ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 28 МАТЕМАТИКИ Выпуск 1
2021

А. А. Карпов [1], **В. О. Миронкин** [2], **М. М. Михайлов** [1] (Москва, [1] Лаборатория ТВП, [2] Национальный исследовательский университет «Высшая школа экономики»). **Об энтропийных характеристиках последовательной процедуры опробования элементов полиномиальной схемы.**

УДК 519.722

Резюме: Получены явные формулы для ряда энтропийных характеристик последовательной процедуры опробования элементов произвольной полиномиальной вероятностной схемы.

Ключевые слова: полиномиальное распределение, энтропия Шеннона, количество информации, последовательное опробование.

Введение. Настоящая работа посвящена задаче оценивания энтропии алгоритма опробования до «успеха», а также ряда его модификаций, и продолжает исследования, начатые в [1]. Текущие результаты обобщают ранее полученные оценки энтропии на случай произвольного полиномиального распределения на опробуемом множестве, а также на случай не единичной вероятности «успеха».

1. Вероятностная модель алгоритма опробования. Пусть для произвольного $n \in \mathbb{N}$ задано некоторое конечное множество $\mathcal{A} = \{\omega_1, \ldots, \omega_n\}$, и на его основе построена невырожденная вероятностная схема $\mathcal{A}^{(1)} = (\mathcal{A}, \overline{p})$ [2]:

$$\mathcal{A}^{(1)} = \begin{pmatrix} \omega_1 & \omega_2 & \ldots & \omega_n \\ p_1 & p_2 & \ldots & p_n \end{pmatrix}, \tag{1}$$

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

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

Пусть далее в соответствии с (2) выбирается некоторый произвольный $\omega \in \mathcal{A}$ и рассматривается классическая задача опробования элементов множества $\mathcal{A}$, упорядоченных по невозрастанию их вероятностей, до наступления «успеха», где под «успехом» понимается определение элемента ω на основе некоторого решающего правила.

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

Отметим, что при опробовании первого элемента ω_1 фактически реализуется исход вероятностной схемы

$$\mathcal{B}^{(1)} = \begin{pmatrix} \omega_1 & \overline{\omega}_1 \\ p_1 & 1 - p_1 \end{pmatrix}.$$

Таким образом, среднее количество информации, полученное при реализации первого шага алгоритма опробования до «успеха», составляет

$$H^{(1)} = \overline{H}^{(1)} = H\left(p_1, 1 - p_1\right),$$

где $H\left(p_1, 1 - p_1\right)$ — энтропия Шеннона дискретной вероятностной схемы $\mathcal{B}^{(1)}$.

Если же $\omega \neq \omega_1$, то процесс опробования продолжается. При этом распределение на оставшемся множестве пересчитывается с использованием формулы Байеса:

$$\mathcal{A}^{(2)} = \begin{pmatrix} \omega_2 & \omega_3 & \dots & \omega_n \\ \frac{p_2}{1-p_1} & \frac{p_3}{1-p_1} & \dots & \frac{p_n}{1-p_1} \end{pmatrix}.$$

Соответственно, при опробовании второго элемента ω_2 уже реализуется исход вероятностной схемы

$$\mathcal{B}^{(2)} = \begin{pmatrix} \omega_2 & \overline{\omega}_2 \\ \frac{p_2}{1-p_1} & 1 - \frac{p_2}{1-p_1} \end{pmatrix}.$$

В этом случае среднее количество информации, полученное при выполнении второго шага алгоритма, составляет

$$H^{(2)} = H\left(\frac{p_2}{1-p_1}, 1 - \frac{p_2}{1-p_1}\right),$$

и, следовательно,

$$\overline{H}^{(2)} = H\left(p_1, 1-p_1\right) + H\left(\frac{p_2}{1-p_1}, 1 - \frac{p_2}{1-p_1}\right).$$

Далее для произвольного $l \in \{1, \dots, n\}$ положим $\pi_l = \sum\limits_{i=1}^{l} p_i$ и $\pi_0 = 0$. Продолжая указанные рассуждая, для произвольного фиксированного $k \in \{1, 2, \dots, n\}$ получим следующие формулы:

$$H^{(k)} = H\left(\frac{p_k}{1-\pi_{k-1}}, 1 - \frac{p_k}{1-\pi_{k-1}}\right),$$

$$\overline{H}^{(k)} = \sum_{i=1}^{k} H\left(\frac{p_i}{1-\pi_{i-1}}, 1 - \frac{p_i}{1-\pi_{i-1}}\right),$$

описывающие теоретико-информационные свойства алгоритма опробования с вероятностью «успеха» 1.

