ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 29 МАТЕМАТИКИ Выпуск 1
2022

Е. А. Щелканова (Москва, МИЭМ НИУ ВШЭ). **Исследование энтропийных характеристик методов опробования.**

УДК DOI https://doi.org/10.52513/08698325_2022_29_1_1

Резюме: Работа посвящена изучению теоретико-информационных свойств алгоритмов опробования элементов произвольного дискретного множества до наступления «успеха», а также его вероятностной модификации — алгоритма усеченного опробования [1].

Для указанного типа алгоритмов получены точные формулы информационных энтропий: Шеннона, Реньи, Цаллиса.

Ключевые слова: Последовательное опробование, опробование с усечением, полиномиальное распределение, энтропия Шеннона, энтропия Реньи, энтропия Цаллиса, количество информации.

Основные результаты

Пусть для произвольного $n \in \mathbb{N}$ задано некоторое конечное множество $A = \{w_1 \ldots w_n\}$, и на его основе построена вероятностная схема $(A, \bar{p})$ [2] :

$$\begin{pmatrix} w_1 & w_2 & \ldots & w_n \\ p_1 & p_2 & \ldots & p_n \end{pmatrix}, \tag{1}$$

где компоненты соответствующего вектора распределения $\bar{p} = \big(p_1, p_2, \ldots, p_n\big)$, удовлетворяют соотношениям

$$p_1 + p_2 + \cdots + p_n = 1 \text{ и } 1 \geqslant p_1 \geqslant p_2 \geqslant \cdots \geqslant p_n > 0. \tag{2}$$

Пусть далее выбирается некоторый произвольный $w \in A$ и рассматривается классическая задача опробования [3, 4] элементов множества A до наступления «успеха», где под «успехом» понимается определение элемента w.

Через $H^{(k)}$ обозначим среднее количество информации, полученное о вероятностной схеме (1) на k-м шаге алгоритма опробования до «успеха», а через $\bar{H}^{(k)}$ — среднее количество информации, полученное за k последовательных шагов.

Для произвольного $l \in \{1, \ldots, n\}$ положим $\pi_l = \sum_{i=1}^{l} p_i$, $\pi_0 = 0$.

Рассмотрим процедуру опробования, представляющую собой схему Бернулли с вероятностью успеха $0 < \pi_l \leqslant 1$, где каждое отдельное испытание заключается в реализации алгоритма усеченного опробования случайного элемента $w \in A$ с надежностью π_l.

Далее через τ_l обозначим случайную величину, равную среднему количеству информации, полученной при реализации указанной процедуры.

Утверждение. *Пусть $n \in \mathbb{N}$ и на множестве A задано распределение* (2). *Тогда для произвольного фиксированного $l \in \{1, \ldots, n\}$ справедливо*

$$\mathbf{E}\tau_l = \sum_{k=1}^{l} \frac{1 - \pi_{k-1}}{\pi_l} H\left(\frac{p_k}{\pi_l - \pi_{k-1}}, 1 - \frac{p_k}{\pi_l - \pi_{k-1}}\right), \tag{3}$$

где H — символ произвольной информационной энтропии.

Д о к а з а т е л ь с т в о. Обозначим через k число шагов описанной процедуры до первого определения искомого элемента w с использованием алгоритма усеченного опробования с надежностью π_l. Тогда случайная величина k имеет геометрическое распределение с параметром π_l :

$$P\{k=t\} = (1-\pi_l)^t\ \pi_l,\ t = 0,1,2,\ldots$$

и при этом

$$\mathbf{E}k = \frac{1-\pi_l}{\pi_l}.$$

Для каждого отдельного испытания, в котором w не попал в число первых l опробуемых элементов, полученное среднее количество информации составляет $\bar{H}_l{}^{(k)}$, в противном случае — $\mathbf{E}\xi_l$.

Следовательно,

$$\begin{aligned}\mathbf{E}\tau_l = \bar{H}_l{}^{(l)}\mathbf{E}k + \mathbf{E}\xi_l &= \frac{1-\pi_l}{\pi_l}\sum_{k=1}^{l} H\left(\frac{p_k}{\pi_l-\pi_{k-1}}, 1-\frac{p_k}{\pi_l-\pi_{k-1}}\right)\\ &+\sum_{k=1}^{l}\left(1-\frac{\pi_{k-1}}{\pi_l}\right) H\left(\frac{p_k}{\pi_l-\pi_{k-1}}, 1-\frac{p_k}{\pi_l-\pi_{k-1}}\right)\\ &= \sum_{k=1}^{l}\frac{1-\pi_{k-1}}{\pi_l} H\left(\frac{p_k}{\pi_l-\pi_{k-1}}, 1-\frac{p_k}{\pi_l-\pi_{k-1}}\right).\end{aligned}$$

Утверждение доказано. ■

Следствие 1. *Пусть $n \in \mathbb{N}$ и на множестве A задано равновероятное распределение. Тогда в терминах информации Шеннона для произвольного фиксированного $l \in \{1,\ldots,n\}$ справедливо*

$$\mathbf{E}\tau_l^{(Sh)} = \frac{n-l}{l}\left(\sum_{i=2}^{l}\log_2 i - \sum_{i=2}^{l-1}\frac{i}{i+1}\log_2 i\right) + \log_2 l.$$

