S©MR ISSN 1813-3304

СИБИРСКИЕ ЭЛЕКТРОННЫЕ
МАТЕМАТИЧЕСКИЕ ИЗВЕСТИЯ

Siberian Electronic Mathematical Reports

http://semr.math.nsc.ru

Том 13, стр. 305–317 (2016) УДК 519.214.5
DOI 10.17377/semi.2016.13.025 MSC 60F05

О ЧИСЛЕ СОВПАДЕНИЙ ЗНАКОВ В ДИСКРЕТНОЙ СЛУЧАЙНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ, УПРАВЛЯЕМОЙ ЦЕПЬЮ МАРКОВА

Н.М. МЕЖЕННАЯ

Abstract. The present paper is devoted to studying the properties of the number of pairs of characters matchings in discrete random sequence controlled by Markov chain with finite number of states. Alongside with the random variable we also consider the number of characters which appears in a random sequence exactly k times, $k = 0, 1, 2$. We derive the estimators for total variation distance between distributions of the considered random variables and Poisson distribution. This results allow to develop Poisson and normal limit theorems.

Keywords: pairs of characters matchings, placing of particles, Poisson limit theorem, normal limit theorem

1. Введение

Пусть $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_T, \ldots)$ — однородная непериодическая неразложимая цепь Маркова с множеством состояний $E_M = \{1, \ldots, M\}$,

$$\pi_k(i) = \mathbf{P}\{Z_i = k\}, \pi_{kl}^{(m)} = \mathbf{P}\{Z_{t+m} = l | Z_t = k\},\ k, l \in E_M,$$

на множестве $A_N = \{1, \ldots, N\}$ заданы M вероятностных распределений $\{p_a^{(j)}\}$, $a \in A_N, j = 1, \ldots, M$.

Mezhennaya, N.M., On the number of characters matchings in discrete random sequence controlled by Markov chain.

Работа выполнена при финансовой поддержке Министерства образования и науки РФ (тема 1.2640.2014).

Поступила 27 января 2016 г., опубликована 6 мая 2016 г.

Известно (см. [1]), что существует единственное стационарное распределение $\tilde{\pi}_k = \lim_{n\to\infty} \pi_k(n)$, $k = 1, \ldots, M$, и константы $C, a > 0$, при которых

$$(1) \qquad \max_{k,l\in E_M} |\pi_{lk}^{(n)} - \tilde{\pi}_k| \le C\tilde{\pi}_k e^{-\alpha n}, \ \max_{k,l\in E_M} |\pi_k(n) - \tilde{\pi}_k| \le C\tilde{\pi}_k e^{-\alpha n}.$$

Рассмотрим последовательность случайных величин $X_1, \ldots, X_T, \ldots$, принимающих значения из множества A_N с вероятностями

$$\mathbf{P}\{X_j = k\} = p_k^{(Z_j)}, \quad k \in A_N, j \in \mathbb{N}.$$

Пусть

$$(2) \qquad \xi_r = \sum_{1\le i_1<\ldots<i_r\le T} I\{X_{i_1} = \ldots = X_{i_r}\}$$

— число r-кратных совпадений знаков в последовательности $X_1, \ldots, X_T$, $r \le T$, μ_r — число знаков из A_N, которые появились в последовательности $X_1, \ldots, X_T$ ровно r раз, $r \in \{0, 1, \ldots, T\}$. Случайные величины ξ_r и μ_r допускают интерпретацию в терминах задачи о размещении частиц (см. [2]). А именно, событие $\{X_j = k\}, k \in A_N$, будем трактовать как попадание j-й частицы в ячейку с номером k. Тогда случайная величина μ_r равна числу ячеек, содержащих ровно r частиц после размещается T частиц по N ячейкам. В отличие от классической схемы размещения вероятности попадания $\{p_a^{(Z_j)}\}$ в ячейки сами являются случайными величинами. В главе 3 книги [2] приведены предельные теоремы для μ_r для равновероятной и неравновероятной схем размещения частиц по ячейкам. В работе [3] при помощи метода Стейна была доказана центральная предельная теорема для числа ячеек, содержащих $d \ge 2$ частиц, в равновероятной схеме размещения частиц по ячейкам с оценкой скорости сходимости. Оценки расстояния по вариации для распределений случайных величин ξ_r и μ_r при размещении частиц комплектами до сопровождающего пуассоновского распределения были получены в работе [4]. Асимптотические свойства распределений указанных случайных величин также изучались в работах [5]—[7], [21].

Распределения величин ξ_r и μ_r в последовательности, образующей цепь Маркова, изучались в работах [8], [9]. В работе [8] был получен предельный закон распределения для μ_r. Более общая постановка задачи о распределении случайных величин, являющихся линейными комбинациями μ_r, в марковской цепи была рассмотрена в работе [9]. Для них были получены оценки скорости сходимости к предельным распределениям в равномерной метрике.

Частным случаем марковской схемы размещения является схема, в которой роль частиц выполняют цепочки заданной длины в последовательности независимых или связанных в цепь Маркова испытаний. Распределение повторений цепочек, в том числе многократных, в последовательности независимых случайных величин было изучено в работах [10]—[15] , в цепи Маркова — в работах [16]—[18]. Доказаны предельные теоремы для числа длинных (в том числе многократных) повторений цепочек в эргодической цепи Маркова.

В настоящей работе изучаются асимптотические свойства случайных величин $\xi_2, \mu_0, \mu_1, \mu_2$ в неоднородной полиномиальной схеме с большим числом исходов, управляемой цепью Маркова с конечным числом состояний. Такая последовательность может трактоваться как последовательность, полученная укрупнением состояний марковской цепи, и как скрытая марковская цепь (см. [19], [20]).

