ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 57 **2021** **Вып. 4**

УДК 621.391 : 519.72

 В.В. Прелов

О МАКСИМУМЕ f-ДИВЕРГЕНЦИИ ВЕРОЯТНОСТНЫХ РАСПРЕДЕЛЕНИЙ ПРИ ЗАДАННОЙ ВЕЛИЧИНЕ ИХ СКЛЕИВАНИЯ[1]

Статья является дополнением к работе автора [1]. Здесь приводятся явные верхние границы, являющиеся в некоторых случаях оптимальными, для максимального значения f-дивергенции $D_f(P\,\|\,Q)$ дискретных распределений вероятностей P и Q при условии, что заданы распределение Q (или его минимальная компонента $q_{\min}$) и величина склеивания P и Q. Получено также явное выражение для максимума дивергенции $D_f(P\,\|\,Q)$ при условии, что задана только величина склеивания распределений P и Q. Результаты [1], относящиеся к дивергенции Кульбака – Лейблера и χ^2-дивергенции, являются частными случаями утверждений, доказанных в данной статье.

Ключевые слова: f-дивергенция, дивергенция Кульбака – Лейблера, χ^2-дивергенция, склеивание дискретных распределений вероятностей.

DOI: 10.31857/S0555292321040021

В данной статье используются обозначения, принятые в [1], однако для удобства читателя мы напомним некоторые из них, а также введем новые, необходимые нам в дальнейшем. Нас будут интересовать величины $D_f^{\max}(Q,\alpha)$, $D_f^{\max}(q_{\min},\alpha)$ и $D_f^{\max}(\alpha)$, определяемые равенствами

$$D_f^{\max}(Q,\alpha)=\max_{P:\,s(P,Q)=\alpha} D_f(P\,\|\,Q), \tag{1}$$

$$D_f^{\max}(q_{\min},\alpha)=\max_{(P,Q):\,s(P,Q)=\alpha,\ \min\limits_{i\in\mathcal{N}} q_i=q_{\min}} D_f(P\,\|\,Q), \tag{2}$$

$$D_f^{\max}(\alpha)=\sup_{(P,Q):\,s(P,Q)=\alpha} D_f(P\,\|\,Q), \tag{3}$$

где $D_f(P\,\|\,Q)=\sum\limits_{i\in\mathcal{N}} q_i f\left(\dfrac{p_i}{q_i}\right)$ – f-дивергенция (см., например, [2]) дискретного распределения вероятностей $P=\{p_i,\ i\in\mathcal{N}\}$ относительно другого дискретного распределения $Q=\{q_i,\ i\in\mathcal{N}\}$ с конечным множеством значений $\mathcal{N}=\{1,2,\ldots,n\}$, $q_{\min}=\min\limits_{i\in\mathcal{N}} q_i$ – минимальная компонента распределения $Q=\{q_i,\ i\in\mathcal{N}\}$, а $s(P,Q)$ – величина склеивания распределений P и Q (напомним, что $s(P,Q)=\Pr\{X=Y\}$, где X и Y – дискретные случайные величины, имеющие распределения P и Q соответственно (см. [1])). Заметим сразу, что максимумы правых частей в определениях (1) и (2), как будет показано ниже, достигаются. Заметим также, что здесь, как и в [1], всегда предполагается, что функция $f\colon(0,\infty)\to\mathbb{R}$, определяющая f-дивергенцию $D_f(P\,\|\,Q)$, является выпуклой дважды дифференцируемой функцией,

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

такой что $f''(x) > 0$, $x \neq 1$ и $f(1) = 0$. При этом всегда по определению предполагается, что $0 \cdot f\left(\dfrac{0}{0}\right) = 0$, $f(0) = \lim\limits_{x \downarrow 0} f(x)$ и $0 \cdot f\left(\dfrac{a}{0}\right) = \lim\limits_{x \downarrow 0} x f\left(\dfrac{a}{x}\right) = a \lim\limits_{t \to \infty} \dfrac{f(t)}{t}$, где $a \neq 0$.

Напомним также, что при заданном распределении вероятностей $Q = \{q_i, i \in \mathcal{N}\}$ всякое равенство

$$\alpha = \sum_{i \in I} q_i + \beta, \quad \text{где} \quad I \subseteq \mathcal{N}, \tag{4}$$

называется *допустимым* (Q, I)*-представлением* α, если либо $\beta = 0$, либо существует индекс $j \in \mathcal{N} \setminus I$, такой что $0 < \beta < q_j$.

Всякое α-склеивание заданного распределения вероятностей $Q = \{q_i,\ i \in \mathcal{N}\}$ с некоторым распределением $P = \{p_i,\ i \in \mathcal{N}\}$ задается с помощью квадратной матрицы $M = \|p_{ij}\|_{i,j=1}^{n}$ с неотрицательными элементами p_{ij}, такой что $\sum\limits_{j=1}^{n} p_{ij} = q_i$ для всех $i \in \mathcal{N}$, $\sum\limits_{i=1}^{n} p_{ij} = p_j$ для всех $j \in \mathcal{N}$ и $\sum\limits_{i=1}^{n} p_{ii} = \alpha$. В этом случае полагаем $D_f(M) = D_f(P \,\|\, Q)$.

