ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 55 **2019** **Вып. 3**

УДК 621.391.1 : 519.2

 В.В. Прелов

ОПТИМАЛЬНЫЕ ВЕРХНИЕ ГРАНИЦЫ ДЛЯ ДИВЕРГЕНЦИИ КОНЕЧНОМЕРНЫХ РАСПРЕДЕЛЕНИЙ ПРИ ЗАДАННОМ ВАРИАЦИОННОМ РАССТОЯНИИ[1]

Рассматривается задача о нахождении максимальных значений дивергенций $D(P\,\|\,Q)$ и $D(Q\,\|\,P)$ дискретных распределений вероятностей P и Q со значениями на конечном множестве $\mathcal{N} = \{1, 2, \ldots, n\}$ при условии, что задано вариационное расстояние $V(P, Q)$ между ними и заданы либо распределение вероятностей Q, либо (в случае $D(P\,\|\,Q)$) лишь значение минимальной компоненты $q_{\min}$ распределения Q. Получены точные выражения для указанных максимумов дивергенций, которые в ряде случаев позволяют выписать для них как явные формулы, так и простые верхние и нижние границы, причем для максимума $D(P\,\|\,Q)$ при заданных $V(P, Q)$ и $q_{\min}$, а также для максимума $D(Q\,\|\,P)$ при заданных Q и $V(P, Q)$ явные формулы получены для всех возможных значений этих параметров.

Ключевые слова: информационная дивергенция, вариационное расстояние, дискретные распределения вероятностей.

DOI: 10.1134/S0555292319030021

Пусть $P = \{p_i,\ i \in \mathcal{N}\}$ и $Q = \{q_i,\ i \in \mathcal{N}\}$ – два распределения вероятностей на конечном множестве $\mathcal{N} = \{1, 2, \ldots, n\}$. Напомним, что вариационное расстояние $V(P, Q)$ между распределениями P и Q определяется равенством

$$V(P, Q) = \sum_{i \in \mathcal{N}} |p_i - q_i|, \tag{1}$$

а информационная дивергенция (или просто дивергенция) $D(P\,\|\,Q)$ распределения P относительно Q – равенством

$$D(P\,\|\,Q) = \sum_{i \in \mathcal{N}} p_i \ln \frac{p_i}{q_i}, \tag{2}$$

где всегда предполагается, что $0 \ln(0/0) = 0$ и $a \ln(a/0) = \infty$, если $a > 0$, а $\ln(\cdot)$ – это натуральный логарифм.

Получению различных нижних границ для минимума дивергенции $D(P\,\|\,Q)$ по всем распределениям P и Q на $\mathcal{N}$ при условии, что задано лишь вариационное расстояние $V(P, Q)$, посвящено много работ (см., например, работу [1] и библиографию в ней). Среди подобных границ снизу для $D(P\,\|\,Q)$ отметим хорошо известные нера-

[1] Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований (номер проекта 19-01-00364).

венства Пинскера [2, задача 3.17]

$$D(P \,\|\, Q) \geqslant \frac{V^2(P,Q)}{2} \tag{3}$$

и Вайды [3]

$$D(P \,\|\, Q) \geqslant \ln \frac{2 + V(P,Q)}{2 - V(P,Q)} - \frac{2V(P,Q)}{2 + V(P,Q)}. \tag{4}$$

В работе [4] неравенство Пинскера (3) улучшается при условии, что задано не только вариационное расстояние $V(P,Q)$, но и распределение вероятностей Q, а в [5] при указанных условиях получено асимптотически точное значение для минимума $D(P \,\|\, Q)$ в случае, когда распределение Q стремится к вырожденному.

