Math-Net.Ru

Общероссийский математический портал

В. А. Иванов, Г. И. Ивченко, Разделимые статистики и моменты останковки в схеме бесповторного выбора, *Дискрет. матем.*, 1997, том 9, выпуск 1, 43–58

DOI: <https://doi.org/10.4213/dm455>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 54.196.121.252

26 апреля 2023 г., 17:53:47

Разделимые статистики и моменты остановки в схеме бесповторного выбора

© 1997 г. В. А. Иванов, Г. И. Ивченко

В рамках обобщенной урновой схемы рассматривается схема бесповторного выбора, когда из урны, содержащей n шаров каждого из N различных цветов, последовательно, равновероятно и независимо друг от друга извлекаются шары, причем извлеченный шар обратно в урну не возвращается. Процесс извлечения шаров прекращается в тот момент, когда впервые частоты (числа появлений) k некоторых цветов достигнут либо превзойдут соответствующие уровни, установленные до начала испытаний. Изучаются разделимые статистики

$$L_{Nk} = \sum_{j=1}^N g(\eta_j),$$

где g — некоторая функция целочисленного аргумента, а $\eta_1, \dots, \eta_N$ — наблюдаемые частоты появления соответствующих цветов в момент остановки испытаний. Получено достаточно полное описание класса предельных распределений статистик L_{Nk} при $N \rightarrow \infty$, $n = n(N)$ и различном характере поведения параметра k .

1. Введение

Пусть в начальный момент времени урна содержит n шаров каждого из N различных цветов $A_1, \dots, A_N$. Шары последовательно и по одному извлекаются из урны так, что в любой момент каждый из находящихся в урне шаров может быть извлечен с одинаковой вероятностью. После извлечения очередного шара фиксируется его цвет, а сам шар обратно в урну не возвращается (схема бесповторного выбора).

Таким образом, если в момент t , $t = 1, 2, \dots$, уже было извлечено k_j шаров цвета A_j , $j = 1, \dots, N$, то вероятность извлечения из урны в момент $t + 1$ шара цвета A_i , $i = 1, \dots, N$, равна

$$(n - k_i) \left(\sum_{j=1}^N (n - k_j) \right)^{-1} \quad (1)$$

Описанной урновой схеме бесповторного выбора можно сопоставить схему размещения частиц по ячейкам, занумерованным числами $1, \dots, N$. Именно, пусть в начальный момент времени все ячейки пусты. Если в момент времени t ,

$t = 1, 2, \dots$, из урны извлекается шар цвета A_j , $j = 1, \dots, N$, то в j -ю ячейку помещается одна частица. Следовательно, вероятность того, что в момент времени $t + 1$ частица попадет в i -ю ячейку, имеет вид, указанный в (1), где $k_1, \dots, k_N$ — содержимое соответствующих ячеек после размещения в них t частиц.

Предположим, что до начала испытаний для j -й ячейки установлен некоторый уровень ν_j , где $\nu, \nu_1, \dots, \nu_N$ — целочисленные одинаково распределенные неотрицательные независимые случайные величины (с.в.), причем значения с.в. ν с вероятностью единица не превышают числа шаров каждого цвета в урне в начальный момент. Будем называть ячейку с номером j заполненной в момент t , $t = 0, 1, 2, \dots$, если $\eta_j(t) \geq \nu_j$, где $\eta_1(t), \dots, \eta_N(t)$ — содержимое соответствующих ячеек после размещения t частиц ($\eta_j(0) = 0$, $j = 1, \dots, N$).

Пусть, далее, задано целое k , $1 \leq k \leq N$. Определим с.в.

$$\nu(N, k) = \min \left\{ t \geq 0: \sum_{j=1}^N I(\eta_j(t) \geq \nu_j) = k \right\}, \quad (2)$$

называемую временем ожидания (до заполнения k ячеек), или моментом остановки (здесь и далее $I(\cdot)$ — индикатор).

Наконец, обозначим через $\eta_j = \eta_j(\nu(N, k))$ содержимое j -й ячейки в момент остановки $\nu(N, k)$ и рассмотрим произвольную разделимую статистику (РС), т.е. с.в. вида

$$L_{Nk} = \sum_{j=1}^N g(\eta_j), \quad (3)$$

где g — некоторая заданная функция целочисленного аргумента.

РС L_{Nk} — это достаточно общий объект, позволяющий охватить и тем самым исследовать с единых позиций большое разнообразие различных характеристик рассматриваемой урновой схемы. В свою очередь, описанная выше схема бесповторного выбора со случайными уровнями укладывается в так называемую обобщенную урновую схему (ОУС), рассматривавшуюся в [5, 11]. В этих работах методом вложения ОУС в подходящий марковский процесс чистого размножения получено замкнутое выражение для характеристической функции (х.ф.) произвольной РС от заполнений ячеек в момент остановки испытаний, специализация которого для схемы бесповторного выбора приводит к интегральному представлению (4), указанному ниже. Базируясь на этом результате, мы в данной работе приводим исчерпывающий асимптотический (при $N \rightarrow \infty$) анализ распределения РС (3), восполняющий пробелы в асимптотическом исследовании схемы бесповторного выбора.

Аналогичная задача для схемы равновероятного повторного выбора рассмотрена в работе [3]. Здесь мы, в основном, следуем идеям этой работы и получаемые нами результаты во многом аналогичны соответствующим результатам в [3]. Из других, близких по тематике и методам работ, отметим [1, 2, 8, 9, 12] и обзоры [4, 10], где изучались различные характеристики в полиномиальной схеме размещения, связанные с моментами остановки. С другой стороны, с позиций прямых задач, т.е. когда число испытаний (объем выборки) фиксировано, схема бесповторного выбора достаточно детально была исследована ранее в работах [6, 7]. Из приведенной библиографии видно, что схема бесповторного выбора,

являющаяся одной из важнейших дискретных вероятностных схем, исследована с рассматриваемых в данной работе позиций значительно менее детально, и в этом аспекте данная работа в определенной мере заполняет эту лакуну.

