Math-Net.Ru

Общероссийский математический портал

А. М. Шойтов, Сложное распределение Пуассона для числа повторений значений дискретной функции от цепочек, *Дискрет. матем.*, 2007, том 19, выпуск 2, 6–26

DOI: <https://doi.org/10.4213/dm16>

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: 52.6.47.48

26 апреля 2023 г., 14:19:38

Сложное распределение Пуассона для числа повторений значений дискретной функции от цепочек

© 2007 г. А. М. Шойтов

Для последовательности $\mathbf{X} = \{X_1, \dots, X_n, \dots\}$ независимых одинаково распределенных случайных величин строятся s -цепочки $Y_i(s) = (X_i, \dots, X_{i+s-1})$, $i = 1, 2, \dots, n$, и рассматриваются случайные величины $\mathbf{F}_i = f(Y_i(s))$, $i = 1, 2, \dots$, где f — функция, заданная на множестве $\mathbf{R}^s$ и принимающая значения из множества натуральных чисел.

В статье рассматривается последовательность $\mathbf{F} = \{\mathbf{F}_1, \mathbf{F}_2, \dots\}$ и изучаются две случайные величины, случайная величина

$$\mathbf{Z}_n(\mathbf{F}) = \sum_{1 \leq i_1 < i_2 \leq n} \mathbf{I}\{\mathbf{F}_{i_1} = \mathbf{F}_{i_2}\},$$

равная числу повторений символов на отрезке длины n последовательности $\mathbf{F}$ (здесь $\mathbf{I}\{\cdot\}$ обозначает индикатор случайного события), и случайная величина

$$\mathbf{Z}'_n(\mathbf{F}) = \sum_{1 \leq i_1 < i_1+s \leq i_2 \leq n} \mathbf{I}\{\mathbf{F}_{i_1} = \mathbf{F}_{i_2}\},$$

равная числу повторений значений функции f от неперекрывающихся s -цепочек отрезка последовательности $\mathbf{X}$ длины $n + s - 1$.

В работы методом Стейна установлены достаточные условия сходимости к сложному распределению Пуассона распределения случайных величин $\mathbf{Z}_n(\mathbf{F})$ и $\mathbf{Z}'_n(\mathbf{F})$ для функции f общего вида. Следствиями этих результатов являются как известные, так и новые предельные теоремы для числа повторений значений функции от цепочек полиномиальной схемы для ряда конкретных типов функций f .

1. Введение

Пусть $\mathbf{X} = \{X_1, \dots, X_n, \dots\}$ — последовательность независимых одинаково распределенных случайных величин. Для натурального s будем рассматривать s -цепочки $Y_i(s) = (X_i, \dots, X_{i+s-1})$, $i = 1, 2, \dots, n$. Будем считать, что на множестве $\mathbf{R}^s$ задана борелевская функция f , принимающая значения из множества натуральных чисел.

Положим $\mathbf{F} = \{\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_n, \dots\}$, $\mathbf{F}_i = f(Y_i(s))$, и определим две случайные величины, случайную величину

$$\mathbf{Z}_n(\mathbf{F}) = \sum_{1 \leq i_1 < i_2 \leq n} \mathbf{I}\{\mathbf{F}_{i_1} = \mathbf{F}_{i_2}\},$$

равную числу повторений символов на отрезке длины n последовательности $\mathbf{F}$ (здесь и далее выражение $\mathbf{I}\{\cdot\}$ обозначает индикатор случайного события или неслучайного множества), и случайную величину

$$Z'_n(\mathbf{F}) = \sum_{1 \leq i_1 < i_1+s \leq i_2 \leq n} \mathbf{I}\{\mathbf{F}_{i_1} = \mathbf{F}_{i_2}\},$$

равную числу повторений значений функции f от неперекрывающихся s -цепочек отрезка последовательности $\mathbf{X}$ длины $n + s - 1$.

Случайная величина $Z_n(\mathbf{F})$ может быть рассмотрена как индикаторная U -статистика от последовательности $(s-1)$ -зависимых случайных величин (см. [6]) или как сумма индикаторов с малым числом зависимостей (см. [9]). Общие условия асимптотической нормальности таких величин получены в работах [7] и [8]. В [15] доказана многомерная (по различным функциям f) центральная предельная теорема для случайных величин $Z_n(\mathbf{F})$ в ситуации, когда растет лишь длина последовательности $\mathbf{X}$.

Рассмотрим отношение эквивалентности $\sim_f$ на множестве s -цепочек, заданное следующим образом: две цепочки принадлежат одному классу $\sim_f$ -эквивалентности, если функция f принимает на них одинаковое значение. Нетрудно видеть, что функции f , для которых отношения $\sim_f$ совпадают, определяют одинаковые случайные величины $Z_n(\mathbf{F})$ и $Z'_n(\mathbf{F})$. Поэтому можно говорить об эквивалентности задач исследования предельного распределения случайной величины $Z_n(\mathbf{F})$ ($Z'_n(\mathbf{F})$) и случайной величины, равной числу пар (числу неперекрывающихся пар) эквивалентных (относительно некоторого отношения эквивалентности, имеющего не более чем счетное множество классов эквивалентности) цепочек отрезка последовательности $\mathbf{X}$ длины $n + s - 1$.

Исследованию достаточных условий сходимости распределения случайных величин $Z_n(\mathbf{F})$ и $Z'_n(\mathbf{F})$ к сложному распределению Пуассона для конкретных видов дискретных функций f (или конкретных видов отношений эквивалентности $\sim_f$) посвящен целый ряд работ, начиная с известной статьи [3] А. М. Зубкова и В. Г. Михайлова об асимптотическом поведении числа точных повторений пар цепочек полиномиальной схемы (см. также [4] и [10]). Кроме задачи о точных повторениях цепочек отметим активно развивающееся в последнее десятилетие направление, связанное с исследованием асимптотического поведения числа пар структурно или перестановочно эквивалентных цепочек (см. [2], [17], [18] и обзор [12]).

В разделе 3 настоящей работы методом Стейна (см. [22]–[24]) установлены достаточные условия сходимости к сложному распределению Пуассона распределения случайной величины $Z_n(\mathbf{F})$, а также случайной величины $Z'_n(\mathbf{F})$, для дискретной функции f общего вида. Следствием этих результатов являются как известные, так и новые предельные сложнопуассоновские теоремы для числа повторений значений функции от цепочек полиномиальной схемы для ряда конкретных типов функций f .

В частности, в разделе 4.1 приведен известный результат А. М. Зубкова и В. Г. Михайлова (см., например, [3], [4] и [10]) о сходимости к сложному закону Пуассона распределения числа точных повторений цепочек полиномиальной схемы, доказательство которого легко получается с помощью результатов раздела 3. В разделе 4.2 приведен известный результат о предельном распределении числа пар $\mathbf{H}$ -эквивалентных (см. [13]) цепочек полиномиальной схемы (см. [16] и [18]). В разделе 4.3 рассмотрены условия сходимости к сложному распределению Пуассона распределения числа пар неперекрывающихся $\mathbf{G}$ -эквивалентных (см. [13]) цепочек полиномиальной схемы, в качестве следствий из результатов раздела 3 выведены новые сложнопуассоновская и нормальная предельные

теоремы для цепочек с одинаковыми векторами частот встречаемости символов (то есть перестановочно эквивалентных цепочек).

2. Обозначения

Условимся обозначать $\rho(\Psi, \Gamma)$ расстояние по вариации между распределениями Ψ и Γ . Для распределений Ψ и Γ , сосредоточенных на множестве $\{0, 1, \dots\}$, справедливо равенство

$$\rho(\Psi, \Gamma) = \frac{1}{2} \sum_{m=0}^{\infty} |\Psi\{m\} - \Gamma\{m\}|.$$

Распределение случайной величины γ будем обозначать $\mathcal{L}(\gamma)$. Пусть $\{\lambda_1, \lambda_2, \dots\}$ — последовательность неотрицательных действительных чисел, образующая сходящийся ряд $\sum_{k=1}^{\infty} \lambda_k < \infty$. Пусть $\{N_1, N_2, \dots\}$ — последовательность независимых случайных величин, причем случайная величина N_k имеет распределение Пуассона с параметром λ_k , $k = 1, 2, \dots$. Обозначим через $\text{CP}(\Lambda)$ сложное распределение Пуассона, совпадающее с распределением случайной величины

$$N = \sum_{k=1}^{\infty} k N_k,$$

причем

$$\Lambda = \sum_{k=1}^{\infty} \lambda_k \delta_k,$$

δ_k — вероятностная мера, сосредоточенная в точке k . Кроме того, пусть

$$\lambda = \mathbf{E} \sum_{k=1}^{\infty} k N_k = \sum_{k=1}^{\infty} k \lambda_k$$

(возможно, $\lambda = \infty$). Тогда распределению $\text{CP}(\Lambda)$ соответствует производящая функция

$$\mathbf{E} z^N = \exp \left(\lambda_0 \left(\sum_{k=1}^{\infty} z^k \lambda_k \lambda_0^{-1} - 1 \right) \right), \quad \lambda_0 = \sum_{k=1}^{\infty} \lambda_k.$$

Пусть $\nu_x(Y_i(s))$ — частота встречаемости символа x , $x \in \mathbf{R}$, в цепочке $Y_i(s)$, то есть

$$\nu_x(Y_i(s)) = \sum_{r=1}^s \mathbf{I}\{X_{i+r-1} = x\}.$$

Для $r = 1, \dots, s$ положим

$$\pi_r(f) = \mathbf{P}\{\mathbf{F}_1 = \mathbf{F}_{1+r}\}, \quad \kappa_r(f) = \max_{a, b \in \mathbf{N}} \mathbf{P}\{\mathbf{F}_1 = a \mid \mathbf{F}_{1+r} = b\}.$$

Нетрудно видеть, что

$$\pi_s(f) = \sum_{a \in \mathbf{N}} \mathbf{P}^2\{\mathbf{F}_1 = a\}, \quad \kappa_s(f) = \max_{a \in \mathbf{N}} \mathbf{P}\{\mathbf{F}_1 = a\}, \quad \mathbf{E} \mathbf{Z}'_n(\mathbf{F}) = \binom{n-s+1}{2} \pi_s(f).$$

