УДК 519.212.2

О времени ожидания в одной схеме размещения частиц по ячейкам

В. П. Чистяков[1]

Поступило в ноябре 2012 г.

Изучается модификация схемы размещения случайного числа частиц по ячейкам, в которой частицы размещаются в соответствии с некоторой полиномиальной схемой “порциями” из случайного числа частиц каждая. Исследованы асимптотические свойства распределения числа испытаний (размещенных “порций” частиц) до заполнения заданного числа ячеек.

DOI: 10.1134/S0371968513030072

ВВЕДЕНИЕ

Исследования поведения случайной величины $\nu(N,k)$, равной наименьшему числу испытаний равновероятной полиномиальной схемы с N исходами, при котором впервые произойдет ровно k различных исходов, приведены в книге [1], вышедшей в 1976 г. В последующие годы интенсивно развивались различные направления, обобщающие эти исследования: изучались время до превышения детерминированного уровня исходов общей полиномиальной схемы, условия сходимости распределения вектора из величин $\nu(N,[xk])$, $k=1,\dots,m$, $0<x<1$, задачи со случайными уровнями и т.д. (см. обзор [2]). В частности, для величины $\nu_m(N,k)$, равной минимальному числу испытаний, при котором какие-либо k исходов реализуются более чем m раз, в работе [3] доказана асимптотическая нормальность при $k,N\to\infty$, $\frac{k}{N}\to\lambda$, $m=\mathrm{const}$.

В работе [4] исследовалась величина X_N, равная времени до появления всех исходов из заданного подмножества $\mathfrak{N}=\{1,\dots,N\}$ исходов $\{1,2,\dots,N,N+1,\dots,M\}$ полиномиальной схемы с вероятностями исходов $\widetilde{p}_i$, $i=1,\dots,M$; были получены верхние и нижние оценки функции распределения величины X_N; была отмечена возможность использования результатов в задачах телекоммуникаций.

Отметим, что распределение величины X_N зависит только от вероятностей исходов заданного подмножества $\mathfrak{N}$. Ограничившись наблюдением только этих ячеек, можно считать, что с вероятностью $p=\widetilde{p}_1+\ldots+\widetilde{p}_N$ частица попадает в одну из ячеек подмножества $\mathfrak{N}$, а с вероятностью $1-p$ теряется (попадает в $\{N+1,\dots,M\}$). Вероятность попадания в ячейку $i\in\mathfrak{N}$ при условии, что частица не потерялась, равна $p_i=\widetilde{p}_i/p$.

1. ПОСТАНОВКА ЗАДАЧИ

Пусть $\{\delta_t\}$ — последовательность целочисленных случайных величин. В t-м испытании при $\delta_t>0$ по полиномиальной схеме с вероятностями

$$p_1,\dots,p_N,\qquad \sum_{i=1}^{N}p_i=1, \tag{1.1}$$

[1]Математический институт им. В.А. Стеклова РАН, Москва, Россия.

по ячейкам $\{1, 2, \ldots, N\}$ размещаются δ_t частиц; при $\delta_t = 0$ в данном испытании частицы не размещаются. При фиксированной последовательности $\{\delta_t\}$ размещения частиц по ячейкам в разных испытаниях независимы. Количество частиц ξ_n, размещенных по N ячейкам за n испытаний, случайно и представляется формулой

$$\xi_n = \delta_1 + \delta_2 + \ldots + \delta_n. \tag{1.2}$$

В приведенной выше схеме размещения из работы [4] величина ξ_n имеет биномиальное распределение.

Обозначим через $\nu(N, k)$, $1 \le k \le N$, наименьшее число испытаний, при котором впервые заполнятся k или более каких-либо ячеек. Отметим, что распределение приведенной выше величины X_N из работы [4] совпадает с распределением величины $\nu(N, N)$ в том случае, если $\{\delta_t\}$ является последовательностью Бернулли.