2. Метод исследования

В дальнейшем нам потребуются некоторые обозначения.

Символом $\mathcal{L}(\xi)$ будем обозначать закон распределения с.в. ξ , будем обозначать $\text{Bi}(n, \alpha)$ биномиальный закон распределения с параметрами n и α , задаваемый вероятностями

$$b_r(n; \alpha) = \binom{n}{r} \alpha^r (1 - \alpha)^{n-r}, \quad r = 0, 1, \dots, n, \quad \alpha \in (0, 1).$$

Пусть θ — с.в., имеющая бета-распределение с плотностью

$$\beta(x; k, N - k + 1) = N \binom{N-1}{k-1} x^{k-1} (1-x)^{N-k}, \quad x \in [0, 1].$$

Через

$$B(p; k, N - k + 1) = \int_0^p \beta(x; k, N - k + 1) dx, \quad p \in [0, 1],$$

обозначим неполную бета-функцию. Известно, что

$$B(p; k, N - k + 1) = \mathbf{P}\{Y(N, p) \geq k\},$$

где $\mathcal{L}(Y(N, p)) = \text{Bi}(N, p)$.

Пусть, далее, $\xi(t)$ — с.в., не зависящая от ν и имеющая распределение $\text{Bi}(n, \alpha(t))$ с $\alpha(t) = 1 - e^{-t}$, $t > 0$, а $\xi_j(t)$, $j = 0, 1, 2$, — независимые с.в., распределения которых задаются следующим образом:

$$\mathbf{P}\{\xi_0(t) = r\} = \frac{(n - r + 1) \mathbf{P}\{\nu = r\} b_{r-1}(n; \alpha(t))}{c(t)}, \quad r = 1, \dots, n,$$

где

$$c(t) = \sum_{m=1}^n (n - m + 1) \mathbf{P}\{\nu = m\} b_{m-1}(n; \alpha(t))$$

— нормирующий множитель,

$$\mathcal{L}(\xi_1(t)) = \mathcal{L}(\xi(t) \mid \xi(t) \geq \nu), \quad \mathcal{L}(\xi_2(t)) = \mathcal{L}(\xi(t) \mid \xi(t) < \nu).$$

Введем также соответствующие х.ф.

$$f_j(\tau; t) = \mathbf{E} \exp\{i\tau g(\xi_j(t))\}, \quad j = 0, 1, 2.$$

Наконец, пусть

$$H(t) = 1 - \bar{H}(t) = \mathbf{P}\{\xi(t) \geq \nu\}, \quad t \geq 0.$$

Обозначим $p = 1 - q = \mathbf{P}\{\nu = 0\} = H(0)$ и будем везде далее считать, что $p < 1$. Тогда при $p > 0$ функция $H(t)$ имеет в нуле скачок величины p , а при $t > 0$

$$H'(t) = h(t) > 0.$$

Следовательно, функция $H(t)$ строго монотонно возрастает от p до 1 и, значит, может быть определена обратная к ней функция

$$t(x) = H^{-1}(x), \quad x \in [p, 1].$$

Для нахождения предельных распределений РС L_{Nk} нам потребуется следующее интегральное представление для ее х.ф. (см. [5], теорема 2):

$$\mathbf{E}e^{i\tau L_{Nk}} = \int_0^1 (I(x \leq p) + I(x > p)f_{Nk}(\tau; t(x)))\beta(x; k, N - k + 1)dx, \quad (4)$$

где

$$f_{Nk}(\tau; t) = f_0(\tau; t)f_1^{k-1}(\tau; t)f_2^{N-k}(\tau; t).$$

Заметим, что (4) можно записать также в виде

$$\mathbf{E}e^{i\tau L_{Nk}} = B(p; k, N - k + 1) + \int_p^1 f_{Nk}(\tau; t(x))\beta(x; k, N - k + 1)dx. \quad (5)$$

Здесь и в дальнейшем без ограничения общности предполагается, что $g(0) = 0$, так как в противном случае можно рассмотреть РС

$$\tilde{L}_{Nk} = L_{Nk} - Ng(0) = \sum_{j=1}^N \tilde{g}(\eta_j),$$

где $\tilde{g}(x) = g(x) - g(0)$.

Как отмечено в [5], представлениям (4)–(5) можно придать следующую вероятностную интерпретацию. Рассмотрим совокупность

$$\{\xi_0(t(\theta)), \xi_{1j}(t(\theta)), j = 1, \dots, k-1, \xi_{2j}(t(\theta)), j = 1, \dots, N-k\}$$

независимых при каждом значении $\theta \in (p, 1]$ с.в., где $\xi_{1j}(t)$ ($\xi_{2j}(t)$) являются независимыми копиями усеченных с.в. $\xi_1(t)$ ($\xi_2(t)$), и предположим, что до начала испытаний производится эксперимент, в котором наблюдается с.в. θ .

Тогда при $\theta \in (p, 1]$ распределение L_{Nk} совпадает с распределением суммы

$$g(\xi_0(t(\theta))) + \sum_{j=1}^{k-1} g(\xi_{1j}(t(\theta))) + \sum_{j=1}^{N-k} g(\xi_{2j}(t(\theta)))$$

условно независимых (т.е. независимых при каждом θ) слагаемых, в противном же случае, при $\theta \leq p$, полагается $L_{Nk} = 0$. Таким образом, по существу мы имеем здесь дело с суммированием условно независимых слагаемых, конструкция которых определяется биномиальной с.в. $\xi(t)$ и ее усечениями случайным уровнем ν .

Для сравнения напомним, что для схемы равновероятного повторного выбора аналогичная конструкция порождается пуассоновской с.в. (см. [3]).

При асимптотическом анализе распределения РС L_{Nk} мы используем метод, изложенный в работе [3]. Напомним вкратце суть этого метода.

Выделим три области изменения параметра k :

- (1) k фиксировано, $k \geq 1$ (левая область),
- (2) $s = N - k + 1$ фиксировано, $s \geq 1$ (правая область),
- (3) $k = \lambda N + o(\sqrt{N})$, $\lambda = 1 - \bar{\lambda} \in (0, 1)$ (средняя область).

