УДК 519.214

Предельные теоремы для числа плотных серий с заданными параметрами в выходной последовательности генератора Пола

© Н.М. Меженная

МГТУ им. Н.Э. Баумана, Москва, 105005, Россия

Работа посвящена изучению случайных величин, связанных с плотными сериями в выходной последовательности генератора Пола. С помощью метода Чена — Стейна получены оценки расстояния по вариации между распределением числа плотных серий заданных длины и веса в выходной последовательности генератора Пола с двумя регистрами и сопровождающим пуассоновским распределением. На основании этих оценок выведены предельные теоремы Пуассона для указанных случайных величин и, как следствие, центральная предельная теорема (в смысле сближения с распределением Пуассона с растущим параметром).

Ключевые слова: *плотные серии, генератор Пола, метод Чена — Стейна, предельная теорема Пуассона, центральная предельная теорема, расстояние по вариации.*

Введение. В настоящей работе рассмотрен генератор Пола, или мультициклический генератор [1], с двумя регистрами взаимно простых длин m и n над кольцом вычетов по модулю M ($M \geqslant 2$). Пусть $(X_0, \ldots, X_{m-1})$ и $(Y_0, \ldots, Y_{n-1})$ — случайные заполнения регистров длин m и n соответственно, состоящие из независимых равномерно распределенных на множестве $\{0, 1, \ldots, M-1\}$ случайных величин. Тогда выходная последовательность генератора Пола образуется в соответствии с правилом

$$Z_t = X_{t \bmod m} + Y_{t \bmod n} \bmod M, \quad t = 0, 1, 2, \ldots \tag{1}$$

Последовательность, полученная согласно формуле (1), имеет гарантированный период длины mn. Основные свойства выходной последовательности генератора Пола описаны в работе [1].

Многие статистические процедуры проверки свойств случайных последовательностей основаны на свойствах серий. В докладе [2] приведена предельная теорема Пуассона для числа цепочек без самоналожения (и, как следствие, для обычных серий) в выходной последовательности генератора Пола.

Настоящая работа посвящена изучению свойств числа плотных серий в выходной последовательности такого генератора.

Предельные теоремы. Напомним, что отрезок последовательности $\{x_1, \ldots, x_k\}$ образует плотную a-цепочку, если $a \in \{x_i, x_{i+1}\}$,

$i = 1, \ldots, k-1$. Плотная a-цепочка называется плотной a-серией, если ее нельзя продлить в обе стороны с сохранением этого свойства (место появления первого знака a будем называть ее началом). Длина плотной a-серии — число знаков в отрезке последовательности, содержащем все ее знаки a, а вес плотной a-серии — число входящих в нее знаков a. Свойства плотных серий в последовательности независимых случайных величин рассмотрены в работах [3, 4].

Пусть $\zeta_{s,w}$ — число плотных a-серий длины s и веса w в выходной последовательности генератора Пола (1), которые начались до момента $T \leqslant mn$. Будем считать, что выполнено естественное требование $w \geqslant 1$, $w \leqslant s \leqslant 2w-1$. Если $s+4 < \max\{m,n\}$, то среднее число плотных серий веса s и длины w

$$\lambda_{s,w} = \mathbf{E}\zeta_{s,w} = T\left(1-\frac{1}{M}\right)^4 p_{s,w}, \quad p_{s,w} = C_{w-1}^{s-w}\frac{1}{M^w}\left(1-\frac{1}{M}\right)^{s-w}. \quad (2)$$

Эта формула получена в работе [4]. Требование $s+4 < \max\{m,n\}$ означает, что плотная a-серия в выходной последовательности Z_t образована независимыми равномерно распределенными случайными величинами из заполнений регистров $(X_0, \ldots, X_{m-1})$ и $(Y_0, \ldots, Y_{n-1})$.

