УДК 519.214 DOI 10.17223/20710410/58/2

# О ТОЧНОСТИ НОРМАЛЬНОЙ АППРОКСИМАЦИИ ДЛЯ РАСПРЕДЕЛЕНИЯ ЧИСЛА ПОВТОРЕНИЙ В СТАЦИОНАРНОЙ ДИСКРЕТНОЙ СЛУЧАЙНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ

В. Г. Михайлов*, Н. М. Меженная**

*Математический институт им. В. А. Стеклова Российской академии наук, г. Москва, Россия

**Московский государственный технический университет им. Н. Э. Баумана, г. Москва, Россия

**E-mail:** mikhail@mi-ras.ru, natalia.mezhennaya@gmail.com

Изучается задача об асимптотической нормальности числа $r$-кратных повторений знаков в отрезке стационарной (в узком смысле) дискретной случайной последовательности на множестве $\{1, 2, \ldots, N\}$, обладающей свойством равномерно сильного перемешивания. Показано, что в случае, когда коэффициент равномерно сильного перемешивания $\varphi(t)$ при произвольно заданном $\alpha > 0$ убывает как $t^{-6-\alpha}$, расстояние в равномерной метрике между функцией распределения числа повторений и функцией распределения стандартного нормального закона с увеличением длины последовательности $n$ убывает со скоростью $O(n^{-\delta})$ для любого $\delta \in (0; \alpha(32 + 4\alpha)^{-1}))$.

**Ключевые слова:** *нормальная аппроксимация, число кратных повторений, стационарная случайная последовательность, равномерно сильное перемешивание, расстояние в равномерной метрике.*

# ABOUT THE RATE OF NORMAL APPROXIMATION FOR THE DISTRIBUTION OF THE NUMBER OF REPETITIONS IN A STATIONARY DISCRETE RANDOM SEQUENCE

V. G. Mikhailov*, N. M. Mezhennaya**

*Steklov Mathematical Institute of Russian Academy of Sciences, Moscow, Russia

**Bauman Moscow State Technical University, Moscow, Russia

The paper presents the problem of asymptotic normality of the number of $r$-fold repetitions of characters in a segment of a (strictly) stationary discrete random sequence on the set $\{1, 2, \ldots, N\}$ with the uniformly strong mixing property. It is shown that in the case when the uniformly strong mixing coefficient $\varphi(t)$ for an arbitrarily given $\alpha > 0$ decreases as $t^{-6-\alpha}$, then the distance in the uniform metric between the distribution function of the number of repetitions and the distribution function of the standard normal law decreases at a rate of $O(n^{-\delta})$ with increasing sequence length $n$ for any $\delta \in (0; \alpha(32 + 4\alpha)^{-1}))$.

**Keywords:** *normal approximation, number of multiple repetitions, stationary random sequence, uniformly strong mixing, distance in uniform metric.*

## Введение

В настоящее время большую роль играет задача генерации случайных или псевдослучайных последовательностей как основы имитационного моделирования изучаемых процессов в науке и технике, что делает необходимой проверку соответствия их свойств требованиям модели. Особенно важно это для криптографических применений. Для такой проверки в криптографической литературе разработаны наборы статистических тестов [1, 2].

Например, отдельно изучены свойства статистик, основанных на числах повторений знаков, в том числе кратных, в независимых и конечно зависимых последовательностях случайных величин. Наиболее успешными оказались исследования условий сходимости распределений чисел повторений к распределению Пуассона [3, 4]. В качестве примера построения достаточных условий асимптотической нормальности числа повторений следует привести работу А. М. Шойтова [5].

В последние годы активно велись исследования повторяемости знаков в дискретных цепях Маркова. Например, в работах [6–8] исследована возможность аппроксимации распределения чисел повторений цепочек в цепи Маркова распределением Пуассона. Аналогичная задача о доказательстве центральной предельной теоремы для таких случайных величин решена в [9]. В частности, в [9, теорема 3] установлена оценка скорости сходимости к нормальному закону для числа кратных совпадений знаков в конечной стационарной цепи Маркова. В настоящей работе показано, что вместо цепей Маркова можно рассматривать более широкий класс — стационарные случайные последовательности, удовлетворяющие условию равномерно сильного перемешивания.

