Труды по дискретной математике. Том 1, стр. 221–226
(Под. ред. В. Н. Сачкова и др.)
Научное издательство «ТВП», 1997 г.

НЕКОТОРЫЕ НЕРАВЕНСТВА ДЛЯ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ ЧИСЛА НЕПОЯВИВШИХСЯ s-ЦЕПОЧЕК

В. Г. МИХАЙЛОВ

В работе получены неравенства, связывающие вероятности больших уклонений для левого и правого хвостов функции распределения числа непоявившихся s-цепочек в последовательности полиномиальных испытаний. Аналогичные неравенства получены и для функции распределения числа повторений s-цепочек в этой схеме.

Пусть $X_1, X_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин, принимающих лишь значения $1, 2, \ldots, N$. Пусть

$$\mathbf{P}\{X_i = k\} = p_k, \qquad k = 1, \ldots, N, \quad i = 1, 2, \ldots.$$

Мы будем считать, что $0 < p_1 \leqslant p_2 \leqslant \cdots \leqslant p_N$.

В последовательности $\{X_t\}$ нас будут интересовать s-цепочки

$$Y_i = (X_i, X_{i+1}, \ldots, X_{i+s-1}), \qquad i = 1, \ldots, n.$$

Пусть $\mu_0(s,n)$ — число непоявившихся элементов множества $\{1, 2, \ldots, N\}^s$ среди $Y_1, \ldots, Y_n$, или, более кратко, число непоявившихся s-цепочек, а $\xi_2(s,n)$ — число повторений в последовательности s-цепочек $Y_1, \ldots, Y_n$, определяемое как

$$\xi_2(s,n) = \sum_{1 \leqslant i < j \leqslant n} I\{Y_i = Y_j\}.$$

Здесь и далее $I\{A\}$ — индикатор случайного события A.

Центральной в предлагаемой работе является

Теорема 1. *Пусть $0 < a \leqslant b$. Тогда*

$$\begin{aligned}
\mathbf{P}\left\{\xi_2(s,n) \leqslant a\right\} &\leqslant \mathbf{P}\left\{\mu_0(s,n) \leqslant a + N^s - n\right\} \\
&\leqslant s \exp\left\{-\frac{(b-a)^2}{2s^2 b}\right\} + \mathbf{P}\left\{\mu_0(s, n-s) \geqslant N^s - \left(\frac{2b}{p_1^s}\right)^{1/2}\right\} \\
&\leqslant s \exp\left\{-\frac{(b-a)^2}{2s^2 b}\right\} + \mathbf{P}\left\{\xi_2(s, n-s) \geqslant n - \left(\frac{2b}{p_1^s}\right)^{1/2}\right\}. \qquad (1)
\end{aligned}$$

© В. Г. Михайлов, 1997 г.

Если при этом

$$0 < a \leqslant b \leqslant (n-s)p_1^s, \tag{2}$$

то

$$\mathbf{P}\left\{\xi_2(s,n) \leqslant a\right\} \leqslant \mathbf{P}\left\{\mu_0(s,n) \leqslant a + N^s - n\right\} \leqslant s\exp\left\{-\frac{(b-a)^2}{2s^2 b}\right\}. \tag{3}$$

Неравенства (1) позволяют оценить вероятности левых хвостов распределений случайных величин $\mu_0(s,n)$ и $\xi_2(s,n)$ через вероятности правых хвостов. Оценивание последних в определенных случаях представляется более простой задачей.

Исследованию свойств случайных величин $\mu_0(s,n)$ и $\xi_2(s,n)$ посвящено значительное число работ (см., например, [1]–[8]). Достаточно полно изучено их предельное поведение при $n, N \to \infty$. Известны условия сходимости их распределений к пуассоновскому и родственным дискретным распределениям и оценка скорости этой сходимости. Хорошо изучены условия асимптотической нормальности случайной величины $\xi_2(s,n)$, чему способствовало то обстоятельство, что она является U-статистикой. Труднее оказалось исследовать случайную величину $\mu_0(s,n)$. Заметное продвижение было достигнуто в рамках исследования свойств марковских размещений [2], и наконец, последний значительный шаг был сделан в [8], где получены достаточные условия асимптотической нормальности величины $\mu_0(s,n)$. Каких-либо оценок для вероятностей уклонений до сих пор получено не было.

