Г. И. Ивченко, Время ожидания и связанные с ним характеристики в полиномиальной схеме, *Дискрет. матем.*, 1993, том 5, выпуск 3, 3–34

Дискретная математика

том 5 выпуск 3 * 1993

УДК 519.24

ВРЕМЯ ОЖИДАНИЯ И СВЯЗАННЫЕ С НИМ ХАРАКТЕРИСТИКИ В ПОЛИНОМИАЛЬНОЙ СХЕМЕ

Г.И. Ивченко

Пусть в полиномиальной схеме с N исходами испытания проводятся до момента, когда впервые появится k исходов, $1 \leqslant k \leqslant N$, частоты которых достигнут или превзойдут заданные уровни, и пусть

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

есть разделимая статистика от частот $\eta_1, ..., \eta_N$ исходов в момент остановки испытаний, где g_j – некоторые функции целочисленного аргумента. Излагаются накопленные к настоящему времени результаты точного и асимптотического (при $N \to \infty$, $k = k(N)$) характера о распределениях статистик L_{Nk} и их различных конкретизаций, а также применения этих результатов к задачам статистического вывода для полиномиальной модели.

1. Введение. Рассмотрим неограниченную последовательность полиномиальных испытаний с N исходами, вероятности которых обозначим $p_1, ..., p_N$ ($p_1 + ... + p_N = 1$), и предположим, что до начала испытаний для j-го исхода установлен некоторый "уровень" ν_j, $j = 1, ..., N$, причем в общем случае будем считать, что $\nu_1, ..., \nu_N$ – независимые неотрицательные целочисленные случайные величины (с.в.). Испытания проводятся до момента, когда впервые появятся k исходов, числа реализаций которых достигнут либо превзойдут соответствующие уровни. Этот момент (момент остановки) обозначим $\nu(N, k)$, введем также с.в. η_j – число реализаций j-го исхода в момент остановки, $j = 1, ..., N$, и рассмотрим разделимую статистику (р.ст.) общего вида от частот $\eta_1, ..., \eta_N$

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

где $g_1, ..., g_N$ – произвольные заданные функции целочисленного аргумента (без потери общности можно предполагать, что $g_j(0) = 0$, $j = 1, ..., N$).

Определенная в (1) с.в. L_{Nk} является сводной характеристикой, соединяющей в себе два основных направления современных исследований полиномиальной схемы (или полиномиальных размещений): направление, связанное с изучением статистик типа времени ожидания $\nu(N, k)$, и направление, получившее в литературе название "разделимые статистики". Отметим, что если в (1) все функции $g_j(x) \equiv x$, то

$L_{Nk} = \sum_{j=1}^{N} \eta_j = \nu(N,k)$, т.е. само время ожидания (момент остановки) также является р.ст. вида (1).

В последние годы оба указанных направления интенсивно разрабатывались, и целью настоящего обзора является изложение накопленных в этой области результатов, связанных с временем ожидания (в обзор включен и ряд еще не опубликованных результатов по данной тематике). Наиболее продвинутыми на сегодня являются исследования самой с.в. $\nu(N, k)$. Первые глубокие результаты здесь связаны с классической схемой: равновероятные испытания $\left(p_1 = \ldots = p_N = \dfrac{1}{N}\right)$ и одинаковый для всех исходов детерминированный уровень ($\nu_1 = \ldots = \nu_N = m,\ m \geqslant 1$) (Эрдеш и Реньи, 1961 г. [27]) – для этого случая для времени ожидания будем использовать обозначение $\nu_m(N, k)$. Литература, посвященная этой с.в. огромна и соответствующие результаты с различной степенью детализации неоднократно освещались в ряде обзоров (В.Ф. Колчин, В.П. Чистяков, 1974 г. [22]; В.Ф. Колчин, Б.А. Севастьянов, В.П. Чистяков, 1976 г. [23]; В.А. Иванов, Г.И. Ивченко, Ю.И. Медведев, 1984 г. [3]; Л. Холст, 1986 г. [32]).

Однако общие обзоры грешат тем, что зачастую некоторые вопросы излагаются в них общими мазками, оставляя заинтересованным читателям непростую проблему добираться до деталей самостоятельно. Поэтому полное и замкнутое изложение конкретной темы (в данном случае – все, что относится к с.в. $\nu_m(N, k)$) представляется, на наш взгляд, самоценным, хотя некоторые повторы при этом, конечно, неизбежны.

Следующим принципиальным этапом было введение случайного уровня (В.А. Иванов, 1982 г. [1]). Время ожидания $\nu(N, k)$ в схеме со случайными уровнями интенсивно изучалось далее в работах В.А. Иванова, Г.И. Ивченко, А.М. Протасова (1984 г. [4]), В.А. Иванова (1986 г. [2]), Г.И. Ивченко (1987 г. [14], 1989 г. [15, 16]).

Другое интенсивное развивавшееся направление в этой области связано с изучением более общих характеристик типа р.ст. (1), которые описывают другие особенности процесса полиномиальных испытаний, связанные с моментами остановки $\nu(N, k)$. Впервые такого рода исследования были предприняты в работе Г.И. Ивченко (1973 г. [11]), где для равновероятной схемы полностью описан класс предельных распределений с.в.

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

где $I(\cdot)$ – индикатор, для момента остановки $\nu_m(N, k)$ при $N \to \infty$, $k = k(N)$ и фиксированном m, а также изучено в этих условиях асимптотическое строение вариационного ряда частот $\eta_{(1)} \leqslant \eta_{(2)} \leqslant \ldots \leqslant \eta_{(N)}$. В этой работе впервые было обнаружено, что в рассматриваемом случае частоты $\eta_1, \ldots, \eta_N$ ведут себя как условно независимые усеченные пуассоновские с.в. с общим случайным параметром (точные формулировки см. ниже), что позволяет записать распределение любой характеристики схемы в виде смеси соответствующего более просто устроенного параметрического семейства распределений. Основываясь на этом факте, в работе Г.И. Ивченко (1975 г. [13]) исследовано асимптотическое поведение эстремальных частот $\eta_{(N)}$ (при $k = N$) и $\eta_{(1)}$ (при $k = 1$) в зависимости от соотношений между уровнем $m = m(N)$ и числом исходов N при $N \to \infty$. В дальнейшем различные вопросы точного и асимптотического анализа распределений р.ст. в момент остановки $\nu_m(N, k)$ рассматривались Холстом (1981 г. [31], 1986 г. [32]), Холстом и Хюслером (1985 г. [33]), Флатто (1982 г. [28]), Янсоном (1983 г. [34]). Наконец, общие результаты для р.ст. (1) получены в работах Г.И. Ивченко (1989 г. [15, 16]).

Наличие общих теорем для с.в. типа времени ожидания $\nu(N, k)$ и р.ст. (1) открывает возможность использования этих с.в. в качестве тестовых статистик для конструирования новых статистических критериев проверки гипотез о вероятностях исходов $p_1, ..., p_N$; первые результаты в этом направлении были получены в работе Г.И. Ивченко (1974 г. [12]). Изложению накопленных к настоящему времени результатов по применению предельных теорем для р.ст. (1) к различным задачам статистического вывода для полиномиальной модели посвящен отдельный раздел 5 работы.

Перейдем к детальному изложению результатов обозначенного круга, идя при этом не в хронологическом порядке, а по принципу "от общего – к частному".

2. Точные результаты. Общее представление для характеристической функции (х.ф.) р.ст. (1) имеет вид [14, 15]

$$\mathsf{E}e^{it L_{Nk}} = \mathsf{P}\left(\sum_{j=1}^{N} I(\nu_j = 0) \geq k\right) + \sum_{l=1}^{N} \sum_{J_{kl}(N)} p_l \int_0^\infty \left(\sum_{r=1}^{\infty} e^{it g_l(r)} \mathsf{P}(\nu_l = r)\pi_{r-1}(p_l t)\right) \times$$
$$\times \prod_{s=1}^{k-1} \left(\sum_{r=0}^{\infty} e^{it g_{j_s}(r)} \mathsf{P}(\nu_{j_s} \leq r)\pi_r(p_{j_s} t)\right) \prod_{j=1}^{N}{}' \left(\sum_{r=0}^{\infty} e^{it g_j(r)} \mathsf{P}(\nu_j > r)\pi_r(p_j t)\right) dt, \quad (2)$$

где использованы обозначения $\pi_r(z) = e^{-z}z^r / r!$, $r = 0,1,\ldots,J_{kl}(N)$ означает область суммирования по всем $j_1, ..., j_{k-1}$, удовлетворяющим условиям $1 \leq j_1 < j_2 < \ldots$

$\ldots < j_{k-1} \leq N$, $j \neq l$, $\prod_{j=1}^{N}{}'$ означает произведение по всем $j \neq l$, $j_1,\ldots,j_{k-1}$. Формальному представлению (2) можно придать прозрачную вероятностную интерпретацию с помощью широко используемого в последнее время приема вложения полиномиальной модели в модель пуассоновского процесса. Именно, рассмотрим на полупрямой $(0, \infty)$ N независимых пуассоновских потоков $\Pi_1, ..., \Pi_N$ с интенсивностями соответственно $p_1, ..., p_N$ и пусть $\xi_j(t)$ обозначает число событий потока Π_j в интервале $(0, t)$: $\mathsf{P}(\xi_j(t) = r) = \pi_r(p_j t)$, $r = 0,1,\ldots$ Тогда объединенный поток $\Pi = \Pi_1 + ... + \Pi_N$ является также пуассоновским с интенсивностью $p_1 + ... + p_N = 1$, и можно представлять себе, что каждый раз, когда в объединенном потоке Π происходит событие, производится очередное полиномиальное испытание, в результате которого реализуется j-й исход, если событие принадлежит потоку Π_j. При такой интепретации модели вычисление вероятности любого события в полиномиальных испытаниях сводится к вычислению вероятности соответствующего события для независимых потоков $\Pi_1, ..., \Pi_N$ в интервале $(0, t)$ с последующим интегрированием по t.

В соответствии с этим, в задачах, связанных с моментами остановки типа $\nu(N, k)$, следует ввести с.в. $T_j = \min\{t\colon \xi_j(t) \geq \nu_j\}$, $j = 1,\ldots,N$, и рассмотреть момент $T_{(k)}$ – их k-ю порядковую статистику. Введем, далее, событие $A_t(l, \bar{j}) = A_t(l, j_1,\ldots,j_{k-1})$, означающее, что в момент $T_{(k)} = t$ в потоке Π_l происходит событие и при этом $\xi_l(t) = \nu_l$, в потоках же $\Pi_{j_1},\ldots,\Pi_{j_{k-1}}$ числа событий достигли либо превзошли соответствующие уровни до момента t, а в остальных потоках Π_j,

$j \neq l,\ j_1, \ldots, j_{k-1}$, это произойдет после t. В силу независимости потоков

$$\mathsf{P}(A_t(l,\bar{j}))dt = p_l \mathsf{P}(\xi_l(t-0) = \nu_l - 1) \prod_{s=1}^{k-1} \mathsf{P}(\xi_{j_s}(t) \geqslant \nu_{j_s}) \prod_j{}' \mathsf{P}(\xi_j(t) < \nu_j)dt. \tag{3}$$

При условии, что событие $A_t(l,\bar{j})$ произошло, число событий $\xi_{1j_s}(t)$ в потоке Π_{j_s} будет иметь усеченное пуассоновское распределение

$$\mathcal{L}(\xi_{1j_s}(t)) = \mathcal{L}(\xi_{j_s}(t) | \xi_{j_s}(t) \geqslant \nu_{j_s}), \tag{4}$$

а число событий $\xi_{2j}(t)$ в потоке Π_j при $j \neq l,\ j_1, \ldots, j_{k-1}$ – распределение

$$\mathcal{L}(\xi_{2j}(t)) = \mathcal{L}(\xi_j(t) | \xi_j(t) < \nu_j); \tag{5}$$

для числа же событий в потоке Π_l по условию выполняется равенство $\xi_l(t) = \nu_l$; при этом с.в. $\xi_{1j_1}(t), \ldots, \xi_{1j_{k-1}}(t),\ \xi_{2j}(t),\ j \neq l,\ j_1, \ldots, j_{k-1}$, независимы. В итоге при фиксированных уровнях $\nu_1, \ldots, \nu_N$ мы можем записать равенство

$$\mathsf{E}e^{i\tau L_{Nk}} = \mathsf{P}\left(\sum_{j=1}^{N} I(\nu_j = 0) \geqslant k \right) + \int_0^\infty \sum_{\{A_t(l,\bar{j})\}} e^{i\tau g_l(\nu_l)} \prod_{s=1}^{k-1} e^{i\tau g_{j_s}(\xi_{1j_s}(t))} \times$$

$$\times \prod_j{}' \mathsf{E}e^{i\tau g_j(\xi_{2j}(t))} \mathsf{P}(A_t(l,\bar{j}))dt, \tag{6}$$

которое с учетом (3)–(5) после усреднения по уровням (с учетом их независимости) сводится к (2).

Выделим представляющие интерес некоторые частные случаи общих представлений (2) и (6).

а) *Симметрическая схема* – под этим будем понимать ситуацию с равновероятными испытаниями $\left(p_1 = \ldots = p_N = \dfrac{1}{N}\right)$, однородными (т.е. одинаково распределенными) уровнями $\nu_1, \ldots, \nu_N$, когда они являются независимыми копиями некоторой неотрицательной целочисленной с.в. ν с $p = \mathsf{P}(\nu = 0) < 1$, и одинаковыми функциями $g_1 = \ldots = g_N = g$. Для этого случая формула (2) принимает вид

$$\mathsf{E}e^{i\tau L_{Nk}} = B(p;\ k, N-k+1) + NC_{N-1}^{k-1} \int_0^\infty \left(\sum_{r=1}^{\infty} e^{i\tau g(r)} \mathsf{P}(\nu = r) \pi_{r-1}(t) \right) \times$$

$$\times \left(\sum_{r=0}^{\infty} e^{i\tau g(r)} \mathsf{P}(\nu \leqslant r) \pi_r(t) \right)^{k-1} \left(\sum_{r=0}^{\infty} e^{i\tau g(r)} \mathsf{P}(\nu > r) \pi_r(t) \right)^{N-k} dt, \tag{7}$$

где

$$B(p;\ k, N-k+1) = \sum_{j=k}^{N} C_N^j p^j q^{N-j}, \quad q = 1 - p.$$

Если ввести функцию $\overline{H}(t) = 1 - H(t) = \mathsf{P}(\xi(t) \geqslant \nu)$, где $\mathcal{L}(\xi(t)) = \Pi(t)$ – пуассоновский закон с параметром $t \geqslant 0$, и определить для $x \in [p, 1]$ обратную к ней функцию $t(x) = \overline{H}^{-1}(x)$, то, произведя в (7) замену переменных по формуле

$\overline{H}(t) = x$, получим представление [15]

$$\mathsf{E}e^{i\tau L_{Nk}} = B(p;\ k, N-k+1) +$$

$$+\int_p^1 f_0(\tau;\ t(x)) f_1^{k-1}(\tau;\ t(x)) f_2^{N-k}(\tau;\ t(x)) \beta(x;\ k, N-k+1)\ dx, \qquad (8)$$

где $\beta(x;\ k, N-k+1) = NC_{N-1}^{k-1} x^{k-1}(1-x)^{N-k}$, $x \in [0,1]$, есть бета-плотность, $f_j(\tau; t)$, $j = 0,\ 1,\ 2$, – характеристические функции независимых с.в. $g(\xi_j(t))$, $j = 0,\ 1,\ 2$, а $\xi_j(t)$ получаются усечениями с.в. $\xi(t)$:

$$\mathcal{L}(\xi_0(t)) = \mathcal{L}(\xi(t) + 1 | \xi(t) = \nu - 1),$$

$$\mathcal{L}(\xi_1(t)) = \mathcal{L}(\xi(t) | \xi(t) \geqslant \nu),\ \ \mathcal{L}(\xi_2(t)) = \mathcal{L}(\xi(t) | \xi(t) < \nu).$$

Представлению (8) можно придать следующую вероятностную интерпретацию. Предположим, что до начала испытаний производится эксперимент, в котором наблюдается с.в. θ, имеющая бета-распределение с плотностью $\beta(x;\ k,\ N - k + 1)$. Тогда при $\theta \leqslant p = \mathsf{P}(\nu = 0)$ полагается $L_{Nk} = 0$, если же наблюдается $p < \theta \leqslant 1$, то статистика L_{Nk} имеет распределение вида

$$\mathcal{L}(L_{Nk}) = \mathcal{L}\left(g(\xi_0(t(\theta))) + \sum_{j=1}^{k-1} g(\xi_{1j}(t(\theta))) + \sum_{j=1}^{N-k} g(\xi_{2j}(t(\theta))) \right), \qquad (9)$$

где слагаемые условно независимы (т.е. при каждм θ они независимы) и $\xi_{1j}(t)(\xi_{2j}(t))$ распределены одинаково с $\xi_1(t)$ (с $\xi_2(t)$). Тем самым (при $p < \theta \leqslant 1$) частоты исходов в момент остановки $\nu(N,\ k)$ ведут себя как условно независимые усеченные пуассоновские с.в. $\xi_0(t(\theta))$, $\xi_{1j}(t(\theta))$, $j = 1, \ldots, k-1$, $\xi_{2j}(t(\theta))$, $j = 1, \ldots, N-k$, и, следовательно, распределение любой характеристики в рассматриваемой модели можно записать в виде смеси соответствующего семейства распределений, определяемого через эти с.в.