Если задано только вариационное расстояние $V(P,Q)$ между P и Q, то очевидно, что дивергенция $D(P \,\|\, Q)$ может принимать сколь угодно большие значения. Поэтому получить нетривиальные верхние границы для дивергенции $D(P \,\|\, Q)$ возможно лишь, если помимо вариационного расстояния $V(P,Q)$ имеются какие-либо другие ограничения на распределения P и Q. В [6, лемма 6.3] показано, что

$$D(P \,\|\, Q) \leqslant \sum_{i \in \mathcal{N}} \frac{(p_i - q_i)^2}{q_i},$$

откуда сразу вытекает следующая верхняя граница для максимума дивергенции $D(P \,\|\, Q)$ при условии, что $q_{\min} = \min\limits_{i \in \mathcal{N}} q_i > 0$

$$D(P \,\|\, Q) \leqslant \frac{V^2(P,Q)}{q_{\min}}. \tag{5}$$

Эта верхняя граница улучшена в работах [7, 8], где показано, что

$$D(P \,\|\, Q) \leqslant \ln\left(1 + \frac{V^2(P,Q)}{2q_{\min}}\right). \tag{6}$$

Авторы [7, 8] назвали верхнюю границу (6) “обратным (reverse) неравенством Пинскера”.

В настоящей статье мы улучшаем границу (6), показав, что для любых распределений $P = \{p_i,\ i \in \mathcal{N}\}$ и $Q = \{q_i,\ i \in \mathcal{N}\}$ на конечном множестве $\mathcal{N}$ с $|\mathcal{N}| \geqslant 3$ и $V(P,Q) = v$ справедливы неравенства

$$D(P \,\|\, Q) \leqslant \left(q_{\min} + \frac{v}{2}\right) \ln\left(1 + \frac{v}{2q_{\min}}\right), \quad \text{если} \quad \frac{v}{2} \geqslant q_{\min}, \tag{7}$$

а если $\dfrac{v}{2} \leqslant q_{\min}$, то

$$D(P \,\|\, Q) \leqslant \left(q_{\min} + \frac{v}{2}\right) \ln\left(1 + \frac{v}{2q_{\min}}\right) + \left(q_{\min} - \frac{v}{2}\right) \ln\left(1 - \frac{v}{2q_{\min}}\right). \tag{8}$$

При этом верхние границы (7), (8) не могут быть улучшены, так как они являются оптимальными, т.е. их правые части – максимальные значения дивергенции $D(P \,\|\, Q)$ при условии, что задано вариационное расстояние $V(P,Q) = v$ и величина минимальной компоненты $q_{\min}$ распределения Q. Это утверждение является следствием (точную формулировку которого см. ниже) доказываемой здесь теоремы 1, в которой получено точное выражение для максимума дивергенции $D(P \,\|\, Q)$ при условии, что заданы распределение Q и вариационное расстояние $V(P,Q)$.

Для формулировки теоремы 1 введем несколько необходимых определений. Обозначим через $\mathcal{N}_s$ подмножество $\mathcal{N}$, не содержащее элемент s, т.е. $\mathcal{N}_s = \mathcal{N} \setminus \{s\}$. Для заданного распределения вероятностей $Q = \{q_i,\ i \in \mathcal{N}\}$ и параметра v, $0 \leqslant v \leqslant 2(1 - q_{\min})$, всякое равенство

$$\frac{v}{2} = \sum_{i \in I_s} q_i + \beta, \quad \text{где} \quad I_s \subseteq \mathcal{N}_s, \tag{9}$$

назовем *допустимым I_s-представлением* $\frac{v}{2}$, если либо $\beta = 0$, либо существует индекс $j \in \mathcal{N}_s \setminus I_s$, такой что $0 < \beta < q_j$. Для всякого допустимого I_s-представления $\frac{v}{2}$ введем величину

$$L_s(I_s, v) = \Big(q_s + \frac{v}{2}\Big) \ln\Big(1 + \frac{v}{2q_s}\Big) + (q_k - \beta) \ln\Big(1 - \frac{\beta}{q_k}\Big), \tag{10}$$

где

$$q_k = \min_{j:\, j \in \mathcal{N}_s \setminus I_s,\ \beta < q_j} \{q_j\}.$$

При этом в случае, когда в представлении (9) $\beta = 0$ и $I_s = \mathcal{N}_s$, по определению считаем, что второе слагаемое в правой части (10) равно нулю.