Обозначим через $\rho(U,V)$ расстояние по вариации между распределениями случайных величин U и V. Для случайных величин, распределенных на множестве неотрицательных целых чисел,

$$\rho(U,V) = \frac{1}{2}\sum_{k=0}^{\infty}|\mathbf{P}\{U=k\} - \mathbf{P}\{V=k\}|.$$

Теорема 1. *Пусть $m,n \geqslant 1$, $T \leqslant mn$, $w \geqslant 1$, $w \leqslant s \leqslant 2w-1$, а случайные величины $X_0, \ldots, X_{m-1}, Y_0, \ldots, Y_{n-1}$ независимы и равномерно распределены на множестве $A_M = \{0, 1, \ldots, M-1\}$. Тогда*

$$\rho(\zeta_{s,w}, \pi_{s,w}) \leqslant 2(m+n)(2s+7)p_{s,w},$$

где $\pi_{s,w}$ — случайная величина, распределенная в соответствии с законом Пуассона с параметром $\lambda_{s,w}$.

Следствие 1. *Пусть случайные величины $X_0, \ldots, X_{m-1}, Y_0, \ldots, Y_{n-1}$ независимы и равномерно распределены на множестве A_M, $m,n,w \to \infty$, $T \leqslant mn$, так что*

$$\lambda_{s,w} \to \lambda \in (0,\infty) \quad \text{и} \quad \frac{\ln m}{n} + \frac{\ln n}{m} \to 0.$$

Тогда закон распределения случайной величины $\zeta_{s,w}$ сходится к закону Пуассона распределения случайной величины с параметром λ.

Следствие 2. *Пусть случайные величины $X_0, \ldots, X_{m-1}, Y_0, \ldots, Y_{n-1}$ независимы и равномерно распределены на множестве A_M, $m,n,w \to \infty$,*

$T \leqslant mn$, *так что*

$$\lambda_{s,\,w} \to \infty \quad \textit{и} \quad s\left(\frac{\lambda_{s,\,w}}{n} + \frac{\lambda_{s,\,w}}{m}\right) \to 0.$$

Тогда закон распределения случайной величины $(\zeta_{s,\,w} - \lambda_{s,\,w})/\sqrt{\lambda_{s,\,w}}$ *сходится к стандартному нормальному закону распределения.*

Обозначим

$$\zeta_w = \sum_{s=w}^{2w-1} \zeta_{s,\,w}$$

число плотных a-серий веса w в выходной последовательности генератора Пола (1), которые начались до момента $T \leqslant mn$.

Из выражения (2) следует, что среднее число плотных a-серий веса w (и любой длины)

$$\lambda_w = \mathbf{E}\zeta_w = \sum_{s=w}^{2w-1} \lambda_{s,\,w} = T \sum_{s=w}^{2w-1} C_{w-1}^{s-w} \frac{1}{M^w}\left(1 - \frac{1}{M}\right)^{s-w+4} =$$

$$= T\left(1 - \frac{1}{M}\right)^4 p_w, \quad p_w = \frac{1}{M^w}\left(2 - \frac{1}{M}\right)^{w-1}.$$

Последняя формула получена в работе [3].

Теорема 2. *Пусть* $m, n \geqslant 1$, $T \leqslant mn$, $w \geqslant 1$, *а случайные величины* $X_0, \ldots, X_{m-1}$, $Y_0, \ldots, Y_{n-1}$ *независимы и равномерно распределены на множестве* A_M. *Тогда*

$$\rho(\zeta_w, \pi_w) \leqslant 2(m+n)(4w+5)p_w,$$

где π_w *— случайная величина, распределенная в соответствии с законом Пуассона с параметром* λ_w.

Следствие 3. *Пусть случайные величины* $X_0, \ldots, X_{m-1}, Y_0, \ldots, Y_{n-1}$ *независимы и равномерно распределены на множестве* A_M, $m, n, w \to \infty$, $T \leqslant mn$, *так что*

$$\lambda_w \to \lambda \in (0, \infty) \quad \textit{и} \quad \frac{\ln m}{n} + \frac{\ln n}{m} \to 0.$$

Тогда закон распределения случайной величины ζ_w *сходится к закону Пуассона распределения случайной величины с параметром* λ.