2. Предельные теоремы

Начнем с изучения свойств случайной величины ξ_2. Ее математическое ожидание определяется формулой $\mathbf{E}\xi_2 = \mathbf{E}\lambda_2(\mathbf{Z})$, где

$$\lambda_2(\mathbf{Z}) = \mathbf{E}(\xi_2|\mathbf{Z}) = \sum_{1\le i<j\le T} \mathbf{P}\{X_i = X_j|\mathbf{Z}\} =$$

$$= \sum_{1\le i<j\le T} \sum_{a\in A_N} \mathbf{P}\{X_i = X_j = a|\mathbf{Z}\} = \sum_{1\le i<j\le T} \sum_{a\in A_N} p_a^{(Z_i)} p_a^{(Z_j)}. \tag{3}$$

Лемма 1. *Пусть $T \to \infty$. Тогда*

$$\mathbf{E}\xi_2 = \lambda_2 \left(1 + O(T^{-1})\right),$$

где

$$\lambda_2 = \frac{T^2}{2} \sum_{k,l\in E_M} \sum_{a\in A_N} p_a^{(k)} p_a^{(l)} \tilde{\pi}_k \tilde{\pi}_l. \tag{4}$$

Перейдем к задаче о поведении распределения случайной величины ξ_2 при фиксированном M и $T \to \infty$.

Будем использовать обозначения $\mathcal{L}(X)$ для закона распределения случайной величины X и $\mathrm{Pois}(\mu)$ для распределения Пуассона с параметром μ. Напомним, что случайная величина X имеет смешанное распределение Пуассона с дискретным случайным параметром Λ, принимающем значения в множестве $\mathcal{E}$, если

$$\mathbf{P}\{X = n\} = \sum_{\lambda\in\mathcal{E}} \frac{\lambda^n}{n!} e^{-\lambda} \mathbf{P}\{\Lambda = \lambda\}.$$

Расстояние по вариации ρ_{TV} между распределениями случайных величин η_1 и η_2, принимающих значения в множестве неотрицательных целых чисел, выражается формулой

$$\rho_{TV}(\mathcal{L}(\eta)_1, \mathcal{L}(\eta_2)) = \frac{1}{2} \sum_{k=0}^{\infty} |\mathbf{P}\{\eta_1 = k\} - \mathbf{P}\{\eta_2 = k\}|.$$

Теорема 1. *Пусть число M фиксировано, $T \to \infty$. Тогда*

$$\rho_{TV}(\mathcal{L}(\xi_2), Pois(\lambda_2)) = O(Tp^*), \tag{5}$$

где $p^* = \max\limits_{j=1,\dots,T} p^{(*j)}$, $p^{(*j)} = \max\limits_{k\in A_N} p_k^{(j)}$.

Теорема 2. *Пусть выполнены условия теоремы 1. Тогда*

$$\rho_{TV}(\mathcal{L}(\mu_2), Pois(\lambda_2)) = O(Tp^*),$$

$$\rho_{TV}(\mathcal{L}(T - N + \mu_0), Pois(\lambda_2)) = O(Tp^*),$$

$$\rho_{TV}\left(\mathcal{L}\left(\frac{T-\mu_1}{2}\right), Pois(\lambda_2)\right) = O(Tp^*\lambda_2).$$

Следствие 1. *Пусть число M фиксировано, $T \to \infty, p^* \to 0$ так, что $\lambda_2 \to a > 0$ и $Tp^* \to 0$. Тогда распределения случайных величин ξ_2, μ_2, $\frac{T-\mu_1}{2}$, $T - N + \mu_0$ сходятся к $Pois(a)$.*

Замечание. В случае, если $M = 1$, утверждение следствия 1 совпадает с теоремой 2 §3 главы 3 книги [2] для рассматриваемых случайных величин в полиномиальной схеме размещения.

Следствие 2. *Пусть число M фиксировано, $T \to \infty$, $p^* \to 0$ так, что $\lambda_2 \to \infty$ и $Tp^*\lambda_2 \to 0$. Тогда распределения случайных величин $\frac{\xi_2-\lambda_2}{\sqrt{\mathbf{E}\xi_2}}$, $\frac{\mu_2-\lambda_2}{\sqrt{\mathbf{E}\xi_2}}$, $\frac{T-\mu_1-2\lambda_2}{2\sqrt{\mathbf{E}\xi_2}}$, $\frac{T-N+\mu_0-\lambda_2}{\sqrt{\mathbf{E}\xi_2}}$ сходятся к стандартному нормальному закону.*

Следствия 1 и 2 очевидным образом вытекают из оценок теорем 1 и 2. Поэтому их доказательства мы не приводим.

3. Доказательства теорем

Доказательство теоремы 1. Воспользуемся неравенством треугольника

$$\rho_{TV}(\mathcal{L}(\xi_2), \mathrm{Pois}(\lambda_2)) \leq \rho_{TV}\left(\mathcal{L}(\xi_2), \mathrm{Pois}(\lambda_2(\mathbf{Z}))\right) + \rho_{TV}\left(\mathrm{Pois}(\lambda_2(\mathbf{Z})), \mathrm{Pois}(\mathbf{E}\xi_2)\right)$$

$$+\rho_{TV}\left(\mathrm{Pois}(\mathbf{E}\xi_2), \mathrm{Pois}(\lambda_2)\right). \tag{6}$$

Для оценки слагаемых в (6) нам понадобятся два вспомогательных утверждения, доказательство которых будет приведено в заключительном разделе статьи.

Лемма 2. *Пусть $T \to \infty$. Тогда*

$$\rho_{TV}\left(Pois(\lambda_2(\mathbf{Z})), Pois(\lambda_2)\right) = O(p^*MT). \tag{7}$$

Рассмотрим вспомогательную постановку задачи. Пусть случайные величины $Y_1, \ldots, Y_T, \ldots$ независимы и принимают значения из множества $A_N = \{1, \ldots, N\}$, причем

$$\mathbf{P}\{Y_j = k\} = p_k^{(j)}, \quad k \in A_N, j \in \mathbb{N}, \tag{8}$$

а наборы $\{p_k^{(j)}\}$ при каждом j удовлетворяют условию $\sum\limits_{k \in A_N} p_k^{(j)} = 1$. Такую последовательность можно рассматривать как последовательность $X_1, \ldots, X_T, \ldots$, описанную в начале работы, при фиксированной последовательности $\mathbf{Z}$.

