Math-Net.Ru

Общероссийский математический портал

В. Г. Михайлов, Оценки точности пуассоновской аппроксимации для распределения числа серий повторений длинных цепочек в цепи Маркова, *Дискрет. матем.*, 2015, том 27, выпуск 4, 67–78

DOI: <https://doi.org/10.4213/dm1348>

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

Параметры загрузки:

IP: 54.174.149.18

26 апреля 2023 г., 10:34:59

Оценки точности пуассоновской аппроксимации для распределения числа серий повторений длинных цепочек в цепи Маркова

© 2015 г. В. Г. Михайлов*

Пусть $X_0, X_1, \dots$ — простая эргодическая цепь Маркова с конечным числом состояний. Получены равномерные по k оценки точности пуассоновской аппроксимации для распределения числа серий k -кратных повторений s -цепочек в начальном отрезке цепи $X_0, X_1, \dots$

Исследование выполнено за счет гранта Российского научного фонда (проект № 14-50-00005).

Ключевые слова: цепь Маркова, k -кратные повторения s -цепочек, точность пуассоновской аппроксимации

1. Введение

Пусть последовательность $\mathbf{X} = \{X_0, X_1, \dots\}$ образует простую неразложимую и ациклическую однородную цепь Маркова с состояниями $1, \dots, N$, матрицей переходных вероятностей $\mathbb{P} = \|p_{a,b}\|$ и произвольным начальным распределением. Отметим, что такая цепь является эргодической. Положим $Y_i(s) = (X_i, \dots, X_{i+s-1})$, $s, i = 1, 2, \dots$, и для натурального $k \geq 2$ рассмотрим события

$$E_{i_1, \dots, i_k} = E_{i_1, \dots, i_k}(s) = \{Y_{i_1}(s) = Y_{i_2}(s) = \dots = Y_{i_k}(s)\}.$$

Пусть

$$\xi_{n,k}(s) = \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} I\{E_{i_1, \dots, i_k}\}$$

— число k -кратных повторений s -цепочек, а

$$\tilde{\xi}_{n,k}(s) = \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} I\{\bar{E}_{i_1-1, \dots, i_k-1} \cap E_{i_1, \dots, i_k}\}$$

*Место работы: Математический институт им. В.А. Стеклова Российской академии наук, e-mail: mikhail@mi.ras.ru

— общее число серий, образованных k -кратными повторениями s -цепочек, начинающихся до момента n включительно (здесь и далее $\bar{E}$ обозначает дополнение к событию E).

В работах [1] – [4] величины $Y_i(s)$, $\xi_{n,k}(s)$ и $\tilde{\xi}_{n,k}(s)$ рассматривались в последовательности полиномиальных испытаний. В [2] – [4] было показано, что при $s, n \rightarrow \infty$ и $\mathbf{E}\tilde{\xi}_{n,k}(s) \rightarrow \lambda \in (0, \infty)$ распределения величин $\xi_{n,k}(s)$ и $\tilde{\xi}_{n,k}(s)$ сходятся к сложному пуассоновскому и пуассоновскому распределениям соответственно.

Задачи, связанные с совпадениями цепочек в двух независимых конечных цепях Маркова, рассматривались в [5]. Задачи о повторениях состояний и цепочек состояний в цепях Маркова в схемах с возрастающим числом исходов изучались в [6] – [8].

Отметим две работы посвященные исследованию числа повторений s -цепочек в неразложимой и ациклической цепи Маркова с фиксированным распределением. В [9] изучалось распределение величины $\tilde{\xi}_{n,2}(s)$ и, в частности, были выведены достаточные условия сходимости (при $n, s \rightarrow \infty$) распределения случайной величины $\tilde{\xi}_{n,2}(s)$ к распределению Пуассона и к нормальному распределению. В качестве следствий этих предельных теорем были получены предельные теоремы для времени первого повторения s -цепочек $\tau(s)$ и для максимальной длины повторившейся цепочки $\mu(n)$.

В недавней работе [10] был рассмотрен случай повторений произвольной кратности. В частности, были получены достаточные условия сходимости распределений случайных величин $\tilde{\xi}_{n,k}(s)$ и $\xi_{n,k}(s)$ к распределению Пуассона и к сложному пуассоновскому распределению соответственно.

В настоящей работе мы продолжаем исследования работ [9] и [10]. Получены равномерные по k оценки точности пуассоновской аппроксимации распределения числа серий k -кратных повторений s -цепочек в начальном отрезке цепи $\mathbf{X}$. Из этих оценок выведены условия выполнения для $\tilde{\xi}_{n,k}(s)$ предельной теоремы Пуассона и достаточные условия асимптотической нормальности этой случайной величины при любом фиксированном $k \geq 2$ и $n, s \rightarrow \infty$.

2. Формулировки основных результатов

Следуя [9] и [10], обозначим через $\pi = (\pi_1, \dots, \pi_N)$ стационарное распределение цепи

$\mathbf{X}$ (здесь все $\pi_r > 0$), $P_k = \sum_{r=1}^N \pi_r^k$, $a, p = \max_{a,b \in \{1, \dots, N\}} p_{a,b}$.

Так как цепь Маркова $\mathbf{X}$ является эргодической, а все стационарные вероятности положительны, то найдутся такие величины $t(\mathbb{P}) < \infty$ и $c(\mathbb{P}) > 0$, что при $x, y \in \{1, \dots, N\}$ и всех $t > t(\mathbb{P})$

$$\left| p_{x,y}^{(t)} - \pi_y \right| < \pi_y e^{-c(\mathbb{P})t}. \quad (1)$$

Пусть $\rho(\Psi, \Gamma)$ обозначает расстояние по вариации между распределениями Ψ и Γ , $\mathcal{L}(Z)$ — распределение случайной величины Z , а $\text{Po}(\lambda)$ — распределение Пуассона с параметром λ .

