D. K. Faddeev, On the concept of entropy of a finite probabilistic scheme, *Uspekhi Mat. Nauk*, 1956, Volume 11, Issue 1(67), 227–231

1956 г. январь—февраль *т. XI, вып. 1 (67)*

УСПЕХИ МАТЕМАТИЧЕСКИХ НАУК

К ПОНЯТИЮ ЭНТРОПИИ КОНЕЧНОЙ ВЕРОЯТНОСТНОЙ СХЕМЫ

Д. К. Фаддеев

1°. Как установлено в статье А. Я. Хинчина [1], энтропия $H(p_1, p_2, \ldots, p_n)$ конечной вероятностной схемы $(A_1, A_2, \ldots, A_n)$ может быть охарактеризована следующей системой аксиом.

1. $H(p_1, p_2, \ldots, p_n)$ непрерывна относительно $p_1, p_2, \ldots, p_n$ в области $0 \leqslant p_i \leqslant 1$, $p_1 + \ldots + p_n = 1$.

2. $H(p_1, p_2, \ldots, p_n)$ симметрична относительно $p_1, p_2, \ldots, p_n$.

3. $H(p_1, p_2, \ldots, p_n, 0) = H(p_1, p_2, \ldots, p_n)$.

4. $H(q_{11}, \ldots, q_{1m}; q_{21}, \ldots, q_{2m}; \ldots; q_{n1}, \ldots, q_{nm}) = H(p_1, p_2, \ldots, p_n) + \sum_{i=1}^{n} p_i H\left(\frac{q_{i1}}{p_i}, \ldots, \frac{q_{im}}{p_i}\right)$,

где

$$p_i = q_{i1} + \ldots + q_{im} \qquad (i = 1, \ldots, n)$$

(эта аксиома сформулирована в статье [1] в других терминах).

5. $H(p_1, p_2, \ldots, p_n) \leqslant H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$.

Этими аксиомами энтропия определяется однозначно с точностью до положительного множителя. Целью настоящей заметки является некоторое упрощение системы аксиом. Именно, предлагаются следующие аксиомы:

1′. $H(p, 1-p)$ непрерывна при $0 \leqslant p \leqslant 1$ и положительна хотя бы в одной точке.

2′. $H(p_1, p_2, \ldots, p_n)$ симметрична относительно $p_1, p_2, \ldots, p_n$.

3′. При $n \geqslant 2$

$$H(p_1, \ldots, p_{n-1}; q_1, q_2) = H(p_1, \ldots, p_{n-1}, p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right).$$

Здесь $p_n = q_1 + q_2$.

Разница в этих двух системах аксиом заключается в том, что, во-первых, аксиома 5 (экстремальность) заменяется требованием положительности энтропии в одной точке и, во-вторых, аксиомы 3 и 4 заменяются одной аксиомой 3′, очень естественной, если рассматривать энтропию как меру неопределённости схемы. Действительно, «неопределённость» схемы

$\begin{pmatrix} A_1, & \ldots, & A_{n-1}; & B_1, & B_2 \\ p_1, & \ldots, & p_{n-1}; & q_1, & q_2 \end{pmatrix}$ отличается от «неопределённости» схемы $\begin{pmatrix} A_1, & \ldots, & A_{n-1}, & A_n \\ p_1, & \ldots, & p_{n-1}, & p_n \end{pmatrix}$ на «неопределённость», происходящую от подразделения события A_n на два подсобытия B_1, B_2 с условными вероятностями $\frac{q_1}{p_n}$, $\frac{q_2}{p_n}$. Эта «неопределённость» должна быть преодолена только в том случае, если реализуется событие A_n, вероятность чего равна p_n.

2°. Прежде всего, установим, что аксиомы 2, 3, 4 равносильны 2′, 3′.

Лемма 1. *Из 2, 3, 4 следует, что* $H(1, 0) = 0$.

