S©MR ISSN 1813-3304

СИБИРСКИЕ ЭЛЕКТРОННЫЕ
МАТЕМАТИЧЕСКИЕ ИЗВЕСТИЯ

Siberian Electronic Mathematical Reports

http://semr.math.nsc.ru

Том 14, стр. 1289–1298 (2017) УДК 519.214.5
DOI 10.17377/semi.2017.14.109 MSC 60F17

ФУНКЦИОНАЛЬНАЯ ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА В БЕСКОНЕЧНОЙ УРНОВОЙ СХЕМЕ ДЛЯ РАСПРЕДЕЛЕНИЙ СО СВЕРХТЯЖЕЛЫМИ ХВОСТАМИ

М.Г. ЧЕБУНИН

Abstract. We study a vector process of a number of urns with fixed quantities of balls in an infinite urn scheme. We assume that probabilities of entering an urn change regularly with exponent minus one. We prove a multidimensional functional central limit theorem for this process.

Keywords: infinite urn scheme; relative compactness; slow variation; functional central limit theorem.

1. Введение

В работе изучается следующая вероятностная модель: имеется n шаров и бесконечное число урн, занумерованных числами $1, 2, \ldots$. Каждый шар попадает в i-ю урну с вероятностью $p_i > 0$ независимо от всех других шаров ($\sum_{i=1}^{\infty} p_i = 1$). Обозначим через X_i номер урны, в которую попал i-й шар, а через $J_i(n)$ количество шаров, попавших в i-ю урну:

$$J_i(n) = \sum_{j=1}^{n} \mathbf{I}(X_j = i).$$

Нас интересуют статистики числа элементов, встретившихся не менее $k \geq 1$ раз в выборке $(X_1, \ldots, X_n)$:

$$R^*_{n,k} = \sum_{i=1}^{\infty} \mathbf{I}(J_i(n) \geq k), \tag{1}$$

Chebunin, M.G., Functional central limit theorem in an infinite urn scheme for distributions with superheavy tails.

Работа выполнена при поддержке РНФ, проект 17-11-01173.

Поступила 10 ноября 2017 г., опубликована 1 декабря 2017 г.

и статистики числа элементов, встретившихся ровно k раз:

$$R_{n,k} = R^*_{n,k} - R^*_{n,k+1}. \tag{2}$$

Будем также использовать обозначение $R_n \stackrel{def}{=} R^*_{n,1} = \sum_{k\geq 1} R_{n,k}$. Нетрудно видеть, что в нашей задаче можно без ограничения общности полагать, что $p_1 \geq p_2 \geq \ldots$.

Наряду с выборкой объема n будем рассматривать выборку случайного объема $\Pi(t)$, где $\{\Pi(t),\ t \geq 0\}$ — пуассоновский процесс с интенсивностью 1 и не зависящий от последовательности $X_1, X_2, \ldots$. Согласно известному свойству расщепления пуассоновских потоков, случайные процессы $\{J_i(\Pi(t)) \stackrel{def}{=} \Pi_i(t),\ t \geq 0\}$ являются пуассоновскими с интенсивностями p_i и взаимно независимы для разных i. Из определения (1) и (2) следует, что

$$R^*_{\Pi(t),k} = \sum_{i=1}^{\infty} \mathbf{I}(\Pi_i(t) \geq k), \quad R_{\Pi(t),k} = \sum_{i=1}^{\infty} \mathbf{I}(\Pi_i(t) = k).$$

Будем предполагать, что функция

$$\alpha(x) = \max\{j \ :\ \ p_j \geq 1/x\} = x^{\theta} L(x) \text{ при } \theta \in [0,1], \tag{3}$$

правильно меняется на бесконечности, где $L(x)$ — медленно меняющаяся функция при $x \to \infty$.

Данная работа является логическим продолжением работы [8], в которой при условии (3) была доказана функциональная центральная предельная теорема (ФЦПТ) для вектора $(R^*_{[nt],1}, R^*_{[nt],2}, \ldots, R^*_{[nt],\nu}, 0 \leq t \leq 1)$ при $\theta \in (0,1)$, а также ФЦПТ для $(R_{[nt]}, 0 \leq t \leq 1)$ при $\theta = 1$.

Обычные (нефункциональные) предельные теоремы для этих статистик при разных предположениях доказаны в работах [2–4, 7, 10–16]. В частности, при самых общих ограничениях на поведение вероятностей p_i (при условии неограниченного роста дисперсий) были доказаны: усиленный закон больших чисел для R_n, асимптотическая нормальность R_n, асимптотическая нормальность случайного вектора $(R_{n,1}, ..., R_{n,\nu})$, локальные предельные теоремы и др. В недавней работе [9] был предложен другой подход к доказательству функциональных теорем для процессов такого типа.

В данной работе удалось преодолеть ряд технических трудностей для распространения ФЦПТ из [8] на важный частный случай $\theta = 1$, соответствующий сколь угодно медленному убыванию хвоста распределения. Распространение ФЦПТ на случай $\theta = 0$ наряду с регулярностью (3) потребует дополнительных ограничений на распределение. Как отмечено в работах [13] и [4], при $\theta = 0$ сходимость дисперсий к бесконечности не гарантируется, возможны разные типы предельного поведения для разных статистик. Можно предположить, что также сходимость дисперсий к бесконечности не гарантирует относительной компактности.