Всякому допустимому (Q, I)-представлению α сопоставляется множество $\mathcal{M}(Q, I)$ матриц $M = \|p_{ij}\|_{i,j=1}^{n}$, осуществляющих α-склеивание заданного распределения вероятностей $Q = \{q_i,\ i \in \mathcal{N}\}$ с некоторым распределением $P = \{p_i,\ i \in \mathcal{N}\}$ и обладающих следующим свойством: на (главной) диагонали каждой такой матрицы стоят числа q_i и β, входящие в данное (Q, I)-представление α, а все остальные ненулевые элементы матрицы находятся в некотором столбце (называемым *главным*) и, возможно, лишь один ненулевой элемент находится вне диагонали и этого главного столбца. В [1] была, в частности, доказана следующая

Теорема. *Для любого распределения вероятностей* $Q = \{q_i,\ i \in \mathcal{N}\}$, $q_1 \geqslant q_2 \geqslant \ldots \geqslant q_n > 0$, *и любого* α, $0 \leqslant \alpha \leqslant 1$, *справедливо равенство*

$$D_f^{\max}(Q, \alpha) = \max_{(Q,I)} \max_{M \in \mathcal{M}(Q,I)} D_f(M), \tag{5}$$

где первый максимум в правой части (5) *берется по всем допустимым* (Q, I)*-представлениям* α. *При этом*

$$D_f^{\max}(Q, \alpha) = \infty, \quad \textit{если } \alpha \leqslant 1 - q_n \textit{ и } f(0) = \infty, \tag{6}$$

и если $\alpha > 1 - q_n$, *то при любом значении* $f(0)$ *справедливо равенство*

$$D_f^{\max}(Q, \alpha) = \max\left\{ q_n f\left(\frac{\alpha - 1 + q_n}{q_n}\right) + q_{n-1} f\left(\frac{1 - \alpha + q_{n-1}}{q_{n-1}}\right), \right.$$
$$\left. q_{n-1} f\left(\frac{\alpha - 1 + q_{n-1}}{q_{n-1}}\right) + q_n f\left(\frac{1 - \alpha + q_n}{q_n}\right) \right\}. \tag{7}$$

В настоящей статье мы дополняем эту теорему нижеследующими предложением и следствием из него для величины $D_f^{\max}(q_{\min}, \alpha)$, определенной в (2).

Предложение 1. *Для любого распределения вероятностей* $Q = \{q_i,\ i \in \mathcal{N}\}$, $q_1 \geqslant q_2 \geqslant \ldots \geqslant q_n > 0$, *для* f*-дивергенции* $D_f^{\max}(Q, \alpha)$ *в случае, когда* $f(0) < \infty$ *и* $f'(0) < \infty$, *справедливы следующие утверждения*:

- *Если* $0 \leqslant \alpha \leqslant q_n$, *то*

$$D_f^{\max}(Q, \alpha) = \max\{A_\alpha(q_n, q_{n-1}), A_\alpha(q_{n-1}, q_n), B_\alpha(q_{n-1}, q_n)\}, \tag{8}$$

где

$$A_\alpha(x,y) = xf\left(\frac{1+\alpha-x}{x}\right) + yf\left(\frac{x-\alpha}{y}\right) + (1-x-y)f(0), \tag{9}$$

$$B_\alpha(x,y) = xf\left(\frac{1-\alpha-x}{x}\right) + yf\left(\frac{x+\alpha}{y}\right) + (1-x-y)f(0); \tag{10}$$

- *Если* $q_n \leqslant \alpha \leqslant 1-q_n$, *то*

$$D_f^{\max}(Q,\alpha) \leqslant q_n f\left(\frac{1-\alpha+q_n}{q_n}\right) + (1-\alpha)f(0), \tag{11}$$

причем эта верхняя граница достигается (т.е. в (11) *имеет место знак равенства), если* $\alpha = q_n + \sum\limits_{i=1}^{n-1} a_i q_i$ *при некоторых* $a_i \in \{0,1\}$.

С л е д с т в и е 1. *Для величины* $D_f^{\max}(q_{\min},\alpha)$, *определенной в* (2), *в случае, когда* $f(0)<\infty$, $f'(0)<\infty$ *и* $|\mathcal{N}| = n \geqslant 3$, *справедливы следующие утверждения*:

- *Для всех* $q_{\min} > 0$ *и* α, $0 \leqslant \alpha \leqslant q_{\min}$, *справедливо равенство*

$$D_f^{\max}(q_{\min},\alpha) = A_\alpha(q_{\min},q_{\min}), \tag{12}$$

где $A_\alpha(x,y)$ *определено в* (9);

- *Для всех* $q_{\min} > 0$ *и* α, $1-q_{\min} \leqslant \alpha \leqslant 1$, *справедливо равенство*

$$D_f^{\max}(q_{\min},\alpha) = q_{\min}\left[f\left(\frac{\alpha-1+q_{\min}}{q_{\min}}\right) + f\left(\frac{1-\alpha+q_{\min}}{q_{\min}}\right)\right]; \tag{13}$$