Доказательство. Подсчитаем $H\left(\frac{1}{2}, \frac{1}{2}, 0, 0\right)$ двумя способами. С одной стороны, в силу 3 имеем $H\left(\frac{1}{2}, \frac{1}{2}, 0, 0\right) = H\left(\frac{1}{2}, \frac{1}{2}\right)$. С другой стороны, $H\left(\frac{1}{2}, \frac{1}{2}, 0, 0\right) = H\left(\frac{1}{2}, 0, \frac{1}{2}, 0\right)$ и в силу 4 получим $H\left(\frac{1}{2}, 0; \frac{1}{2}, 0\right) = H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2}H(1, 0) + \frac{1}{2}H(1, 0)$. Следовательно, $H(1, 0) = 0$.

Лемма 2. *Из 2, 3, 4 следует* 3′.

Доказательство. $H(p_1, 0; \ldots; p_{n-1}, 0; q_1, q_2)$ равно в силу 2 и 3 $H(p_1, 0; \ldots; p_{n-1}, 0; q_1, q_2)$, что равно в силу 4 $H(p_1, \ldots, p_{n-1}, p_n) + p_1 H(1, 0) + \ldots + p_{n-1}H(1, 0) + p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right) = H(p_1, \ldots, p_{n-1}, p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right)$.

Лемма 3. *Из 2′, 3′ следует, что* $H(1, 0) = 0$.

Доказательство. Подсчитаем двумя способами $H\left(\frac{1}{2}, \frac{1}{2}, 0\right)$. С одной стороны, в силу 3′ имеем $H\left(\frac{1}{2}; \frac{1}{2}, 0\right) = H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2}H(1, 0)$. С другой стороны, $H\left(\frac{1}{2}, \frac{1}{2}, 0\right) = H\left(0; \frac{1}{2}, \frac{1}{2}\right) = H(0, 1) + H\left(\frac{1}{2}, \frac{1}{2}\right)$. Следовательно, $H(1, 0) = 0$.

Лемма 4. *Из 2′, 3′ следует* 3.

Доказательство. $H(p_1, \ldots, p_{n-1}; p_n, 0) = H(p_1, \ldots, p_{n-1}, p_n) + p_n H(1, 0) = H(p_1, \ldots, p_{n-1}, p_n)$.

Лемма 5. *Из 3′ следует:*

$$H(p_1, \ldots, p_{n-1}; q_1, \ldots, q_m) = H(p_1, \ldots, p_{n-1}, p_n) + p_n H\left(\frac{q_1}{p_n}, \ldots, \frac{q_m}{p_n}\right).$$

Здесь $n \geqslant 2$, $p_n = q_1 + \ldots + q_m$.

Доказательство. При $m = 2$ утверждение леммы совпадает с аксиомой 3′, при $m \geqslant 3$ лемма легко доказывается индукцией по m.

Лемма 6. *Из 2′, 3′ следует:*

$$H(q_{11}, \ldots, q_{1m_1}; q_{21}, \ldots, q_{2m_2}; \ldots; q_{n1}, \ldots, q_{nm_n}) = \\ = H(p_1, \ldots, p_n) + \sum_{i=1}^{n} p_i H\left(\frac{q_{i1}}{p_i}, \ldots, \frac{q_{im_i}}{p_i}\right).$$

Здесь $p_i = q_{i1} + \ldots + q_{im_i}$.

Доказательство. Нужно n раз применить лемму 5 к каждой группе аргументов, что возможно в силу симметрии.

Положив $m_1 = m_2 = \ldots = m_n = m$, получим, что из 2′, 3′ следует 4.

Итак, мы установили равносильность групп аксиом 2, 3, 4 и 2′, 3′.

3°. Положим, следуя статье [1], $F(n) = H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$ при $n \geqslant 2$ и $F(1) = 0$. Применяя лемму 6 к случаю $m_1 = \ldots = m_n = m$, $q_{ij} = \frac{1}{mn}$, получим:

$$F(mn) = F(m) + F(n) \tag{1}$$

при $m \geqslant 2$, $n \geqslant 2$. При $m = 1$ или $n = 1$ соотношение (1) тривиально. Далее, применяя лемму 5 к $H\left(\frac{1}{n}; \frac{1}{n}, \ldots, \frac{1}{n}\right)$, получим:

$$H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) = H\left(\frac{1}{n}, \frac{n-1}{n}\right) + \frac{n-1}{n} H\left(\frac{1}{n-1}, \ldots, \frac{1}{n-1}\right),$$