Определим случайную величину

$$\varsigma_2 = \sum_{1 \leq i < j \leq T} I\{Y_i = Y_j\}$$

— число пар совпавших знаков в последовательности $Y_1, \ldots, Y_T$. Ее математическое ожидание определяется формулой

$$\mathbf{E}\varsigma_2 = \sum_{1 \leq i < j \leq T} \mathbf{P}\{Y_i = Y_j\} = \sum_{1 \leq i < j \leq T} \sum_{a \in A_N} \mathbf{P}\{Y_i = Y_j = a\} = \sum_{1 \leq i < j \leq T} \sum_{a \in A_N} p_a^{(i)} p_a^{(j)}.$$

Лемма 3. *При $T \geq 2$ выполнена оценка*

$$\rho_{TV}(\mathcal{L}(\varsigma_2), Pois(\mathbf{E}\varsigma_2)) \leq 4Tp^*. \tag{9}$$

Вернемся к доказательству теоремы 1. Из формулы (9) следует, что

$$\rho_{TV}\left(\mathcal{L}(\xi_2), \mathrm{Pois}(\lambda_2(\mathbf{Z}))\right) \leq 4Tp^*. \tag{10}$$

Для оценки последнего слагаемого в (6), воспользуемся теоремой 1.C книги [21], согласно которой

$$\rho_{TV}(\mathrm{Pois}(\mathbf{E}\xi_2), \mathrm{Pois}(\lambda_2)) \le \min\left\{1, \frac{1}{\sqrt{\lambda_2}}\right\} |\mathbf{E}\xi_2 - \lambda_2|.$$

Так как

$$\lambda_2 \le \frac{T^2}{2} p^* \sum_{k,l \in E_M} \sum_{a \in A_N} p_a^{(l)} \tilde{\pi}_k \tilde{\pi}_l = \frac{T^2}{2} p^*,$$

то

$$\mathbf{E}\xi_2 - \lambda_2 = O(T^{-1}\lambda_2) = O(Tp^*),$$

$$\rho_{TV}(\mathcal{L}(\xi_2), \mathrm{Pois}(\lambda_2)) = O(Tp^*). \tag{11}$$

Подставляя (7), (10) и (11) в правую часть неравенства (6), получаем оценку (5). Теорема 1 доказана. □

Нам понадобится еще одно вспомогательное утверждение.

Лемма 4. *При $T \ge 3$ выполнена оценка*

$$\mathbf{E}\xi_3 \le 2Tp^*\lambda_2. \tag{12}$$

Доказательство теоремы 2. Согласно определениям

$$\sum_{r=0}^{T} \mu_r = N, \quad \sum_{r=1}^{T} r\mu_r = T. \tag{13}$$

Случайные величины ξ_2 и ξ_3 могут быть выражены через $\mu_2, \ldots, \mu_T$ (см. [4]), а именно,

$$\xi_2 = \mu_2 + C_3^2\mu_3 + C_4^2\mu_4 + \ldots = \sum_{k=2}^{T} C_k^2 \mu_k,$$

$$\xi_3 = \mu_3 + C_4^3\mu_4 + \ldots = \sum_{k=3}^{T} C_k^3 \mu_k.$$

Эти соотношения и лемма 4 позволяют найти предельные законы распределения для случайных величин μ_0, μ_1, μ_2. Заметим сначала, что

$$\xi_2 = \mu_2 + \sum_{k=3}^{T} C_k^2 \mu_k = \mu_2 + \sum_{k=3}^{T} \frac{3}{k-2} C_k^2 \mu_k \le \mu_2 + 3\sum_{k=3}^{T} C_k^3 \mu_k = \mu_2 + 3\xi_3,$$

значит,

$$0 \le \xi_2 - \mu_2 \le 3\xi_3.$$

Из последнего соотношения и (12) получаем

$$\rho_{TV}(\mathcal{L}(\xi_2), \mathcal{L}(\mu_2)) \le \mathbf{P}\{\xi_3 > 0\} \le \mathbf{E}\xi_3 \le 2Tp^*\lambda_2.$$

Тогда

$$\rho_{TV}(\mathcal{L}(\mu_2), \mathrm{Pois}(\lambda_2)) \le \rho_{TV}(\mathcal{L}(\mu_2), \mathcal{L}(\xi_2)) + \rho_{TV}(\mathcal{L}(\xi_2), \mathrm{Pois}(\lambda_2)) = O(Tp^*\lambda_2). \tag{14}$$

Теперь рассмотрим случайную величину μ_1. Из второго равенства (13) имеем

$$\frac{T-\mu_1}{2} - \mu_2 = \frac{1}{2}\sum_{k=3}^{T} k\mu_k = \sum_{k=3}^{T} \frac{3}{(k-1)(k-2)} C_k^3 \mu_k \le \frac{3}{2}\sum_{k=3}^{T} C_k^3 \mu_k = \frac{3}{2}\xi_3.$$

Аналогично (14) получаем

$$\rho_{TV}\left(\mathcal{L}\left(\frac{T-\mu_1}{2}\right), \mathrm{Pois}(\lambda_2)\right) \le \rho_{TV}\left(\mathcal{L}\left(\frac{T-\mu_1}{2}\right), \mathcal{L}(\mu_2)\right)$$

$$+\rho_{TV}(\mathcal{L}(\mu_2), \mathcal{L}(\xi_2)) = O(Tp^*\lambda_2). \tag{15}$$

Из формул (13) следует, что

$$\begin{aligned} \mu_0 &= N - \mu_1 - (\mu_2 + \ldots + \mu_T) \\ &= N - (T - 2\mu_2 - 3\mu_3 - \ldots - T\mu_T) - (\mu_2 + \ldots + \mu_T) \\ &= N - T + \mu_2 + 2\mu_3 + \ldots + (T-1)\mu_T \\ &= N - T + \mu_2 + \sum_{k=3}^{T} \frac{6}{k(k-2)} C_k^3 \mu_k \le N - T + \mu_2 + 2\xi_3. \end{aligned}$$