Использование формулы Байеса позволяет выписать аналогичные формулы для алгоритмов усеченного опробования с вероятностью «успеха» π_l, где $0 < \pi_l \leqslant 1$. Так, для произвольных фиксированных $l \in \{1, \dots, n\}$ и $k \in \{1, \dots, l\}$ имеем

$$H_l^{(k)} = H\left(\frac{p_k}{\pi_l - \pi_{k-1}}, 1 - \frac{p_k}{\pi_l - \pi_{k-1}}\right),$$

$$\overline{H}_l^{(k)} = \sum_{i=1}^{k} H\left(\frac{p_i}{\pi_l - \pi_{i-1}}, 1 - \frac{p_i}{\pi_l - \pi_{i-1}}\right),$$

где величины $H_l^{(k)}$ и $\overline{H}_l^{(k)}$ определяют среднее количество информации, полученное о вероятностной схеме

$$\widehat{\mathcal{A}}_l^{(1)} = \begin{pmatrix} \omega_1 & \omega_2 & \dots & \omega_l \\ \frac{p_1}{\pi_l} & \frac{p_2}{\pi_l} & \dots & \frac{p_l}{\pi_l} \end{pmatrix}$$

на k-м шаге и соответственно за k шагов алгоритма усеченного опробования с вероятностью «успеха» π_l. При этом математическое ожидание случайной величины ξ_l с распределением

$$\xi_l \sim \begin{pmatrix} \overline{H}_l^{(1)} & \overline{H}_l^{(2)} & \dots & \overline{H}_l^{(l)} \\ \frac{p_1}{\pi_l} & \frac{p_2}{\pi_l} & \dots & \frac{p_l}{\pi_l} \end{pmatrix}$$

представляет собой среднее количество информации, полученное при реализации алгоритма усеченного опробования с вероятностью «успеха» π_l.

Основные результаты. Здесь и далее для произвольных $i_0, i_1 \in \mathbb{Z}$: $i_0 > i_1$ положим $\sum\limits_{j=i_0}^{i_1} (\cdot) \equiv 0$.

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

$$\mathbf{E}\xi_l = H\left(\frac{p_1}{\pi_l}, \frac{p_2}{\pi_l}, \ldots, \frac{p_l}{\pi_l}\right).$$

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

$$\mathbf{E}\xi_l = \log_2 l.$$

Здесь следует отметить, что в ряде криптографических приложений возникает необходимость в оценке среднего количества информации, полученного до первого определения ω с использованием усеченного алгоритма.

С этой целью рассмотрим процедуру опробования, представляющую собой схему независимых испытаний Бернулли с параметром $0 < \pi_l \leqslant 1$, где каждое отдельное испытание заключается в реализации алгоритма усеченного опробования с вероятностью «успеха» π_l переменного для каждого испытания (как случайная величина) элемента $\omega \in \mathcal{A}$.

Через τ_l обозначим случайную величину, равную среднему количеству информации, полученной при реализации указанной процедуры с использованием алгоритма усеченного опробования с вероятностью «успеха» π_l.

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

$$\mathbf{E}\tau_l = \frac{1-\pi_l}{\pi_l} \sum_{i=1}^{l} H\left(\frac{p_i}{\pi_l - \pi_{i-1}}, 1 - \frac{p_i}{\pi_l - \pi_{i-1}}\right) + H\left(\frac{p_1}{\pi_l}, \frac{p_2}{\pi_l}, \ldots, \frac{p_l}{\pi_l}\right).$$

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

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

На рис. представлен график зависимости $\mathbf{E}\tau_l$ от значений параметра l в случае равновероятного распределения на $\mathcal{A}$.

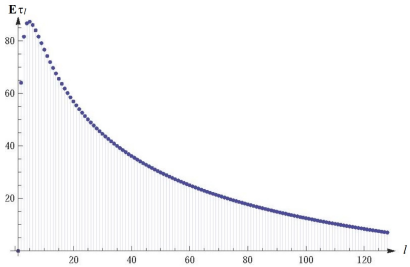

Рис. Зависимость $\mathbf{E}\tau_l$ от l при равновероятном распределении на $\mathcal{A}$

Отметим, что при $\pi_l = 1$ указанная процедура опробования в явном виде совпадает с алгоритмом опробования до «успеха» и требует для своей реализации (см. рис.) минимальное среднее количество информации об исходной вероятностной схеме, что согласуется с результатами работ [3, 4].

СПИСОК ЛИТЕРАТУРЫ

1. *Миронкин В. О.*, *Михайлов М. М.* Об энтропии последовательной процедуры опробования дискретной вероятностной схемы. — Обозрение прикл. и промышл. матем., 2020, т. 27, в. 1, с. 76–79. // *Mironkin V. O., Mikhailov M. M.* On the entropy of a sequential procedure for testing elements of a discrete probabilistic scheme. — OP&PM Surv. Appl. Ind. Math., Moscow, 2020, v. 27, is. 1, p. 76–79. (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 p. (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. *Arbekov I. M.* Lower bounds for the practical secrecy of a key. — 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 519.722

Karpov A. A. [1], ***Mironkin V. O.*** [2], ***Mikhailov M. M.*** [1] ([1] TVP Laboratory, Russia [2] National Research University Higher School of Economics, Russia)**. On the entropy characteristics of a sequential procedure for testing elements of a polynomial scheme**

Abstract: Explicit formulas are obtained for a number of entropy characteristics of a sequential procedure for testing an arbitrary polynomial probabilistic scheme.

Keywords: Polynomial scheme, Shannon entropy, information content, sequential testing.