, будут также достаточными и экстремальное распределение $\mathbf{Q}$ является решением системы уравнений

$$\frac{\partial \mathbf{J}(\mathbf{Q}, \mathbf{P})}{\partial Q(k)} = \lambda, \ \textit{при всех } k\textit{, для которых } Q(k) > 0\,,$$

$$\frac{\partial \mathbf{J}(\mathbf{Q}, \mathbf{P})}{\partial Q(k)} \leq \lambda, \ \textit{при всех } k\textit{, для которых } Q(k) = 0\,.$$

Множитель Лагранжа λ выбирается таким образом, чтобы максимум в (11) имел место при

$$\sum_{k=0}^{K-1} Q(k) = 1.$$

Наша дальнейшая цель — доказательство следующих двух теорем, которые называются прямыми теоремами Шеннона для ДКБП.

Теорема 1. (**Теорема Шеннона о средней ошибке**). *Если* $R < C$, *то существует* (N, R) *–код* X *и декодирование* $D = D(\mathbf{y})$, *для которых* $\mathcal{P}(X) \to 0$ *при* $N \to \infty$.

Теорема 2. (**Теорема Шеннона о максимальной ошибке**). *Если* $R < C$, *то существует* (N, R) *–код* X^* *и декодирование* $D^* = D^*(\mathbf{y})$, *для которых* $\mathcal{P}^*(X^*) \to 0$ *при* $N \to \infty$.

Из теорем 1 и 2 вытекает

Следствие. (**Прямая теорема Шеннона**). *Если* $0 < R < C$, *то существует*

$$\lim_{N\to\infty} \mathcal{E}_N(\mathbf{P}, R) = \lim_{N\to\infty} \mathcal{E}_N^*(\mathbf{P}, R) = 0.$$

6.2 Доказательства теорем 1 и 2

6.2.1 Пороговое декодирование

Зафиксируем произвольное распределение вероятностей (р.в.) $\mathbf{Q} = (Q(0),\ldots,Q(K-1))$ на входном алфавите канала (1) и обозначим через

$$Q_N(\mathbf{x}) := \prod_{n=1}^{N} Q(x_n) \ - \tag{12}$$

соответствующее *распределение–произведение* на входных словах $\mathbf{x} = (x_1,\ldots,x_N)$ канала (1). На выходных словах $\mathbf{y} = (y_1,\ldots,y_N)$ введем распределение вероятностей

$$\tilde{Q}_N(\mathbf{y}) := \sum_{\mathbf{x}} Q_N(\mathbf{x}) P_N(\mathbf{y}|\mathbf{x}), \tag{13}$$

которое возникает на выходе канала (1), если на его входе слово $\mathbf{x}$ имеет р.в. (12). Воспользовавшись известным правилом

$$\sum_{\mathbf{x}} \prod_{i=1}^{N} f(x_i,y_i) = \prod_{i=1}^{N} \Big(\sum_{x=0}^{K-1} f(x,y_i)\Big),$$

которое справедливо для любой функции двух переменных $f(x,y)$, можем написать, что

$$\tilde{Q}_N(\mathbf{y}) = \prod_{i=1}^{N} \tilde{Q}(y_i), \tag{14}$$

где р.в. $\tilde{\mathbf{Q}} = (\tilde{Q}(0),\ldots,\tilde{Q}(J-1))$ на выходных символах канала (1) определено (9).

На множестве пар $(\mathbf{x},\mathbf{y})$ определим функцию

$$I_N(\mathbf{x},\mathbf{y}) := \log \frac{P_N(\mathbf{y}|\mathbf{x})}{\tilde{Q}_N(\mathbf{y})} = \sum_{i=1}^{N} \log \frac{P_N(y_i|x_i)}{\tilde{Q}(y_i)}, \tag{15}$$

где второе равенство вытекает из (1) и (14). Для фиксированного кода $X = (\mathbf{x}(1),\ldots,\mathbf{x}(M))$ рассмотрим *пороговый декодер*, который работает следующим образом. При заданной принятой последовательности $\mathbf{y}$ декодер вычисляет статистику

$$I(m) = I_N(\mathbf{x}(m),\mathbf{y})$$

для каждого кодового слова $\mathbf{x}(m)$, $m = \overline{1, M}$. Декодер сравнивает все статистики $I(m)$ с числом $T{\cdot}N$, где T — некоторый фиксированный параметр (порог). Если имеется одно и только одно значение m, для которого $I(m) > T{\cdot}N$, то декодер принимает решение о том, что по каналу передавалось слово $\mathbf{x}(m)$. В противном случае *производит стирание, отказываясь принимать решение.*

Для данного способа принятия решения при передаче слова $\mathbf{x}(m)$ ошибочное решение (необнаруженая ошибка) происходит тогда и только тогда, когда статистика $I(m) {\leq} T{\cdot}N$ и существует ровно одно значение $m'{\neq}m$ для которого $I(m') > T{\cdot}N$. Правильное решение при передаче слова $\mathbf{x}(m)$ происходит тогда и только тогда, когда статистика $I(m) > T{\cdot}N$ и для всех значений $m'{\neq}m$ статистика $I(m'){\leq}T{\cdot}N$.

