Общероссийский математический портал

Н. М. Меженная, Предельные теоремы для числа плотных серий в случайной последовательности, *Дискрет. матем.*, 2009, том 21, выпуск 1, 105–116

DOI: <https://doi.org/10.4213/dm1041>

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: 54.237.206.68

26 апреля 2023 г., 12:51:15

Предельные теоремы для числа плотных серий в случайной последовательности

© 2009 г. Н. М. Меженная

В работе изучается совместное распределение числа плотных серий в случайной последовательности со значениями в конечном алфавите. При помощи метода Чена–Стейна найдены оценки расстояния по вариации между распределением вектора из числа плотных серий единиц заданных длин и сопровождающим многомерным пуассоновским распределением. Из этих оценок выведены предельные теоремы пуассоновского типа для числа плотных серий единиц заданных длин и длины, не меньшей заданной, для числа плотно заполненных единицами отрезков, предельная теорема для максимальной длины плотной серии единиц, а также доказана пуассоновская предельная теорема для числа плотных серий единиц заданного веса.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, проект 08–01–00078а.

1. Введение

Пусть $Y_n = \{y_1, \dots, y_n\}$ и $Z_{t+m} = \{z_1, \dots, z_{t+m}\}$ — последовательности знаков алфавита $\{1, \dots, N\}$. Согласно [2], последовательность Y_n плотно вкладывается в последовательность Z_{t+m} на месте t , если найдутся натуральные числа $t = i_1 < \dots < i_n \leq t + m$, $i_{k+1} - i_k \in \{1, 2\}$, $k = 1, \dots, n - 1$, такие, что $y_k = z_{i_k}$, $k = 1, \dots, n$.

Задача о плотном вложении двоичных последовательностей была рассмотрена в [2]. Там была получена верхняя оценка для вероятности плотного вложения двоичной последовательности в равновероятную случайную двоичную последовательность. В [3] результаты работы [2] были распространены на последовательности над произвольным конечным алфавитом и было показано, что вероятность плотного вложения принимает свое минимальное значение, когда вкладываемая последовательность состоит из одинаковых знаков.

Упомянутое свойство минимальности в контексте криптографических проблем, упомянутых в [2], привлекает интерес к специальному изучению плотно заполненных отрезков в случайной последовательности. Отрезок $x_1, \dots, x_k$ последовательности знаков алфавита $\{1, \dots, N\}$ мы называем плотно заполненным знаком $a \in \{1, \dots, N\}$, если $a \in \{x_i, x_{i+1}\}$ при всех $i = 1, \dots, k - 1$, то есть пропуски между знаками a в нем могут состоять не более чем из одного другого знака. Распределение числа плотно заполненных отрезков в случайной последовательности можно вывести из распределения числа плотно заполненных серий. Плотно заполненный (знаком a) отрезок назовем плотно заполненной (знаком a) серией (или просто плотной серией знаков a), если он не содержится ни в каком плотно заполненном отрезке большей длины. Без ограничения общности будем в

дальнейшем рассматривать только плотные серии единиц, при этом для простоты изложения иногда будем говорить о плотных сериях, подразумевая тем самым плотные серии из единиц. Длиной плотной серии назовем расстояние между крайними единицами в ней, а весом плотной серии — число единиц, входящих в ее запись.

Настоящая работа посвящена изучению распределения числа плотных серий заданных длин и плотных серий заданного веса в последовательности независимых одинаково распределенных случайных величин. При помощи метода Чена—Стейна найдены оценки расстояния по вариации между распределением вектора из числа плотных серий единиц заданных длин и сопровождающим многомерным пуассоновским распределением. Из этих оценок выведены предельные теоремы пуассоновского типа для числа плотных серий единиц заданных длин и длины, не меньшей заданной, для числа плотно заполненных единицами отрезков, предельная теорема для максимальной длины плотной серии единиц, а также доказана пуассоновская предельная теорема для числа плотных серий единиц заданного веса.

Статья организована следующим образом. В разделе 2 приводятся постановка задачи и формулировки основных результатов. Их доказательства содержатся в разделе 3. Раздел 4 посвящен изучению числа плотных серий единиц заданного веса.

2. Формулировки основных результатов

Пусть $X = \{X_{-1}, X_0, X_1, \dots, X_T, \dots\}$ — последовательность независимых случайных величин, принимающих значения из множества $\{1, \dots, N\}$ с вероятностями $p_1, \dots, p_N$. Нас интересуют отрезки длины s , плотно заполненные некоторым знаком $a \in \{1, \dots, N\}$. Далее будем считать, что $a = 1$, $p_1 = p \in (0, 1)$. Будем говорить, что знаки $X_t, \dots, X_{t+s-1}$ образуют плотную цепочку единиц длины s , $s \geq 3$, если отрезок последовательности $X_t, \dots, X_{t+s-1}$ плотно заполнен единицами. Будем говорить, что в момент t началась плотная серия единиц длины s , $s \geq 1$, если $X_{t-2} \neq 1$, $X_{t-1} \neq 1$, $X_t = X_{t+s-1} = 1$, $X_{t+s} \neq 1$, $X_{t+s+1} \neq 1$, и отрезок последовательности $X_t, \dots, X_{t+s-1}$ плотно заполнен единицами. Заметим, что согласно этим определениям длина плотной серии на две единицы меньше длины отвечающей ей максимальной плотной цепочки. Пусть событие $E_{t,s}$ означает, что в момент t началась плотная серия единиц длины s .

Пусть

$$c = \frac{p(1-p)^4}{\sqrt{p(4-3p)}}, \quad q = \frac{p + \sqrt{p(4-3p)}}{2}, \quad q_1 = \frac{p - \sqrt{p(4-3p)}}{2}.$$

Нетрудно проверить, что $q > p$ и $q_1 < 0$ при $p \in (0, 1)$.