Ниже исследуется поведение случайной величины $\nu(N, k)$ при $k, N \to \infty$. При этом предполагается, что при некоторых постоянных $a > 0$, $\sigma > 0$

$$\lim_{n\to\infty} \mathbf{P}\left\{\frac{\xi_n - na}{\sigma\sqrt{n}} < x\right\} = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-u^2/2}\, du, \qquad -\infty < x < \infty. \tag{1.3}$$

Таким образом, из рассмотрения не исключаются последовательности $\{\delta_t\}$ из зависимых величин, от которых только требуется выполнение условия (1.3).

Изучение распределения величины $\nu(N, k)$ основывается на равенстве

$$\mathbf{P}\big(\nu(N, k) \le n\big) = \mathbf{P}\big(\mu_0(\xi_n, N) \le N - k\big), \tag{1.4}$$

где $\mu_0(\xi_n, N)$ — число ячеек, оставшихся пустыми после размещения случайного числа ξ_n частиц по полиномиальной схеме с вероятностями исходов (1.1). Исследование величины $\mu_0(\xi_n, N)$ проводится в трех областях изменения параметров n, N при $n, N \to \infty$:

в центральной области

$$0 < \underline{\alpha} \le \alpha_N = \frac{n}{N} \le \overline{\alpha} < \infty; \tag{1.5}$$

в правой области

$$a\alpha_N = \frac{na}{N} = \ln N - \ln \overline{\lambda} + o(1); \tag{1.6}$$

в левой области

$$\frac{n^2}{2N} \to \lambda < \infty, \tag{1.7}$$

где $\underline{\alpha}$, $\overline{\alpha}$, $\overline{\lambda}$, λ — некоторые числовые параметры.

2. ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ В ЦЕНТРАЛЬНОЙ ОБЛАСТИ

Введем обозначения

$$\begin{gathered} m(A) = \frac{1}{N}\sum_{k=1}^{N} e^{-ANp_k}, \qquad \sigma_\mu^2 = \alpha_N \left(\frac{1}{N}\sum_{k=1}^{N} Np_k e^{-ANp_k}\right)^2 \sigma^2 + B^2(A), \\ B^2(A) = \frac{1}{N}\sum_{k=1}^{N} e^{-ANp_k} - \frac{1}{N}\sum_{k=1}^{N} e^{-2ANp_k} - A\left(\frac{1}{N}\sum_{k=1}^{N} Np_k e^{-ANp_k}\right)^2, \\ A = a\alpha_N, \qquad \alpha_N = \frac{n}{N}. \end{gathered} \tag{2.1}$$

Нам понадобится следующая теорема об асимптотической нормальности величины $\mu_0(\xi_n, N)$.

Теорема 1. *Если при* $n, N \to \infty$ *выполнены условия* (1.3), (1.5) *и* $Np_k \le C < \infty$, $k = 1, \dots, N$, *то случайная величина*

$$\mu_0^*(N) = \frac{\mu_0(\xi_n, N) - Nm(A)}{\sigma_\mu \sqrt{N}} \tag{2.2}$$

асимптотически нормальна с параметрами $(0,1)$.

Доказательство этой теоремы мы опускаем, поскольку оно не отличается от доказательства аналогичной теоремы для равновероятной полиномиальной схемы (см. [1, гл. VI, § 2, теорема 1]).

Если при $n, N \to \infty$ распределение вероятностей (1.1) сближается с равномерным, то для $m(A)$ и σ_μ^2 можно выписать более простые выражения. Для статистических приложений полезно рассматривать следующие предположения о характере изменения $\{p_k\}$ при $N \to \infty$:

$$\begin{aligned} &p_k = \frac{1}{N}\left(1 + \frac{b_k(N)}{\sqrt[4]{N}}\right), \qquad |b_k(N)| \le C < \infty, \qquad k = 1, \dots, N, \\ &b_N^2 = \frac{1}{N}\sum_{k=1}^{N} b_k^2(N) \to b^2 > 0. \end{aligned} \tag{2.3}$$