Замечание. Для рассматриваемого порогового декодирования области декодирования $D_m^{(T)}$, $m = \overline{1, M}$, при использовании порога $T > 0$ задается следующим образом

$$D_m^T = \{\mathbf{y} \,:\, I_N(\mathbf{x}(m), \mathbf{y}) > T{\cdot}N, \textit{ и для любого } m'{\neq}m \textit{ значение}$$

$$I_N(\mathbf{x}(m'), \mathbf{y}) \leq T{\cdot}N\}.$$

Пример. Пороговое декодирование для ДСК с вероятностью ошибки q, $0 < q < 1/2$. Выберем в качестве р.в. $\mathbf{Q} = (1/2, 1/2)$, на котором достигается пропускная способность ДСК без памяти $C = \log 2 - h(q)$. Тогда очевидно $\tilde{\mathbf{Q}} = (1/2, 1/2)$ и для любого $\mathbf{y}$ вероятность $\tilde{Q}_N(\mathbf{y}) = 2^{-N}$. Поэтому, согласно определению (15) и определению ДСК в §2,

$$I_N(\mathbf{x}, \mathbf{y}) = N \log 2{\cdot}(1 - q) + \rho(\mathbf{x}, \mathbf{y}) \log \frac{q}{1 - q}, \quad 0 < q < 1/2.$$

Следовательно, положив

$$T^* = T^*(q, T) = \frac{\log 2(1 - q) - T}{\log \frac{1-q}{q}},$$

области декодирования $D_m^{(T)}$, $m = \overline{1, M}$, можно записать в виде

$$D_m^{(T)} = \{\mathbf{y} \,:\, \rho(\mathbf{x}(m), \mathbf{y}) < T^* N, \textit{ а для любого } m' \neq m$$

$$\textit{расстояние Хэмминга} \quad \rho(\mathbf{x}(m'), \mathbf{y}) \geq T^* N\}.$$

Пусть

$$\pi_1^{(N)} = \pi_1(\mathbf{x}(m), T) \;-$$

условная вероятность того, что статистика $I(m) \le T{\cdot}N$ при условии, что по каналу было передано слово $\mathbf{x}(m)$. Для $m' \ne m$ введем

$$\pi_2^{(N)}(m') = \pi_2(\mathbf{x}(m'), T) \; -$$

условную вероятность того, что статистика $I(m') > T{\cdot}N$ при условии, что по каналу было передано слово $\mathbf{x}(m)$. Имеем

$$\pi_1^{(N)} = \sum_{\mathbf{y}} P_N(\mathbf{y}|\mathbf{x}(m))\chi_T(\mathbf{x}(m), \mathbf{y}),$$

$$\pi_2^{(N)}(m') = \sum_{\mathbf{y}} P_N(\mathbf{y}|\mathbf{x}(m)){\cdot}[1 - \chi_T(\mathbf{x}(m'), \mathbf{y})], \quad m' \ne m, \tag{16}$$

где характеристическая функция

$$\chi_T(\mathbf{x}, \mathbf{y}) := \begin{cases} 1, \textit{если } I_N(\mathbf{x}, \mathbf{y}) \le T{\cdot}N, \\ 0, \textit{если } I_N(\mathbf{x}, \mathbf{y}) > T{\cdot}N, \end{cases} \tag{17}$$

а $I_N(\mathbf{x}, \mathbf{y})$ определена (15).

Далее под *ошибкой* будем подразумевать ситуацию либо *стирания* либо *необнаруженной ошибки.* Из описанного выше алгоритма порогового декодирования вытекает

Лемма 1. (**Граница ошибки порогового декодера**). *Для любого фиксированного кода* $X = (x(1), \ldots, x(M))$ *при пороговом декодировании условная вероятность ошибки* (4) *удовлетворяет неравенству*

$$\mathcal{P}_m(X) \le \pi_1^{(N)} + \sum_{m' \ne m}^{M} \pi_2^{(N)}(m'). \tag{18}$$

Для получения верхних оценок вероятностей $\pi_1^{(N)}$ и $\pi_2^{(N)}(m')$ применяется предложенный Шенноном *метод случайного кодирования.*

6.2.2 Метод случайного кодирования

Рассмотрим *ансамбль* (N, R) –кодов $X = (\mathbf{x}(1), \ldots, \mathbf{x}(M))$, где р.в. $Q(X)$ на ансамбле X задается соотношением

$$Q(X) := \prod_{m=1}^{M} Q_N(\mathbf{x}(m)), \tag{19}$$

где $Q_N(\mathbf{x})$ определено (12). При этом элементы $x_i(m)$, $i = \overline{1,N}$, $m = \overline{1,M}$, кодовой матрицы $X = \|x_i(m)\|$ являются независимыми одинаково распределенными случайными величинами с р.в. $\mathbf{Q} = (Q(0),\ldots,Q(K-1))$. Как и в **§2**, *чертой сверху* будем обозначать усреднение (математическое ожидание) по ансамблю (19).

Пусть $\overline{\pi_1^{(N)}}$ и $\overline{\pi_2^{(N)}(m')}$ — среднии по ансамблю значения соответствующих вероятностей. В силу (16), имеем

$$\overline{\pi_1^{(N)}} = \sum_X \sum_{\mathbf{y}} Q(X)\cdot P_N(\mathbf{y}|\mathbf{x}(m))\cdot\chi_T(\mathbf{x}(m),\mathbf{y}).$$

Из определения (19) следует, что $\overline{\pi_1^{(N)}}$ не зависит от $m = \overline{1,M}$ и вычисляется по формуле

$$\overline{\pi_1^{(N)}} = \sum_{\mathbf{x}} \sum_{\mathbf{y}} Q_N(\mathbf{x})\cdot P_N(\mathbf{y}|\mathbf{x})\cdot\chi_T(\mathbf{x},\mathbf{y}). \tag{20}$$

В силу (16), для любого $m' \neq m$ число $\overline{\pi_2^{(N)}(m')}$ не зависит от $m' = \overline{1,M}$, $m' \neq m$, и

$$\overline{\pi_2^{(N)}(m')} = \sum_X \sum_{\mathbf{y}} Q(X)\cdot P_N(\mathbf{y}|\mathbf{x}(m))[1-\chi_T(\mathbf{x}(m'),\mathbf{y})].$$