Далее везде мы предполагаем (не оговаривая это каждый раз), что распределение уровня ν фиксировано и не зависит от N , что же касается параметра n , то в некоторых случаях он будет предполагаться фиксированным, а в других допускается, что $n = n(N) \rightarrow \infty$ при $N \rightarrow \infty$.

Хорошо известно, что при $N \rightarrow \infty$ и фиксированном $p \in (0, 1)$ в левой области

$$B(p; k, N - k + 1) \rightarrow 1,$$

в правой области

$$B(p; k, N - k + 1) \rightarrow 0,$$

в средней области

$$B(p; k, N - k + 1) \rightarrow \begin{cases} 1, & \lambda < p, \\ 1/2, & \lambda = p, \\ 0, & \lambda > p. \end{cases}$$

Таким образом, при $p \neq 0$ в левой области и в средней области при $\lambda < p$ и $N \rightarrow \infty$ имеет место сходимость по вероятности $L_{Nk} \rightarrow 0$ (напомним, что $g(0) = 0$).

Эти тривиальные случаи из дальнейших рассуждений исключаются.

Далее, если положить $x_N = a_N + x b_N$, то при соответствующем выборе констант a_N и b_N справедливо соотношение

$$\beta_N(x) = \beta(x_N; k, N - k + 1) |b_N| \rightarrow \beta(x),$$

где в качестве $\beta(x)$ выступает либо плотность экстремальных значений, либо нормальная плотность. При этом в левой области

$$a_N = 0, \quad b_N = 1/N, \quad \beta(x) = \pi_{k-1}(x) = x^{k-1} e^{-x} / (k-1)!, \quad x > 0,$$

в правой области

$$a_N = 1, \quad b_N = -1/N, \quad \beta(x) = \pi_{s-1}(x), \quad x > 0,$$

в средней области

$$a_N = \lambda, \quad b_N = \frac{1}{N} (\lambda \bar{\lambda})^{1/2}, \quad \beta(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbf{R}.$$

Положив $t_N = t(x_N)$ и произведя в (4) соответствующую замену переменных, для левой, правой и средней областей, соответственно, получим, что

$$\begin{aligned} \mathbf{E} e^{i\tau L_{Nk}} &= \int_0^\infty (I(x \leq Np) + I(Np < x \leq N) f_{Nk}(\tau; t_N)) \beta_N(x) dx, \\ \mathbf{E} e^{i\tau L_{Nk}} &= \int_0^\infty (I(Nq \leq x \leq N) + I(x < Nq) f_{Nk}(\tau; t_N)) \beta_N(x) dx, \\ \mathbf{E} e^{i\tau L_{Nk}} &= \int_{-\infty}^\infty \left(I \left(-\sqrt{\lambda N / \bar{\lambda}} \leq x \leq (p - \lambda) \sqrt{N / (\lambda \bar{\lambda})} \right) \right. \\ &\quad \left. + I \left((p - \lambda) \sqrt{N / (\lambda \bar{\lambda})} < x \leq \sqrt{\lambda N / \bar{\lambda}} \right) f_{Nk}(\tau; t_N) \right) \beta_N(x) dx. \end{aligned} \tag{6}$$

Предположим теперь, что для любого x из соответствующей области существует

$$f(\tau; x) = \lim_{N \rightarrow \infty} f_{Nk}(\tau; t_N),$$

и воспользуемся следующим вариантом теоремы о мажорируемой сходимости [13]: если β_N и β — функции плотности, $|\psi_N| \leq 1$, $|\psi| \leq 1$ и имеет место поточечная сходимость $\beta_N(x) \rightarrow \beta(x)$, $\psi_N(x) \rightarrow \psi(x)$ при $N \rightarrow \infty$, то

$$\int \psi_N(x) \beta_N(x) dx \rightarrow \int \psi(x) \beta(x) dx.$$

Отсюда следует, что если при соответствующих условиях существует предел $f(\tau; x)$, то предельная х.ф. РС L_{Nk} будет иметь вид

$$\int_0^\infty f(\tau; x) \beta(x) dx,$$

за исключением средней области при $\lambda = p$, когда этот предел равен

$$\frac{1}{2} + \int_0^\infty f(\tau; x) \varphi(x) dx.$$

Таким образом, задача отыскания предельных распределений РС L_{Nk} распадается на три части:

- (а) нахождение условий существования и вида пределов $f(\tau; x)$,
- (б) вычисление предельных х.ф.,
- (с) обращение предельных х.ф., полученных в (б).

В следующих разделах проводятся соответствующие аналитические вычисления для указанных выше трех областей изменения параметра k .

3. Асимптотические распределения в левой области

Напомним, что в рассматриваемом случае $p = \mathbf{P}\{\nu = 0\} = 0$, $g(0) = 0$. Пусть целые m_1 , l определяются условиями

$$m_1 = \min \{1 \leq r \leq n: \mathbf{P}\{\nu = r\} > 0\},$$

$$l = \min \{1 \leq j \leq n: g(j) \neq 0\}$$

и $a = 1 - \bar{a} = \mathbf{P}\{\nu = m_1\}$.