3. Основные результаты

Положим

$$J_r = \{\bar{i} = (i_1, i_2): 1 \leq i_1 < i_1 + rn \leq i_2 \leq n\}, \quad r = 1, 2, 3,$$

и для любой пары $\bar{i} = (i_1, i_2)$ из J_2 определим множество

$$\Gamma_{\bar{i}}^{nvs} = \{\bar{j} \in J_2 \setminus \{\bar{i}\}: \max\{|i_1 - j_1|, |i_2 - j_2|\} < s\}.$$

Для $\Gamma_{\bar{i}}^{vs} \subseteq \Gamma_{\bar{i}}^{nvs}$, $\bar{i} \in J_2$, положим

$$\begin{aligned} \Gamma^{vs} &= \{(\bar{i}, \Gamma_{\bar{i}}^{vs}), \bar{i} \in J_2\}, & U(\Gamma_{\bar{i}}^{vs}) &= \sum_{\bar{j} \in \Gamma_{\bar{i}}^{vs}} \mathbf{I}\{\mathbf{F}_{j_1} = \mathbf{F}_{j_2}\}, \\ V(\Gamma_{\bar{i}}^{vs}) &= \sum_{\bar{j} \in \Gamma_{\bar{i}}^{nvs} \setminus \Gamma_{\bar{i}}^{vs}} \mathbf{I}\{\mathbf{F}_{j_1} = \mathbf{F}_{j_2}\}, & \psi(\Gamma^{vs}) &= \sum_{\bar{i} \in J_2} \mathbf{E}\{\mathbf{I}\{\mathbf{F}_{i_1} = \mathbf{F}_{i_2}\} V(\Gamma_{\bar{i}}^{vs})\}. \end{aligned}$$

Рассмотрим сложное распределение Пуассона $\text{CP}(\Lambda_n(\Gamma^{vs}))$, где

$$\begin{aligned} \Lambda_n(\Gamma^{vs}) &= \sum_{k=1}^{(2s-1)^2} \lambda_k(\Gamma^{vs}) \delta_k, \\ \lambda_k(\Gamma^{vs}) &= k^{-1} \sum_{\bar{i} \in J_2} \mathbf{E}\{\mathbf{I}\{\mathbf{F}_{i_1} = \mathbf{F}_{i_2}\} \mathbf{I}\{\mathbf{I}\{\mathbf{F}_{i_1} = \mathbf{F}_{i_2}\} + U(\Gamma_{\bar{i}}^{vs}) = k\}\} \end{aligned}$$

при $k \leq (2s-1)^2$ и $\lambda_k(\Gamma^{vs}) = 0$ при $k > (2s-1)^2$.

Следующая теорема дает оценку близости распределений случайных величин $\mathbf{Z}_n(\mathbf{F})$ и $\mathbf{Z}'_n(\mathbf{F})$ и сложного распределения Пуассона $\text{CP}(\Lambda_n(\Gamma^{vs}))$.

Теорема 1. Пусть $n \geq 2s + 1$. Тогда

$$(a) \quad \rho(\mathcal{L}(\mathbf{Z}_n(\mathbf{F})), \mathcal{L}(\mathbf{Z}'_n(\mathbf{F}))) \leq n \sum_{r=1}^{s-1} \pi_r(f),$$

(б) и для любого выбора множества Γ^{vs} , $\Gamma_{\bar{i}}^{vs} \subseteq \Gamma_{\bar{i}}^{nvs}$, $\bar{i} \in J_2$, справедливы оценка

$$\begin{aligned} \rho(\mathcal{L}(\mathbf{Z}'_n(\mathbf{F})), \text{CP}(\Lambda_n(\Gamma^{vs}))) \\ \leq 3ns\pi_s(f) + c(\Gamma^{vs})(\mathbf{E}\mathbf{Z}'_n(\mathbf{F})sn(25\pi_s(f) + 8\kappa_s(f)) + \psi(\Gamma^{vs})), \end{aligned}$$

где

$$c(\Gamma^{vs}) \leq \min \left\{ 1, \frac{1}{\lambda_1(\Gamma^{vs})} \right\} \exp \left\{ \sum_{k=1}^{(2s-1)^2} \lambda_k(\Gamma^{vs}) \right\},$$

причем, если

$$k\lambda_k(\Gamma^{vs}) > (k+1)\lambda_{k+1}(\Gamma^{vs}), \quad k = 1, \dots, (2s-1)^2 - 1,$$

или

$$\sum_{k=1}^{(2s-1)^2} k^2 \lambda_k(\Gamma^{vs}) < \frac{3}{2} \sum_{k=1}^{(2s-1)^2} k \lambda_k(\Gamma^{vs}),$$

то

$$c(\Gamma^{vs}) = \min\{1, (\lambda_1(\Gamma^{vs}) - 2\lambda_2(\Gamma^{vs}))^{-1}((4(\lambda_1(\Gamma^{vs}) - 2\lambda_2(\Gamma^{vs})))^{-1} + (\ln 2(\lambda_1(\Gamma^{vs}) - 2\lambda_2(\Gamma^{vs})))_+)\}.$$

$$(в) \left| \mathbf{E}\mathbf{Z}'_n(\mathbf{F}) - \sum_{k=1}^{(2s-1)^2} k\lambda_k(\Gamma^{vs}) \right| \leq sn\pi_s(f);$$

(г) справедливо неравенство

$$\left| \mathbf{D}\mathbf{Z}'_n(\mathbf{F}) - \sum_{k=1}^{(2s-1)^2} k^2\lambda_k(\Gamma^{vs}) \right| \leq 5ns^3\pi_s(f) + \mathbf{E}\mathbf{Z}'_n(\mathbf{F})sn(25\pi_s(f) + 8\kappa_s(f)) + \left(\binom{n-2s+1}{2} + (2s-1)^2 \right) \psi(\Gamma^{vs}).$$

Из теоремы 1 вытекает следующая теорема, описывающая достаточные условия сходимости распределений случайных величин $\mathbf{Z}_n(\mathbf{F})$ и $\mathbf{Z}'_n(\mathbf{F})$ к сложному распределению Пуассона $\text{CP}(\Lambda_n(\Gamma^{vs}))$.

Теорема 2. Если при $n \rightarrow \infty$ параметр s , функция f , распределение случайной последовательности $\mathbf{X}$ и множество Γ^{vs} меняются так, что существуют пределы

$$\begin{aligned} \lambda_k &= \lim_{n \rightarrow \infty} \lambda_k(\Gamma^{vs}), & k &= 1, 2, \dots, \\ \lambda_0 &= \lim_{n \rightarrow \infty} \sum_{k=1}^{(2s-1)^2} \lambda_k(\Gamma^{vs}), & 0 < \lambda_0 < \infty, \end{aligned}$$

и выполнено условие

$$\mathbf{E}\mathbf{Z}'_n(\mathbf{F})sn\kappa_s(f) + \psi(\Gamma^{vs}) \rightarrow 0, \quad (1)$$

то распределение случайной величины $\mathbf{Z}'_n(\mathbf{F})$ сходится к распределению $\text{CP}(\Lambda)$, определяемому последовательностью $\{\lambda_1, \lambda_2, \dots\}$, причем

$$\rho(\mathcal{L}(\mathbf{Z}'_n(\mathbf{F})), \text{CP}(\Lambda_n(\Gamma^{vs}))) = O(\mathbf{E}\mathbf{Z}'_n(\mathbf{F})sn\kappa_s(f) + \psi(\Gamma^{vs})).$$

Если при этом выполнено условие

$$n \sum_{r=1}^{s-1} \pi_r(f) \rightarrow 0, \quad (2)$$

то распределение случайной величины $\mathbf{Z}_n(\mathbf{F})$ также сходится к распределению $\text{CP}(\Lambda)$, причем

$$\rho(\mathcal{L}(\mathbf{Z}_n(\mathbf{F})), \text{CP}(\Lambda_n(\Gamma^{vs}))) = O\left(n \left(\mathbf{E}\mathbf{Z}'_n(\mathbf{F})s\kappa_s(f) + \sum_{r=1}^{s-1} \pi_r(f) \right) + \psi(\Gamma^{vs})\right).$$

Замечание 1. В условиях теоремы 2 допустим случай, когда $\mathbf{EZ}'_n(\mathbf{F}) \rightarrow \infty$ (соответствующий пример приведен в конце статьи). Если же существует $\overline{\lim}_{n \rightarrow \infty} \mathbf{EZ}'_n(\mathbf{F}) < \infty$, то условие (1) эквивалентно условию

$$s\kappa_s(f)\pi_s(f)^{-1/2} + \psi(\Gamma^{vs}) \rightarrow 0, \quad (3)$$

а (2) — условию

$$\pi_s(f)^{-1/2} \sum_{r=1}^{s-1} \pi_r(f) \rightarrow 0. \quad (4)$$

Замечание 2. Асимптотическое поведение случайных величин $\mathbf{Z}_n(\mathbf{F})$ и $\mathbf{Z}'_n(\mathbf{F})$ может различаться, если не выполнено условие (2) теоремы 2. Так будет, если вероятности совпадения значений функции f от неперекрывающихся цепочек убывают по порядку существенно быстрее, чем значения f от перекрывающихся цепочек. Содержательный пример такой ситуации содержится в [17] и [19].

Замечание 3. Выбор множества Γ^{vs} , для которого может быть установлено сближение распределения $\mathbf{Z}'_n(\mathbf{F})$ и $\text{CP}(\Lambda_n(\Gamma^{vs}))$, вообще говоря, неоднозначен и зависит от вида функции f (см. раздел 4).

Теорема 1 позволяет получить достаточные условия сходимости распределения случайной величины $\mathbf{Z}'_n(\mathbf{F})$ к нормальному распределению.