- *Для всех* $q_{\min} > 0$ *и* α, $q_{\min} \leqslant \alpha \leqslant 1-q_{\min}$, *справедлива верхняя граница*

$$D_f^{\max}(q_{\min},\alpha) \leqslant q_{\min} f\left(\frac{1-\alpha+q_{\min}}{q_{\min}}\right) + (1-\alpha)f(0), \tag{14}$$

причем эта верхняя граница достигается, т.е. в (14) *имеет место равенство, если* $2q_{\min} \leqslant \alpha < 1-q_{\min}$ *и* $q_{\min} \leqslant \dfrac{1}{n+1}$, *а также если* $q_{\min} \leqslant \dfrac{1}{n}$ *и* $\alpha = kq_{\min}$, *где* k – *любое целое, такое что* $2 \leqslant k \leqslant n-1$.

П р е д л о ж е н и е 2. *Для величины* $D_f^{\max}(\alpha)$, *определенной в* (3), *и любого* α, $0 \leqslant \alpha < 1$, *справедливо равенство*

$$D_f^{\max}(\alpha) = (1-\alpha)[f(0) + f^*(0)], \tag{15}$$

где $f^*(x) = xf\left(\dfrac{1}{x}\right)$ *и* $f^*(0) = \lim\limits_{x\downarrow 0} f^*(x) = \lim\limits_{t\to\infty} \dfrac{f(t)}{t}$.

Замечание. Отметим, что дивергенция Кульбака–Лейблера

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

(известная также под многими другими названиями, например, *информационная дивергенция*, *относительная энтропия*, *энтропия меры по мере*, *информация для различения* или просто *дивергенция*) и χ^2-дивергенция

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

являются важнейшими частными случаями f-дивергенции, когда $f(t) = t\log t$ и $f(t) = (t-1)^2$ соответственно (см., например, [3, 4]). Результаты в [1], относящиеся к дивергенции Кульбака – Лейблера и χ^2-дивергенции, являются частными случаями результатов, полученных в приведенных выше предложении 1 и следствии 1. Отметим, например, что для дивергенции Кульбака – Лейблера и χ^2-дивергенции в формуле (8) максимальной является величина $A(q_n, q_{n-1})$, откуда и вытекают соответствующие формулы в [1]. В общем же случае для f-дивергенции не удается выяснить, какая из трех величин в (8) является максимальной, поскольку это существенно зависит от вида и свойств функции f.

Отметим также, что аналог равенства (15) в случае, когда вместо условия на величину склеивания распределений P и Q накладывается условие на вариационное расстояние $V(P,Q)$ между ними, был получен в работах [3,5]. А именно, в них было показано, что справедливо равенство

$$\sup_{(P,Q)\colon V(P,Q)=v} D_f(P\,\|\,Q) = \frac{v}{2}[f(0) + f^*(0)].$$

Доказательство предложения 1. **1.** Докажем вначале формулу (8), когда предполагается, что $0 \leqslant \alpha \leqslant q_n$. В этом случае в соответствии с (5) оптимальная матрица M_{opt}, задающая α-склеивание данного распределения вероятностей $Q = \{q_i,\, i \in \mathcal{N}\}$ с некоторым распределением $P = \{p_i,\, i \in \mathcal{N}\}$, для которой $D_f(M_{\mathrm{opt}}) = D_f(P\,\|\,Q) = D_f^{\max}(Q,\alpha)$, находится среди матриц $M = \|p_{ij}\|_{i,j=1}^n$, у которых на диагонали в некотором столбце стоит α, остальные диагональные элементы равны нулю, а главный столбец либо содержит α, либо нет. Поэтому, пользуясь формулой

$$D_f(P\,\|\,Q) = \sum_{i\colon p_i>0} q_i f\left(\frac{p_i}{q_i}\right) + \left(\sum_{i\colon p_i=0} q_i\right) f(0), \tag{16}$$

получаем, что для таких матриц величина $D_f(M)$ может иметь три разных вида:

- Если α стоит на диагонали в k-м (главном) столбце, а $q_k - \alpha$ – в k-й строке и ℓ-м столбце, то

$$\begin{aligned} D_f(M) &= A_\alpha(q_k, q_\ell) = \\ &= q_k f\left(\frac{1+\alpha-q_k}{q_k}\right) + q_\ell f\left(\frac{q_k-\alpha}{q_\ell}\right) + (1-q_k-q_\ell)f(0); \end{aligned} \tag{17}$$

- Если на диагонали в главном k-м столбце стоит ноль, α стоит на диагонали в ℓ-м столбце и q_k стоит в k-й строке и ℓ-м столбце, то

$$\begin{aligned} D_f(M) &= B_\alpha(q_k, q_\ell) = \\ &= q_k f\left(\frac{1-\alpha-q_k}{q_k}\right) + q_\ell f\left(\frac{q_k+\alpha}{q_\ell}\right) + (1-q_k-q_\ell)f(0); \end{aligned} \tag{18}$$