Теорема 1. *Для любого распределения вероятностей $Q = \{q_i,\ i \in \mathcal{N}\}$, такого что $q_1 \geqslant q_2 \geqslant \ldots \geqslant q_n > 0$, и любого v, $0 \leqslant v \leqslant 2(1 - q_n)$, для величины*

$$\widehat{D}(Q, v) = \max_{P:\, V(P,Q) = v} D(P \,\|\, Q)$$

справедливы следующие утверждения:

- *Для любого v справедливо равенство*

$$\widehat{D}(Q, v) = \max_{s, I_s} L_s(I_s, v), \tag{11}$$

  где $L_s(I_s, v)$ определены в (10), а максимум берется по всем s и всем допустимым I_s-представлениям $\frac{v}{2}$, таким что $t \leqslant s \leqslant n$, где $t = \min\Big\{i :\ q_i + \frac{v}{2} \leqslant 1\Big\}$;
- *Если $0 \leqslant v/2 \leqslant q_n$, то*

$$\widehat{D}(Q, v) = \Big(q_{n-1} + \frac{v}{2}\Big) \ln\Big(1 + \frac{v}{2q_{n-1}}\Big) + \Big(q_n - \frac{v}{2}\Big) \ln\Big(1 - \frac{v}{2q_n}\Big); \tag{12}$$

- *Если $1 - q_{n-1} \leqslant v/2 \leqslant 1 - q_n$, то*

$$\widehat{D}(Q, v) = \Big(q_n + \frac{v}{2}\Big) \ln\Big(1 + \frac{v}{2q_n}\Big) + \Big(1 - q_n - \frac{v}{2}\Big) \ln \frac{1 - q_n - \frac{v}{2}}{q_{n-1}}; \tag{13}$$

- *Для любых v справедлива верхняя граница*

$$\widehat{D}(Q, v) \leqslant \Big(q_n + \frac{v}{2}\Big) \ln\Big(1 + \frac{v}{2q_n}\Big); \tag{14}$$

- *Если $q_n < v/2 < 1 - q_{n-1}$ и $\frac{v}{2} = \sum\limits_{i=1}^{n-1} a_i q_i + \beta$, где $a_i \in \{0, 1\}$, и либо $\beta = 0$, либо существует индекс j, такой что $a_j = 0$ и $0 < \beta < q_j$, то справедлива нижняя граница*

$$\widehat{D}(Q, v) \geqslant \Big(q_n + \frac{v}{2}\Big) \ln\Big(1 + \frac{v}{2q_n}\Big) + (q_j - \beta) \ln\Big(1 - \frac{\beta}{q_j}\Big); \tag{15}$$

- *Верхняя и нижняя границы в* (14) *и* (15) *совпадают, так что*

$$\widehat{D}(Q,v) = \Bigl(q_n + \frac{v}{2}\Bigr)\ln\Bigl(1 + \frac{v}{2q_n}\Bigr), \tag{16}$$

если $\dfrac{v}{2} = \sum\limits_{i=1}^{n-1} a_i q_i$ *при любых* $a_i \in \{0,1\}$.

Доказательство этой теоремы приведено в Приложении.

С л е д с т в и е. *Пусть* $P = \{p_i, i \in \mathcal{N}\}$ *и* $Q = \{q_i, i \in \mathcal{N}\}$ *– распределения вероятностей на одном и том же конечном множестве* $\mathcal{N}$ *с* $|\mathcal{N}| \geqslant 3$, *и пусть*

$$q_{\min} = \min_{i\in\mathcal{N}} q_i > 0.$$

Тогда для величины

$$D^*(q_{\min}, v) = \max_{(P,Q)\colon V(P,Q)=v,\ \min\limits_{i\in\mathcal{N}} q_i = q_{\min}} D(P\,\|\,Q)$$

при любом v, $0 \leqslant v \leqslant 2(1-q_{\min})$, *справедливы следующие равенства*:

- *Если* $\dfrac{v}{2} \geqslant q_{\min}$, *то*

$$D^*(q_{\min}, v) = \Bigl(q_{\min} + \frac{v}{2}\Bigr)\ln\Bigl(1 + \frac{v}{2q_{\min}}\Bigr); \tag{17}$$