Теорема 1. Пусть $2s^3 + 4s^2 < n$, $s^2 > t(\mathbb{P})$ и для некоторого $1 < \gamma < 2$ выполнены неравенства $s \geq \gamma^{-1} \sqrt{|\ln(p)|/c(\mathbb{P})}$, $\ln n < c(\mathbb{P})s^\gamma$, $s^3np^s < 1$. Тогда найдется такая константа $C < \infty$, что

$$\rho(\mathcal{L}(\tilde{\xi}_{n,2}(s)), \text{Po}(\mathbf{E}\tilde{\xi}_{n,2}(s))) < Cs^3np^s, \quad (2)$$

Теорема 2. Пусть $k \geq 3$, $ks^3 + 2ks^2 < n$, $s^2 > t(\mathbb{P})$ и для некоторого γ , $1 < \gamma < 2$, выполнены неравенства $s \geq \gamma^{-1} \sqrt{|\ln(p)|/c(\mathbb{P})}$, $k \leq s^{2-\gamma}/2$, $\ln n < c(\mathbb{P})^\gamma$ и $(2s^2)^{k-1}knp^s < 1$. Тогда найдется такая константа $C' < \infty$, что

$$\rho(\mathcal{L}(\tilde{\xi}_{n,k}(s)), \text{Po}(\mathbf{E}\tilde{\xi}_{n,k}(s))) < C' (2s^2)^{k-1}knp^s. \quad (3)$$

Из теорем 1 и 2 вытекают следующие утверждения.

Следствие 1. Пусть $k = 2$, $s, n \rightarrow \infty$ и $s^3np^s \rightarrow 0$. Если $\mathbf{E}\tilde{\xi}_{n,2}(s) \rightarrow \lambda \in (0, \infty)$, то $\mathcal{L}(\tilde{\xi}_{n,2}(s)) \rightarrow \text{Po}(\lambda)$. Если же $\mathbf{E}\tilde{\xi}_{n,2}(s) \rightarrow \infty$, то функция распределения случайной величины $(\mathbf{E}\tilde{\xi}_{n,2}(s))^{-1/2}(\tilde{\xi}_{n,2}(s) - \mathbf{E}\tilde{\xi}_{n,2}(s))$ сходится к функции стандартного нормального распределения.

Следствие 2. Пусть число $k \geq 3$ фиксировано, $s, n \rightarrow \infty$ и $s^{2(k-1)}np^s \rightarrow 0$. Если $\mathbf{E}\tilde{\xi}_{n,k}(s) \rightarrow \lambda \in (0, \infty)$, то $\mathcal{L}(\tilde{\xi}_{n,k}(s)) \rightarrow \text{Po}(\lambda)$. Если же $\mathbf{E}\tilde{\xi}_{n,k}(s) \rightarrow \infty$, то функция распределения случайной величины $(\mathbf{E}\tilde{\xi}_{n,k}(s))^{-1/2}(\tilde{\xi}_{n,k}(s) - \mathbf{E}\tilde{\xi}_{n,k}(s))$ сходится к функции стандартного нормального распределения.

Замечание 1. Пусть $\mathbb{P}_k = \|p_{a,b}^k\|$, а $\rho_k = \rho(\mathbb{P}_k) > 0$ — максимальное характеристическое число матрицы $\mathbb{P}_k$. В работе [10] было показано, что при переходе к пределу в условиях следствий 1 и 2 найдутся такие числа $0 < B_1 < B_2 < \infty$, что

$$B_1 C_n^k \rho_k^s < \mathbf{E}\tilde{\xi}_{n,k}(s) < B_2 C_n^k \rho_k^s. \quad (4)$$

Нетрудно убедиться, что в силу (4) одновременное выполнение условий $s^3np^s \rightarrow 0$ (или $s^{2(k-1)}np^s \rightarrow 0$) и $\mathbf{E}\tilde{\xi}_{n,k}(s) \rightarrow \lambda > 0$ следствий 1 и 2 возможно лишь при предположении, что $\rho_k < p^k$ и существует достаточно малая константа $c > 0$, при которой $\mathbf{E}\tilde{\xi}_{n,k}(s) = O(e^{cs})$ (подробнее см. [9]).

3. Доказательства теорем 1 и 2

Следуя [10], положим $J = \{\bar{i} = (i_1, \dots, i_k) : 1 \leq i_1 < \dots < i_k \leq n\}$,

$$J_{s,n} = \{\bar{i} \in J : \min_{1 \leq j \leq k} (i_j - i_{j-1}) > s^3\},$$

где считается, что $i_0 = 1$, и определим случайные величины

$$\delta_{n,k}(s) = \sum_{\bar{i} \in J_{s,n}} I\{\bar{E}_{i_1-1, \dots, i_k-1} \cap E_{i_1, \dots, i_k}\}. \quad (5)$$

Для того чтобы величина $\delta_{n,k}(s)$ могла быть отлична от нуля, мы используем условие $n > ks^3 + 2ks^2$. Оно также гарантирует содержательность некоторых построений, использованных при доказательстве теоремы.

Наш план следующий. Сначала оценим величины $\rho(\mathcal{L}(\tilde{\xi}_{n,k}(s)), \mathcal{L}(\delta_{n,k}(s)))$ и $\rho(\text{Po}(\mathbf{E}\delta_{n,k}(s)), \text{Po}(\mathbf{E}\tilde{\xi}_{n,k}(s)))$, затем — величину $\rho(\mathcal{L}(\delta_{n,k}(s)), \text{Po}(\mathbf{E}\delta_{n,k}(s)))$. После этого с помощью неравенства треугольника построим искомую оценку для $\rho(\mathcal{L}(\tilde{\xi}_{n,k}(s)), \text{Po}(\mathbf{E}\tilde{\xi}_{n,k}(s)))$.