откуда следует:

$$\eta_n = H\left(\frac{1}{n}, \frac{n-1}{n}\right) = F(n) - \frac{n-1}{n} F(n-1).$$

Лемма 7. *При* $n \to \infty$

$$\mu_n = \frac{F(n)}{n} \to 0 \quad \text{и} \quad \lambda_n = F(n) - F(n-1) \to 0.$$

Доказательство. В силу непрерывности $H(p, 1-p)$ при $p = \eta_n = H\left(\frac{1}{n}, \frac{n-1}{n}\right) \to H(0, 1) = 0$.

Далее, $n\eta_n = nF(n) - (n-1)F(n-1)$, откуда

$$nF(n) = \sum_{k=1}^{n} k\eta_k$$

и

$$\frac{F(n)}{n} = \frac{1}{n^2} \sum_{k=1}^{n} k\eta_k = \frac{n+1}{2n} \cdot \frac{2}{n(n+1)} \sum_{k=1}^{n} k\eta_k.$$

Но $\frac{2}{n(n+1)} \sum_{k=1}^{n} k\eta_k$ есть среднее арифметическое первых $\frac{n(n+1)}{2}$ членов стремящейся к нулю последовательности $\eta_1, \eta_2, \eta_2, \eta_3, \eta_3, \eta_3, \ldots$ Следовательно, $\frac{2}{n(n+1)} \sum_{k=1}^{n} k\eta_k \to 0$ и $\mu_n = \frac{F(n)}{n} \to 0$.

Далее, $\lambda_n = F(n) - F(n-1) = \eta_n - \frac{1}{n} F(n-1) \to 0$. Лемма доказана.

В силу соотношения (1) $F(n)$ будет определено при всех натуральных n, как только заданы значения функции F на простых числах. Именно, если $n = p_1^{\alpha_1}, \ldots, p_s^{\alpha_s}$, где $p_1, \ldots, p_s$ простые, то $F(n) = \alpha_1 F(p_1) + \ldots + \alpha_s F(p_s)$.

Положим $F(p) = c_p \lg p$. Тогда

$$F(n) = \alpha_1 c_{p_1} \lg p_1 + \ldots + \alpha_s c_{p_s} \lg p_s.$$

Лемма 8. *Среди чисел* c_p $(p=2, 3, 5, 7, \ldots)$ *существует наибольшее.*

Доказательство. Допустим противное и покажем, что это предположение вступит в противоречие с предположением о непрерывности $H(p, 1-p)$ при $p=0$.

Действительно, если в последовательности c_p $(p=2, 3, 5, 7, \ldots)$ нет наибольшего числа, то можно построить бесконечную последовательность простых чисел $p_1, p_2, \ldots, p_i$ так, что $p_1=2$, p_i есть наименьшее простое число, такое, что $c_{p_i} > c_{p_{i-1}}$. Из способа построения этой последовательности ясно, что как только простое число q меньше p_i, то $c_q \leqslant c_{p_i}$.

Пусть $i>1$ и $q_1^{\alpha_1} \ldots q_s^{\alpha_s}$ есть каноническое разложение числа p_i-1. Рассмотрим

$$\begin{aligned}\lambda_{p_i} &= F(p_i) - F(p_i-1) = \\ &= F(p_i) - \frac{F(p_i)}{\lg p_i}\lg(p_i-1) + c_{p_i}\lg(p_i-1) - F(p_i-1) = \\ &= \frac{F(p_i)}{p_i}\cdot\frac{p_i}{\lg p_i}\lg\frac{p_i}{p_i-1} + \sum_{j=1}^{s}\alpha_j(c_{p_i}-c_{q_j})\lg q_j.\end{aligned}$$

Ясно, что

$$\sum_{j=1}^{s}\alpha_j(c_{p_i}-c_{q_j})\lg q_j \geqslant (c_{p_i}-c_2)\lg 2 \geqslant (c_{p_2}-c_2)\lg 2,$$

ибо $q_j < p_i$, $c_{q_j} \leqslant c_{p_i}$ и в силу чётности p_i-1 среди q_j присутствует число 2 с ненулевым показателем.