В нашем случае $x_N = x/N$, $x > 0$, поэтому $t_N \rightarrow \infty$ при $N \rightarrow \infty$, и из представления

$$H(t) = \sum_{r=m_1}^n \mathbf{P}\{\nu \leq r\} b_r(n, \alpha(t)) = a \binom{n}{m_1} t^{m_1} (1 + o(1)),$$

справедливого при $nt \rightarrow \infty$ (здесь допускается, что $n = n(N) \rightarrow \infty$ при $N \rightarrow \infty$), следует, что при $N \rightarrow \infty$

$$t_N = \left(x / \left(a \binom{n}{m_1} N \right) \right)^{1/m_1} (1 + o(1)).$$

Отсюда получим, что

$$\begin{aligned} f_0(\tau; t_N) &= \frac{1}{c(t_N)} \sum_{r=m_1}^n e^{i\tau g(r)} \mathbf{P}\{\nu = r\} (n - r + 1) b_{r-1}(n, \alpha(t_N)) \\ &= e^{i\tau g(m_1)} (1 + o(1)), \\ f_1(\tau; t_N) &= \frac{1}{H(t_N)} \sum_{r=0}^n e^{i\tau g(r)} \mathbf{P}\{\nu \leq r \mid b_r(n, \alpha(t_N))\} \\ &= \frac{N}{x} e^{i\tau g(m_1)} a \binom{n}{m_1} t_N^{m_1} (1 + o(1)) \\ &= e^{i\tau g(m_1)} (1 + o(1)). \end{aligned}$$

Для х.ф. $f_2(\tau; t_N)$ можно записать представление

$$\begin{aligned} f_2(\tau; t_N) &= \frac{1}{\bar{H}(t_N)} \sum_{r=0}^n e^{i\tau g(r)} \mathbf{P}\{\nu > r\} b_r(n, \alpha(t_N)) \\ &= 1 + \frac{1}{1 - x_N} \sum_{r=l}^n (e^{i\tau g(r)} - 1) \mathbf{P}\{\nu > r\} b_r(n, \alpha(t_N)), \end{aligned}$$

из которого следует, что

$$f_2(\tau; t_N) = 1 + (e^{i\tau g(l)} - 1) \mathbf{P}\{\nu > l\} \binom{n}{l} t_N^l (1 + o(1)).$$

Отсюда следует, что если $l > m_1$, то $f_2^N(\tau; t_N) \rightarrow 1$, если $l = m_1$, то

$$f_2^N(\tau; t_N) \rightarrow \exp \left\{ (e^{i\tau g(m_1)} - 1) \bar{a}x/a \right\},$$

если же $l < m_1$, то

$$f_2^N \left(N^{(l-m_1)/m_1} \binom{n}{l}^{-1} \binom{n}{m_1}^{l/m_1} \tau; t_N \right) \rightarrow \exp \left\{ i\tau g(l) (x/a)^{l/m_1} \right\}.$$

Итак, в данном случае для любого $x > 0$ все возможные пределы для х.ф. $f_{Nk}(\tau; t_N)$ исчерпываются следующими возможностями:

если $l > m_1$, то $f_{Nk}(\tau; t_N) \rightarrow 1$,

если $l = m_1$, то

$$f_{Nk}(\tau; t_N) \rightarrow \exp \left\{ i\tau kg(m_1) + (e^{i\tau g(m_1)} - 1) \bar{a}x/a \right\},$$

если же $l < m_1$, то

$$f_{Nk} \left(N^{(l-m_1)/m_1} \binom{n}{l}^{-1} \binom{n}{m_1}^{l/m_1} \tau; t_N \right) \rightarrow \exp \left\{ i\tau g(l) (x/a)^{l/m_1} \right\}.$$

После обращения соответствующих х.ф.

$$\int_0^\infty f(\tau; x) \pi_{k-1}(x) dx$$

получаем следующую классификацию асимптотических распределений РС L_{Nk} .

Теорема 1. Пусть $N \rightarrow \infty$, k фиксировано, $p = 0$. Тогда при $l > m_1$ РС L_{Nk} асимптотически вырождается в 0, при $l = m_1$ с.в. $g^{-1}(m_1)L_{Nk} - k$ имеет в пределе отрицательное биномиальное распределение $\overline{\text{Bi}}(k, a)$,

$$\mathbf{P}\{g^{-1}(m_1)L_{Nk} - k = r\} \rightarrow \binom{k+r-1}{r} a^k \bar{a}^r, \quad r = 0, 1, \dots,$$

при $l < m_1$

$$\mathbf{P}\left\{N^{(l-m_1)/m_1} \binom{n}{m_1}^{l/m_1} \left(g(l) \binom{n}{l}\right)^{-1} L_{Nk} \leq x\right\} \rightarrow \Gamma_k(ax^{m_1/l}), \quad x > 0,$$

где

$$\Gamma_k(t) = \int_0^t \pi_{k-1}(y) dy.$$

Замечание 1. Сравнение теоремы 1 с соответствующим результатом для схемы равновероятного повторного выбора (теорема 2 в [3]) показывает, что некоторое отличие мы имеем здесь лишь в нормировке для случая $l < m_1$ при фиксированном n , если же $n = n(N) \rightarrow \infty$ при $N \rightarrow \infty$, то и это различие исчезает.

Рассмотрим некоторые примеры применения теоремы 1.

Пример 1. Пусть $g(x) = x$. Тогда $L_{Nk} = \nu(N, k)$, и для времени ожидания получаем, что при $m_1 = 1$

$$\mathcal{L}(\nu(N, k) - k) \rightarrow \overline{\text{Bi}}(k, a)$$

(при $a = 1$, т.е. для вырожденного уровня, $\nu(N, k)$ сходится по вероятности к k), а при $m_1 > 1$

$$\mathbf{P}\left\{N^{(1-m_1)/m_1} n^{-1} \binom{n}{m_1}^{1/m_1} \nu(N, k) \leq x\right\} \rightarrow \Gamma_k(ax^{m_1}), \quad x > 0.$$

Пример 2. Пусть $g(x) = I(x = r)$. Тогда

$$L_{Nk} = \mu_r = \sum_{j=1}^N I(\eta_j = r),$$

и при $r > m_1$ случайная величина μ_r сходится по вероятности к нулю,

$$\mathcal{L}(\mu_{m_1} - k) \rightarrow \overline{\text{Bi}}(k, a)$$

(при $a = 1$ случайная величина μ_{m_1} сходится по вероятности к k), при $1 \leq r < m_1$

$$\mathbf{P} \left\{ N^{(r-m_1)/m_1} \binom{n}{m_1}^{r/m_1} \binom{n}{r}^{-1} \mu_r \leq x \right\} \rightarrow \Gamma_k(ax^{m_1/r}), \quad x > 0,$$

а с.в.

$$N - \mu_0 = \sum_{j=1}^N I(\eta_j > 0)$$

имеет в пределе такое же распределение, как и μ_1 .