Значит, $0 \le T - N + \mu_0 - \mu_2 \le 2\xi_3$, а следовательно,

$$\rho_{TV}(\mathcal{L}(N - T + \mu_0), \mathcal{L}(\mu_2)) \le \mathbf{P}\{\xi_3 > 0\} \le \mathbf{E}\xi_3 = 2Tp^*\lambda_2,$$

что вместе с (14) дает оценку

$$\rho_{TV}(\mathcal{L}(N - T + \mu_0), \mathrm{Pois}(\lambda_2)) = O(Tp^*\lambda_2). \tag{16}$$

Утверждение теоремы следует из оценок (14), (15) и (16). Теорема 2 доказана. □

4. Доказательства лемм

Доказательство леммы 1. Из формулы (3) следует, что

$$\begin{aligned} \mathbf{E}\xi_2 = \mathbf{E}\lambda_2(\mathbf{Z}) &= \mathbf{E} \sum_{1 \le i < j \le T} \sum_{a \in A_N} p_a^{(Z_i)} p_a^{(Z_j)} \\ &= \sum_{k,l \in E_M} \sum_{1 \le i < j \le T} \sum_{a \in A_N} p_a^{(k)} p_a^{(l)} \mathbf{P}\{Z_i = k, Z_j = l\} \\ &= \sum_{k,l \in E_M} \sum_{1 \le i < j \le T} \sum_{a \in A_N} p_a^{(k)} p_a^{(l)} \mathbf{P}\{Z_i = k\} \mathbf{P}\{Z_j = l | Z_i = k\} \\ &= \sum_{k,l \in E_M} \sum_{1 \le i < j \le T} \sum_{a \in A_N} p_a^{(k)} p_a^{(l)} \pi_k(i) \pi_{kl}^{(j-i)} \\ &= \sum_{k,l \in E_M} \sum_{a \in A_N} p_a^{(k)} p_a^{(l)} \sum_{1 \le i < j \le T} \pi_k(i) \pi_{kl}^{(j-i)}. \end{aligned}$$

Из оценки (1) получаем, что

$$\begin{aligned} \mathbf{E}\xi_2 \le \bar{\lambda}_2 &= \sum_{k,l \in E_M} \sum_{a \in A_N} p_a^{(k)} p_a^{(l)} \sum_{1 \le i < j \le T} \tilde{\pi}_k \tilde{\pi}_l \left(1 + Ce^{-\alpha i}\right) \left(1 + Ce^{-\alpha(j-i)}\right) \\ &= \sum_{k,l \in E_M} \sum_{a \in A_N} p_a^{(k)} p_a^{(l)} \tilde{\pi}_k \tilde{\pi}_l \sum_{1 \le i < j \le T} \left(1 + Ce^{-\alpha i}\right) \left(1 + Ce^{-\alpha(j-i)}\right) \\ &= \sum_{k,l \in E_M} \sum_{a \in A_N} p_a^{(k)} p_a^{(l)} \tilde{\pi}_k \tilde{\pi}_l \sum_{1 \le i < j \le T} \left(1 + Ce^{-\alpha i} + Ce^{-\alpha(j-i)} + Ce^{-\alpha j}\right). \end{aligned}$$

Аналогично,

$$\mathbf{E}\xi_2 \geq \underline{\lambda}_2 = \sum_{k,l \in E_M} \sum_{a \in A_N} p_a^{(k)} p_a^{(l)} \tilde{\pi}_k \tilde{\pi}_l \sum_{1 \leq i < j \leq T} \Big(1 - Ce^{-\alpha i} - Ce^{-\alpha(j-i)} + Ce^{-\alpha j}\Big).$$

Так как

$$B_1 = \sum_{1 \leq i < j \leq T} e^{-\alpha j} = \frac{1}{(e^\alpha - 1)^2} \left(1 + Te^{-\alpha T}(1 - e^\alpha) - e^{-\alpha T}\right), \tag{17}$$

то

$$\left| \lambda_2 - (C_T^2 + CB_1) \sum_{k,l \in E_M} \sum_{a \in A_N} p_a^{(k)} p_a^{(l)} \tilde{\pi}_k \tilde{\pi}_l \right| \leq \frac{1}{2}(\bar{\lambda}_2 - \underline{\lambda}_2). \tag{18}$$

Рассмотрим разность

$$\begin{aligned}\bar{\lambda}_2 - \underline{\lambda}_2 &= 2C \sum_{k,l \in E_M} \sum_{a \in A_N} p_a^{(k)} p_a^{(l)} \tilde{\pi}_k \tilde{\pi}_l \sum_{1 \leq i < j \leq T} \Big(e^{-\alpha i} + e^{-\alpha(j-i)}\Big) \\ &= 4C \sum_{k,l \in E_M} \sum_{a \in A_N} p_a^{(k)} p_a^{(l)} \tilde{\pi}_k \tilde{\pi}_l \sum_{1 \leq i < j \leq T} e^{-\alpha i}.\end{aligned}$$

Из равенств

$$\sum_{1 \leq i < j \leq T} e^{-\alpha i} = \sum_{1 \leq i < j \leq T} e^{-\alpha(j-i)} = \frac{1}{(e^\alpha - 1)^2} \Big(T(e^\alpha - 1) - e^\alpha + e^{-\alpha(T-1)}\Big) \tag{19}$$

следует, что

$$\bar{\lambda}_2 - \underline{\lambda}_2 \leq 4C_1 T \sum_{k,l \in E_M} \sum_{a \in A_N} p_a^{(k)} p_a^{(l)} \tilde{\pi}_k \tilde{\pi}_l.$$

Подставив последнюю оценку в правую часть формулы (18), получим

$$\left| \mathbf{E}\xi_2 - (C_T^2 + CB_1) \sum_{k,l \in E_M} \sum_{a \in A_N} p_a^{(k)} p_a^{(l)} \tilde{\pi}_k \tilde{\pi}_l \right| \leq 4C_1 T \sum_{k,l \in E_M} \sum_{a \in A_N} p_a^{(k)} p_a^{(l)} \tilde{\pi}_k \tilde{\pi}_l.$$

Из последней формулы следует, что при $T \to \infty$ имеет место (4). Лемма 1 доказана. □