Перейдем к доказательству теоремы 1. Для этого нам понадобится ряд вспомогательных результатов и, в частности, следующее утверждение, доказанное в работе [9].

Пусть на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbf{P})$ заданы последовательность σ-алгебр $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subset \cdots \subset \mathcal{F}$ и последовательность случайных величин $\xi_1, \xi_2, \ldots$, удовлетворяющие условиям:

1) $0 \leqslant \xi_i \leqslant 1$;

2) случайная величина ξ_i измерима относительно $\mathcal{F}_i$, $i = 1, 2, \ldots$.

Пусть $M_i = \mathbf{E}\{\xi_i \mid \mathcal{F}_{i-1}\}$, а τ — некоторый марковский момент относительно последовательности σ-алгебр $\{\mathcal{F}_i\}$. Тогда при $0 < a \leqslant b$ вероятности

$$\mathbf{P}\left\{\sum_{i=1}^{\tau} \xi_i \leqslant a, \sum_{i=1}^{\tau} M_i \geqslant b\right\}, \quad \mathbf{P}\left\{\sum_{i=1}^{\tau} \xi_i \geqslant b, \sum_{i=1}^{\tau} M_i \leqslant a\right\}$$

не превосходят $\exp\{-\frac{(b-a)^2}{2b}\}$.

Далее в качестве марковского момента мы будем рассматривать неслучайное натуральное число: $\tau = n$, и положим

$$M_i^{(k)} = \mathbf{E}\{\xi_i \mid \mathcal{F}_{i-k}\}, \qquad i > k.$$

Тогда

$$\xi_i = M_i^{(0)}, \quad M_i = M_i^{(1)}, \quad M_i^{(k)} = \mathbf{E}\{M_i^{(k-1)} \mid \mathcal{F}_{i-k}\}.$$

Теорема 2. *Пусть* $0 < a \leqslant b$, *а случайные величины* $\xi_1, \ldots, \xi_n$ *удовлетворяют условиям* 1) *и* 2). *Тогда*

$$\begin{aligned} &\mathbf{P}\left\{\sum_{i=1+r}^{n} \xi_i \leqslant a,\ \sum_{i=1+r}^{n} M_i^{(r)} \geqslant b\right\} \leqslant r \exp\left\{-\frac{(b-a)^2}{2r^2 b}\right\}, \\ &\mathbf{P}\left\{\sum_{i=1+r}^{n} \xi_i \geqslant b,\ \sum_{i=1+r}^{n} M_i^{(r)} \leqslant a\right\} \leqslant r \exp\left\{-\frac{(b-a)^2}{2r^2 b}\right\}. \end{aligned} \tag{4}$$

Доказательство теоремы 2. Для сокращения записи введем обозначение

$$M^{(k)} = \sum_{i=1+r}^{n} M_i^{(k)}.$$

Пусть $0 < a \leqslant c \leqslant b$. Очевидно, что

$$\begin{aligned} &\mathbf{P}\{M^{(0)} \leqslant a, M^{(2)} \geqslant b\} \\ &\quad \leqslant \mathbf{P}\{M^{(0)} \leqslant a, M^{(1)} \geqslant c, M^{(2)} \geqslant b\} \\ &\qquad + \mathbf{P}\{M^{(0)} \leqslant a, M^{(1)} \leqslant c, M^{(2)} \geqslant b\} \\ &\quad \leqslant \mathbf{P}\{M^{(0)} \leqslant a, M^{(1)} \geqslant c\} \\ &\qquad + \mathbf{P}\{M^{(1)} \leqslant c, M^{2} \geqslant b\}. \end{aligned}$$

Аналогично, разместив на отрезке $[a, b]$ не одну точку c, а сразу $r-1$ точек

$$c_k = \frac{(r-k)a + kb}{r}, \qquad k = 1, \ldots, r-1,$$

получим

$$\begin{aligned} &\mathbf{P}\{M^{0} \leqslant a, M^{(r)} \geqslant b\} \\ &\quad \leqslant \sum_{k=0}^{r-1} \mathbf{P}\left\{M^{(k)} \leqslant \frac{(r-k)a + kb}{r}, M^{(k+1)} \geqslant \frac{(r-k-1)a + (k+1)b}{r}\right\}. \end{aligned}$$