1. Оценим $\rho(\mathcal{L}(\tilde{\xi}_{n,k}(s)), \mathcal{L}(\delta_{n,k}(s)))$. Так как

$$\rho(\mathcal{L}(\tilde{\xi}_{n,k}(s)), \mathcal{L}(\delta_{n,k}(s))) \leq \mathbf{P}\{\tilde{\xi}_{n,k}(s) \neq \delta_{n,k}(s)\},$$

а $\mathbf{P}\{\tilde{\xi}_{n,k}(s) \geq \delta_{n,k}(s)\} = 1$, то

$$\begin{aligned} \rho(\mathcal{L}(\tilde{\xi}_{n,k}(s)), \mathcal{L}(\delta_{n,k}(s))) &\leq \mathbf{P}\{\tilde{\xi}_{n,k}(s) \neq \delta_{n,k}(s)\} \leq \mathbf{E}\tilde{\xi}_{n,k}(s) - \mathbf{E}\delta_{n,k}(s) = \\ &= \sum_{\bar{i} \in J \setminus J_{s,n}} \mathbf{P}\{E_{i_1, \dots, i_k}\}. \end{aligned} \quad (6)$$

Для любого $\bar{i}$ из J определим множество

$$U(\bar{i}) = \bigcup_{t=1}^k [i_t, i_t + s - 1].$$

Для всех $r = 1, \dots, k$ положим

$$J(r) = \{\bar{i} \in J, \text{ множество } U(\bar{i}) \text{ состоит ровно из } r \text{ связных компонент}\}.$$

Очевидно, что $J = \bigcup_{r=1}^k J(r)$ и, следовательно,

$$J \setminus J_{s,n} = \bigcup_{r=1}^k H(r), \quad \text{где } H(r) = (J \setminus J_{s,n}) \cap J(r). \quad (7)$$

Сначала рассмотрим случай $r = k$. Используя неравенство $\mathbf{P}\{X_j = b | X_i = a\} \leq p$, $j > i$, получаем для неперекрывающихся цепочек длины s ($i_r \geq i_{r-1} + s$, $r = 2, \dots, k$) оценки (подробности см. в [10])

$$\begin{aligned} \mathbf{P}\{E_{i_1, \dots, i_k}\} &= \mathbf{P}\{Y_{i_1}(s) = Y_{i_2}(s) = \dots = Y_{i_k}(s)\} \leq \\ &\leq \sum_{(x_1, \dots, x_s) \in \{1, \dots, N\}^s} \mathbf{P}\{X_{i_1} = x_1\} p^{s(k-1)} p_{x_1, x_2} p_{x_2, x_3} \cdot \dots \cdot p_{x_{s-1}, x_s} = p^{(k-1)s}. \end{aligned} \quad (8)$$

Заметим, что $|H(k)| < |J \setminus J_{s,n}|$ и

$$C_n^k - C_{n-m}^k = C_n^k \left(1 - \prod_{j=0}^{k-1} \left(1 - \frac{m}{n-j} \right) \right) <$$

$$< C_n^k \left(1 - \left(1 - \frac{m}{n-k+1} \right)^k \right) < C_n^k \frac{km}{n-k}. \quad (9)$$

Поэтому

$$|H(k)| \leq |J \setminus J_{s,n}| = C_n^k - C_{n-ks^3}^k < C_n^k \frac{k^2 s^3}{n-k}. \quad (10)$$

Из (8) и (10) вытекает неравенство

$$\sum_{\tilde{i} \in H(k)} \mathbf{P} \left\{ E_{i_1, \dots, i_k} \right\} < C_n^k \frac{k^2 s^3}{n-k} p^{(k-1)s}. \quad (11)$$

Теперь рассмотрим случай $1 \leq r \leq k-1$. Рассуждаем так же, как в [10]. При $1 \leq r \leq k-1$ цепочки $Y_{i_1}(s), Y_{i_2}(s), \dots, Y_{i_k}(s)$ перекрываются. Пусть для определенности перекрываются цепочки $Y_{i_1}(s)$ и $Y_{i_2}(s)$ (возможно, не только они). Возьмем из всех других компонент связности множества $U(\tilde{i})$ по одной представляющей их цепочке: $Y_{i_{j_2}}(s), \dots, Y_{i_{j_r}}(s)$. Заметим, что при $i_2 = i_1 + 1, \dots, i_1 + s - 1$ фиксация вектора $Y_{i_1}(i_2 - i_1) = (x_1, \dots, x_{i_2 - i_1})$ однозначно определяет все компоненты вектора $Y_{i_2}(s) = (x_1, \dots, x_{i_2 - i_1}, x_1, \dots)$ и векторов $Y_{i_{j_2}}(s), \dots, Y_{i_{j_r}}(s)$. Поэтому

$$\begin{aligned} & \mathbf{P} \left\{ Y_{i_1}(s) = Y_{i_2}(s) = \dots = Y_{i_k}(s) \right\} \leq \\ & \leq \mathbf{P} \left\{ Y_{i_1}(s) = Y_{i_2}(s) = Y_{i_{j_2}}(s) = \dots = Y_{i_{j_r}}(s) \right\} \leq p^{rs}, \end{aligned}$$

а так как

$$|H(r)| \leq |J(r)| \leq C_n^r C_{k-1}^{r-1} s^{k-r}, \quad (12)$$

то

$$\sum_{\tilde{i} \in H(r)} \mathbf{P} \left\{ E_{i_1, \dots, i_k} \right\} < C_n^r C_{k-1}^{r-1} s^{k-r} p^{rs}, \quad 1 \leq r \leq k-1. \quad (13)$$

Из (6), (7), (11), (13) следует, что

$$\begin{aligned} & \rho \left(\mathcal{L}(\tilde{\xi}_{n,k}(s)), \mathcal{L}(\delta_{n,k}(s)) \right) \leq \mathbf{E} \tilde{\xi}_{n,k}(s) - \mathbf{E} \delta_{n,k}(s) < \\ & < C_n^k \frac{k^2 s^3}{n-k} p^{(k-1)s} + \sum_{r=1}^{k-1} C_n^r C_{k-1}^{r-1} s^{k-r} p^{rs}. \end{aligned} \quad (14)$$

В частности,

$$\begin{aligned} & \rho \left(\mathcal{L}(\tilde{\xi}_{n,2}(s)), \mathcal{L}(\delta_{n,2}(s)) \right) \leq \mathbf{E} \tilde{\xi}_{n,2}(s) - \mathbf{E} \delta_{n,2}(s) < \\ & < \left(\frac{2s^2}{1-2/n} + 1 \right) snp^s < \frac{8}{3} s^3 np^s, \end{aligned} \quad (15)$$

если $n \geq 8, s \geq 2$.