Доказательство леммы 2. Воспользуемся теоремой 1.C книги [21], которая в нашем случае формулируется следующим образом:

$$\rho_{TV}\left(\mathrm{Pois}(\lambda_2(\mathbf{Z})), \mathrm{Pois}(\lambda_2)\right) \leq \frac{1 - e^{-\lambda_2}}{\lambda_2} \mathbf{D}\lambda_2(\mathbf{Z}). \tag{20}$$

Вычислим сначала второй момент случайной величины $\lambda_2(\mathbf{Z})$. Согласно (3)

$$\begin{aligned}\mathbf{E}(\lambda_2(\mathbf{Z}))^2 &= \mathbf{E}\left(\sum_{1 \leq i < j \leq T} \sum_{a \in A_N} p_a^{(Z_i)} p_a^{(Z_j)}\right)^2 \\ &= \sum_{1 \leq i < j \leq T} \sum_{a \in A_N} \sum_{1 \leq i' < j' \leq T} \sum_{b \in A_N} \mathbf{E} p_a^{(Z_i)} p_a^{(Z_j)} p_b^{(Z_{i'})} p_b^{(Z_{j'})}.\end{aligned}$$

Введем обозначения

$$\mathbf{P}\{Z_i = k, Z_j = l\} = p_{i,j}(k,l), \quad \mathbf{P}\{Z_i = k, Z_j = l, Z_{i'} = k'\} = p_{i,j,i'}(k,l,k'),$$
$$\mathbf{P}\{Z_i = k, Z_j = l, Z_{i'} = k', Z_{j'} = l'\} = p_{i,j,i',j'}(k,l,k',l')$$

при $1 \le i < j \le T$, $1 \le i' < j' \le T$, $k, l, k', l' \in E_M$. Очевидно, что

$$p_{i,j}(k,l) = p_{i,j,i}(k,l,k) = p_{i,j,i,j}(k,l,k,l),$$

$$p_{i,j,i'}(k,l,k') = p_{i,j,i',j}(k,l,k',l), \quad p_{i,j,j'}(k,l,l') = p_{i,j,i,j'}(k,l,k,l').$$

Тогда

$$\mathbf{E}(\lambda_2(\mathbf{Z}))^2$$

$$(21) \quad = \sum_{1\le i<j\le T} \sum_{a\in A_N} \sum_{1\le i'<j'\le T} \sum_{b\in A_N} \sum_{k,l,k',l'\in E_M} p_a^{(k)} p_a^{(l)} p_b^{(k')} p_b^{(l')} p_{i,j,i',j'}(k,l,k',l').$$

Разобьем сумму в правой части (21) на отдельные слагаемые, которые обозначим

$$D_1 = \sum_{1\le i<j\le T} \sum_{a\in A_N} \sum_{b\in A_N} \sum_{k,l,k',l'\in E_M} p_a^{(k)} p_a^{(l)} p_b^{(k')} p_b^{(l')} p_{i,j,i,j}(k,l,k',l'),$$

$$D_2 = \sum_{1\le i<j\le T} \sum_{a\in A_N} \sum_{\substack{i<j'\le T,\\ j'\ne j}} \sum_{b\in A_N} \sum_{k,l,k',l'\in E_M} p_a^{(k)} p_a^{(l)} p_b^{(k')} p_b^{(l')} p_{i,j,i,j'}(k,l,k',l'),$$

$$D_3 = \sum_{1\le i<j\le T} \sum_{a\in A_N} \sum_{\substack{1\le i'<j,\\ i'\ne i}} \sum_{b\in A_N} \sum_{k,l,k',l'\in E_M} p_a^{(k)} p_a^{(l)} p_b^{(k')} p_b^{(l')} p_{i,j,i',j}(k,l,k',l'),$$

$$D_4 = \sum_{1\le i<j\le T} \sum_{a\in A_N} \sum_{\substack{1\le i'<j'\le T,\\ |\{i,j,i',j'\}|=4}} \sum_{b\in A_N} \sum_{k,l,k',l'\in E_M} p_a^{(k)} p_a^{(l)} p_b^{(k')} p_b^{(l')} p_{i,j,i',j'}(k,l,k',l').$$

Начнем с оценивания D_1. Так как $p_{i,j,i,j}(k,l,k',l') = 0$ при $k' \ne k$ или $l' \ne l$, и $p_{i,j,i,j}(k,l,k',l') = p_{i,j}(k,l) \ge 0$ только при $k' = k$, $l' = l$, то

$$\begin{aligned} D_1 &= \sum_{1\le i<j\le T} \sum_{a\in A_N} \sum_{b\in A_N} \sum_{k,l\in E_M} p_a^{(k)} p_a^{(l)} p_b^{(k)} p_b^{(l)} p_{i,j}(k,l) \\ &\le \sum_{a\in A_N} \sum_{k,l\in E_M} p_a^{(k)} p_a^{(l)} p^* \sum_{1\le i<j\le T} p_{i,j}(k,l) \sum_{b\in A_N} p_b^{(k)} \\ &= p^* \sum_{a\in A_N} \sum_{k,l\in E_M} p_a^{(k)} p_a^{(l)} \sum_{1\le i<j\le T} p_{i,j}(k,l) = p^* \lambda_2. \end{aligned}$$

Оценим D_2. Аналогично предыдущему случаю, $p_{i,j,i,j'}(k,l,k',l') = p_{i,j,j'}(k,l,l') \ge 0$ при $k = k'$. Поэтому

$$\begin{aligned} D_2 &= \sum_{1\le i<j\le T} \sum_{a\in A_N} \sum_{\substack{i<j'\le T,\\ j'\ne j}} \sum_{b\in A_N} \sum_{k,l,l'\in E_M} p_a^{(k)} p_a^{(l)} p_b^{(k)} p_b^{(l')} p_{i,j,j'}(k,l,l') \\ &= \sum_{1\le i<j\le T} \sum_{a\in A_N} \sum_{b\in A_N} \sum_{k,l,l'\in E_M} p_a^{(k)} p_a^{(l)} p_b^{(k)} p_b^{(l')} \left(\sum_{j<j'\le T} p_{i,j}(k,l) \pi_{ll'}^{(j'-j)} \right. \end{aligned}$$

$$(22) \qquad \left. + \sum_{i<j'<j} p_{i,j'}(k,l') \pi_{l'l}^{(j-j')} \right).$$