Введем следующие обозначения: при $\theta = 1,\ k \geq 1$

$$L^*(n) = \int_0^{\infty} \frac{e^{-1/y}}{y} L(ny)dy, \quad \beta(n) = \beta_k(n) = \begin{cases} nL^*(n), & \text{если } k = 1; \\ \alpha(n), & \text{если } k > 1; \end{cases}$$

$$Y^*_{n,k}(t) = \frac{R^*_{[nt],k} - \mathbf{E}R^*_{[nt],k}}{(\beta(n))^{1/2}}, \quad Y_{n,k}(t) = \frac{R_{[nt],k} - \mathbf{E}R_{[nt],k}}{(\beta(n))^{1/2}},$$

$$Z^*_{n,k}(t) = \frac{R^*_{\Pi(nt),k} - \mathbf{E}R^*_{\Pi(nt),k}}{(\beta(n))^{1/2}}, \qquad K_{m,s} = \begin{cases} \frac{(m+s-2)!}{m!s!}, & \text{если } m+s \geq 2; \\ 0, & \text{если } m+s \leq 1; \end{cases}$$

где $\{\Pi(nt), t \in [0,1]\}$ — пуассоновский процесс с интенсивностью n.

Справедлива следующая основная теорема.

Теорема 1. *Пусть при $\theta = 1$ выполнено условие (3), $\nu \geq 1$ — целое число, тогда случайный процесс*

$$\left(Y^*_{n,1}(t), \ldots, Y^*_{n,\nu}(t),\ 0 \leq t \leq 1\right)$$

слабо сходится в равномерной метрике в $D((0,1)^\nu)$ к ν-мерному гауссовскому процессу с непрерывными траекториями, нулевым математическим ожиданием и ковариационной функцией $C^*(\tau,t) = (c^*_{ij}(\tau,t))^\nu_{i,j=1}$ такой, что

$$c^*_{11}(\tau,t) = \min(\tau,t), \quad c^*_{1j}(\tau,t) = 0 \quad \textit{при } j > 1,$$

и для любых $i,j \in \{2,\ldots,\nu\}$, $\tau \leq t$

$$c^*_{ij}(\tau,t) = \tau + \sum_{s=0}^{i-1}\sum_{m=0}^{j-s-1} \tau^s(t-\tau)^m t^{1-m-s} K_{m,s} - \sum_{s=0}^{i-1}\sum_{m=0}^{j-1} \tau^s t^m (t+\tau)^{1-m-s} K_{m,s} \quad \textit{при } i < j,$$

$$c^*_{ij}(\tau,t) = \tau + t\sum_{m=0}^{j-1} K_{m,0} - \sum_{s=0}^{i-1}\sum_{m=0}^{j-1} \tau^s t^m (t+\tau)^{1-m-s} K_{m,s} \quad \textit{при } i \geq j,$$

$$c^*_{ji}(t,\tau) = c^*_{ij}(\tau,t).$$

Ясно, что предельный ν-мерный гауссовский процесс самоподобен с параметром Херста $H = 1/2$. Также из теоремы 1 легко получить следующее следствие.

Следствие 1. *В условиях теоремы 1 случайный процесс*

$$((Y_{n,0}(t), Y_{n,1}(t), \ldots, Y_{n,\nu}(t)),\ 0 \leq t \leq 1)$$

при $Y_{n,0}(t) := Y^*_{n,1}(t)$, слабо сходится в равномерной метрике в $D((0,1)^{\nu+1})$ к $(\nu+1)$-мерному гауссовскому процессу с непрерывными траекториями, нулевым математическим ожиданием и ковариационной функцией $C(\tau,t) = (c_{ij}(\tau,t))^\nu_{i,j=0}$,

$$c_{00}(\tau,t) = c_{01}(\tau,t) = c_{11}(\tau,t) = \min\{\tau,t\}, \quad c_{0i}(\tau,t) = c_{1i}(\tau,t) = 0 \quad \textit{при } i \geq 2,$$

$$c_{ij}(\tau,t) = K_{i,j-i}\tau^i(t-\tau)^{j-i}t^{1-j} - K_{i,j}\tau^i t^j (t+\tau)^{1-i-j} \quad \textit{при } 2 \leq i \leq j,\ \tau \leq t,$$

$$c_{ij}(\tau,t) = -K_{i,j}\tau^i t^j (t+\tau)^{1-i-j} \quad \textit{при } i > j \geq 2,\ \tau \leq t,$$

$c_{ji}(t,\tau) = c_{ij}(\tau,t)$.