## 1. Оценка скорости сходимости в центральной предельной теореме

Пусть $X_1, \ldots, X_n$ — отрезок стационарной (в узком смысле) случайной последовательности со значениями из множества $\{1, \ldots, N\}$, стационарным распределением $\mathsf{P}[X_1 = k] = p_k \in (0; 1)$, $k = 1, \ldots, N$, $p_1 + \ldots + p_N = 1$, удовлетворяющей условию равномерно сильного перемешивания.

Напомним [10, с. 361], что процесс $\{X_t\}_{-\infty}^{\infty}$ называется стационарным (в узком смысле), если распределения случайных векторов вида $(X_{t_1+h}, \ldots, X_{t_s+h})$ при всех натуральных $s$ не зависят от $h$.

Стационарная последовательность $\{X_t\}_{-\infty}^{\infty}$ обладает свойством *равномерно сильного перемешивания* [10, с. 391], если

$$\varphi(t) = \sup_{A \in \mathcal{F}_{-\infty}^{0}, B \in \mathcal{F}_{t}^{\infty}} |\mathsf{P}[B|A] - \mathsf{P}[B]| \downarrow 0, \quad t \to \infty,$$

где $\mathcal{F}_a^b$ — $\sigma$-алгебра событий, порождённая величинами $X_a, \ldots, X_b$.

В работе [9] доказаны теоремы, описывающие условия асимптотической нормальности при $n \to \infty$ распределения случайной величины

$$\xi_r = \sum_{1 \leqslant j_1 < \ldots < j_r \leqslant n} I\{X_{j_1} = \ldots = X_{j_r}\}$$

— числа $r$-кратных повторений знаков в отрезке $X_1, \ldots, X_n$ (здесь и далее $I\{A\}$ — индикатор события $A$). В частности, доказана следующая теорема 1.

Введём случайные величины

$$\zeta_k = \sum_{t=1}^{n} I\{X_t = k\}, \quad \eta_k = \zeta_k - np_k, \quad k = 1, \ldots, N,$$

$$U_n = \frac{1}{(r-1)!} \sum_{k=1}^{N} p_k^{r-1} \eta_k.$$

Рассмотрим центрированную и нормированную случайную величину

$$\xi_r^* = \frac{\xi_r}{n^{r-1}\sqrt{\mathsf{D}U_n}} - \frac{1}{n^{r-1} r! \sqrt{\mathsf{D}U_n}} \sum_{k=1}^{N} \left( \prod_{j=0}^{r-1} (np_k - j) \right).$$

**Теорема 1** [9, теорема 3]. Пусть $X_1, \ldots, X_n$ — отрезок простой стационарной цепи Маркова с множеством состояний $\{1, \ldots, N\}$ и положительной матрицей переходных вероятностей. Тогда для любого $0 < \delta < 1/4$

$$\sup_{-\infty < x < \infty} |\mathsf{P}[\xi_r^* \leqslant x] - \Phi(x)| = O\left(n^{-\delta}\right), \quad n \to \infty.$$

Оказывается, что аналогичный результат может быть получен в несколько более общем случае — для стационарных последовательностей, удовлетворяющих условию равномерно сильного перемешивания. Основной результат работы имеет следующий вид:

**Теорема 2.** Пусть $X_1, \ldots, X_n$ — отрезок стационарной (в узком смысле) случайной последовательности со значениями из множества $\{1, \ldots, N\}$, со стационарным распределением $\mathsf{P}[X_1 = k] = p_k \in (0; 1)$, $k = 1, \ldots, N$, $p_1 + \ldots + p_N = 1$. Пусть среди вероятностей $p_1, \ldots, p_N$ есть различные, а для коэффициента равномерно сильного премешивания $\varphi$ последовательности $X_1, \ldots, X_n$ при некотором $\alpha > 0$ выполнено условие

$$\varphi(t) \leqslant t^{-6-\alpha}.$$

Тогда для любого $0 < \delta < \dfrac{\alpha}{4(8+\alpha)}$

$$\sup_{-\infty < x < \infty} |\mathsf{P}[\xi_r^* < x] - \Phi(x)| = O\left(n^{-\delta}\right), \quad n \to \infty,$$

где $\Phi(\cdot)$ — функция распределения стандартного нормального закона.

**Замечание 1.** В условиях теоремы 2 выполнены неравенства $0 < \delta < 1/4$.

## 2. Доказательство теоремы 2

Ясно, что последовательности

$$I\{X_1 = k\} - p_k, \ \ldots, \ I\{X_n = k\} - p_k$$

при всех $k = 1, \ldots, N$ являются стационарными и обладают свойством равномерно сильного перемешивания. Такими же свойствами обладают последовательности

$$u_1 = \sum_{k=1}^{N} p_k^{r-1} (I\{X_1 = k\} - p_k), \ \ldots, \ u_n = \sum_{k=1}^{N} p_k^{r-1} (I\{X_n = k\} - p_k).$$

Нетрудно проверить, что

$$U_n = \frac{1}{(r-1)!} (u_1 + u_2 + \ldots + u_n).$$

**Замечание 2.** В равновероятном случае, когда $p_1 = \ldots = p_N = 1/N$, величины $U_n$, $u_1, u_2, \ldots, u_n$ с вероятностью единица равны нулю. Условия теоремы 2 исключают эту ситуацию.

Нам понадобится ряд вспомогательных утверждений.

**Лемма 1** [9, теорема 2]. Пусть выполнены условия теоремы 2 и $R_n$ — натуральное число, удовлетворяющее условию $R_n < n$. Тогда

$$D_n = \sup_{-\infty<x<\infty} \left| \mathsf{P}\left[\frac{U_n}{\sqrt{\mathsf{D}U_n}} < x\right] - \Phi(x) \right| \leqslant$$
$$\leqslant \sqrt[4]{\frac{2}{\pi}} \left( 4N^3 \frac{nR_n^2}{(\sqrt{\mathsf{D}U_n})^3} + \frac{6}{\pi(r-1)!} \frac{n^2\varphi(R_n)}{\sqrt{\mathsf{D}U_n}} + \frac{2N^2}{\pi} \frac{n^2\varphi(R_n)}{\mathsf{D}U_n} \right)^{1/2}. \tag{1}$$

**Лемма 2.** Пусть выполнены условия теоремы 2. Тогда для любого $0 < \delta < \dfrac{\alpha}{4(8+\alpha)}$

$$D_n = \sup_{-\infty<x<\infty} \left| \mathsf{P}\left[\frac{U_n}{\sqrt{\mathsf{D}U_n}} < x\right] - \Phi(x) \right| = O\left(\frac{1}{n^\delta}\right), \quad n \to \infty.$$

***Доказательство.*** Обозначим

$$A_n = 4N^3 \frac{nR_n^2}{(\sqrt{\mathsf{D}U_n})^3}; \tag{2}$$

$$B_n = \frac{6}{\pi(r-1)!} \frac{n^2\varphi(R_n)}{\sqrt{\mathsf{D}U_n}}; \tag{3}$$

$$C_n = \frac{2N^2}{\pi} \frac{n^2\varphi(R_n)}{\mathsf{D}U_n}. \tag{4}$$

Тогда неравенство (1) можно переписать в виде

$$\sqrt{\frac{\pi}{2}} D_n^2 \leqslant A_n + B_n + C_n. \tag{5}$$

Заметим, что величина $U_n$ является суммой элементов стационарной последовательности. Применим к ней утверждение [10, теорема 18.2.3, с. 413]: если вещественная стационарная последовательность $X_1, \ldots, X_n$ удовлетворяет условию равномерно сильного перемешивания и $\lim\limits_{n\to\infty} \mathsf{D}S_n = \infty$, то

$$\mathsf{D}S_n = nh(n), \tag{6}$$

где $h(n)$ — медленно меняющаяся функция.