Далее, при $i \to \infty$ $\frac{F(p_i)}{p_i} \to 0$, $\frac{p_i}{\lg p_i}\lg\frac{p_i}{p_i-1} \to 0$ и $\lambda_{p_i} \to 0$.

Следовательно, $(c_{p_2}-c_2)\lg 2 \leqslant 0$, что невозможно.

Совершенно таким же образом устанавливается, что среди c_p существует наименьшее.

Лемма 9. $F(n) = c\lg n$, *где* c — *постоянное.*

Доказательство. Достаточно установить, что все c_p равны друг другу.

Допустим, что найдётся такое простое число p', что $c_{p'} > c_2$. Обозначим через p то простое число, для которого c_p принимает наибольшее значение. Тогда $c_p > c_2$ и $p > 2$.

Пусть m — натуральное число и $q_1^{\alpha_1} \ldots q_s^{\alpha_s}$ есть каноническое разложение p^m-1. Рассмотрим

$$\begin{aligned}\lambda_{p^m} &= F(p^m) - F(p^m-1) = \\ &= F(p^m) - \frac{F(p^m)}{\lg p^m}\lg(p^m-1) + c_p\lg(p^m-1) - F(p^m-1) = \\ &= \frac{F(p^m)}{p^m}\cdot\frac{p^m}{\lg p^m}\lg\frac{p^m}{p^m-1} + \sum_{j=1}^{s}\alpha_j(c_p-c_{q_j})\lg q_j \geqslant \\ &\geqslant \frac{F(p^m)}{p^m}\cdot\frac{p^m}{\lg p^m}\lg\frac{p^m}{p^m-1} + (c_p-c_2)\lg 2.\end{aligned}$$

При переходе к пределу при $m \to \infty$ получим:

$$(c_p - c_2) \lg 2 \leqslant 0,$$

что невозможно. Таким образом, для всех p' имеет место $c_{p'} \leqslant c_2$.

Совершенно таким же образом устанавливается, что $c_{p'} \geqslant c_2$ при всех p'. Лемма доказана.

Теорема.

$$H(p_1, p_2, \ldots, p_n) = c \sum_{i=1}^{n} p_i \lg \frac{1}{p_i} \qquad (c > 0).$$

Доказательство. Рассмотрим сначала случай $n = 2$. Пусть $p = \frac{r}{s}$ при целых r и s.

Применим лемму 6 к $H\left(\frac{1}{s}, \ldots, \frac{1}{s}\right)$, соединив аргументы в две группы из r и $s-r$ элементов. Получим:

$$H\left(\frac{1}{s}, \ldots, \frac{1}{s}\right) =$$
$$= H\left(\frac{r}{s}, 1 - \frac{r}{s}\right) + \frac{r}{s} H\left(\frac{1}{r}, \ldots, \frac{1}{r}\right) + \frac{s-r}{s} H\left(\frac{1}{s-r}, \ldots, \frac{1}{s-r}\right),$$

откуда

$$\begin{aligned} H(p, 1-p) &= F(s) - pF(r) - (1-p) F(s-r) = \\ &= c \lg s - pc \lg r - (1-p) c \lg (s-r) = \\ &= c\left(p \lg \frac{s}{r} + (1-p) \lg \frac{s}{s-r}\right) = \\ &= c\left(p \lg \frac{1}{p} + (1-p) \lg \frac{1}{1-p}\right). \end{aligned}$$

В силу непрерывности $H(p, 1-p)$ полученная формула может быть распространена и на иррациональные значения p. $c > 0$ в силу положительности $H(p, 1-p)$ хотя бы в одной точке.

Переход к общему случаю осуществляется методом математической индукции, на основании аксиомы 3′.

Поступило в редакцию 21 июля 1955 г.

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

[1] А. Я. Хинчин, Понятие энтропии в теории вероятностей, УМН VIII, вып. 3 (55), (1953).