Теорема 3. Пусть при $n \rightarrow \infty$ параметр s , функция f и распределение случайной последовательности $\mathbf{X}$ меняются так, что выполнены условия

$$\overline{\lim}_{n \rightarrow \infty} \frac{\mathbf{DZ}'_n(\mathbf{F})}{\mathbf{EZ}'_n(\mathbf{F})} < \frac{3}{2}, \quad (5)$$

$$sn\kappa_s(f) \ln \mathbf{EZ}'_n(\mathbf{F}) \rightarrow 0, \quad (6)$$

$$\lim_{n \rightarrow \infty} \frac{s^{4+\varepsilon}}{\mathbf{EZ}'_n(\mathbf{F})} = 0 \quad (7)$$

для некоторого положительного ε . Тогда распределение случайной величины

$$\frac{\mathbf{Z}'_n(\mathbf{F}) - \mathbf{EZ}'_n(\mathbf{F})}{\sqrt{\mathbf{DZ}'_n(\mathbf{F})}}$$

сходится к стандартному нормальному закону.

Замечание 4. Условия (6) и (7) теоремы 3 могут быть несколько ослаблены в том случае, когда известны главные члены асимптотического разложения величин $\lambda_k(\Gamma^{vs})$, $k = 1, 2, \dots, (2s-1)^2$. Так, условие (7) можно заменить на более слабое условие: для некоторого натурального $r_0 \geq 3$ и всех $r \geq r_0$ выполнены соотношения

$$\lim_{n \rightarrow \infty} \sum_{k=1}^{(2s-1)^2} k^r \lambda_k(\Gamma^{vs}) / \left(\sum_{k=1}^{(2s-1)^2} k^2 \lambda_k(\Gamma^{vs}) \right)^{r/2} = 0. \quad (8)$$

Кроме того, если вместо условия (5) использовать условие

$$k\lambda_k(\Gamma^{vs}) \downarrow 0 \quad (9)$$

при $k \uparrow \infty$, то условие (6) можно заменить на более слабое условие

$$c(\Gamma^{vs})sn(\mathbf{EZ}'_n(\mathbf{F})\kappa_s(f) + \psi(\Gamma^{vs})) \rightarrow 0. \quad (10)$$

4. Следствия из основных результатов

Из теоремы 2 легко получается теорема, описывающая достаточные условия сходимости распределения случайных величин $\mathbf{Z}_n(\mathbf{F})$ и $\mathbf{Z}'_n(\mathbf{F})$ к обычному распределению Пуассона.

Теорема 4. Если при $n \rightarrow \infty$ параметр s , функция f и распределение случайной последовательности $\mathbf{X}$ меняются так, что $\mathbf{E}\mathbf{Z}'_n(\mathbf{F}) \rightarrow \lambda$, $0 < \lambda < \infty$, и выполнено условие

$$sn\kappa_s(f) + \sum_{\bar{i} \in J_2} \sum_{\bar{j} \in \Gamma_{\bar{i}}^{nvs}} \mathbf{E}\{\mathbf{I}\{\mathbf{F}_{i_1} = \mathbf{F}_{i_2}\} \mathbf{I}\{\mathbf{F}_{j_1} = \mathbf{F}_{j_2}\}\} \rightarrow 0,$$

то распределение случайной величины $\mathbf{Z}'_n(\mathbf{F})$ сходится к распределению Пуассона с параметром λ . Если при этом выполнено условие

$$n \sum_{r=1}^{s-1} \pi_r(f) \rightarrow 0,$$

то распределение случайной величины $\mathbf{Z}_n(\mathbf{F})$ также сходится к распределению Пуассона с параметром λ .

Отметим, что для исследования пуассоновской сходимости распределений случайных величин $\mathbf{Z}'_n(\mathbf{F})$ и $\mathbf{Z}_n(\mathbf{F})$ лучше использовать метод Чена–Стейна (см., например, [9], [21]), который позволяет получать более точные оценки скорости сближения соответствующих распределений. В [19] этим методом установлены достаточные условия асимптотической пуассоновости распределений $\mathbf{Z}_n(\mathbf{F})$ и $\mathbf{Z}'_n(\mathbf{F})$. Получены оценки скорости сближения распределений. Исследованию условий сходимости этой случайной величины к распределению Пуассона для некоторых конкретных функций f посвящены, например, работы [1], [3], [4], [12], [16] и [18].

Приведем примеры использования теорем 2 и 3 соответственно для описания достаточных условий сходимости случайных величин $\mathbf{Z}_n(\mathbf{F})$ и $\mathbf{Z}'_n(\mathbf{F})$ к сложному распределению Пуассона и к нормальному распределению для некоторых частных случаев функций f .

4.1. Совпадающие цепочки

Будем предполагать, что последовательность $\mathbf{X}$ получена по полиномиальной схеме испытаний с вероятностями исходов $p_m = \mathbf{P}\{X_1 = m\}$, $m = 1, 2, \dots$. Рассмотрим случайную величину

$$\delta_n = \sum_{1 \leq i_1 < i_2 \leq n} \mathbf{I}\{Y_{i_1}(s) = Y_{i_2}(s)\},$$

равную числу точных повторений s -цепочек на отрезке последовательности $\mathbf{X}$ длины $n + s - 1$. Подобные случайные величины исследовались в работах А. М. Зубкова и В. Г. Михайлова (см., например, [3], [4], [10]). В качестве следствия из теоремы 2 может быть получен известный результат (см. [10]) о сходимости распределения δ_n к сложному распределению Пуассона.

Следствие 1. Если при $n \rightarrow \infty$ параметр s и вероятности $p_1, p_2, \dots$ меняются так,

что

$$\sum_{m=1}^{\infty} p_m^2 \rightarrow \rho, \quad 0 \leq \rho < 1,$$

$$\binom{n}{2} \left(\sum_{m=1}^{\infty} p_m^2 \right)^s (1 - \rho) \rightarrow \lambda, \quad 0 < \lambda < \infty,$$

и выполнено условие

$$ns(\max_{m \in \mathbb{N}} p_m)^s \rightarrow 0,$$

то

$$\mathbf{E} z^{\delta_n} \rightarrow \exp \left\{ \lambda \left(\frac{(1 - \rho)z}{1 - \rho z} - 1 \right) \right\}.$$

Для доказательства следствия 1 функция f в теореме 2 определяется так, чтобы она принимала одинаковые натуральные значения только для одинаковых цепочек. Нетрудно проверить, что в таком случае справедливы равенства

$$\kappa_s(f) = (\max_{m \in \mathbb{N}} p_m)^s, \quad \pi_s(f) = \left(\sum_{m=1}^{\infty} p_m^2 \right)^s.$$

Кроме того, нетрудно доказывается (см. [2]), что

$$\pi_r(f) \leq (\max_{m \in \mathbb{N}} p_m)^s.$$

Далее, полагая

$$\Gamma_{\bar{i}}^{vs} = \{\bar{j} \in \Gamma_{\bar{i}}^{nvs} : i_1 - j_1 = i_2 - j_2\}, \bar{i} \in J_2,$$

можно убедиться в справедливости асимптотических равенств

$$\lim_{n \rightarrow \infty} \lambda_k(\Gamma^{vs}) = \lambda(1 - \rho)\rho^{k-1}, \quad k = 1, 2, \dots$$

Бесконечная малость величины $\psi(\Gamma^{vs})$ требует отдельного доказательства, которое повторяет соответствующие рассуждения работ [3] и [10].

4.2. Цепочки с одинаковой структурой

Будем предполагать, что последовательность $\mathbf{X}$ получена по полиномиальной схеме с исходами в множестве $A = \{1, \dots, N\}$. Пусть $\mathbf{H}$ — подгруппа симметрической группы подстановок S_N . Две s -цепочки $\bar{a} = (a_1, \dots, a_s)$ и $\bar{b} = (b_1, \dots, b_s)$ из A^s называются $\mathbf{H}$ -эквивалентными (см. монографию [13]), если существует такая подстановка h из $\mathbf{H}$, что для всех $j = 1, \dots, s$ справедливо равенство $h(a_j) = b_j$. В этом случае будем писать $\bar{a} \mathbf{H} \bar{b}$. Если $\mathbf{H} = S_N$, то будем называть такие цепочки структурно эквивалентными (см. обзор [12]). Нетрудно видеть, что для $\mathbf{H}$ -эквивалентных цепочек $\bar{a}$ и $\bar{b}$ справедливо свойство

$$a_r = a_q \iff b_r = b_q, \quad r, q = 1, \dots, s,$$

которое превращается в критерий для структурной эквивалентности цепочек.

Определим случайную величину

$$\xi_n(\mathbf{H}) = \sum_{1 \leq i_1 < i_2 \leq n} \mathbf{I}\{Y_{i_1}(s) \mathbf{H} Y_{i_2}(s)\},$$

равную числу пар $\mathbf{H}$ -эквивалентных s -цепочек в отрезке последовательности $\mathbf{X}$ длины $n + s - 1$. Подобные случайные величины исследовались в работах С. М. Буравлева (см. [1], [2]), В. Г. Михайлова (см., например, [11]) и автора (см., например, [16], [18]), подробное описание известных результатов об асимптотическом поведении таких величин можно найти в обзоре [12]. В качестве следствия из теоремы 2 легко получается результат работы [18] о сходимости распределения случайной величины $\xi_n(\mathbf{H})$ к сложному распределению Пуассона.

Следствие 2. Если при $n \rightarrow \infty$, параметр s стремится к бесконечности, параметр N , группа $\mathbf{H}$ и вероятности $p_1, p_2, \dots, p_N$ фиксированы, $p_j > 0$, $j = 1, \dots, N$, и существует положительный предел

$$\lambda = \lim_{n \rightarrow \infty} \binom{n}{2} \left(\sum_{m=1}^N p_m^2 \right)^s \left(1 - \sum_{m=1}^N p_m^2 \right) |\mathbf{H}(p_1, \dots, p_N)|,$$

где $\mathbf{H}(p_1, \dots, p_N) = \{h \in \mathbf{H} : p_j = p_{h(j)}, j = 1, \dots, N\}$, то

$$\lim_{n \rightarrow \infty} \mathbf{E} z^{\xi_n(H)} = \exp \left(\lambda(z-1) / \left(1 - z \sum_{m=1}^N p_m^2 \right) \right).$$

Замечание 5. Сложное распределение Пуассона, возникающее в следствиях 1 и 2, соответствует распределению пуассоновского числа независимых одинаково распределенных величин, имеющих геометрическое распределение. При этом в следствиях 1 и 2 множество Γ^{vs} выбирается одинаково. Такая ситуация наблюдается не всегда (см. раздел 4.3). При этом отличительной особенностью случая $\mathbf{H}$ -эквивалентных цепочек является то, что асимптотическое поведение случайной величины $\xi_n(\mathbf{H})$ определяется не самой группой $\mathbf{H}$, а ее подгруппой $\mathbf{H}(p_1, \dots, p_N)$.