- *Если* $\dfrac{v}{2} \leqslant q_{\min}$, *то*

$$D^*(q_{\min}, v) = \Bigl(q_{\min} + \frac{v}{2}\Bigr)\ln\Bigl(1 + \frac{v}{2q_{\min}}\Bigr) + \Bigl(q_{\min} - \frac{v}{2}\Bigr)\ln\Bigl(1 - \frac{v}{2q_{\min}}\Bigr). \tag{18}$$

Действительно, очевидно, что

$$D^*(q_{\min}, v) = \max_{Q\colon \min\limits_{i\in\mathcal{N}} q_i = q_{\min}} \widehat{D}(Q,v) \leqslant \Bigl(q_{\min} + \frac{v}{2}\Bigr)\ln\Bigl(1 + \frac{v}{2q_{\min}}\Bigr), \tag{19}$$

так как из (10) и (11) следует, что для любых $Q = \{q_i, i \in \mathcal{N}\}$ с $\min\limits_{i\in\mathcal{N}} q_i = q_{\min}$ и любых v, $0 \leqslant v \leqslant 2(1-q_{\min})$, справедливо неравенство

$$\widehat{D}(Q,v) \leqslant \Bigl(q_{\min} + \frac{v}{2}\Bigr)\ln\Bigl(1 + \frac{v}{2q_{\min}}\Bigr).$$

С другой стороны, если $v/2 \geqslant q_{\min}$, то существует распределение $Q' = \{q'_i,\ i \in \mathcal{N}\}$, $q'_1 \geqslant q'_2 \geqslant \ldots \geqslant q'_n > 0$, для которого $q'_n = q_{\min}$ и $\dfrac{v}{2} = \sum\limits_{i=1}^{n-1} a_i q'_i$ при некоторых $a_i \in \{0,1\}$, а тогда из (15) следует, что

$$D^*(q_{\min}, v) \geqslant \widehat{D}(Q',v) \geqslant \Bigl(q_{\min} + \frac{v}{2}\Bigr)\ln\Bigl(1 + \frac{v}{2q_{\min}}\Bigr). \tag{20}$$

Из (19) и (20) теперь следует (17).

Если же $v/2 \leqslant q_{\min}$, то с одной стороны, из (12) следует, что для любого распределения $Q = \{q_i,\ i \in \mathcal{N}\}$, $q_1 \geqslant q_2 \geqslant \ldots \geqslant q_n > 0$, для которого $q_n = q_{\min}$, справедливо неравенство

$$D^*(q_{\min}, v) = \max_{Q\colon \min\limits_{i\in\mathcal{N}} q_i = q_{\min}} \widehat{D}(Q,v) \leqslant$$
$$\leqslant \Bigl(q_{\min} + \frac{v}{2}\Bigr)\ln\Bigl(1 + \frac{v}{2q_{\min}}\Bigr) + \Bigl(q_{\min} - \frac{v}{2}\Bigr)\ln\Bigl(1 - \frac{v}{2q_{\min}}\Bigr),$$

так как по предположению $q_{n-1} \geqslant q_n$, а функция $\left(x + \frac{v}{2}\right) \ln\left(1 + \frac{v}{2x}\right)$ убывает по x. С другой стороны, в случае, когда $|\mathcal{N}| \geqslant 3$, существует распределение $Q'' = \{q''_i,\ i \in \mathcal{N}\}$, $q''_1 \geqslant q''_2 \geqslant \ldots \geqslant q''_n > 0$, для которого $q''_{n-1} = q''_n = q_{\min}$, а тогда из (12) следует, что

$$D^*(q_{\min}, v) \geqslant \widehat{D}(Q'', v) = \left(q_{\min} + \frac{v}{2}\right) \ln\left(1 + \frac{v}{2q_{\min}}\right) + \left(q_{\min} - \frac{v}{2}\right) \ln\left(1 - \frac{v}{2q_{\min}}\right),$$

что и доказывает равенство (18).

В теореме 1 было приведено точное (а в некоторых случаях и явное) значение для максимума дивергенции $D(P \,\|\, Q)$ при условии, что заданы распределение Q и вариационное расстояние $V(P,Q)$, а в следующей теореме приводится явное выражение для максимума дивергенции $D(Q \,\|\, P)$ при тех же условиях.

