А.М.ЗУБКОВ, О.К.ШИБАНОВ

ПУАССОНОВСКАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА ДЛЯ ДВУХЭТАПНОЙ РАВНОВЕРОЯТНОЙ СХЕМЫ РАЗМЕЩЕНИЯ ЧАСТИЦ ПО ЯЧЕЙКАМ

1. Введение

Схема случайного размещения частиц по ячейкам является удобной математической моделью различных реальных процессов в физике, технике, биологии и т.п. Различные ее варианты рассматриваются в математической статистике. Предельные теоремы для ряда вариантов схемы случайного размещения частиц по ячейкам можно найти в книге [1]. В обычной схеме равновероятного размещения частиц по ячейкам предполагается, что частицы независимо одна от другой размещаются по N ячейкам так, что вероятность попадания k-й частицы в j-ю ячейку равна $1/N$ для любых $k = 1, \ldots, n$ и $j = 1, \ldots, N$. Пусть $\eta_j = \eta_j(n, N)$ – число частиц, попавших в j-ю ячейку после размещения n частиц, и $\mu_r(n, N)$ – число ячеек, в которые попало ровно r частиц. В [1] доказано, в частности, что если $r > 1$ – константа, а $n, N \to \infty$ так, что

$$\frac{1}{r!}\,\frac{n^r}{N^{r-1}}\,e^{-n/N} \to \lambda \in (0, \infty),$$

то

$$\mathbf{P}(\mu_r(n, N) = k) \to \frac{\lambda^k}{k}\,e^{-\lambda}, \quad k = 0, 1, \ldots$$

Приведенная формулировка охватывает два случая: левую область (когда $n/N \to 0$) и правую область (когда $n/N \to \infty$).

В настоящей работе мы рассматриваем более сложную (двухэтапную, или двуслойную) схему равновероятного размещения частиц по ячейкам. Схема задается не двумя, а тремя параметрами, которые мы будем обозначать N_0, N_1 и N_2. На первом этапе N_0 исходных частиц независимо и равновероятно размещаются по N_1 ячейкам первого слоя. На втором этапе N_1 ячеек первого слоя рассматриваются как частицы, и они независимо и равновероятно размещаются по N_2 ячейкам второго слоя вместе с содержащимися в них исходными частицами. Обозначим через $\eta_j^{(i)} = \eta_j^{(i)}(N_0, N_1, N_2)$ число исходных частиц, попавших в j-ю ячейку i-го слоя, и через $\mu_r^{(i)}(N_0, N_1, N_2) = \mu_r^{(i)}$ – число ячеек i-го слоя, в которые попало ровно r исходных частиц.

Теорема. *Если в равновероятной двухэтапной схеме размещения частиц $r > 0$ фиксировано, $N_0, N_1, N_2 \to \infty$ так, что $N_0 = o(N_2)$, $N_0 = O(N_1)$, и*

$$\mathbf{E}\mu_r^{(2)} \to \lambda \in (0, \infty),$$

то

$$\mathbf{P}(\mu_r^{(2)} = k) \to \frac{\lambda^k}{k!}\, e^{-\lambda}, \qquad k = 0, 1, \ldots$$

Результаты анонсировались в [1] – [2].

Прежде чем переходить к доказательству, отметим одну интерпретацию двухэтапной схемы равновероятного размещения частиц по ячейкам, демонстрирующую ее аналогию с процедурами, используемыми при построении оценок по методу бутстрэпа.

Сначала по обычной равновероятной схеме разместим N_1 частиц по N_2 ячейкам; пусть $\eta_j(N_1, N_2)$ – число частиц, попавших в j-ю ($j = 1, \ldots, N_2$) ячейку. Отношения $\pi_j = \frac{\eta_j(N_1,N_2)}{N_1}, j = 1, \ldots, N_2$, образуют случайное распределение вероятностей. Теперь проведем размещение N_0 частиц по N_2 ячейкам в соответствии с распределением π_j. Нетрудно убедиться в том, что описанная схема эквивалентна двухэтапной равновероятной схеме размещения частиц. Действительно, схему размещения, в которой (случайные) вероятности попадания в ячейки выбираются в соответствии с вышеприведенной процедурой, можно описать иначе следующим образом. Первый этап размещения N_1 частиц по N_2 ячейкам соответствует случайному разбиению N_1 ячеек на N_2 подмножеств $M_1, \ldots, M_{N_2}$, содержащих $\eta_1(N_1, N_2), \ldots, \eta_{N_2}(N_1, N_2)$ элементов соответственно. Если после этого разбиения провести равновероятное размещение N_0 частиц по N_1 ячейкам, то вероятность попадания частицы в множество M_j (и, следовательно, в j-ю ячейку второго слоя) будет равна $\frac{\eta_j(N_1,N_2)}{N_1}$.