Ясно, что $L(n) \to 0$ при $n \to \infty$, если $\theta = 1$. Карлин ([13], лемма 4) доказал, что $L^*(n) \to 0$ при $n \to \infty$, медленно меняется, и

$$\lim_{n\to\infty} \frac{L^*(n)}{L(n)} \stackrel{def}{=} \lim_{n\to\infty} \delta_n = 0. \tag{4}$$

Поэтому, чтобы вычислить ковариационную функцию $C(\tau,t) = (c_{ij}(\tau,t))^\nu_{i,j=0}$, достаточно воспользоваться следующими формулами при $i,j \in \{2,...,\nu\}, \tau \leq t$:

$$c_{00}(\tau,t) = c^*_{11}(\tau,t), \quad c_{01}(\tau,t) = c^*_{11}(\tau,t) - c^*_{1,2}(\tau,t) \cdot \lim_{n\to\infty}\sqrt{\delta_n},$$

$$c_{11}(\tau,t) = c^*_{11}(\tau,t) - 2c^*_{1,2}(\tau,t) \cdot \lim_{n\to\infty}\sqrt{\delta_n} + c^*_{2,2}(\tau,t) \cdot \lim_{n\to\infty}\delta_n,$$

$$c_{0i}(\tau,t) = c^*_{1,i}(\tau,t) - c^*_{1,i+1}(\tau,t),$$
$$c_{1i}(\tau,t) = c^*_{1,i}(\tau,t) - c^*_{1,i+1}(\tau,t) - (c^*_{2,i}(\tau,t) - c^*_{2,i+1}(\tau,t)) \cdot \lim_{n\to\infty} \sqrt{\delta_n},$$
$$c_{ij}(\tau,t) = c^*_{ij}(\tau,t) - c^*_{i+1,j}(\tau,t) - c^*_{i,j+1}(\tau,t) + c^*_{i+1,j+1}(\tau,t).$$

2. Доказательство теоремы 1

Напомним пару вспомогательных утверждений, доказательство которых можно найти в [8].

Лемма 1. *Пусть $\tau \le t$, тогда для любого $k \ge 1$*

$$\mathbf{E}(R^*_{\Pi(t),k} - R^*_{\Pi(\tau),k}) \le \mathbf{E}R_{\Pi(t-\tau)}.$$

Лемма 2. *Для любой пары чисел $\varepsilon,\ \delta \in (0,1)$ существует число $N = N(\varepsilon,\delta)$ такое, что для любого $n \ge N$,*

$$\mathbf{P}(\forall t \in [0,1] \quad \exists \tau : |\tau - t| \le \delta,\ \Pi(n\tau) = [nt]) \ge 1 - \varepsilon.$$

Докажем еще одно вспомогательное утверждение.

Лемма 3. *Пусть при $\theta = 1$ выполнено условие (3), $t_2 - t_1 = \delta \ge 0$, тогда существуют $n_0 \ge 1$ и $C < \infty$ такие, что*

$$\frac{\mathbf{E}(R^*_{\Pi(t_2 n),k} - R^*_{\Pi(t_1 n),k})}{\alpha(n)} \le C\delta^{1/2},$$

для любых $\delta \in [0,1],\ n \ge n_0,\ k \ge 2$.

Доказательство. Положим $t = nt_1$ и $\tau = nt_2$, тогда

$$\mathbf{E}(R^*_{\Pi(t),k} - R^*_{\Pi(\tau),k}) = \mathbf{E}\left(\sum_{i=1}^{\infty} \mathbf{I}(\Pi_i(t) \ge k, \Pi_i(\tau) < k)\right)$$
$$= \sum_{i=1}^{\infty}\sum_{j=0}^{k-1} \mathbf{P}(\Pi_i(\tau) = j)\mathbf{P}(\Pi_i(t) - \Pi_i(\tau) \ge k - j)$$
$$\le \sum_{i=1}^{\infty} \mathbf{P}(\Pi_i(\tau) = k-1)\mathbf{P}(\Pi_i(t-\tau) = 1) + \sum_{i=1}^{\infty} \mathbf{P}(\Pi_i(t-\tau) \ge 2)$$
$$= \sum_{i=1}^{\infty} \frac{p_i^k \tau^{k-1}(t-\tau)}{(k-1)!} e^{-p_i t} + \mathbf{E}R^*_{\Pi(t-\tau),2} \le \frac{k(t-\tau)}{t}\mathbf{E}R_{\Pi(t),k} + \mathbf{E}R^*_{\Pi(t-\tau),2}.$$

В силу того, что ([13], формула (23))

$$\lim_{x\to\infty} \frac{\mathbf{E}R_{\Pi(x),k}}{\alpha(x)} = \frac{1}{k(k-1)}, \quad \lim_{x\to\infty} \frac{\mathbf{E}R^*_{\Pi(x),2}}{\alpha(x)} = 1$$

существует $x_1 > 1$ такое, что при $x \ge x_1$

$$\mathbf{E}R_{\Pi(x),k} < \mathbf{E}R^*_{\Pi(x),2} < 2\alpha(x).$$

В силу представления Карамата (теорема 2.1, приложение 6, неравенство (A6.2.10) в [6]), найдется $x_2 > 0$ такое, что при всех x и $\delta \in (0,1]$ удовлетворяющих неравенству $x\delta \ge x_2$, выполняется

$$\frac{L(x\delta)}{L(x)} \le 2\delta^{-1/2}.$$

Пусть $nt_2 > n\delta > \max\{x_1, x_2\} = x_0$, тогда

$$\frac{\mathbf{E}R^*_{\Pi(n\delta),2}}{\alpha(n)} \le 2\frac{n\delta L(n\delta)}{nL(n)} \le 4\delta^{1/2}, \quad \frac{\mathbf{E}R_{\Pi(nt_2),k}}{\alpha(n)} \le 2\frac{nt_2 L(nt_2)}{nL(n)} \le 4t_2^{1/2}.$$