Из (14) также следует, что найдется такая константа $C_0 < \infty$, что при всех $k, s, n, 3 \leq k \leq s \leq n/k$,

$$\rho \left(\mathcal{L}(\tilde{\xi}_{n,k}(s)), \mathcal{L}(\delta_{n,k}(s)) \right) \leq \mathbf{E} \tilde{\xi}_{n,k}(s) - \mathbf{E} \delta_{n,k}(s) < C_0 s^{k-1} np^s. \quad (16)$$

Теперь оценим расстояние $\rho(\text{Po}(\mathbf{E}\delta_{n,k}(s)), \text{Po}(\mathbf{E}\tilde{\xi}_{n,k}(s)))$. Для этого воспользуемся теоремой 1.С и замечанием 1.1.4 из [13], согласно которым

$$\rho(\text{Po}(\mathbf{E}\delta_{n,k}(s)), \text{Po}(\mathbf{E}\tilde{\xi}_{n,k}(s))) \leq |\mathbf{E}\tilde{\xi}_{n,k}(s) - \mathbf{E}\delta_{n,k}(s)|. \quad (17)$$

Поэтому для этого расстояния выполнены оценки (15) – (16).

2. Оценим точность аппроксимации распределения случайной величины $\delta_n(s)$ распределением Пуассона с параметром $\mathbf{E}\delta_n(s)$. Для этого представим случайную величину $\delta_n(s)$ в виде суммы случайных индикаторов:

$$\delta_n(s) = \sum_{\bar{i} \in J_{s,n}} \eta(\bar{i}), \quad (18)$$

где $\eta(\bar{i}) = I\{\bar{E}_{i_1-1, \dots, i_k-1} \cap E_{i_1, \dots, i_k}\}$.

Для исследования распределения случайной величины $\delta_n(s)$ воспользуемся методом Чена – Стейна (см. например, [13] или [14]). Следуя этому методу, для любого $\bar{i} = (i_1, \dots, i_k)$ из $J_{s,n}$ определим множество $\Gamma(\bar{i}) \subset J_{s,n} \setminus \{\bar{i}\}$ таких $\bar{j} = (j_1, \dots, j_k)$ из $J_{s,n}$, $\bar{j} \neq \bar{i}$, что хотя бы одно из расстояний между отрезками $[j_1 - 1, j_1 + s - 1], \dots, [j_k - 1, j_k + s - 1]$ и $[i_1 - 1, i_1 + s - 1], \dots, [i_k - 1, i_k + s - 1]$ меньше s^2 . Для $\bar{i}$ из $J_{s,n}$ положим

$$V_{\bar{i}} = \sum_{\bar{j} \in J_{s,n} \setminus (\Gamma(\bar{i}) \cup \{\bar{i}\})} \eta(\bar{j}).$$

Согласно теореме 2.5 из [14] в нашем случае справедлива верхняя оценка

$$\begin{aligned} & \rho(\mathcal{L}(\delta_{n,k}(s)), \text{Po}(\mathbf{E}\delta_{n,k}(s))) \leq \\ & \leq \frac{1 - e^{-\mathbf{E}\delta_{n,k}(s)}}{\mathbf{E}\delta_{n,k}(s)} \left(\sum_{\bar{i} \in J_{s,n}} \left(\mathbf{E}\eta(\bar{i}) \left(\mathbf{E}\eta(\bar{i}) + \sum_{\bar{j} \in \Gamma(\bar{i})} \mathbf{E}\eta(\bar{j}) \right) \right) + \sum_{\bar{i} \in J_{s,n}} \sum_{\bar{j} \in \Gamma(\bar{i})} \mathbf{E}(\eta(\bar{j})\eta'(\bar{i})) \right) + \\ & + \min\left(1, \sqrt{\frac{2}{e\mathbf{E}\delta_{n,k}(s)}}\right) \sum_{\bar{i} \in J_{s,n}} \mathbf{E}|\mathbf{E}\eta(\bar{i}) - \mathbf{E}\{\eta(\bar{i})|V_{\bar{i}}\}|. \end{aligned} \quad (19)$$

Оценим сумму

$$\Sigma_1 = \sum_{\bar{i} \in J_{s,n}} \left(\mathbf{E}\eta(\bar{i}) \left(\mathbf{E}\eta(\bar{i}) + \sum_{\bar{j} \in \Gamma(\bar{i})} \mathbf{E}\eta(\bar{j}) \right) \right).$$

Заметим, что

$$\mathbf{E}\eta(\bar{j}) \leq \mathbf{P}\{Y_{j_1}(s) = Y_{j_2}(s) = \dots = Y_{j_k}(s)\} \leq p^{(k-1)s}.$$

Поэтому

$$\Sigma_1 \leq \sum_{\bar{i} \in J_{s,n}} (1 + |\Gamma(\bar{i})|) p^{(k-1)s}.$$

Аналогично (10) из (9) получаем (здесь, конечно, используется условие $n > ks^3 + 2ks^2$)

$$|\Gamma(\bar{i})| \leq C_{n-ks^3}^k - C_{n-ks^3-2ks^2}^k < C_{n-ks^3}^k \frac{2k^2 s^2}{n - ks^3 - k} \leq \frac{2ks^2 n^{k-1}}{(k-1)!} \quad (20)$$

для всех $\bar{i} \in J_{s,n}$ (в последнем переходе было использовано, что при наших предположениях $s^3 \geq k+1$). Собрав приведенные оценки, получим

$$\Sigma_1 \leq \mathbf{E} \delta_{n,k}(s) \left(p^{(k-1)s} + \frac{2ks^2}{(k-1)!} (np^s)^{k-1} \right) < \mathbf{E} \delta_{n,k}(s) (4s^2 + 1) (np^s)^{k-1}. \quad (21)$$