Следствие 1. *Пусть выполнены условия теоремы* 1 *и условия* (2.3). *Тогда случайная величина* $\mu_0^*(N)$, *для которой*

$$\begin{gathered} m(A) = e^{-A}\left(1 + A^2\frac{b_N^2}{2\sqrt{N}}\right), \qquad B^2(A) = e^{-A}\bigl(1 - (1+A)e^{-A}\bigr), \\ \sigma_\mu^2 = \alpha_N \sigma^2 e^{-2A} + B^2(A), \qquad A = a\alpha_N, \end{gathered} \tag{2.4}$$

асимптотически нормальна с параметрами $(0,1)$.

Теорема 2. *Если при* $N \to \infty$ *выполнены условия* (1.3), (2.3) *и* $\gamma_N = \frac{k}{N} \to \gamma$, $0 < \gamma < 1$, *то случайная величина* $\nu(N,k)$ *асимптотически нормальна с математическим ожиданием*

$$\frac{N}{a}\ln\frac{1}{1-\gamma_N} + \sqrt{N} b_N^2 \frac{1}{2a}\ln^2(1-\gamma_N)$$

и дисперсией

$$N\left(\frac{1}{a^2}\frac{\gamma_N}{1-\gamma_N} - \frac{1}{a^2}\left(1 - \frac{\sigma^2}{a}\right)\ln\frac{1}{1-\gamma_N}\right).$$

Доказательство. Положим

$$x_N = \frac{N - k - Nm(\alpha_N a)}{\sigma_\mu\sqrt{N}}, \qquad \alpha_N > 0, \tag{2.5}$$

и пусть при $N \to \infty$

$$x_N \to x, \qquad -\infty < x < \infty. \tag{2.6}$$

Из равенства (1.4), положив $n = \alpha_N N$ (α_N определяется равенством (2.5)), получим

$$\mathbf{P}\bigl(\nu(N,k) \le \alpha_N N\bigr) = \mathbf{P}\bigl(\mu_0^*(N) \le x_N\bigr). \tag{2.7}$$

Покажем, что в условиях доказываемой теоремы для величины α_N выполняется условие (1.5).

Используя (2.4) и (2.5), получаем

$$e^{-\alpha_N a}\left(1+\alpha_N^2 a^2 \frac{b_N^2}{2\sqrt{N}}\right) = 1-\gamma_N - \frac{x_N}{\sqrt{N}}\sqrt{\alpha_N\sigma^2 e^{-2\alpha_N a} + e^{-\alpha_N a}(1-(1+\alpha_N a)e^{-\alpha_N a})}. \quad (2.8)$$

Отсюда и из неравенств $\alpha_N > 0$, $0 < \underline{\gamma} < \gamma_N < \overline{\gamma} < 1$ следует, что

$$e^{-\alpha_N a} = 1-\gamma_N + O\left(\frac{1}{\sqrt{N}}\right), \qquad \alpha_N = \frac{1}{a}\ln\frac{1}{1-\gamma_N} + O\left(\frac{1}{\sqrt{N}}\right) \quad (2.9)$$

и выполняется условие (1.5). Таким образом, согласно теореме 1 из равенства (2.7) получаем

$$\lim_{N\to\infty}\mathbf{P}\big(\nu(N,k)\le \alpha_N N\big) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-u^2/2}\,du. \quad (2.10)$$

Используя (2.8) и (2.9), выразим α_N через γ_N:

$$\alpha_N a = \ln\frac{1}{1-\gamma_N} + \frac{1}{1-\gamma_N}\frac{1}{\sqrt{N}}\left(\frac{1}{2}\alpha_N^2 a^2 b_N^2 + x_N\sqrt{\alpha_N\sigma^2 e^{-2\alpha_N a} + e^{-\alpha_N a}(1-(1+\alpha_N)e^{-\alpha_N a})}\right) + \\ + O\left(\frac{1}{N}\right).$$