Если $n\delta \le x_0$ и $nt_2 \le x_0$, то

$$\frac{\mathbf{E}R^*_{\Pi(n\delta),2}}{\alpha(n)} \le \frac{\mathbf{E}\Pi(n\delta)}{\alpha(n)} = \frac{n\delta}{nL(n)}, \quad \frac{\mathbf{E}R_{\Pi(nt_2),k}}{\alpha(n)} \le \frac{\mathbf{E}\Pi(nt_2)}{\alpha(n)} = \frac{nt_2}{nL(n)}.$$

Выберем n_0 такое, что для всех $n \ge n_0$ выполнено $nL(n) \ge n^{1/2}$, тогда

$$\frac{n\delta}{nL(n)} \le \frac{n\delta}{n^{1/2}} = (n\delta)^{1/2}\delta^{1/2} \le x_0^{1/2}\delta^{1/2}, \quad \frac{nt_2}{nL(n)} \le x_0^{1/2}t_2^{1/2}.$$

Положим $c = \max\{4, x_0^{1/2}\}$, тогда (т. к. $t_2 > \delta$) для всех $n \ge n_0$

$$\frac{\mathbf{E}(R^*_{\Pi(t_2 n),k} - R^*_{\Pi(t_1 n),k})}{\alpha(n)} \le \frac{\frac{k(t-\tau)}{t}\mathbf{E}R_{\Pi(t),k} + \mathbf{E}R^*_{\Pi(t-\tau),2}}{\alpha(n)}$$

$$\le \frac{k\delta \cdot ct_2^{1/2}}{t_2} + c\delta^{1/2} \le (k+1)c\delta^{1/2} = C\delta^{1/2}.$$

□

Приступим к доказательству теоремы 1.

Шаг 1 (ковариации)

Вычислим ковариацию $R^*_{\Pi(\tau),1}$ и $R^*_{\Pi(t),j}$ при $j \ge 2$:

$$\widetilde{c}_{1j}(\tau, t) = \mathbf{cov}(R^*_{\Pi(\tau),1}, R^*_{\Pi(t),j})$$

$$= \sum_{k=1}^{\infty}\Big(\mathbf{P}(\Pi_k(\tau) \ge 1, \Pi_k(t) \ge j) - \mathbf{P}(\Pi_k(\tau) \ge 1)\mathbf{P}(\Pi_k(t) \ge j)\Big).$$

Ясно, что $|\widetilde{c}_{1j}(\tau, t)| \le \mathbf{E}R^*_{\Pi(t),j}$. Тогда из (4) и того, что $\mathbf{E}R^*_{\Pi(nt),j} \sim ntL(n)$ при $n \to \infty$ ([13], формула (23)), получаем

$$|c^*_{1j}(\tau, t)| = \lim_{n\to\infty}\frac{|\widetilde{c}_{1j}(n\tau, nt)|}{n\sqrt{L^*(n)L(n)}} \le \lim_{n\to\infty}\sqrt{\frac{L(n)}{L^*(n)}} = 0.$$

Вычислим ковариацию $R^*_{\Pi(\tau),i}$ и $R^*_{\Pi(t),j}$ при $\tau \le t$ и $i, j \ge 2$:

$$\widetilde{c}_{ij}(\tau, t) = \mathbf{cov}(R^*_{\Pi(\tau),i}, R^*_{\Pi(t),j})$$

$$= \sum_{k=1}^{\infty}\Big(\mathbf{P}(\Pi_k(\tau) \ge i, \Pi_k(t) \ge j) - \mathbf{P}(\Pi_k(\tau) \ge i)\mathbf{P}(\Pi_k(t) \ge j)\Big)$$

$$= \sum_{k=1}^{\infty}\Big(\mathbf{P}(\Pi_k(\tau) < i,\ \Pi_k(t) < j) - \mathbf{P}(\Pi_k(\tau) < i)\mathbf{P}(\Pi_k(t) < j)\Big).$$

Если $i < j$, то

$$\widetilde{c}_{ij}(\tau, t) = \sum_{k=1}^{\infty}\sum_{s=0}^{i-1}\mathbf{P}(\Pi(\tau p_k) = s)\Big(\mathbf{P}(\Pi((t-\tau)p_k) < j - s) - \mathbf{P}(\Pi(tp_k) < j)\Big)$$

$$= \sum_{k=1}^{\infty}\sum_{s=0}^{i-1}\frac{(\tau p_k)^s}{s!}e^{-\tau p_k}\left(\sum_{m=0}^{j-s-1}\frac{((t-\tau)p_k)^m}{m!}e^{-(t-\tau)p_k} - \sum_{m=0}^{j-1}\frac{(tp_k)^m}{m!}e^{-tp_k}\right)$$