4.3. $\mathbf{G}$ -эквивалентные цепочки

Будем предполагать, что последовательность $\mathbf{X}$ получена по равновероятной полиномиальной схеме с исходами в множестве $A = \{1, \dots, N\}$. Пусть $\mathbf{G}$ — подгруппа симметрической группы подстановок S_s . Две s -цепочки $\bar{a} = (a_1, \dots, a_s)$ и $\bar{b} = (b_1, \dots, b_s)$ из A^s называются $\mathbf{G}$ -эквивалентными (см. [13]), если существует такая подстановка g из $\mathbf{G}$, что для всех $j = 1, \dots, s$ справедливо равенство $a_{g(j)} = b_j$. В этом случае будем писать $\bar{a}\mathbf{G}\bar{b}$. Если $\mathbf{G} = S_s$, то будем называть такие цепочки перестановочно эквивалентными. Нетрудно видеть, что у $\mathbf{G}$ -эквивалентных цепочек совпадают векторы частот встречаемости символов, то есть $v_r(Y_{i_1}(s)) = v_r(Y_{i_2}(s))$, $r = 1, \dots, s$. Причем последнее условие является критерием перестановочной эквивалентности цепочек.

Введем случайную величину

$$\zeta'_n(\mathbf{G}) = \sum_{1 \leq i_1 < i_1 + s \leq i_2 \leq n} \mathbf{I}\{Y_{i_1}(s)\mathbf{G}Y_{i_2}(s)\},$$

равную числу пар неперекрывающихся $\mathbf{G}$ -эквивалентных s -цепочек на отрезке последовательности $\mathbf{X}$ длины $n + s - 1$. Рассмотрим два частных случая $\mathbf{G}$ -эквивалентности, определяемые выбором группы $\mathbf{G}$.

4.3.1. Цепочки, совпадающие с точностью до циклического сдвига

Пусть группа подстановок $\mathbf{G}$ состоит из подстановок циклического сдвига, то есть $\mathbf{G} = T_s$, где

$$T_s = \left\{ \begin{pmatrix} 1 & 2 & \dots & n-1 & n \\ 1 & 2 & \dots & n-1 & n \end{pmatrix}, \begin{pmatrix} 1 & 2 & \dots & n-1 & n \\ 2 & 3 & \dots & n & 1 \end{pmatrix}, \dots, \begin{pmatrix} 1 & 2 & \dots & n-1 & n \\ n & 1 & \dots & n-2 & n-1 \end{pmatrix} \right\}.$$

Следствие 3. Если при $n \rightarrow \infty$ параметр N фиксирован, а параметр $s \rightarrow \infty$ так, что $\binom{n}{2}(1 - N^{-1})^2 s N^{-s} \rightarrow \lambda$, $0 < \lambda < \infty$, то распределение случайной величины $\zeta'_n(T_s)$ сходится к распределению суммы

$$u_1 v_1 + u_2 v_2 + \dots + u_\pi v_\pi,$$

где случайные величины π , u_k и v_k независимы, величина π распределена по закону Пуассона с параметром λ , а величины u_l и v_l , $l = 1, 2, \dots$, имеют геометрическое распределение с вероятностью успеха $1 - N^{-1}$.

Замечание 6. Следствие 3 доказано на основе теоремы 2. При этом, в отличие от случаев точных повторений и цепочек с одинаковой структурой, в качестве подмножеств Γ_i^{vs} берутся сами множества Γ_i^{nvs} , $i \in J_2$. Такой выбор объясняется различием вида предельных распределений в следствиях 1, 2 и в следствии 3. Непосредственное доказательство следствия 3 содержится в [17].

4.3.2. Цепочки с одинаковыми частотами встречаемости символов

Пусть группа подстановок $\mathbf{G}$ совпадает с симметрической группой подстановок S_s . Тогда случайная величина

$$\zeta'_n(S_s) = \sum_{1 \leq i_1 < i_1 + s \leq i_2 \leq n} \mathbf{I}\{v_r(Y_{i_1}(s)) = v_r(Y_{i_2}(s)), \quad r = 1, \dots, N\},$$

равна числу пар неперекрывающихся s -цепочек с совпадающими частотами встречаемости символов на отрезке последовательности $\mathbf{X}$ длины $n + s - 1$.

Для натурального $N \geq 6$ обозначим через μ_N сумму сходящегося ряда (при $N \leq 5$ ряд расходится)

$$4 \sum_{m=1}^{\infty} (m+1) N^{-2m} \sum_{\alpha_1 + \dots + \alpha_N = m, \alpha_j \geq 0} \left(\frac{m!}{\alpha_1! \dots \alpha_N!} \right)^2.$$

Следствие 4. Пусть $n \rightarrow \infty$.

(а) Если параметр $N \geq 6$ фиксирован, а параметр $s \rightarrow \infty$ так, что

$$\binom{n}{2} N^{N/2} (4\pi s)^{(1-N)/2} \rightarrow \lambda, \quad 0 < \lambda < \infty,$$

то

$$\rho(\mathcal{L}(\zeta'_n(S_s)), \text{CP}(\Lambda_n(\Gamma^{vs}))) = o(1),$$

где $\Gamma_i^{vs} = \Gamma_i^{nvs}$, $i \in J_2$, причем

$$\lim_{n \rightarrow \infty} \mathbf{E} \zeta'_n(S_s) = \lambda, \quad \lim_{n \rightarrow \infty} \mathbf{D} \zeta'_n(S_s) = \lambda(1 + \mu_N).$$

(б) Если параметр $N \geq 14$ фиксирован так, что $\mu_N < 1/8$, а параметр s стремится к бесконечности так, что выполнены условия

$$\frac{n \ln n}{s^{(N-3)/2}} \rightarrow 0, \quad \frac{s^{\varepsilon+(N+7)/2}}{n} \rightarrow 0,$$

для некоторого положительного ε , то распределение случайной величины

$$\frac{\zeta'_n(S_s) - \binom{n}{2} N^{N/2} (4\pi s)^{(1-N)/2}}{\sqrt{\binom{n}{2} N^{N/2} (4\pi s)^{(1-N)/2} (1 + \mu_N)}}$$

сходится к стандартному нормальному закону.

Замечание 7. Утверждение пункта (а) следствия 4 не может быть распространено на случай $N \leq 5$, так как кроме неограниченности математического ожидания величины $\zeta'_n(S_s)$ в этом случае не будет выполнено условие (1) теоремы 2.

Замечание 8. К сожалению, пока не удалось исследовать асимптотическое поведение величин $\lambda_i(\Gamma^{vs})$ и полностью описать предельное сложное распределение Пуассона для случайных величин $\zeta'_n(S_s)$.

Замечание 9. В условиях следствий 3 и 4 при $s \geq 2$ асимптотическое поведение случайной величины $\zeta'_n(\mathbf{G})$ и случайной величины $\zeta_n(\mathbf{G})$ различается. Можно доказать, что асимптотическое поведение $\zeta_n(\mathbf{G})$ определяется парами перекрывающихся перестановочно эквивалентных цепочек $Y_i(s)$ и $Y_{i+1}(s)$, $i = 1, \dots, n-1$, и в условиях следствий 3 и 4 величина $\mathbf{E} \zeta_n(\mathbf{G})$ стремится к бесконечности (см. [12] и [17]).

В условиях следствия 4

$$\mathbf{E} \zeta'_n(S_s) = \binom{n-s+1}{2} \pi_s(S_s), \quad \pi_s(S_s) = \sum_{k_1 + \dots + k_N = s, k_i \geq 0} \left(\frac{s!}{k_1! \dots k_N!} \right)^2. \quad (11)$$

Асимптотическое поведение величин $\pi_s(S_s)$ для различных вариантов изменения параметров s и N исследовано в работе [14].

5. Доказательства

Доказательство теоремы 1. Пункт (а) очевидным образом следует из неравенства

$$\rho(\mathcal{L}(\mathbf{Z}_n(\mathbf{F})), \mathcal{L}(\mathbf{Z}'_n(\mathbf{F}))) \leq \mathbf{E} \mathbf{Z}_n(\mathbf{F}) - \mathbf{E} \mathbf{Z}'_n(\mathbf{F}).$$

Докажем утверждения пунктов (б), (в) и (г) теоремы 1. Определим случайную величину

$$\mathbf{Z}''_n(\mathbf{F}) = \sum_{(i_1, i_2) \in J_2} \mathbf{I}\{\mathbf{F}_{i_1} = \mathbf{F}_{i_2}\}.$$

Заметим, что

$$\begin{aligned} \mathbf{P}\{\mathbf{Z}'_n(\mathbf{F}) \neq \mathbf{Z}''_n(\mathbf{F})\} &\leq \mathbf{E} \sum_{(i_1, i_2) \in J_1} \mathbf{I}\{\mathbf{F}_{i_1} = \mathbf{F}_{i_2}\} - \mathbf{E} \sum_{(i_1, i_2) \in J_2} \mathbf{I}\{\mathbf{F}_{i_1} = \mathbf{F}_{i_2}\} \\ &= \sum_{(i_1, i_2) \in J_1 \setminus J_2} \mathbf{P}\{\mathbf{F}_{i_1} = \mathbf{F}_{i_2}\} = |J_1 \setminus J_2| \pi_s(f) \leq sn \pi_s(f), \end{aligned} \quad (12)$$