Отсюда, используя (13) и (19), получаем, что $\overline{\pi_2^{(N)}(m')}$ для любого $m' \neq m$ вычисляется по формуле

$$\overline{\pi_2^{(N)}(m')} = \sum_{\mathbf{x}} \sum_{\mathbf{y}} \tilde{Q}_N(\mathbf{y})\cdot Q_N(\mathbf{x})[1-\chi_T(\mathbf{x},\mathbf{y})]. \tag{21}$$

Усредняя неравенство (18) по распределению ансамбля $Q(X)$ и подставляя результат в определение вероятности ошибки (5), получим

$$\overline{\mathcal{P}(X)} \le \overline{\pi_1^{(N)}} + (M-1)\cdot\overline{\pi_2^{(N)}(m')}, \tag{22}$$

где $\overline{\mathcal{P}(X)}$ — усреднение вероятности ошибки $\mathcal{P}(X)$ по распределению (19). Поскольку случайная величин должна принимать хотя бы одно значение, не превышающее ее среднее (математическое ожидание), то из (20)–(22) вытекает

Лемма 2. (**Граница случайного кодирования**). *Существует код X объЯма M такой, что при декодировании с порогом T средняя вероятность ошибки*

$$\mathcal{P}(X) \le \mathcal{P}_N^{(1)} + (M-1)\mathcal{P}_N^{(2)}, \tag{23}$$

где

$$\mathcal{P}_N^{(1)} = \mathcal{P}_N^{(1)}(T,Q) := \sum_{\mathbf{x}} \sum_{\mathbf{y}} Q_N(\mathbf{y}) \cdot P_N(\mathbf{y}|\mathbf{x}) \chi_T(\mathbf{x},\mathbf{y}), \tag{24a}$$

$$\mathcal{P}_N^{(2)} = \mathcal{P}_N^{(2)}(T,Q) := \sum_{\mathbf{x}} \sum_{\mathbf{y}} \tilde{Q}_N(\mathbf{y}) \cdot Q_N(\mathbf{x}) \cdot [1 - \chi_T(\mathbf{x},\mathbf{y})]. \tag{25a}$$

6.2.3 Верхние границы $\mathcal{P}_N^{(i)}$, $i = \overline{1,2}$

Введем независимые одинаково распределенные величины η_n, $n = \overline{1,N}$, где р.в. η_n задается соотношениями

$$\Pr\{\eta_n = \log \frac{P(j|k)}{\tilde{Q}(j)}\} = Q(k)P(j|k), \quad k = \overline{0,K-1}, \quad j = \overline{0,J-1}. \tag{24б}$$

Из определений (10), (15) и (24а) следует, что

$$\mathcal{P}_N^{(1)} = \Pr\{\sum_{n=1}^{N} \eta_n \le TN\}, \tag{24в}$$

а математическое ожидание $\bar{\eta}_n = \mathbf{J}(\mathbf{Q},\mathbf{P})$, $n = \overline{1,N}$, где $\mathbf{J}(\mathbf{Q},\mathbf{P})$ — значение информационной функции канала (10) для рассматриваемого при пороговом декодировании р.в. $\mathbf{Q}$.

Замечание. Для величины $\mathcal{P}_N^{(2)}$ имеет место следующая вероятностная интерпретация

$$\mathcal{P}_N^{(2)} = \Pr\{\sum_{n=1}^{N} \eta'_n > TN\}, \tag{25б}$$

где η'_1, η'_2, ..., η'_N — независимые случайные величины с одинаковым р.в.

$$\Pr\{\eta'_n = \log \frac{P(j|k)}{\tilde{Q}(j)}\} = Q(k)\tilde{Q}(j), \quad k = \overline{0,K-1}, j = \overline{0,J-1}. \tag{25в}$$

Зафиксируем число δ, $0 < \delta < \mathbf{J}(\mathbf{Q},\mathbf{P})$, и выберем порог $T{:=}\mathbf{J}(\mathbf{Q},\mathbf{P}) - \delta$. Тогда, в силу закона больших чисел

$$\mathcal{P}_N^{(1)} = \Pr\Big\{\frac{\sum_{n=1}^{N} \eta_n}{N} \le \mathbf{J}(\mathbf{Q},\mathbf{P}) - \delta\Big\} \Rightarrow 0, \quad N \Rightarrow \infty. \tag{26}$$

Из (15), (17) и (25a) следует, что

$$\mathcal{P}_N^{(2)} = \sum_{(\mathbf{x},\mathbf{y}):\log \frac{P_N(\mathbf{y}|\mathbf{x})}{\tilde{Q}_N(\mathbf{y})} > T\cdot N} Q_N(x)\tilde{Q}_N(y),$$

т.е суммирование идет по тем парам $(\mathbf{x},\mathbf{y})$, для которых $\tilde{Q}_N(\mathbf{y}) \le P_N(\mathbf{y}|\mathbf{x})\exp\{-T\cdot N\}$, где уже выбрали $T = \mathbf{J}(\mathbf{Q},\mathbf{P}) - \delta$. Поэтому

$$\mathcal{P}_N^{(2)} \le \exp\{-N[\mathbf{J}(\mathbf{Q},\mathbf{P}) - \delta]\} \sum_{(\mathbf{x},\mathbf{y})} Q_N(\mathbf{x})\cdot P_N(\mathbf{y}|\mathbf{x}) = \exp\{-N[\mathbf{J}(\mathbf{Q},\mathbf{P}) - \delta]\}.$$

Определение (N,R) –кода (3) означает, что $M-1 < \exp\{NR\}$. Поэтому

$$(M-1)\cdot\mathcal{P}_N^{(2)} \le \exp\{-N[\mathbf{J}(\mathbf{Q},\mathbf{P}) - R - \delta]\}. \tag{27}$$

Пусть распределение $\mathbf{Q}$ в описании порогового декодера совпадает с тем р.в., на котором достигается максимум в определении (11), т.е $\mathbf{J}(\mathbf{Q},\mathbf{P}) = C$ — пропукслая способность ДКБП. Если $R < C$, то положим $\delta = (C-R)/2$. Тогда (27) означает, что при выборе порога $T := (C+R)/2$ величина

$$(M-1)\cdot\mathcal{P}_N^{(2)} \le \exp\{-\frac{C-R}{2}N\}. \tag{28}$$

Подставляя (26) и (28) в (23) получаем утверждение теоремы 1.