Всюду в работе $c, c_0, \ldots$ обозначают некоторые положительные константы. Случайная величина $\theta_j = \theta_j(r)$, $j = 1, \ldots, N_2$, – индикатор события, состоящего в том, что j-ю ячейку второго слоя попало ровно r частиц; таким образом,

$$\mu_r^{(2)} = \sum_{j=1}^{N_2} \theta_j. \qquad (1)$$

Для натурального $r \geq 2$ будем использовать обозначение
$K(r) = \{\mathbf{k} = (k_1, \ldots, k_r) \in Z_0^r : k_1 + 2k_2 + \ldots + rk_r = r\}$, где
$Z_0 = \{0, 1, \ldots\}$. Число элементов множества $K(r)$ не превосходит числа целых неотрицательных решений уравнения $m_1 + m_2 + \ldots + m_r = r$, которое равно C_{2r-1}^r.

2. Пуассоновские предельные распределения $\mu_r^{(2)}$

Нетрудно проверить, что

$$\{\eta_j^{(2)} = r\} = \bigcup_{\mathbf{k} \in K(r)} C_j(\mathbf{k}),$$

где

$$C_j(\mathbf{k}) = \{\text{в } j\text{-ю ячейку второго слоя попало ровно } k_i \text{ ячеек первого слоя, содержащих ровно } i \text{ частиц, } i = 1, \ldots, r\}.$$

Пусть $\overline{\mu}_1^{(1)} = \mu_1^{(1)} + \ldots + \mu_{N_0}^{(1)}$ – количество непустых ячеек первого слоя. Очевидно, что $0 \leq \overline{\mu}_1^{(1)} \leq \min(N_0, N_1)$.

В лемме 1 приводятся асимптотические соотношения для $\mathbf{E}\mu_r^{(2)}$ при некоторых способах изменения параметров N_0, N_1, N_2.

Лемма 1. *В равновероятной двухэтапной схеме при любом $r \geq 2$*

$$\mathbf{E}\mu_r^{(2)} = \frac{N_0^r}{r! N_1^{r-1}} e^{-\frac{N_0}{N_1}} \left(1 + O\left(\frac{N_1}{N_2} + \frac{1}{N_0} + \frac{1}{N_1}\right)\right),$$

если $N_0 \to \infty$, $N_0 = O(N_1)$, $N_1 = O(N_2)$, а если $N_0 \to \infty$, $N_0 = o(\min\{N_1, N_2\})$, $N_2 = O(N_1)$, то

$$\mathbf{E}\mu_r^{(2)} = \frac{N_0^r}{r! N_2^{r-1}} \left(1 + O\left(\frac{N_0}{N_2} + \frac{1}{N_0} + \frac{N_2}{N_1}\right)\right).$$

Доказательство леммы 1. Для равновероятной двухэтапной схемы, очевидно, индикаторы $\theta_1, \ldots, \theta_{N_2}$ одинаково распределены; следовательно,

$$\mathbf{E}\mu_r^{(2)} = \sum_{i=1}^{N_2} \mathbf{P}(\theta_i = 1) = N_2 \mathbf{P}(\theta_1 = 1). \tag{2}$$

В силу равенства (2) достаточно найти асимптотику $\mathbf{P}(\theta_1 = 1)$. Пусть $\mathbf{k}_0 = (0, \ldots, 0, 1)$, тогда

$$\mathbf{P}(\theta_1 = 1) = \mathbf{P}(C_1(\mathbf{k}_0)) + \sum_{\mathbf{k} \in K(r) \setminus \{\mathbf{k}_0\}} \mathbf{P}(C_1(\mathbf{k})).$$

Вероятность отдельного события равна

$$\mathbf{P}(C_1(\mathbf{k})) = \frac{\mathbf{E}(\mu_1^{(1)})^{[k_1]} \ldots (\mu_r^{(1)})^{[k_r]}}{k_1! \ldots k_r! N_2^{k_1 + \ldots + k_r}} \left(1 - \frac{1}{N_2}\right)^{\overline{\mu}_1^{(1)} - k_1 - \ldots - k_r}. \tag{3}$$