$$= \int_0^\infty \sum_{s=0}^{i-1} \frac{\tau^s x^{-s}}{s!} e^{-\tau/x} \left(\sum_{m=0}^{j-s-1} \frac{(t-\tau)^m x^{-m}}{m!} e^{-(t-\tau)/x} - \sum_{m=0}^{j-1} \frac{t^m x^{-m}}{m!} e^{-t/x} \right) d\alpha(x)$$

$$= \int_0^\infty \sum_{s=0}^{i-1} \frac{\tau^s}{s!} \left(\sum_{m=0}^{j-s-1} \frac{(t-\tau)^m x^{-m-s}}{m!} e^{-t/x} - \sum_{m=0}^{j-1} \frac{t^m x^{-m-s}}{m!} e^{-(t+\tau)/x} \right) d\alpha(x).$$

Проинтегрируем по частям и разобьем на два интеграла:

$$\widetilde{c}_{ij}(\tau,t) = \int_0^\infty \sum_{s=0}^{i-1} \frac{\tau^s}{s!} \sum_{m=0}^{j-s-1} \frac{(t-\tau)^m((m+s)x^{-m-s-1} - tx^{-m-s-2})}{m!} e^{-t/x}\alpha(x)dx$$

$$- \int_0^\infty \sum_{s=0}^{i-1} \frac{\tau^s}{s!} \sum_{m=0}^{j-1} \frac{t^m((m+s)x^{-m-s-1} - (t+\tau)x^{-m-s-2})}{m!} e^{-(t+\tau)/x}\alpha(x)dx.$$

В первом интеграле сделаем замену $y = x/t$, во втором $y = x/(t+\tau)$:

$$\widetilde{c}_{ij}(\tau,t) = \sum_{s=0}^{i-1} \sum_{m=0}^{j-s-1} \frac{\tau^s (t-\tau)^m t^{1-m-s}}{s!m!} \int_0^\infty ((m+s)y^{-m-s} - y^{-m-s-1}) e^{-1/y} L(ty)dy$$

$$- \sum_{s=0}^{i-1} \sum_{m=0}^{j-1} \frac{\tau^s t^m (t+\tau)^{1-m-s}}{s!m!} \int_0^\infty ((m+s)y^{-m-s} - y^{-m-s-1}) e^{-1/y} L((t+\tau)y)dy.$$

Разобьем двойные суммы на слагаемые, в которых $m+s \le 1$ и $m+s \ge 2$. Заметим, что для любого целого $r \ge 2$ при $t \to \infty$

$$\int_0^\infty y^{-r} e^{-1/y} L(ty)dy \sim L(t) \int_0^\infty y^{-r} e^{-1/y} dy = L(t)(r-2)!,$$

$$\int_0^\infty (ry^{-r} - y^{-r-1}) e^{-1/y} dy = (r-2)!.$$

Тогда (так как $\frac{\alpha(nt)}{\alpha(n)} \to t$ при $n \to \infty$)

$$c^*_{ij}(\tau,t) = \lim_{n\to\infty} \frac{\widetilde{c}_{ij}(n\tau, nt)}{\alpha(n)}$$

$$= \tau + \sum_{s=0}^{i-1} \sum_{m=0}^{j-s-1} \tau^s (t-\tau)^m t^{1-m-s} K_{m,s} - \sum_{s=0}^{i-1} \sum_{m=0}^{j-1} \tau^s t^m (t+\tau)^{1-m-s} K_{m,s}.$$

Аналогично при $i \ge j$

$$\widetilde{c}_{ij}(\tau,t) = \sum_{k=1}^{\infty} \Big(\mathbf{P}(\Pi_k(t) < j) - \mathbf{P}(\Pi_k(\tau) < i)\mathbf{P}(\Pi_k(t) < j) \Big)$$

$$= \int_0^\infty \sum_{m=0}^{j-1} \frac{t^m x^{-m}}{m!} e^{-t/x} \left(1 - \sum_{s=0}^{i-1} \frac{\tau^s x^{-s}}{s!} e^{-\tau/x} \right) d\alpha(x)$$

$$= \sum_{m=0}^{j-1} \left(\frac{t}{m!} \int_0^\infty (my^{-m} - y^{-m-1}) e^{-1/y} L(ty)dy \right.$$

$$\left. - \sum_{s=0}^{i-1} \frac{t^m \tau^s (t+\tau)^{1-m-s}}{s!m!} \int_0^\infty ((m+s)y^{-m-s} - y^{-m-s-1}) e^{-1/y} L((t+\tau)y)dy \right),$$

$$c^*_{ij}(\tau,t) = \tau + t\sum_{m=0}^{j-1} K_{m,0} - \sum_{s=0}^{i-1} \sum_{m=0}^{j-1} \tau^s t^m (t+\tau)^{1-m-s} K_{m,s}.$$

Шаг 2 (сходимость конечномерных распределений)

По аналогии с доказательством теоремы 1 (см. [10]), можно показать, что для любых $m \geq 1, \quad 0 < t_1 < t_2 < \ldots < t_m \leq 1$ треугольный массив $m\nu$-мерных случайных векторов (независимых по k для каждого n)

$$\left\{ \frac{\mathbf{I}(\Pi_k(nt_j) \geq i) - \mathbf{P}(\Pi_k(nt_j) \geq i)}{\sqrt{\alpha(n)}}, \ i \leq \nu, \ j \leq m, \ k \leq n \right\}_{n \geq 1}$$

удовлетворяет условию Линдеберга (см. [6], теорема 6.2, стр. 204).