Теорема 1. При любых $s \geq 3$, $1 \leq t \leq T$ и $p \in (0, 1)$ выполнены соотношения

$$\mathbf{P}\{E_{t,s}\} = c(q^s - q_1^s), \quad (1)$$

$$\mathbf{P}\{E_{t,s}\} > \mathbf{P}\{E_{t,s+1}\}. \quad (2)$$

Заметим, что формула (1) верна при любых $s \geq 1$, а именно, $\mathbf{P}\{E_{1,1}\} = p(1-p)^4$, $\mathbf{P}\{E_{1,2}\} = p^2(1-p)^4$. Кроме того, при $s = 2$ в неравенстве (2) имеет место знак равенства.

При $k \geq 1$, $s \geq 1$ определим случайные величины

$$\zeta_{k,T} = \sum_{t=1}^T I\{E_{t,k}\}, \quad (3)$$

$$\xi_{s,T} = \sum_{k=s}^{\infty} \zeta_{k,T}, \quad (4)$$

равные числу плотных серий, которые начинаются в моменты времени от 1 до T и которые имеют длину k , не меньшую s , соответственно. Число плотных цепочек единиц длины s обозначим через $\vartheta_{s,T}$. Ясно, что при $s \geq 3$

$$\vartheta_{s,T} = \sum_{k=0}^{\infty} (k+1) \zeta_{s+k-2,T}.$$

Пусть

$$\lambda_k = \mathbf{E} \zeta_{s+k,T} = T \mathbf{P}\{E_{1,s+k}\}, \quad k = 0, 1, \dots, r. \quad (5)$$

Обозначим через $\pi_{\lambda_0}, \dots, \pi_{\lambda_r}$ независимые случайные величины, распределенные по закону Пуассона с параметрами $\lambda_0, \dots, \lambda_r$. Пусть $\rho(X, Y)$ — расстояние по вариации между распределениями случайных величин X и Y .

Теорема 2. Пусть $s \geq 3, r \geq 0$. Тогда

$$\rho((\zeta_{s,T}, \dots, \zeta_{s+r,T}), (\pi_{\lambda_0}, \dots, \pi_{\lambda_r})) \leq \frac{2}{T} (r+1)^2 \lambda_0^2 \left(s+r+4 + \frac{2-p}{(1-p)^2} \right). \quad (6)$$

При $r = 0$, используя одномерный вариант метода Чена–Стейна (см. [1]), можно получить более точную оценку. Пусть $s \geq 3$. Тогда

$$\rho(\zeta_{s,T}, \pi_{\lambda_0}) \leq \frac{2\lambda_0(1-e^{-\lambda_0})}{T} \left(s+4 + \frac{2-p}{(1-p)^2} \right). \quad (7)$$

Теорема 3. Пусть $s \geq 3, r \geq 0$. Тогда при любом $d \geq 1$

$$\begin{aligned} & \rho((\zeta_{s,T}, \dots, \zeta_{s+r,T}, \xi_{s+r+1,T}), (\pi_{\lambda_0}, \dots, \pi_{\lambda_r}, \pi_{\Lambda_{r+1}})) \\ & \leq \frac{2}{T} (r+d+1)^2 \lambda_0^2 \left(s+r+d+4 + \frac{2-p}{(1-p)^2} \right) + 2 \sum_{k=r+d+1}^{\infty} \lambda_k. \end{aligned} \quad (8)$$

Здесь $\pi_{\Lambda_{r+1}}$ — случайная величина, распределенная по закону Пуассона с параметром $\Lambda_{r+1} = \sum_{k=r+1}^{\infty} \lambda_k$ и не зависящая от случайных величин $\pi_{\lambda_0}, \dots, \pi_{\lambda_r}$.

Взяв в правой части неравенства (8) $r+d$ приблизительно равным s , получим следующее утверждение.

Следствие 1. Пусть $s \geq 3, 0 \leq r \leq s-2$. Тогда найдется такая константа $C < \infty$, что

$$\rho((\zeta_{s,T}, \dots, \zeta_{s+r,T}, \xi_{s+r+1,T}), (\pi_{\lambda_0}, \dots, \pi_{\lambda_r}, \pi_{\Lambda_{r+1}})) < C T s^3 q^{2s}. \quad (9)$$

Можно показать, что неравенство (9) выполнено при $C = 0,5$.

Оценки (6)–(9) позволяют доказать следующие предельные теоремы.

Теорема 4. Пусть $s, T \rightarrow \infty$,

$$\lambda_0 \rightarrow \lambda, \quad 0 < \lambda < \infty. \quad (10)$$

Тогда

- (1) при любом целом неотрицательном числе r случайные величины $\zeta_{s,T}, \zeta_{s+1,T}, \dots, \zeta_{s+r,T}, \xi_{s+r+1,T}$ асимптотически независимы в совокупности и имеют в пределе распределения Пуассона с параметрами $\lambda, q\lambda, \dots, q^r\lambda, (1-q)^{-1}q^{r+1}\lambda$ соответственно,
- (2) распределение случайной величины $\vartheta_{s,T}$ в пределе совпадает с распределением $\sum_{k=0}^{\infty} (k+1)\bar{\pi}_k$, где $\bar{\pi}_k, k = 0, 1, \dots$, — независимые случайные величины, имеющие распределение Пуассона с параметрами $q^{k-2}\lambda$ соответственно.

Замечание 1. Из теоремы 1 следует, что (10) эквивалентно условию $Tcq^s \rightarrow \lambda$.