Для случая вырожденного уровня, когда $\mathsf{P}(\nu = m) = 1$ для некоторого $m \geqslant 1$, формула (8) принимает вид [31, 33]

$$\mathsf{E}e^{i\tau L_{Nk}} = e^{i\tau g(m)} \int_0^1 f_1^{k-1}(\tau;\ t_m(x)) f_2^{N-k}(\tau;\ t_m(x)) \beta(x;\ k, N-k+1)\ dx =$$

$$= e^{i\tau g(m)} NC_{N-1}^{k-1} \int_0^1 \left(\sum_{r \geqslant m} e^{i\tau g(r)} \pi_r(t_m(x)) \right)^{k-1} \left(\sum_{r<m} e^{i\tau g(r)} \pi_r(t_m(x)) \right)^{N-k} dx, \qquad (10)$$

где $t_m(x)$ – функция, обратная к

$$\overline{\Pi}_{m-1}(t) = \mathsf{P}(\xi(t) \geqslant m) = 1 - \sum_{r=0}^{m-1} \pi_r(t) = 1 - \Pi_{m-1}(x)$$

(эти обозначения используются везде в дальнейшем).

б) *Время ожидания.* Из (8) при $g(x) \equiv x$ следует интегральное представление для производящей функции времени ожидания $\nu(N,\ k)$ в симметрической схеме [1, 4]

$$\mathsf{E}z^{\nu(N,k)} = B(p;\ k, N-k+1) + \int_p^1 \exp\{(1 - z^{-1}) N t(x)\} \beta(x;\ k, N-k+1)\ dx. \qquad (11)$$

Для с.в. $\nu_m(N,\ k)$ (вырожденный в точке m уровень) формула (11) конкретизируется следующим образом:

$$\mathsf{E}z^{\nu_m(N,k)} = \int_0^1 \exp\{(1-z^{-1})Nt_m(x)\}\beta(x;\ k, N-k+1)\ dx, \tag{11$'$}$$

где $t_m(x)$ указано в (10). Впервые представление (11′) было найдено А. Бекеши в 1964 г. [25], и позже "переоткрыто" в [31]. Для частного случая $m = 1$ формула (11′) сводится к хорошо известному виду

$$\mathsf{E}z^{\nu_1(N,k)} = \frac{N!}{(N-k)!} z^k \prod_{j=0}^{k-1} (N-jz)^{-1}.$$

В случае неравновероятных испытаний аналогом (11′) является представление [4, 8]

$$\mathsf{E}z^{\nu_m(N,k)} = 1 - \frac{1-z}{z} \int_0^\infty e^{t(1-z^{-1})} \sum_{r=0}^{k-1} \sum_{1 \leqslant j_1 < \ldots < j_r \leqslant N} \prod_{s=1}^{r} \overline{\Pi}_{m-1}(p_{j_s}t) \prod_j{}' \Pi_{m-1}(p_j t)\ dt \tag{12}$$

(штрих означает, что произведение берется по $j \neq j_1, \ldots j_r$), для простейшего случая $m = 1$ оно сводится к виду [8]

$$\mathsf{E}z^{\nu_1(N,k)} = 1 + (1-z)\sum_{r=0}^{k-1} (-1)^{k-r} C_{N-r-1}^{N-k} \sum_{1 \leqslant j_1 < \ldots < j_r \leqslant N} \frac{1}{1 - z(p_{j_1} + \ldots + p_{j_r})}. \tag{12$'$}$$

в) *Индикаторные характеристики.* Пусть A – некоторое подмножество $\{0,\ 1,\ 2, \ldots\}$ и $\mu_A = \sum_{j=1}^{N} I(\eta_j \in A)$, при $A = \{r\}$ пишем μ_r. Тогда из (9) следует, что $\mu_A = NI(0 \in A)$, если $\theta \leqslant p$, а при $\theta > p$

$$\mathcal{L}(\mu_A) = \mathcal{L}(Y_0(1, p_{0A}(\theta)) + Y_1(k-1, p_{1A}(\theta)) + Y_2(N-k, p_{2A}(\theta))), \tag{13}$$

где $p_{jA}(x) = \mathsf{P}(\xi_j(t(x)) \in A)$, $j = 0,1,2$, слагаемые условно независимы и $\mathcal{L}(Y.(n, \alpha)) = \mathrm{Bi}(n, \alpha)$ – биномиальный закон с параметрами n и α.

Если $\mathsf{P}(\nu = m) = 1$ для некоторого $m \geqslant 1$, то из (13) для с.в. μ_r получаем следующие представления [11]:

$$\mathcal{L}(\mu_r) = \begin{cases} \mathrm{Bi}\left(N-k,\ \dfrac{1}{1-\theta}\pi_r(t_m(\theta))\right), & \text{если } r < m, \\ \mathrm{Bi}\left(k-1,\ \dfrac{1}{\theta}\pi_r(t_m(\theta))\right), & \text{если } r > m, \end{cases} \tag{13$'$}$$

наконец,

$$\mathcal{L}(\mu_m - 1) = \mathrm{Bi}\left(k-1,\ \frac{1}{\theta}\pi_m(t_m(\theta))\right).$$

Тем самым в данном случае распределения с.в. μ_r представляют собой смеси биномиальных распределений, в которых вероятности успехов являются специфическими функциями от с.в. θ. С.в. μ_1 для случая $m = 1$, $k = N$ изучалась также в [31], а моменты $E\mu_r$ в общем случае – в [32] (см. п.г) раздела 6).

г) *Порядковые статистики.* Рассмотрим для примера максимальную частоту $\eta_{(N)} = \max(\eta_1, \ldots, \eta_N)$. Введем множества $A(s) = \{0,\ 1, \ldots, s\}$ и $\overline{A}(s) = \{j\!:\ j > s\}$, $s \geqslant 0$. Тогда с учетом (13) для распределения с.в. $\eta_{(N)}$ справедливо представление [15]

$$\mathsf{P}(\eta_{(N)} \leqslant s) = \mathsf{P}(\mu_{\overline{A}(s)} = 0) = B(p;\ k, N-k+1) +$$

$$+\int_p^1 p_{0A(s)}(x)(p_{1A(s)}(x))^{k-1}(p_{2A(s)}(x))^{\nu-k}\beta(x;\ k, N-k+1)\ dx. \tag{14}$$

Для случая вырожденного в точке m уровня формула (14) конкретизируется следующим образом [11]: $\mathsf{P}(\eta_{(N)} \leqslant s) = 0$ при $s < m$, а при $s \geqslant m$

$$\mathsf{P}(\eta_{(N)} \leqslant s) = \int_0^1 \left(1 - \frac{1}{x}\overline{\Pi}_s(t_m(x))\right)^{k-1} \beta(x;\ k, N-k+1)\ dx. \tag{14$'$}$$

Аналогичные представления можно записать и для других членов вариационного ряда $\eta_{(1)} \leqslant \eta_{(2)} \leqslant \ \dots\ \leqslant \eta_{(N)}$, построенного по с.в. $\eta_1, \dots, \eta_N$. В качестве иллюстрации укажем еще вид распределения с.в. $\eta_{(1)}$ в случае вырожденного уровня: $\mathsf{P}(\eta_{(1)} > s) = 0$ при $s \geqslant m-1$, а при $s < m-1$ (что возможно лишь при $m > 1$)

$$\mathsf{P}(\eta_{(1)} > s) = \int_0^1 \left(1 - \frac{1}{1-x}\Pi_s(t_m(x))\right)^{N-k} \beta(x;\ k, N-k+1)\ dx; \tag{14$''$}$$

тем самым с.в. $\eta_{(1)}$ принимает лишь значения 0, 1, ..., $m-1$.

д) *Схема с выделенными исходами.* В ряде приложений представляют интерес схемы с "неравноправными" исходами, т.е. когда следят за частотами лишь некоторых выделенных $N_1 < N$ исходов (далее это будут исходы с номерами 1, ..., N_1). Таким образом, в данном случае испытания проводятся до тех пор, пока впервые среди выделенных N_1 исходов появятся k, $1 \leqslant k \leqslant N_1$, исходов, частоты которых достигнут либо превзойдут соответствующие уровни $\nu_1, \dots, \nu_{N_1}$, устанавливаемые до начала испытаний. Чтобы изучать подобного рода схемы, в представлении (2) надо формально положить для $j > N_1$ $\mathsf{P}(\nu_j = \infty) = 1$. Итогом является общая формула для х.ф. [15]

$$\mathsf{E}e^{i\tau L_{Nk}} = \mathsf{P}\left(\sum_{j=1}^{N_1} I(\nu_j = 0) \geqslant k\right) + \sum_{l=1}^{N_1} \sum_{J_{kl}(N_1)} p_l \int_0^\infty \left(\sum_{r=1}^\infty e^{i\tau g_l(r)} \mathsf{P}(\nu_l' = r)\pi_{r-1}(p_l t)\right) \times$$

$$\times \prod_{s=1}^{k-1} \left(\sum_{r=0}^\infty e^{i\tau g_{j_s}(r)} \mathsf{P}(\nu_{j_s} \leqslant r)\pi_r(p_{j_s} t)\right) {\prod_{j=1}^{N_1}}' \left(\sum_{r=0}^\infty e^{i\tau g_j(r)} \mathsf{P}(\nu_j > r)\pi_r(p_{j_s} t)\right) \times$$

$$\times \prod_{j=N_1+1}^{N} \left(\sum_{r=0}^\infty e^{i\tau g_j(r)} \pi_r(p_j t)\right)\ dt,$$

которая для симметрической по отношению к выделенным исходам ситуации ($p_1 = \ \dots\ = p_{N_1} = \alpha$, уровни $\nu_1, \dots, \nu_{N_1}$ – независимые копии с.в. ν, функции $g_1 = \ \dots\ = g_{N_1} = g$) принимает вид

$$\mathsf{E}\exp\left\{i\tau\left(\sum_{j=1}^{N_1} g(\eta_j) + \sum_{j=N_1+1}^{N} g_j(\eta_j)\right)\right\} = B(p;\ k, N_1 - k + 1) +$$

$$+\int_p^1 f_0(\tau;\ t(x)) f_1^{k-1}(\tau;\ t(x)) f_2^{N_1-k}(\tau;\ t(x)) \times$$

$$\times \prod_{j=N_1+1}^{N} \varphi_j(\tau;\ t(x))\beta(x;\ k, N_1-k+1)\ dx, \tag{15}$$

где дополнительно к обозначениям в (8)

$$\varphi_j(\tau;\ t) = \mathsf{E}\exp\left\{i\tau g_j\left(\xi\left(\frac{p_j}{\alpha}t\right)\right)\right\},\quad j = N_1+1,\dots,N.$$

В качестве примера использования представления (15) в [15] получено следующее выражение для совместной х.ф. частот $\eta_2, \dots, \eta_N$ в момент достижения частотой 1-го исхода заданного уровня ν с $\mathsf{E}z^{\nu} = A(z)$:

$$\mathsf{E}\exp\left\{i\sum_{j=2}^{N} z_j\eta_j\right\} = A\left(p_1\left(1-\sum_{j=2}^{N} p_j e^{iz_j}\right)^{-1}\right). \tag{16}$$

Распределение с х.ф. (16) называется обобщенным отрицательным полиномиальным распределением и его детальный вероятностный и статистический анализ проведен в [19, 20] (см. п.д) раздела 5).

3. Предельные теоремы. Основное внимание в рассматриваемых задачах уделяется исследованию асимптотического при $N \to \infty$ поведения распределений соответствующих характеристик. При общем подходе основой для этого может служить представление (2) (или (6)) для х.ф. произвольной р.ст. (1). Однако, в общем случае эти представления "перегружены" степенями свободы (произвольные функции g_j, произвольные уровни ν_j, произвольные вероятности исходов p_j), поэтому содержательные результаты в асимптотической теории удается получить, как это обычно бывает в подобных случаях, лишь при "разумных" конкретизациях модели. Важным для приложений примером подобной конкретизации (и в то же время сохраняющим достаточную общность модели) является симметрическая схема, описанная в п.а) предыдущего раздела. Излагаемые далее в этом разделе результаты относятся именно к этой схеме [15, 16].

Итак, пусть при $N \to \infty$ параметр $k = k(N)$ меняется так, что выполняется одно из трех условий:

1) $k \geqslant 1$ фиксировано, 2) $s = N - k \geqslant 0$ фиксировано, 3) $k = \lambda N + o(\sqrt{N})$, $\lambda = 1 - \bar{\lambda} \in (0,1)$. В случаях 1) и 2) мы говорим об "экстремальных" значениях k, а в случае 3) – о "средних" значениях. Относительно распределения уровня ν далее предполагается, что оно фиксировано и не зависит от N; кроме того, будем считать далее, что в случае 1) параметр $p = \mathsf{P}(\nu = 0) = 0$, а в случае 3) $\lambda \geqslant p$ (это исключает "тривиальные" ситуации); напомним также, что $g(0) = 0$.

1) *Первый экстремальный случай.* Определим целые $m, l \geqslant 1$ условиями

$$m = \min\{r \geqslant 1:\ \mathsf{P}(\nu = r) > 0\},\quad l = \min\{r:\ g(r) \neq 0\}.$$

Оказывается, что асимптотическое поведение р.ст. L_{Nk} в этом случае определяется лишь этими локальными характеристиками.