4. Асимптотические распределения в правой области

В этом случае задача состоит не только в определении предела $f(\tau; x)$, но и в установлении условий его существования.

В рассматриваемой области ключевым является соотношение

$$x_N = 1 - \frac{x}{N} = H(t_N),$$

и основную роль здесь играет поведение функции $\bar{H}(t) = 1 - H(t)$ при $t \rightarrow \infty$.

Положим $m_2 = \max \{r \leq n: \mathbf{P}\{\nu = r\} > 0\}$ и $b = 1 - \bar{b} = \mathbf{P}\{\nu = m_2\}$. Тогда при $t \rightarrow \infty$

$$\bar{H}(t) = \sum_{r=0}^{m_2-1} \mathbf{P}\{\nu > r\} b_r(n, \alpha(t)) = b b_{m_2-1}(n, \alpha(t))(1 + o(1)).$$

Отсюда следует, что

$$\frac{x}{N} = b b_{m_2-1}(n, \alpha(t_N))(1 + o(1)),$$

и для $t_N = t(1 - x/N)$ получаем при $N \rightarrow \infty$ асимптотическое представление

$$t_N = \frac{1}{n - m_2 + 1} (c_N - \ln x + o(1)), \quad x > 0,$$

где

$$c_N = \ln \left(bN \binom{n}{m_2-1} \right),$$

а также, что

$$N b_r(n, \alpha(t_N)) \rightarrow \begin{cases} 0, & r < m_2 - 1, \\ x/b, & r = m_2 - 1, \\ \infty & r \geq m_2. \end{cases}$$

С учетом всего вышесказанного получаем, что при $N \rightarrow \infty$

$$f_0(\tau; t_N) = e^{i\tau g(m_2)}(1 + o(1)),$$

$$f_2(\tau; t_N) = e^{i\tau g(m_2-1)}(1 + o(1)),$$

а для х.ф. $f_1(\tau; t_N)$ справедливо асимптотическое представление

$$\begin{aligned} f_1(\tau; t_N) &= 1 + \frac{1}{x_N} \left(\mathbf{E} e^{i\tau g(\xi(t_N))} - 1 - \sum_{r=0}^{m_2-1} (e^{i\tau g(r)} - 1) \mathbf{P}\{\nu > r\} b_r(n, \alpha(t_N)) \right) \\ &= 1 + (\mathbf{E} e^{i\tau g(\xi(t_N))} - 1)(1 + o(1)) - \frac{x}{N} (e^{i\tau g(m_2-1)} - 1)(1 + o(1)). \end{aligned}$$

Пусть функция $g = g_N$ такова, что для любого $x > 0$ существует предел

$$\lim_{N \rightarrow \infty} N(\mathbf{E} e^{i\tau g_N(\xi(t_N))} - 1) = a(\tau; x). \quad (7)$$

Выражение в левой части (7) можно записать в виде

$$(e^{i\tau g(m_2-1)} - 1) \frac{x}{b} + N \sum_{r=m_2}^n (e^{i\tau g_N(r)} - 1) b_r(n, \alpha(t_N)) + o(1),$$

поэтому условие (7) означает, в частности, что существует

$$\lim_{N \rightarrow \infty} g_N(m_2 - 1) = g(m_2 - 1)$$

и $g_N(r) = o(1)$ при $r \geq m_2$.

Если выполнено (7), то

$$f_1^N(\tau; t_N) \rightarrow \exp\{a(\tau; x) - x(e^{i\tau g(m_2-1)} - 1)\}$$

и, следовательно, в рассматриваемом случае

$$f_{Nk}(\tau; t_N) \rightarrow f(\tau; x) \quad (8)$$

где

$$f(\tau; x) = \exp\{i\tau(s-1)g(m_2-1) - x(e^{i\tau g(m_2-1)} - 1) + a(\tau; x)\}.$$

В итоге мы можем сформулировать следующий общий результат.

Теорема 2. Пусть $N \rightarrow \infty$ и $s = N - k + 1$ фиксировано, $s \geq 1$. Тогда, если функция $g = g_N$ такова, что выполняется условие (7), то

$$\mathbf{E} e^{i\tau L_{Nk}} \rightarrow \int_0^\infty f(\tau; x) \pi_{s-1}(x) dx,$$

где функция $f(\tau; x)$ определена в (8).

Значение этого результата заключается в том, что из него уже достаточно просто можно выводить различные, представляющие интерес частные случаи, когда удастся получить явное выражение для предела в (7).

Пусть, например, функция g обладает следующим свойством: при некотором фиксированном r , $1 \leq r \leq n$, $g(r) \neq 0$ и $g(j) = 0$ для всех $j > r$.

Тогда, если $r < m_2 - 1$, то в (7) $a(\tau; x) \equiv 0$ и, следовательно, $f(\tau; x) = 1$; если $r = m_2 - 1$, то $a(\tau; x) = (e^{i\tau g(m_2-1)} - 1)x/b$, т.е.

$$f(\tau; x) = \exp\{i\tau(s-1)g(m_2-1) + (e^{i\tau g(m_2-1)} - 1)\bar{b}x/b\}.$$

Наконец, если $r \geq m_2$, то условие (7) сводится к условию

$$Nb_r(n, \alpha(t_N))(e^{i\tau g_N(r)} - 1) \rightarrow a(\tau; x).$$

Отсюда находим, что если положить $g_N(j) = d_N g(j)$, где

$$d_N = \left(N \binom{n}{m_2 - 1} \right)^{(n-r)/(n-m_2+1)} \left(N \binom{n}{r} \right)^{-1},$$

то для фиксированного n

$$a(\tau; x) = i\tau g(r) \left(\frac{x}{b} \right)^{(n-r)/(n-m_2+1)},$$

а при $n \rightarrow \infty$

$$a(\tau; x) = i\tau g(r) \frac{x}{b}.$$

В свою очередь, это означает, что функция $f_1^N(\tau d_N; t_N)$, а вместе с ней и х.ф. $f_{Nk}(\tau d_N; t_N)$, имеют пределом $\exp\{a(\tau; x)\}$. Таким образом, все возможные пределы для х.ф. f_{Nk} выписываются в данном случае в явном виде, и их обращение приводит к следующей классификации соответствующих предельных распределений:

если $r < m_2 - 1$, то L_{Nk} сходится по вероятности к нулю,

если $r = m_2 - 1$, то

$$\mathcal{L}(g^{-1}(m_2 - 1)L_{Nk} - s + 1) \rightarrow \overline{\text{Bi}}(s, b),$$

если $r \geq m_2$, то

$$\mathbf{P}\{g^{-1}(r)d_N L_{Nk} \leq x\} \rightarrow \Gamma(bx^\alpha),$$

где $\alpha = (n - m_2 + 1)/(n - r)$ при n фиксированном и $\alpha = 1$ при $n \rightarrow \infty$.