Теорема 2. *Для любого распределения вероятностей $Q = \{q_i,\ i \in \mathcal{N}\}$, такого что $q_1 \geqslant q_2 \geqslant \ldots \geqslant q_n > 0$, и любого v, $0 \leqslant v \leqslant 2(1 - q_n)$, для величины $\widetilde{D}(Q, v) = \max\limits_{P\colon V(P,Q)=v} D(Q \,\|\, P)$ справедливо равенство*

$$\widetilde{D}(Q, v) = \begin{cases} \infty, & \text{если } \dfrac{v}{2} \geqslant q_n, \\ q_n \ln\left(1 + \dfrac{v}{2q_n - v}\right) + q_{n-1} \ln\left(1 - \dfrac{v}{2q_{n-1} - v}\right), & \text{если } \dfrac{v}{2} < q_n. \end{cases} \tag{21}$$

Эта теорема также доказывается в Приложении.

ПРИЛОЖЕНИЕ

Доказательство теоремы 1 основано на следующем утверждении.

Лемма. *Существует оптимальное распределение $P^* = \{p^*_i,\ i \in \mathcal{N}\}$ (т.е. распределение, на котором достигается значение $\widehat{D}(Q, v) = \max\limits_{P\colon V(P,Q)=v} D(P \,\|\, Q)$), для которого $p^*_s = q_s + v/2$ для некоторого $s \in \mathcal{N}$, а все остальные $p^*_i \leqslant q_i$, $i \in \mathcal{N}_s$. При этом все эти p^*_i равны либо нулю, либо соответствующим q_i, за исключением, возможно, лишь одного, скажем, p^*_j, $j \in \mathcal{N}_s$, такого что $0 < p^*_j < q_j$.*

Доказательство. Пусть $P = \{p_i,\ i \in \mathcal{N}\}$ – произвольное распределение вероятностей, для которого $V(P,Q) = v$, и пусть $A = A(P,Q) = \{i \in \mathcal{N} :\ p_i > q_i\}$. Покажем, что в случае, когда A содержит по крайней мере два различных элемента, скажем, k и j, существует другое распределение вероятностей $\widehat{P} = \{\widehat{p}_i,\ i \in \mathcal{N}\}$, для которого $V(\widehat{P}, Q) = v$ и $D(\widehat{P} \,\|\, Q) > D(P \,\|\, Q)$. Пусть для определенности $q_k \geqslant q_j$. Тогда таким распределением $\widehat{P} = \{\widehat{p}_i, i \in \mathcal{N}\}$ служит распределение с компонентами

$$\widehat{p}_i = \begin{cases} p_i & \text{при} \quad i \in \mathcal{N} \setminus \{k, j\}, \\ q_k & \text{при} \quad i = k, \\ p_j + p_k - q_k & \text{при} \quad i = j. \end{cases}$$

Действительно, очевидно, что $V(\widehat{P}, Q) = V(P, Q)$ и

$$D(P \,\|\, Q) - D(\widehat{P} \,\|\, Q) = p_k \ln \frac{p_k}{q_k} + p_j \ln \frac{p_j}{q_j} - (p_j + p_k - q_k) \ln \frac{p_j + p_k - q_k}{q_j}.$$

Поэтому для доказательства того, что $D(\widehat{P} \,\|\, Q) > D(P \,\|\, Q)$, достаточно воспользоваться неравенством

$$p_k \ln \frac{p_k}{q_k} + p_j \ln \frac{p_j}{q_j} < (p_j + p_k - q_k) \ln \frac{p_j + p_k - q_k}{q_j},$$

справедливость которого немедленно следует их того, что функция

$$f(x) = (q_k + x) \ln \frac{p_k + x}{p_k} + p_j \ln \frac{p_j}{q_j} - (p_j + x) \ln \frac{p_j + x}{q_j}$$

строго убывает по x при $x \geqslant 0$, а при $x = 0$ она равна нулю. Действительно,

$$f'(x) = \ln \frac{q_k q_j + x q_j}{q_k p_j + x q_k} < 0,$$

так как $q_j < p_j$, а $q_j \leqslant q_k$ по условию. Поэтому существует оптимальное распределение $P^* = \{p_i^*,\ i \in \mathcal{N}\}$, для которого множество $A(P^*, Q) = \{i \in \mathcal{N} :\ p_i^* > q_i\}$ содержит лишь один элемент, скажем, s, такой что $p_s^* = q_s + v/2$, а все остальные $p_i^* \leqslant q_i$, $i \in \mathcal{N}_s$.