Шаг 3 (относительная компактность)

Определим для любых положительных чисел $\tau_1 \leq \tau_2$ при $k \geq 2$ (относительная компактность при $k = 1$ доказана в работе [8])

$$R^*_{\Pi(\tau_2),k} - R^*_{\Pi(\tau_1),k} = \sum_{i=1}^{\infty} \mathbf{I}(\Pi_i(\tau_2) \geq k, \ \Pi_i(\tau_1) < k) \stackrel{def}{=} \sum_{i=1}^{\infty} \mathbf{I}_i(\tau_1, \tau_2) = \sum_{i=1}^{\infty} \mathbf{I}_i,$$

$$P_i = P_i(\tau_1, \tau_2) \stackrel{def}{=} \mathbf{P}(\mathbf{I}_i).$$

Далее мы будем использовать обозначения $\mathbf{I}_i$ и, соответственно, P_i для различных значений положительных чисел $\tau_1 < \tau_2$. Также далее нам потребуется новый процесс

$$Z^{**}_{n,k}(t) = \frac{R^*_{\Pi([nt]),k} - \mathbf{E}R^*_{\Pi([nt]),k}}{(\alpha(n))^{1/2}}.$$

Будем придерживаться следующего плана:
(**a**) докажем непрерывность предельного процесса;
(**b**) докажем, что $Z^*_{n,k}$ и $Z^{**}_{n,k}$ достаточно «близки»;
(**c**) докажем относительную компактность $Z^{**}_{n,k}$.

(**a**) Положим $\tau_1 = nt_1, \ \tau_2 = nt_2$ для $0 < t_1 < t_2 < 0$, тогда

$$\mathbf{E}(Z^*_{n,k}(t_2) - Z^*_{n,k}(t_1))^2 = \sum_{i=1}^{\infty} \mathbf{E}(\mathbf{I}_i - P_i)^2/\alpha(n)$$

$$\leq \sum_{i=1}^{\infty} P_i/\alpha(n) = \mathbf{E}(R^*_{\Pi(\tau_2),k} - R^*_{\Pi(\tau_1),k})/\alpha(n) \leq C(t_2 - t_1)^{1/2}.$$

Выше мы воспользовались независимостью слагаемых, а также фактом, что дисперсия индикатора меньше, чем его математическое ожидание, и леммой 3.

Используя вычисленные на шаге 1 ковариации и теорему 1.4 в [1], мы доказываем, что k-я компонента предельного гауссовского процесса почти наверное имеет непрерывную модификацию на $[0, 1]$.

Так как траектории предельного гауссовского процесса принадлежат п. н. классу $C((0, 1)^\nu)$, то (см. [5]) слабая сходимость в топологии Скорохода влечет слабую сходимость в равномерной метрике. Следовательно, нам достаточно доказать относительную компактность $\{Z^*_{n,k}\}_{n \geq n_0}$ (где n_0 — число, определенное в лемме 3) в топологии Скорохода.

(**b**) Так как с вероятностью 1

$$R^*_{\Pi(nt),k} - R^*_{\Pi([nt]),k} \leq \Pi(nt) - \Pi([nt]) \leq \Pi([nt] + 1) - \Pi([nt]),$$

то

$$\mathbf{E}(R^*_{\Pi(nt),k} - R^*_{\Pi([nt]),k}) \leq \mathbf{E}(\Pi([nt] + 1) - \Pi([nt])) = 1.$$

Следовательно, для любого $\eta > 0$

$$\mathbf{P}(\sup_{0\le t\le 1} |Z^*_{n,k}(t) - Z^{**}_{n,k}(t)| > \eta)$$

$$\le \mathbf{P}(\sup_{0\le t\le 1} (R^*_{\Pi(nt),k} - R^*_{\Pi([nt]),k} + \mathbf{E}(R^*_{\Pi(nt),k} - R^*_{\Pi([nt]),k})) > \eta\sqrt{\alpha(n)})$$

$$\le \mathbf{P}(\sup_{0\le t\le 1} (\Pi([nt]+1) - \Pi([nt]) + 1) > \eta\sqrt{\alpha(n)})$$

$$= \mathbf{P}(\sup_{0\le m\le n} (\Pi(m+1) - \Pi(m) + 1) > \eta\sqrt{\alpha(n)})$$

$$\le \sum_{m=0}^{n} \mathbf{P}(\Pi(m+1) - \Pi(m) + 1 > \eta\sqrt{\alpha(n)})$$

$$\le \sum_{m=0}^{n} \frac{\mathbf{E}e^{\Pi(m+1)-\Pi(m)+1}}{e^{\eta\sqrt{\alpha(n)}}} = (n+1)\frac{\mathbf{E}e^{\Pi(1)}}{e^{\eta\sqrt{\alpha(n)}-1}} = (n+1)e^{e-\eta\sqrt{\alpha(n)}} \to 0$$

при $n \to \infty$.