Теорема 1 доказана.

Замечание. Вероятность $\mathcal{P}_N^{(1)}$, определенная (24) при $T < \bar{\eta}_n = \mathbf{J}(\mathbf{Q},\mathbf{P})$, называется *вероятностью больших уклонений* для суммы независимых случайных величин η_n, $n = \overline{1,N}$. Аналогичный смысл имеет и вероятность $\mathcal{P}_N^{(2)}$, определенная (25) при $T > \overline{\eta_n'}$. В разделе **6.3** будет построена общая верхняя граница для вероятности больших уклонений суммы независимых случайных величин, с помощью которой будет доказано, что при $0 < R \le T < \mathbf{J}(\mathbf{Q},\mathbf{P})$ правая часть (23) экспоненциально убывает с ростом N. Это означает, что при $0 < R \le T < C$ существует (N,R) –код и T –пороговый декодер с вероятностью ошибки, экспоненциально убывающей с ростом N.

6.2.4 Доказательство теоремы 2

Имеет место очевидное (доказательство от противного) утверждение:

Пусть $\alpha > 0$ и $a_1, a_2, \ldots, a_{2M}$, $a_i \geq 0$, — действительные неотрицательные числа, среднее арифметическое которых

$$\frac{1}{2M}\sum_{m=1}^{2M} a_m \leq \alpha.$$

Тогда среди a_m, $m = \overline{1,2M}$, обязательно найдутся M чисел $a_{m_1}, a_{m_2}, \ldots, a_{m_M}$, $1 \leq m_1 < m_2 < \ldots < m_M \leq 2M$, которые не превосходят 2α, т.е. $a_{m_i} \leq 2\alpha$, $i = \overline{1,M}$.

Из леммы 2 следует существование кода $X = (\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(2M))$ длины N, объёма $2M$, для которого при пороговом декодировании средняя вероятность ошибки (5) удовлетворяет неравенству

$$\mathcal{P}(X) \leq \mathcal{P}_N^{(1)} + (2M-1)\cdot\mathcal{P}_N^{(2)} := \alpha.$$

Поскольку

$$\mathcal{P}(X) = \frac{1}{2M}\sum_{m=1}^{2M} \mathcal{P}_m(X),$$

то, в силу леммы 3, существует код $X^* = (\mathbf{x}(m_1), \mathbf{x}(m_2), \ldots, \mathbf{x}(m_M))$, для которого все условные вероятности ошибки

$$\mathcal{P}_{m_i}(X^*) \leq 2\alpha < 2\mathcal{P}_N^{(1)} + 4M\cdot\mathcal{P}_N^{(2)},\ i = \overline{1,M}.$$

Здесь мы учли, что при переходе от кода X объёма $2M$ к коду X^* объёма M и использовании порогового декодера, область декодирования слова $\mathbf{x}(m_i)$ $D_{m_i} \subseteq D^*_{m_i}$, а дополнение $\overline{D^*_{m_i}} \subseteq \overline{D_{m_i}}$. Поэтому условная вероятность ошибки $\mathcal{P}_m(X^*) \leq \mathcal{P}_{m_i}(X)$.

Следовательно, для максимальной вероятности ошибки (6) кода X^* при пороговом декодировании имеем

$$\mathcal{P}^*(X^*) \leq 2\mathcal{P}_N^{(1)} + \frac{4M}{M-1}\cdot[\mathcal{P}_N^{(2)}\cdot(M-1)]. \tag{29}$$

Подставляя (26) и (28) в (29) получаем утверждение теоремы 2 о максимальной вероятности ошибки.

Теорема 2 доказана.

6.3 Экспоненциальная граница вероятности ошибки

6.3.1 Верхние границы вероятностей больших уклонений

Пусть $\mathcal{Z}$ — конечное множество объёма $|\mathcal{Z}|$ с элементами $z \in \mathcal{Z}$, а $P = \{p(z),\ z \in \mathcal{Z}\}$, $0 \leq p(z) < 1$, $\sum\limits_{\mathcal{Z}} p(z) = 1$, — произвольное фиксированное р.в. на $\mathcal{Z}$. Числовой функции $f = f(z)$, $z \in \mathcal{Z}$, сопоставим последовательность $\xi = (\xi_1, \ldots, \xi_N)$ независимых случайных величин с одним и тем же р.в. $p(z) = \Pr\{\xi_i = f(z)\}$, $z \in \mathcal{Z}$. Используя черту сверху как символ математического ожидания, введём:

$$m = \overline{\xi_i} = \sum_{\mathcal{Z}} f(z) \cdot p(z),\ i = \overline{1, N},$$

а также функции:

$$\phi(h) = \overline{\exp\{h \cdot \xi_i\}} = \sum_{\mathcal{Z}} p(z) \cdot \exp\{h \cdot f(z)\},$$

$$\mu(h) = \log \phi(h),$$

действительного параметра h, которые называются *производящей функцией моментов* и *производящей функцией семиинвариантов*, соответственно.