- Если на диагонали в главном k-м столбце стоит ноль, q_k стоит в k-й строке и ℓ-м столбце, а α – на диагонали в m-м $(m \neq \ell)$ столбце, то

$$\begin{aligned} D_f(M) &= C_\alpha(q_k, q_\ell, q_m) = \\ &= q_k f\left(\frac{1-\alpha-q_k}{q_k}\right) + q_\ell f\left(\frac{q_k}{q_\ell}\right) + q_m f\left(\frac{\alpha}{q_m}\right) + (1-q_k-q_\ell-q_m)f(0). \end{aligned} \tag{19}$$

Поэтому

$$D_f^{\max}(Q,\alpha) = \max\{A_\alpha, B_\alpha, C_\alpha\}, \tag{20}$$

где

$$A_\alpha = \max_{k,\ell} A_\alpha(q_k, q_\ell), \quad B_\alpha = \max_{k,\ell} B_\alpha(q_k, q_\ell), \quad C_\alpha = \max_{k,\ell,m} C_\alpha(q_k, q_\ell, q_m), \tag{21}$$

а максимумы в (21) берутся по всем различным между собой натуральным числам k, ℓ и m из множества $\mathcal{N}$. Докажем теперь следующее утверждение.

Лемма 1. *Справедливо неравенство $C_\alpha \leqslant B_\alpha$, где B_α и C_α определены в* (21).

Доказательство. Достаточно показать, что при любом фиксированном значении k справедливо неравенство

$$\max_{\ell:\,\ell\neq k} q_\ell f\left(\frac{q_k+\alpha}{q_\ell}\right) \geqslant \max_{(\ell,m):\,\ell\neq k,\, m\neq k}\left[q_\ell f\left(\frac{q_k}{q_\ell}\right) + q_m f\left(\frac{\alpha}{q_m}\right) - q_m f(0)\right]. \tag{22}$$

Для этого заметим, что функция $xf\left(\dfrac{a}{x}\right) - xf(0)$ убывает по x, если $a = \mathrm{const} > 0$ и $x > 0$. Действительно,

$$\left[xf\left(\frac{a}{x}\right) - xf(0)\right]'_x = f\left(\frac{a}{x}\right) - \left(\frac{a}{x}\right)f'\left(\frac{a}{x}\right) - f(0) < 0,$$

так как функция $f(u) - uf'(u) - f(0)$ убывает по u, а при $u = 0$ она равна нулю. Поэтому, полагая

$$s = \begin{cases} n, & \text{если } k \neq n, \\ n-1, & \text{если } k = n, \end{cases}$$

получаем, что для доказательства (22) достаточно показать, что

$$q_s f\left(\frac{q_k+\alpha}{q_s}\right) - q_s f(0) \geqslant q_s f\left(\frac{q_k}{q_s}\right) - q_s f(0) + q_s f\left(\frac{\alpha}{q_s}\right) - q_s f(0),$$

т.е. что

$$f\left(\frac{q_k+\alpha}{q_s}\right) \geqslant f\left(\frac{q_k}{q_s}\right) + f\left(\frac{\alpha}{q_s}\right) - f(0). \tag{23}$$

Справедливость неравенства (23), а значит, и леммы 1, следует из того, что разность левой и правой частей в (23) возрастает по α, а при $\alpha = 0$ она равна нулю. ▲

Лемма 2. *Справедливо равенство*

$$\max\{A_\alpha, B_\alpha\} = \max\bigl\{A_\alpha(q_n, q_{n-1}), A_\alpha(q_{n-1}, q_n), B_\alpha(q_{n-1}, q_n)\bigr\},$$

где A_α и B_α определены в (21), *а $A_\alpha(x,y)$ и $B_\alpha(x,y)$ – в* (9) *и* (10).

Доказательство. Прежде всего заметим, что $A_\alpha(q_k, q_\ell)$ и $B_\alpha(q_k, q_\ell)$ являются выпуклыми функциями от q_k и убывающими от q_ℓ. Поэтому (см. (21))

$$A_\alpha = \max\bigl\{A_\alpha(q_{n-1}, q_n), A_\alpha(q_1, q_n), A_\alpha(q_n, q_{n-1})\bigr\}, \tag{24}$$

$$B_\alpha = \max\bigl\{B_\alpha(q_{n-1}, q_n), B_\alpha(q_1, q_n), B_\alpha(q_n, q_{n-1})\bigr\}. \tag{25}$$

Покажем теперь, что

$$A_\alpha(q_1, q_n) \leqslant \max\bigl\{A_\alpha(q_{n-1}, q_n), B_\alpha(q_n, q_{n-1})\bigr\}, \tag{26}$$

$$B_\alpha(q_1, q_n) \leqslant \max\bigl\{B_\alpha(q_{n-1}, q_n), A_\alpha(q_n, q_{n-1})\bigr\}. \tag{27}$$

Действительно, для доказательства (26) заметим, что если $[A_\alpha(x, q_n)]'_x\big|_{x=q_1} < 0$, то $A_\alpha(x, q_n)$ убывает по x при всех $0 < x < q_1$, и тогда $A_\alpha(q_1, q_n) < A(q_{n-1}, q_n)$.