Замечание 1. В условиях следствия 1 и 3 вес $w = O(\ln T)$.

Следствие 4. *Пусть случайные величины* $X_0, \ldots, X_{m-1}, Y_0, \ldots, Y_{n-1}$ *независимы и равномерно распределены на множестве* A_M, $m, n, w \to \infty$, $T \leqslant mn$, *так что*

$$\lambda_w \to \infty \quad \textit{и} \quad w\left(\frac{\lambda_w}{n} + \frac{\lambda_w}{m}\right) \to 0.$$

Тогда закон распределения случайной величины $(\zeta_w - \lambda_w)/\sqrt{\lambda_w}$ *сходится к стандартному нормальному закону распределения.*

Замечание 2. Можно показать, что условия следствия 4 выполнены, если $T = mn$, $m > n$, $n \to \infty$,

$$w = \left[\frac{\ln mn - \ln\ln mn}{\ln P^{-1}}\right], \quad P = \frac{1}{M}\left(2 - \frac{1}{M}\right), \quad \frac{\ln m \ln\ln m}{n} \to 0.$$

Доказательство теоремы 1. Отметим, что распределение последовательности

$$\tilde{Z}_t = a - X_{t \bmod m} + Y_{t \bmod n} \bmod M, \quad t = 0, 1, 2, \ldots, \tag{3}$$

совпадает с распределением последовательности (1), так как в условиях теоремы распределения наборов

$$(X_0, \ldots, X_{m-1}) \quad \text{и} \quad (a - X_0, \ldots, a - X_{m-1})$$

совпадают. Знак минус означает вычитание по модулю M. Формула (3) удобнее тем, что

$$\{\tilde{Z}_t = a\} = \{X_{t \bmod m} = Y_{t \bmod n}\}.$$

Пусть $W_{i,j}^{s,w}$ — событие, состоящее в том, что в момент t в последовательности $\tilde{Z}_0, \tilde{Z}_1, \ldots$ началась плотная a-серия длины s и веса w, $i = t \bmod m$, $j = t \bmod n$ (такая серия начинается со знаков X_i и Y_j).

Обозначим $U_T = \{(t \bmod m, t \bmod n)\colon t = 0, 1, \ldots, T-1\}$, при $(i,j) \in U_T$

$$U_{i,j} = U_i \cup U_j,$$
$$U_i = \{(i', j') \in U_T\colon |i' - i| \leqslant s + 3\},$$
$$U_j = \{(i', j') \in U_T\colon |j' - j| \leqslant s + 3\}.$$

Тогда

$$|U_{i,j}| \leqslant |U_i| + |U_j| = \big(2(s+3)+1\big)(m+n). \tag{4}$$

Согласно методу Чена — Стейна [5],

$$\rho(\zeta_{s,w}, \pi_{s,w}) \leqslant \frac{1 - e^{-\lambda_{s,w}}}{\lambda_{s,w}}(S_1 + S_2), \tag{5}$$

где

$$S_1 = \sum_{(i,j)\in U_T} \sum_{(i',j')\in U_{i,j}} \mathbf{P}\{W_{i,j}^{s,w}\}\mathbf{P}\{W_{i',j'}^{s,w}\},$$
$$S_2 = \sum_{(i,j)\in U_T} \sum_{(i',j')\in \dot{U}_{i,j}} \mathbf{P}\{W_{i,j}^{s,w} W_{i',j'}^{s,w}\}. \tag{6}$$

Оценим сумму S_1. В условиях теоремы 1

$$\mathbf{P}\{W_{i,j}^{s,w}\} = \mathbf{P}\{W_{0,0}^{s,w}\} = \left(1 - \frac{1}{M}\right)^4 p_{s,w} < p_{s,w},$$

поэтому из (4) имеем

$$S_1 = \sum_{(i,j)\in U_T} \sum_{(i',j')\in U_{i,j}} \mathbf{P}\{W_{i,j}^{s,w}\}\mathbf{P}\{W_{i',j'}^{s,w}\} <$$

$$< \sum_{(i,j)\in U_T} |U_{i,j}| p_{s,w}^2 \left(1 - \frac{1}{M}\right)^4 \leqslant \left(1 - \frac{1}{M}\right)^4 T\,(2s+7)\,(m+n) p_{s,w}^2 =$$

$$= \lambda_{s,w}\,(2s+7)\,(m+n) p_{s,w}. \quad (7)$$

Теперь оценим сумму S_2 с использованием следующей леммы.