Пусть

$$S_N = \sum_{i=1}^{N} \xi_i$$

стандартное обозначение суммы независимых случайных величин. Наша цель — построить верхнюю границу вероятности больших уклонений, т.е. величины

$$\pi_N(T) = \begin{cases} \Pr(S_N \geq N \cdot T), & \text{если } T \geq m, \\ \Pr(S_N \leq N \cdot T), & \text{если } T \leq m. \end{cases}$$

Далее эта граница будет использована при построении верхних границ вероятностей $\mathcal{P}_N^{(i)}$, $i = \overline{1, 2}$, определённых соотношениями $(24')$, $(25')$.

Лемма 3. 1) *Если $T \geq m$, то для любого $h \geq 0$ вероятность*

$$\pi_N(T) \leq \exp\{-N \cdot [h \cdot T - \mu(h)]\},\ h \geq 0.$$

2) *Если $T \leq m$, то для любого $h \leq 0$ вероятность*

$$\pi_N(T) \leq \exp\{-N \cdot [h \cdot T - \mu(h)]\},\ h \leq 0.$$

Доказательство. Докажем лишь утверждение 1), поскольку утверждение 2) устанавливается аналогично. При $T \geq m$ *для любого* $h \geq 0$ имеем

$$\pi_N(T) = \Pr(h \cdot S_N \geq h \cdot N \cdot T) = \Pr(\exp\{h \cdot S_N\} \geq \exp\{h \cdot N \cdot T\}) \leq$$
$$\leq \overline{\exp\{h \cdot S_N\}} \cdot \exp\{-h \cdot N \cdot T\},$$

где воспользовались неравенством Маркова:

$$\Pr(\xi{\geq}X) \leq \frac{\overline{\xi}}{X},\ X > 0,$$

которое справедливо для любой случайной величины $\xi{\geq}0$. В силу независимости ξ_i имеем

$$\overline{\exp\{h{\cdot}S_N\}} = \overline{\exp\{\sum\nolimits_{i=1}^{N} h{\cdot}\xi_i\}} = \prod_{i=1}^{N} \overline{\exp\{h{\cdot}\xi_i\}} = (\phi(h))^N = \exp\{N{\cdot}\mu(h)\}.$$

Подставив это выражение в верхнюю оценку $\pi_N(T)$, получим доказываемое утверждение 1).

Лемма 3 доказана.

Из леммы 3 вытекает

Лемма 4. *Вероятность*

$$\pi_N(T) \leq \exp\{-N{\cdot}E(T)\},$$

$$E(T) = \begin{cases} \max\limits_{h\geq 0}\{h{\cdot}T - \mu(h)\},\ \textit{если}\ T{\geq}m, \\ \max\limits_{h\leq 0}\{h{\cdot}T - \mu(h)\},\ \textit{если}\ T{\leq}m. \end{cases}$$

Граница леммы 4 называется *границей Чернова* для вероятности больших уклонений.

Свойства функции $\mu(h)$ устанавливает:

Лемма 5. *Функция* $\mu(h)$ *выпукла вниз,* т.е. $\mu''(h) > 0$. *Кроме того*

$$\mu(0) = 0,\ \ \mu'(0) = m.$$

Доказательство. Равенства леммы очевидны. Для доказательства выпуклости, как при выводе аналогичной леммы из §4, рассмотрим произвольные числа $h,\ h_1,\ h_2$ и $\lambda,\ 0 < \lambda < 1$, такие что

$$h = \lambda{\cdot}h_1 + (1-\lambda){\cdot}h_2,$$

и покажем, что

$$\mu(h){\leq}\lambda{\cdot}\mu(h_1) + (1-\lambda){\cdot}\mu(h_2). \tag{30}$$

Согласно определению $\phi(h)$, имеем

$$\phi(h) = \sum_{\mathcal{Z}} \exp\{h{\cdot}f(z)\}{\cdot}p(z) =$$
$$= \sum_{\mathcal{Z}} [\exp\{h_1{\cdot}f(z)\}{\cdot}p(z)]^{\lambda} \cdot [\exp\{h_2{\cdot}f(z)\}{\cdot}p(z)]^{1-\lambda} \leq$$
$$\leq \phi(h_1)^{\lambda} \cdot \phi(h_2)^{1-\lambda},$$

где пользовались неравенством Гёльдера:

$$\sum_i a_i^{\lambda} \cdot b_i^{1-\lambda} \leq \left(\sum_i a_i\right)^{\lambda} \cdot \left(\sum_i b_i\right)^{1-\lambda}.$$

Логарифмируя полученное неравенство и учитывая определение функции $\mu(h)$, приходим к неравенству (30).

Лемма 5 доказана.

С помощью леммы 5 легко доказываются ниже формулируемые леммы 6 и 7, устанавливающие свойства показателя экспоненты $E(T)$ в границе леммы 4.

Лемма 6. *Пусть*

$$A = \max_{\mathcal{Z}} f(z) = f(z_A).$$

Тогда при $m \leq T \leq A$ *функция* $E(T)$ *принимает конечные значения и описывается параметрическими соотношениями*

$$\begin{aligned} &E(T) = E_0(h) = h \cdot \mu'(h) - \mu(h), \\ &T = T_0(h) = \mu'(h), \; h \geq 0. \end{aligned} \tag{31}$$

При этом производные $E'(T)$ *и* $E''(T)$ *также задаются параметрически*

$$\begin{cases} E'(T) = \frac{E_0'(h)}{T_0'(h)} = \frac{h \cdot \mu''(h)}{\mu''(h)} = h, \\ T = \mu'(h), \; h \geq 0, \end{cases} \tag{32}$$

$$\begin{cases} E''(T) = \frac{1}{\mu''(h)}, \\ T = \mu'(h), \; h \geq 0. \end{cases} \tag{33}$$

Лемма 6 означает, что при $m \leq T \leq A$ функция $E(T)$ монотонно возрастает, выпукла вниз, $E(m) = 0$, $E'(m) = 0$ и

$$E(A) = \lim_{h \to \infty} E_0(h) = -\ln p(z_A).$$

Лемма 7. *Пусть*

$$a = \min_{\mathcal{Z}} f(z) = f(z_a).$$

Тогда при $a \leq T \leq m$ *функция* $E(T)$ *принимает конечные значения и описывается вместе со своими производными* $E'(T)$ *и* $E''(T)$ *параметрическими соотношениями* (31) – (33), *в которых параметр* $h \leq 0$.