Поэтому нам достаточно доказать относительную компактность $\{Z^{**}_{n,k}\}_{n\ge n_0}$ (где n_0 — число, определенное в лемме 3) в топологии Скорохода.

(**c**) Пусть $t_1,\ t_2 \in [0,1]$ и $t_2 - t_1 \ge \frac{1}{2n}$, тогда

$$[nt_2] - [nt_1] \le n(t_2 - t_1) + 1 \le 3n(t_2 - t_1).$$

Положим $\tau_1 = [nt_1],\ \tau_2 = [nt_2]$.

Используя независимость слагаемых и неравенство Розенталя, получаем, что для всех $n \ge n_0$

$$\mathbf{E}|Z^{**}_{n,k}(t_2) - Z^{**}_{n,k}(t_1)|^{18} = \frac{\mathbf{E}\left|\sum\limits_{i=1}^{\infty} \mathbf{I}_i - P_i\right|^{18}}{(\alpha(n))^9}$$

$$\le \frac{c}{(\alpha(n))^9}\left(\sum_{i=1}^{\infty} \mathbf{E}|\mathbf{I}_i - P_i|^{18} + \left(\sum_{i=1}^{\infty} \mathbf{E}(\mathbf{I}_i - P_i)^2\right)^9\right)$$

$$\le \frac{c}{(\alpha(n))^9}\left(\sum_{i=1}^{\infty} P_i + \left(\sum_{i=1}^{\infty} P_i\right)^9\right)$$

$$= \frac{c}{(\alpha(n))^9}\left(\mathbf{E}(R^*_{\Pi([nt_2]),k} - R^*_{\Pi([nt_1]),k}) + \left(\mathbf{E}(R^*_{\Pi([nt_2]),k} - R^*_{\Pi([nt_1]),k})\right)^9\right)$$

$$\le c\left(\frac{24n^4(t_2-t_1)^4}{(\alpha(n))^9} + C^9\left(\frac{[nt_2]-[nt_1]}{n}\right)^{9/2}\right) \le \widetilde{C}(t_2 - t_1)^4,$$

где c и $\widetilde{C}$ абсолютные постоянные.

Выше мы использовали факт, что дисперсия индикатора меньше, чем его математическое ожидание, а также неравенство

$$\mathbf{E}(R^*_{\Pi([nt_2]),k} - R^*_{\Pi([nt_1]),k}) \le \mathbf{E}(\Pi([nt_2]) - \Pi([nt_1])) = [nt_2] - [nt_1]$$

$$\le 3n(t_2 - t_1) \le 3n(t_2 - t_1)\cdot(2n(t_2 - t_1))^3 = 24n^4(t_2 - t_1)^4,$$

лемму 1 и лемму 3.

Если $0 \le t_2 - t_1 < \frac{1}{n}$, то $[nt_1] = [nt]$ или $[nt_2] = [nt]$ для любого $t \in [t_1, t_2]$. Таким образом

$$Q \stackrel{def}{=} \mathbf{E}(|Z^{**}_{n,k}(t) - Z^{**}_{n,k}(t_1)|^9 |Z^{**}_{n,k}(t_2) - Z^{**}_{n,k}(t)|^9) = 0 \le (t_2 - t_1)^2.$$

Если $t_2 - t_1 \ge 1/n$, то возможны следующие 3 ситуации:

1) если $t_2 - t \ge \frac{1}{2n}$, $t - t_1 \ge \frac{1}{2n}$, то с помошью неравенства Коши—Буняковского получаем, что

$$Q \le \widetilde{C}(t_2 - t)^2 \cdot (t - t_1)^2 \le \widetilde{C}(t_2 - t_1)^2;$$

2) если $t_2 - t \ge \frac{1}{2n}$, $t - t_1 < \frac{1}{2n}$, то так как

$$R_{\Pi([nt]),k)} - R_{\Pi([nt_1]),k)} \le_{\text{п. н.}} \Pi([nt]) - \Pi([nt_1]) \le_{st} \Pi(1),$$

неравенство Коши—Буняковского влечет оценку

$$Q \le \left(\widetilde{C}(t_2 - t)^4 \cdot \mathbf{E} \left(\frac{\Pi(1) + 1}{\sqrt{\alpha(n)}} \right)^{18} \right)^{1/2} \le \widehat{C}(t_2 - t_1)^2;$$

3) если $t_2 - t < \frac{1}{2n}$, $t - t_1 \ge \frac{1}{2n}$, то так как

$$R_{\Pi([nt_2]),k)} - R_{\Pi([nt]),k)} \le_{\text{п. н.}} \Pi([nt_2]) - \Pi([nt]) \le_{st} \Pi(1),$$

неравенство Коши—Буняковского влечет оценку

$$Q \le \left(\mathbf{E} \left(\frac{\Pi(1) + 1}{\sqrt{\alpha(n)}} \right)^{18} \cdot \widetilde{C}(t - t_1)^4 \right)^{1/2} \le \widehat{C}(t_2 - t_1)^2.$$

Из доказанного следует (см. [5], теорема 13.5) плотность семейства (относительная компактность) k-й компоненты, а следовательно, и всего случайного вектора.