Каждое слагаемое мы оцениваем с помощью цитированного результата из [9] и после небольшого загрубления приходим к первому из неравенств в (4):

$$\begin{aligned} &\mathbf{P}\{M^{(0)} \leqslant a, M^{(r)} \geqslant b\} \\ &\quad \leqslant \sum_{k=0}^{r-1} \exp\left\{-\frac{([b-a]/r)^2}{2[(r-k+1)a + (k+1)b]/r}\right\} \\ &\quad \leqslant r \exp\left\{-\frac{(b-a)^2}{2r^2 b}\right\}. \end{aligned}$$

Аналогично доказывается и второе неравенство в (4). Теорема 2 доказана.

Рассмотрим теперь случайную величину

$$Z = \sum_{k=1}^{n} I\{Y_k \in \{Y_1, \ldots, Y_{k-1}\}\}.$$

Лемма. *Пусть* $0 < a \leqslant b$. *Тогда*

$$\mathbf{P}\{Z \leqslant a\} \leqslant s \exp\left\{-\frac{(b-a)^2}{2s^2 b}\right\} + \mathbf{P}\left\{\mu_0(s, n-s) \geqslant N^s - \left(\frac{2b}{p_1^s}\right)^{1/2}\right\}. \tag{5}$$

Если же выполнено условие (2), *то*

$$\mathbf{P}\{Z \leqslant a\} \leqslant a \exp\left\{-\frac{(b-a)^2}{2s^2 b}\right\}. \tag{6}$$

Доказательство леммы. Введем вспомогательные случайные величины

$$u_k = I\big\{Y_k \in \{Y_1, \ldots, Y_{k-s}\}\big\}, \qquad k = s+1, \ldots, n,$$
$$U = u_{s+1} + u_{s+2} + \cdots + u_n.$$

Очевидно, что $u_k \leqslant I\{Y_k \in \{Y_1, \ldots, Y_{k-1}\}\}$. Поэтому

$$U \leqslant Z. \tag{7}$$

Введем случайные величины

$$v_k = \mathbf{E}\{u_k \mid \mathcal{F}_{k-s}\}, \qquad k = s+1, \ldots, n,$$

где $\mathcal{F}_t$ есть σ-алгебра, порожденная случайными величинами $Y_1, \ldots, Y_t$ (или, что то же самое, — случайными величинами $X_1, \ldots, X_{t+s-1}$). Положим

$$V = v_{s+1} + v_{s+2} + \cdots + v_n.$$

Заметим, что к вероятности $\mathbf{P}\{U \leqslant a, V \geqslant b\}$ можно применить теорему 2 с $r = s$.

Легко проверить, что с вероятностью 1

$$\begin{aligned} v_k &= \mathbf{P}\left\{Y_k \in \{Y_1, \ldots, Y_{k-s}\} \mid X_1, \ldots, X_{k-1}\right\} \\ &\geqslant p_1^s \left|\{Y_1, \ldots, Y_{k-s}\}\right|, \end{aligned}$$

где $|A|$ обозначает число элементов множества A. Это обстоятельство позволяет оценить снизу случайную величину V следующим образом (все неравенства выполняются с вероятностью 1):

$$\begin{aligned} V &\geqslant p_1^s \sum_{t=1}^{n-s} \left|\{Y_1, \ldots, Y_t\}\right| \\ &\geqslant p_1^s \sum_{t=1}^{n-s} \max\left\{1, \left|\{Y_1, \ldots, Y_{n-s}\}\right| - \big((n-s) - t\big)\right\} \\ &\geqslant \max\left\{(n-s)p_1^s, \frac{p_1^s}{2}\left|\{Y_1, \ldots, Y_{n-s}\}\right|^2\right\}. \end{aligned}$$

Другими словами, с вероятностью 1

$$V \geqslant (n-s)p_1^s, \tag{8}$$

$$V \geqslant \frac{p_1^s}{2}\big(N^s - \mu_0(s, n-s)\big)^2. \tag{9}$$

Если выполнено условие (2), то из (8) следует, что $\mathbf{P}\{V \geqslant b\} = 1$. Значит,

$$\mathbf{P}\{U \leqslant a\} = \mathbf{P}\{U \leqslant a, V \geqslant b\}. \tag{10}$$

Правую часть здесь оцениваем с помощью теоремы 2, а затем применяем (7):

$$\mathbf{P}\{Z \leqslant a\} \leqslant \mathbf{P}\{U \leqslant a\} \leqslant s \exp\left\{-\frac{(b-a)^2}{2s^2 b}\right\}.$$

Итак, неравенство (6) доказано.