Замечание 2. Случайная величина $\sum_{k=0}^{\infty} (k+1)\bar{\pi}_k$ имеет сложное пуассоновское распределение с производящей функцией

$$\begin{aligned} \exp \left\{ \sum_{k=-2}^{\infty} q^k \lambda (z^{k+3} - 1) \right\} &= \exp \left\{ \frac{\lambda}{q^2(1-q)} \left(\sum_{k=1}^{\infty} (1-q)q^{k-1} z^k - 1 \right) \right\} \\ &= \exp \left\{ \frac{\lambda}{q^2(1-q)} \left(\frac{z(1-q)}{1-qz} - 1 \right) \right\}. \end{aligned}$$

Поэтому распределение случайной величины $\sum_{k=0}^{\infty} (k+1)\bar{\pi}_k$ совпадает с распределением суммы $Z_1 + \dots + Z_L$, где случайные величины $L, Z_1, Z_2, \dots$ независимы в совокупности, случайная величина L распределена по закону Пуассона с параметром $\lambda/(q^2(1-q))$, а $Z_1, Z_2, \dots$ — случайные величины, имеющие геометрическое распределение с параметром q : $\mathbf{P}\{Z_i = k\} = (1-q)q^{k-1}, k = 1, 2, \dots, i = 1, 2, \dots$

Пусть $\nu = \nu(T)$ — максимальная из длин отрезков, плотно заполненных единицами, которые начинаются в моменты времени от 1 до T . Из следствия 1 непосредственно вытекает следующее утверждение.

Следствие 2. Пусть $s \geq 3$. Тогда $|\mathbf{P}\{\nu < s\} - e^{-\Lambda_0}| \leq C T s^3 q^{2s}$.

Пусть

$$s_0 = s_0(T) = \left[\frac{\ln T}{\ln(1/q)} \right] = \frac{\ln T}{\ln(1/q)} - \left\{ \frac{\ln T}{\ln(1/q)} \right\},$$

где через $[x]$ и $\{x\}$ обозначены целая и дробная части числа x соответственно.

Теорема 5. Пусть $T \rightarrow \infty$, а параметр $k(T)$ удовлетворяет соотношениям

$$k(T) + \frac{1}{2 \ln(1/q)} (\ln T - 3 \ln \ln T) \rightarrow \infty, \quad (11)$$

$$k(T) + \left(\frac{1}{\ln(1/q)} - \frac{1}{\ln(1/|q_1|)} \right) \ln T \rightarrow \infty. \quad (12)$$

Тогда

$$\mathbf{P}\{\nu < s_0(T) + k\} = \exp \left\{ -\frac{Tcq^{s_0(T)+k}}{1-q} \right\} + o(1)$$

равномерно по $k \geq k(T)$.

Замечание 3. Пусть $\tilde{p}$ — решение уравнения $p^3 - 4p^2 + 5p - 1 = 0$. Можно показать, что это уравнение имеет единственный действительный корень, который приближенно равен 0,245. При $p < \tilde{p}$ из соотношения (11) следует (12). Если $p \geq \tilde{p}$, то (12) влечет (11).

Наконец отметим, что оценки (6)–(9) позволяют получить достаточные условия асимптотической нормальности случайных величин $\zeta_{s,T}, \dots, \zeta_{s+r,T}, \xi_{s+r+1,T}$.

Следствие 3. Пусть число $r, 0 \leq r \leq s-2$, фиксировано, а $s, T \rightarrow \infty$ так, что $Tq^s \rightarrow \infty$, $Ts^3q^{2s} \rightarrow 0$. Тогда случайный вектор

$$\left(\frac{\zeta_{s,T} - \lambda_0}{\sqrt{\lambda_0}}, \dots, \frac{\zeta_{s+r,T} - \lambda_r}{\sqrt{\lambda_r}}, \frac{\xi_{s+r+1,T} - \Lambda_{r+1}}{\sqrt{\Lambda_{r+1}}} \right)$$

имеет в пределе многомерное нормальное распределение с нулевым вектором средних и единичной матрицей ковариаций.

Оценки одномерного варианта метода Чена–Стейна (9) позволяют получить более общие условия асимптотической нормальности случайной величины $\zeta_{s,T}$.

Следствие 4. Пусть $s, T \rightarrow \infty$ так, что $Tq^s \rightarrow \infty$. Тогда распределение случайной величины $(\zeta_{s,T} - \lambda_0)/\sqrt{\lambda_0}$ в пределе совпадает со стандартным нормальным законом.

3. Доказательства

Доказательство теоремы 1. При вычислении вероятности появления плотной серии важно только число знаков, которые разделяют группы из единиц в плотной серии, а значения самих этих элементов роли не играют. Поэтому знаки алфавита, отличные от единицы, можно заменить одним общим знаком, например нулем, вероятность выпадения которого равна $p_2 + \dots + p_N$. Поэтому далее последовательность $\{X_{-1}, X_0, X_1, \dots, X_T, \dots\}$ будем считать последовательностью Бернулли с вероятностью успеха p . В силу однородности последовательности X , $\mathbf{P}\{E_{t,s}\} = \mathbf{P}\{E_{1,s}\}$. Вычислим $\mathbf{P}\{E_{1,s}\}$.