так как

$$|J_1| = \binom{n-s+1}{2}, \quad |J_2| = \binom{n-2s+1}{2}, \quad |J_1 \setminus J_2| = s(n - (3s-1)/2).$$

По свойствам метрики полной вариации

$$\begin{aligned} \rho(\mathcal{L}(\mathbf{Z}'_n(\mathbf{F})), \text{CP}(\Lambda_n(\Gamma^{vs}))) &\leq \rho(\mathcal{L}(\mathbf{Z}''_n(\mathbf{F})), \text{CP}(\Lambda_n(\Gamma^{vs}))) + \rho(\mathcal{L}(\mathbf{Z}''_n(\mathbf{F})), \mathcal{L}(\mathbf{Z}'_n(\mathbf{F}))) \\ &\leq \rho(\mathcal{L}(\mathbf{Z}''_n(\mathbf{F})), \text{CP}(\Lambda_n(\Gamma^{vs}))) + \mathbf{P}\{\mathbf{Z}'_n(\mathbf{F}) \neq \mathbf{Z}''_n(\mathbf{F})\} \\ &\leq \rho(\mathcal{L}(\mathbf{Z}''_n(\mathbf{F})), \text{CP}(\Lambda_n(\Gamma^{vs}))) + sn \pi_s(f). \end{aligned}$$

Далее оценим расстояние по вариации между распределением случайной величины $\mathbf{Z}''_n(\mathbf{F})$ и распределением $\text{CP}(\Lambda_n(\Gamma^{vs}))$. Будем использовать метод Стейна оценки точности сложной пуассоновской аппроксимации (см. [22]) и в определенной мере придерживаться схемы доказательства теоремы 1 из [10].

Пусть Γ — произвольный конечный набор индексов, I_a , $a \in \Gamma$, — случайные индикаторы, $W = \sum_{a \in \Gamma} I_a$. Для каждого I_a разделим некоторым образом множество Γ на четыре непересекающихся множества $\{a\}$, Γ_a^{vs} , Γ_a^b , Γ_a^{vw} , и положим

$$U_a = \sum_{b \in \Gamma_a^{vs}} I_b, \quad V_a = \sum_{b \in \Gamma_a^b} I_b.$$

Определим набор $\Lambda = (\lambda_1, \dots, \lambda_{D+1}, 0, \dots)$, где

$$\lambda_k = k^{-1} \sum_{a \in \Gamma} \mathbf{E}\{I_a \mathbf{I}\{I_a + U_a = k\}\}, \quad D = \max_a |\Gamma_a^{vs}|.$$

Введем обозначение

$$\varphi = \sum_{a \in \Gamma} \sum_{k=1}^{|\Gamma_a^{vs}|+1} \varphi_{ak},$$

где

$$\varphi_{ak} = \mathbf{E}[\mathbf{E}\{I_a \mathbf{I}\{I_a + U_a = k\} \mid (I_b : b \in \Gamma_a^{vw})\} - \mathbf{E}\{I_a \mathbf{I}\{I_a + U_a = k\}\}].$$

Теорема 5 ([22]). *При любом выборе указанных выше множеств*

$$\rho(\mathcal{L}(W), \text{CP}(\Lambda)) \leq c_1(\Lambda) \varphi + c_2(\Lambda) \sum_{a \in \Gamma} ((\mathbf{E} I_a)^2 + \mathbf{E} I_a \mathbf{E}(U_a + V_a) + \mathbf{E} I_a V_a), \quad (13)$$

где

$$\max\{c_1(\Lambda), c_2(\Lambda)\} \leq \min \left\{ 1, \frac{1}{\lambda_1} \right\} \exp \left\{ \sum_{k=1}^{\infty} \lambda_k \right\}.$$

Если $k\lambda_k \downarrow 0$, при $k \uparrow \infty$, или

$$\left(\sum_{k \geq 1} k\lambda_k \right)^{-1} \sum_{k \geq 1} k(k-1)\lambda_k < 1/2,$$

то

$$c_1(\Lambda) = \begin{cases} 1, & \lambda_1 - 2\lambda_2 \leq 1, \\ (\lambda_1 - 2\lambda_2)^{-1/2}(2 - (\lambda_1 - 2\lambda_2)^{-1/2}), & \lambda_1 - 2\lambda_2 > 1, \end{cases}$$

$$c_2(\Lambda) = \min\{1, (\lambda_1(n) - 2\lambda_2(n))^{-1}((4(\lambda_1(n) - 2\lambda_2(n)))^{-1} + (\ln 2(\lambda_1(n) - 2\lambda_2(n))))_+\}.$$

Оценка (13) доказана в работе [22] с помощью результатов работы [23], неравенства и выражения для коэффициентов получены в [23] и [24].

В нашем случае для любой пары $\bar{i} = (i_1, i_2)$ из J_2 разделим множество $\Gamma = J_2$ на четыре непересекающихся подмножества

$$\{\bar{i}\}, \quad \Gamma_i^{vw} = \{\bar{j} = (j_1, j_2) \in \Gamma : \min\{|i_1 - j_1|, |i_2 - j_2|, |i_1 - j_2|, |i_2 - j_1|\} > 2s - 1\},$$

$$\Gamma_i^{vs} \subseteq \Gamma_i^{nvs} = \{\bar{j} \in \Gamma : \max\{|i_1 - j_1|, |i_2 - j_2|\} < s, \}, \quad \Gamma_i^b = \Gamma \setminus (\Gamma_i^{vw} \cup \Gamma_i^{vs} \cup \{\bar{i}\}).$$

Тогда нетрудно видеть, что $\varphi = 0$. Кроме того, сложное распределение Пуассона определяемое таким разбиением множества пар цепочек в соответствии с теоремой 4 отличается от распределения $\text{CP}(\Lambda_n(\Gamma^{vs}))$ только за счет цепочек таких $\bar{i} = (i_1, i_2)$ из J_2 , для которых $i_1 < s$ или $i_2 > n - s + 1$, суммарное количество которых не превосходит $2ns$.

Следовательно,

$$\rho(\mathcal{L}(\mathbf{Z}_n''(\mathbf{F})), \text{CP}(\Lambda_n(\Gamma^{vs}))) \leq 2ns\pi_s(f) + c_2(\Lambda_n(\Gamma^{vs}))$$

$$\times \left(\sum_{\bar{i} \in \Gamma} \left(\mathbf{E}\eta_{\mathbf{F}}(\bar{i}) \left(\mathbf{E}\eta_{\mathbf{F}}(\bar{i}) + \sum_{\bar{j} \in \Gamma_i^{vs}} \mathbf{E}\eta_{\mathbf{F}}(\bar{j}) + \sum_{\bar{j} \in \Gamma_i^b} \mathbf{E}\eta_{\mathbf{F}}(\bar{j}) \right) \right) \right.$$

$$\left. + \sum_{\bar{i} \in \Gamma} \sum_{\bar{j} \in \Gamma_i^b} \mathbf{E}(\eta_{\mathbf{F}}(\bar{i})\eta_{\mathbf{F}}(\bar{j})) \right), \quad (14)$$

где

$$\eta_{\mathbf{F}}(\bar{k}) = \mathbf{I}(\mathbf{F}_{k_1} = \mathbf{F}_{k_2}), \quad \bar{k} = (k_1, k_2) \in \Gamma.$$

Тогда, учитывая, что

$$|\Gamma \setminus \Gamma_i^{vw}| \leq 4(4s - 3)n$$

для всех $\bar{i} \in \Gamma$, получаем, что

$$\sum_{\bar{i} \in \Gamma} \left(\mathbf{E}\eta_{\mathbf{F}}(\bar{i}) \left(\mathbf{E}\eta_{\mathbf{F}}(\bar{i}) + \sum_{\bar{j} \in \Gamma_i^{vs}} \mathbf{E}\eta_{\mathbf{F}}(\bar{j}) \right) + \sum_{\bar{j} \in \Gamma_i^b} \mathbf{E}\eta_{\mathbf{F}}(\bar{j}) \right) \leq \mathbf{E}\mathbf{Z}_n'(\mathbf{F})16sn\pi_s(f).$$

Далее для любых $\bar{i}, \bar{j} \in \Gamma$ определим множества

$$U(\bar{i}) = [i_1, i_1 + s - 1] \cup [i_2, i_2 + s - 1], \quad U(\bar{i}, \bar{j}) = U(\bar{i}) \cup U(\bar{j}).$$

Если $\bar{i} \in \Gamma$ и $\bar{j} \in \Gamma_{\bar{i}}^b$, то множество $U(\bar{i}, \bar{j})$ состоит из двух, трех или из четырех компонент связности. Обозначим через $\Gamma_{\bar{i}}^b(m)$, $\bar{i} \in \Gamma$, множество таких пар $\bar{j}, \bar{j} \in \Gamma_{\bar{i}}^b$, что $U(\bar{i}, \bar{j})$ состоит из m компонент связности, $m = 2, 3, 4$.