Перейдем к доказательству неравенства (5). Из (9) следует, что

$$\mathbf{P}\{V \leqslant b\} \leqslant \mathbf{P}\left\{\mu_0(s, n-s) \geqslant N^s - \left(\frac{2b}{p_1^s}\right)^{1/2}\right\}.$$

Поэтому с учетом неравенства (7) и теремы 2

$$\begin{aligned}\mathbf{P}\{Z \leqslant a\} &\leqslant \mathbf{P}\{U \leqslant a\} \\ &\leqslant \mathbf{P}\{U \leqslant a, V \geqslant b\} + \mathbf{P}\{V \leqslant b\} \\ &\leqslant s \exp\left\{-\frac{(b-a)^2}{2s^2 b}\right\} + \mathbf{P}\left\{\mu_0(s, n-s) \geqslant N^s - \left(\frac{2b}{p_1^s}\right)^{1/2}\right\}.\end{aligned}$$

Неравенство (5), а вместе с ним и лемма доказаны.

Доказательство теоремы 1. Рассмотрим число появившихся цепочек $N^s - \mu_0(s, n)$. Очевидно, что

$$\begin{aligned}N^s - \mu_0(s, n) &= 1 + \sum_{k=2}^{n} I\{Y_k \notin \{Y_1, \ldots, Y_{k-1}\}\} \\ &= 1 + \sum_{k=2}^{n} \Big(1 - I\{Y_k \in \{Y_1, \ldots, Y_{k-1}\}\}\Big) = n - Z.\end{aligned}$$

Поэтому $Z = \mu_0(s, n) - N^s + n$. Далее,

$$I\{Y_k \in \{Y_1, \ldots, Y_{k-1}\}\} \leqslant \sum_{i=1}^{k-1} I\{Y_i = Y_k\}.$$

Значит,

$$Z \leqslant \sum_{k=1}^{n} \sum_{i=1}^{k-1} I\{Y_i = Y_k\} = \xi_2(s, n).$$

Из (11), (12), (5) и (6) следуют неравенства (1) и (3).

Теорема 1 доказана.

СПИСОК ЛИТЕРАТУРЫ

1. *Беляев П. Ф.* О вероятности непоявления заданного числа s-цепочек в сложных цепях Маркова. — Теория вероятн. и ее примен., 1965, т. X, в. 3, с. 547–551.
2. *Зубков А. М.* Неравенства для вероятностей переходов с запрещениями и их применения. — Матем. сб., 1979, т. 109 (151), № 4 (8), с. 491–532.
3. *Зубков А. М.*, *Михайлов В. Г.* Предельные распределения случайных величин, связанных с длинными повторениями в последовательности независимых испытаний. — Теория вероятн. и ее примен., 1974, т. XIX, в. 1, с. 173–181.
4. *Зубков А. М.*, *Михайлов В. Г.* О повторениях s-цепочек в последовательности независимых величин. — Теория вероятн. и ее примен., 1979, т. XXIV, в. 2, с. 267–279.
5. *Колчин В.Ф.*, *Чистяков В.П.* Предельные распределения числа непоявившихся s-цепочек в полиномиальной схеме. — Теория вероятн. и ее примен., 1974, т. XIX, в. 4, с. 855–864.
6. *Михайлов В. Г.* Асимптотическая нормальность в схеме конечно-зависимого размещения частиц по ячейкам. — Матем. сб., 1982, т. 119 (161), № 4 (12), с. 509–520.
7. *Михайлов В. Г.* Асимптотическая нормальность разделимых статистик от частот m-цепочек. — Дискретн. матем., 1989, т. 1, № 4, с. 92–103.
8. *Тихомирова М. И.* Об асимптотической нормальности числа непоявившихся s-цепочек. — Дискретн. матем., 1992, т. 4, № 2, с. 122–129.
9. *Freedman D.* Another note on the Borel–Cantelli lemma and strong law, with Poisson approximation as a by-product. — Ann. Probab., 1973, v. 1, № 6, p. 910–925.