Лемма 7 означает, что при $a\le T\le m$ функция $E(T)$ монотонно убывает, выпукла вниз, $E(m)=0$, $E'(m)=0$ и

$$E(A)=\lim_{h\to-\infty} E_0(h)=-\ln p(z_A).$$

Замечание. Можно показать, что граница леммы 4 асимптотически точна, когда $N\to\infty$, т.е. существует предел

$$\lim_{N\to\infty}\frac{-\ln\pi_N(T)}{N}=E(T).$$

Задача 6.1. Рассмотрите в качестве примера случай, когда S_N — число успехов в испытаниях Бернули. Проверьте, что параметрическое определение показателя экспоненты (31) совпадает с результатом задачи 2.4 из §2. Нарисуйте графики функций $\mu(h)$ и $E(T)$.

Задача 6.2. Нарисуйте вид графиков функций $\mu(h)$ и $E(T)$ в общем случае. Для определённости рассмотрите, например, ситуацию, когда $0<a<m<A$.

6.3.2 Вероятность ошибки порогового декодирования

Для рассматриваемого в разделе **6.2** ДКБП $\mathbf{P}=\|P(j|k)\|$, $k=\overline{0,K-1}$, $j=\overline{0,J-1}$, распределения вероятностей (р.в.) на входе $\mathbf{Q}=(Q(k),\, k=\overline{0,K-1})$ и соответствующего р.в. на выходе $\tilde{\mathbf{Q}}=(\tilde{Q}(j),\ j=\overline{0,J-1})$, где

$$\tilde{Q}(j)=\sum_{k=0}^{K-1}Q(k)\cdot P(j|k),$$

введём функцию действительного параметра h

$$\mu_{\mathbf{Q}}(h)=\log\left[\sum_{k=0}^{K-1}\sum_{j=0}^{J-1}Q(k)\cdot P(j|k)\cdot\left(\frac{P(j|k)}{\tilde{Q}(j)}\right)^h\right]. \qquad (34)$$

Эта функция является производящей функцией семиинвариантов случайной величины η_n, $n=\overline{0,N}$, определённой (24б).

Пусть $\mathbf{J}(\mathbf{Q},\mathbf{P})=\overline{\eta}_n$ — информационная функция канала (10), а определяемые (24)–(25) величины $\mathcal{P}_N^{(i)}$, $i=\overline{1,2}$, задают верхнюю границу (23) для вероятности ошибки T–порогового декодирования. Имеет место

Лемма 8. 1) *Если* $0 \le T \le \mathbf{J}(\mathbf{Q},\mathbf{P})$, *то*

$$\mathcal{P}_N^{(1)} \le \exp\bigl\{-NE_{\mathbf{Q}}(T)\bigr\}, \tag{35}$$

$$\mathcal{P}_N^{(2)} \le \exp\bigl\{-N\bigl[T+E_{\mathbf{Q}}(T)\bigr]\bigr\}, \tag{36}$$

$$E_{\mathbf{Q}}(T) \le \max_{-1\le h\le 0}\bigl\{Th-\mu_{\mathbf{Q}}(h)\bigr\}, \tag{37}$$

где $\mu_{\mathbf{Q}}(h)$ *определена* (34). 2)*При* $0 < T < \mathbf{J}(\mathbf{Q},\mathbf{P})$ *функция* $E_{\mathbf{Q}}(T) > 0$, *монотонно убывает, выпукла вниз и при* $T=\mathbf{J}(\mathbf{Q},\mathbf{P})$ *число*

$$E_{\mathbf{Q}}(\mathbf{J}(\mathbf{Q},\mathbf{P})) = \frac{\partial E_{\mathbf{Q}}(\mathbf{J}(\mathbf{Q},\mathbf{P}))}{\partial T} = 0.$$

Доказательство. 1) Применяя утверждение 2) леммы 3 для получения верхней оценки (24), можем написать, что для любого $T \le \mathbf{J}(\mathbf{Q},\mathbf{P})$ и любого $h \le 0$

$$\mathcal{P}_N^{(1)} \le \exp\{-N(hT-\mu_{\mathbf{Q}}(h))\}, \tag{38}$$

где $\mu_{\mathbf{Q}}(h)$ определена (34).

Прежде чем оценивать $\mathcal{P}_N^{(2)}$, заметим, что в силу (25 в) математическое ожидание

$$\bar{\eta'_n} = \sum_{k,j} Q(k)\tilde{Q}(j)\ln\frac{P(j/k)}{\tilde{Q}(j)}, \quad n=1,\ldots,N. \tag{39}$$

Применяя неравенство $\ln x \le x-1$ для оценки (39), получаем

$$\bar{\eta'_n} \le \sum_{k,j} Q(k)\tilde{Q}(j)\bigl[\frac{P(j/k)}{\tilde{Q}(j)}-1\bigr] = 0.$$

Поэтому для любого $T \ge 0$ и любого $\tilde{h} \ge 0$, в силу (25) и утверждения 1) леммы 3, число

$$\mathcal{P}_N^{(2)} \le \exp\{-N(\tilde{h}T-\tilde{\mu}(\tilde{h}))\}, \tag{40}$$

