ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 22 **МАТЕМАТИКИ** **Выпуск 4**
2015

А. В. Иванов (Москва, МИЭМ НИУ ВШЭ). **О распределении минимального расстояния между номерами ячеек с некоторым свойством в равновероятной урновой схеме.**

При изучении проблем, связанных, в частности, с «парадоксом дня рождения» (см., например, [4, с. 67]), возникает ряд задач, в которых требуется знать распределение расстояний между номерами ячеек с некоторым заранее заданным свойством (или функций от расстояний). При рассмотрении подобных задач в качестве теоретико-вероятностной модели, как правило, используется обобщенная урновая схема Маркова–Пойа [6], включающая в себя различные варианты схем размещений, или урновых схем [1]. Одна из таких задач решена в настоящей заметке. Приведем ее математическую постановку.

Пусть имеется N различимых ячеек, последовательно занумерованных числами от 1 до N, в которые равновероятно, последовательно и независимо размещаются n различимых частиц без ограничений на заполнения ячеек (схема упорядоченного выбора с возвращением [3]). Вероятность попадания частицы в любую ячейку равна N^{-1}.

Введем следующее

Определение. Расстоянием $\nu_{i,j}$ между ячейками с номерами $i,j \in \{1,2,\ldots,N\}$ будем называть величину

$$\nu_{i,j} = \begin{cases} j-i, & \text{если } 1 \leqslant i \leqslant j \leqslant N, \\ j-i+N, & \text{если } 1 \leqslant j < i \leqslant N. \end{cases}$$

Подобные задачи могут интерпретироваться в терминах размещения дуг на окружности с конечным числом точек. В такой постановке изучаются длины соответствующих дуг (см, например, [2, 5]).

Предположим, что после размещения частиц имеется ровно l непустых ячеек с номерами $1 \leqslant i_1 < i_2 < \cdots < i_l \leqslant N$. Обозначим соответствующие расстояния через $\nu_1 = \nu_{i_1,i_2}, \nu_2 = \nu_{i_2,i_3}, \ldots, \nu_l = \nu_{i_l,i_1}$.

Введем в рассмотрение случайную величину $\mu(N,n)$, равную числу непустых ячеек после размещений, и положим $p_{N,n}(t) = \mathbf{P}\{\mu(N,n)=t\}, t=1,2,\ldots,\min(N,n)$.

Обозначим через $i_1, i_2, \ldots, i_{\mu(N,n)}, 1 \leqslant i_1 < i_2 < \cdots < i_{\mu(N,n)} \leqslant N$ номера соответствующих ячеек и рассмотрим расстояния $\nu_1, \nu_2, \ldots, \nu_{\mu(N,n)}$.

Введем случайную величину $\theta(\mu(N,n)) = \min_{1 \leqslant j \leqslant \mu(N,n)} \nu_j$. Для распределения случайной величины $\theta(\mu(N,n))$ верно следующее

Утверждение 1. *Для произвольного* $m = 1,2,\ldots,N-1$

$$\mathbf{P}\{\theta(N,n) \geqslant m\} = \sum_{t=2}^{\min\left(n,\left[\frac{N}{m}\right]\right)} \left[\left(\frac{t}{N}\right)^{n-1} \binom{N-(m-1)t-1}{t-1} \right.$$

$$\times\left(\sum_{j=0}^{t}(-1)^j\binom{N-t}{t}\left(1-\frac{j}{N-t}\right)^n\right)\Bigg],$$

где $[\alpha]$ *— целая часть числа* α *.*

С помощью полученной формулы можно найти вероятность

$$\mathbf{P}\big\{\theta(N,n)=m\big\}=\mathbf{P}\big\{\theta(N,n)\geqslant m\big\}-\mathbf{P}\big\{\theta(N,n)\geqslant m+1\big\}.$$

Сформулируем данный результат в виде следствия.

Следствие 1. *Для произвольного* $m=1,2,\ldots,N-1$

$$\mathbf{P}\{\theta(N,n)=m\}=\sum_{l=m}^{m+1}(-1)^{m+2-l}\times\sum_{t=2}^{\min\left(n,\left[\frac{N}{l}\right]\right)}\left[\left(\frac{t}{N}\right)^{n-1}\binom{N-(m-1)t-1}{t-1}\times\right.$$

$$\left.\times\left(\sum_{j=0}^{t}(-1)^j\binom{N-t}{t}\left(1-\frac{j}{N-t}\right)^n\right)\right].$$

В практических приложениях, существенно использующих результаты, так или иначе связанных с классическим «парадоксом дня рождения» и различных его модификаций и обобщений, иногда требуется найти распределение минимального расстояния между ячейками, каждая из которых содержит по 2 частицы.