Покажем теперь, что существует оптимальное распределение $P^* = \{p_i^*,\ i \in \mathcal{N}\}$, которое наряду с первым удовлетворяет и второму утверждению леммы. Для этого достаточно воспользоваться следующим утверждением: пусть $P = \{p_i,\ i \in \mathcal{N}\}$ – некоторое распределение вероятностей, для которого $V(P, Q) = v$, $p_s = q_s + v/2$, $p_i \leqslant q_i$ для всех $i \in \mathcal{N}_s$ и существует два различных индекса k и j из $\mathcal{N}_s$, такие что $0 < p_k < q_k$ и $0 < p_j < q_j$. Тогда существует распределение вероятностей $\widehat{P} = \{\widehat{p}_i,\ i \in \mathcal{N}\}$, для которого:

1) $\widehat{p}_i = p_i$ при всех $i \in \mathcal{N} \setminus \{k, j\}$;

2) $\widehat{p}_k \in \{0, q_k\}$ или $\widehat{p}_j \in \{0, q_j\}$;

3) $V(\widehat{P}, Q) = v$;

4) $D(\widehat{P} \,\|\, Q) \geqslant D(P \,\|\, Q)$.

Действительно, рассмотрим распределения $P(x) = \{p_i(x),\ i \in \mathcal{N}\}$ с компонентами

$$p_i(x) = \begin{cases} p_i & \text{при } i \in \mathcal{N} \setminus \{k, j\}, \\ p_k - x & \text{при } i = k, \\ p_j + x & \text{при } i = j, \end{cases}$$

где $x \in [\max\{-p_j, p_k - q_k\}, \min\{p_k, q_j - p_j\}]$. Тогда очевидно, что для всех таких распределений $P(x)$ имеет место равенство $V(P(x), Q) = v$, а максимум $D(P(x) \,\|\, Q)$, больший или равный $D(P(0) \,\|\, Q) = D(P \,\|\, Q)$, достигается на одном из концов указанного выше отрезка значений параметра x, поскольку функция

$$(p_k - x) \ln \frac{p_k - x}{q_k} + (p_j + x) \ln \frac{p_j + x}{q_j}$$

является выпуклой по x, откуда и следует сформулированное выше утверждение. ▲

Из этой леммы первое утверждение теоремы 1, т.е. формула (11), немедленно следует, если заметить, что функция $(x - \beta) \ln\left(1 - \dfrac{\beta}{x}\right)$ убывает по x.

Справедливость формулы (12) также легко следует из утверждения леммы. Действительно, если $0 \leqslant v/2 < q_n$, то в I_s-допустимом представлении $v/2$ величина $I_s = \varnothing$, а $\beta = v/2$. Поэтому очевидно, что

$$\max_{s, I_s} L_s(I_s, v) = \max\{L_n(\varnothing, v), L_{n-1}(\varnothing, v)\}.$$

Замечая, что разность $L_n(\varnothing, v) - L_{n-1}(\varnothing, v)$, как легко проверить, убывает по v, а при $v = 0$ она равна нулю, получаем, что $\max\limits_{s, I_s} L_s(I_s, v) = L_{n-1}(\varnothing, v)$, а это равносильно формуле (12). Легко видеть, что формула (12) справедлива и в случае, когда $v/2 = q_n$.

Для доказательства (13) следует заметить, что в случае, когда $1 - q_{n-1} < v/2 \leqslant \leqslant 1 - q_n$, имеем $t = \min\{i: \; q_i + v/2 \leqslant 1\} = n$, и поэтому

$$\widehat{D}(Q, v) = \max_{I_n} L_n(I_n, v) = \left(q_n + \frac{v}{2}\right) \ln\left(1 + \frac{v}{2q_n}\right) +$$

$$+ \max_{k:\, k \neq n} \left\{ \left(1 - q_n - \frac{v}{2}\right) \ln \frac{1 - q_n - \frac{v}{2}}{q_k} \right\} =$$

$$= \left(q_n + \frac{v}{2}\right) \ln\left(1 + \frac{v}{2q_n}\right) + \left(1 - q_n - \frac{v}{2}\right) \ln \frac{1 - q_n - \frac{v}{2}}{q_{n-1}}.$$

Если же $1 - q_{n-1} = v/2$, то справедливость равенства (13) также легко проверяется непосредственно.