В частности, здесь содержится результат для индикаторной с.в.

$$\mu_r = \sum_{j=1}^N I(\eta_j = r)$$

(при $g(r) = 1$), сравнение которого с соответствующим результатом для схемы равновероятного повторного выбора (теорема 3 в [3]) показывает, что при $n = n(N) \rightarrow \infty$ различие между этими схемами в пределе нивелируется, в то время как при фиксированном n в случае $r \geq m_2$ соответствующие асимптотики различаются.

В качестве еще одного примера, представляющего самостоятельный интерес, рассмотрим время ожидания $\nu(N, k)$. Пусть сначала $m_2 < n$. Обозначив

$$\varepsilon_N = N^{-(n-m_2)/(n-m_2+1)}$$

и положив $g_N(x) = \varepsilon_N(n - x)/n$, получим в (7) при фиксированном n

$$N(\mathbf{E}e^{i\tau g_N(\xi(t_N))} - 1) = N((1 + e^{-tN}(e^{i\tau \varepsilon_N/n} - 1))^n - 1) = i\tau N \varepsilon_N e^{-tN}(1 + o(1)),$$

что с учетом асимптотики для t_N приводит к равенству

$$a(\tau; x) = i\tau \left(x \left(b \binom{n}{m_2 - 1} \right)^{-1} \right)^{1/(n-m_2+1)}$$

Кроме того, здесь $g(m_2 - 1) = 0$, и поэтому функция $f(\tau; x)$ в (8) равна $\exp\{a(\tau; x)\}$. Тем самым по теореме 2

$$\mathbf{E} \exp\{i\tau \varepsilon_N(nN - \nu(N, k))/n\} \rightarrow \int_0^\infty e^{a(\tau; x)} \pi_{s-1}(x) dx.$$

В случае $m_2 = n$ надо положить $g(x) = n = x$ при $x \leq n$ и $g(x) = 0$ при $x > n$. В итоге здесь

$$f(\tau; x) = \exp\{i\tau(s-1) + (e^{i\tau} - 1)\bar{b}x/b\}.$$

Следовательно,

$$\mathbf{E} \exp\{i\tau(nN - s + 1 - \nu(N, k))\} \rightarrow \int_0^\infty \exp\{(e^{i\tau} - 1)\bar{b}x/b\} \pi_{s-1}(x) dx.$$

Обращая предельные х.ф., можно выписать и явный вид предельных функций распределения. Ввиду важности рассматриваемой характеристики сформулируем окончательный результат в виде теоремы.

Теорема 3. Пусть $N \rightarrow \infty$, а $s = N - k + 1$ и n фиксированы. Тогда для времени ожидания $\nu(N, k)$ имеют место следующие предельные соотношения: если $m_2 \leq n$, то

$$\mathbf{P}\{d_N(nN - \nu(N, k)) \leq x\} \rightarrow \Gamma_s(x^{n-m_2+1}), \quad x > 0,$$

где

$$d_N = \left(b \binom{n}{m_2 - 1} \right)^{1/(n-m_2+1)} \left(nN^{(n-m_2)/(n-m_2+1)} \right)^{-1},$$

если $m_2 = n$, то

$$\mathcal{L}(Nn - s + 1 - \nu(N, k)) \rightarrow \overline{\text{Bi}}(s, b).$$

Подчеркнем, что этот результат не имеет аналога для схемы равновероятного повторного выбора, так что с позиций данной характеристики схемы повторного и бесповторного выбора различаются наиболее существенно. Так, в частности, время ожидания $\nu(N, N)$ до появления всех цветов по крайней мере по одному разу ($m_2 = 1$, $b = 1$) при бесповторном выборе имеет предельное распределение

$$\mathbf{P}\{(nN - \nu(N, N))/(nN^{(n-1)/n}) \leq x\} \rightarrow 1 - e^{-x^n}, \quad x > 0,$$

в то время как в схеме повторного выбора соответствующее соотношение имеет вид [3]

$$\mathbf{P}\{N^{-1}\nu(N, N) - \ln N \leq x\} \rightarrow e^{-e^{-x}}, \quad -\infty < x < \infty.$$

5. Асимптотические распределения в средней области

В этом параграфе мы будем считать n фиксированным. Как уже отмечалось в параграфе 3, при

$$k = \lambda N + o(\sqrt{N}), \quad \lambda = 1 - \bar{\lambda} \in (0, 1),$$

асимптотическое поведение РС L_{Nk} существенно различается для случаев $\lambda > p$ и $\lambda = p > 0$.

В первом случае имеет место асимптотическая нормальность.

Теорема 4. Пусть $N, k \rightarrow \infty$ так, что $k = \lambda N + o(\sqrt{N})$ и $\lambda > p$. Тогда

$$\mathcal{L}(L_{Nk}) \sim \mathcal{N}(N\mu_\lambda, N\sigma_\lambda^2),$$

где

$$\begin{aligned} \mu_\lambda &= \mathbf{E}g(\xi_\lambda), & \sigma_\lambda^2 &= \mathbf{D}\tilde{g}(\xi_\lambda), \\ \xi_\lambda &= \xi(t(\lambda)), & \alpha_\lambda &= \alpha(t(\lambda)), \\ \tilde{g}(\xi_\lambda) &= g(\xi_\lambda) - d_\lambda(g)I(\xi_\lambda \geq \nu), \\ d_\lambda(g) &= \frac{1}{\alpha_\lambda} \text{cov}(g(\xi_\lambda), \xi_\lambda)t'(\lambda). \end{aligned}$$

Доказательство. Доказательство основывается на использовании центральной предельной теоремы и аналогично доказательству теоремы 4 в [3], поэтому отметим лишь некоторые технические моменты, отражающие специфику схемы бесповторного выбора.