Последний множитель (с учетом неравенства $k_1 + \ldots + k_r \leq \overline{\mu_1^{(1)}} \leq \min(N_0, N_1)$) можно оценить так:

$$1 \geq \left(1 - \frac{1}{N_2}\right)^{\overline{\mu}_1^{(1)} - k_1 - \ldots - k_r} \geq 1 - \frac{\min(N_0, N_1) - k_1 - \ldots - k_r}{N_2}. \tag{4}$$

Согласно [3, стр.45] при $k_1 + \ldots + rk_r = r$

$$\mathbf{E}(\mu_1^{(1)})^{[k_1]} \ldots (\mu_r^{(1)})^{[k_r]} = \frac{N_0^{[r]} N_1^{[k_1 + \ldots + k_r]}}{N_1^r (1!)^{k_1} \ldots (r!)^{k_r}} \left(1 - \frac{k_1 + \ldots + k_r}{N_1}\right)^{N_0 - r}. \tag{5}$$

Так как $-x - x^2 < \ln(1 - x) < -x$ при $0 < x < \frac{1}{2}$, то

$$e^{-\frac{kN_0}{N_1}} e^{-\frac{kN_0}{N_1^2}} < \left(1 - \frac{k}{N_1}\right)^{N_0 - r} < e^{-\frac{k(N_0 - r)}{N_1}}$$

при $N_1 > 2k$. Заменяя факториальные степени обычными и учитывая условия леммы 1, находим

$$\mathbf{P}(C_1(\mathbf{k}_0)) = \frac{N_0^r}{r! N_1^{r-1} N_2} e^{-\frac{N_0}{N_1}} \left(1 + O\left(\frac{\min(N_0, N_1)}{N_2} + \frac{1}{N_0}\right)\right).$$

С другой стороны, если $\mathbf{k} \in K(r)\backslash\{\mathbf{k}_0\}$, то $2 \le k_1+\ldots+k_r \le r$; учитывая (3) и (5), находим:

$$\mathbf{P}(C_1(\mathbf{k})) = \frac{N_0^{[r]} N_1^{[k_1+\ldots+k_r]}}{N_1^r (1!)^{k_1} \ldots (r!)^{k_r} k_1! \ldots k_r! N_2^{k_1+\ldots+k_r}} \times$$

$$\times \left(1 - \frac{k_1+\ldots+k_r}{N_1}\right)^{N_0-r} \left(1 - \frac{1}{N_2}\right)^{\overline{\mu}_1^{(1)}-k_1-\ldots-k_r} \le$$

$$\le \frac{N_0^r}{N_2^{k_1+\ldots+k_r} N_1^{r-k_1-\ldots-k_r}}\, e^{-\frac{(k_1+\ldots+k_r)(N_0-r)}{N_1}} =$$

$$= O\left(\frac{N_0^r}{N_2^2 N_1^{r-2}}\, e^{-\frac{2N_0}{N_1}}\right) = O\left(\frac{N_1}{N_2}\mathbf{P}(C_1(\mathbf{k}_0))\right),$$

и в силу конечности множества $K(r)$ при условиях $N_0 = O(N_1)$, $N_1 = O(N_2)$

$$\mathbf{E}\mu_r^{(2)} = N_2 \mathbf{P}(\theta_1 = 1) = \frac{N_0^r}{r! N_1^{r-1}}\, e^{-\frac{N_0}{N_1}} \left(1 + O\left(\frac{N_0+N_1}{N_2} + \frac{1}{N_0} + \frac{1}{N_1}\right)\right).$$

Для доказательства второй формулы положим $\mathbf{k}_1 = (r, 0, \ldots, 0)$; тогда

$$\mathbf{P}(\theta_1 = 1) = \mathbf{P}(C_1(\mathbf{k}_1)) + \sum\nolimits_{\mathbf{k} \in K(r)\backslash\{\mathbf{k}_1\}} \mathbf{P}(C_1(\mathbf{k})).$$

Первое слагаемое есть

$$\mathbf{P}(C_1(\mathbf{k}_1)) = \frac{N_0^r}{r! N_2^r}\left(1 + O\left(\frac{N_0}{N_2} + \frac{N_0}{N_1} + \frac{1}{N_0}\right)\right).$$