Если же $[A_\alpha(x,q_n)]'_x\big|_{x=q_1} > 0$, то $A_\alpha(x,q_n)$ возрастает по x при $x > q_1$, и тогда $A_\alpha(q_1,q_n) < A(1-q_n,q_n)$.

Поэтому для доказательства (26) необходимо только показать, что

$$A_\alpha(1-q_n,q_n) \leqslant B_\alpha(q_n,q_{n-1}). \tag{28}$$

Сравнивая выражения для $A_\alpha(1-q_n,q_n)$ и $B_\alpha(q_n,q_{n-1})$ (см. (9) и (10)), мы видим, что для доказательства (28) достаточно лишь убедиться, что

$$(1-q_n)f\left(\frac{\alpha+q_n}{1-q_n}\right) - (1-q_n)f(0) \leqslant q_{n-1}f\left(\frac{\alpha+q_n}{q_{n-1}}\right) - q_{n-1}f(0).$$

А это неравенство следует из того, что, как было показано выше при доказательстве леммы 1, функция $xf\left(\frac{a}{x}\right) - xf(0)$ убывает по x, если $a = \mathrm{const} > 0$ и $x > 0$. Неравенство (26) доказано. Неравенство (27) доказывается аналогично. Поэтому, учитывая (24)–(27), получаем, что

$$\max\{A_\alpha, B_\alpha\} = \max\bigl\{A_\alpha(q_{n-1},q_n), A_\alpha(q_n,q_{n-1}), B_\alpha(q_{n-1},q_n), B_\alpha(q_n,q_{n-1})\bigr\}.$$

Таким образом, для доказательства леммы 2 достаточно лишь показать, что $B_\alpha(q_n,q_{n-1}) \leqslant A_\alpha(q_n,q_{n-1})$. Для доказательства этого последнего неравенства следует лишь убедиться, что производная разности $A_\alpha(q_n,q_{n-1}) - B_\alpha(q_n,q_{n-1})$ по α положительна, а при $\alpha = 0$ эта разность равна нулю. ▲

Наконец, формула (8) теперь следует из (20) и лемм 1 и 2.

2. Докажем теперь верхнюю границу (11) в случае, когда $q_n < \alpha < 1 - q_n$. Для этого в соответствии с (5) достаточно показать, что для всех матриц M, осуществляющих α-склеивание заданного распределения вероятностей $Q = \{q_i,\ i \in \mathcal{N}\}$ с некоторым распределением $P = \{p_i,\ i \in \mathcal{N}\}$ и принадлежащих множествам $\mathcal{M}(Q,I)$, соответствующих всевозможным допустимым (Q,I)-представлениям α, справедливо неравенство

$$D_f(M) \leqslant q_n f\left(\frac{1-\alpha+q_n}{q_n}\right) + (1-\alpha)f(0). \tag{29}$$

Пусть $\alpha = \sum\limits_{i\in I} q_i + \beta$, где $\beta = 0$ или существует j, такое что $0 < \beta < q_j$ для некоторого $j \in \mathcal{N} \setminus I$, – произвольное допустимое (Q,I)-представление α (отметим, что в случае, когда $I = \varnothing$, считаем, что $\sum\limits_{i\in I} q_i = 0$). Тогда величина $D_f(M)$, соответствующая различным матрицам $M \in \mathcal{M}(Q,I)$, может принимать разные значения в зависимости от расположения главного столбца (диагональным элементом которого является либо некоторое q_k, $k \in I$, либо ноль, либо β) и единственного ненулевого элемента вне главного столбца и диагонали (хотя при некоторых (Q,I)-представлениях α такой элемент может вообще отсутствовать). Рассмотрим каждый из этих случаев в отдельности и покажем, что для каждого из них справедливо неравенство (29).

а) Если диагональным элементом главного (k-го) столбца матрицы $M \in \mathcal{M}(Q,I)$ служит q_k, где $k \in I$, а элемент β расположен на диагонали в j-м столбце, то воспользовавшись формулой (16), получим, что

$$D_f(M) = q_k f\left(\frac{1-\alpha+q_k}{q_k}\right) + q_j f\left(\frac{\beta}{q_j}\right) + (1-\alpha-q_j+\beta)f(0). \tag{30}$$

Справедливость верхней границы (29) в этом случае следует из того, что функция $q_k f\left(\dfrac{1-\alpha+q_k}{q_k}\right)$ убывает по q_k, а при любых q_j и β, $0 \leqslant \beta \leqslant q_j$, имеет место неравенство

$$q_j f\left(\frac{\beta}{q_j}\right) - (q_j - \beta) f(0) \leqslant 0. \tag{31}$$

Действительно, неравенство (31) следует из того, что его левая часть является выпуклой функцией β, а при $\beta = 0$ и $\beta = q_j$ она равна нулю (так как по условию $f(1) = 0$).

b1) Если диагональным элементом главного k-го столбца матрицы $M \in \mathcal{M}(Q, I)$ служит ноль, элементы q_k и q_j, где $k \notin I$, $j \in I$, находятся в j-м столбце, а элемент β расположен на диагонали в i-м столбце, то для такой матрицы

$$\begin{aligned} D_f(M) = {} & q_k f\left(\frac{1-\alpha-q_k}{q_k}\right) + q_j f\left(\frac{q_k+q_j}{q_j}\right) + q_i f\left(\frac{\beta}{q_i}\right) + \\ & + (1-\alpha+\beta-q_k-q_i) f(0). \end{aligned} \tag{32}$$