Значит, найдется такая константа $C_1 < \infty$, при которой для всех k, s, n , $n > ks^3 + 2ks^2$, $k \geq 2$, выполнено неравенство

$$\Sigma_1 < C_1 \mathbf{E} \delta_{n,k}(s) s^2 (np^s)^{k-1}. \quad (22)$$

Теперь оценим сумму

$$\Sigma_2 = \sum_{\bar{i} \in J_{s,n}} \sum_{\bar{j} \in \Gamma(\bar{i})} \mathbf{E}(\eta(\bar{j}) \eta'(\bar{i})).$$

Следующие построения мы заимствуем из работы [10]. Для любого $\bar{i}$ из $J_{s,n}$ разобьем множество $\Gamma(\bar{i})$ на несколько подмножеств. Определим их следующим образом:

$\Gamma_a(\bar{i})$ — ровно a отрезков $[j_1 - 1, j_1 + s + l], \dots, [j_k - 1, j_k + s + l]$ лежат на расстоянии менее s^2 от отрезков $[i_1 - 1, i_1 + s + m], \dots, [i_k - 1, i_k + s + m]$, а остальные отрезки лежат на расстоянии не менее s^2 друг от друга, здесь $0 \leq a \leq k-1$,

$\Gamma'_k(\bar{i})$ — каждый из k отрезков $[j_1 - 1, j_1 + s + l], \dots, [j_k - 1, j_k + s + l]$ лежит на расстоянии менее s^2 от соответствующего отрезка среди $[i_1 - 1, i_1 + s + m], \dots, [i_k - 1, i_k + s + m]$, при этом хотя бы один не пересекается с этим отрезком более чем на одну позицию,

$$\Gamma''_k(\bar{i}) = \{\bar{j} \in \Gamma_k(\bar{i}) \setminus \Gamma'_k(\bar{i}) : j_1 - i_1 = \dots = j_k - i_k\},$$

$$\Gamma'''_k(\bar{i}) = \Gamma_k(\bar{i}) \setminus (\Gamma'_k(\bar{i}) \cup \Gamma''_k(\bar{i})).$$

Из определений и оценки (20) следует, что

$$|\Gamma_a(\bar{i})| \leq C_k^a (2(s^2 + s + 1))^a C_{n-(k-a)s^3-2(k-a)s^2}^{k-a} <$$

$$< \frac{k! (2(s^2 + s + 1))^a}{a!((k-a)!)^2} n^{k-a}, \quad 1 \leq a \leq k-1,$$

$$|\Gamma'_k(\bar{i})| < (2(s^2 + s + 1))^k, \quad |\Gamma'''_k(\bar{i})| < (2(s^2 + s + 1))^k.$$

Кроме того,

$$\mathbf{P}\{\eta(\bar{j}) = 1 | \eta(\bar{i}) = 1\} \leq p^{(k-a)s}, \quad \bar{j} \in \Gamma_a(\bar{i}),$$

$$\mathbf{P}\{\eta(\bar{j}) = 1 | \eta(\bar{i}) = 1\} \leq p^{2s}, \quad \bar{j} \in \Gamma'_k(\bar{i}),$$

$$\mathbf{P}\{\eta(\bar{j}) = 1 | \eta(\bar{i}) = 1\} = 0, \quad \bar{j} \in \Gamma''_k(\bar{i}),$$

$$\mathbf{P}\{\eta(\bar{j}) = \eta(\bar{i}) = 1\} \leq p^{ks}, \quad \bar{j} \in \Gamma'''_k(\bar{i}).$$

Первые три оценки вытекают непосредственно из определений. Последняя оценка доказывается индукцией по $k \geq 2$.

Воспользуемся этими результатами при построении оценки для Σ_2 . Заметим, что при $a \leq s$ выполнено неравенство $(s^2 + s + 1)^a < s^{2a} e^{(a+1)/s}$. Поэтому, используя приведенные выше оценки и (18), получаем:

$$\begin{aligned} \sum_{\bar{i} \in J_{s,n}} \sum_{\beta \in \Gamma_a(\bar{i})} \mathbf{E}(\eta(\bar{i})\eta(\bar{j})) &= \sum_{\bar{i} \in J_{s,n}} \mathbf{P}\{\eta(\bar{i}) = 1\} \sum_{\beta \in \Gamma_a((\bar{i}))} \mathbf{P}\{\eta(\bar{j}) = 1 | \eta(\bar{i}) = 1\} < \\ &< \mathbf{E}\delta_{n,k}(s) \frac{k!}{a!((k-a)!)^2} (2(s^2 + s + 1))^a n^{k-a} p^{(k-a)s} < \\ &< \mathbf{E}\delta_{n,k}(s) \frac{1}{((k-a)!)^2} (2s^2)^a e^{k/s} (knp^s)^{k-a}, \quad a = 1, \dots, k-1. \end{aligned}$$

Из этих оценок следует, что при $knp^s < 1$

$$\begin{aligned} \sum_{a=1}^{k-1} \sum_{\bar{i} \in J_{s,n}} \sum_{\beta \in \Gamma_a(\bar{i})} \mathbf{E}(\eta(\bar{i})\eta(\bar{j})) &< \\ &< (2s^2)^{k-1} e^{k/s} \mathbf{E}\delta_{n,k}(s) \left(knp^s + I\{k \geq 3\} (knp^s)^2 \sum_{a=1}^{k-2} \frac{1}{((k-a)!)^2} \right) < \\ &< (2s^2)^{k-1} e^{k/s} \mathbf{E}\delta_{n,k}(s) \left(knp^s + I\{k \geq 3\} \frac{1}{3} (knp^s)^2 \right). \end{aligned}$$