Лемма 1. *Если* $(i',j') \in U_{i,j} \setminus \{(i,j)\}$, *то*

$$\mathbf{P}\{W_{i,j}^{s,w} W_{i',j'}^{s,w}\} \leqslant p_{s,w}^2 \left(1 - \frac{1}{M}\right)^4 = \left(C_{w-1}^{s-w} \frac{1}{M^w} \left(1 - \frac{1}{M}\right)^{s-w+2}\right)^2. \quad (8)$$

С помощью формулы (8) оценим правую часть выражения (6):

$$S_2 = \sum_{(i,j)\in U_T} \sum_{(i',j')\in \dot{U}_{i,j}} \mathbf{P}\{W_{i,j}^{s,w} W_{i',j'}^{s,w}\} < \sum_{(i,j)\in U_T} |U_{i,j}| p_{s,w}^2 \left(1 - \frac{1}{M}\right)^4 \leqslant$$

$$\leqslant \left(1 - \frac{1}{M}\right)^4 T\,(2s+7)\,(m+n) p_{s,w}^2 = \lambda_{s,w}\,(2s+7)\,(m+n) p_{s,w}. \quad (9)$$

Подставляя (7) и (9) в формулу (5), получаем

$$\rho(\zeta_{s,w}, \pi_{s,w}) \leqslant 2 \frac{1 - e^{-\lambda_{s,w}}}{\lambda_{s,w}} \lambda_{s,w}\,(2s+7)\,(m+n) p_{s,w} <$$

$$< 2\,(2s+7)\,(m+n) p_{s,w}.$$

Теорема 1 доказана. □

Доказательство теоремы 2. Доказательство теоремы 2 проведем аналогично доказательству теоремы 1. Так как при весе w максимальная длина серии $s = 2w - 1$, окрестности определяются формулами

$$\bar{U}_{i,j} = \bar{U}_i \cup \bar{U}_j,$$
$$\bar{U}_i = \{(i',j') \in U_T : |i' - i| \leqslant 2w + 2\},$$
$$\bar{U}_j = \{(i',j') \in U_T : |j' - j| \leqslant 2w + 2\}.$$

Из формулы (4) получаем

$$|\bar{U}_{i,j}| \leqslant |\bar{U}_i| + |\bar{U}_j| = \big(2(2w+2)+1\big)(m+n) = (4w+5)\,(m+n).$$

Далее все рассуждения полностью повторяют рассуждения при доказательстве теоремы 1, в которых окрестности $U_{i,j}$ заменены на $\bar{U}_{i,j}$, а вероятности $p_{s,w}$ на p_w. Имеем

$$S_1 < \sum_{(i,j)\in U_T} |\bar{U}_{i,j}| p_w^2 \left(1-\frac{1}{M}\right)^4 \leqslant$$

$$\leqslant \left(1-\frac{1}{M}\right)^4 T\,(4w+5)\,(m+n)p_w^2 = \lambda_w\,(4w+5)\,(m+n)p_w,$$

$$S_2 < \sum_{(i,j)\in U_T} |\bar{U}_{i,j}| p_w^2 \left(1-\frac{1}{M}\right)^4 \leqslant \lambda_w\,(4w+5)\,(m+n)p_w,$$

значит, расстояние по вариации $\rho(\zeta_w, \pi_w) \leqslant (4w+5)\,(m+n)p_w$.