Разобьем сумму в скобках на два слагаемых, которые обозначим $D_{2,1}$ и $D_{2,2}$. Из (1) имеем $\mathbf{P}\{Z_{j'} = l' | Z_j = l\} = \pi_{ll'}{}^{(j'-j)} \le \pi_{l'}(1 + Ce^{-a(j'-j)})$ при $j' > j$. Значит,

$$D_{2,1} = \sum_{1\le i<j\le T} \sum_{a\in A_N} \sum_{b\in A_N} \sum_{k,l,l'\in E_M} p_a^{(k)} p_a^{(l)} p_b^{(k)} p_b^{(l')} \sum_{j<j'\le T} p_{i,j}(k,l) \pi_{ll'}^{(j'-j)}$$

$$\le \sum_{1\le i<j<j'\le T} \sum_{a\in A_N} \sum_{b\in A_N} \sum_{k,l\in E_M} p_a^{(k)} p_a^{(l)} p_b^{(k)} p^* p_{i,j}(k,l) \sum_{l'\in E_M} \pi_{ll'}^{(j'-j)}.$$

Так как $\sum\limits_{l'\in E_M} \pi_{ll'}^{(j'-j)} = 1$ и $\sum\limits_{b\in A_N} p_b^{(k)} = 1$, то

$$D_{2,1} \le p^* \sum_{1\le i<j<j'\le T} \sum_{a\in A_N} \sum_{k,l\in E_M} p_a^{(k)} p_a^{(l)} p_{i,j}(k,l) \le T p^* \lambda_2.$$

Аналогично,

$$D_{2,2} = \sum_{1\le i<j'<j\le T} \sum_{a\in A_N} \sum_{b\in A_N} \sum_{k,l,l'\in E_M} p_a^{(k)} p_a^{(l)} p_b^{(k)} p_b^{(l')} p_{i,j'}(k,l') \pi_{l'l}^{(j-j')}$$

$$\le \sum_{1\le i<j'<j\le T} \sum_{a\in A_N} \sum_{b\in A_N} \sum_{k,l'\in E_M} p_a^{(k)} p^* p_b^{(k)} p_b^{(l')} p_{i,j'}(k,l') \sum_{l\in E_M} \pi_{l'l}^{(j-j')}$$

$$= p^* \sum_{1\le i<j'<j\le T} \sum_{b\in A_N} \sum_{k,l'\in E_M} p_b^{(k)} p_b^{(l')} p_{i,j'}(k,l') \le p^* T \lambda_2.$$

Подставляя полученные оценки для $D_{2,1}$ и $D_{2,2}$ в правую часть (22), получаем оценку

$$D_2 \le 2p^* \lambda_2 T.$$

Оценка для третьей суммы D_3 в формуле (21) проводится следующим образом:

$$D_3 = \sum_{1\le i<j\le T} \sum_{a\in A_N} \sum_{\substack{1\le i'<j,\\ i'\ne i}} \sum_{b\in A_N} \sum_{k,l,k'\in E_M} p_a^{(k)} p_a^{(l)} p_b^{(k')} p_b^{(l)} p_{i,j,i'}(k,l,k')$$

$$\le \sum_{1\le i<j\le T} \sum_{a\in A_N} \sum_{\substack{1\le i'<j,\\ i'\ne i}} \sum_{b\in A_N} \sum_{k,l,k'\in E_M} p_a^{(k)} p_a^{(l)} p^* p_b^{(l)} p_{i,j}(k,l)$$

$$= Mp^* \sum_{1\le i<j\le T} \sum_{a\in A_N} \sum_{\substack{1\le i'<j,\\ i'\ne i}} \sum_{k,l\in E_M} p_a^{(k)} p_a^{(l)} p_{i,j}(k,l) \le Mp^* T \lambda_2.$$

Наконец, оценим D_4. Заметим, что в силу (1)

$$D_4 \le \left(\sum_{a\in A_N} \sum_{k,l\in E_M} p_a^{(k)} p_a^{(l)} \tilde{\pi}_k \tilde{\pi}_l \sum_{1\le i<j\le T} \left(1 + Ce^{-\alpha i} + Ce^{-\alpha(j-i)} + Ce^{-\alpha j}\right) \right)^2.$$

Из последнего неравенства и формул (17), (19) следует (аналогично (4)), что

$$D_4 \le \left(\lambda_2(1 + O(T^{-1}))\right)^2 = \lambda_2^2(1 + O(T^{-1})).$$

Подставляя полученные оценки для слагаемых в (21), приходим к неравенству

$$\mathbf{E}(\lambda_2(\mathbf{Z}))^2 \le p^*\lambda_2(1+(M+2)T) + \lambda_2^2(1+O(T^{-1})).$$

Значит,

$$\mathbf{D}\lambda_2(\mathbf{Z}) \le p^*\lambda_2(1+(M+2)T) + \lambda_2^2 O(T^{-1}).$$

Подставляя эту оценку в правую часть (20), получаем (7). Лемма 2 доказана. □

Доказательство леммы 3. Для каждой пары индексов $(i,j) : 1 \le i < j \le T$ определим множество $O(i,j)$ равенством

$$O(i,j) = \{(i,k) : i \le k \le T\} \cup \{(k,j) : 1 \le k \le j\}.$$

Тогда случайный индикатор $I\{Y_i = Y_j\}$ и набор случайных индикаторов

$$\left(I\{Y_{i'} = Y_{j'}\}\right), (i',j') : 1 \le i' < j' \le T, (i',j') \notin O(i,j),$$