где

$$\begin{aligned}\tilde{\mu}(\tilde{h}) &= \ln\Bigl[\textstyle\sum_{k,j} Q(k)\tilde{Q}(j)\exp\bigl\{\tilde{h}\ln\frac{P(j/k)}{\tilde{Q}(j)}\bigr\}\Bigr] = \\ &= \ln\Bigl[\textstyle\sum_{k,j} Q(k)\tilde{Q}(j)^{1-\tilde{h}}P(j/k)^{\tilde{h}}\Bigr], \quad \tilde{h}>0.\end{aligned} \tag{41}$$

Пусть $0 \le \tilde{h} \le 1$ и $-1 \le h \le 0$. Из (41) вытекает, что после замены переменной $\tilde{h} = 1+h$ можно написать

$$\tilde{h}T-\tilde{\mu}(\tilde{h}) = T+hT-\mu_{\mathbf{Q}}(h), \tag{42}$$

где $\mu_{\mathbf{Q}}(h)$, $-1 \le h \le 0$, определено (34). Подставив (42) в (40) и учитывая (38), получим, что для любого T , $0 < T < J(Q)$, и любого h , $-1 \le h \le 0$, одновременно выполняются два неравенства

$$\begin{aligned} \mathcal{P}_N^{(1)} &\le \exp\{-N(hT - \mu_{\mathbf{Q}}(h))\}, \\ \mathcal{P}_N^{(2)} &\le \exp\{-N(T + hT - \mu_{\mathbf{Q}}(h))\}. \end{aligned} \tag{43}$$

Из (43) вытекает утверждение 1) леммы 8.

2) Свойство функции $E_{\mathbf{Q}}(T)$, сформулированные в утверждении 2) леммы 8, являются очевидным следствием леммы 7.

Лемма 8 доказана.

Из леммы 2 (см. неравенство (23)), леммы 8 и неравенства $M - 1 \le \exp\{RN\}$ вытекает экспоненциальная верхняя граница вероятности ошибки порогового декодирования, даваемая теоремой 3.

Теорема 3. *Если* $0 < R < \mathbf{J}(\mathbf{Q},\mathbf{P})$, *то для ДКБП (1) существует* (N,R) *–код* X , *такой что при декодировании с порогом* T , $R \le T < \mathbf{J}(\mathbf{Q},\mathbf{P})$ *,для средней вероятности ошибки* $\mathcal{P}(X)$ *справедлива экспоненциально убывающая (с ростом* N *)верхняя граница*

$$\mathcal{P}(X) \le \exp\{-NE_{\mathbf{Q}}(T)\} + \exp\bigl\{-N\bigl[E_{\mathbf{Q}}(T) + T - R\bigr]\bigr\}.$$

Пусть C — пропускная способность ДКБП (1) и распределение $\mathbf{Q}$ совпадает с тем р.в., на котором достигается максимум в определении (11) пропускной способности C , т.е. $\mathbf{J}(\mathbf{Q},\mathbf{P}) = C$. Если выбрать порог $T = R < C$, то из теоремы 3 вытекает верхняя граница для $\mathcal{E}_N(\mathbf{P},R)$ — средней вероятности ошибки оптимального (наилучшего) (N,R)–кода, даваемая теоремой 4.

Теорема 4. *Для любого* R , $0 < R \le C$,

$$\mathcal{E}_N(\mathbf{P},R) \le 2\exp\{-NE_{\mathbf{Q}}(R)\},$$

где показатель экспоненты $E_{\mathbf{Q}}(R)$ *определяется соотношениями* (34) *и* (37) *при* $T = R$. *При* $0 < R < C$ *функция* $E_{\mathbf{Q}}(R) > 0$ *монотонно убывает, выпукла вниз и* $E_{\mathbf{Q}}(C) = 0$.

6.4 Выпуклость информационной функции

В этом разделе мы покажем, что определенная формулой (10) информационная функция

$$\mathbf{J}(\mathbf{Q},\mathbf{W}) = \sum_{x=0}^{K-1}\sum_{y=0}^{J-1} Q(x)\cdot W(y|x)\cdot \log\frac{W(y|x)}{\sum_{k=0}^{K-1} Q(k)\cdot W(y|k)} \tag{44}$$

является выпуклой *вверх* функцией аргумента $\mathbf{Q} = \{Q(x),\ \ x = \overline{0, K-1}\}$ и выпуклой *вниз* функцией аргумента $\mathbf{W} = \{W(y|x), x = \overline{0, K-1}, y = \overline{0, J-1}\}$. Выпуклость вверх имеет существенное значение для максимизации информационной функции по $\mathbf{Q}$ при вычислении пропускной способности, а выпуклость вниз — для минимизации информационной функции по $\mathbf{W}$ при вычислении ε–энтропии.

6.4.1 Выпуклость вверх по Q

Пусть $\mathbf{q} = (q(z),\ z = \overline{0,1})$ — произвольное р.в. на $\{0,1\}$. Зафиксируем два произвольных р.в. на $\{0, 1, \ldots, K-1\}$:

$$\mathbf{Q}_0 = (\mathbf{Q}(x|0), \quad x = \overline{0, K-1}),$$

$$\mathbf{Q}_1 = (\mathbf{Q}(x|1), \quad x = \overline{0, K-1})$$

и обозначим через

$$\mathbf{Q} = q(0)\mathbf{Q}_0 + q(1)\mathbf{Q}_1,$$

$$\mathbf{Q} = \big(Q(x), \quad x = \overline{0, K-1}\big),$$

$$Q(x) = \sum_{z=0}^{1} q(z)Q(x|z),$$

смесь этих распределений. Выпуклость вверх информационной функции по аргументу $\mathbf{Q}$ означает, что

$$\mathbf{J}(\mathbf{Q}, \mathbf{W}) \geq q(0){\cdot}\mathbf{J}(\mathbf{Q}_0, \mathbf{W}) + q(1){\cdot}\mathbf{J}(\mathbf{Q}_1, \mathbf{W}). \tag{45}$$