Следствие 2. *Пусть $n \in \mathbb{N}$ и на множестве A задано равновероятное распределение. Тогда в терминах информации Реньи для произвольного фиксированного $l \in \{1,\ldots,n\}$ справедливо*

$$\mathbf{E}\tau_l^{(R)} = \frac{1}{1-a}\sum_{i=0}^{l-1}\frac{n-i}{l}\left(\log_2\left(1+(l-i-1)^a\right) - a\log_2(l-i)\right).$$

Следствие 3. *Пусть $n \in \mathbb{N}$ и на множестве A задано равновероятное распределение. Тогда в терминах информации Цаллиса для произвольного фиксированного $l \in \{1,\ldots,n\}$ справедливо*

$$\mathbf{E}\tau_l^{(Ts)} = \frac{k}{q-1}\sum_{i=0}^{l-1}\frac{n-i}{l}\left(1-\frac{1+(l-i-1)^q}{(l-i)^q}\right).$$

Для равновероятного распределения на исходном алфавите A имеет место следующая зависимость: если значения параметров a и q в формулах для $\mathbf{E}\tau_l^{(R)}$ и $\mathbf{E}\tau_l^{(Ts)}$ в окрестности единицы, максимум информации приходится на начальные этапы проведения процедуры опробования. С увеличением указанных значений или при приближении от единицы к нулю количество информации начинает «распределяться» по всем шагам процедуры. Указанная зависимость согласовывается с работой [1] поскольку энтропия Реньи приближается к энтропии Шеннона при $a \to 1$.

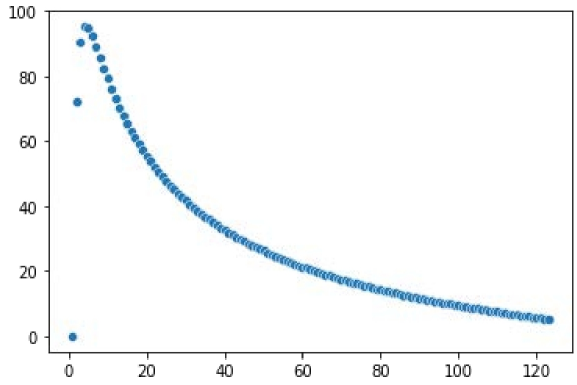

Рис. 1. Зависимость среднего количества информации от мощности алфавита при равновероятном распределении и q, находящимся в окрестности единицы

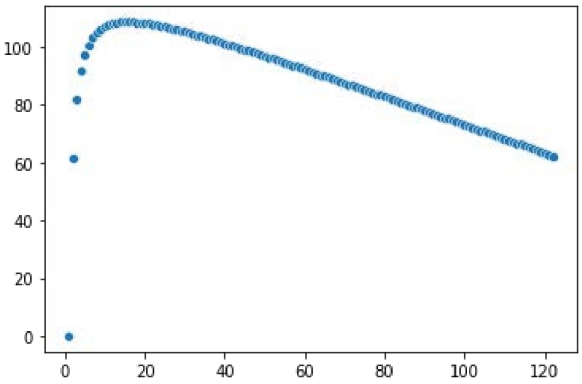

Рис. 2. Зависимость среднего количества информации от мощности алфавита при равновероятном распределении и больших q, а также при q, стремящемся от единицы к нулю

СПИСОК ЛИТЕРАТУРЫ

1. *Карпов А. А.*, *Миронкин В. О.*, *Михайлов М. М.* Об энтропийных характеристиках последовательной процедуры опробования элементов полиномиальной схемы. — Обозрение прикл. и промышл. матем., 2021, т. 28, в. 1, с. 10–14. // *Karpov A. A.*, *Mironkin V. O.*, *Mikhailov M. M.* On the entropy characteristics of a sequential procedure for testing elements of a polynomial scheme. — OPPM Surv. Appl. Ind. Math., Moscow, 2021, v. 28, is. 1, p. 10–14. (In Russian.)

2. *Духин А. А.,* Теория информации: Учебное пособие, М: Гелиос АРВ, 2007, 248 с. // *Dukhin A. A.,* Information Theory: A Textbook, Moscow: Gelios ARV, 2007, 248 pp. (In Russian.)

3. *Арбеков И. М.* Критерии секретности ключа. — Математические вопросы криптографии, 2016, т. 7, в. 1 с. 39–56. // *Arbekov I. M.,* Criteria of key security. — Mathematical Aspects of Cryptography, 2016, v. 7, is. 1, p. 39–56. (In Russian.)

4. *Арбеков И. М.* Нижние оценки для практической секретности ключа. — Математические вопросы криптографии, 2017, т. 8, в. 2, с. 29–38. // *Arbekov I. M.,* Lower bounds for the practical secrecy of a key. — Mathematical Aspects of Cryptography, 2017, v. 8, is. 2, p. 29–38. (in Russian.)

UDC DOI https://doi.org/10.52513/08698325_2022_29_1_1

Shelkanova E. A. (Moscow, MIEM, National Research University Higher School of Economics)**. On the application of Mallows inequality in testing numerical sequences.**

Abstract: The work is devoted to the study of information-theoretic properties of algorithms for testing elements of an arbitrary discrete set before the onset of «success», as well as its probabilistic modification – the truncated testing algorithm.

Keywords: Sequential testing, truncated testing, polynomial distribution, Shannon entropy, Renyi entropy, Tsallis entropy, amount of information.