Теорема 1. *Если $N \to \infty$, а $k \geqslant 1$ фиксировано, то при $l > m$ $L_{Nk} \xrightarrow{\mathsf{P}} 0$; при $l = m$*

$$\mathcal{L}(g^{-1}(m)L_{Nk} - k) \to \overline{\mathrm{Bi}}(k, a),$$

где $a = \mathsf{P}(\nu = m)$ и $\overline{\mathrm{Bi}}(k,a)$ – отрицательное биномиальное распределение, задаваемое вероятностями $C_{k+n-1}^{n} a^k (1-a)^n$, $n = 0,1,\dots,$; при $l < m$

$$\mathsf{P}(N^{(l-m)/m} l!\, g^{-1}(l) L_{Nk} \leqslant x) \to \overline{\Pi}_{k-1}\left(\frac{a}{m!} x^{m/l}\right),\quad x > 0.$$

Применяя эту теорему к времени ожидания $\nu(N, k)$ находим, что (здесь $l = 1$) при $m = 1$

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

а при $m > 1$

$$\mathsf{P}(\nu(N,k)/N^{1-1/m} \leqslant x) \to \overline{\Pi}_{k-1}\left(\frac{a}{m!}x^m\right), \quad x > 0. \tag{17}$$

Впервые эти результаты для с.в. $\nu(N, k)$ были получены в [1]. Подчеркнем также, что классический случай вырожденного в точке $m \geqslant 1$ уровня соответствует значению $a = 1$, в этом случае под $\overline{\mathrm{Bi}}(k,a)$ понимается вырожденное в нуле распределение. Таким образом, "случайность" уровня ν делает схему "богаче" – появляется новый предельный закон $\overline{\mathrm{Bi}}(k,a)$. Утверждение (17) для с.в. $\nu_m(N, k)$, $m > 1$, (т.е. при $a = 1$) впервые было установлено в [10].

В качестве еще одного примера рассмотрим применение теоремы 1 к с.в. μ_r (здесь $l = r$): при $r > m$ с.в. $\mu_r \xrightarrow{\mathsf{P}} 0$; $\mathcal{L}(\mu_m - k) \to \overline{\mathrm{Bi}}(k,a)$; при $1 \leqslant r < m$

$$\mathsf{P}(N^{(r-m)/m} r!\mu_r \leqslant x) \to \overline{\Pi}_{k-1}\left(\frac{a}{m!}x^{m/r}\right), \quad x > 0;$$

с.в. $N - \mu_0$ имеет такое же предельное распределение, как и с.в. μ_1. Для случая вырожденного уровня (т.е. при $a = 1$) эти результаты были получены в [11], где имеется также и следующее "многомерное" утверждение: если $1 \leqslant r_1 < r_2 < \ldots$ $\ldots < r_\tau < m$, то

$$\mathsf{P}(N^{(r_j-m)/m}\mu_{r_j} \leqslant x_j, \; j = 1,\ldots,\tau) \to \overline{\Pi}_{k-1}\left(\frac{1}{m!} \min_{1\leqslant j\leqslant\tau} (r_j! x_j)^{m/r_j}\right).$$

2) *Второй экстремальный случай.* Содержательные результаты здесь можно получить лишь при определенных ограничениях на поведение функции $H(t) = \mathsf{P}(\xi(t) < \nu)$ при $t \to \infty$ (см. (8)). Для случая ограниченного уровня, т.е. когда $m = \max\{r\colon \mathsf{P}(\nu = r) > 0\} < \infty$, общий результат имеет следующий вид.

Обозначим $t_N = t\left(1 - \dfrac{x}{N}\right)$ (см. (8)), $b = 1 - \overline{b} = \mathsf{P}(\nu = m)$ и пусть функция $g = g_N$ удовлетворяет условиям (A): существуют $\lim\limits_{N\to\infty} g_N(m-1) = a$ и для любого $x > 0$

$$\lim_{N\to\infty} N \sum_{r=m}^{\infty} (e^{i\tau g_N(r)} - 1)\pi_r(t_N) = a(i\tau; x).$$

Т е о р е м а 2. *Если N, $k \to \infty$ так, что $s = N - k \geqslant 0$ постоянно, и выполнено условие* (A), *то*

$$\lim_{N\to\infty} \mathsf{E}e^{i\tau L_{Nk}} = \int_0^\infty f(\tau; x)\pi_s(x)\, dx,$$

где

$$f(\tau; x) = \exp\left\{i\tau s a + \frac{\overline{b}x}{b}(e^{i\tau a} - 1) + a(i\tau; x)\right\}.$$

В этой теореме содержится результат работы [34] для случая вырожденного уровня (т.е. при $b = 1$), где дополнительно предполагаются выполненными условия $\lim_{N\to\infty} g_N(r) = 0,\ r = 0,1,\ldots,m-1$.

Применяя теорему 2 к с.в. $\nu(N, k)$ находим, что (здесь $g_N(x) = (x - c_N)/N$, $c_N = \ln N + (m-1)\ln\ln N - \ln((m-1)!/b)$):

$$\mathsf{P}(N^{-1}\nu(N,k) - c_N \leqslant x) \to \Pi_s(e^{-x}),\quad -\infty < x < \infty. \tag{18}$$

Классический результат для с.в. $\nu_m(N, k)$ соответствует здесь значению $b = 1$ и получен впервые в [27] для случая $s = 0$, а для произвольного s – в [25].

Более детальный анализ для с.в. $\nu(N, k)$ (по сравнению с теоремой 2) проведен в работах [1] и [4]. Оказалось, что, помимо классического типа (18), здесь возникает и предельный закон нового типа, а именно, если для любого $\tau > 0$ выполняется условие $H(t)/H(\tau t) \to \tau^\alpha$ при $t \to \infty$, то

$$\mathsf{P}\left(\frac{\nu(N,k)}{Nt\left(1-\frac{1}{N}\right)} \leqslant x\right) \to \Pi_s(x^{-\alpha}),\quad x > 0. \tag{19}$$

Так, если при некотором целом $d \geqslant 1$

$$\mathsf{P}(\nu > n) = \frac{d!n!}{(n+d)!},\quad n = 0,1,\ldots,$$

то $H(t) = d!(1-\Pi_{d-1}(t))t^{-d}$ и будет выполняться (19) с $\alpha = d$ и $t(1 - 1/N) \sim (d!N)^{1/d}$. В этом случае у с.в. ν существуют (конечные) моменты лишь до порядка $d - 1$; в частности, если ν имеет бесконечное среднее ($d = 1$), то $\nu(N, k)$ растет как $O(N^2)$, если ν имеет конечное среднее, но бесконечный второй момент ($d = 2$), то $\nu(N, k)$ растет как $O(N^{3/2})$ и т.д.

Наиболее общие условия, приводящие к соотношению типа (18), формулируются следующим образом: определим две последовательности констант $b_N = t(1 - 1/N)$ и $a_N = t(1 - 1/(eN)) - t(1 - 1/N)$ и пусть $b_N / a_N \to \infty$, $b_N / (Na_N^2) \to 0$ при $N \to \infty$. Тогда

$$\mathsf{P}\left(\frac{\nu(N,k) - Nb_N}{Na_N} \leqslant x\right) \to \Pi_s(e^{-x}),\quad -\infty < x < \infty. \tag{20}$$

Указанные условия выполняются, например, [1] когда при $t \to \infty$

$$H(t) = at^\alpha e^{-\beta t}(1 + o(1)),\quad a,\beta > 0,\ \alpha \geqslant 0,$$

для этого случая $a_N = \beta^{-1} + o(1)$,

$$b_N = \frac{1}{\beta}\left(\ln N + \alpha \ln\ln N - \ln\frac{\beta^\alpha}{a}\right) + o(1).$$

В [1] приведены примеры встречающихся в приложениях конкретных схем размещения частиц по ячейкам, для которых имеют место соотношения (19)–(20).

Рассмотрим, наконец, применение теоремы 2 к с.в. μ_r: при $r < m - 1$ с.в. $\mu_r \xrightarrow{\mathsf{P}} 0$; $\mathcal{L}(\mu_{m-1} - s) \to \Pi(\bar{b}\zeta_s / b)$, где с.в. ζ_s имеет гамма-плотность $\pi_s(x)$, $x > 0$;

при $r \geqslant m$

$$\mathsf{P}\left(\frac{r!b}{(m-1)!}(\ln N)^{m-r-1}\mu_r \leqslant x\right) \to \overline{\Pi}_s(x), \quad x > 0. \tag{21}$$

Для случая вырожденного уровня (т.е. при $b = 1$) эти утверждения впервые были получены в [11] и позже "переоткрывались" для различных частных случаев в [28, 31, 32, 34]. Так, в [31] для случая $m = 1$, $s = 0$ установлена не только слабая сходимость с.в. $\mu_1/\ln N$ к экспоненциальному распределению $\overline{\Pi}_0(x)$, но и соответствующая сходимость производящей функции моментов, т.е. сходимость всех моментов.

В [11] доказан также следующий многомерный вариант утверждения (21) (при $b = 1$): если $m \leqslant r_1 < r_2 < \ldots < r_\tau$, то

$$\mathsf{P}\left(\frac{r_j!}{(m-1)!}(\ln N)^{m-1-r_j}\mu_{r_j} \leqslant x_j,\ j = 1,\ldots,\tau\right) \to \overline{\Pi}_s\left(\min_{1\leqslant j\leqslant\tau} x_j\right).$$

Близкий к (21) результат для случая $b = 1$ и $s = 0$ установлен в [28]: при $N \to \infty$ соотношение

$$\mathsf{P}(|r!e^x(\ln N)^{m-r-1}\mu_r - 1| > \varepsilon | \nu_m(N,N) = [N(\ln N + (m-1)\ln\ln N + x)]) \to 0$$

выполняется для любого $\varepsilon > 0$ равномерно по x в любом конечном интервале.

3) *Средние значения* k. При $k = \lambda N + o(\sqrt{N})$, $\lambda = 1 - \overline{\lambda} \in (0,1)$, асимптотическое поведение р.ст. L_{Nk} существенно различается для случаев $\lambda > p$ и $\lambda = p > 0$. Для первого из этих случаев характерна асимптотическая нормальность р.ст.

Теорема 3. *Пусть N, $k \to \infty$ так, что $\lambda > p$, и функция g такова, что* $\mathsf{D}g(\xi(t)) < \infty$ *для всех $t > 0$. Тогда* $\mathcal{L}(L_{Nk}) \sim \mathcal{N}(Na_\lambda(g), N\sigma_\lambda^2(g))$, *где* $a_\lambda(g) = \mathsf{E}g(\xi(t(\lambda)))$,

$$\sigma_\lambda^2(g) = \mathsf{D}\tilde{g}(\xi(t(\lambda))),\ \ \tilde{g}(\xi) = g(\xi) - d(g)I(\xi \geqslant \nu),$$

$$d(g) = \frac{t'(\lambda)}{t(\lambda)}\operatorname{cov}(g(\xi(t(\lambda))),\ \xi(t(\lambda))).$$

При этом параметры асимптотического распределения являются главными значениями среднего и дисперсии L_{Nk} соответственно.

Для случая вырожденного уровня ν этот результат совпадает с соответствующим результатом работы [33].

Если в теореме 3 $g(x) \equiv x$, то $a_\lambda(g) = t(\lambda)$, $\sigma_\lambda^2(g) = \lambda\overline{\lambda}(t'(\lambda))^2 - t(\lambda)$, и мы получаем утверждение об асимптотической нормальности времени ожидания $\nu(N, k)$ [1]; случай вырожденного уровня впервые был исследован в [25], а более детально – в [10]. Так, в [10] установлено, что если $m \geqslant 2$ фиксировано, то соотношение

$$\mathcal{L}(\nu_m(N,k)) \sim \mathcal{N}(Nt_m(\lambda),\ N(\lambda\overline{\lambda}(t_m'(\lambda))^2 - t_m(\lambda))) \tag{22}$$

справедливо при $N \to \infty$ для всех k, удовлетворяющих условиям k, $N - k \to \infty$ ($t_m(\lambda)$ определено в (10)). Случай $m = 1$ имеет некоторые особенности, впервые детально проанализированные в работе [24] (см. также [18]), а именно, асимптотическая нормальность с.в. $\nu_1(N, k)$ в зоне $k/N \to 0$ имеет место лишь при достаточно быстром росте k, а именно при $k/\sqrt{N} \to \infty$, и она сохраняется, пока $N - k \to \infty$.

Наконец, приведем конкретизацию параметров теоремы 3 для с.в. μ_r: $a_\lambda(g) = \pi_r(t(\lambda))$,

$$\sigma_\lambda^2(g) =$$

$$= \pi_r(t(\lambda))\left(1 - \pi_r(t(\lambda))\left(1 + 2(\mathsf{P}(\nu \leqslant r) - \lambda)(r - t(\lambda))\frac{t'(\lambda)}{t(\lambda)} - \lambda\bar{\lambda}(r - t(\lambda))^2\left(\frac{t'(\lambda)}{t(\lambda)}\right)^2\right)\right);$$

для случая вырожденного уровня этот результат получен в [11], где имеются также многомерная нормальная предельная теорема для этих с.в. и теоремы о предельных распределениях в зонах $k/N \to 0$ и $k/N \to 1$ (см. п.в) раздела 6).

"Граничные" эффекты для случая $\lambda = p > 0$ описываются следующим образом. Дополнительно к обозначениям в теореме 1 положим

$$c = \frac{pq(m!)^2}{a^2(2m)!}, \quad d_l = \frac{g(l)}{l!}\left(m!\frac{\sqrt{pq}}{a}\right)^{l/m}$$

и пусть $E(x)$ есть функция вырожденного в 0 распределения, а $F(x)$ – функция распределения модуля стандартной нормальной с.в.

Т е о р е м а 4. *Пусть $p > 0$ и $N, k \to \infty$ так, что $k = pN + o(\sqrt{N})$. Тогда $L_{Nk} \xrightarrow{\mathsf{P}} 0$ при $l > 2m$; при $l = 2m$ с.в. $g^{-1}(2m)L_{Nk}$ имеет в пределе дискретное распределение*

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

при $l < 2m$

$$\mathsf{P}(d_l^{-1}L_{Nk} / N^{1-l/2m} \leqslant x) \to \frac{1}{2}(\mathsf{E}(x) + F(x^{m/l})), \quad x > 0.$$

Для с.в. $\nu(N, k)$ утверждение теоремы 4 было получено впервые в [1].

4. Время ожидания (продолжение). Время ожидания $\nu_m(N, k)$ до не менее m-кратной реализации некоторых k исходов в полиномиальной схеме – весьма популярная в литературе по дискретным вероятностным задачам характеристика, и она исследована гораздо детальнее, чем это отражено в предыдущих разделах. Здесь мы приводим результаты, не следующие из общих теорем.

а) *Многомерные распределения.* (1) Введем нормированные с.в.

$$\nu_j^*(N) = \nu_j(N,1)N^{-1+1/j}(j!)^{-1/j}, \quad j = 2,\ldots,m.$$

Из (17) следует, что предельная при $N \to \infty$ функция распределения с.в. $\nu_j^*(N)$ есть $1 - e^{-x^j}$, $x > 0$. В работе [32] доказано, что при любом фиксированном $m > 2$ эти с.в. асимптотически независимы и $\mathsf{E}(\nu_j^*(N))^r \to \Gamma(1 + r/j)$.

(2) Рассмотрим теперь другой экстремальный случай и положим

$$\nu_j'(N) = N^{-1}\nu_j(N,N) - \ln N - (j-1)\ln\ln N + \ln(j-1)!, \quad j = 1,\ldots,m.$$

В работе [28] доказано, что при любом фиксированном $m > 1$ эти с.в. асимптотически (при $N \to \infty$) независимы и одинаково распределены с функцией распределения $e^{-e^{-x}}$ (см. (18)); при этом $\mathsf{E}(\nu_j'(N))^r \to (-1)^r\Gamma^{(r)}(1)$ (см. также [32]). Асимптотическая независимость с.в. $\nu_m(N, 1)$ и $\nu_m(N, N)$ при любом $m = o(N)$ установлена в [13].