Верхняя граница (14) немедленно следует из (10) и 11), а нижняя граница (15) также вытекает из (10) и (11), если заметить, что $\widehat{D}(Q, v) \geqslant L_n(I_n, v)$, где I_n – множество индексов суммирования i в сумме $v/2 = \sum\limits_{i=1}^{n-1} a_i q_i + \beta$, для которых $a_i \neq 0$. Наконец, равенство (16) сразу следует из (14) и (15), так как в этом случае в нижней границе (15) величина β равна нулю. ▲

Доказательство теоремы 2. Пусть заданы распределение вероятностей $Q = \{q_i, i \in \mathcal{N}\}$, $q_1 \geqslant q_2 \geqslant \ldots \geqslant q_n > 0$, и параметр v, $0 \leqslant v \leqslant 2(1 - q_n)$. Прежде всего заметим, что первое равенство в (21) почти очевидно, так как в этом случае легко построить распределение вероятностей $P = \{p_i, i \in \mathcal{N}\}$, такое что $V(P, Q) = v$, а одна из компонент $p_i = 0$. Действительно, если v удовлетворяет условию $2q_n \leqslant \leqslant v \leqslant 2(1 - q_{n-1})$, то полагаем

$$P = \{q_1 - \varepsilon_1, q_2 - \varepsilon_2, \ldots, q_{n-2} - \varepsilon_{n-2}, q_{n-1} + v/2, 0\},$$

где ε_i, $i = 1, 2, \ldots, n-2$, таковы что $0 \leqslant \varepsilon_i \leqslant q_i$ и $\sum\limits_{i=1}^{n-2} \varepsilon_i = v/2 - q_n$. Такие ε_i существуют, так как $v/2 - q_n \leqslant \sum\limits_{i=2}^{n-2} q_i$, поскольку по условию $v \leqslant 2(1 - q_{n-1})$. Если же v удовлетворяет условию $2(1 - q_{n-1}) \leqslant v \leqslant 2(1 - q_n)$, то полагаем

$$P = \{q_1 - \varepsilon_1, q_2 - \varepsilon_2, \ldots, q_{n-2} - \varepsilon_{n-2}, 0, q_n + v/2\},$$

где ε_i, $i = 1, 2, \ldots, n-2$, таковы что $0 \leqslant \varepsilon_i \leqslant q_i$ и $\sum\limits_{i=1}^{n-2} \varepsilon_i = v/2 - q_{n-1}$. Такие ε_i в данном случае тоже существуют, так как $v/2 - q_{n-1} \leqslant \sum\limits_{i=2}^{n-2} q_i$. Таким образом, первая из формул в (21) доказана.

Докажем теперь вторую формулу в (21), когда предполагается, что $0 \leqslant v/2 < q_n$. В этом случае мы вначале хотим показать, что максимум $D(Q \,\|\, P)$ достигается на одном из двух распределений

$$P_1 = \{q_1, q_2, \ldots, q_{n-2}, q_{n-1} - v/2, q_n + v/2\}$$

или

$$P_2 = \{q_1, q_2, \ldots, q_{n-2}, q_{n-1} + v/2, q_n - v/2\},$$

а затем выберем из них *оптимальное*, т.е. то, на котором достигается значение $\widetilde{D}(Q, v)$.