Отсюда и из (2.9) заключаем, что

$$\alpha_N = \frac{1}{a}\ln\frac{1}{1-\gamma_N} + \frac{1}{\sqrt{N}}\left(\frac{b_N^2}{2a}\ln^2(1-\gamma_N) + x_N\sqrt{\frac{\sigma^2}{a^3}\ln\frac{1}{1-\gamma_N} + \frac{1}{a^2}\frac{\gamma_N}{1-\gamma_N} - \frac{1}{a^2}\ln\frac{1}{1-\gamma_N}}\right) + \\ + O\left(\frac{1}{N}\right).$$

Отсюда и из (2.10) следует утверждение теоремы. □

Асимптотическая нормальность $\nu(N,k)$ сохранится и без предположения (2.3). В этом случае величины, нормирующие и центрирующие $\nu(N,k)$, будут зависеть от функции $\alpha_N = \alpha_N(\gamma_N)$, заданной уравнением (2.8).

3. ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ В ПРАВОЙ ОБЛАСТИ

Условие (1.6) можно записать в виде

$$n = \frac{N}{a}(\ln N - \overline{a}_N), \qquad \text{где} \quad \overline{a}_N \to \ln\overline{\lambda}. \quad (3.1)$$

Положим

$$\eta_n = \eta_n(\xi_n) = \frac{\xi_n - na}{\sigma\sqrt{n}}$$

и, используя (3.1), представим ξ_n в следующем виде:

$$\xi_n = N(\ln N - a_N^*),$$

где

$$a_N^* = \overline{a}_N - \eta_n\sigma\sqrt{\frac{\ln N - \overline{a}_N}{Na}}. \quad (3.2)$$

Теорема 3. *Если выполнены условия* (1.3), (3.1) *и при* $N \to \infty$

$$p_k = \frac{1}{N}\left(1 + \frac{b_k(N)}{\ln N}\right), \qquad b_N^* = \frac{1}{N}\sum_{k=1}^{N} e^{-b_k(N)} \to b^*,$$
$$|b_k(N)| \le B < \infty, \qquad k = 1, 2, \dots, N, \tag{3.3}$$

то

$$\lim_{N\to\infty} \mathbf{P}\big(\mu_0(\xi_n, N) = k\big) = \frac{\lambda^k}{k!}e^{-\lambda}, \qquad \lambda = b^*\overline{\lambda}, \quad k = 0, 1, \dots .$$

Доказательство. Вероятность $\mathbf{P}(\mu_0(\xi_n, N) = k)$ представим в виде

$$\mathbf{P}(\mu_0(\xi_n, N) = k) = \Sigma_1 + \Sigma_2, \tag{3.4}$$

где

$$\Sigma_1 = \sum_{|\eta_n(m)|\le C} \mathbf{P}(\xi_n = m)\mathbf{P}\big(\mu_0(m, N) = k\big), \qquad \Sigma_2 = \sum_{|\eta_n(m)|> C} \mathbf{P}(\xi_n = m)\mathbf{P}\big(\mu_0(m, N) = k\big),$$

а $\mu_0(m, N)$ — число ячеек, оставшихся пустыми после размещения m частиц по полиномиальной схеме с вероятностями исходов (3.3).

При $n \to \infty$ и любой постоянной $C > 0$

$$|\Sigma_2| \le \mathbf{P}\big(|\eta_n| > C\big) \to \frac{2}{\sqrt{2\pi}}\int_C^{\infty} e^{-x^2/2}\,dx. \tag{3.5}$$

Если C фиксировано, то в Σ_1 в каждом слагаемом вероятность $\mathbf{P}(\mu_0(m, N) = k)$ сходится при $N \to \infty$ к $\lambda^k e^{-\lambda}/k!$ (см. [1, гл. III, § 3, теорема 3]), так как $m = na + \eta_n(m)\sigma\sqrt{n}$ и при $N \to \infty$

$$\frac{m}{N} = \ln N - \overline{a}_N + \eta_n(m)\sigma\frac{\sqrt{n}}{N} = \ln N - \ln\overline{\lambda} + o(1)$$