Пусть $p_{s,1}$ — вероятность того, что наблюдается отрезок длины s , плотно заполненный единицами, который начинается и заканчивается единицей, а $p_{s,0}$ — вероятность того, что наблюдается отрезок длины s , плотно заполненный единицами, который начинается единицей и заканчивается нулем. Тогда вероятность того, что наблюдается последовательность длины s , плотно заполненная единицами и начинающаяся единицей, равна $p_s = p_{s,1} + p_{s,0}$. Очевидно, что

$$p_{s+1,0} = p_{s,1}(1-p), \quad (13)$$

$$p_{s+1,1} = pp_{s,1} + pp_{s,0}. \quad (14)$$

С помощью формулы (13) выразим $p_{s,0}$ и подставим в формулу (14), получаем равенство

$$p_{s+1,1} = pp_{s,1} + p(1-p)p_{s-1,1}. \quad (15)$$

Начальные условия имеют вид

$$p_{1,1} = p, \quad p_{2,1} = p^2. \quad (16)$$

Решение рекуррентного уравнения (15) будем искать в виде $p_{s,1} = z^s$. Подставив эту функцию в (15), получим квадратное уравнение $z^2 - pz - p(1-p) = 0$, решениями

которого являются числа $(p \pm \sqrt{p(4-3p)})/2$. Значит, общее решение уравнения (15) имеет вид

$$p_{s,1} = c_1 \left(\frac{p + \sqrt{p(4-3p)}}{2} \right)^s + c_2 \left(\frac{p - \sqrt{p(4-3p)}}{2} \right)^s.$$

Согласно начальным условиям (16)

$$c_1 = \frac{p}{\sqrt{p(4-3p)}}, \quad c_2 = -\frac{p}{\sqrt{p(4-3p)}}.$$

Таким образом,

$$p_{s,1} = \frac{p}{\sqrt{p(4-3p)}}(q^s - q_1^s), \quad p_{s,0} = \frac{p(1-p)}{\sqrt{p(4-3p)}}(q^{s-1} - q_1^{s-1}). \quad (17)$$

Теперь выразим вероятность $\mathbf{P}\{E_{1,s}\}$ через величины $p_{s,1}$ и $p_{s,0}$. Отрезок, плотно заполненный единицами, который начинается и заканчивается единицей, отличается от плотной серии из единиц тем, что последняя ограничена с обеих концов двумя нулевыми элементами. Поэтому

$$\mathbf{P}\{E_{1,s}\} = (1-p)^4 p_{s,1} = \frac{p(1-p)^4}{\sqrt{p(4-3p)}}(q^s - q_1^s) = c(q^s - q_1^s).$$

Таким образом, соотношение (1) доказано.

Индукцией по s покажем, что $\mathbf{P}\{E_{1,s}\} > \mathbf{P}\{E_{1,s+1}\}$ при $s \geq 3$. Проверим, что $p_{3,1} > p_{4,1} > p_{5,1}$. Согласно (17) $p_{3,1} = p^2$, $p_{4,1} = p^3(2-p)$, $p_{5,1} = p^3(1+p-p^2)$, поэтому при $p \in (0, 1)$ справедливы соотношения $p_{3,1} - p_{4,1} = p^2 - 2p^3 - p^4 = p^2(1-p)^2 > 0$, $p_{4,1} - p_{5,1} = 2p^3 - p^4 - p^3 - p^4 + p^5 = p^3(1-p)^2 > 0$. Пусть при $s \geq 6$ имеют место неравенства $p_{s-2,1} > p_{s-1,1} > p_{s,1}$. Покажем, что тогда $p_{s,1} > p_{s+1,1}$. Согласно (15) $p_{s,1} - p_{s+1,1} = p(p_{s-1,1} - p_{s,1}) + p(1-p)(p_{s-2,1} - p_{s-1,1}) > 0$ в силу предположения индукции. Значит, $p_{s,1} > p_{s+1,1}$ при $s \geq 3$, и следовательно, $\mathbf{P}\{E_{1,s}\} > \mathbf{P}\{E_{1,s+1}\}$. Теорема 1 доказана.

Доказательство теоремы 2. Пусть $U = \{\alpha = (t, u): t = 1, \dots, T; u = s, \dots, s+r\}$, $W = (E_\alpha)_{\alpha \in U}$, $\Pi = (\pi_\alpha)_{\alpha \in U}$, где случайные величины, образующие набор Π , независимы в совокупности и имеют распределения Пуассона с параметрами $\mathbf{E}\pi_\alpha = \mathbf{P}\{E_\alpha\}$, $\alpha \in U$.

Для оценивания расстояния по вариации между распределениями вероятностей наборов $(\zeta_{s,T}, \dots, \zeta_{s+r,T})$ и $(\pi_{\lambda_0}, \dots, \pi_{\lambda_r})$ применим функциональный вариант метода Чена–Стейна (см. [1], гл. 10). Для каждой точки $(t, u) \in U$ выберем такое множество $\mathbb{O}(t, u) \subset U$, что событие $E_{t,u}$ не зависит от совокупности событий $\{E_{t',v}, (t', v) \in U \setminus \mathbb{O}(t, u)\}$. В нашем случае в множество $\mathbb{O}(t, u)$, $(t, u) \in U$, попадут все такие точки (t', v) , при которых события $E_{t,u}$ и $E_{t',v}$ зависят хотя бы от одного общего члена последовательности $X_{-1}, X_0, X_1, \dots, X_T, \dots$. Поэтому

$$\mathbb{O}(t, u) = \{(t', v): \max\{1, t-v-3\} \leq t' \leq t+u+3, \quad v = s, \dots, s+r\}.$$

Тогда, согласно теореме 10.A книги [1], расстояние по вариации между наборами W и Π оценивается неравенством

$$\rho(W, \Pi) \leq S_1 + S_2, \quad (18)$$

где

$$S_1 = \sum_{(t,u) \in U} \sum_{(t',v) \in \mathbb{O}(t,u)} \mathbf{P}\{E_{t,u}\} \mathbf{P}\{E_{t',v}\}, \quad S_2 = \sum_{(t,u) \in U} \sum_{(t',v) \in \mathbb{O}(t,u) \setminus \{(t,u)\}} \mathbf{P}\{E_{t,u} E_{t',v}\}.$$