Напомним, что функция $h = h(x)$ вещественного аргумента $x$ называется медленно меняющейся, если она удовлетворяет условию

$$\lim_{x\to\infty} \frac{h(tx)}{h(x)} = 1 \quad \text{для всех } t > 0.$$

Для примера отметим, что медленно меняющимися являются функции $(\ln x)^\alpha$ при $\alpha \in (-\infty, \infty)$.

Тогда, возвращаясь к нашей задаче, согласно (6), имеем

$$\mathsf{D}U_n = nh(n), \tag{7}$$

где $h(n)$ — медленно меняющаяся при $n \to \infty$ функция.

Из (2)–(4) и (7) следует, что

$$A_n = 4N^3 \frac{nR_n^2}{(\sqrt{\mathsf{D}U_n})^3} = \frac{R_n^2}{\sqrt{n}}\, h_1(n); \tag{8}$$

$$B_n = \frac{6}{\pi(r-1)!}\, \frac{n^2\varphi(R_n)}{\sqrt{\mathsf{D}U_n}} = \frac{n^{3/2}}{R_n^{6+\alpha}}\, h_2(n); \tag{9}$$

$$C_n = \frac{2N^2}{\pi}\, \frac{n^2\varphi(R_n)}{\mathsf{D}U_n} = \frac{n}{R_n^{6+\alpha}}\, h_3(n). \tag{10}$$

Здесь $h_1(n), h_2(n), h_3(n)$ — медленно меняющиеся при $n \to \infty$ функции.

Подставив оценки (8)–(10) в (5), получим

$$\sqrt{\frac{\pi}{2}} D_n^2 \leqslant \left( \frac{R_n^2}{\sqrt{n}} + \frac{n^{3/2}}{R_n^{6+\alpha}} \right) \tilde{h}(n), \tag{11}$$

где $\tilde{h}(n)$ — некоторая медленно меняющаяся при $n \to \infty$ функция.

Положив

$$R_n = \left[ n^{2/(8+\alpha)} \right],$$

получим, что слагаемые в правой части (11) равны по порядку с точностью до медленно меняющихся функций и

$$D_n = O\left( \frac{R_n}{n^{1/4}} \right) = O\left( n^{-\alpha/(4(8+\alpha))} \right), \quad n \to \infty. \tag{12}$$

Из (12) следует утверждение леммы 2. ■

**Лемма 3.** Пусть выполнены условия теоремы 2 и для распределения случайной величины $U_n$ при некоторых $c > 0$ и $0 < \delta < \alpha/(4(8+\alpha))$ справедливо неравенство

$$\sup_{-\infty<x<\infty} \left| \mathsf{P}\left[ \frac{U(n)}{\sqrt{\mathsf{D}U(n)}} < x \right] - \Phi(x) \right| \leqslant \frac{c}{n^\delta}. \tag{13}$$

Тогда найдётся такое $c_1 < \infty$, что

$$\sup_{-\infty<x<\infty} |\mathsf{P}[\xi_r^* \leqslant x] - \Phi(x)| \leqslant \frac{c_1}{n^\delta}. \tag{14}$$

***Доказательство.*** Лемма 3 доказывается аналогично лемме 3 в [9]. Пусть

$$V = \xi_r^* - \frac{U_n}{\sqrt{\mathsf{D}U_n}},$$

$\{v_k : k \geqslant 1\}$ — множество значений величины $V$. Тогда для $x \in \mathbb{R}$

$$\mathsf{P}[\xi_r^* < x] = \mathsf{P}\left[ \frac{U_n}{\sqrt{\mathsf{D}U_n}} + V < x \right] = \mathsf{P}\left[ \frac{U_n}{\sqrt{\mathsf{D}U_n}} + \sum_{k\geqslant 1} v_k I\{V = v_k\} < x \right] =$$

$$= \sum_{k\geqslant 1} \mathsf{P}\left[ \frac{U_n}{\sqrt{\mathsf{D}U_n}} < x - v_k \right] \mathsf{P}[V = v_k].$$