Нетрудно видеть, что

$$\sum_{\bar{i} \in \Gamma} \sum_{\bar{j} \in \Gamma_{\bar{i}}^b(2)} \mathbf{E}(\eta_{\mathbf{F}}(\bar{i})\eta_{\mathbf{F}}(\bar{j})) \leq \psi(\Gamma^{vs}).$$

Пусть для некоторого $\bar{i} \in \Gamma$ справедливо включение $\bar{j} \in \Gamma_{\bar{i}}^b(3)$. В этом случае ровно одна из цепочек $Y_{i_1}(s)$, $Y_{i_2}(s)$ пересекается ровно с одной из цепочек $Y_{j_1}(s)$, $Y_{j_2}(s)$, пусть, не ограничивая общности, это будут цепочки $Y_{i_1}(s)$ и $Y_{j_1}(s)$, причем $i_1 \geq j_1$. Тогда

$$\begin{aligned} \mathbf{P}\{\mathbf{F}_{i_1} = \mathbf{F}_{i_2}, \mathbf{F}_{j_1} = \mathbf{F}_{j_2}\} &= \sum_{b \in \mathbf{N}} (\mathbf{P}\{\mathbf{F}_{j_1} = \mathbf{F}_{j_2} \mid \mathbf{F}_{i_1} = \mathbf{F}_{i_2} = b\} \mathbf{P}\{\mathbf{F}_{i_1} = \mathbf{F}_{i_2} = b\}) \\ &\leq \sum_{b \in \mathbf{N}} (\max_{a \in \mathbf{N}} \mathbf{P}\{\mathbf{F}_{j_2} = a \mid \mathbf{F}_{i_1} = \mathbf{F}_{i_2} = b\} \mathbf{P}\{\mathbf{F}_{i_1} = \mathbf{F}_{i_2} = b\}) \\ &\leq \sum_{b \in \mathbf{N}} (\max_{a \in \mathbf{N}} \mathbf{P}\{\mathbf{F}_{j_2} = a\} \mathbf{P}\{\mathbf{F}_{i_1} = \mathbf{F}_{i_2} = b\}) \\ &= \kappa_s(f) \sum_{b \in \mathbf{N}} \mathbf{P}\{\mathbf{F}_{i_1} = \mathbf{F}_{i_2} = b\} = \kappa_s(f) \pi_s(f). \end{aligned}$$

Тогда, учитывая очевидную оценку $|\Gamma_{\bar{i}}^b(3)| \leq 8sn$, $\bar{i} \in \Gamma$, получаем, что

$$\sum_{\bar{i} \in \Gamma} \sum_{\bar{j} \in \Gamma_{\bar{i}}^b(3)} \mathbf{E}(\eta_{\mathbf{F}}(\bar{i})\eta_{\mathbf{F}}(\bar{j})) \leq \sum_{\bar{i} \in \Gamma} \sum_{\bar{j} \in \Gamma_{\bar{i}}^b(3)} \pi_s(f) \kappa_s(f) \leq \mathbf{E}Z'_n(\mathbf{F}) 8sn \kappa_s(f).$$

Рассмотрим теперь случай $\bar{j} \in \Gamma_{\bar{i}}^b(4)$, $\bar{i} \in \Gamma$. Очевидно, что

$$\mathbf{P}\{\mathbf{F}_{i_1} = \mathbf{F}_{i_2}, \mathbf{F}_{j_1} = \mathbf{F}_{j_2}\} = \pi_s(f)^2$$

и $|\Gamma_{\bar{i}}^b(4)| \leq 8sn$, $\bar{i} \in \Gamma$. Следовательно,

$$\sum_{\bar{i} \in \Gamma} \sum_{\bar{j} \in \Gamma_{\bar{i}}^b(4)} \mathbf{E}(\eta_{\mathbf{F}}(\bar{i})\eta_{\mathbf{F}}(\bar{j})) \leq \sum_{\bar{i} \in \Gamma} \sum_{\bar{j} \in \Gamma_{\bar{i}}^b(4)} \pi_s(f)^2 \leq \mathbf{E}Z'_n(\mathbf{F}) 8sn \pi_s(f).$$

Подставляя эти оценки в (14), получаем утверждение пункта (б) теоремы 1.

Для доказательства пунктов (в) и (г) теоремы 1 воспользуемся еще одним результатом работы [22]. Пусть g — произвольная ограниченная функция, заданная на множестве неотрицательных целых чисел, со значениями в $\mathbf{R}$. Тогда, в соответствии с результатами из [22], в обозначениях теоремы 4 справедлива оценка

$$\begin{aligned} \left| \mathbf{E} \left\{ Wg(W) - \sum_{k=1}^{\infty} k \lambda_k(\Gamma^{vs}) g(W+k) \right\} \right| \\ \leq \|g\| \varphi + \|\Delta g\| \sum_{a \in \Gamma} ((\mathbf{E}I_a)^2 + \mathbf{E}I_a \mathbf{E}(U_a + V_a) + \mathbf{E}I_a V_a), \end{aligned}$$

где

$$\|g\| = \sup_{j \geq 1} |g(j)|, \quad \|\Delta g\| = \sup_{j \geq 1} |g(j+1) - g(j)|.$$

Выбирая в качестве функции g константу единица, с учетом (12) получим утверждение пункта (в) теоремы 1. Для доказательства пункта (г) теоремы 1 достаточно задать функцию g следующим образом:

$$g(x) = \begin{cases} x, & \text{если } x \leq \binom{n-2s+1}{2} + (2s-1)^2, \\ \binom{n-2s+1}{2} + (2s-1)^2 & \text{в противном случае,} \end{cases}$$

и воспользоваться несложно проверяемой оценкой

$$|\mathbf{DZ}'_n(\mathbf{F}) - \mathbf{DZ}''_n(\mathbf{F})| \leq 5ns^3\pi_s(f).$$

Теорема 1 доказана.

Доказательство теоремы 2. Применим теорему 1, используя единицу в качестве верхней оценки для величины $c(\Gamma^{vs})$. Тогда с учетом неравенства $\kappa_s(f) \geq \pi_s(f)$ для завершения доказательства достаточно заметить, что из условий теоремы 2 следует сходимость распределения $\text{CP}(\Lambda_n(\Gamma^{vs}))$ к распределению $\text{CP}(\Lambda)$. Теорема 2 доказана.

Доказательство теоремы 3. Положим $\Gamma_i^{vs} = \Gamma_i^{nvs}$, $i \in J_2$, и воспользуемся теоремой 1. Очевидно, что в таком случае справедливо равенство $\psi(\Gamma^{vs}) = 0$. Следовательно,

$$\rho(\mathcal{L}(\mathbf{Z}'_n(\mathbf{F})), \text{CP}(\Lambda_n(\Gamma^{vs}))) \leq 3ns\pi_s(f) + c(\Gamma^{vs})\mathbf{EZ}'_n(\mathbf{F})sn(25\pi_s(f) + 8\kappa_s(f)). \quad (15)$$

Далее из условия (7) следует, что $s = o(n^2)$, что с учетом условия (6) и пунктов (в) и (г) теоремы 1 гарантирует справедливость оценок

$$\sum_{k=1}^{(2s-1)^2} k\lambda_k(\Gamma^{vs}) = \mathbf{EZ}'_n(\mathbf{F})(1 + o(1)), \quad (16)$$

$$\sum_{k=1}^{(2s-1)^2} k^2\lambda_k(\Gamma^{vs}) = \mathbf{DZ}'_n(\mathbf{F})(1 + o(1)). \quad (17)$$

Из оценок (16) и (17) с учетом условия (5) следует, что для величины $c(\Gamma_i^{nvs})$ можно использовать оценку

$$c(\Gamma^{vs}) = \min\{1, (\lambda_1(\Gamma^{vs}) - 2\lambda_2(\Gamma^{vs}))^{-1}((4(\lambda_1(\Gamma^{vs}) - 2\lambda_2(\Gamma^{vs})))^{-1} + (\ln 2(\lambda_1(\Gamma^{vs}) - 2\lambda_2(\Gamma^{vs})))_+)\}. \quad (18)$$

Кроме того, из условия (5) и оценок (16) и (17) следует, что

$$\frac{\mathbf{EZ}'_n(\mathbf{F})}{\lambda_1(\Gamma^{vs}) - 2\lambda_2(\Gamma^{vs})} = O(1). \quad (19)$$

Подставляя оценку (18) в (15) с учетом оценки (19) и условия (6), с помощью пункта (б) теоремы 1 убеждаемся в том, что расстояние по вариации между распределением случайной величины $\mathbf{Z}'_n(\mathbf{F})$ и сложным распределением Пуассона $\text{CP}(\Lambda_n(\Gamma^{vs}))$ стремится к нулю.

Докажем, что из условий теоремы 3 следует сходимость к нормальному закону распределения $\text{CP}(\Lambda_n(\Gamma^{vs}))$. Воспользуемся теоремой 1 работы [25]. В указанной теореме сказано, что если для некоторой последовательности случайных величин $\varepsilon = \{\varepsilon_1, \dots, \varepsilon_n, \dots\}$ при $n \rightarrow \infty$ для семиинвариантов $\kappa_j(\varepsilon_n)$, $j = 1, 2, \dots$, справедливы асимптотические равенства $\kappa_1(\varepsilon_n) \rightarrow \mu$, $\zeta_2(\varepsilon_n) \rightarrow \sigma^2 > 0$ и $\kappa_r(\varepsilon_n) \rightarrow 0$ для всех $r \geq r_0$ при некотором фиксированном $r_0 \geq 3$, то распределение ε_n сходится к нормальному распределению со средним μ и дисперсией σ^2 . Применим эту теорему к центрированной средним и нормированной дисперсией последовательности частичных сумм последовательности $\{N_1(\Gamma^{vs}), 2N_2(\Gamma^{vs}), \dots, kN_k(\Gamma^{vs}), \dots\}$, где случайная величина $N_k(\Gamma^{vs})$ распределена по закону Пуассона с параметром $\lambda_k(\Gamma^{vs})$, $k = 1, 2, \dots$.

Прямыми вычислениями убеждаемся, что в рассматриваемом случае r -й семиинвариант имеет вид

$$\left(\sum_{k=1}^{(2s-1)^2} k^2 \lambda_k(\Gamma^{vs}) \right)^{-r/2} \sum_{k=1}^{(2s-1)^2} k^r \lambda_k(\Gamma^{vs}). \quad (20)$$

Далее, из условия (5) с учетом пунктов (в) и (г) теоремы 1 следует, что для достаточно большого n справедливы оценки

$$\lambda_1(\Gamma^{vs}) \geq k \lambda_k(\Gamma^{vs}), \quad k = 2, 3, \dots$$

Поэтому

$$\sum_{k=1}^{(2s-1)^2} k^r \lambda_k(\Gamma^{vs}) \leq \sum_{k=1}^{(2s-1)^2} k^{r-1} \lambda_1(\Gamma^{vs}) \leq \lambda_1(\Gamma^{vs}) (2s-1)^{2r}. \quad (21)$$

Подставляя оценку (21) в оценку (20) и используя условие (7), убеждаемся, что условия теоремы 1 из [25] выполнены, и следовательно, устанавливаем сходимость распределения $\text{CP}(\Lambda_n(\Gamma^{vs}))$ к нормальному закону.