Воспользовавшись тем, что $q_i f\left(\dfrac{\beta}{q_i}\right) - (q_i - \beta) f(0) \leqslant 0$ (см. (31)), получаем оценку

$$D_f(M) \leqslant q_k f\left(\frac{1-\alpha-q_k}{q_k}\right) + q_j f\left(\frac{q_k+q_j}{q_j}\right) + (1-\alpha) f(0) - q_k f(0). \tag{33}$$

Поэтому для доказательства верхней границы (29) в рассматриваемом случае достаточно показать, что при любых q_k и q_j справедливо неравенство

$$q_n f\left(\frac{1-\alpha+q_n}{q_n}\right) \geqslant q_k f\left(\frac{1-\alpha-q_k}{q_k}\right) + q_j f\left(\frac{q_k+q_j}{q_j}\right) - q_k f(0). \tag{34}$$

Для доказательства неравенства (34) заметим, что разность его левой и правой частей убывает по α, а при максимальном значении $\alpha = 1 - q_k$ (так как в рассматриваемом случае на диагонали в k-м столбце стоит ноль) эта разность, равная $q_n f\left(\dfrac{q_k+q_n}{q_n}\right) - q_j f\left(\dfrac{q_k+q_j}{q_j}\right)$, положительна, поскольку функция $q_j f\left(\dfrac{q_k+q_j}{q_j}\right)$ убывает по q_j.

b2) Если диагональным элементом главного k-го столбца матрицы $M \in \mathcal{M}(Q, I)$ служит ноль, q_k находится в i-м столбце, в котором на диагонали стоит β, то для такой матрицы

$$\begin{aligned} D_f(M) &= q_k f\left(\frac{1-\alpha-q_k}{q_k}\right) + q_i f\left(\frac{q_k+\beta}{q_i}\right) + (1-\alpha+\beta-q_k-q_i) f(0) = \\ &= q_k f\left(\frac{1-\alpha-q_k}{q_k}\right) + q_i f\left(\frac{q_k+\beta}{q_i}\right) + (1-\alpha) f(0) - q_k f(0) - q_i f\left(\frac{\beta}{q_i}\right) + \\ &+ \left[q_i f\left(\frac{\beta}{q_i}\right) - (q_i-\beta) f(0)\right] \leqslant q_k f\left(\frac{1-\alpha-q_k}{q_k}\right) + q_i f\left(\frac{q_k+q_i}{q_i}\right) + \\ &+ (1-\alpha) f(0) - q_k f(0). \end{aligned} \tag{35}$$

Неравенство в (35) следует из (31) и того факта, что

$$q_i f\left(\frac{q_k+q_i}{q_i}\right) \geqslant q_i f\left(\frac{q_k+\beta}{q_i}\right) - q_i f\left(\frac{\beta}{q_i}\right). \tag{36}$$

Действительно, справедливость неравенства (36) является следствием того, что его правая часть является возрастающей функцией β, а при максимальном значении $\beta = q_i$ левая и правая части (36) равны. Теперь заметим, что правая часть неравенства (35) совпадает с выражением правой части в (33), которую мы уже оценили сверху в п. b1), показав, что она не превосходит нужной нам оценки (29).

b3) Если диагональным элементом главного k-го столбца матрицы $M \in \mathcal{M}(Q, I)$ служит ноль, q_k находится в ℓ-м столбце, диагональным элементом которого является ноль, а β находится на диагонали в i-м столбце, то для такой матрицы

$$
\begin{aligned}
D_f(M) &= q_k f\left(\frac{1-\alpha-q_k}{q_k}\right) + q_i f\left(\frac{\beta}{q_i}\right) + q_\ell f\left(\frac{q_k}{q_\ell}\right) + \\
&+ (1-\alpha+\beta-q_k-q_i-q_\ell) f(0) \leqslant q_k f\left(\frac{1-\alpha-q_k}{q_k}\right) + q_i f\left(\frac{\beta}{q_i}\right) + \\
&+ q_\ell f\left(\frac{q_k+q_\ell}{q_\ell}\right) + (1-\alpha+\beta-q_k-q_i) f(0).
\end{aligned} \tag{37}
$$

Неравенство в (37) следует из того, что

$$
q_\ell f\left(\frac{q_k}{q_\ell}\right) - q_\ell f(0) \leqslant q_\ell f\left(\frac{q_k+q_\ell}{q_\ell}\right). \tag{38}
$$

Действительно, справедливость (38) является следствием того, что ввиду выпуклости функции $f(x)$ и условия $f(1) = 0$ имеет место неравенство $f(1+x) \geqslant f(x) - f(0)$. Правая часть неравенства (37) совпадает с выражением правой части в (32), которую мы уже оценили сверху в п. b1), показав, что она не превосходит оценки (29).