Пусть $P = \{p_i,\ i \in \mathcal{N}\}$ – некоторое распределение вероятностей, такое что $V(P, Q) = v$. Положим

$$
\begin{aligned}
&A = A(P, Q) = \{i \in \mathcal{N} :\ p_i > q_i\},\\
&B = B(P, Q) = \{i \in \mathcal{N} :\ p_i < q_i\}.
\end{aligned}
$$

Тогда, как и при доказательстве теоремы 1, нетрудно показать, что для оптимального распределения P множества $A(P, Q)$ и $B(P, Q)$ содержат лишь по одному элементу, причем либо $A(P, Q) = \{n\}$ и $B(P, Q) = \{n - 1\}$, либо $A(P, Q) = \{n - 1\}$ и $B(P, Q) = \{n\}$. Действительно, если $A(P, Q)$ содержит по крайней мере два элемента j и k, $j \neq k$, то P не может быть оптимальным распределением. Это утверждение немедленно следует из неравенства

$$
q_j \ln \frac{q_j}{p_j} + q_k \ln \frac{q_k}{p_k} < q_k \ln \frac{q_k}{p_k - (q_j - p_j)}, \quad \text{если} \quad q_k \leqslant q_j,
$$

которое, в свою очередь, является следствием того, что функция

$$
f(x) = q_j \ln \frac{q_j}{q_j - x} + q_k \ln \frac{q_k}{p_k} - q_k \ln \frac{q_k}{p_k - x}
$$

строго убывает по x, $0 < x < v/2$.

Аналогично доказывается, что множество $B(P, Q)$ для оптимального распределения P тоже содержит лишь один элемент. Наконец, утверждение о том, что оптимальным распределением является P_1 или P_2, теперь следует из того, что

$$
q_k \ln \frac{q_k}{q_k - v/2} \quad \text{и} \quad q_k \ln \frac{q_k}{q_k + v/2}
$$

являются убывающими функциями q_k.

Для доказательства второго равенства в (21) нам осталось лишь показать, что на самом деле оптимальным является распределение P_2, т.е. что $D(Q \,\|\, P_2) \geqslant D(Q \,\|\, P_1)$. А это неравенство следует из того, что разность

$$
D(Q \,\|\, P_2) - D(Q \,\|\, P_1)
$$

является возрастающей функцией v, $v \in [0, 2q_n)$, так как легко проверить, что

$$
[D(Q \,\|\, P_2) - D(Q \,\|\, P_1)]'_{v/2} = \left(v^2/2\right) \left[(q_{n-1}^2 - v^2/4)(q_n^2 - v^2/4\right]^{-1} (q_{n-1}^2 - q_n^2) \geqslant 0,
$$

а при $v = 0$ эта разность равна нулю. ▲

СПИСОК ЛИТЕРАТУРЫ

1. *Fedotov A.A., Harremöes P., Topsøe F.* Refinements of Pinsker's Inequality // IEEE Trans. Inform. Theory. 2003. V. 49. № 6. P. 1491–1498.
2. *Чисар И., Кёрнер Я.* Теория информации: Теория кодирования для дискретных систем без памяти. М.: Мир, 1985.
3. *Vajda I.* Note on Discrimination Information and Variation // IEEE Trans. Inform. Theory. 1970. V. 16. № 6. P. 771–773.
4. *Ordentlich E., Weinberger M.J.* A Distribution Dependent Refinement of Pinsker's Inequality // IEEE Trans. Inform. Theory. 2005. V. 51. № 5. P. 1836–1840.
5. *Прелов В.В.* О склеивании вероятностных расределений и оценивании дивергенции через вариацию // Пробл. передачи информ. 2017. Т. 53. № 3. С. 16–22.

6. *Csiszár I., Talata Z.* Context Tree Estimation for Not Necessarily Finite Memory Processes via BIC and MDL // IEEE Trans. Inform. Theory. 2006. V. 52. № 3. P. 1007–1016.
7. *Sason I., Verdú S.* Upper Bounds on the Relative Entropy and Rényi Divergence as a Function of Total Variation Distance for Finite Alphabets // Proc. 2015 IEEE Information Theory Workshop (ITW'2015). Jeju, Korea. October 11–15, 2015. P. 214–218.
8. *Sason I., Verdú S.* f-Divergence Inequalities // IEEE Trans. Inform. Theory. 2016. V. 62. № 11. P. 5973–6006.

Прелов Вячеслав Валерьевич
Институт проблем передачи информации
им. А.А. Харкевича РАН
prelov@iitp.ru

Поступила в редакцию
21.05.2019
После доработки
03.07.2019
Принята к публикации
05.07.2019