равномерно по $m \in \{m\colon |\eta_n(m)| < C\}$,

$$\mathbf{E}\mu_0(m, N) = \sum_{k=1}^{N}(1 - p_k)^m \to \lambda, \qquad m\min_{1\le i\le N} p_i \ge \frac{m}{N}\left(1 - \frac{B}{\ln N}\right) \to \infty.$$

Следовательно, при $N \to \infty$

$$\Sigma_1 \to \frac{\lambda}{\sqrt{2\pi}}\int_{-C}^{C} e^{-x^2/2}\,dx.$$

Отсюда и из (3.5) при $C \to \infty$ следует утверждение теоремы. $\square$

Теорема 4. *Если выполнены условия* (3.1), (3.3) *и* $k = N - c$, *где* $c \ge 0$ — *фиксированное целое число, то при* $N \to \infty$

$$\mathbf{P}\left(\frac{a\nu(N,k)}{N} - \ln N \ge x\right) \to \sum_{l=0}^{c}\frac{(\lambda(x))^l}{l!}e^{-\lambda(x)}, \qquad \lambda(x) = b^*e^{-x}.$$

Доказательство. В равенстве (1.4), положив

$$k = N - c, \qquad n = \frac{N}{a}(\ln N + x_N), \qquad x_N \to x,$$

получим

$$\mathbf{P}\Big(\nu(N,k) \le \frac{N}{a}(\ln N + x_N)\Big) = \sum_{l=0}^{c} \mathbf{P}\big(\mu_0(\xi_n, N) = l\big). \tag{3.6}$$

По теореме 3 распределение $\mu_0(\xi_n, N)$ сходится к распределению Пуассона с параметром $\lambda(x) = b^* e^{-x}$. Отсюда и из равенства (3.6) следует утверждение теоремы. $\square$

4. ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ В ЛЕВОЙ ОБЛАСТИ

Будем предполагать, что при $N \to \infty$

$$p_k = \frac{1}{N}(1 + \Delta_k(N)), \quad k = 1, \dots, N, \qquad \Delta^2(N) = \frac{1}{N}\sum_{k=1}^{N} \Delta_k^2(N) \to \Delta^2. \tag{4.1}$$

Теорема 5. *Если выполнены предположения* (1.3), (1.7), (4.1), *то при* $N \to \infty$

$$\mathbf{P}\big(\overline{\mu}_0(N) = l\big) \to \frac{\Lambda^l}{l!} e^{-\Lambda}, \qquad l = 0, 1, 2, \dots,$$

где

$$\overline{\mu}_0(N) = \mu_0(\xi_n, N) - N + \xi_n, \qquad \Lambda = \lambda a^2(1 + \Delta^2).$$

Доказательство теоремы проводится аналогично доказательству теоремы 3. Вероятность $\mathbf{P}(\overline{\mu}_0(N) = l)$ представим в виде $\mathbf{P}(\overline{\mu}_0(N) = l) = \Sigma_1 + \Sigma_2$, где

$$\Sigma_1 = \sum_{|\eta_n(m)| \le C} \mathbf{P}(\xi_n = m)\mathbf{P}\big(\mu_0(m, N) - N + m = l\big),$$

$$\Sigma_2 = \sum_{|\eta_n(m)| > C} \mathbf{P}(\xi_n = m)\mathbf{P}\big(\mu_0(m, N) - N + m = l\big).$$

При фиксированном C при $N \to \infty$ в Σ_1 (см. [1, гл. III, § 3, теорема 2])

$$\mathbf{P}\big(\mu_0(m, N) - N + m = l\big) \to \frac{\Lambda^l e^{-\Lambda}}{l!}, \qquad l = 0, 1, \dots,$$

так как

$$\frac{m^2}{2N} = \frac{n^2 a^2}{2N} + \frac{a\sigma n\sqrt{n}}{N}\eta_n + \frac{\sigma^2 n \eta_n(m)}{2N} \to \lambda a^2$$

