ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 21 МАТЕМАТИКИ Выпуск 4
2014

В. О. Мирончкин (Москва, ТВП). **Вероятность t-коллизий для двух случайных подстановок.**

В настоящей работе изучается вероятностная характеристика совпадения значений двух случайных подстановок из S_N на некоторых заданных равномощных множествах $\{V_0,\dots,V_{t-1}\}$ и $\{U_0,\dots,U_{t-1}\}$, $t\in\mathbf{N}$, $V_i\in\mathbf{Z}_N$, $U_i\in\mathbf{Z}_N$, $i\in\{0,1,\dots,t-1\}$. Элементы указанных множеств формируются по следующему правилу: U_0, V_0 — произвольные элементы из $\mathbf{Z}_N$,

$$V_{i+1}=V_i+1\,(\mathrm{mod}\,N),\quad i=1,2,-2,$$

$$U_{i+1}=U_i+1\,(\mathrm{mod}\,N),\quad i=1,2,\dots,t-2.$$

Наличие двух одинаковых отрезков длины $t\in\mathbf{N}$ при реализации двух случайных подстановок $\pi_1,\pi_2\in S_N$ равносильно следующему событию:

$$\bigl\{\pi_1(V+i)=\pi_2(V'+i),\quad i=0,1,\dots,t-1\bigr\},$$

где V, V' — произвольные элементы из $\mathbf{Z}_N$.

Вычислим вероятность этого события. Для этого сформулируем ряд определений и вспомогательных лемм.

О п р е д е л е н и е 1. Будем говорить, что две подстановки $\pi_1,\pi_2\in S_N$ образуют *коллизию* длины t (подстановочную t-коллизию), где $2\leqslant t\leqslant N$, если t является максимальным числом, для которого в нижней строке обеих подстановок существуют такие вектор $(\alpha_1,\alpha_2,\dots,\alpha_t)$ и значения x и y, что справедливы равенства:

$$\begin{aligned}\pi_1(x)=\pi_2(y)&=\alpha_1,\\ \pi_1(x+1\,(\mathrm{mod}\,N))=\pi_2(y+1\,(\mathrm{mod}\,N))&=\alpha_2,\\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots&\\ \pi_1(x+t-1\,(\mathrm{mod}\,N))=\pi_2(y+t-1\,(\mathrm{mod}\,N))&=\alpha_t.\end{aligned}\tag{1}$$

Зададим на множестве подстановок S_N равномерное распределение и рассмотрим две случайные подстановки $\pi_1,\pi_2\in S_N$.

Обозначим через $A_{\pi_1,\pi_2}(N,t)$ событие, заключающееся в том, что подстановки π_1 и π_2 образуют коллизию длины $t\in\{2,3,\dots,N\}$.

Далее будем использовать представление подстановок в двустрочной записи. Подстановки π_1, π_2 можно представить в следующем виде:

$$\pi_1=\begin{pmatrix}0&1&\dots&N-1\\ x_0&x_1&\dots&x_{N-1}\end{pmatrix}\quad\text{и}\quad\pi_2=\begin{pmatrix}0&1&\dots&N-1\\ y_0&y_1&\dots&y_{N-1}\end{pmatrix}.$$

Обозначим $[i]_N\equiv i(\mathrm{mod}\,N)$, $i=1,2,\dots,N-1$, тогда событие $A_{\pi_1,\pi_2}(N,t)$ можно представить следующим образом:

$$A_{\pi_1,\pi_2}(N,t)=\left\{\begin{array}{c}\exists i,j:\ \text{в }\pi_1\text{ и }\pi_2\ \exists\text{ блоки макс. длины }t\\ \begin{pmatrix}[i]_N&[i+1]_N&\dots&[i+t-1]_N\\ \dots&\dots&\dots&\dots\\ x_{[i]_N}&x_N&\dots&x_{[i+t-1]_N}\end{pmatrix},\begin{pmatrix}[j]_N&[j+1]_N&\dots&[j+t-1]_N\\ \dots&\dots&\dots&\dots\\ x_{[i]_N}&x_{[i+1]_N}&\dots&x_{[i+t-1]_N}\end{pmatrix}\end{array}\right\}.$$

З а м е ч а н и е. В силу последнего соотношения с учетом того, что элементы верхних строк подстановок π_1, π_2 берутся по модулю N, можно рассматривать только нижние строки подстановок, считая множества $\{x_0, x_1, \ldots, x_N\}$ и $\{y_0, y_1, y_{N-1}\}$ окружностями, соответствующими подстановкам π_1, π_2.