независимы. Тогда согласно теореме 1 работы [22] расстояние по вариации между распределением случайной величины ς_2 и сопровождающим пуассоновским распределением $\mathrm{Pois}(\mu_2)$ оценивается как

$$\rho_{TV}(\mathcal{L}(\varsigma_2), \mathrm{Pois}(\mathbf{E}\varsigma_2)) \le \frac{1-e^{-\mathbf{E}\varsigma_2}}{\mathbf{E}\varsigma_2}(S_1+S_2), \tag{23}$$

$$S_1 = \sum_{1\le i<j\le T} \sum_{(i',j')\in O(i,j)} \mathbf{P}\{Y_i = Y_j\}\mathbf{P}\{Y_{i'} = Y_{j'}\},$$

$$S_2 = \sum_{1\le i<j\le T} \sum_{(i',j')\in O(i,j)\setminus\{(i,j)\}} \mathbf{P}\{Y_i = Y_j, Y_{i'} = Y_{j'}\}.$$

Оценим S_1 в правой части формулы (23). Для этого заметим, что

$$S_1 = \sum_{1\le i<j\le T} \sum_{(i',j')\in O(i,j)} \sum_{a\in A_N} \sum_{b\in A_N} \mathbf{P}\{Y_i = Y_j = a\}\mathbf{P}\{Y_{i'} = Y_{j'} = b\}$$

$$= \sum_{1\le i<j\le T} \sum_{(i',j')\in O(i,j)} \sum_{a\in A_N} \sum_{b\in A_N} p_a^{(i)} p_a^{(j)} p_b^{(i')} p_b^{(j')}$$

$$= \sum_{1\le i<j\le T} \sum_{a\in A_N} \sum_{b\in A_N} p_a^{(i)} p_a^{(j)} \left[\sum_{k=i+1}^{T} p_b^{(i)} p_b^{(k)} + \sum_{k=1}^{j-1} p_b^{(k)} p_b^{(j)} + p_a^{(i)} p_a^{(j)}\right]$$

$$= \sum_{1\le i<j\le T} \sum_{a\in A_N} \sum_{b\in A_N} p_a^{(i)} p_a^{(j)} \left[p_b^{(i)} \sum_{k=i}^{T} p_b^{(k)} + p_b^{(j)} \sum_{k=1}^{j} p_b^{(k)}\right].$$

Для оценки выражения в квадратных скобках в правой части последнего неравенства воспользуемся тем, что $p_b^{(i)} \le p^*$ и

$$p_b^{(i)} \sum_{k=i}^{T} p_b^{(i)} + p_b^{(j)} \sum_{k=1}^{j} p_b^{(j)} \le p^*(p_b^{(i)}(T-i+1) + j p_b^{(j)}).$$

Таким образом,

$$S_1 \le p^* \sum_{1\le i<j\le T} \sum_{a\in A_N} \sum_{b\in A_N} p_a^{(i)} p_a^{(j)} (p_b^{(i)}(T-i+1) + j p_b^{(j)})$$

$$\leq p^* \sum_{1 \leq i < j \leq T} \sum_{a \in A_N} p_a^{(i)} p_a^{(j)} (T - i + 1 + j)$$

$$\leq 2Tp^* \sum_{1 \leq i < j \leq T} \sum_{a \in A_N} p_a^{(i)} p_a^{(j)} \leq 2Tp^* \mathbf{E}\varsigma_2. \tag{24}$$

Перейдем в оцениванию второй суммы S_2 в правой части (23). В силу определения множеств $O(i,j)$ имеем

$$S_2 = \sum_{1 \leq i < j \leq T} \left[\sum_{\substack{k=i+1,\ldots,T, \\ k \neq j}} \mathbf{P}\{Y_i = Y_j, Y_i = Y_k\} + \sum_{\substack{k=1,\ldots,j-1, \\ k \neq i}} \mathbf{P}\{Y_i = Y_j, Y_k = Y_j\} \right]$$

$$= \sum_{1 \leq i < j \leq T} \sum_{a \in A_N} \left[\sum_{\substack{k=i+1,\ldots,T, \\ k \neq j}} \mathbf{P}\{Y_i = Y_j = Y_k = a\} + \sum_{\substack{k=1,\ldots,j-1, \\ k \neq i}} \mathbf{P}\{Y_i = Y_k = Y_j = a\} \right]$$

$$= \sum_{1 \leq i < j \leq T} \sum_{a \in A_N} p_a^{(i)} p_a^{(j)} \left[\sum_{\substack{k=i+1,\ldots,T, \\ k \neq j}} p_a^{(k)} + \sum_{\substack{k=1,\ldots,j-1, \\ k \neq i}} p_a^{(k)} \right].$$

Аналогично оценке для S_1 получим, что

$$\sum_{\substack{k=i+1,\ldots,T, \\ k \neq j}} p_a^{(k)} + \sum_{\substack{k=1,\ldots,j-1, \\ k \neq i}} p_a^{(k)} \leq p^*(T - i + j - 1) \leq 2Tp^*.$$

Тогда

$$S_2 \leq 2Tp^* \sum_{1 \leq i < j \leq T} \sum_{a \in A_N} p_a^{(i)} p_a^{(j)} = 2Tp^* \mathbf{E}\varsigma_2. \tag{25}$$

Подставляя оценки (24) и (25) в неравенство (23), получаем (9). Лемма 3 доказана. □

Доказательство леммы 4. Согласно определению

$$\mathbf{E}\xi_3 = \mathbf{E} \sum_{1 \leq i < j < k \leq T} \mathbf{P}\{X_i = X_j = X_k | \mathbf{Z}\} = \mathbf{E} \sum_{1 \leq i < j < k \leq T} \sum_{a \in A_N} p_a^{(Z_i)} p_a^{(Z_j)} p_a^{(Z_k)}$$

$$= \sum_{l,m,n \in E_M} \sum_{1 \leq i < j < k \leq T} \sum_{a \in A_N} p_a^{(l)} p_a^{(m)} p_a^{(n)} \mathbf{P}\{Z_i = l, Z_j = m, Z_k = n\}$$

$$= \sum_{l,m,n \in E_M} \sum_{1 \leq i < j < k \leq T} \sum_{a \in A_N} p_a^{(l)} p_a^{(m)} p_a^{(n)} \pi_l(i) \pi_{lm}^{(j-i)} \pi_{mn}^{(k-j)}$$

$$= \sum_{l,m,n \in E_M} \sum_{a \in A_N} p_a^{(l)} p_a^{(m)} p_a^{(n)} \sum_{1 \leq i < j < k \leq T} \pi_l(i) \pi_{lm}^{(j-i)} \pi_{mn}^{(k-j)}.$$