Если $\mathbf{k} \in K(r)\backslash\{\mathbf{k}_1\}$, то $1 \le k_1 + \ldots + k_r \le r-1$ и при $N_2 = O(N_1)$

$$\mathbf{P}(C_1(\mathbf{k})) \le \frac{N_0^r}{N_2^{k_1+\ldots+k_r} N_1^{r-k_1-\ldots-k_r}} \le \frac{N_0^r}{N_1 N_2^{r-1}} = O\left(\frac{N_2}{N_1}\,\mathbf{P}(C_1(\mathbf{k}_1))\right).$$

Так же, как в первом случае, из последних оценок с учетом соотношения $N_0 = O(N_2)$ следует, что

$$\mathbf{E}\mu_r^{(2)} = N_2 \mathbf{P}(C_1(\mathbf{k}_1))\left(1 + O\left(\frac{N_2}{N_1}\right)\right) = \frac{N_0^r}{r! N_2^{r-1}}\left(1 + O\left(\frac{N_0}{N_2} + \frac{1}{N_0} + \frac{N_2}{N_1}\right)\right).$$

Тем самым лемма 1 доказана.

Лемма 2. *В равновероятной двухэтапной схеме размещения частиц для любых* $r \in \{2, 3, \ldots\}$, $m \in \{1, 2, \ldots\}$ *при* $N_0, N_1, N_2 \to \infty$, $N_0 = o(N_2)$, $N_0 = O(N_1)$

$$\mathbf{E}\mu_r^{[m]} = (\mathbf{E}\mu_r)^m \left(1 + O\left(\frac{N_0}{N_2} + \frac{1}{N_0} + \frac{1}{N_1}\right)\right).$$

Доказательство леммы 2. Как в лемме 1,

$$\mathbf{E}\mu_r^{(2)} = N_2 \mathbf{P}(\theta_1 = 1),$$

$$\mathbf{E}(\mu_r^{(2)})^{[m]} = N_2^{[m]} \mathbf{P}(\theta_1 = \ldots = \theta_m = 1),$$
$$\mathbf{P}(\theta_1 = 1) = \sum_{\mathbf{k} \in K(r)} \mathbf{P}(C_1(\mathbf{k})),$$

и

$$\mathbf{P}(\theta_1 = \ldots = \theta_m = 1) = \sum_{\mathbf{k}_1, \ldots, \mathbf{k}_m \in K(r)} \mathbf{P}(C_1(\mathbf{k}_1) \ldots C_m(\mathbf{k}_m)).$$

Поэтому утверждение леммы следует из равенства

$$\mathbf{P}(C_1(\mathbf{k}_1) \ldots C_m(\mathbf{k}_m)) = \tag{6}$$

$$= \mathbf{P}(C_1(\mathbf{k}_1)) \ldots \mathbf{P}(C_m(\mathbf{k}_m)) \left(1 + O\left(\frac{N_0}{N_2} + \frac{1}{N_0} + \frac{1}{N_1}\right)\right),$$

где $\mathbf{k}_i = (k_{i1}, \ldots, k_{ir})$, $k_{i1} + 2k_{i2} + \ldots + rk_{ir} = r$, $i = 1, \ldots, m$.

Докажем (6). Условная вероятность события $C_1(\mathbf{k}_1) \ldots C_m(\mathbf{k}_m)$ при условии $\mu_1 = v_1, ..., \mu_r = v_r$ равна

$$\frac{v_1^{[\kappa_1]} \ldots v_r^{[\kappa_r]}}{k_{11}! \ldots k_{mr}! N_2^{\kappa_1 + \ldots + \kappa_r}} \left(1 - \frac{m}{N_2}\right)^{\overline{\mu}_1^{(1)} - \kappa_1 - \ldots - \kappa_r},$$

где $\kappa_j = k_{1j} + \ldots + k_{mj}, j = 1, \ldots, r$. Поэтому вероятность в левой части (6) равна

$$\mathbf{E}\mu_1^{[\kappa_1]} \ldots \mu_r^{[\kappa_r]} \frac{1}{k_{11}! \ldots k_{mr}! N_2^{\kappa_1 + \ldots + \kappa_r}} \left(1 - \frac{m}{N_2}\right)^{\overline{\mu}_1^{(1)} - \kappa_1 - \ldots - \kappa_r}.$$