Далее, при тех же условиях

$$\begin{aligned} \sum_{\bar{i} \in J_{s,n}} \sum_{\beta \in \Gamma'_k(\bar{i})} \mathbf{E}(\eta(\bar{i})\eta(\bar{j})) &< \mathbf{E}\delta_{n,k}(s) (2s^2)^{k-1} e^{k/s} p^{2s}, \\ \sum_{\bar{i} \in J_{s,n}} \sum_{\beta \in \Gamma''_a(\bar{i})} \mathbf{E}(\eta(\bar{i})\eta(\bar{j})) &= 0, \quad \sum_{\bar{i} \in J_{s,n}} \sum_{\beta \in \Gamma'''_k(\bar{i})} \mathbf{E}(\eta(\bar{i})\eta(\bar{j})) < C_n^k (2s^2)^{k-1} e^{k/s} p^{ks}. \end{aligned}$$

Из (18) и этих оценок следует, что

$$\begin{aligned} \Sigma_2 &= \sum_{\bar{i} \in J_{s,n}} \sum_{\beta \in \Gamma(\bar{i})} \mathbf{E}(\eta(\bar{i})\eta(\bar{j})) < \\ &< \mathbf{E}\delta_{n,k} (2s^2)^{k-1} e^{k/s} \left(knp^s + I\{k \geq 3\} \frac{1}{3} (knp^s)^2 + p^{2s} \right) + \frac{1}{k!} (2s^2)^{k-1} e^{k/s} (np^s)^k < \\ &< \mathbf{E}\delta_{n,k} e (2s^2)^{k-1} \left(knp^s + \frac{1}{2} (snp^s)^2 \right) + \frac{e}{k!} (2s^2)^{k-1} (np^s)^k. \end{aligned} \quad (23)$$

В последнем переходе использованы неравенства

$$I\{k \geq 3\} \frac{1}{3} (knp^s)^2 + p^{2s} \leq \frac{1}{2} (snp^s)^2, \quad e^{k/s} \leq e,$$

вытекающие из соотношений $s^2 n^2 > 6$ и $k \leq s$, следующих из условий теоремы.

Отметим, что в силу (23) найдется такая константа $C_2 < \infty$, при которой для всех k, s, n , $2 \leq k \leq s$, $knp^s < 1$, выполнено неравенство

$$\Sigma_2 < C_2 (2s^2)^{k-1} \left(\mathbf{E} \delta_{n,k} knp^s + \frac{1}{k!} (np^s)^k \right). \quad (24)$$

Теперь оценим величину

$$\Sigma_3 = \sum_{\bar{i} \in J_{s,n}} \mathbf{E} \left| \mathbf{E} \eta(\bar{i}) - \mathbf{E} \{ \eta(\bar{i}) | V_{\bar{i}} \} \right| \quad (25)$$

где, напомним,

$$V_{\bar{i}} = \sum_{\bar{j} \in J_{s,n} \setminus (\Gamma(\bar{i}) \cup \{\bar{i}\})} \eta(\bar{j}).$$

Зададим число $0 \leq v \leq |J_{s,n} \setminus (\Gamma(\bar{i}) \cup \{\bar{i}\})|$, величину

$$E(\bar{i}, v) = \left| \mathbf{E} \eta(\bar{i}) - \mathbf{E} \{ \eta(\bar{i}) | V_{\bar{i}} = v \} \right|, \quad (26)$$

и набор чисел

$$T(\bar{i}) = (i_1 - s^2, i_1 + s^2 + s - 1, i_2 - s^2, i_2 + s^2 + s - 1, \dots, i_k - s^2, i_k + s^2 + s - 1).$$

Положим

$$\begin{aligned} F(\bar{i}, (u_1, v_1, \dots, u_k, v_k)) = \\ = \left| \mathbf{E} \eta(\bar{i}) - \mathbf{E} \{ \eta(\bar{i}) | (X_t, t \in T(\bar{i})) = (u_1, v_1, \dots, u_k, v_k) \} \right|, \end{aligned} \quad (27)$$

где $u_1, v_1, \dots, u_k, v_k \in \{1, \dots, N\}$. Согласно марковскому свойству цепи $\mathbf{X}$

$$\max_v E(\bar{i}, v) \leq \max_{u_1, v_1, \dots, u_k, v_k} F(\bar{i}, (u_1, v_1, \dots, u_k, v_k)).$$

Поэтому

$$\Sigma_3 \leq C_n^k \max_{u_1, v_1, \dots, u_k, v_k} F(\bar{i}, (u_1, v_1, \dots, u_k, v_k)). \quad (28)$$

Введем величину

$$f(u_1, v_1, \dots, u_k, v_k) = \mathbf{E} \{ \eta(\bar{i}) | (X_t, t \in T(\bar{i})) = (u_1, v_1, \dots, u_k, v_k) \}. \quad (29)$$

Так как

$$\min_{u_1, v_1, \dots, u_k, v_k} f(u_1, v_1, \dots, u_k, v_k) \leq \mathbf{E} \eta(\bar{i}) \leq \max_{u_1, v_1, \dots, u_k, v_k} f(u_1, v_1, \dots, u_k, v_k),$$

то

$$\begin{aligned} F(\bar{i}, (u_1, v_1, \dots, u_k, v_k)) \leq \\ \max_{u_1, v_1, \dots, u_k, v_k} f(u_1, v_1, \dots, u_k, v_k) - \min_{u_1, v_1, \dots, u_k, v_k} f(u_1, v_1, \dots, u_k, v_k). \end{aligned} \quad (30)$$

Оценим функцию $f(u_1, v_1, \dots, u_k, v_k)$, используя марковское свойство цепи $\mathbf{X}$:

$$f(u_1, v_1, \dots, u_k, v_k) = \quad (31)$$

$$= \sum_{x_1, \dots, x_s \in \{1, \dots, N\}} \prod_{j=1}^k \mathbf{P}\{Y_{i_j}(s) = (x_1, \dots, x_s) | X_{i_j-s^2} = u_j, X_{i_j+s^2+s-1} = v_j\}.$$