c1) Если диагональным элементом главного k-го столбца матрицы $M \in \mathcal{M}(Q, I)$ служит β, а $q_k - \beta$ находится в j-м столбце, диагональным элементом которого является q_j, $j \in I$, то для такой матрицы

$$
\begin{aligned}
&D_f(M) = \\
&= q_k f\left(\frac{1-\alpha-q_k+2\beta}{q_k}\right) + q_j f\left(\frac{q_k+q_j-\beta}{q_j}\right) + (1-\alpha+\beta-q_k) f(0).
\end{aligned} \tag{39}
$$

Так как правая часть (39) является выпуклой функцией β, то ее максимум достигается либо при $\beta = 0$, либо при $\beta = q_k$. В случае $\beta = 0$ правая часть (39) совпадает с правой частью (33), которую мы уже оценивали сверху в п. b1), показав, что она не превосходит оценки (29). Если же $\beta = q_k$, то максимум правой части (39) очевидно совпадает с доказываемой оценкой (29).

c2) Если диагональным элементом главного k-го столбца матрицы $M \in \mathcal{M}(Q, I)$ служит β, а $q_k - \beta$ находится в i-м столбце, диагональным элементом которого является ноль, то для такой матрицы

$$
\begin{aligned}
&D_f(M) = \\
&= q_k f\left(\frac{1-\alpha-q_k+2\beta}{q_k}\right) + q_i f\left(\frac{q_k-\beta}{q_i}\right) + (1-\alpha+\beta-q_k-q_i) f(0).
\end{aligned} \tag{40}
$$

Снова очевидно, что правая часть (40) является выпуклой функцией β, а тогда опять все сводится к рассмотренным ранее случаям (см. (37), (39)).

Верхняя граница (11) теперь полностью доказана. Наконец, очевидно, что верхняя граница (11) достигается, если $\alpha = q_n + \sum\limits_{i=1}^{n-1} a_i q_i$ при любых $a_i \in \{0, 1\}$, так как в этом случае $\beta = 0$ и в (30) при $q_k = q_n$ получаем нужное нам равенство. ▲

Доказательство следствия 1. Равенство (12) для случая, когда $0 \leqslant \alpha \leqslant q_n$ и $|\mathcal{N}| = n \geqslant 3$, следует из того, что значение $D_f^{\max}(q_{\min}, \alpha)$ достигается для распределения $Q = \{q_i,\ i \in \mathcal{N}\}$, у которого $q_{n-1} = q_n = q_{\min}$, в чем нетрудно убедиться, просматривая доказательство равенства (8). При этом этот максимум равен $A_\alpha(q_n, q_n)$, где $A_\alpha(x, y)$ определено в (9). Действительно, в случае, когда $q_{n-1} = q_n = q_{\min}$, имеем $A_\alpha(q_n, q_{n-1}) = A_\alpha(q_{n-1}, q_n) = A_\alpha(q_n, q_n)$, и нетрудно видеть, что $A_\alpha(q_n, q_n) \geqslant B_\alpha(q_n, q_n)$, для чего следует лишь заметить, что разность $A_\alpha(q_n, q_n) - B_\alpha(q_n, q_n)$ возрастает по α, а при $\alpha = 0$ эта разность равна нулю.

Равенство (13) также легко доказать, заметив, что максимумы каждого из двух выражений в (7) достигаются при $q_{n-1} = q_n = q_{\min}$, а в этом случае оба этих выражения совпадают.

Наконец, верхняя граница (14) является прямым следствием (11), а справедливость равенства в (14) также следует из утверждения предложения об условиях равенства в (11), поскольку при сформулированных там условиях нетрудно предъявить соответствующие распределения вероятностей $Q = \{q_i,\ i \in \mathcal{N}\}$, при которых имеет место равенство $\alpha = \sum\limits_{i=1}^{n-1} a_i q_i + q_n$ при некоторых $a_i \in \{0, 1\}$ (см. доказательство следствия 1 в [1]). ▲

Доказательство предложения 2. Заметим вначале, что из определения f-дивергенции сразу следует, что $D_f^{\max}(\alpha) = 0$, если $\alpha = 1$. Далее заметим, что в случае, когда $0 \leqslant \alpha < 1$ и либо $f(0) = \infty$, либо $f^*(0) = \infty$, равенство (15) очевидно, так как всегда можно построить матрицу $M = \|p_{ij}\|_{i,j=1}^n$, задающую α-склеивание некоторого распределения вероятностей $P = \{p_i,\ i \in \mathcal{N}\}$ с другим распределением $Q = \{q_i,\ i \in \mathcal{N}\}$, такую что либо некоторое $p_i = 0$, а соответствующее $q_i \neq 0$ (в случае, когда $f(0) = \infty$), либо некоторое $q_i = 0$, а $p_i \neq 0$ (в случае $f^*(0) = \infty$). Поэтому в дальнейшем при доказательстве равенства (15) будем всегда предполагать, что и $f(0) < \infty$, и $f^*(0) < \infty$.