Учитывая (4) и то, что $\min(N_0, N_1) = O(N_0)$ в силу условия $N_0 = O(N_1)$, получаем

$$\mathbf{P}(C_1(\mathbf{k}_1)\dots C_m(\mathbf{k}_m)) =$$
$$= \mathbf{E}\mu_1^{[\kappa_1]}\dots\mu_r^{[\kappa_r]}\frac{1}{k_{11}!\dots k_{mr}!N_2^{\kappa_1+\dots+\kappa_r}}\left(1+O\left(\frac{N_0}{N_2}\right)\right).$$

Согласно [3, стр.45] при $\kappa_1+\dots+r\kappa_r = rm$

$$\mathbf{E}\mu_1^{[\kappa_1]}\dots\mu_r^{[\kappa_r]} = \frac{N_0^{[rm]}N_1^{[\kappa_1+\dots+\kappa_r]}}{N_1^{rm}}\frac{1}{(1!)^{\kappa_1}\dots(r!)^{\kappa_r}}\left(1-\frac{\kappa_1+\dots+\kappa_r}{N_1}\right)^{N_0-rm}.$$

Проводя те же оценки, что в лемме 1, получаем соотношение

$$\mathbf{P}\left(C_1(\mathbf{k}_1)\dots C_m(\mathbf{k}_m)\right) =$$
$$= \frac{N_0^{rm}}{k_{11}!\dots k_{mr}!N_2^{\kappa_1+\dots+\kappa_s}N_1^{rm-\kappa_1-\dots-\kappa_r}(1!)^{\kappa_1}\dots(r!)^{\kappa_r}}\times$$
$$\times e^{-\frac{(\kappa_1+\dots+\kappa_r)N_0}{N_1}}\left(1+O\left(\frac{N_0}{N_2}+\frac{1}{N_0}+\frac{1}{N_1}\right)\right). \qquad (7)$$

Аналогично преобразуем сомножители в правой части (6):

$$\mathbf{P}(C_i(\mathbf{k}_i)) = \mathbf{E}\mu_1^{[k_{i1}]}\dots\mu_r^{[k_{ir}]}\frac{1}{k_{i1}!\dots k_{ir}!N_2^{k_{i1}+\dots+k_{ir}}}\left(1+O\left(\frac{N_0}{N_2}\right)\right) =$$
$$= \frac{N_0^r}{k_{i1}!\dots k_{ir}!N_2^{k_{i1}+\dots+k_{ir}}N_1^{r-k_{i1}-\dots-k_{ir}}(1!)^{k_{i1}}\dots(r!)^{k_{ir}}}e^{-\frac{(k_{i1}+\dots+k_{ir})N_0}{N_1}}\times$$
$$\times\left(1+O\left(\frac{N_0}{N_2}+\frac{1}{N_0}+\frac{1}{N_1}\right)\right),$$

значит,

$$\mathbf{P}(C_1(\mathbf{k}_1)\dots C_m(\mathbf{k}_m)) =$$
$$= \mathbf{P}(C_1(\mathbf{k}_1))\dots\mathbf{P}(C_m(\mathbf{k}_m))\left(1+O\left(\frac{N_0}{N_2}+\frac{1}{N_0}+\frac{1}{N_1}\right)\right).$$

Отметим, что постоянные в оценках равномерны по всем событиям, так как при фиксированных r, m суммы по $\mathbf{k}_1,\dots,\mathbf{k}_m$ состоят из конечного числа слагаемых. Лемма 2 доказана.

Доказательство теоремы. Из леммы 2 следует, что при условиях теоремы
$$\mathbf{E}\mu_r^{[m]} \to \lambda^m,$$
следовательно, $\mu_r^{(2)}$ сходится к случайной величине, имеющей распределение Пуассона с параметром λ. Теорема доказана.

Литература

1. Зубков А.М., Шибанов О.К. Многоступенчатые схемы размещения частиц по ячейкам. – Обозрение прикл. и промышл. математики, 2002, т.9, вып. 1, с. 115 – 116.

2. Зубков А.М., Шибанов О.К. Двухступенчатая схема размещения частиц по ячейкам. – Обозрение прикл. и промышл. математики, 2002, т.9, вып. 2, с. 378 – 379.

3. Колчин В.Ф., Севастьянов Б.А., Чистяков В.П. Случайные размещения. – М., "Наука", 1976.