Воспользуемся оценкой (13). Получаем

$$\begin{aligned}\mathsf{P}[\xi_r^* < x] &\leqslant \sum_{k\geqslant 1} \mathsf{P}[V = v_k] \left(\Phi(x - v_k) + \frac{c}{n^\delta}\right) = \\ &= \left(\sum_{k\geqslant 1} \mathsf{P}[V = v_k] \left(\Phi(x - v_k) + \frac{c}{n^\delta}\right) - \Phi(x)\right) + \Phi(x).\end{aligned} \tag{15}$$

Так как $\sum\limits_{k\geqslant 1} \mathsf{P}\{V = v_k\} = 1$, то из (15) следует

$$\begin{aligned}\mathsf{P}[\xi_r^* < x] &\leqslant \Phi(x) + \frac{c}{n^\delta} + \sum_{k\geqslant 1} \mathsf{P}[V = v_k] \left(\Phi(x - v_k) - \Phi(x)\right) \leqslant \\ &\leqslant \Phi(x) + \frac{c}{n^\delta} + \frac{1}{\sqrt{2\pi}} \sum_{k\geqslant 1} |v_k| \, \mathsf{P}[V = v_k] = \Phi(x) + \frac{c}{n^\delta} + \frac{1}{\sqrt{2\pi}} \mathsf{E}|V|.\end{aligned} \tag{16}$$

Согласно [9, оценка (15)], выполнено соотношение $\mathsf{E}|V| = O\left(h(n)n^{-1/2}\right)$, где $h(n)$ — некоторая медленно меняющаяся функция. Поэтому из (16) вытекает, что найдётся такое $c' < \infty$, при котором

$$\mathsf{P}[\xi_r^* < x] \leqslant \Phi(x) + \frac{c'}{n^\delta}. \tag{17}$$

Аналогично доказывается, что

$$\mathsf{P}[\xi_r^* < x] \geqslant \Phi(x) - \frac{c''}{n^\delta}. \tag{18}$$

Из (17) и (18) следует (14) с $c_1 = \max\{c', c''\}$. ■

Этим доказательство теоремы 2 завершено.

## Заключение

Получена оценка скорости сходимости распределения числа $r$-кратных повторений знаков в стационарной случайной последовательности, удовлетворяющей условию равномерно сильного перемешивания с коэффициентом $\varphi(t) \leqslant t^{-6-\alpha}$ при некотором $\alpha > 0$, к нормальному распределению. Установлено, что в этом случае расстояние в равномерной метрике между функцией распределения центрированного и нормированного специальным образом числа повторений и сопровождающим нормальным распределением с увеличением длины последовательности $n$ убывает со скоростью $O(n^{-\delta})$ для любого $\delta \in (0, \alpha(32 + 4\alpha)^{-1}))$.

## ЛИТЕРАТУРА

1. *Иванов М. А., Чугунков И. В.* Теория, применение и оценка качества генераторов псевдослучайных последовательностей. М.: КУДИЦ-ОБРАЗ, 2003. 240 с.
2. *Rukhin A., Soto J., Nechvatal J., et al.* A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications. NIST, Apr. 2010. `https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-22r1a.pdf`.
3. *Михайлов В. Г.* Предельная теорема пуассоновского типа для числа пар почти полностью совпавших цепочек // Теория вероятностей и ее применения. 2008. Т. 53. Вып. 1. С. 59–71.
4. *Михайлов В. Г., Шойтов А. М.* О числах множеств эквивалентных цепочек в последовательности независимых случайных величин // Математические вопросы криптографии. 2013. Т. 4. Вып. 1. С. 77–86.
5. *Шойтов А. М.* Нормальное приближение в задаче об эквивалентных цепочках // Труды по дискретной математике. 2007. Т. 10. С. 326–349.