(3) Наконец, для области средних значений k в работе [9] установлен следующий результат: при $N \to \infty$ нормированный процесс

$$\nu^*_{m,N}(\lambda) = N^{-1/2}(\nu_m(N,[\lambda N]) - Nt_m(\lambda)),\ \lambda \in [\delta, 1-\delta],\ \delta > 0,$$

сходится (в смысле сходимости конечномерных распределений) к центрированному гауссовскому процессу $\nu_m(\lambda)$ с непрерывными траекториями и ковариацией

$$V_m(\lambda_1,\lambda_2) =$$
$$= t'_m(\lambda_1)t'_m(\lambda_2)\left(\lambda_1\bar{\lambda}_2 - \frac{t_m(\lambda_1)}{t'_m(\lambda_1)}\left(\bar{\lambda}_2 - \sum_{j=0}^{m-2}\pi_j(t_m(\lambda_1))\Pi_{m-2-j}(t_m(\lambda_2) - t_m(\lambda_1))\right)\right),$$

$\lambda_1 \leqslant \lambda_2$ (при $m = 1$ пустая сумма заменяется нулем, $t_m(\lambda)$ указано в (22)). Для частного случая $m = 1$ функция $t_1(\lambda) = -\ln\bar{\lambda}$ и $V_1(\lambda_1,\lambda_2) = V_1^2(\lambda_1)$, $\lambda_1 \leqslant \lambda_2$, где $V_1^2(\lambda) = \lambda/\bar{\lambda} + \ln\bar{\lambda}$. Отсюда нормированная ковариационная функция $\upsilon_1(\lambda_1,\lambda_2) = V_1(\lambda_1)/V_1(\lambda_2)$, т.е. предельный процесс $\nu_1(\lambda)$ является марковским.

б) *Растущий уровень.* Пусть теперь $m = m(N) \to \infty$ при $N \to \infty$. Асимптотическое поведение времени ожидания $\nu_m(N, k)$ в этом случае рассматривалось в [10, 13]. Приведем основные результаты этих работ.

Что касается зоны средних значений параметра k, то в [10] доказано, что соотношение (22) продолжает оставаться справедливым и при $m \to \infty$, $m = O(\ln N)$. Для крайних же значений k картина оказывается значительно сложнее: предельное соотношение типа (18) продолжает быть справедливым при $m \to \infty$ лишь если $m = o(\sqrt{\ln N / \ln\ln N})$, а соотношение (17) при $m \to \infty$ уже не имеет места.

Приведем точные формулировки соответствующих утверждений.

Т е о р е м а 5. *Если $N, k \to \infty$ так, что $s = N - k \leqslant c < \infty$, а $m = o(\ln N)$, то*

$$\mathsf{P}\left(\frac{\nu_{m+1}(N,k)}{N} - \ln N - m\ln\ln N + \ln m! - mu\left(\frac{m\ln\ln N - \ln m!}{\ln N}, \frac{m}{\ln N}\right) \leqslant x\right) \to$$
$$\to \Pi_s(e^{-x}),\ -\infty < x < \infty, \tag{23}$$

где функция $u(t, a)$ есть стремящееся к нулю при $t \to 0$ решение уравнения

$$e^u - au - 1 = t,\ a \geqslant 0,\ a \neq 1.$$

При малых t для $u(t, a)$ можно выписать представление в виде степенного ряда:

$$u(t,a) = \sum_{n=1}^{\infty} d_n(a)t^n.$$

Если $m = o\left(\left(\frac{\ln N}{\ln\ln N}\right)^{\frac{i+1}{i+2}}\right)$ при некотором целом $i \geqslant 0$, то в (23) вместо ряда для $u(t, a)$ достаточно оставить i его первых членов. В частности, при $m = o(\sqrt{\ln N / \ln\ln N})$ центрирующую величину $mu(t, a)$ в (23) можно исключить и мы приходим к соотношению типа (18).

Теорема 6. *Пусть $k \geqslant 1$ фиксировано, $m \to \infty$, $m/\ln N \to 0$ при $N \to \infty$. Тогда*

$$\mathsf{P}\left(m \sqrt[m]{\frac{N}{m!}} \left(\frac{\nu_m(N,k)}{N} - m\upsilon\left(\frac{1}{m} \sqrt[m]{\frac{m!}{N}} \right) \right) \leqslant x \right) \to \overline{\Pi}_{k-1}(e^x), \quad -\infty < x < \infty, \tag{24}$$

где $\upsilon(t)$ есть стремящееся к нулю при $t \to 0$ решение уравнения $\upsilon e^{-\upsilon} = t$.

При малых t для функции $\upsilon(t)$ имеет место представление

$$\upsilon(t) = \sum_{n=1}^{\infty} \frac{n^{n-1}}{n!} t^n.$$

Если при некотором целом $i \geqslant 1$ величина $i(\ln N)/m - \ln m \to \infty$, то в (24) вместо ряда $\upsilon(t)$ достаточно оставить i его первых членов. В частности, если $m \leqslant \ln N/\ln \ln N$, то центрирование в (24) упрощается:

$$\mathsf{P}\left(m^m \sqrt{\frac{N}{m!}} \nu_m(N,k) / N - m \leqslant x \right) \to \overline{\Pi}_{k-1}(e^x).$$

Чтобы описать асимптотическое поведение с.в. $\nu_m(N, k)$ в экстремальных зонах параметра k при m, растущем как $\ln N$, введем некоторые вспомогательные величины. Рассмотрим уравнение

$$a(x - \ln(1+x)) - 1 = 0, \quad a > 0.$$

Оно имеет два вещественных корня $\rho \in (0, \infty)$ и $\delta \in (-1, 0)$. Определим, далее, при $x = \rho, \delta$ функцию $\tau_x(t)$ как стремящееся к нулю при $t \to 0$ решение уравнения

$$\tau - \ln(1 + \tau / (1+x)) = t.$$

Для этой функции можно выписать при малых $|t|$ представление в виде степенного ряда:

$$\tau_x(t) = \frac{1+x}{x} t + \sum_{r=2}^{\infty} g_r(x) t^r.$$

Теорема 7. *Пусть N, $m \to \infty$ так, что $m/\ln N \to a > 0$. Тогда при $s = N - k \leqslant c < \infty$*

$$\mathsf{P}\left(\frac{\nu_{m+1}(N,k)}{N} - m\left(1 + \rho + \tau_\rho\left(\frac{a \ln N - m - a \ln\sqrt{2\pi m}}{am} \right) \right) \leqslant x \right) \to \Pi_s\left(\frac{1+\rho}{\rho} e^{-x\frac{\rho}{1+\rho}} \right),$$

$$-\infty < x < \infty;$$

если же $k \leqslant c < \infty$, то

$$\mathsf{P}\left(\frac{\nu_m(N,k)}{N} - m\left(1 + \delta + \tau_\delta\left(\frac{a \ln N - m - a \ln\sqrt{2\pi m}}{am} \right) \right) \leqslant x \right) \to \overline{\Pi}_{k-1}\left(-\frac{1}{\delta} e^{-x\frac{\delta}{1+\delta}} \right),$$

$$-\infty < x < \infty.$$

Если при некотором целом $i \geqslant 1$ выполняется соотношение $m - a\ln N = o((\ln N)^{i/(i+1)})$, то в этих соотношениях вместо рядов $\tau_x(t)$ достаточно оставить i их первых членов.

Наконец, чтобы сформулировать соответствующие утверждения для случая $m/\ln N \to \infty$, введем две функции $\tau^{(1)}(t) \geqslant 0$ и $\tau^{(2)}(t) \leqslant 0$ как решения уравнения

$$\tau - \ln(1+\tau) = t, \quad t \geqslant 0.$$

При малых t для них существуют представления в виде степенных рядов:

$$\tau^{(1)}(t) = \sqrt{2t} + \sum_{r=2}^{\infty} d_r t^{r/2}, \quad \tau^{(2)}(t) = -\sqrt{2t} + \sum_{r=2}^{\infty} (-1)^r d_r t^{r/2}.$$

Теорема 8. *Пусть $N, m \to \infty$ так, что $m/\ln N \to \infty$, $m = o(N)$. Тогда при $s = N - k \leqslant c < \infty$*

$$\mathsf{P}\left(\sqrt{\frac{2\ln N}{m}}\left(\frac{\nu_{m+1}(N,k)}{N} - m - m\tau^{(1)}\left(\frac{\ln N - \frac{1}{2}\ln\ln N}{m}\right)\right) + \ln\sqrt{4\pi} \leqslant x\right) \to \Pi_s(e^{-x}),$$

$$-\infty < x < \infty;$$

если же $k \leqslant c < \infty$, то

$$\mathsf{P}\left(\sqrt{\frac{2\ln N}{m}}\left(\frac{\nu_{m}(N,k)}{N} - m - m\tau^{(2)}\left(\frac{\ln N - \frac{1}{2}\ln\ln N}{m}\right)\right) - \ln\sqrt{4\pi} \leqslant x\right) \to \overline{\Pi}_{k-1}(e^{x}),$$

$$-\infty < x < \infty.$$

Если $m / (\ln N)^{(i+2)/i} \to \infty$ при некотором целом $i \geqslant 1$, то в этих соотношениях вместо рядов $\tau^{(j)}(t)$ достаточно оставить i их первых членов.

В работе [13] исследовано также асимптотическое поведение разности $\tau_m(N) = \nu_m(N,N) - \nu_m(N,1)$ при $N \to \infty$, $m = m(N)$: с.в. $\tau_m(N)$ измеряет время, прошедшее после m-кратной реализации какого-либо одного исхода до, по крайней мере, m-кратной реализации всех N исходов. Из соответствующих результатов отметим следующий: $\tau_m(N) / a(m,N) \xrightarrow{\mathsf{P}} 1$, где $a(m, N) = N \ln N$ при $m = o(\ln N)$, $a(m, N) = (\rho - \delta)aN \ln N$ при $m/\ln N \to a > 0$ (ρ, δ определены в теореме 7), наконец, $a(m, N) = 2N(2m \ln N)^{1/2}$ при $m/\ln N \to \infty$, $m = o(N)$.

в) *Несимметрическая схема.* Если вероятности исходов $p_j = a_j / N$, $j = 1,\ldots,N$, при $N \to \infty$ удовлетворяют условию регулярности

$$0 < c_1 \leqslant \min_j a_j \leqslant \max_j a_j \leqslant c_2 < \infty, \tag{25}$$

то многие результаты об асимптотическом поведении времени ожидания $\nu_m(N, k)$ в равновероятной схеме ($a_j = 1$ для всех j) переносятся и на общий случай. Приведем точные формулировки соответствующих утверждений.

Пусть вероятности p_j задаются с помощью непрерывной на [0, 1] плотности распределения $p(y) > 0$:

$$p_j = \int_{(j-1)/N}^{j/N} p(y)dy, \quad j = 1,\ldots,N.$$

Определим для каждого $x \in [0, 1]$ функцию $t_m(x)$ как решение уравнения

$$\int_0^1 \overline{\Pi}_{m-1}(tp(y))\ dy = x \tag{26}$$

(для равновероятной схемы, т.е. при $p(y) \equiv 1$, эта функция совпадает с введенной в (10)), и пусть $\{\nu_m(x), x \in [0, 1]\}$ есть центрированный гауссовский процесс с непрерывными траекториями и ковариацией

$$R_m(x_1,x_2)=t'_m(x_1)t'_m(x_2)\left(1-x_2-\int_0^1 \Pi_{m-1}(t_m(x_1)p(y))\Pi_{m-1}(t_m(x_2)p(y))\ dy-\right.$$

$$-\frac{t_m(x_1)}{t'_m(x_1)}\int_0^1\Big(\Pi_{m-1}(t_m(x_2)p(y))-$$

$$\left.\left.-\sum_{j=0}^{m-2}\pi_j(t_m(x_1)p(y))\Pi_{m-j-2}((t_m(x_2)-t_m(x_1))p(y))\right)p(y)\ dy\right),\quad x_1\leqslant x_2.$$

Теорема 9. *При $N\to\infty$ нормированный процесс*

$$\nu^*_{m,N}(x)=N^{-1/2}(\nu^\cdot_m(N,[xN])-Nt_m(x)),\quad x\in[\delta,1-\delta],\quad \delta>0,$$

сходится (в смысле сходимости конечномерных распределений) к процессу $\nu_m(x)$, где $t_m(x)$ определено в (26).

Данная теорема представляет собой непосредственный перенос результата (3) п.а) на регулярную схему и доказана в работе [9].

Для экстремальных значений k справедливы следующие утверждения [17].

Теорема 10. *Пусть $N,\ k\to\infty$ так, что $s=N-k\geqslant 0$ фиксировано, а вероятности p_j удовлетворяют условию $p_j=(1+g_j/\ln N)/N$, $|g_j|=o(\ln N)$, $j=1,\dots,N$,*

$$g(N)=\frac{1}{N}\sum_{j=1}^{N}e^{-g_j}\leqslant c<\infty.$$

Тогда

$$\mathsf{P}\left(\frac{\nu_m(N,k)}{N}-\ln N-(m-1)\ln\ln N-\ln\frac{g(N)}{(m-1)!}\leqslant x\right)\to\Pi_s(e^{-x}),\quad -\infty<x<\infty.$$

Теорема 11. *Если $N\to\infty$, k фиксировано и выполнено* (25), *то при $m\geqslant 2$*

$$\mathsf{P}\left(\left(\sum_{j=1}^{N}p_j^m\right)^{\frac{1}{m}}\nu_m(N,k)\leqslant x\right)\to\overline{\Pi}_{k-1}\left(\frac{x^m}{m!}\right),\quad x>0,$$

а с.в. $\nu_1(N,k)$ в пределе вырождается в точке k.

Наконец, ситуация, когда с.в. $\nu_1(N,k)$ перестает быть в пределе вырожденной, описывается следующим утверждением.

Теорема 12. *Пусть выполнено* (25) *и при $N,\ k\to\infty$*

$$\frac{k^2}{2}\sum_{j=1}^{N}p_j^2\to\lambda,\quad 0<\lambda<\infty.$$

Тогда $\mathcal{L}(\nu_1(N,k)-k)\to\Pi(\lambda)$.

Этот результат принадлежит Холсту [30], который установил также асимптотическую нормальность с.в. $\nu_1(N,k)$ при $k^2/N\to\infty$, $\overline{\lim}\ k/N<1$.

Если несимметричность схемы связана и с неоднородностью уровней ν_j, то ситуация значительно усложняется и для этого случая известны лишь отдельные результаты [4]. Введем функции $R_{\nu_j}(t)=\mathsf{P}(\xi(t)<\nu_j)$, где, напомним, $\mathcal{L}(\xi(t))=\Pi(t)$, и предположим, что при каждом $t > 0$ существует

$$\lim_{N\to\infty} \prod_{j=1}^{N} R_{\nu_j}(a_j t) = R(t) > 0. \tag{27}$$

Тогда при $N \to \infty$

$$\mathsf{P}(\nu(N,1)/N \leqslant t) \to 1 - R(t), \quad t > 0.$$

Условие (27) означает, что выполнено хотя бы одно из условий $\lim\limits_{j\to\infty} \mathsf{P}(\nu_j > n) = 1$, $n = 1,2,\ldots$, и $\lim\limits_{j\to\infty} a_j = 0$, т.е. с бóльшей вероятностью первыми будут заполняться ячейки с малыми номерами. При этом условии с.в. типа номера первой заполненной ячейки, второй заполненной ячейки и т.д. будут иметь собственные предельные распределения. Например, при $a_1 = \ldots = a_N = 1$

$$\lim_{N\to\infty} \mathsf{P}(S(N) = j,\ \nu(N,1)/N \leqslant t) \to -\int_0^t \frac{R'_{\nu_j}(\tau)}{R_{\nu_j}(\tau)} R(\tau)\ d\tau, \quad j = 1,2,\ldots,$$

где $S(N)$ означает номер первой заполненной ячейки.