Сначала оценим сумму S_1 . В силу определений,

$$S_1 = \sum_{t=1}^T \sum_{u=s}^{s+r} \sum_{v=s}^{s+r} \sum_{t'=\max\{1, t-v-3\}}^{t+u+3} \mathbf{P}\{E_{t,u}\} \mathbf{P}\{E_{t',v}\} \leq \sum_{t=1}^T \sum_{u=s}^{s+r} \sum_{v=s}^{s+r} \sum_{t'=t-(s+r)-3}^{t+s+r+3} \mathbf{P}\{E_{1,u}\} \mathbf{P}\{E_{1,v}\}.$$

Согласно (5) и теореме 1,

$$\mathbf{P}\{E_{1,s+u}\} = \frac{\lambda_u}{T} \leq \frac{\lambda_0}{T}. \quad (19)$$

Таким образом,

$$S_1 \leq \frac{2(s+r)+7}{T} \sum_{u=0}^r \sum_{v=0}^r \lambda_u \lambda_v \leq \frac{2(s+r)+7}{T} (r+1)^2 \lambda_0^2. \quad (20)$$

Теперь перейдем к оцениванию суммы S_2 . Нетрудно проверить, что события $E_{t,u}$ и $E_{t',v}$, $(t', v) \in \mathbb{O}(t, u)$, несовместны при $t' \notin \{t-v-3, t-v-2, t, t+u+2, t+u+3\}$. При $t' \in \{t-v-3, t+u+3\}$

$$\mathbf{P}\{E_{t,u} E_{t',v}\} = \frac{1}{1-p} \mathbf{P}\{E_{t,u}\} \mathbf{P}\{E_{t',v}\} = \frac{1}{1-p} \mathbf{P}\{E_{1,u}\} \mathbf{P}\{E_{1,v}\},$$

а при $t' \in \{t-v-2, t+u+2\}$

$$\mathbf{P}\{E_{t,u} E_{t',v}\} = \frac{1}{(1-p)^2} \mathbf{P}\{E_{t,u}\} \mathbf{P}\{E_{t',v}\} = \frac{1}{(1-p)^2} \mathbf{P}\{E_{1,u}\} \mathbf{P}\{E_{1,v}\}.$$

С учетом двух последних равенств, оценки (19) и определения суммы S_2 , получим, что

$$\begin{aligned} S_2 &\leq 2 \sum_{t=1}^T \sum_{u=s}^{s+r} \sum_{v=s}^{s+r} \left(\frac{1}{1-p} \mathbf{P}\{E_{1,u}\} \mathbf{P}\{E_{1,v}\} + \frac{1}{(1-p)^2} \mathbf{P}\{E_{1,u}\} \mathbf{P}\{E_{1,v}\} \right) \\ &= \frac{2(2-p)}{(1-p)^2} \sum_{t=1}^T \sum_{u=s}^{s+r} \sum_{v=s}^{s+r} \mathbf{P}\{E_{1,u}\} \mathbf{P}\{E_{1,v}\} \leq \frac{2(2-p)}{T(1-p)^2} (r+1)^2 \lambda_0^2. \end{aligned} \quad (21)$$

Подставив оценки (20) и (21) в правую часть неравенства (18), получим, что

$$\rho(W, \Pi) \leq \frac{2}{T} (r+1)^2 \lambda_0^2 \left(s+r+4 + \frac{2-p}{(1-p)^2} \right).$$

Заметим, что вектор $(\zeta_{s,T}, \dots, \zeta_{s+r,T})$ может быть получен из набора W в результате применения к нему определенной функции ϕ . Распределение вектора $(\pi_{\lambda_0}, \dots, \pi_{\lambda_r})$ выбрано так, что оно совпадает с распределением вектора $\varphi(\Pi)$. Поэтому

$$\rho((\zeta_{s,T}, \dots, \zeta_{s+r,T}), (\pi_{\lambda_0}, \dots, \pi_{\lambda_r})) = \rho(\phi(W), \phi(\Pi)) \leq \rho(W, \Pi).$$

Теорема 2 доказана.

Доказательство теоремы 3. Пусть d — некоторое фиксированное положительное целое число. Согласно неравенству треугольника,

$$\begin{aligned} & \rho((\zeta_{s,T}, \dots, \zeta_{s+r,T}, \xi_{s+r+1,T}), (\pi_{\lambda_0}, \dots, \pi_{\lambda_r}, \pi_{\Lambda_{r+1}})) \\ & \leq \rho\left((\zeta_{s,T}, \dots, \zeta_{s+r,T}, \xi_{s+r+1,T}), \left(\zeta_{s,T}, \dots, \zeta_{s+r,T}, \sum_{k=r+1}^{r+d} \zeta_{s+k,T}\right)\right) \\ & \quad + \rho\left(\left(\zeta_{s,T}, \dots, \zeta_{s+r,T}, \sum_{k=r+1}^{r+d} \zeta_{s+k,T}\right), (\pi_{\lambda_0}, \dots, \pi_{\lambda_r}, \tilde{\pi})\right) \\ & \quad + \rho((\pi_{\lambda_0}, \dots, \pi_{\lambda_r}, \tilde{\pi}), (\pi_{\lambda_0}, \dots, \pi_{\lambda_r}, \pi_{\Lambda_{r+1}})), \end{aligned} \quad (22)$$

где $\tilde{\pi}$ есть случайная величина, распределенная по закону Пуассона с параметром $\sum_{k=r+1}^{r+d} \lambda_k$.