Теорема 3 доказана.

Доказательство следствия 4. Зададим функцию f , зависящую от цепочек полиномиальной схемы так, чтобы она принимала одинаковые натуральные значения только для цепочек с совпадающими частотами встречаемости символов. Докажем, что для такой функции f выполнено условие (1) теоремы 2 при $\Gamma_i^{vs} = \Gamma_i^{nvs}$, $i \in J_2$ ($\psi(\Gamma^{vs}) = 0$).

Применяя результаты работы [14] для оценки величины $\pi_s(f)$ (см. равенства (11) замечания 4 выше), получаем, что

$$\pi_s(f) = N^{N/2} (4\pi s)^{(1-N)/2} (1 + O(s^{-1})),$$

а значит,

$$\mathbf{E} \zeta'_n = \lambda(1 + o(1)).$$

Можно показать (см. [19]), что

$$\kappa_s(f) = N^{-s} \frac{s!}{(\lfloor s/N \rfloor!)^N (\lfloor s/N \rfloor + 1)^{s-N\lfloor s/N \rfloor}},$$

где $\lfloor x \rfloor$ — ближайшее целое снизу для x . Из этой формулы нетрудно вывести, что $\kappa_{r+1}(f) \leq \kappa_r(f)$, $r = 1, 2, \dots$, и, следовательно, $\kappa_s(f) \leq \kappa_{(N \lfloor s/N \rfloor)}(f)$. Применяя формулу Стирлинга для факториалов, стоящих в числителе и знаменателе дроби, представляющей $\kappa_{(N \lfloor s/N \rfloor)}(f)$, получаем, что

$$\kappa_{(N \lfloor s/N \rfloor)}(f) = N^{N/2} (2\pi N \lfloor s/N \rfloor)^{(1-N)/2} (1 + o(1)).$$

Из выведенных оценок, с учетом замечания 1, следует сближение распределений случайной величины $\zeta'_n(S_s)$ и ее сопровождающего распределения $\text{CP}(\Lambda_n(\Gamma^{vs}))$. Действительно, из условий следствия 4 с учетом равенства $\psi(\Gamma^{vs}) = 0$ вытекает справедливость оценки (3).

Для завершения доказательства пункта (а) следствия 4 остается получить асимптотическую оценку дисперсии $\zeta'_n(S_s)$.

Лемма 1. В условиях следствия 4 для любых натуральных чисел r и l , любых размещений $1 \leq j_1 < \dots < j_r \leq s$, $1 \leq m_1 < \dots < m_l \leq s$ и любых наборов $a_1, \dots, a_r$, $b_1, \dots, b_l$, $a_i, b_k = 1, \dots, N$, $i = 1, \dots, r$, $k = 1, \dots, l$, при $s \rightarrow \infty$ справедливо асимптотическое равенство

$$\begin{aligned} \mathbf{P}\{v_r(Y_1(s)) = v_r(Y_{s+1}(s)), r = 1, \dots, N \mid X_{j_i} = a_i, X_{s+m_k} = b_k, \\ i = 1, \dots, r, k = 1, \dots, l\} = \pi_s(f)(1 + O((r+l)s^{-1/3})). \end{aligned}$$

Доказательство леммы 1. Не ограничивая общности, можно считать, что $j_i = i$, $m_k = k$, $i = 1, \dots, r$, $k = 1, \dots, l$, $a_1 = 1$. Обозначим через $\bar{a}$ вектор $a_2, \dots, a_r$, а вектор $b_1, \dots, b_l$ через $\bar{b}$. Кроме того, пусть

$$D(\bar{a}, \bar{b}) = \{X_i = a_i, i = 2, \dots, r, X_{s+k} = b_k, k = 1, \dots, l\}.$$

Тогда достаточно доказать, что для любых наборов $a_2, \dots, a_r$, $b_1, \dots, b_l$, $a_i, b_k = 1, \dots, N$, $i = 2, \dots, r$, $k = 1, \dots, l$, справедливо соотношение

$$\begin{aligned} \mathbf{P}\{v_r(Y_1(s)) = v_r(Y_{s+1}(s)), r = 1, \dots, N \mid X_1 = 1, D(\bar{a}, \bar{b})\} \\ = \mathbf{P}\{v_r(Y_1(s)) = v_r(Y_{s+1}(s)), r = 1, \dots, N \mid X_1 = 2, D(\bar{a}, \bar{b})\}(1 + O(s^{-1/3})). \end{aligned}$$

По неравенству Чебышева

$$\mathbf{P}\{|v_r(Y_1(s)) - sN^{-1}| \geq s^{2/3}\} \leq \frac{N^{-1}(1 - N^{-1})}{s^{1/3}}, \quad r = 1, \dots, N.$$

Пусть

$$\Upsilon = \{|v_r(Y_1(s)) - sN^{-1}| \geq s^{2/3}, r = 1, \dots, N\}.$$

Тогда $\mathbf{P}\{\Upsilon\} \leq s^{-1/3}$. Следовательно,

$$\begin{aligned} \mathbf{P}\{v_r(Y_1(s)) = v_r(Y_{s+1}(s)), r = 1, \dots, N \mid X_1 = 1, D(\bar{a}, \bar{b})\} \\ - \mathbf{P}\{v_r(Y_1(s)) = v_r(Y_{s+1}(s)), r = 1, \dots, N, \bar{\Upsilon} \mid X_1 = 1, D(\bar{a}, \bar{b})\} \\ = \mathbf{P}\{v_r(Y_1(s)) = v_r(Y_{s+1}(s)), r = 1, \dots, N, \Upsilon \mid X_1 = 1, D(\bar{a}, \bar{b})\} \\ = O(\kappa_s(f)s^{-1/3}) = O(\pi_s(f)s^{-1/3}). \end{aligned}$$

Пусть $\beta_r = v_r(\bar{a})$, $\gamma_r = v_r(\bar{b})$, $r = 1, \dots, N$. Тогда

$$\begin{aligned}
\mathbf{P}\{v_r(Y_1(s)) = v_r(Y_{s+1}(s)), r = 1, \dots, N, \bar{\gamma} \mid X_1 = 1, D(\bar{a}, \bar{b})\} \\
= N^{-2s+r+l} \sum_{\substack{\alpha_1 + \dots + \alpha_N = m \\ |\alpha_j - sN^{-1}| < s^{5/8}, \alpha_j \geq 0}} \frac{m!}{(\alpha_1 - \gamma_1)! \dots (\alpha_N - \gamma_N)!} \\
\times \frac{m!}{(\alpha_1 - 1 - \beta_1)! (\alpha_2 - \beta_2)! \dots (\alpha_N - \beta_N)!} \\
= N^{-2s+r+l} (1 + O(s^{-1/3})) \sum_{\substack{\alpha_1 + \dots + \alpha_N = m \\ |\alpha_j - sN^{-1}| < s^{5/8}, \alpha_j \geq 0}} \frac{m!}{(\alpha_1 - \gamma_1)! \dots (\alpha_N - \gamma_N)!} \\
\times \frac{m!}{(\alpha_1 - \beta_1)! (\alpha_2 - 1 - \beta_2)! \dots (\alpha_N - \beta_N)!} \\
= \mathbf{P}\{v_r(Y_1(s)) = v_r(Y_{s+1}(s)), r = 1, \dots, N, \bar{\gamma} \mid X_1 = 2, D(\bar{a}, \bar{b})\}.
\end{aligned}$$

Лемма 1 доказана.

Оценим дисперсию случайной величины ζ'_n . Нетрудно видеть, что

$$\begin{aligned}
\mathbf{D}\zeta'_n &= \binom{n-2s+1}{2} (1 + o(1)) \sum_{j_1=1}^{2s-1} \sum_{j_2=3s-1}^{5s-3} \mathbf{P}\{F_s = F_{4s-2}, F_{j_1} = F_{j_2}\} \\
&= (1 + o(1)) \left(\binom{n-2s+1}{2} \mathbf{P}\{F_s = F_{4s-2}\} \right. \\
&\quad \left. + \binom{n-2s+1}{2} 4 \sum_{m=1}^{2s-2} \sum_{j_1=s}^{m+s} \mathbf{P}\{F_s = F_{4s-2}, F_{j_1} = F_{5s-2+m-j_1}\} \right).
\end{aligned}$$

Для выполнения условия $F_{j_1} = F_{5s-2+m-j_1}$ при условии $F_s = F_{4s-2}$ необходимо и достаточно (подробнее см. [19]) выполнения условия

$$v_r(Y_s(j_1 - s), Y_{5s-2}(s + m - j_1)) = v_r(Y_{2s}(j_1 - s), Y_{4s-2}(s + m - j_1)), \quad r = 1, \dots, N,$$

которое обозначим через Ξ .

С использованием леммы 1 находим, что

$$\begin{aligned}
\mathbf{P}\{F_s = F_{4s-2}, F_{j_1} = F_{5s-2+m-j_1}\} &= \mathbf{P}\{F_s = F_{4s-2}, \Xi\} \\
&= \mathbf{E}(\mathbf{P}\{F_s = F_{4s-2}, \Xi \mid Y_s(j_1 - s), Y_{4s-2}(s + m - j_1)\}) \\
&= \mathbf{E}(\mathbf{P}\{F_s = F_{4s-2} \mid Y_s(j_1 - s), Y_{4s-2}(s + m - j_1)\} \\
&\quad \times \mathbf{P}\{\Xi \mid Y_s(j_1 - s), Y_{4s-2}(s + m - j_1)\}) \\
&= \mathbf{E}(\pi_s(f)(1 + O(ms^{-1/3})) \mathbf{P}\{\Xi \mid Y_s(j_1 - s), Y_{4s-2}(s + m - j_1)\}) \\
&= \pi_s(f)(1 + O(ms^{-1/3})) \mathbf{P}\{\Xi\} \\
&= \pi_s(f)(1 + O(ms^{-1/3})) \sum_{\alpha_1 + \dots + \alpha_N = m, \alpha_j \geq 0} \left(\frac{m!}{\alpha_1! \dots \alpha_N!} \right)^2.
\end{aligned}$$

Тогда

$$\begin{aligned} & \left| \mathbf{D}\zeta'_n(S_s) - \lambda + \binom{n-2s+1}{2} 4 \sum_{m=1}^{\lfloor s^{1/4} \rfloor} (m+1) \pi_s(f) (1 + O(ms^{-1/3})) \right. \\ & \quad \times \left. \sum_{\alpha_1 + \dots + \alpha_N = m, \alpha_j \geq 0} \left(\frac{m!}{\alpha_1! \dots \alpha_N!} \right)^2 \right| \\ & \leq (1 + o(1)) \binom{n-2s+1}{2} 4 \sum_{m=\lfloor s^{1/4} \rfloor + 1}^{2s-2} (m+1) \kappa_{\lfloor m/2 \rfloor}(f) \mathbf{P}\{F_s = F_{4s-2}\} = o(\mathbf{D}\zeta'_n(S_s)), \end{aligned}$$

где последнее равенство справедливо в силу сходимости ряда $\sum_{m=1}^{\infty} m \kappa_m(f)$. Таким образом, доказан пункт (а) следствия 4.