Для доказательства (45) рассмотрим тройку случайных величин (Z, X, Y) с р.в.

$$p(z, x, y) = q(z)Q(x|z)W(y|x), \tag{46}$$

которое определено на тройках (z, x, y), $z = \overline{0,1}$, $x = \overline{0, K-1}$, $y = \overline{0, J-1}$. Отметим следующие три очевидные свойства р.в. (46):

1. условное р.в. $p(y|x,z) = p(y|x) = W(y|x)$, т.е. тройка (Z, X, Y) — цепь Маркова;

2. маргинальное р.в. $p(x)$ совпадает с $Q(x)$, так как

$$p(x) = \sum_{z,y} p(z, x, y) = \sum_{z} q(z)Q(x|z) = Q(x);$$

3. при фиксированном $z = \overline{0,1}$ условное р.в. $p(x,y|z)$ является р.в. на парах (x,y) с маргинальным р.в.

$$p(y|z) = \sum_x p(x,y|z) = \sum_x Q(x|z)W(y|x).$$

Из 1) и 2) следует, что для пары случайных величин (X,Y) информация

$$J(X,Y) = \mathbf{J}(\mathbf{Q},\mathbf{W}). \tag{47}$$

Из 1) и 3) следует, что при фиксированном $z = \overline{0,1}$ для информации I_z р.в. $p(x,y|z)$ имеем

$$\begin{gathered} I_z = \sum_{x,y} p(x,y|z) \log \frac{P(y|xz)}{P(y|z)} = \\ = \sum_{x,y} Q(x|z)W(y|x) \log \frac{W(y|x)}{\sum_k Q(k|z)W(y|k)} = \mathbf{J}(\mathbf{Q}_z,\mathbf{W}). \end{gathered} \tag{48}$$

Для р.в. (46) условная информация

$$J(X;Y|Z) = \sum_{z=0}^{1} q(z) I_z.$$

Поэтому (48) означает, что для р.в. (46) условная информация

$$J(X;Y|Z) = \sum_{z=0}^{1} q(z)\mathbf{J}(\mathbf{Q}_z,\mathbf{W}). \tag{49}$$

Поскольку для цепи Маркова (Z,X,Y) условная информация $J(X;Y|Z) \le J(X;Y)$ (см. **§5**), то из (47) и (49) вытекает доказываемое неравенство (45).

6.4.2 Выпуклость вниз по W

Пусть $\mathbf{q} = (q(z),\ z = \overline{0,1})$ — произвольное р.в. на $(0,1)$. Зафиксируем два произвольных условных р.в. на парах (x,y), $x = \overline{0,K-1}$, $y = \overline{0,J-1}$:

$$\mathbf{W}_0 = (W_0(y|x), \quad x = \overline{0,K-1}, \quad y = \overline{0,J-1}),$$

$$\mathbf{W}_1 = (W_1(y|x), \quad x = \overline{0,K-1}, \quad y = \overline{0,J-1}),$$

и обозначим через

$$\mathbf{W} = q(0)\mathbf{W}_0 + q(1)\mathbf{W}_1,$$

$$\mathbf{W} = (W(y|x), \quad x = \overline{0, K-1}, \quad y = \overline{0, J-1}),$$

$$W(y|x) = \sum_{z=0}^{1} q(z) W_z(y|x),$$

смесь этих условных распределений. Выпуклость вниз информационной функции по аргументу $\mathbf{W}$ означает, что

$$\mathbf{J}(\mathbf{Q}, \mathbf{W}) \leq q(0) \cdot \mathbf{J}(\mathbf{Q}, \mathbf{W}_0) + q(1) \cdot \mathbf{J}(\mathbf{Q}, \mathbf{W}_1). \tag{50}$$

Для доказательства (50) рассмотрим тройку случайных величин (Z, X, Y) с р.в.

$$p(z, x, y) = q(z) \cdot Q(x) \cdot W_z(y|x), \tag{51}$$

которое определено на тройках (z, x, y), $z = \overline{0, 1}$, $x = \overline{0, K-1}$, $y = \overline{0, J-1}$. Отметим следующие три очевидные свойства р.в. (51):

1. р.в. $p(z, x) = p(z) \cdot P(x) = q(z) Q(x)$, т.е величины X и Z — независимы;

2. для пары (X, Y) распределение вероятностей

$$p(x, y) = \sum_z p(z, x, y) = \sum_z q(z) \cdot Q(x) W_z(y|x) = Q(x) \cdot W(y|x);$$

3. при фиксированном $z = \overline{0, 1}$ справедливы равенства

$$p(x, y|z) = Q(x) W_z(y|x), \quad p(x|z) = Q(x), \quad p(y|z) = \sum_x Q(x) W_z(y|x).$$

Из 2) следует, что для пары случайных величин (X, Y) информация

$$J(X; Y) = \mathbf{J}(\mathbf{Q}, \mathbf{W}). \tag{52}$$

Свойство 3) означает, что для р.в. (51) условная информация

$$\begin{gathered} J(X; Y|Z) = \textstyle\sum_{x,y,z} p(z, x, y) \log \frac{p(x,y|z)}{p(x|z) p(y|z)} = \\ = \textstyle\sum_z q(z) \sum_{x,y} Q(x) W_z(y|x) \log \frac{W(y|x)}{\sum_k Q(k) W_z(y|k)} = \\ = \textstyle\sum_z q(z) \mathbf{J}(\mathbf{Q}, \mathbf{W}_z). \end{gathered} \tag{53}$$

Поскольку для независимых величин X и Z условная информация $J(X; Y|Z) \geq J(X; Y)$ (см.**§5**), то из (52) и (53) вытекает доказываемое неравенство (50).