Используя неравенство

$$\rho(X, Y) \leq \mathbf{P}\{X \neq Y\} \quad (23)$$

при оценке первого слагаемого в правой части неравенства (22), получаем, что

$$\begin{aligned} & \rho\left((\zeta_{s,T}, \dots, \zeta_{s+r,T}, \xi_{s+r+1,T}), \left(\zeta_{s,T}, \dots, \zeta_{s+r,T}, \sum_{k=r+1}^{r+d} \zeta_{s+k,T}\right)\right) \\ & \leq \mathbf{P}\left\{\xi_{s+r+1,T} \neq \sum_{k=r+1}^{r+d} \zeta_{s+k,T}\right\} = \mathbf{P}\left\{\sum_{k=r+d+1}^{\infty} \zeta_{s+k,T} \geq 1\right\} \\ & \leq \mathbf{E}\left(\sum_{k=r+d+1}^{\infty} \zeta_{s+k,T}\right) = \sum_{k=r+d+1}^{\infty} \lambda_k. \end{aligned}$$

Вектор $(\zeta_{s,T}, \dots, \zeta_{s+r,T}, \sum_{k=r+1}^{r+d} \zeta_{s+k,T})$ может быть получен в результате применения к вектору $(\zeta_{s,T}, \dots, \zeta_{s+r+d,T})$ некоторой функции ψ и распределение вектора $(\pi_{\lambda_0}, \dots, \pi_{\lambda_r}, \tilde{\pi})$ совпадает с распределением вектора $\psi(\pi_{\lambda_0}, \dots, \pi_{\lambda_{r+d}})$, поэтому в силу теоремы 2 (где вместо r берем $r+d$)

$$\begin{aligned} & \rho\left(\left(\zeta_{s,T}, \dots, \zeta_{s+r,T}, \sum_{k=r+1}^{r+d+1} \zeta_{s+k,T}\right), (\pi_{\lambda_0}, \dots, \pi_{\lambda_r}, \tilde{\pi})\right) \\ & \leq \frac{2}{T}(r+d+1)^2 \lambda_0^2 \left(s+r+d+4 + \frac{2-p}{(1-p)^2}\right). \end{aligned}$$

Оценим последнее слагаемое в правой части неравенства (22). Используя для этого неравенство (23), получим, что

$$\begin{aligned} & \rho((\pi_{\lambda_0}, \dots, \pi_{\lambda_r}, \tilde{\pi}), (\pi_{\lambda_0}, \dots, \pi_{\lambda_r}, \pi_{\Lambda_{r+1}})) \leq \mathbf{P}\{\tilde{\pi} \neq \pi_{\Lambda_{r+1}}\} \\ & = \mathbf{P}\left\{\sum_{k=r+d+1}^{\infty} \pi_{\lambda_k} \geq 1\right\} \leq \mathbf{E}\left(\sum_{k=r+d+1}^{\infty} \pi_{\lambda_k}\right) = \sum_{k=r+d+1}^{\infty} \lambda_k. \end{aligned}$$

Подставив оценки всех слагаемых в правую часть неравенства (22), получим (8). Теорема 3 доказана.

Доказательство следствия 1. Положим $d = s - r - 1$ и заметим, что $\lambda_0 \leq 2cTq^s$. Тогда выражение в правой части неравенства (8) оценивается следующим образом:

$$\begin{aligned} & \frac{2}{T}(r+d+1)^2\lambda_0^2 \left(s+r+d+4 + \frac{2-p}{(1-p)^2} \right) + 2 \sum_{k=r+d+1}^{\infty} \lambda_k \\ & \leq \frac{2}{T}s^2(2cTq^s)^2 \left(2s+3 + \frac{2-p}{(1-p)^2} \right) + 2cT \sum_{k=s}^{\infty} (q^{s+k} - q_1^{s+k}) \\ & \leq 8Ts^3c^2q^{2s} \left(2 + \frac{3}{s} + \frac{2-p}{s(1-p)^2} + \frac{1}{4c(1-q)s^3} \right). \end{aligned}$$

Следствие 1 доказано.

Замечание 4. Метод Чена–Стейна можно применить к несколько более широкому набору случайных индикаторов, чем это делалось при доказательстве теоремы 2. Такая замена приводит к заметно более громоздким выкладкам, но позволяет получить более точные в ряде случаев оценки. Например, выражение в правой части оценки (9) следствия 1 при $s \geq 3$ принимает вид $\tilde{C}T(r+3)^3sq^{2s}$, где $\tilde{C} < 1$.

Доказательство теоремы 4. Первое утверждение следует непосредственно из оценки (9) и условия (10). Из него в свою очередь следует, что при любом целом неотрицательном числе r распределение случайной величины $\sum_{k=0}^r (k+1)\zeta_{s+k-2,T}$ сходится к распределению случайной величины $\sum_{k=0}^r (k+1)\tilde{\pi}_k$.

При $r \geq 0$

$$\begin{aligned} \mathbf{P} \left\{ \vartheta_{s,T} \neq \sum_{k=0}^r (k+1)\zeta_{s+k-2,T} \right\} &= \mathbf{P} \left\{ \sum_{k=r+1}^{\infty} \zeta_{s+k-2,T} \geq 1 \right\} \\ &\leq \mathbf{E} \sum_{k=r+1}^{\infty} \zeta_{s+k-2,T} = \sum_{k=r+1}^{\infty} q^{k-2}\lambda. \end{aligned}$$

Так как $\sum_{k=0}^{\infty} q^k\lambda < \infty$, выражения в правой части этой цепочки неравенств стремятся к нулю при $r \rightarrow \infty$. Значит, выполнено второе утверждение теоремы. Теорема 4 доказана.