Здесь

$$\begin{aligned} & \mathbf{P}\{Y_{i_j}(s) = (x_1, \dots, x_s) | X_{i_j-s^2} = u_j, X_{i_j+s^2+s-1} = v_j\} = \\ &= \frac{\mathbf{P}\{X_{i_j-s^2} = u_j, Y_{i_j}(s) = (x_1, \dots, x_s), X_{i_j+s^2+s-1} = v_j\}}{\mathbf{P}\{X_{i_j-s^2} = u_j, X_{i_j+s^2+s-1} = v_j\}}. \end{aligned} \quad (32)$$

Для краткости положим $w = e^{-c(\mathbb{P})s^2}$. Из (1) следует оценка

$$\begin{aligned} \mathbf{P}\{X_{i_j-s^2} = u_j, X_{i_j+s^2+s-1} = v_j\} &= \mathbf{P}\{X_{i_j-s^2} = u_j\} p_{u_j, v_j}^{(2s^2+s-1)} = \\ &= \mathbf{P}\{X_{i_j-s^2} = u_j\} \pi_{v_j} (1 + \theta_j w^2) \end{aligned}$$

при некотором $|\theta_j| \leq 1$. Аналогично,

$$\begin{aligned} & \mathbf{P}\{X_{i_j-s^2} = u_j, Y_{i_j}(s) = (x_1, \dots, x_s), X_{i_j+s^2+s-1} = v_j\} = \\ &= \mathbf{P}\{X_{i_j-s^2} = u_j\} p_{u_j, x_1}^{(s^2)} \mathbf{P}\{Y_{i_j}(s) = (x_1, \dots, x_s) | X_{i_j} = x_1\} p_{x_s, v_j}^{(s^2)} = \\ &= \mathbf{P}\{X_{i_j-s^2} = u_j\} \pi_{x_1} (1 + \theta'_j w) \mathbf{P}\{Y_{i_j}(s) = (x_1, \dots, x_s) | X_{i_j} = x_1\} \pi_{v_j} (1 + \theta''_j w). \end{aligned}$$

Подставив эти выражения в (32), получим

$$\begin{aligned} & \mathbf{P}\{Y_{i_j}(s) = (x_1, \dots, x_s) | X_{i_j-s^2} = u_j, X_{i_j+s^2+s-1} = v_j\} = \\ &= \frac{\pi_{x_1} \mathbf{P}\{Y_{i_j}(s) = (x_1, \dots, x_s) | X_{i_j} = x_1\} (1 + \theta'_j w) (1 + \theta''_j w)}{1 + \theta_j w^2}. \end{aligned} \quad (33)$$

Из (31) и (33) следует формула

$$\begin{aligned} & f(u_1, v_1, \dots, u_k, v_k) = \\ &= \sum_{x_1, \dots, x_s \in \{1, \dots, N\}} \prod_{j=1}^k \frac{\pi_{x_1} \mathbf{P}\{Y_{i_j}(s) = (x_1, \dots, x_s) | X_{i_j} = x_1\} (1 + \theta'_j w) (1 + \theta''_j w)}{1 + \theta_j w^2}. \end{aligned} \quad (34)$$

Чтобы упростить последнюю формулу, воспользуемся следующими непосредственно проверяемыми утверждениями.

Лемма 1. Пусть $0 < w < 1$. Тогда для всякого числа θ , $|\theta| \leq 1$, найдется такое число $\tilde{\theta}$, $|\tilde{\theta}| \leq 1/2$, что

$$\frac{1}{1 + \theta w^2} = 1 + \tilde{\theta} w.$$

Лемма 2. Пусть $0 < w < 1$ и $m \in \{2, 3, \dots\}$. Тогда для любого набора чисел $\theta_1, \dots, \theta_m$, $|\theta_j| \leq 1$ ($j = 1, \dots, m$), найдется такое число $\tilde{\theta}$, $|\tilde{\theta}| \leq 1$, что

$$\prod_{j=1}^m (1 + \theta_j w) = (1 + \tilde{\theta} w)^m.$$

Эти леммы позволяют переписать (34) в следующем виде

$$f(u_1, v_1, \dots, u_k, v_k) = (1 + \tilde{\theta}w)^{3k} \sum_{x_1, \dots, x_s \in \{1, \dots, N\}} \prod_{j=1}^k \pi_{x_1} \mathbf{P}\{Y_{i_j}(s) = (x_1, \dots, x_s) | X_{i_j} = x_1\}. \quad (35)$$

Из (30) и (35) получаем

$$F(\tilde{i}, (u_1, v_1, \dots, u_k, v_k)) \leq \left| (1 + \tilde{\theta}'w)^{3k} - (1 + \tilde{\theta}''w)^{3k} \right| \sum_{x_1, \dots, x_s \in \{1, \dots, N\}} \prod_{j=1}^k \pi_{x_1} \mathbf{P}\{Y_{i_j}(s) = (x_1, \dots, x_s) | X_{i_j} = x_1\},$$

где $|\tilde{\theta}'|, |\tilde{\theta}''| \leq 1$. Следовательно,

$$\begin{aligned} F(\tilde{i}, (u_1, v_1, \dots, u_k, v_k)) &\leq (1 + w)^{3k} - (1 - w)^{3k} = 2w \sum_{l=0}^{[(3k-1)/2]} C_{3k}^{2l+1} w^{2l} < \\ &< 2w \sum_{l=0}^{\infty} \frac{(3k)^{2l+1}}{(2l+1)!} w^{2l} < 6kw \sum_{l=0}^{\infty} \frac{(3kw)^{2l}}{(2l+1)!} < 6kw \sum_{l=0}^{\infty} \frac{1}{(2l+1)!} < 8kw \end{aligned} \quad (36)$$

(в оценках использовано неравенство $3kw \leq 1$, вытекающее из условия $\ln n < c(\mathbb{P})s^\gamma$). Используя (35), (36) и (28), получаем, что $\Sigma_3 \leq 8kC_n^k w$.