Для другой экстремальной зоны значений параметра k может быть установлен следующий результат. Пусть выполнено условие регулярности (25) и дополнительное условие на функции $R_{\nu_j}(t)$: $R_{\nu_j}(t) \leqslant e^{-\delta t}$ при некотором $\delta \in (0, 1)$ и всех $t > 0$, $j = 1, \ldots, N$. Введем функцию

$$Q_N(x) = \sum_{j=1}^{N} R_{\nu_j}(a_j x)$$

и пусть x_N есть корень уравнения $Q_N(x) = 1$ (при больших N для x_N справедливо представление $x_N = c_N \ln N$, где $0 < d_1 \leqslant c_N \leqslant d_2 < \infty$ при некоторых d_1, d_2). Тогда, если для любого фиксированного x существует

$$\lim_{N\to\infty} Q_N(x_N + x) = Q(x),$$

то при N, $k \to \infty$ и фиксированном $s = N - k \geqslant 0$

$$\mathsf{P}(\nu(N,k)/N - x_N \leqslant x) \to \Pi_s(Q(x)), \quad -\infty < x < \infty.$$

Наконец, приведем некоторые результаты для нерегулярных схем. Рассмотрим время ожидания $\nu_m(N, 1)$ до m-кратной реализации какого-нибудь исхода ($m \geqslant 2$), и пусть $\max\limits_j p_j = p_1 = \alpha \to 0$ при $N \to \infty$. Обозначим также

$$|p|_m = \left(\sum_{j=2}^{N} p_j^m\right)^{1/m}, \quad \lambda_m = \frac{\alpha}{|p|_m}, \quad \gamma_m = |p|_{m+1}/|p|_m.$$

Т е о р е м а 13. *Если* $\lambda_m \to \infty$, *то* $\mathcal{L}(\alpha\nu_m(N,1)) \to \mathcal{L}(\xi_m)$, *где с.в.* ξ_m *имеет гамма-плотность* $\pi_{m-1}(x)$, $x > 0$; *если* $\lambda_m \to 0$, *то* $\mathcal{L}(|p|_m \nu_m(N,1)) \to \mathcal{L}(\eta_m)$, *где* $\mathsf{P}(\eta_m \leqslant x) = 1 - e^{-x^m/m!}$, $x > 0$; *если* $\lambda_m \to \delta > 0$, *то* $\mathcal{L}(|p|_m \nu_m(N,1)) \to \mathcal{L}(\zeta_m)$, *где* $\zeta_m = \min(\delta\xi_m, \eta_m)$ *и* ξ_m, η_m *независимы и имеют указанные выше распределения.*

г) *Моменты.* В симметрической схеме из формулы (11) следует представление для произвольного факториального момента с.в. $\nu(N, k)$:

$$\mathsf{E}\nu^{[r]}(N,k) = N^r \int_p^1 t^r(x)\beta(x;k,N-k+1)\ dx, \tag{28}$$

где $a^{[r]} = a(a+1)\dots(a+r-1)$, $r \geqslant 1$. Для случая вырожденного уровня аналог представления (28) при $r = 1,2$ был установлен в [25], а в общем случае – в [39] (см. также [32]). Для случайного уровня и $r = 1,2$ формула (28) получена в [2], где найдена также асимптотика среднего и дисперсии при $N \to \infty$ и каждом из трех случаев: k фиксировано, $k = \lambda N$, $0 < \lambda < 1$, и $N - k$ фиксировано.

Для иллюстрации приведем некоторые результаты для случая вырожденного уровня: $\mathsf{P}(\nu = m) = 1$ при некотором $m \geqslant 1$.

Если $m = 1$, то

$$\mathsf{E}\nu_1(N,k) = N \sum_{n=N-k+1}^{N} \frac{1}{n},$$

$$\mathsf{D}\nu_1(N,k) = N^2 \sum_{n=N-k+1}^{N} \frac{1}{n^2} - \mathsf{E}\nu_1(N,k)$$

и при $N \to \infty$ справедливы следующие асимптотические соотношения [18, 24]:

если $k / \sqrt{N} \leqslant c < \infty$, то

$$\mathsf{E}\nu_1(N,k) = k + \frac{k(k-1)}{2N} + O\left(\frac{k^3}{N^2}\right),$$

$$\mathsf{D}\nu_1(N,k) = \frac{k(k-1)}{2N} + O\left(\frac{k^3}{N^2}\right);$$

если $k / \sqrt{N} \to \infty$, $N - k \to \infty$, то

$$\mathsf{E}\nu_1(N,k) = N \ln\frac{1}{1-\lambda} + \frac{2-\lambda}{2(1-\lambda)} + o(1), \quad \lambda = \frac{k}{N},$$

$$\mathsf{D}\nu_1(N,k) = N\left(\frac{\lambda}{1-\lambda} - \ln\frac{1}{1-\lambda}\right) + O(1);$$

если $s = N - k \leqslant c < \infty$, то

$$\mathsf{E}\nu_1(N,k) = N\left(\ln N + \gamma - \sum_{n=1}^{s} \frac{1}{n}\right) + \frac{1}{2} + O\left(\frac{1}{N}\right),$$

$$\mathsf{D}\nu_1(N,k) = N^2\left(\frac{\pi^2}{6} - \sum_{n=1}^{s} \frac{1}{n^2}\right) - N\ln N - N\left(\gamma - \sum_{n=2}^{s} \frac{1}{n}\right) + O(1),$$

где $\gamma = 0{,}5772\dots$ – константа Эйлера (пустые суммы при $s = 0,1$ заменяются нулем). Для общего случая $m \geqslant 2$ имеют место следующие асимптотики [10]: при $1 \leqslant k \leqslant$ $\leqslant c < \infty$

$$\mathsf{E}\nu_m(N,k) = \frac{N}{(k-1)!}\left(\Gamma(k+1/m)\left(\frac{m!}{N}\right)^{1/m} + \frac{\Gamma(k+2/m)}{m+1}\left(\frac{m!}{N}\right)^{2/m} + O(N^{-3/m})\right),$$

$$\mathsf{D}\nu_m(N,k) = \frac{N^2}{(k-1)!}\left(\left(\Gamma(k+2/m) - \frac{\Gamma^2(k+1/m)}{(k-1)!}\right)\left(\frac{m!}{N}\right)^{2/m} + O(N^{-3/m})\right);$$

при $k = \lambda N$, $0 < \lambda < 1$,

$$\mathsf{E}\nu_m(N,k) = Nt_m(\lambda) - \lambda t'_m(\lambda) + \frac{\lambda(1-\lambda)}{2} t''_m(\lambda) + O\left(\frac{1}{N}\right),$$

$$\mathsf{D}\nu_m(N,k) = N[\lambda(1-\lambda)(t'_m(\lambda))^2 - t_m(\lambda)] + O(1)$$

($t_m(\lambda)$ указано в (22));

при $s = N - k \leqslant c < \infty$

$$\mathsf{E}\nu_m(N,k) = N\left(\ln N + (m-1)\ln\ln N + \gamma - \ln(m-1)! - \sum_{n=1}^{s} \frac{1}{n} + O\left(\frac{\ln\ln N}{\ln N}\right)\right),$$

$$\mathsf{D}\nu_m(N,k) = N^2\left(\frac{\pi^2}{6} - \sum_{n=1}^{s} \frac{1}{n^2} + O\left(\frac{\ln\ln\ln N}{\ln\ln N}\right)\right).$$

История "борьбы" за эти формулы такова. Формула для $\mathsf{E}\nu_m(N, N)$ выводилась в работах [26, 27, 36], при этом константа $\gamma - \ln(m-1)!$ (при $s = 0$ константа $\sum_{n=1}^{s} \frac{1}{n}$ отсутствует) была вычислена впервые в [27], а остаточный член – в [26]. Формула для $\mathsf{D}\nu_m(N, N)$ получена в [26], а формулы для моментов при $k = \lambda N$ – в [25]. Для остальных случаев формулы доказаны в [10].

Общие представления для первых двух моментов с.в. $\nu_m(N, k)$ в неравновероятной схеме получены в [8]. Так, формула для среднего имеет вид

$$\mathsf{E}\nu_m(N,k) = \sum_{r=0}^{k-1} \sum_{1\leqslant j_1<\ldots<j_r\leqslant N} \int_0^\infty \prod_{s=1}^{r} \overline{\Pi}_{m-1}(p_{j_s}t) \prod_j{}' \Pi_{m-1}(p_j t)\ dt$$

(см. обозначения в (12)), которая для случая $k = N$ сводится к виду [26]

$$\mathsf{E}\nu_m(N,N) = \int_0^\infty \left[1 - \prod_{j=1}^{N} \overline{\Pi}_{m-1}(p_j t)\right] dt,$$

а для случая $m = 1$ – к виду (см. (12′)) [38]

$$\mathsf{E}\nu_1(N,k) = \sum_{r=0}^{k-1} (-1)^{k-r-1} C_{N-r-1}^{N-k} \sum_{1\leqslant j_1<\ldots<j_r\leqslant N} \frac{1}{1 - p_{j_1} - \ldots - p_{j_r}} =$$

$$= \sum_{r=N-k+1}^{N} (-1)^{r-N+k-1} C_{r-1}^{N-k} \sum_{1\leqslant j_1<\ldots<j_r\leqslant N} \frac{1}{p_{j_1} + \ldots + p_{j_r}}.$$

Для второго момента справедливо следующее общее представление

$$\mathsf{E}\nu_m(N,k)(\nu_m(N,k)+1) = 2\sum_{r=0}^{k-1} \sum_{1\leqslant j_1<\ldots<j_r\leqslant N} \int_0^\infty t \prod_{s=1}^{r} \overline{\Pi}_{m-1}(p_{j_s}t) \prod_j{}' \Pi_{m-1}(p_j t)\ dt,$$

а в [26] показано, что

$$\mathsf{E}\nu_m(N,N)(\nu_m(N,N)+1) = 2\int_0^\infty t\left[1 - \prod_{j=1}^{N} \overline{\Pi}_{m-1}(p_j t)\right] dt.$$

Некоторые асимптотические результаты содержатся в [26, 37] (см. также следующий раздел).

Для частного случая $k = 1$ при $m \geqslant 2$ можно установить следующие асимптотические представления для среднего и дисперсии. Пусть выполняются предположения теоремы 13.

Тогда, если $\lambda_m \to 0$, то

$$\mathsf{E}\nu_m(N,1) = \frac{(m!)^{1/m}}{|p|_m}\left(\Gamma\left(1+\frac{1}{m}\right) + \frac{(m!)^{1/m}}{m+1}\Gamma\left(1+\frac{2}{m}\right)\gamma_m^{m+1} + O(\lambda_m^2)\right),$$

$$\mathsf{D}\nu_m(N,1) = \frac{(m!)^{2/m}}{|p|_m^2}\left(\Gamma\left(1+\frac{2}{m}\right) - \Gamma^2\left(1+\frac{1}{m}\right)\right)\left(1+O(|p|_m + \gamma_m^{m+1})\right)$$

(этот случай примыкает к регулярному случаю (25));

если $\lambda_m \to \delta > 0$ и $p_2/|p|_m \to 0$, то

$$\mathsf{E}\nu_m(N,1) = \frac{\delta_1}{\alpha}\int_0^\infty e^{-x^m}\Pi_{m-1}(x\delta_1)\ dx\left(1+O(\gamma_m^{m+1})\right),\ \ \delta_1 = \delta(m!)^{1/m},$$

$$\mathsf{D}\nu_m(N,1) = \frac{\delta_1^2}{\alpha^2}\int_0^\infty (2x^2-1)e^{-x^m}\Pi_{m-1}(x\delta_1)\ dx\left(1+O(\gamma_m^{m+1})\right);$$

если $\lambda_m \to \infty$, то

$$\mathsf{E}\nu_m(N,1) = \frac{m}{\alpha}\left(1 - \frac{1}{m}C_{2m}^{m+1}\frac{1}{\lambda_m^m}\left(1+O\left(\frac{1}{\lambda_m}\right)\right)\right),$$

$$\mathsf{D}\nu_m(N,1) = \frac{m}{\alpha^2}\left(1-\alpha+2C_{2m}^{m+1}\frac{m(m+1)+1}{m(m+2)}\ \frac{1}{\lambda_m^m}\left(1+O\left(\alpha+\frac{1}{\lambda_m}\right)\right)\right).$$

5. Статистические выводы. Этот раздел посвящен применению изложенных выше результатов к различным задачам статистического вывода для полиномиальной модели. В основном нас будет интересовать интенсивно обсуждаемая в статистической литературе проблема проверки гипотезы о равновероятности исходов H_0: $p_j = 1/N$, $j = 1,\ldots,N$, при числе исходов $N \to \infty$. Альтернативой при этом является общая гипотеза H_1: $(p_1,\ldots,p_N) \neq (1/N,\ldots,1/N)$. В качестве тестовых статистик мы будем здесь рассматривать симметрические р.ст.

$$L_{Nk}(g) = \sum_{j=1}^{N} g(\eta_j), \tag{29}$$

соответствующие моменту остановки $\nu_m(N, k)$. Поскольку здесь имеется три "степени свободы" – параметры k, m и порождающая функция g, то мы имеем целый класс тестовых статистик и основной задачей является оптимальный выбор статистики из этого класса или каких-то его подклассов. В дальнейшем о критерии согласия для гипотезы H_0, порождаемом р.ст. (29), будем говорить кратко как о критерии g и под $\mathcal{G}_\alpha$ понимать класс всех таких критериев размера α.

Впервые идея использования такого типа критериев была предложена в работе [12], где рассмотрен случай $g(x) \equiv x$ (т.е. тестовой статистикой является время ожидания $\nu_m(N, k)$); в дальнейшем исследования в этом направлении были продолжены в [5, 7, 9]. Установление в последнее время общих предельных теорем для р.ст. $L_{Nk}(g)$ (эти теоремы при нулевой гипотезе H_0 изложены в разделе 3) позволило сделать дальнейший шаг и "поднять" проблему на уровень всего класса $\mathcal{G}_\alpha$.