равномерно по $m \in \{m \colon |\eta_n(m)| \le C\}$ и

$$\frac{m^2}{2}\sum_{i=1}^{N} p_i^2 = \frac{m^2}{2N}(1 + \Delta^2(N)) \to \lambda a^2(1 + \Delta^2) = \Lambda.$$

Отсюда теми же рассуждениями, что были использованы в доказательстве теоремы 3, получаем утверждение теоремы 5. $\square$

Теорема 6. *Если при* $n \to \infty$ *выполнены условия* (1.3), (4.1) *и* $\frac{k^2}{2N} = \lambda_N \to \lambda$, *то распределение случайной величины*

$$\nu^*(N, k) = \Big(\nu(N, k) - \frac{k}{a}\Big)\Big(\frac{\sigma}{a}\sqrt{\frac{k}{a}}\Big)^{-1}$$

сходится к нормальному распределению с параметрами $(0, 1)$.

Доказательство. Положим
$$n = \frac{k}{a} + x_N \frac{\sigma}{a}\sqrt{\frac{k}{a}}, \tag{4.2}$$
где $x_N \to x$ при $N \to \infty$. Тогда
$$\xi_n = k + x_N\sigma\sqrt{\frac{k}{a}} + \eta_n\sigma\sqrt{\frac{k}{a}}\biggl(1 + O\biggl(\frac{1}{\sqrt{k}}\biggr)\biggr). \tag{4.3}$$
Подставляя (4.2), (4.3) в (1.4), получаем
$$\mathbf{P}\biggl(\nu(N,k) \le \frac{k}{a} + x_N\sigma\sqrt{\frac{k}{a}}\biggr) = \mathbf{P}\biggl(\overline{\mu}_0(N) \le x_N\sigma\sqrt{\frac{k}{a}} + \eta_n\sigma\sqrt{\frac{k}{a}}\biggl(1 + O\biggl(\frac{1}{\sqrt{k}}\biggr)\biggr)\biggr) =$$
$$= \mathbf{P}\biggl(\eta_n\biggl(1 + O\biggl(\frac{1}{\sqrt{k}}\biggr)\biggr) \ge -x_N + \frac{\overline{\mu}_0(N)}{\sigma\sqrt{k/a}}\biggr). \tag{4.4}$$

Из равенства (4.2) следует, что для $\overline{\mu}_0(N)$ выполнены условия теоремы 5. Согласно этой теореме $\overline{\mu}_0(N)$ имеет в пределе собственное распределение и, следовательно, величина $\overline{\mu}_0(N)/\sigma\sqrt{k/a}$ стремится по вероятности к нулю. Таким образом, отсюда, из равенств (4.4) и условия (1.3) следует, что
$$\mathbf{P}\bigl(\nu^*(N,k) \le x_N)\bigr) = \mathbf{P}\biggl(\eta_n\biggl(1 + O\biggl(\frac{1}{\sqrt{k}}\biggr)\biggr) \ge -x_N\biggr) \to \frac{1}{\sqrt{2\pi}}\int_{-x}^{\infty} e^{-u^2/2}\,du.$$
Из этих соотношений вытекает утверждение теоремы. □

СПИСОК ЛИТЕРАТУРЫ

1. *Колчин В.Ф., Севастьянов Б.А., Чистяков В.П.* Случайные размещения. М.: Наука, 1976.
2. *Ивченко Г.И.* Время ожидания и связанные с ним характеристики в полиномиальной схеме // Дискрет. математика. 1993. Т. 5, №3. С. 3–34.
3. *Ивченко Г.И.* Время ожидания и проверка гипотез в полиномиальной схеме // Теория вероятн. и ее примен. 1974. Т. 19, №4. С. 839–844.
4. *Shioda S.* Some upper and lower bounds on the coupon collector problem // J. Comput. Appl. Math. 2007. V. 200. P. 154–167.