Для доказательства пункта (б) следствия 4 заметим, что из условий этого пункта с учетом полученных выше оценок для среднего и дисперсии величины $\zeta'_n(S_s)$ вытекают условия (5), (6) и (7) теоремы 3, и воспользуемся этой теоремой.

Следствие 4 доказано.

В заключение приведем пример к замечанию 1, соответствующий случаю, когда величина $\mathbf{E}\mathbf{Z}'_n(\mathbf{F})$ стремится к бесконечности. Будем предполагать, что последовательность $\mathbf{X}$ получена по равновероятной полиномиальной схеме с исходами в множестве $\{1, \dots, N\}$ и M — целое неотрицательное число, не превосходящее N . Определим функцию от s -цепочек последовательности $\mathbf{X}$ следующим образом. Если две цепочки не содержат символов из множества $\{1, \dots, M\}$, то функция f принимает на них одинаковые натуральные значения тогда и только тогда, когда они совпадают. На цепочках, содержащих символы из множества $\{1, \dots, M\}$, f принимает одинаковые натуральные значения, только если они имеют одинаковые частоты встречаемости символов из множества $\{1, \dots, M\}$, то есть $\nu_r(Y_{i_1}(s)) = \nu_r(Y_{i_2}(s))$, $r = 1, \dots, M$, для цепочек $Y_{i_1}(s)$ и $Y_{i_2}(s)$, для которых $\sum_{r=1}^M \nu_r(Y_{i_1}(s)) \geq 1$. Обозначим случайную величину $\mathbf{Z}'_n(\mathbf{F})$ для определенной выше функции f через $D'_{n,M}$. Нетрудно видеть, что $D'_{n,0} = \delta'_n$, $D'_{n,N} = \zeta'_n(S_s)$.

Следствие 5. Если при $n \rightarrow \infty$ параметры N , M и s стремятся к бесконечности так, что $s^8 = o(M)$, $sM = o(N)$ и выполнено условие

$$\binom{n-2s+1}{2} MN^{-2} \rightarrow \lambda, \quad 0 < \lambda < \infty, \quad (22)$$

то распределение случайной величины $s^{-2}D'_{n,M}$ сходится к распределению Пуассона с параметром λ .

Замечание 10. Из доказательства следствия 5 вытекает сближение распределения случайной величины $D'_{n,M}$ и сложного распределения Пуассона $CP(\Lambda_n)$, для которого

$$\Lambda_n = \binom{n-2s+1}{2} Ms^{-1} N^{-2} \delta_{s^2},$$

где δ_{s^2} — мера, сосредоточенная в точке s^2 , то есть все члены последовательности $\lambda_1, \lambda_2, \dots$ кроме λ_{s^2} равны нулю, а величина $s^2 \lambda_{s^2}$ стремится к бесконечности, причем

$$\mathbf{E}D'_{n,M} = s^2 \lambda_{s^2} (1 + o(1)).$$

Доказательство следствия 5. Заметим, что из условия $sM = o(N)$ и условия (22) следует, что $s = o(n)$. Поэтому справедлива оценка $n^2 N^{-s} = o(1)$, которая гарантирует отсутствие в последовательности $\mathbf{X}$ точных повторений s -цепочек. Далее, нетрудно проверить, что условие $sM = o(N)$ обеспечивает отсутствие цепочек с одинаковыми наборами символов из множества $\{1, \dots, M\}$, содержащих более одного символа из этого множества. Таким образом, можно считать, что двум s -цепочкам соответствуют одинаковые значения функции f только в том случае, когда в каждой из них встретился один раз один и тот же символ из множества $\{1, \dots, M\}$ и других символов из множества $\{1, \dots, M\}$ в них не встретилось.

Применим теорему 2 при $\Gamma_i^{vs} = \Gamma_i^{nvs}$, $i \in J_2$. Нетрудно убедиться, что асимптотически при $n \rightarrow \infty$ справедливы оценки

$$\kappa_s(f) = sN^{-1}(1 + o(1)), \quad \pi_s(f) = s^2 MN^{-2}(1 + o(1)).$$

Следовательно, справедлива оценка

$$\mathbf{E}D'_{n,M} = \lambda s^2(1 + o(1))$$

и условие (1) теоремы 2 приобретает вид $s^3 n N^{-1} = o(1)$. Легко видеть, что это условие вытекает из условия (22) и условия $s^8 = o(M)$ следствия 5.

Кроме того, нетрудно видеть что, все члены последовательности $\lambda_1, \lambda_2, \dots$ кроме λ_{s^2} равны нулю, а

$$\lambda_{s^2} = \binom{n - 2s + 1}{2} MN^{-2}(1 + o(1)).$$

Следствие 5 доказано.

Автор признателен В. Г. Михайлову за полезные замечания.

Список литературы

1. Буравлев С. М., Повторения с точностью до перестановок в последовательности независимых испытаний. *Дискретная математика* (1999) **11**, №1, 53–75.
2. Буравлев С. М., Повторения с точностью до перестановок, образующих латинский прямоугольник. *Дискретная математика* (2000) **12**, №1, 24–46.
3. Зубков А. М., Михайлов В. Г., Предельные распределения случайных величин, связанных с длинными повторениями в последовательности независимых испытаний. *Теория вероятностей и ее применения* (1974) **19**, №1, 173–181.
4. Зубков А. М., Михайлов В. Г., О повторениях s -цепочек в последовательности независимых величин. *Теория вероятностей и ее применения* (1979) **24**, №2, 267–279.
5. Колчин В. Ф., Севастьянов Б. А., Чистяков В. П., *Случайные размещения*. Наука, Москва, 1976.
6. Королюк В. С., Боровских Ю. В., *Теория U-статистик*. Наукова думка, Киев, 1989.
7. Михайлов В. Г., Асимптотическая нормальность в схеме конечно-зависимого размещения частиц по ячейкам. *Матем. сб.* (1982) **119**, №4, 509–520.
8. Михайлов В. Г., Асимптотическая нормальность разделимых статистик от частот m -цепочек. *Дискретная математика* (1989) **1**, №4, 92–103.

9. Михайлов В. Г., Явные оценки в предельных теоремах для сумм случайных индикаторов. *Обзор прикладной и промышленной математики* (1994) **1**, №4, 580–617.
10. Михайлов В. Г., Оценка точности сложной пуассоновской аппроксимации для распределения числа совпадающих цепочек. *Теория вероятностей и ее применения* (2003) **46**, №4, 713–723.
11. Михайлов В. Г., Об асимптотических свойствах распределения числа пар H -связанных цепочек. *Дискретная математика* (2002) **14**, №3, 122–129.
12. Михайлов В. Г., Шойтов А. М., Структурная эквивалентность s -цепочек в случайных дискретных последовательностях. *Дискретная математика* (2003) **15**, №4, 7–34.
13. Сачков В. Н., *Введение в комбинаторные методы дискретной математики*. Наука, Москва, 1982.
14. Тимашев А. Н. О вероятности совпадения векторов частот исходов независимых полиномиальных схем. В сб.: *Третья Всероссийская школа-коллоквиум по стохастическим методам. Тезисы докладов*. ТВП, Москва, 1996, с. 156–157.
15. Шойтов А. М., Повторения значений функции от отрезков последовательности независимых испытаний. *Дискретная математика* (2000) **12**, №3, 49–59.
16. Шойтов А. М., Предельные распределения числа наборов H -эквивалентных отрезков в равновероятной полиномиальной схеме серий. *Дискретная математика* (2002) **14**, №1, 82–98.
17. Шойтов А. М., О перестановочной эквивалентности цепочек полиномиальной схемы. *Вестник ИКСИ Академии ФСБ РФ, серия К, спец. вып.* 2003, 107–113.
18. Шойтов А. М., Предельные распределения случайных величин, характеризующих связь цепочек полиномиальной схемы структурной эквивалентностью. *Труды по дискретной математике* (2004) **8**, 312–326.
19. Шойтов А. М., Пуассоновское приближение для числа повторений значений дискретной функции от цепочек. *Дискретная математика* (2005) **17**, №2, 56–69.
20. Шойтов А. М., Нормальное приближение для распределения числа пар эквивалентных цепочек в равновероятной схеме. *Обзор прикладной и промышленной математики* (2005) **12**, №1, 204–205.
21. Barbour A. D., Holst L., Janson S., *Poisson approximation*. Oxford Univ. Press, Oxford, 1992.
22. Roos M., Stein method for compound Poisson approximation: The local approach. *Ann. Appl. Probab.* (1994) **4**, 1177–1187.
23. Barbour A. D., Chen L. H. Y., Loh W.-L., Compound Poisson approximation for non-negative random variables via Stein's method. *Appl. Probab.* (1992) **20**, №4, 1843–1866.
24. Barbour A. D., Xia A., Poisson perturbation. *Probab. Statist.* (1999) **3**, 131–150.
25. Janson S., Normal convergence by higher semiinvariants with applications to sums of dependent random variables and random graphs. *Ann. Probab.* (1998) **16**, №1, 305–312.

Статья поступила 14.06.2006.