Непосредственно устанавливается, что при $x_N = \lambda + x\sqrt{\lambda\bar{\lambda}/N}$

$$b_r(n, \alpha(t_N)) = b_r(n, \alpha_\lambda) \left(1 + (r/\alpha_\lambda - n)xt'(\lambda)\sqrt{\lambda\bar{\lambda}/N} \right) + O(1/N),$$

откуда находим, что

$$\mathbf{E}g(\xi_j(t_N)) = \mu_{j\lambda} + \frac{xc_j}{\sqrt{N}} + O\left(\frac{1}{N}\right), \quad j = 1, 2,$$

где

$$\begin{aligned} \mu_{j\lambda} &= \mathbf{E}g(\xi_{j\lambda}), & \xi_{j\lambda} &= \xi_j(t(\lambda)), \\ c_1 &= \left(\frac{t'(\lambda)}{\alpha_\lambda} \mathbf{E}(\xi_{1\lambda} - n\alpha_\lambda)g(\xi_{1\lambda}) - \frac{\mu_{1\lambda}}{\lambda} \right) \sqrt{\lambda\bar{\lambda}}, \\ c_2 &= \left(\frac{t'(\lambda)}{\alpha_\lambda} \mathbf{E}(\xi_{2\lambda} - n\alpha_\lambda)g(\xi_{2\lambda}) + \frac{\mu_{2\lambda}}{\lambda} \right) \sqrt{\lambda\bar{\lambda}}. \end{aligned}$$

Далее устанавливается, что предельная х.ф. для нормированной статистики

$$\frac{1}{\sqrt{N}}(L_{Nk} - N(\lambda\mu_{1\lambda} + \bar{\lambda}\mu_{2\lambda}))$$

равна $\exp\{-\sigma_\lambda^2 \tau^2 / 2\}$, где

$$\sigma_\lambda^2 = \lambda \sigma_1^2 + \bar{\lambda} \sigma_2^2 + (\lambda c_1 + \bar{\lambda} c_2)^2, \quad \sigma_j^2 = \mathbf{D}g(\xi_{j\lambda}), \quad j = 1, 2.$$

Наконец, непосредственными вычислениями получаем, что

$$\mu_\lambda = \lambda \mu_{1\lambda} + \bar{\lambda} \mu_{2\lambda}, \quad \sigma_\lambda^2 = \mathbf{D}g(\xi_\lambda) + \lambda \bar{\lambda} d_\lambda(g)(d_\lambda(g) + 2(\mu_{2\lambda} - \mu_{1\lambda})).$$

Остается заметить, что

$$\mu_{1\lambda} = \frac{1}{\lambda} \mathbf{E}g(\xi_\lambda)I(\xi_\lambda \geq \nu), \quad \mu_{2\lambda} = \frac{1}{\bar{\lambda}} \mathbf{E}g(\xi_\lambda)I(\xi_\lambda < \nu),$$

откуда

$$\lambda \bar{\lambda} (\mu_{1\lambda} - \mu_{2\lambda}) = \mathbf{E}g(\xi_\lambda)(\bar{\lambda}I(\xi_\lambda \geq \nu) - \lambda I(\xi_\lambda < \nu)) = \text{cov}(g(\xi_\lambda), I(\xi_\lambda \geq \nu)),$$

так как $\mathbf{E}I(\xi_\lambda \geq \nu) = H(t(\lambda)) = \lambda$. В итоге для σ_λ^2 получаем указанный в формулировке теоремы вид.

В качестве примера конкретных применений теоремы 4 рассмотрим время ожидания $\nu(N, k)$. В этом случае $g(x) = x$, поэтому

$$\begin{aligned} \mu_\lambda &= \mathbf{E}\xi_\lambda = n\alpha_\lambda = n(1 - e^{-t(\lambda)}), \\ d_\lambda(g) &= d_\lambda = (\mathbf{D}\xi_\lambda)t'(\lambda)/\alpha_\lambda = nt'(\lambda)e^{-t(\lambda)}. \end{aligned} \quad (9)$$

Далее, из формулы

$$H'(t) = \sum_{r=0}^n \mathbf{P}\{\nu \leq r\} b_r(n, \alpha(t)) \left(\frac{r}{\alpha(t)} - n \right) = \frac{1}{\alpha(t)} \mathbf{E}\xi(t)I(\xi(t) \geq \nu) - nH(t)$$

при $t = t(\lambda)$ следует, что

$$\mathbf{E}\xi_\lambda I(\xi_\lambda \geq \nu) = (H'(t(\lambda)) + \lambda n)\alpha_\lambda = \left(\frac{1}{t'(\lambda)} + \lambda n \right) \alpha_\lambda.$$

Отсюда

$$\text{cov}(\xi_\lambda, I(\xi_\lambda \geq \nu)) = \alpha_\lambda / t'(\lambda)$$

и, следовательно,

$$\begin{aligned} \sigma_\lambda^2 &= \mathbf{D}\xi_\lambda + \lambda \bar{\lambda} d_\lambda^2 - 2d_\lambda \text{cov}(\xi_\lambda, I(\xi_\lambda \geq \nu)) \\ &= \lambda \bar{\lambda} n^2 (t'(\lambda))^2 e^{-2t(\lambda)} - ne^{-t(\lambda)}(1 - e^{-t(\lambda)}). \end{aligned} \quad (10)$$

Таким образом, в условиях теоремы 4 с.в. $\nu(N, k)$ асимптотически имеет нормальное распределение $\mathcal{N}(N\mu_\lambda, N\sigma_\lambda^2)$, где μ_λ и σ_λ^2 имеют соответственно вид (9) и (10).