Пользуясь оценкой (1), получаем

$$\mathbf{E}\xi_3 \leq \sum_{l,m,n \in E_M} \sum_{a \in A_N} p_a^{(l)} p_a^{(m)} p_a^{(n)} \sum_{1 \leq i < j < k \leq T} \pi_l(i) \pi_{lm}^{(j-i)} \tilde{\pi}_n \left(1 + Ce^{-\alpha(k-j)}\right)$$

$$\leq p^* \sum_{l,m\in E_M} \sum_{a\in A_N} p_a^{(l)} p_a^{(m)} \sum_{1\leq i<j<k\leq T} \pi_l(i)\pi_{lm}^{(j-i)} \left(1 + Ce^{-\alpha(k-j)}\right)$$
$$\leq 2Tp^* \sum_{l,m\in E_M} \sum_{a\in A_N} p_a^{(l)} p_a^{(m)} \sum_{1\leq i<j\leq T} \pi_l(i)\pi_{lm}^{(j-i)} = 2Tp^*\lambda_2.$$

Лемма 4 доказана. □

Автор выражает глубокую признательность рецензенту за полезные замечания и комментарии.

References

[1] Yu. A. Rozanov, *Stochastic processes. Short course*, M.: Nauka, 1979, 184 pp. (in Russian).

[2] V. F. Kolchin, B. A. Sevast'yanov, V. P. Chistyakov, *Random allocations*, Scripta Series in Mathematics, V. H. Winston & Sons, Washington, DC, 1978, 262 pp. Zbl 0464.60003

[3] J. Bartroff, L. Goldstein, *A Berry-Esseen bound for the uniform multinomial occupancy model*, Electron. J. Probab., **18** (2013), 1–29. Zbl 1287.60031

[4] V. G. Mikhailov, *An estimate of the rate of convergence to the Poisson distribution in group placing of particles*, Theory Probab. Appl., **22**:3 (1978), 554—562.

[5] A. D. Barbour, A. V. Gnedin, *Small counts in the infinite occupancy scheme*, Electron. J. Probab. , **14**:13 (2009), 365–384. Zbl 1189.60048

[6] G. Englund, *A remainder term estimate for the normal approximation in classical occupancy*, Ann. Prob. **9**:4 (1981), 684–692. Zbl 0464.60025

[7] H.-K. Hwang, S. Janson, *Local limit theorems for finite and infinite urn models*, Ann. Prob., **36**:3 (2008), 992–1022. Zbl 1138.60027

[8] P. F. Belyaev, *On the joint frequency distribution of outcomes in Markov chains with large number of states*, Theory Probab. Appl., **22**:3 (1977), 521–532. Zbl 0389.60051

[9] A. M. Zubkov, *Inequalities for transition probabilities with taboos and their applications*, Math. USSR-Sb., **37**:4 (1980), 451—488. Zbl 0453.60066

[10] A. M. Zubkov, V. G. Mikhailov, *Limit distributions of random variables connected with long duplications in a sequence of independent trials*, Theory Probab. Appl., **19**:1 (1974), 172—179. Zbl 0326.60024

[11] V. F. Kolchin, V. P. Chistyakov, *Limit distributions of the number of not appeared s-tuples in a multinomial scheme*, Theory Probab. Appl., **19**:4 (1975), 822—830.

[12] V. G. Mikhailov, *Limit distributions of random variables connected with multiple long duplications in a sequence lof independent trials*, Theory Probab. Appl., **19**:1 (1974), 180—184.

[13] A. M. Zubkov, V. G. Mikhailov, *On the accuracy of Poisson approximation in an occupancy problem*, Theory Probab. Appl., **23**:4 (1979), 789—794.

[14] A. M. Zubkov, V. G.Mikhailov, *On the repetitions of s-tuples in a sequence of independent trials*, Theory Probab. Appl., **24**:2 (1979), 269—282.

[15] L. Goldstein, *Berry Esseen Bounds for Combinatorial Central Limit Theorems and Pattern Occurrences, using Zero and Size Biasing*, Journal of Applied Probability, **42** (2005), 661–683. Zbl 1087.60021

[16] V. G. Mikhailov, A. M. Shoitov, *On repetitions of long tuples in a Markov chain*, Discrete Math. Appl., **25**:5 (2015), 295—303.

[17] V. G. Mikhailov, A. M. Shoitov, *On multiple repetitions of long tuples in a Markov chain*, Mat. Vopr. Kriptogr., **6**:3 (2015), 117—133 (in Russian).

[18] V. G. Mikhailov, *On the estimation of the accuracy of Poisson approximation for distribution of number of series of repetitions of long tuples in a Markov chain*, Discrete Math. Appl., **27**:4 (2015), 67—78.

[19] R.J. Elliott, L. Aggoun, J.B. Moore, *Hidden Markov models*, Applications of Mathematics, **29**, New-York, Springer-Verlag, 1995. Zbl 0819.60045

[20] A. A. Petushko, *On Markov random fields and their link with Markov chains*, Intellect. systemy. Theory and Appl., **14**:1–4 (2010), 225–236 (in Russian).

[21] A. D. Barbour, L. Holst, S. Janson, *Poisson Approximation*, Oxford, Oxford Univ. Press, 1992. Zbl 0746.60002

[22] R. Arratia, L. Goldstein, L. Gordon, *Two Moments Suffice for Poisson Approximations: The Chen-Stein Method*, Ann. Prob., **17**:1 (1989), 9–25. Zbl 0675.60017

Natalia Mikhailovna Mezhennaya
Bauman Moscow State Technical University,
ul. Baumanskaya 2-ya, 5/1,
105005, Moscow, Russia
E-mail address: natalia.mezhennaya@gmail.com