Доказательство теоремы 5. Для доказательства воспользуемся следствием 1, в силу которого $\rho(\xi_{s_0+k,T}, \pi_{\Lambda_{s_0+k}}) \leq CT(s_0+k)^3q^{2(s_0+k)}$. Так как $\mathbf{P}\{v < s_0+k\} = \mathbf{P}\{\xi_{s_0+k,T} = 0\}$, справедливо равенство

$$\mathbf{P}\{v < s_0+k\} = e^{-\Lambda_{s_0+k}} + R_1, \quad (24)$$

где $R_1 = \theta_1 T(s_0+k)^3q^{2(s_0+k)}$ при некотором $|\theta_1| \leq C$. Заметим, что в силу (11)

$$s_0(T) + k(T) = \frac{1}{2\ln(1/q)} (\ln T + 3\ln \ln T) + a_1(T), \quad (25)$$

где $a_1(T) \rightarrow \infty$ при $T \rightarrow \infty$. Поэтому при любом $k \geq k(T)$

$$\begin{aligned} R_1 &= \theta_1 T (s_0 + k)^3 q^{2(s_0 + k)} \\ &= \theta_1 T \left(\frac{1}{2 \ln(1/q)} (\ln T + 3 \ln \ln T) + a_1(T) \right)^3 q^{2a_1(T) + (\ln T + 3 \ln \ln T) / \ln(1/q)} \\ &= \theta_1 \left(\frac{1}{2 \ln(1/q)} \left(1 + \frac{3 \ln \ln T}{\ln T} \right) + \frac{a_1(T)}{\ln T} \right)^3 q^{2a_1(T)}. \end{aligned}$$

Таким образом, для достаточно больших T

$$R_1 = \tilde{\theta}_1 \left(\frac{a_1(T)}{\ln T} \right)^3 q^{2a_1(T)} = o \left(\frac{1}{\ln^3 T} \right). \quad (26)$$

Далее заметим, что

$$e^{-\Lambda_{s_0+k}} = e^{-cT((q^{s_0+k})/(1-q) - q_1^{s_0+k}/(1-q_1))} = e^{-(cTq^{s_0+k}/(1-q) - cTq_1^{s_0+k}/(1-q_1))}.$$

Из соотношения (12) следует, что

$$s_0(T) + k(T) = a_2(T) + \frac{\ln T}{\ln(1/|q_1|)} \rightarrow \infty.$$

Поэтому

$$\begin{aligned} Tq_1^{s_0+k} &= (-1)^{s_0+k} T|q_1|^{s_0+k} = (-1)^{s_0+k} T|q_1|^{a_2(T) + (\ln T)/\ln(1/|q_1|)} \\ &= (-1)^{s_0+k} |q_1|^{a_2(T)} \rightarrow 0 \end{aligned}$$

при $T \rightarrow \infty$. Значит, при некоторых ограниченных функциях $A = A(T, q_1)$, $\tilde{A} = \tilde{A}(T, q_1)$ и $\tilde{\tilde{A}} = \tilde{\tilde{A}}(T, q_1)$ выполнены равенства

$$\begin{aligned} e^{-\Lambda_{s_0+k}} &= e^{-cTq^{s_0+k}/(1-q)} (1 + ATq_1^{s_0+k}) \\ &= e^{-cTq^{s_0+k}/(1-q)} (1 + \tilde{A}|q_1|^{a_2(T)}) = e^{-cTq^{s_0+k}/(1-q)} + \tilde{\tilde{A}}|q_1|^{a_2(T)}. \end{aligned} \quad (27)$$

Подставляя соотношения (26) и (27) в формулу (24), получаем, что при $T \rightarrow \infty$

$$\begin{aligned} \mathbf{P}\{v < s_0(T) + k\} &= \exp \left\{ -\frac{Tcq^{s_0(T)+k}}{1-q} \right\} + \tilde{\theta}_1 \left(\frac{a_1(T)}{\ln T} \right)^3 q^{2a_1(T)} + \tilde{\tilde{A}}|q_1|^{a_2(T)} \\ &= \exp \left\{ -\frac{Tcq^{s_0(T)+k}}{1-q} \right\} + o(1). \end{aligned}$$

Теорема 5 доказана.

4. Дополнения и замечания

Использованный при доказательстве теорем 2 и 3 метод Чена–Стейна успешно работает и в других задачах. Рассмотрим, например, задачу о числе серий знаков в последовательности независимых случайных величин $X = \{X_0, X_1, \dots, X_n, \dots\}$, для которых

$$\mathbf{P}\{X_i = 1\} = 1 - \mathbf{P}\{X_i = 0\} = p. \quad (28)$$

Определим события

$$\begin{aligned} B_{1,t,k} &= \{X_{t-1} = 0, X_t = \dots = X_{t+k-1} = 1, X_{t+k} = 0\}, \\ B_{0,t,k} &= \{X_{t-1} = 1, X_t = \dots = X_{t+k-1} = 0, X_{t+k} = 1\}. \end{aligned}$$

Событие $B_{1,t,k}$ ($B_{0,t,k}$) означает, что в момент t начинается серия единиц (нулей) длины k . Тогда случайные величины

$$\varsigma_1(k, n) = \sum_{t=1}^n I\{B_{1,t,k}\}, \quad (29)$$

$$\varsigma_0(k, n) = \sum_{t=1}^n I\{B_{0,t,k}\}, \quad (30)$$

равны числу тех серий единиц и нулей соответственно, которые начинаются в моменты времени от 1 до n и имеют длину k .

Определим случайные величины

$$\xi_\delta(s, n) = \sum_{k=0}^{\infty} \varsigma_\delta(s + k, n), \quad \delta = 0, 1, \quad (31)$$

которые равны числу серий знаков δ в последовательности X , начинающихся в моменты времени от 1 до n и имеющих длину не меньше s .