Воспользуемся теперь условиями теоремы 3 о том, что для некоторого $1 < \gamma < 2$ при $s \geq \gamma^{-1} \sqrt{\ln(p)/c(\mathbb{P})}$ выполнены неравенства $s^{2-\gamma} \geq 2k$ и $\ln n < c(\mathbb{P})s^\gamma$, и вспомним, что $w = e^{-c(\mathbb{P})s^2}$. Тогда

$$C_n^k w < ne^{c(\mathbb{P})((k-1)s^\gamma - s^2)} < ne^{-c(\mathbb{P})(k-1)s^\gamma} \leq ne^{-|\ln(p)|s} = np^s.$$

Значит,

$$\Sigma_3 \leq 8knp^s. \quad (37)$$

Подставим оценки (22), (24) и (37) в формулу (19). При этом используем неравенство $(1 - e^{-\mathbf{E}\delta_{n,k}(s)})\mathbf{E}\delta_{n,k}(s)^{-1} < \mathbf{E}\delta_{n,k}(s)^{-1}$ в тех случаях, когда подставляемая оценка для Σ_1 и часть оценки для Σ_2 содержат множитель $\mathbf{E}\delta_{n,k}(s)$, и используем неравенство $(1 - e^{-\mathbf{E}\delta_{n,k}(s)})\mathbf{E}\delta_{n,k}(s)^{-1} < 1$ в том случае, когда в части оценки для Σ_2 такого множителя нет. Получаем

$$\begin{aligned} \rho(\mathcal{L}(\delta_{n,k}(s)), \text{Po}(\mathbf{E}\delta_{n,k}(s))) &\leq \frac{1 - e^{-\mathbf{E}\delta_{n,k}(s)}}{\mathbf{E}\delta_{n,k}(s)} (\Sigma_1 + \Sigma_2) + \min\left(1, \sqrt{\frac{2}{e\mathbf{E}\delta_{n,k}(s)}}\right) \Sigma_3 < \\ &< C_1 s^2 (np^s)^{k-1} + C_2 (2s^2)^{k-1} knp^s + \frac{C_2}{k!} (2s^2)^{k-1} k(np^s)^k + 8knp^s < \\ &C_1 s^2 (np^s)^{k-1} + 2C_2 (2s^2)^{k-1} knp^s + 8knp^s. \end{aligned} \quad (38)$$

3. Из соотношений (15), (17), (38) и неравенства треугольника следует, что

$$\rho(\mathcal{L}(\tilde{\xi}_{n,2}(s)), \text{Po}(\tilde{\xi}_{n,2}(s))) < (C_1 + 2C_2 + 8)s^3 np^s. \quad (39)$$

Из соотношений (16), (17), (38) и неравенства треугольника следует, что при $k \geq 3$

$$\rho\left(\mathcal{L}(\tilde{\xi}_{n,k}(s)), \text{Po}(\mathbf{E}\tilde{\xi}_{n,k}(s))\right) < (2C_0 + C_1 + 2C_2 + 2)(2s^2)^{k-1} knp^s. \quad (40)$$

Здесь C_0, C_1, C_2 — константы, введенные в соотношениях (22), (24) и (38). Теорема 1 доказана.

Доказательства следствий проводятся стандартными рассуждениями (см., например, [9]), и мы их не приводим.

Автор признателен А. М. Зубкову и А. М. Шойтову за интерес к работе и полезные замечания.

Список литературы

1. Беляев П.Ф., “О совместном распределении частот длинных s -цепочек в мультиномиальной схеме с равновероятными исходами”, *Теория вероятн. и ее примен.*, **14**:3 (1969), 540-546.
2. Зубков А.М., Михайлов В.Г., “Предельные распределения случайных величин, связанных с длинными повторениями в последовательности независимых испытаний”, *Теория вероятн. и ее примен.*, **19**:1 (1974), 173-181.
3. Зубков А.М., Михайлов В.Г., “О повторениях s -цепочек в последовательности независимых величин.”, *Теория вероятн. и ее примен.*, **24**:2 (1979), 267-279.
4. Михайлов В.Г., “Предельные распределения случайных величин, связанных с многократными длинными повторениями в последовательности независимых испытаний”, *Теория вероятн. и ее примен.*, **19**:1 (1974), 182-187.
5. Hansen N. R., “Local alignment of Markov chains”, *Ann. Appl. Probab.*, **16**:3 (2006), 1262-1296.
6. Беляев П.Ф., “К вопросу о совместном распределении частот s -цепочек в сложных цепях Маркова с большим числом состояний”, *Теория вероятн. и ее примен.*, **14**:2 (1969), 333-339.
7. Беляев П.Ф., “О совместном распределении частот исходов в цепях Маркова с большим числом состояний”, *Теория вероятн. и ее примен.*, **22**:3 (1977), 534-545.
8. Зубков А.М., “Неравенства для вероятностей переходов с запрещениями и их применения”, *Матем. сб.*, **109(151)**:4(8) (1979), 491-532.
9. Михайлов В.Г., Шойтов А.М., “О длинных повторениях цепочек в цепи Маркова”, *Дискретная математика*, **26**:3 (2014), 79-89.
10. Михайлов В.Г., Шойтов А.М., “Многократные повторения длинных цепочек в конечной цепи Маркова”, *Математические вопросы криптографии*, **6**:3 (2015).
11. Гантмахер Ф.Р., *Теория матриц*, 5-е изд., М.: ФИЗМАТЛИТ, 2004, 560 с.
12. Шойтов А.М., “Пуассоновское приближение для числа повторений значений дискретной функции от цепочек”, *Дискретная математика*, **17**:2 (2005), 56-69.
13. Barbour A.D., Holst L., Janson S., *Poisson Approximation*, Oxford University Press, 1992, 277 pp.
14. Erhardsson T., “Stein’s method for Poisson and compound Poisson approximation”, In “*An introduction to Stein’s method*”, eds. Barbour A. D., Chen L. H. Y., Singapore Univ. Press, 2005, 61–113.