Далее для доказательства (15) воспользуемся следующими неравенствами для выпуклой функции $f(x)$, задающей f-дивергенцию $D_f(P \,\|\, Q)$:

$$f(x) \leqslant \begin{cases} (1-x)f(0), & \text{если } 0 < x < 1, \\ (x-1)f^*(0), & \text{если } x > 1. \end{cases} \tag{41}$$

Действительно, первое неравенство следует из определения выпуклости функции $f(x)$ и условия $f(1) = 0$, второе – из того, что функция $f^*(x) = xf(1/x)$ тоже выпукла и $f^*(1) = 0$.

Теперь, поскольку

$$D_f^{\max}(\alpha) = \sup_{q_{\min}\colon 0 < q_{\min} \leqslant 1/n} D_f^{\max}(q_{\min}, \alpha), \tag{42}$$

то для получения оценки сверху для $D_f^{\max}(\alpha)$ достаточно оценить сверху функции от $q_{\min}$, стоящие в правых частях равенств (12), (13) и неравенства (14) при заданном параметре α, $0 \leqslant \alpha < 1$, и любом возможном значении $q_{\min}$. Замечая, что функции в правых частях равенств (12),(13) являются выпуклыми относительно $q_{\min}$, а функция в правой части неравенства (14) убывает по $q_{\min}$, то очевидно имеем

$$A(q_{\min}, q_{\min}) \leqslant \max\{A_\alpha(\alpha, \alpha), A_\alpha(1, 1)\}, \tag{43}$$

так как в (12) предполагается, что $q_{\min} \geqslant \alpha$,

$$\begin{aligned} q_{\min}\left[f\left(\frac{\alpha - 1 + q_{\min}}{q_{\min}}\right) + f\left(\frac{1 - \alpha + q_{\min}}{q_{\min}}\right)\right] \leqslant \\ \leqslant \max\bigl\{(1-\alpha)[f(0) + f(2)], f(\alpha) + f(2-\alpha)\bigr\}, \end{aligned} \tag{44}$$

так как в (13) предполагается, что $q_{\min} \geqslant 1-\alpha$, и

$$q_{\min} f\left(\frac{1-\alpha+q_{\min}}{q_{\min}}\right) \leqslant 0 \cdot f\left(\frac{1-\alpha}{0}\right) = (1-\alpha) f^*(0), \tag{45}$$

так как в (14) предполагается, что $q_{\min} \leqslant \alpha$.

Поэтому, воспользовавшись неравенствами (41) для оценивания сверху выражений в (43)–(45) и учитывая (12)–(14) и (42), получаем, что

$$D_f^{\max}(\alpha) \leqslant (1-\alpha)[f(0)+f^*(0)]. \tag{46}$$

Таким образом, для доказательства равенства (15) необходимо показать, что на самом деле знак неравенства в (46) можно заменить на знак равенства при любом α, $0 \leqslant \alpha < 1$. Покажем это. Действительно, если $\alpha = 0$, то из (12) следует, что

$$\lim_{q_{\min}\to 0} D_f^{\max}(q_{\min}, \alpha) = \lim_{x\to 0} A_0(x,x) = \lim_{t\to\infty} \frac{f(t)}{t+1} + f(0) = f^*(0) + f(0).$$

Это означает, что при $\alpha = 0$ в (46) имеет место знак равенства, т.е. равенство (15) доказано для $\alpha = 0$. Если же α, $0 < \alpha < 1$, – любое фиксированное число, то в (14) при достаточно малых $q_{\min}$ имеет место знак равенства, а тогда

$$\lim_{q_{\min}\to 0} D_f^{\max}(q_{\min}, \alpha) = \lim_{x\to 0}\left[x f\left(\frac{1-\alpha+x}{x}\right)\right] + (1-\alpha) f(0) =$$
$$= \lim_{t\to\infty} (1-\alpha)\frac{f(t)}{t-1} + (1-\alpha) f(0) = (1-\alpha)[f^*(0)+f(0)],$$

так что снова в (46) имеет место знак равенства, а значит, равенство (15) справедливо и при любом α, $0 < \alpha < 1$. ▲

СПИСОК ЛИТЕРАТУРЫ

1. *Прелов В.В.* f-дивергенция и склеивание вероятностных распределений // Пробл. передачи информ. 2021. Т. 57. № 1. С. 64–80. `https://doi.org/10.31857/S0555292321010034`
2. *Csiszár I.* Information-type Measures of Difference of Probability Distributions and Indirect Observations // Studia Sci. Math. Hungar. 1967. V. 2. № 3–4. P. 299–318.
3. *Sason I., Verdú S.* f-Divergence Inequalities // IEEE Trans. Inform. Theory. 2016. V. 62. № 11. P. 5973–6006. `https://doi.org/10.1109/TIT.2016.2603151`
4. *Макур А., Чжен Л.* Сравнение коэффициентов сжатия для f-дивергенций // Пробл. передачи информ. 2020. Т. 56. № 2. С. 3–62. `https://doi.org/10.31857/S0555292320020011`
5. *Basu A., Shioya Y., Park C.* Statistical Inference: The Minimum Distance Approach. Boca Raton, FL: CRC Press, 2011.

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

Поступила в редакцию
12.11.2021
После доработки
16.11.2021
Принята к публикации
16.11.2021