О п р е д е л е н и е 2. *Окружностью* длины N, соответствующую произвольной подстановке $\pi \in S_N$, назовем упорядоченный набор $(x_0, x_1, \ldots, x_{N-1})$, составленный из элементов нижней строки этой подстановки.

О п р е д е л е н и е 3. Дугой $D_i(t)$ длины $1 \leqslant t \leqslant N$ с началом в точке x_i, $x_i = \pi(i)$, $i \in \{0, 1, \ldots, N-1\}$, на окружности $\pi \in S_N$ назовем упорядоченный набор из t элементов $(x_{[i]_N}, x_{[i+1]_N}, \ldots, x_{[i+t-1]_N})$.

О п р е д е л е н и е 4. Будем говорить, что для двух произвольных подстановок $\pi_1, \pi_2 \in S_N$ соответствующие дуги $D_i(t)$ и $D_j(t)$, где $i, j \in \{0, 1, \ldots, N-1\}$, совпадают, если $x_{[i+k]_N} = y_{[j+k]_N}$, $k = \{0, 1, \ldots, t-1\}$.

О п р е д е л е н и е 5. Расстоянием между дугами $D_i(t_1)$ и $D_j(t_2)$, $i < j$, произвольной подстановки $\pi \in S_N$ назовем величину

$$d_{i,j}^{(t_1)} = \begin{cases} j - i - t_1, & j - i - t_1 > 0 \\ 0, & j - i - t_1 \leqslant 0 \end{cases}.$$

Будем говорить, что дуги $D_i(t_1)$ и $D_j(t_2)$ пересекаются, если $d_{i,j}^{(t_1)} \leqslant 0$.

Согласно сделанному замечанию и введенным определениям имеет место соотношение:

$$A_{\pi_1,\pi_2}(N,t) = \left\{ \begin{array}{c} \text{на окружностях подстановок } \pi_1 \text{ и } \pi_2 \,\exists \text{ одинаковые} \\ \text{непересекающиеся дуги максимальной длины } t \end{array} \right\}. \qquad (3)$$

Исследуем вопрос о максимально возможном количестве одинаковых непересекающихся дуг длины $t \in \{2, 3, \ldots, N\}$ на окружностях произвольных подстановок $\pi_1, \pi_2 \in S_N$, при котором выполняется соотношение (3).

Лемма 1. *Пусть $t \in \{2, 3, \ldots, N-1\}$ и $\pi \in S_N$ — одна из двух подстановок, удовлетворяющих* (3), *тогда на соответствующей ей окружности может одновременно существовать не более $\alpha = [N/t + 1]$ непересекающихся дуг.*

З а м е ч а н и е. Если $t = N$ в условиях леммы 1, то существует ровно одна такая дуга длины N, содержащая все точки рассматриваемой окружности.

Лемма 2. *Пусть $\pi \in S_N$ есть одна из двух подстановок, удовлетворяющих* (3), *тогда вероятность $P_{t,l,N}^{(\pi)}$ разместить $l \in \{1, 2, \ldots, \alpha$ непересекающихся дуг длины t на соответствующей ей окружности равна*

$$P_{t,l,N}^{(\pi)} = \frac{1}{N^l} \left(\begin{array}{c} N - lt - 1 \\ N - l\,(t+1) \end{array} \right), \quad t < N, \qquad P_{N,l,N}^{(\pi)} = \frac{1}{N^l}, \quad t = N.$$