6. *Михайлов В. Г.* Оценки точности пуассоновской аппроксимации для распределения числа серии повторений длинных цепочек в цепи Маркова // Дискретная математика. 2015. Т. 27. Вып. 4. С. 67–78.
7. *Михайлов В. Г., Шойтов А. М.* О длинных повторениях цепочек в цепи Маркова // Дискретная математика. 2014. Т. 26. Вып. 3. С. 79–89.
8. *Михайлов В. Г., Шойтов А. М.* Многократные повторения длинных цепочек в цепи Маркова // Математические вопросы криптографии. 2015. Т. 6. Вып. 3. С. 117–134.
9. *Михайлов В. Г., Меженная Н. М., Волгин А. В.* Об условиях асимптотической нормальности числа повторений в стационарной случайной последовательности // Дискретная математика. 2021. Т. 33. Вып. 3. С. 64–78.
10. *Ибрагимов И. А., Линник Ю. В.* Независимые и стационарно связанные случайные величины М.: Наука, 1965. 816 с.

REFERENCES

1. *Ivanov M. A. and Chugunkov I. V.* Teoriya, primenenie i otsenka kachestva generatorov psevdosluchaynykh posledovatel'nostey [Theory, Application and Evaluation of the Quality of Pseudo-Random Sequence Generators]. Moscow, KUDITs-OBRAZ, 2003. 240 p. (in Russian)
2. *Rukhin A., Soto J., Nechvatal J., et al.* A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications. NIST, Apr. 2010. `https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-22r1a.pdf`.
3. *Mikhailov V. G.* A Poisson-type limit theorem for the number of pairs of matching sequences. Theory of Probability and its Appl., 2009, vol. 53, no. 1, pp. 106–116.
4. *Mikhaylov V. G. and Shoytov A. M.* O chislakh mnozhestv ekvivalentnykh tsepochek v posledovatel'nosti nezavisimykh sluchaynykh velichin [On the numbers of equivalent tuples sets in a sequence of independent random variables]. Matematicheskie Voprosy Kriptografii, 2013, vol. 4, iss. 1, pp. 77–86. (in Russian)
5. *Shoytov A. M.* Normal'noe priblizhenie v zadache ob ekvivalentnykh tsepochkakh [Normal approximation in a problem on equivalent tuples]. Trudy po Diskretnoy Matematike, 2007, vol. 10, pp. 326–349. (in Russian)
6. *Mikhailov V. G.* Estimates of accuracy of the Poisson approximation for the distribution of number of runs of long string repetitions in a Markov chain. Discr. Math. Appl., 2016, vol. 26, no. 2, pp. 105–113.
7. *Mikhailov V. G. and Shoitov A. M.* On repetitions of long tuples in a Markov chain. Discr. Math. Appl., 2015, vol. 25, no. 5, pp. 295–303.
8. *Mikhaylov V. G. and Shoytov A. M.* Mnogokratnye povtoreniya dlinnykh tsepochek v tsepi Markova [On multiple repetitions of long tuples in a Markov chain]. Matematicheskie Voprosy Kriptografii, 2015, vol. 6, iss. 3, pp. 117–134. (in Russian)
9. *Mikhaylov V. G., Mezhennaya N. M., and Volgin A. V.* Ob usloviyakh asimptoticheskoy normal'nosti chisla povtoreniy v statsionarnoy sluchaynoy posledovatel'nosti [Conditions for asymptotic normality of the number of repetitions in a discrete stationary random sequence]. Diskretnaya Matematika, 2021, vol. 33, iss. 3, pp. 64–78.
10. *Ibragimov I. A. and Linnik. Yu. V.* Independent and Stationary Sequences of Random Variables. Wolters-Noordhoff, Groningen, Netherlands, 1971.