Теорема 2 доказана. □

Доказательство леммы 1. Отметим, если $1 \leqslant |i-i'| = |j-j'| \leqslant s$, то $\mathbf{P}\{W_{i,j}^{s,w} W_{i',j'}^{s,w}\} = 0$. Например, при $1 \leqslant i-i' = j-j' \leqslant s$ одна из серий начинается со знаков X_i и Y_j и $X_{i-2} \neq Y_{i-2}$, $X_{i-1} \neq Y_{i-1}$, но знаки X_{i-2}, Y_{i-2}, X_{i-1}, Y_{i-1} образуют плотную серию совпадений знаков, начинающихся со знаков $X_{i'}$ и $Y_{j'}$.

Пусть $\tilde{W}_{i,j}^{s,w}$ — событие, состоящее в том, что в момент t в последовательности $\tilde{Z}_0, \tilde{Z}_1, \ldots$ началась плотная a-цепочка длины s и веса w. Очевидно, что

$$\begin{aligned} &\mathbf{P}\{W_{i,j}^{s,w}\} \leqslant P\{\tilde{W}_{i,j}^{s,w}\}\left(1-\frac{1}{M}\right)^2, \\ &\mathbf{P}\{W_{i,j}^{s,w} W_{i',j'}^{s,w}\} \leqslant \mathbf{P}\{\tilde{W}_{i,j}^{s,w}\tilde{W}_{i',j'}^{s,w}\}\left(1-\frac{1}{M}\right)^4. \end{aligned} \tag{10}$$

Если $s+1 \leqslant |i-i'| = |j-j'| \leqslant s+3$, то

$$\mathbf{P}\{W_{i,j}^{s,w} W_{i',j'}^{s,w}\} \leqslant \mathbf{P}\{\tilde{W}_{i,j}^{s,w}\tilde{W}_{i',j'}^{s,w}\}\left(1-\frac{1}{M}\right)^4 \leqslant$$

$$\leqslant \mathbf{P}\{\tilde{W}_{i,j}^{s,w}\}\mathbf{P}\{\tilde{W}_{i',j'}^{s,w}\}\left(1-\frac{1}{M}\right)^4.$$

Таким образом, формула (8) для случая $1 \leqslant |i-i'| = |j-j'| \leqslant s+3$ доказана. Далее будем считать, что $|i-i'| \neq |j-j'|$.

Зафиксируем конфигурации C_1 и C_2 совпадающих и несовпадающих знаков отдельно на отрезках $(X_i, \ldots, X_{i+s-1})$ и $(Y_j, \ldots, Y_{j+s-1})$ и на отрезках $(X_{i'}, \ldots, X_{i'+s-1})$ и $(Y_{j'}, \ldots, Y_{j'+s-1})$ соответственно. Предположим, что обе конфигурации совпадений могут осуществиться вместе (в противном случае вероятность их совместного появления равна 0 и лемма доказана). Число таких различных конфигураций для

каждой пары отрезков равно C_{w-1}^{s-w}. Поэтому общее число конфигураций совпадений $C = C_1 \cap C_2$, которые могут осуществиться вместе, не превосходит $(C_{w-1}^{s-w})^2$.

Обозначим через $\tilde{W}_{i,j,C_1}^{s,w}$ и $\tilde{W}_{i',j',C_2}^{s,w}$ события $\tilde{W}_{i,j}^{s,w}$ и $\tilde{W}_{i',j'}^{s,w}$ при фиксированных конфигурациях C_1 и C_2. Тогда

$$\tilde{W}_{i,j}^{s,w} = \bigcup_{C_1} \tilde{W}_{i,j,C_1}^{s,w}, \quad \tilde{W}_{i',j'}^{s,w} = \bigcup_{C_2} \tilde{W}_{i',j',C_2}^{s,w}. \tag{11}$$

Рассмотрим граф $\Gamma = (G, V)$, в котором множество вершин $G = I \cup J$,

$$I = \{i, \ldots, i+s-1\} \cup \{i', \ldots, i'+s-1\},$$
$$J = \{j, \ldots, j+s-1\} \cup \{j', \ldots, j'+s-1\}.$$