Пусть ξ_N — случайная величина, равная числу ячеек, каждая из которых содержит по 2 частицы. Положим $p_N(t)=p_N(2,t)=\mathbf{P}\{\xi_N=t\}$, где $t=0,1,\ldots,\min\left(\left[\frac{n}{2}\right],N\right)$.

После размещения n частиц обозначим через

$$i_1,i_2,\ldots,i_{\xi_N};\quad 1\leqslant i_1<i_2<\cdots<i_{\xi_N}\leqslant N$$

номера ячеек, каждая из которых содержит по 2 частицы, и рассмотрим расстояния $\nu_1,\nu_2,\ldots,\nu_{\xi_N}$. Определим случайную величину $\theta(\xi_N)=\min_{1\leqslant j\leqslant \xi_N}\nu_j$.

Для распределения случайной величины $\theta(\xi_N)$ верно следующее

Утверждение 2. *Для произвольного* $m=1,2,\ldots,N-1$

$$\mathbf{P}\{\theta(\xi_N)\geqslant m\}=\sum_{t=2}^{\left[\frac{n}{2}\right]}\Big[\prod_{j=1}^{t-1}\Big(\frac{N-mt+j}{N-t+j}\Big)\binom{N}{t}\frac{n^{[2t]}}{2^tN^{2t}}\Big(1-\frac{t}{N}\Big)^{n-2t}\times$$

$$\times\sum_{j=0}^{N-t-\left[\frac{n}{2}\right]-1}(-1)^j\binom{N-t}{j}\frac{(n-2t)^{[2j]}}{2^j(N-t)^{2j}}\Big(1-\frac{j}{N-t}\Big)^{n-2(t+j)}\Big],$$

где $k^{[l]}=\prod_{j=1}^{l-1}(k-j),k\geqslant l$ *— обобщенная факториальная степень.*

С использованием утверждения 2 нетрудно получить

Следствие 2. *Для произвольного* $m=1,2,\ldots,N-1$

$$\mathbf{P}\{\theta(\xi_N)=m\}=\sum_{l=m}^{m+1}\left[(-1)^{m+2-l}\sum_{t=2}^{\left[\frac{n}{2}\right]}\left[\prod_{j=1}^{t-1}\left(\frac{N-mt+j}{N-t+j}\right)\binom{N}{t}\frac{n^{[2t]}}{2^tN^{2t}}\Big(1-\frac{t}{N}\Big)^{n-2t}\right.\right.$$

$$\left.\left.\times\sum_{j=0}^{N-t-\left[\frac{n}{2}\right]-1}(-1)^j\binom{N-t}{j}\frac{(n-2t)^{[2j]}}{2^j(N-t)^{2j}}\Big(1-\frac{j}{N-t}\Big)^{n-2(t+j)}\right]\right].$$

Отметим, что асимптотические распределения числа ячеек с определенным свойством получены в работах [1, 3], а также в приложении «Appendix B» к диссертации [7, p. 275].

В дальнейшем планируется найти асимптотику полученных в работе вероятностей при различном характере изменения n,N , а также рассмотреть аналогичные вероятности для ячеек с другими свойствами.

СПИСОК ЛИТЕРАТУРЫ

1. *Иванов В. А.*, *Ивченко Г. И.*, *Медведев Ю. И.* Дискретные задачи в теории вероятностей. — В сб.: Теория вероятностей. Математическая статистика. Теоретическая кибернетика. Т. 22 (Итоги науки и техники ВИНИТИ АН СССР). М.: ВИНИТИ, 1984, с. 3–60.
2. *Ивченко Г. И.* Предельные теоремы в дискретной задаче покрытия окружности. — В сб.: Вероятностные задачи дискретной математики. М.: МИЭМ, 1990, с. 27–34.
3. *Колчин В. Ф.*, *Севастьянов Б. А.*, *Чистяков В. П.* Случайные размещения. М.: Наука, 1976, 224 с.
4. *Секей Г.* Парадоксы в теории вероятностей и математической статистике. М.: МИР, 1990, 240 с.
5. *Сачков В. Н.* Курс комбинаторного анализа. М.–Ижевск: НИЦ «Регулярная и хаотическая динамика», 2013, 336 с.
6. *Ivanov A. V.*, *Ivchenko G. I.* On waiting time in the Markov-Polya scheme. — J. mathematical sciences, 1998, v. 91, № 3, p. 2904–2916.
7. *Preenel B.* Analysis and Design of Cryptographic Hash Functions. PhD, 2003, 324 p.