Шаг 4 (приближение исходного процесса)

В силу монотонности $\Pi(t)$ и УЗБЧ известно (см. лемму 2), что для каждой пары чисел ε, $\delta \in (0, 1)$ существует целое $N = N(\varepsilon, \delta)$ такое, что при $n \ge N$

$$\mathbf{P}(\forall t \in [0, 1] \;\; \exists \tau : |\tau - t| \le \delta, \; \Pi(n\tau) = [nt]) \stackrel{def}{=} \mathbf{P}(A(n)) \ge 1 - \varepsilon.$$

Из относительной компактности распределений процессов

$$\{Z^*_{n,k}\}_{n \ge n_0, k \ge 1}$$

получаем, что для каждой пары чисел $\varepsilon \in (0, 1)$, $\eta > 0$ существуют $\delta \in (0, 1)$ и целое $N_1 = N_1(\varepsilon, \eta)$ такие, что для всех $n \ge N_1$

$$\mathbf{P}(\sup_{|t-\tau| \le \delta} \left| Z^*_{n,k}(\tau) - Z^*_{n,k}(t) \right| \ge \eta) \le \varepsilon.$$

Следовательно, так как

$$\mathbf{P}(Y^*_{n,k}(t) = Z^*_{n,k}(\tau) | \Pi(n\tau) = [nt]) = 1,$$

то для всех $n \ge \max(N, N_1)$

$$\mathbf{P} \left(\sup_{0 \le t \le 1} \left| Y^*_{n,k}(t) - Z^*_{n,k}(t) \right| \ge \eta \right) \le \mathbf{P} \left(\sup_{0 \le t \le 1} \left| Y^*_{n,k}(t) - Z^*_{n,k}(t) \right| \ge \eta, A(n) \right) + \varepsilon$$

$$\le \mathbf{P} \left(\sup_{|t-\tau| \le \delta} \left| Z^*_{n,k}(\tau) - Z^*_{n,k}(t) \right| \ge \eta \right) + \varepsilon \le 2\varepsilon.$$

Автор благодарит А. П. Ковалевского за конструктивные комментарии и предложения.

References

[1] R.J. Adler, *An introduction to continuity, extrema, and related topics for general Gaussian processes,* Institute of Math. Stat., Hayward, California, 1990. MR1088478

[2] R. R. Bahadur, *On the number of distinct values in a large sample from an infinite discrete distribution,* Proceedings of the National Institute of Sciences of India, **26A**:2 (1960), 67–75. MR0137256

[3] A. D. Barbour, *Univariate approximations in the infinite occupancy scheme,* ALEA Lat. Am. J. Probab. Math. Stat., **6** (2009), 415–433. MR2576025

[4] A. D. Barbour, A. V. Gnedin, *Small counts in the infinite occupancy scheme,* Electronic Journal of Probability, **14**:13 (2009), 365–384. MR2480545

[5] P. Billingsley, *Convergence of Probability Measures,* Second Edition, New York: John Wiley & Sons, Inc., 1999. MR1700749

[6] A. A. Borovkov, *Probability Theory,* London: Universitext, Springer, 2013. MR3086572

[7] M. G. Chebunin, *Estimation of parameters of probabilistic models which is based on the number of different elements in a sample,* Sib. Zh. Ind. Mat., **17**:3 (2014), 135–147. MR3364413

[8] M. Chebunin, A. Kovalevskii, *Functional central limit theorems for certain statistics in an infinite urn scheme,* Statistics and Probability Letters, **119** (2016), 344–348. MR3555307

[9] O. Durieu, Y. Wang, *From infinite urn schemes to decompositions of self-similar Gaussian processes,* Electronic Journal of Probability, **21**:43 (2016), 1–23. MR3530320

[10] M. Dutko, *Central limit theorems for infinite urn models,* Ann. Probab., **17** (1989), 1255–1263. MR1009456

[11] A. Gnedin, B. Hansen, J. Pitman, *Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws,* Probability Surveys, **4** (2007), 146–171. MR2318403

[12] H.-K. Hwang, S. Janson, *Local Limit Theorems for Finite and Infinite Urn Models,* The Annals of Probability, **36**:3 (2008), 992–1022. MR2408581

[13] S. Karlin, *Central Limit Theorems for Certain Infinite Urn Schemes,* Jounal of Mathematics and Mechanics, **17**:4 (1967), 373–401. MR0216548

[14] E. S. Key, *Rare Numbers,* Journal of Theoretical Probability, **5**:2 (1992), 375–389. MR1157991

[15] E. S. Key, *Divergence rates for the number of rare numbers,* Journal of Theoretical Probability, **9**:2 (1996), 413–428. MR1385405

[16] N. S. Zakrevskaya, A. P. Kovalevskii, *One-parameter probabilistic models of text statistics,* Sib. Zh. Ind. Mat., **4**:2 (2001), 142–153. MR1965927

Mikhail Georgievich Chebunin
Novosibirsk State University,
str. Pirogova, 2,
630090, Novosibirsk, Russia
E-mail address: chebuninmikhail@gmail.com