При любом целом положительном числе m определим векторы

$$\begin{aligned} Z'_1(s_1, m) &= (\varsigma_1(s_1, n), \dots, \varsigma_1(s_1 + m - 1, n), \xi_1(s_1 + m, n)), \\ Z'_0(s_0, m) &= (\varsigma_0(s_0, n), \dots, \varsigma_0(s_0 + m - 1, n), \xi_0(s_0 + m, n)), \\ \Pi'_1(m) &= (\pi_0^1, \dots, \pi_{m-1}^1, \pi^1), \quad \Pi'_0(m) = (\pi_0^0, \dots, \pi_{m-1}^0, \pi^0), \end{aligned}$$

где π_j^l и π^l — независимые в совокупности случайные величины, распределенные по закону Пуассона с параметрами

$$\lambda_i^{(1)}(p, n) = \mathbf{E}\pi_i^1 = np^{s_1+i}(1-p)^2, \quad i = 0, 1, \dots, m-1, \quad (32)$$

$$\lambda_j^{(0)}(p, n) = \mathbf{E}\pi_j^0 = n(1-p)^{s_0+j}p^2, \quad j = 0, 1, \dots, m-1, \quad (33)$$

$$\Lambda_{1,m}(p, n) = \mathbf{E}\pi^1 = np^{s_1+m}(1-p), \quad (34)$$

$$\Lambda_{0,m}(p, n) = \mathbf{E}\pi^0 = n(1-p)^{s_0+m}p. \quad (35)$$

При $m = 0$ векторы $Z'_1(s_1, 0)$, $Z'_0(s_0, 0)$, $\pi'_1(0)$ и $\pi'_0(0)$ понимаются как состоящие из одного последнего элемента. Рассуждениями, аналогичными тем, что использовались в разделе 2, удается доказать следующее утверждение.

Теорема 6. Пусть m — целое неотрицательное число, $s^* = \max\{s_0, s_1\} \geq m + 2$. Тогда

$$\begin{aligned} &\rho((Z'_1(s_1, m), Z'_0(s_0, m)), (\Pi'_1(m), \Pi'_0(m))) \\ &\leq 6(s^*)^2(\Lambda_{1,0}(p, n) + \Lambda_{0,0}(p, n))(p^{s_1-1} + (1-p)^{s_0-1}). \end{aligned} \quad (36)$$

Теорема 6 позволяет исследовать асимптотические свойства такой характеристики плотной серии, как ее вес. Весом плотной серии единиц мы называем число единиц, входящих в ее запись.

Пусть $Y = \{Y_{-1}, Y_0, Y_1, \dots, Y_T, \dots\}$ — последовательность независимых случайных величин, для которых $\mathbf{P}\{Y_i = 1\} = 1 - \mathbf{P}\{Y_i = 0\} = p$. Построим по ней последовательность $Y' = \{Y'_1, Y'_2, \dots, Y'_n\}$ исходов независимых геометрически распределенных случайных величин, выражающих расстояния между соседними единицами в Y , $\mathbf{P}\{Y'_l = l\} = p(1-p)^{l-1}$, $l = 1, 2, \dots$. По Y' определим последовательность независимых одинаково распределенных случайных индикаторов $Y'' = \{Y''_1, Y''_2, \dots, Y''_n\}$, где $Y''_t = I\{Y'_t \leq 2\}$. Тогда $\mathbf{P}\{Y''_t = 1\} = 1 - \mathbf{P}\{Y'_t = 0\} = P$, где $P = p + p(1-p)$. Плотными сериями веса k в последовательности Y отвечают цепочки единиц длины $k-1$ в последовательности Y'' . Оценка теоремы 6 влечет многомерную предельную теорему Пуассона для числа серий единиц заданных длин, начала которых лежат в отрезке последовательности Y'' длины $n = [Tp]$. Стандартными рассуждениями эта теорема переносится на плотные серии единиц заданного веса, начала которых лежат в отрезке последовательности Y длины T .

Пусть $\zeta'_{w,T}$ и $\xi'_{w,T}$ — число плотных серий веса w и число плотных серий веса не меньше w в последовательности Y соответственно. Получаем следующее утверждение.

Теорема 7. Пусть $T \rightarrow \infty$, m — целое неотрицательное число, а параметр w меняется так, что $Tr(1-P)^2P^{w-1} \rightarrow \lambda$, $0 < \lambda < \infty$. Тогда случайные величины $\zeta'_{w,T}, \zeta'_{w+1,T}, \dots, \zeta'_{w+m-1,T}, \xi'_{w+m,T}$ асимптотически независимы в совокупности и имеют в пределе распределения Пуассона с параметрами $\lambda, P\lambda, \dots, P^{m-1}\lambda, P^m\lambda/(1-P)$ соответственно.

Автор выражает признательность В. Г. Михайлову за постоянное внимание к работе и полезные замечания.

Список литературы

1. Barbour A. D., Holst L., Janson S., *Poisson approximation*. Oxford Univ. Press, Oxford, 1992.
2. Golić J. D., Constrained embedding probability for two binary strings. *SIAM J. Discrete Math.* (1996) **9**, №3, 360–364.
3. Михайлов В. Г., Меженная Н. М., Оценки для вероятности плотного вложения одной дискретной последовательности в другую. *Дискретная математика* (2005) **17**, №3, 19–27.
4. Меженная Н. М., Предельные теоремы для числа плотных серий в случайной последовательности. В сб.: *Научно-техническая конференция студентов, аспирантов и молодых специалистов МИЭМ. МИЭМ, Москва, 2008*, с. 11–13.
5. Меженная Н. М., О скорости сходимости распределения числа серий различной длины в случайной последовательности с «разладками» к распределению Пуассона. *Обозрение прикладной и промышленной математики* (2007) **14**, №2, 248–249.

Статья поступила 15.10.2008.