а) Уже при исследовании случая $g(x) \equiv x$ в [12] выяснилось, что наиболее чувствительными (к отклонениям от H_0) в рассматриваемом классе являются критерии, соответствующие "средним" значениям параметра k: $k = \lambda N$, $0 < \lambda < 1$. При этом. критерий, порождаемый тестовой статистикой $\nu_m(N, \lambda N)$, способен отличать от H_0 лишь близкие (при $N \to \infty$) альтернативы вида

$$H_{1N}^{(1)}:\ p_j = \frac{1}{N}\left(1 + \frac{b_j(N)}{N^{1/4}}\right),\quad j = 1,\dots,N, \tag{30}$$

где $\max\limits_j |b_j(N)| \le c < \infty$ и $\frac{1}{N}\sum\limits_{j=1}^{N} b_j^2(N) \to b^2 > 0$: при заданном размере α мощность $W_N(\cdot)$ такого критерия удовлетворяет асимптотическому соотношению

$$W_N(H_{1N}^{(1)}) \to \gamma,\quad \alpha < \gamma < 1. \tag{31}$$

Для статистики $\nu_m(N, \lambda N)$ предельная мощность γ в (31) равна $\Phi(b^2\rho_m(\lambda)/2 - u_\alpha)$, где Φ – стандартная нормальная функция распределения, $\Phi(-u_\alpha) = \alpha$ и

$$\rho_m(\lambda) = t_m(\lambda)|t_m(\lambda) - m + 1|/(\lambda\bar{\lambda}(t_m'(\lambda))^2 - t_m(\lambda))^{1/2}$$

(обозначения см. в (22)). Отметим также, что для простейшего случая $m = 1$ функция

$$\rho_1(\lambda) = \left(\ln\frac{1}{\bar{\lambda}}\right)^2 \Big/ \left(\frac{\lambda}{\bar{\lambda}} - \ln\frac{1}{\bar{\lambda}}\right)^{1/2}$$

имеет один локальный максимум в точке $\lambda^* = 0{,}972\dots$, $\rho_1(\lambda^*) = 1{,}145\dots$, а соответствующий критерий задается критической областью

$$S_\alpha^* = \left\{\nu_1(N, \lambda^* N) > N\ln\frac{1}{\bar{\lambda}^*} + u_\alpha\sqrt{N\left(\frac{\lambda^*}{\bar{\lambda}^*} - \ln\frac{1}{\bar{\lambda}^*}\right)}\right\}.$$

Аналогичное исследование для произвольного критерия $g \in \mathcal{G}_\alpha$ проведено И.С. Мальцевой (неопубликованный результат). Ею установлено, что в обозначениях теоремы 3

$$\begin{aligned} &\mathsf{E}(L_{N,\lambda N}(g)|H_{1N}^{(1)}) = Na_{\lambda m}(g) + \sqrt{N}\ \frac{b^2}{2} Eg(\xi_{\lambda m})\Psi_m(\xi_{\lambda m})(1 + o(1)),\\ &\mathsf{D}\ (L_{N,\lambda N}(g)|H_{1N}^{(1)}) = N\sigma_{\lambda m}^2(g)(1 + o(1)), \end{aligned} \tag{32}$$

где индекс m отражает вырожденность уровня ν в точке m, $\xi_{\lambda m} = \xi(t_m(\lambda))$ и

$$\Psi_m(\xi_{\lambda m}) = \xi_{\lambda m}(\xi_{\lambda m} - t_m(\lambda) - 1) + (m-1)(t_m(\lambda) - \xi_{\lambda m}); \tag{33}$$

что же касается распределения р.ст. $L_{N,\lambda N}(g)$ при альтернативах $H_{1N}^{(1)}$ вида (30), то оно продолжает оставаться асимптотически нормальным с указанными в (32) параметрами.

Эти результаты позволяют построить соответствующий g-критерий согласия для гипотезы H_0 и вычислить его предельную мощность γ в (31). Именно, если $\mathsf{E}g(\xi_{\lambda m})\Psi_m(\xi_{\lambda m}) > 0$, то соответствующая критическая область $S_\alpha(g)$ имеет вид

$$S_\alpha(g) = \{L_{N,\lambda N}(g) > Na_{\lambda m}(g) + \sqrt{N}\sigma_{\lambda m}(g)u_\alpha\}, \tag{34}$$

а мощность

$$W_N(H_{1N}^{(1)}) \to \Phi(b^2\rho_m(\lambda;g)/2-u_\alpha), \tag{35}$$

где

$$\rho_m(\lambda;g)=\sigma_{\lambda m}^{-1}(g)\mathrm{E}g(\xi_{\lambda m})\Psi_m(\xi_{\lambda m})$$

есть показатель эффективности критерия g.

Вопрос об оптимальном критерии решается следующим утверждением.

Теорема 14. *Асимптотически оптимальным в классе* $\mathcal{G}_\alpha$ *при любых фиксированных* $m \geqslant 1$ *и* $\lambda \in (0,1)$ *является критерий* $g^* = \Psi_m$ (*см.* (33)); *критическая область* (34) *принимает в этом случае вид*

$$S_\alpha(g^*)=\left\{\sum_{j=1}^{N}\eta_j^2-(t_m(\lambda)+m)\nu_m(N,\lambda N) > -N(m-1)t_m(\lambda)+\sqrt{N}\sigma_{\lambda m}(\Psi_m)u_\alpha\right\},$$

а показатель эффективности в (35) *равен*

$$\rho_m(\lambda;\Psi_m)=t_m(\lambda)((t_m(\lambda)-m+1)^2+2t_m(\lambda))/\sigma_{\lambda m}(\Psi_m),$$

где

$$\sigma_{\lambda m}^2(\Psi_m)=(t_m(\lambda)-m+1)^2(\lambda\bar{\lambda}(t_m'(\lambda))^2+t_m(\lambda))+2t_m^2(\lambda).$$

В частности,

$$\rho_1(\lambda;\Psi_1)=\left(2+\ln\frac{1}{\lambda}\right)\ln\frac{1}{\lambda}\Big/\left(\ln\frac{1}{\lambda}+\frac{\lambda}{\bar{\lambda}}+2\right)^{1/2}$$

и $\max\limits_{\lambda}\rho_1(\lambda;\Psi_1)=\rho_1(\lambda_1^*;\Psi_1)=3{,}14\ldots,\ \lambda_1^*=0{,}97\ldots$

В следующей таблице приведены значения $\lambda_m^* = \arg\max\rho_m(\lambda;\Psi_m)$ и $\rho_m^*=\rho_m(\lambda_m^*;\Psi_m)$ для некоторых m.

m	1	2	3	4	5	6	7	8	9	10
λ_m^*	0,973	0,960	0,952	0,945	0,940	0,935	0,931	0,927	0,924	0,921
ρ_m^*	3,141	4,997	6,698	8,325	9,906	11,456	12,984	14,494	15,989	17,473

Отметим также, что при $m \geqslant 2$ расчет параметров g^*-критерия значительно упрощается при выборе $\lambda=\lambda_m$, где λ_m – корень уравнения $t_m(\lambda) = m - 1$, т.е. $\lambda_m=\overline{\Pi}_{m-1}(m-1)$: в этом случае

$$S_\alpha(g^*)=\left\{\sum_{j=1}^{N}\eta_j^2-(2m-1)\nu_m(N,\lambda_m N) > -N(m-1)^2+\sqrt{2N}(m-1)u_\alpha\right\},$$

а $\rho_m=\rho_m(\lambda_m;\Psi_m)=\sqrt{2}(m-1)$. Численные значения таких λ_m и ρ_m приведены в таблице

m	2	3	4	5	6	7	8	9	10
λ_m	0,264	0,323	0,353	0,371	0,384	0,394	0,401	0,407	0,413
ρ_m	1,414	2,828	4,243	5,657	7,071	8,485	9,899	11,314	12,728

В работе [9] построен многошаговый критерий последовательного типа, основанный на многомерной статистике $(\nu_m(N,\lambda_1 N),\ldots,\nu_m(N,\lambda_r N))'$, $\lambda_1<\ldots<\lambda_r$, критическая область которого имеет вид $\{\nu_m(N,\lambda_j N) > h_j(N),\ j=1,\ldots,r\}$. Основой для расчета этого критерия является теорема 9, и он также отличает от H_0 лишь альтернативы вида (30).

Наконец, случай проверки неравновероятной нулевой гипотезы рассмотрен в [7], где показано, что критерий, порождаемый тестовой статистикой $\nu_m(N, \lambda N)$, $0 < \lambda < 1$, позволяет различать гипотезы, сближающиеся со скоростью $N^{-1/2}$. Однако такой критерий перестает быть критерием согласия: некоторый подкласс таких близких альтернатив он не отличает в пределе от нулевой гипотезы.

б) Рассмотрим теперь при $N \to \infty$ значения k вида $k = N - s$, где $s \geqslant 0$ фиксировано, и пусть близкие к H_0 альтернативы задаются в виде

$$H_{1N}^{(2)}:\ p_j = \frac{1}{N}\left(1+\frac{g_j}{\ln N}\right),\ j=1,\ldots,N, \tag{36}$$

где $\max\limits_j |g_j| = o(\ln N)$ и

$$\frac{1}{N}\sum_{j=1}^{N} e^{-g_j} \to g,\ 1<g<\infty.$$

В работе [5] построен критерий согласия для H_0, основанный на тестовой статистике $\nu_m(N, N-s)$. Этот критерий способен отличать от H_0 альтернативы вида (36); он задается критической областью

$$S_{\alpha,s} = \{N^{-1}\nu_m(N,N-s) > \ln N + (m-1)\ln\ln N - \ln((m-1)!\,t_{s+1}(\alpha))\},$$

и его мощность удовлетворяет асимптотическому соотношению

$$W_N(H_{1N}^{(2)}) \to \overline{\Pi}_s(g t_{s+1}(\alpha)) \tag{37}$$

(обозначения см. в (10)); при этом предельная мощность возрастает с ростом s.

Подчеркнем, что этот критерий оказывается менее чувствительным к "равномерным" отклонениям от H_0 (сравни (30) и (36)), однако в некоторых ситуациях такого типа критерий может оказаться полезным. Так, его естественно использовать против альтернатив, при которых некоторые вероятности p_j равны нулю. Например, критерий, порождаемый статистикой $\nu_1(N, N)$, с вероятностью 1 улавливает альтернативу, при которой хотя бы один исход имеет нулевую вероятность реализации. Отметим также, что предельная мощность в (37) не зависит от параметра m, поэтому при практическом использовании этого критерия следует выбирать $m = 1$.

Теоретической базой для расчета изложенного критерия служит теорема 10.

в) Опишем, наконец, свойства критериев, порождаемых статистиками $\nu_m(N, k)$, $m \geqslant 2$, при ограниченных k (при $m = 1$ эти с.в. асимптотически вырождаются, см. теорему 11). В [5] показано, что соответствующий критерий нечувствителен к малым отклонениям от H_0 в подклассе регулярных альтернатив (25). Однако по большому числу независимых реализаций с.в. $\nu_m(N, k)$ уже можно построить достаточно чувствительный критерий. Опишем конструкцию наилучшего в рассматриваемом подклассе критерия, который соответствует значению $m = 2$.

Будем рассматривать в исходной последовательности полиномиальных испытаний момент остановки $\nu_2(N, k)$ как момент реализации соответствующего рекуррентного события $\mathcal{E}$, и пусть ξ_n есть число реализаций событий $\mathcal{E}$ в n испытаниях.

В качестве тестовой статистики используем с.в. ξ_n. Тогда, если $n/\sqrt{N} \to \infty$, $n = o(N^{3/2})$ при $N \to \infty$, то критерий согласия размера α для гипотезы H_0 задается критической областью

$$S_{\alpha,k}=\left\{\xi_n>c_1(k)\frac{n}{\sqrt{2N}}+c_2(k)u_\alpha\left(\frac{n}{\sqrt{2N}}\right)^{1/2}\right\},$$

где $c_1(k)=\dfrac{(k-1)!}{\Gamma\left(k+\frac{1}{2}\right)}$, $c_2(k)=\dfrac{(k-1)!}{\Gamma^{3/2}\left(k+\frac{1}{2}\right)}\left(k!-\dfrac{\Gamma^2\left(k+\frac{1}{2}\right)}{(k-1)!}\right)^{1/2}$. Этот критерий способен отличать от H_0 близкие альтернативы вида

$$H_{1N}^{(3)}:\ p_j=\frac{1}{N}\left(1+d_j(N)\left(\frac{n}{\sqrt{N}}\right)^{-1/4}\right),\quad j=1,\dots,N,$$

где $\max\limits_j|d_j(N)|\leq c<\infty$, $\dfrac{1}{N}\sum\limits_{j=1}^{N}d_j^2(N)\to d^2>0$, и его мощность удовлетворяет асимптотическому соотношению

$$W_N(H_{1N}^{(3)})\to\Phi(2^{-5/4}d^2\psi(k)-u_\alpha),$$

где $\psi^2(k)=\dfrac{(k-1)!\,\Gamma\left(k+\frac{1}{2}\right)}{(k-1)!\,k!-\Gamma^2\left(k+\frac{1}{2}\right)}$ – возрастающая по k функция.

г) Рассмотрим еще ситуацию, когда гипотеза H_0 является не обязательно регулярной. Именно, пусть H_0: $p_1,\dots,p_N$ таковы, что $\max\limits_j p_j/\|p\|\to 0$ при $N\to\infty$, где $\|p\|=\left(\sum\limits_{j=1}^{N}p_j^2\right)^{1/2}$. Рассмотрим критерий проверки H_0, задаваемый критической областью

$$S_\alpha=\{\nu_2(N,1)>\|p\|^{-1}(-2\ln\alpha)^{1/2}\},\quad \alpha\in(0,1).$$

Основываясь на теореме 13, можно показать, что асимптотически этот критерий будет иметь размер α, он будет несмещенным против альтернатив H': $p_1',\dots,p_N'$, $\max\limits_j p_j'/\|p'\|\to 0$, $\|p'\|<\|p\|$, и его мощность $W_N(H')\sim\alpha^{\|p'\|^2/\|p\|^2}$.

Более того, если мы имеем n независимых наблюдений $X_1(N),\dots,X_n(N)$ над с.в. $\nu_2(N,1)$, то критическая область

$$S_{\alpha,n}=\left\{\sum_{i=1}^{n}X_i(N)>\|p\|^{-1}\chi^2_{1-\alpha,2n}\right\}$$

задает критерий асимптотического размера α и мощности

$$W_N(H')\sim K_{2n}\left(\frac{\|p'\|^2}{\|p\|^2}\chi^2_{1-\alpha,2n}\right),$$

где $K_{2n}(x)$ есть функция распределения закона $\chi^2(2n)$, а $\chi^2_{\gamma,2n}$ – его γ-квантиль.

д) В этом пункте мы опишем некоторые статистические результаты для схемы с фиксированным числом исходов N, полученные в работах [19, 20,40]. Именно, мы рассматриваем схему с одним выделенным исходом (пусть это будет исход с номером 1 – см. п. д) раздела 2), в которой испытания проводятся до момента m-кратной реализации этого исхода. Наблюдаемой статистикой является здесь вектор $\bar{\eta}=(\eta_2,\dots,\eta_N)$, где η_j есть число реализаций j-го исхода в момент остановки; его распределение дано в (16), откуда для функции правдоподобия данных следует представление

$$L(\bar{k},\bar{p})\equiv \mathsf{P}_{\bar{p}}(\bar{\eta}=\bar{k})=\frac{(k+m-1)!}{(m-1)!k_2!\dots k_N!}p_1^m\prod_{j=2}^{N}p_j^{k_j},$$

$\bar{p}=(p_2,\dots,p_N)$, $p_1=1-p$, $\bar{k}=(k_2,\dots,k_N)$, $k_j=0,1,\dots$, $j=2,\dots,N$ (здесь и далее в этом пункте для любого вектора $\bar{a}=(a_2,\dots,a_N)$ используется обозначение $a = a_2 + \dots + a_N$). Этот вектор является полной достаточной статистикой для параметра $\bar{p}$, параметрическое множество которого есть

$$\mathcal{P}=\{\bar{p}\colon\ 0\leqslant p_j\leqslant 1,\ j=2,\dots,N,\ p_2+\dots+p_N\leqslant 1\},$$

и мы кратко обсудим задачи оценивания и проверки гипотез для параметра $\bar{p}$ по информации, доставляемой статистикой $\bar{\eta}$.