Множество ребер построим следующим образом. Пусть $u, v \in G$. Если при конфигурации совпадений C знаки X_u и X_v связаны отношением равенства или неравенства, то вершины u и v будут связаны ребром первого или второго типа соответственно. При таком построении каждой вершине инцидентны не более двух ребер. Таким образом, множество ребер E имеет вид

$$E = \{(i+k, j+k),\ k = 0, 1, \ldots, s-1\} \cup \{(i'+k, j'+k),\ k = 0, 1, \ldots, s-1\}.$$

Ребрам первого типа припишем метку $p = \{X_1 = Y_1\} = 1/M$, а ребрам второго типа — метку $1 - p = 1 - 1/M$.

Приведем следующие важные для нас свойства графа Γ:

1) граф Γ является двудольным с долями I и J;

2) каждой вершине инциденты не более двух ребер;

3) граф Γ не имеет циклов и параллельных ребер;

4) граф Γ состоит из одной или нескольких цепей ребер.

Из этих свойств графа Γ и независимости всех случайных величин $\{X_0, \ldots, X_{m-1}\}$ и $\{Y_0, \ldots, Y_{n-1}\}$ следует, что вероятность $\mathbf{P}\{\tilde{W}_{i,j}^{s,w} \tilde{W}_{i',j'}^{s,w}\}$ равна произведению меток всех ребер графа Γ, поэтому

$$\mathbf{P}\{\tilde{W}_{i,j}^{s,w} \tilde{W}_{i',j'}^{s,w}\} = p^{2w}(1-p)^{2(s-w)}. \tag{12}$$

Из выражений (10)–(12) следует формула (8).

Лемма 1 доказана. □

Заключение. С помощью известного метода Чена — Стейна установлена скорость сближения распределения числа плотных серий заданных длины и веса в выходной последовательности генератора Пола с двумя регистрами с их сопровождающими пуассоновскими распределениями. Это позволяет получить пуассоновскую и нормальную предельные теоремы для указанных случайных величин при определенном изменении параметров схемы и указать скорость сходимости в них. Эти результаты дополняют полученные ранее в работах [2–4].

Автор выражает признательность ведущему научному сотруднику Математического института им. В. А. Стеклова РАН В. Г. Михайлову за идею доказательства леммы 1 и ряд полезных замечаний.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект 11-01-00139-а).

ЛИТЕРАТУРА

[1] Pohl P. Description of MCV, a Pseudo-random Number Generator. *Scand. Actuarial J.*, 1976, no. 1, pp. 1–14.

[2] Меженная Н.М., Михайлов В.Г. Вероятностные свойства выходной последовательности генератора Пола. *Семинар отдела дискретной математики МИАН*. URL: http://www.mathnet.ru/php/seminars.phtml ?option_lang=rus&presentid=6239 (дата обращения 10.05.2013).

[3] Меженная Н.М. Предельные теоремы для числа плотных серий в случайной последовательности. *Дискретная математика*, 2009, т. 21, вып. 1, с. 105–116.

[4] Меженная Н.М. Предельная теорема Пуассона для числа плотных серий заданной длины и веса. *Вестник МГТУ им. Н. Э. Баумана. Сер. Естественные науки*, 2011, спец. вып. *Прикладная математика*, с. 75–82.

[5] Barbour A.D., Holst L., Janson S. *Poisson Approximation*. Oxford, Oxford University Press, 1992, 277 p.

Статья поступила в редакцию 15.05.2013

Ссылку на эту статью просим оформлять следующим образом:

Меженная Н.М. Предельные теоремы для числа плотных серий с заданными параметрами в выходной последовательности генератора Пола. *Инженерный журнал: наука и инновации, 2013, вып. 4*. URL: http://engjournal.ru/catalog/fundamentals/math/661.html

Меженная Наталья Михайловна — канд. физ.-мат. наук, доц. кафедры «Прикладная математика» МГТУ им. Н. Э. Баумана. e-mail: natalia.mezhennaya@gmail.com