Следствие. *При выполнении условия леммы 2 справедливо неравенство*

$$P_{t,l,N}^{(\pi)} \geqslant \frac{1}{N(l-1)!} \left(1 - \frac{(l-1)lt}{N} - \frac{(l-1)l}{2N} \right), \quad t < N.$$

Лемма 3. *Пусть $\pi_1, \pi_2 \in S_N$, и пусть на соответствующих им окружностях выделено по $l \in \{1, 2, \ldots, \alpha\}$ непересекающихся дуг длины $t < N$. Тогда вероятность совпадения этих дуг равна $l!/(N)_{lt}$.*

Обозначим $\xi_{\pi_1,\pi_2} \in \{2, 3, \ldots, N\}$ случайную величину, равную длине одинаковых блоков в нижних строках подстановок $\pi_1, \pi_2 \in S_N$.

Событие $\{\xi_{\pi_1,\pi_2} \geqslant t\}$ означает, что в нижних строках подстановок π_1, π_2 существуют одинаковые блоки длины t, или, что то же самое, одинаковые непересекающиеся дуги длины t на окружностях, соответствующих подстановкам π_1, π_2.

Согласно (3) имеет место равенство

$$\mathbf{P}\left\{A_{\pi_1,\pi_2}(N,t)\right\} = \mathbf{P}\left\{\xi_{\pi_1,\pi_2} \geqslant t\right\} - \mathbf{P}\left\{\xi_{\pi_1,\pi_2} \geqslant t+1\right\}. \qquad (4)$$

Из этого равенства, вычислив распределение случайной величины ξ_{π_1,π_2}, получим искомую вероятность. Результат следующего предложения следует из приведенных выше лемм 1–3.

Предложение. *Для случайных подстановок* $\pi_1, \pi_2 \in S_N$

$$\mathbf{P}\{\xi_{\pi_1,\pi_2} \geqslant t\} = \sum_{l=1}^{\left[\frac{N}{t+1}\right]} \frac{l!}{N^{2l}(N)_{lt}} \left(\begin{array}{c} N-lt-1 \\ N-l(t+1) \end{array} \right)^2, \quad 1 < t < N, \quad \mathbf{P}\{\xi_{\pi_1,\pi_2} = N\} = \frac{1}{N!}.$$

Обозначим $\varkappa_t$, $t \in \{2, 3, \ldots, N\}$, случайную величину, равную числу пар подстановок $\pi_1, \pi_2 \in S_N$, для которых существуют одинаковые блоки длины t, т. е. когда $\xi_{\pi_1,\pi_2} \geqslant t$.

Следствие 1. *Для случайных подстановок* $\pi_1, \pi_2 \in S_N$

$$\mathbf{E}\varkappa_t = (N!)^2 \sum_{l=1}^{\left[\frac{N}{t+1}\right]} \frac{l!}{N^{2l}(N)_{lt}} \left(\begin{array}{c} N-lt-1 \\ N-l(t+1) \end{array} \right)^2, \quad 1 < t < N, \qquad \mathbf{E}\varkappa_N = (N!)^2.$$

Таким образом, из приведенного выше предложения и соотношения (4) получаем искомое выражение для вероятности t-коллизии двух случайных подстановок.

Следствие 2. *Для случайных подстановок* $\pi_1, \pi_2 \in S_N$

$$\mathbf{P}\{A_{\pi_1,\pi_2}(N,t)\} = \sum_{l=1}^{\left[\frac{N}{t+1}\right]} \frac{\left(\begin{array}{c} N-lt-1 \\ N-l(t+1) \end{array} \right) l!}{N^{2l}(N)_{lt}} - \sum_{l=1}^{\left[\frac{N}{t+2}\right]} \frac{\left(\begin{array}{c} N-l(t+1)-1 \\ N-l(t+2) \end{array} \right) l!}{N^{2l}(N)_{l(t+1)}}, \quad t < N.$$

$$\mathbf{P}\{A_{\pi_1,\pi_2}(N,t)\} = \frac{1}{N!}, \quad t = N.$$

СПИСОК ЛИТЕРАТУРЫ

1. *Сачков В. Н.* Курс комбинаторного анализа. М.–Ижевск: РХД, 2013.