Рассмотрим прежде всего вопрос о несмещенном оценивании. В данном случае оцениваемыми являются все аналитические функции $\tau(\bar{p})$, $\bar{p}\in\mathcal{P}$. Для примера укажем вид несмещенных оценок с минимальной дисперсией (н.о.м.д.) для некоторых таких функций. Пусть

$$\tau_{s,\bar{l}}(\bar{p})=p_1^s\prod_{j=2}^{N}p_j^{l_j}$$

при произвольных целых $0\leqslant s\leqslant m$, $l_j\geqslant 0$, $j=2,\dots,N$, тогда н.о.м.д. для нее есть

$$\tau^*_{s,\bar{l}}=\frac{(m-1)_s}{(\eta+m-1)_{s+l}}\prod_{j=2}^{N}(\eta_j)_{l_j},$$

где $(a)_k=a(a-1)\dots(a-k+1)$, $(a)_0=1$. В частности, н.о.м.д. произвольной степени p_j^l, $j\geqslant 2$, есть $(\eta_j)_l/(\eta+m-1)_l$. Н.о.м.д. для функции

$$\tau_{\bar{s}}(\bar{p})=p_1^{-s}\prod_{j=2}^{N}p_j^{s_j}$$

имеет вид

$$\tau^*_{\bar{s}}=\frac{1}{(s+m-1)_s}\prod_{j=2}^{N}(\eta_j)_{s_j}.$$

В частности, если $\tau_j(\bar{p})=p_j/p_1$, $j\geqslant 2$, то $\tau_j^*=\eta_j/m$, а $(1/p_1)^*=1+\eta/m$.

Для сравнения укажем вид оценок максимального правдоподобия (о.м.п.) параметра $\bar{p}$: $\hat{p}=(\hat{p}_2,\dots,\hat{p}_N)=\bar{\eta}/(\eta+m)$.

Удобно провести перепараметризацию модели, введя новые параметры $\theta_j=p_j/p_1$, $j=2,\dots,N$. Тогда о.м.п. $\hat{\theta}=(\hat{\theta}_2,\dots,\hat{\theta}_N)=\bar{\eta}/m$ и, следовательно, о.м.п.

параметров θ_j совпадают с их н.о.м.д. τ_j^*. В [19] доказано, что при $m \to \infty$

$$\mathcal{L}_{\overline{\theta}}(\sqrt{m}(\hat{\theta}-\overline{\theta})) \to \mathcal{N}(\overline{0}, \Sigma), \tag{38}$$

где $\Sigma = \Sigma(\overline{\theta}) = \|\theta_i(\theta_j + \delta_{ij})\|_2^N$, δ_{ij} – символ Кронекера.

Этот результат позволяет строить асимптотически нормальные о.м.п. $\hat{\tau} = \tau(\hat{\theta})$ для гладких функций $\tau(\overline{\theta})$, а также рассчитывать соответствующие асимптотические доверительные интервалы и области как для самого параметра $\overline{\theta}$ и отдельных его координат, так и для функций $\tau(\overline{\theta})$. Так, асимптотический γ-доверительный интервал для θ_j имеет вид

$$\left(\frac{\eta_j}{m} \mp c_\gamma \sqrt{\eta_j(\eta_j + m)/m^3}\right), \quad \Phi(c_\gamma) = \frac{1+\gamma}{2},$$

а γ-доверительное множество для вектора $\overline{\theta}$ имеет вид

$$\left\{\overline{\theta}:\ \sum_{j=2}^{N} (\hat{\theta}_j - \theta_j)^2 / \hat{\theta}_j - (\hat{\theta}-\theta)^2/(1+\hat{\theta}) \leq \frac{1}{m}\chi^2_{\gamma,N-1}\right\}.$$

В [20] построены также байесовские оценки для $\overline{p}$ и $\overline{\theta}$, когда априорным распределением параметров является распределение Дирихле, а функция потерь выбирается в виде евклидова расстояния между оценкой и параметром.

Рассмотрим, наконец, задачу проверки простой гипотезы H_0: $\overline{p} = \overline{p}^0$, где $\overline{p}^0$ – заданная внутренняя точка множества $\mathcal{P}$. Из (38) следует, что при $m \to \infty$ статистика

$$X_m^2 = p_1^0 \sum_{j=2}^{N} \frac{(\eta_j - mp_j^0/p_1^0)^2}{mp_j^0} - \frac{p_1^0}{m}\left(\eta - m\frac{p^0}{p_1^0}\right)^2$$

имеет при гипотезе H_0 предельное распределение $\chi^2(N-1)$. Это позволяет предложить следующий критерий типа хи-квадрат для гипотезы H_0: $\{X_m^2 > \chi^2_{1-\alpha,N-1}\}$. Этот критерий имеет асимптотически размер α, состоятелен и его мощность при близких альтернативах H_{1m}: $\overline{p} = \overline{p}(m) = \overline{p}^0 + \overline{\beta}/\sqrt{m}$, где $\overline{\beta} = (\beta_2, \ldots, \beta_N)$ – ненулевой фиксированный вектор, задающий отклонение от гипотезы H_0, удовлетворяет асимптотическому соотношению

$$W_m(H_{1m}) \to 1 - K_{N-1}(\chi^2_{1-\alpha,N-1}; \lambda^2),$$

где

$$\lambda^2 = \frac{1}{p_1^0}\sum_{j=1}^{N} \beta_j^2 / p_j^0,$$

$\beta_1 = -\beta$ и $K_n(x;\ \lambda^2)$ есть функция нецентрального распределения хи-квадрат с n степенями свободы и параметром нецентральности λ^2.

е) Коснемся кратко еще одного круга статистических задач, связанных с проблемой упорядочения исходов полиномиальной схемы в соответствии с величинами вероятностей их реализации. В простейшем своем варианте задача состоит в разработке такой процедуры, которая с достаточно высокой вероятностью правильного решения позволяла бы выделить наиболее вероятный исход. В общем случае формулируется задача выделения такого подмножества исходов, чтобы оно содержало заданное число s из k наиболее вероятных исходов. Полный обзор работ по данной тематике имеется в [29], здесь же мы рассмотрим в качестве иллюстрации одну из этого круга задач, решение которой связано с временем ожидания.

Предположим, что вероятности исходов удовлетворяют условию

$$p_1 \geqslant p_2 \geqslant \ldots \geqslant p_k > p_{k+1} \geqslant p_{k+2} \geqslant \ldots \geqslant p_N, \quad k \geqslant 1,$$

и требуется выделить группу из k наиболее вероятных исходов (при этом для простоты будем считать, что сами вероятности p_j известны, но неизвестно, какому именно исходу соответствует каждая вероятность). Одним из вариантов решения подобной задачи является использование процедуры последовательного типа с некоторым правилом остановки. В рассматриваемом случае такой процедурой может быть следующая: испытания проводятся до тех пор, пока какие-либо k исходов реализуются впервые не менее $m \geqslant 2$ раз каждый, – соответствующие исходы и принимаются за искомые. Обозначим эту процедуру $R_m(N, k)$. Основными характеристиками процедуры являются вероятность правильного решения $\mathsf{P}(R_m(N, k))$ и число испытаний до момента остановки – время ожидания $\nu_m(N, k)$, детально рассмотренное выше. Процедура $R_m(N, k)$ подробно проанализирована в [21]. Так, для вероятности правильного решения имеет место представление

$$\mathsf{P}(R_m(N,k)) = \sum_{j=1}^{k} p_j \int_0^{\infty} \pi_{m-1}(p_j t) \prod_{\substack{i=1 \\ i \neq j}}^{k} \overline{\Pi}_{m-1}(p_j t) \prod_{s=k+1}^{N} \Pi_{m-1}(p_s t)\ dt,$$

которое при $k = 1$ (для задачи выделения наиболее вероятного исхода) принимает вид

$$\mathsf{P}(R_m(N,1)) = \int_0^{\infty} \pi_{m-1}(t) \prod_{s=2}^{N} \Pi_{m-1}(tp_s / p_1)\ dt.$$

Если $N \to \infty$, $k \geqslant 1$ и $m \geqslant 2$ фиксированы, $p_1 = \ldots = p_k = \alpha \to 0$ и

$$\lambda_{m,k} = \lambda_{m,k}(N) = \alpha \Big/ \left(\sum_{j=k+1}^{N} p_j^m \right)^{1/m} \to \infty, \tag{39}$$

то справедливо асимптотическое разложение (являющееся обоснованием рассматриваемой процедуры)

$$\mathsf{P}(R_m(N,k)) = 1 - A_{m,k} \lambda_{m,k}^{-m} + O(\lambda_{m,k}^{-m-1}),$$

где константа

$$A_{m,k} = \frac{k}{m!} \int_0^{\infty} t^m \pi_{m-1}(t) (\overline{\Pi}_{m-1}(t))^{k-1}\ dt.$$

В частности, при $k = 1$ справедливо равенство $A_{m,1} = C_{2m-1}^{m}$. Асимптотика распределения и моментов с.в. $\nu_m(N, 1)$ в условиях (39) указаны в теореме 13 и в п. г) раздела 4.

6. Разные задачи. а) Отметим интересный результат, установленный в [35] для времени ожидания $\nu_m(N, k)$ в равновероятной схеме: при $N \to \infty$ равномерно по x, $-\infty < x < \infty$,

$$\max_{1 \leqslant k \leqslant N} |\mathsf{P}(\nu_m(N,k)/N - \ln N - (m-1)\ln\ln N + \ln(m-1)! \leqslant x) - \Pi_{N-k}(e^{-x})| = o(1).$$

б) В работах [11, 13] исследовано асимптотическое поведение экстремальных порядковых статистик $\eta_{(N)}$ и $\eta_{(1)}$ для момента остановки $\nu_m(N, k)$ в равновероятной схеме (их точные распределения указаны в (14′) и (14″)); при этом в [11] рассмотрен случай фиксированного уровня m и различных $k = k(N)$, а в [13] – случай $m =$

$= m(N) \to \infty$ и $k = N$ для с.в. $\eta_{(N)}$, либо $k = 1$ для с.в. $\eta_{(1)}$. Ограничимся качественным описанием соответствующих результатов.

Рассмотрим момент $\nu_m(N, N)$, т.е. момент, когда все исходы реализуются по крайней мере m раз каждый. Тогда, если m фиксировано, а $N \to \infty$, то максимальная частота $\eta_{(N)}$ будет иметь асимптотически порядок $e \ln N + (em - m - e + 1/2)\ln \ln N$, т.е. возрастает медленно и ее рост слабо зависит от m. Асимптотический порядок $\eta_{(N)} = O(\ln N)$ сохраняется и при растущем уровне, пока $m/\ln N \leqslant c < \infty$; при этом предельными распределениями для соответствующим образом центрированной с.в. $\eta_{(N)}$ являются некоторые специальные дискретные распределения. При очень высоком уровне ($m/\ln N \to \infty$) с.в. $\eta_{(N)}$ имеет при соответствующей нормировке некоторое распределение непрерывного типа. В качестве иллюстрации приведем одно наиболее просто формулируемое утверждение такого рода: если $m/(\ln N)^3 \to \infty$, то

$$\mathsf{P}\left(\sqrt{\frac{2\ln N}{m}}\eta_{(N)} - \sqrt{2m\ln N} - 4\ln N + \ln\ln N + \ln 4\pi \leqslant x\right) \to \int_0^\infty \exp\left\{-\theta - \frac{e^{-x}}{\theta}\right\} d\theta.$$

Интересно также ведет себя асимптотически минимальная частота $\eta_{(1)}$ в момент остановки $\nu_m(N, 1)$, т.е. когда какая-нибудь частота достигает уровня m. Пока уровень m удовлетворяет условию $\overline{\lim\limits_{N\to\infty}}\, m/\ln N < e$, с.в. $\eta_{(1)}$ асимптотически вырождается в 0; таким образом, $\eta_{(1)}$ "отрывается от нуля" лишь при $m/\ln N \to e$, и в этом случае ее распределение асимптотически сосредоточено либо в одной, либо в двух соседних целочисленных точках. При более высоких уровнях характер асимптотики $\eta_{(1)}$ похож на сказанное выше о с.в. $\eta_{(N)}$.

Возвращаясь к случаю фиксированного уровня m, отметим еще следующие факты: для значений параметра k, близких к N, $N - k \leqslant c < \infty$, число исходов с максимальной частотой $\eta_{(N)}$ будет асимптотически ограничено, если же $N - k \to \infty$, то их число также будет неограниченно возрастать; для средних значений k, $k = \lambda N$, $0 < \lambda < 1$, максимальная частота $\eta_{(N)}$ имеет порядок $\ln N/\ln \ln N$ и ее распределение асимптотически является либо одноточечным, либо двухточечным; наконец, если $k^{m+1}/N \to 0$, то $\mathsf{P}(\eta_{(N)} = m) \to 1$, т.е. k исходов будут иметь частоту m, а остальные – частоту, меньшую m.

Что касается минимальной частоты $\eta_{(1)}$ в случае фиксированного $m \geqslant 2$, то $\mathsf{P}(\eta_{(1)} = m-1) \to 1$ при $N - k = o(\ln N)$, $\mathsf{P}(\eta_{(1)} = 0) \to 1$ при $(N-k)/(\ln N)^{m-1} \to \infty$, а при остальных значениях k распределение с.в. $\eta_{(1)}$ асимптотически является либо одноточечным, либо двухточечным.

в) Особенности асимптотического поведения с.в. μ_r для момента остановки $\nu_m(N, k)$ в промежуточных зонах (I) $k \to \infty$, $\lambda = k/N \to 0$, и (II) $s = N - k \to \infty$, $\overline{\lambda} = s/N \to 0$, описываются следующими утверждениями [11]:

для зоны (I) определим целое $r_0 \geqslant m + 1$ условием $k^{r_0}/N^{r_0 - m} \to \gamma$, $0 < \gamma < \infty$, тогда при $r > r_0$ с.в. $\mu_r \xrightarrow{\mathsf{P}} 0$,

$$\mathcal{L}(\mu_{r_0}) \to \Pi\left(\frac{1}{r_0!}\left(m!^{r_0}\gamma\right)^{1/m}\right),$$

при $m < r < r_0$

$$\mathcal{L}(\mu_r) \sim \mathcal{N}\left(N\frac{1}{r!}(m!\lambda)^{r/m},\ N\frac{1}{r!}(m!\lambda)^{r/m}\right),$$

при $1 \leqslant r < m$

$$\mathcal{L}(\mu_r) \sim \mathcal{N}\left(N\frac{1}{r!}(m!\lambda)^{r/m},\ N\lambda^{(2r-m)/m}(m!)^{2r/m}/((r-1)!m)\right),$$

а с.в. $N-k-\mu_0$ и μ_1 также, как и с.в. $k-1-\mu_m$ и μ_{m+1} распределены асимптотически одинаково;

для зоны (II) определим целое $r_0 \leqslant m-2$ (при $m \geqslant 2$) условием $s/(\ln N)^{m-1-r_0} \to \delta,\ 0<\delta<\infty$, тогда при $r<r_0$ с.в. $\mu_r \xrightarrow{\mathsf{P}} 0$,

$$\mathcal{L}(\mu_{r_0}) \to \Pi\left(\frac{(m-1)!}{r_0!}\delta\right),$$

при $r_0 < r \leqslant m-2$

$$\mathcal{L}(\mu_r) \sim \mathcal{N}\left(s\frac{(m-1)!}{r!}\left(\ln\frac{1}{\lambda}\right)^{r-m+1},\ s\frac{(m-1)!}{r!}\left(\ln\frac{1}{\lambda}\right)^{r-m+1}\right),$$

при $r \geqslant m$

$$\mathcal{L}(\mu_r) \sim \mathcal{N}\left(s\frac{(m-1)!}{r!}\left(\ln\frac{1}{\lambda}\right)^{r-m+1},\ s\left(\frac{(m-1)!}{r!}\right)^2\left(\ln\frac{1}{\lambda}\right)^{2(r-m+1)}\right),$$

а с.в. $s-\mu_{m-1}$ и μ_{m-2} распределены асимптотически одинаково.