Рассмотрим, наконец, случай $\lambda = p > 0$. Пусть m_1 , l и a определены так же, как в параграфе 3, и $E(x)$ есть функция вырожденного в нуле распределения, а $F(x)$ — функция распределения модуля стандартной нормальной с.в. Введем дополнительно обозначения

$$c = pq \binom{n}{2m_1} \left(a \binom{n}{m_1} \right)^{-2}, \quad d_l = g(l) \binom{n}{l} (\sqrt{pq})^{l/m_1} \left(a \binom{n}{m_1} \right)^{-l/m_1}.$$

Теорема 5. Пусть $p > 0$ и $N, k \rightarrow \infty$ так, что $k = pN + o(\sqrt{N})$. Тогда если $l > 2m_1$, то L_{Nk} сходится по вероятности к нулю, если $l = 2m_1$, то с.в. $g^{-1}(2m_1)L_{Nk}$ имеет в пределе дискретное распределение $\{p_r, r = 0, 1, \dots\}$, где

$$p_0 = \frac{1}{2}(1 + (1 + 2c)^{-1/2}), \quad p_r = \frac{(2r)!(1 + 2c)^{-1/2}}{(r!)^2 2^{2r+1}} \left(\frac{2c}{1 + 2c} \right)^r, \quad r \geq 1,$$

если $l < 2m_1$, то

$$\mathbf{P}\{d_l^{-1}N^{(l-2m_1)/(2m_1)}L_{Nk} \leq x\} \rightarrow \frac{1}{2}(E(x) + F(x^{m_1/l})), \quad x > 0.$$

Доказательство. Доказательство повторяет ход рассуждений теоремы 1. Здесь

$$x_N = p + x\sqrt{pq/N}, \quad x > 0,$$

и $t_N \rightarrow 0$ при $N \rightarrow \infty$. Из представления

$$H(t) = p + \binom{n}{m_1} at^{m_1}(1 + o(1)), \quad t \rightarrow +0,$$

находим, что при $N \rightarrow \infty$

$$t_N = \left(x\sqrt{pq/N} \left(a \binom{n}{m_1} \right)^{-1} \right)^{1/m_1} (1 + o(1)).$$

Отсюда следует, что

$$f_0(\tau; t_N) = e^{i\tau g(m_1)}(1 + o(1)),$$

$$f_1(\tau; t_N) = 1 + \frac{1}{x_N} \sum_{r=l}^n (e^{i\tau g(r)} - 1) \mathbf{P}\{\nu \leq r\} b_r(n, \alpha(t_N))$$

$$= 1 + (e^{i\tau g(l)} - 1) \mathbf{P}\{\nu \leq l\} \frac{1}{p} \binom{n}{l} t_N^l (1 + o(1)),$$

$$f_2(\tau; t_N) = 1 + \frac{1}{1 - x_N} \sum_{r=l}^n (e^{i\tau g(r)} - 1) \mathbf{P}\{\nu > r\} b_r(n, \alpha(t_N))$$

$$= 1 + (e^{i\tau g(l)} - 1) \mathbf{P}\{\nu > l\} \frac{1}{q} \binom{n}{l} t_N^l (1 + o(1)).$$

Отсюда, в свою очередь, следует, что при $l > 2m_1$

$$f_{Nk}(\tau; t_N) \rightarrow 1,$$

при $l = 2m_1$

$$f_{Nk}(\tau; t_N) \rightarrow \exp\{(e^{i\tau g(2m_1)} - 1)cx^2\},$$

при $l < 2m_1$

$$f_{Nk}(N^{(l-2m_1)/(2m_1)}\tau; t_N) \rightarrow \exp\{i\tau d_l x^{l/m_1}\}.$$

Обращая соответствующие х.ф.

$$\frac{1}{2} + \int_0^\infty f(\tau; x) \varphi(x) dx$$

для каждого из этих случаев, получаем утверждение теоремы.

Список литературы

1. Иванов В. А. Предельные теоремы в схеме размещения со случайными уровнями. *Матем. заметки* (1982) **31**, №4, 619–631.
2. Ивченко Г. И. Разделимые статистики в схеме со случайными уровнями. В кн.: *Вероятностные задачи дискретной математики*. МИЭМ, Москва 1987, с. 151–161.
3. Ивченко Г. И. Разделимые статистики в обратной задаче о размещении. *Дискретная математика* (1989) **1**, №1, 60–73.
4. Ивченко Г. И. Время ожидания и связанные с ним характеристики в полиномиальной схеме. *Дискретная математика* (1993) **5**, №3, 3–34.
5. Ивченко Г. И., Иванов А. В. Разделимые статистики в обратных урновых задачах. *Дискретная математика* (1995) **7**, №2, 103–117.
6. Ивченко Г. И., Левин В. В. Асимптотическая нормальность в схеме выбора без возвращения. *Теория вероятностей и ее применения* (1978) **23**, №1, 97–108.
7. Трунов А. Н. Об асимптотической нормальности класса разделимых статистик в схеме выбора без возвращения из мультимножеств. В кн.: *Вероятностные процессы и их приложения*. МИЭМ, Москва, 1985, с. 109–116.
8. Flatto L. Limit theorems for some random variables associated with urn models. *Ann. Probab.* (1982) **10**, №4, 927–934.
9. Holst L. On sequential occupancy problems. *J. Appl. Probab.* (1981) **18**, №2, 435–442.
10. Holst L. On birthday, collector's, occupancy and other classical urn problems. *Intern. Statist. Rev.* (1986) **54**, №1, 15–27.
11. Holst L., Husler J. Sequential urn schemes and birth processes. *Adv. Appl. Probab.* (1985) **17**, 257–279.
12. Janson S. Limit theorems for some sequential occupancy problems. *J. Appl. Probab.* (1983) **20**, №3, 545–553.
13. Pratt W. On interchanging limits and integrals. *Ann. Math. Statist.* (1960) **31**, №1, 74–77.

Статья поступила 16.08.95.