г) Ряд интересных соотношений для средних значений с.в. μ_r, подсчитываемых в момент остановки $\nu_m(N,k)$ (см. (13′)), получен в работе [32]. Обозначим для $j \geqslant 0$

$$\zeta_{kj}^{+} = \sum_{r \geqslant m+j} \mu_r,\quad \zeta_{kj}^{-} = \sum_{r<m-j} \mu_r$$

(ясно, что $\zeta_{k0}^{+}=k,\ \zeta_{k0}^{-}=N-k,\ \zeta_{1j}^{+}=\zeta_{Nj}^{-}=0$ при $j \geqslant 1$), и пусть $S_1,\dots,S_N$ – независимые с.в. с гамма-плотностью $\pi_{m-1}(t)$, а $S_{(1)} \leqslant \dots \leqslant S_{(N)}$ – их вариационный ряд. Тогда справедливы представления

$$\mathsf{E}\zeta_{kj}^{+} = \mathsf{E}(S_{(1)}^{j} + \dots + S_{(k-1)}^{j})/m^{[j]},\quad \mathsf{E}\zeta_{kj}^{-} = \mathsf{E}(S_{(k+1)}^{-j} + \dots + S_{(N)}^{-j})(m-1)_j,$$

где, напомним, $m^{[j]} = m(m+1)\dots(m+j-1)$, $(m-1)_j = (m-1)(m-2)\dots(m-j)$ и $j \geqslant 1$; кроме того

$$\mathsf{E}\nu_m^{[r]}(N,k) = N^r \mathsf{E}S_{(k)}^{r}.$$

Отсюда также следует, что при $k=N$

$$\mathsf{E}\sum_{i=0}^{j-1}\mu_{m+i} = \mathsf{E}S_{(N)}^{j}/m^{[j]}, \tag{40}$$

а при $k=1$

$$\mathsf{E}\sum_{i=0}^{j}\mu_{m-i} = \mathsf{E}S_{(1)}^{-j}(m-1)_j.$$

Отметим любопытный частный случай формулы (40): среднее число исходов, реализовавшихся ровно m раз каждый в тот момент, когда все исходы реализовались по крайней мере m раз каждый, удовлетворяет соотношению

$$\mathsf{E}\mu_m = \mathsf{E}S_{(N)}/m = \mathsf{E}\nu_m(N,N)/Nm.$$

Для $m = 1$ это принимает вид [13]

$$\mathsf{E}\mu_1 = \mathsf{E}\nu_1(N,N)/N = \sum_{n=1}^{N} \frac{1}{n}$$

(см. раздел 4, п. г)).

В [32] установлен также следующий асимптотический результат, согласующийся с утверждением (21): при $N \to \infty$, $k = N - s$, где $s \geqslant 0$ фиксировано,

$$m^{[j+1]}\mathsf{E}\mu_{m+j} / (\ln N)^{j+1} \to s+1.$$

д) Для неравновероятной схемы с выделенными исходами (см. п. д) раздела 2) в работе [8] получены интегральные представления для производящей функции и моментов времени ожидания $\nu'_m(N,k)$ до момента не менее m-кратной реализации каждого из $k \leqslant N$ выделенных исходов:

$$\mathsf{E}z^{\nu'_m(N,k)} = \frac{1-z}{z} \int_0^\infty e^{t(1-z^{-1})} \prod_{j=1}^{k} \overline{\Pi}_{m-1}(p_j t)\ dt$$

(это является конкретизацией соотношения (15) для рассматриваемой статистики и при $k = N$ совпадает с (12)),

$$\mathsf{E}\nu'_m(N,k) = \int_0^\infty \left[1 - \prod_{j=1}^{k} \overline{\Pi}_{m-1}(p_j t) \right] dt,$$

$$\mathsf{E}\nu'_m(N,k)(\nu'_m(N,k)+1) = 2\int_0^\infty t \left[1 - \prod_{j=1}^{k} \overline{\Pi}_{m-1}(p_j t) \right] dt.$$

Отметим частный случай этих представлений при $m = 1$ (ср. с (12′)):

$$\mathsf{E}z^{\nu'_1(N,k)} = 1 + (1-z)\sum_{r=1}^{k} (-1)^r \sum_{1 \leqslant j_1 < \ldots < j_r \leqslant k} \frac{1}{1 - z(1 - p_{j_1} - \ldots - p_{j_r})},$$

$$\mathsf{E}\nu'_1(N,k) = \sum_{r=1}^{k} (-1)^{r-1} \sum_{1 \leqslant j_1 < \ldots < j_r \leqslant k} (p_{j_1} + \ldots + p_{j_r})^{-1},$$

$$\mathsf{E}\nu'_1(N,k)(\nu'_1(N,k)+1) = 2\sum_{r=1}^{k} (-1)^{r-1} \sum_{1 \leqslant j_1 < \ldots < j_r \leqslant k} (p_{j_1} + \ldots + p_{j_r})^{-2};$$

в случае равновероятной схемы справедливы представления

$$\mathsf{E}z^{\nu'_1(N,k)} = k!\, z^k \prod_{j=1}^{k} \frac{1}{N - (N-j)z},$$

$$\mathsf{E}\nu'_1(N,k) = N \sum_{n=1}^{k} \frac{1}{n},$$

$$\mathsf{D}\nu'_1(N,k) = N^2 \sum_{n=1}^{k} \frac{1}{n^2} - N \sum_{n=1}^{k} \frac{1}{n}.$$

е) Исследование времени ожидания $\nu_m(N_1, N, k)$ до не менее m-кратной реализации некоторых k из N_1 выделенных исходов (см. п. д) раздела 2) проведено в работе [6]. Используя соотношение

$$\mathsf{E}z^{\nu_m(N_1,N,k)} = \mathsf{E}'\left(\frac{\alpha_N z}{1 - (1-\alpha_N)z} \right)^{\nu_m(N_1,k)},$$

где $\alpha_N = p_1 + \ldots + p_{N_1} \leqslant 1$ и E' означает среднее относительно распределения

$p'_j = p_j / \alpha_N$, $j = 1,\ldots,N_1$, в обычной (без выделения исходов) схеме с N_1 исходами, в [6] все результаты, известные для стандартной с.в. $\nu_m(N_1,k)$ (см. (12) и пп. в) и г) раздела 4), перенесены на с.в. $\nu_m(N_1,N,k)$. В частности, выяснено влияние параметра $1-\alpha_N$ на асимптотическое поведение этой с.в. – его влияние на тип предельного распределения начинает сказываться лишь при $\sqrt{N}(1-\alpha_N) \geqslant c > 0$. В качестве иллюстрации приведем соответствующие утверждения для зоны "малых" значений параметра k (ср. с теоремой 12). Пусть $N_1, N, k \to \infty$ так, что

$$\frac{k^2}{2}\sum_{j=1}^{N_1} p_j^2 \to \lambda < \infty,$$

тогда при $(1-\alpha_N)k \to \beta < \infty$

$$\mathcal{L}(\nu_1(N_1,N,k)-k) \to \Pi(\lambda+\beta),$$

а при $(1-\alpha_N)k \to \infty$

$$\mathcal{L}(\nu_1(N_1,N,k)) \sim \mathcal{N}(k/\alpha_N, k(1-\alpha_N)/\alpha_N).$$

СПИСОК ЛИТЕРАТУРЫ

1. Иванов В.А. Предельные теоремы в схеме размещения со случайными уровнями // Матем. заметки. – 1982. – Т. 31, вып. 4. – С. 619–631.
2. Иванов В.А. Асимптотика моментов в схеме размещения со случайными уровнями // Статистич. методы оценивания и проверки гипотез: Межвуз. сб. – Пермь, 1986. – С. 79–83.
3. Иванов В.А., Ивченко Г.И., Медведев Ю.И. Дискретные задачи в теории вероятностей // Итоги науки и техники. Сер. Теория вероятн.; мат. статист., теор. киберн. Т. 22. – М.: ВИНИТИ, 1984. – С. 3–60.
4. Иванов В.А., Ивченко Г.И., Протасов А.М. О времени достижения случайного уровня в полиномиальной схеме // Вероятн. процессы и их прил.: Межвуз. сб. – М.: МИЭМ, 1984. – С. 89–100.
5. Иванова Т.В. О некоторых статистических критериях в полиномиальной схеме // Тр. Моск. ин-та электрон. машиностр. – 1976. – Вып. 57. – С. 149–162.
6. Иванова Т.В. Предельные теоремы в схеме последовательного заполнения ячеек при разбиении их на группы // Вероятн. процессы и управление: Межвуз. сб. – Краснодар, 1977. – С. 24–32.
7. Иванова Т.В. О статистических критериях, основанных на времени ожидания // Статистич. методы оценивания и проверки гипотез: Межвуз. сб. – Пермь, 1984. – С. 23–28.
8. Иванова Т.В., Ивченко Г.И. О времени ожидания в полиномиальной схеме // Тр. Моск. ин-та электрон. машиностр. – 1975. – Вып. 44. – С. 77–95.
9. Иванова Т.В., Ивченко Г.И. Об одном критерии последовательного типа // Теория вероятностей и ее применения.– 1976. –Т. 21, вып. 4. – С. 854–857.
10. Ивченко Г.И. Предельные теоремы в задаче о размещении // Теория вероятностей и ее применения. – 1971. – Т. 16, вып. 2. – С. 292–305.
11. Ивченко Г.И. Время ожидания и вариационный ряд частот в полиномиальной схеме // Тр. Моск. ин-та электрон. машиностр. – 1973. – Вып. 32. – С. 39–64.
12. Ивченко Г.И. Время ожидания и проверка гипотез в полиномиальной схеме // Теория вероятностей и ее применения. – 1974. –Т. 19, вып. 4. – С. 839–844.
13. Ивченко Г.И. О некоторых распределениях, связанных с временем ожидания в полиномиальной схеме // Теория вероятностей и ее применения. – 1975. – Т. 20, вып. 3. – С. 557–570.
14. Ивченко Г.И. Разделимые статистики в схеме со случайными уровнями // Вероятн. задачи дискретной математики: Межвуз. сб. – М.: МИЭМ, 1987. – С. 151–161.

15. И в ч е н к о Г.И. Разделимые статистики в обратной задаче о размещении // Дискретная математика. – 1989. – Т. 1, вып. 1. – С. 60–73.
16. И в ч е н к о Г.И. Разделимые статистики и один класс моментов остановки для полиномиальных испытаний // Докл. АН СССР. – 1989. – Т. 305, № 1. – С. 26–30.
17. И в ч е н к о Г.И., И в а н о в а Т.В. Предельные теоремы для времени ожидания // Тр. Моск. ин-та электрон. машиностр. – 1975. – Вып. 44. – С. 96–100.
18. И в ч е н к о Г.И., М е д в е д е в Ю.И. Асимптотическое поведение числа комплектов частиц в классической задаче о размещении // Теория вероятностей и ее применения. – 1966. – Т. 11, вып. 4. – С. 701–708.
19. И в ч е н к о Г.И., М о р о з о в а Н.М. Отрицательное полиномиальное распределение и его вероятностный анализ // Вероятн. процессы и их прил.: Межвуз. сб. – М.: МИЭМ, 1989. – С. 3–8.
20. И в ч е н к о Г.И., М о р о з о в а Н.М. Оценивание для отрицательного полиномиального распределения // Вероятн. процессы и их прил.: Межвуз. сб. – М.: МИЭМ, 1991. – С. 3–8.
21. И в ч е н к о Г.И., Ш а р о в а Е.Е. Статистические процедуры для выделения наиболее вероятных исходов // Тр. Моск. ин-та электрон. машиностр. – 1976. – Вып. 57. – С. 186–200.
22. К о л ч и н В.Ф., Ч и с т я к о в В.П. Комбинаторные задачи теории вероятностей // Итоги науки и техники. Сер. Теория вероятн., мат. статист., теор. киберн. Т. 11. – М.: ВИНИТИ, 1974. – С. 5–45.
23. К о л ч и н В.Ф., С е в а с т ь я н о в Б.А., Ч и с т я к о в В.П. Случайные размещения. – М.: Наука, 1976. – 224 с.
24. B a u m L.E., B i l l i n g s l e y P. Asymptotic distributions for the coupon collector's problem // Ann. Math. Statist. – 1965. – V. 36, № 6. – P. 1835–1839.
25. B e k e s s y A. On classical occupancy problems.II // Magy. Tud. Acad. Mat. Kutató Int. Közl. – 1964. – V. 9A, № 1–2. – P. 133–141.
26. B r a y t o n R.K. On the asymptotic behavior of the number of trials necessary to complete a set with random selection // J. Math. Anal. Appl. – 1963. – V. 7, № 1. – P. 31–61.
27. E r d ö s P., R e n y i A. On a classical problem of probability theory // Magy. Tud. Akad. Mat. Kutató Int. Közl. – 1961. – V. 6A, № 1–2. – P. 215–220.
28. F l a t t ö L. Limit theorems for some random variables assotiated with urn models // Ann. Probab. – 1982. – V. 10, № 4. – P. 927–934.
29. G u p t a S.S., P a n c h a p a k e s a n S. Subset selection procedures: review and assessment // Amer. J. Math. and Manag. Sci. – 1985. – V. 5. – P. 235–311.
30. H o l s t L. Limit theorems for some occupancy and sequential occupancy problems // Ann. Math. Statist. – 1971. – V. 42, № 5. – P. 1671–1680.
31. H o l s t L. On sequential occupancy problems // J. Appl. Probab. – 1981. – V. 18, № 2. – P. 435–442.
32. H o l s t L. On birthday, collector's, occupancy and other classical urn problems // Intern. Statist. Rev. – 1986.– V. 54, №1. – P. 15–27.
33. H o l s t L., H ü s l e r J. Sequential urn schemes and birth processes // Adv. Appl. Probab. – 1985.– V. 17. – P. 257–279.
34. J a n s o n S. Limit theorems for some sequential occupancy problems // J. Appl. Probab. – 1983. – V 20, № 3. – P. 545–553.
35. K a p l a n N. A generalization of a result of Erdös and Renyi // J. Appl. Probab. – 1977. – V. 14, № 1. – P. 212–216.
36. N e w m a n D., S h e p p L. The double dixie cap problem // Amer. Math. Month. – 1960. – V. 67, № 1. – P. 58–61.
37. R o s e n B. On the coupon collector's waiting time // Ann. Math. Statist. – 1970. – V. 41, № 6. – P. 1952–1969.
38. von S c h e l l i n g H. Coupon collecting for unequal probabilities // Amer. Math. Month. – 1954. – V. 61, № 5. – P. 306–311.
39. Y o u n g D.H. Moment relations for order statictics of the standardized gamma distribution and the inverse multinomial distribution // Biometrika. – 1971. – V. 58. – P. 637–640.
40. И в ч е н к о Г .И. , М о р о з о в а И.М. Отрицательное полиномиальное распределение // Дискретная математика.– 1993.– Т.5, вып.2. – С.138– 149.(Добавлено при корректуре.)

Статья поступила 01.09.92