OP&PM SURVEYS ON APPLIED AND INDUSTRIAL MATHEMATICS
Volume 27 Issue 2
2020

IV INTERNATIONAL BALTIC SYMPOSIUM ON APPLIED AND INDUSTRIAL MATHEMATICS

N. M. Mezhennaya, V. G. Mikhailov (Moscow, Bauman Moscow State Technical University (BMSTU); Moscow, Steklov Mathematical Institute of Russian Academy of Sciences). **On the distribution of multiple repetitions in stationary sequence satisfying uniform mixing condition.**

УДК 519.214

Abstract: The paper presents limit theorems for the number of repetitions of letters in a segment of a stationary random sequence of length n satisfying the uniform mixing condition for $n \to \infty$. The fulfillment of the condition that the stationary distribution is equiprobable essentially changes the form of the limit law, namely, in the unequiprobable case, the asymptotic distribution for the number of multiple repetitions is normal, and in the equiprobable case it converges to a quadratic form from the normal random vector.

Keywords: multiple repetitions, limit theorem, uniform mixing, discrete random sequence.

Let $\mathbb{Z}_+ = \{0, 1, 2, \ldots, n, \ldots\}$, $\{X_t, t \in \mathbb{Z}_+\}$ be the random sequence, and F_a^b be the σ-algebra, generated by the random variables $\{X_a, \ldots, X_b\}$ for $a, b \in \mathbb{Z}_+$ and $a < b$.

A random sequence $\{X_t, t \in \mathbb{Z}_+\}$ is called *strictly stationary* [1, p. 284] if for any natural number n the distribution of the random vector $(X_{t_1+h}, \ldots, X_{t_n+h})$ does not depend on h where $t_1 + h, \ldots, t_n + h \in \mathbb{Z}_+$.

The stationary sequence $\{X_t, t \in \mathbb{Z}_+\}$ satisfy the *uniform mixing condition* [1, p. 309] if

$$\phi(\tau) = \sup_{\substack{A \in F_0^t, \\ B \in F_{t+\tau}^{+\infty}}} \frac{|\,\mathsf{P}(AB) - \mathsf{P}(A)\,\mathsf{P}(B)|}{\mathsf{P}(A)} \to 0 \qquad (\tau \to \infty).$$

The non-increasing function $\phi(\tau)$ is called the *uniform mixing coefficient.*

Let

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

be the number of repetitions of multiplicity r of letters $X_1, \ldots, X_n$.

Let us formulate the following

Condition A. Let $\{X_1, \ldots, X_n\}$ be the strictly stationary random sequence with values in the set $\{1, \ldots, N\}$ and the stationary distribution $\mathsf{P}\{X_j = k\} = p_k$, $k = 1, \ldots, N$, satisfying uniform mixing condition with the uniform mixing coefficient ϕ.

The properties of the number ξ_r of multiple repetitions are closely related to the problem of random allocations [2], the properties of ξ_r in sequences of a special kind are well studied (for example, in [3, 4, 5] for independent random variables, in [6, 7] for Markov chains) and are used for the analysis of the properties of random sequences [8, 9].

It is known that for a sequence of independent random variables, the condition

$$p_1 = \cdots = p_N = 1/N \tag{1}$$

(i.e., the observed random sequence has an equiprobable stationary distribution) fundamentally affects the limit distribution of the number ξ_r of multiple repetitions of letters. A similar result in the present paper will be obtained for sequences of a more general form, namely, sequences satisfying the uniform mixing condition.

Consider the sequence

$$\left\{ u_j = \frac{1}{(r-1)!} \sum_{k=1}^{N} p_k^{r-1} \left(I\{X_j = k\} - p_k \right), \qquad j = 1, \ldots, n \right\} \tag{2}$$

and its sum

$$U_n = \sum_{j=1}^{n} u_j .$$

It is clear that the sequence (2) satisfies Condition A and $\mathsf{P}\{u_j = 0\} = 1$ if and only if the condition (1) holds. Therefore, the properties of the sequence (2) and U_n will be used to formulate and prove limit theorems.

In present paper the following notation is used: $\mathcal{N}(\mathbf{m}; \Sigma)$ for the multivariate normal distribution with mean vector $\mathbf{m}$ and covariance matrix Σ, $\mathcal{N}(0;1)$ for the standard normal distribution, $\mathcal{L}(\xi)$ for the distribution law of the random variable ξ, $\xrightarrow{d}$ for the convergence in distribution, $C_n^r = \frac{n!}{r!(n-r)!}$ for the Binomial coefficient, and $tr\Sigma$ for the trace of matrix Σ.

Theorem 1. *Let Condition A be fulfilled and $n \to \infty$. If the condition* (1) *is violated and* $\mathrm{D}U_n \to \infty$, *then*

$$\xi_r^* = \frac{\xi_r}{n^{r-1}\sqrt{\mathrm{D}U_n}} - \frac{n}{r!\sqrt{\mathrm{D}U_n}} \sum_{k=1}^{N} p_k^r \xrightarrow{d} \mathcal{N}(0;1).$$

In the proof of the limit theorem in the equiprobable case, we use the additional assumption that the numbers $\varsigma_1, \ldots, \varsigma_N$ of letters in $X_1, \ldots, X_n$ for $n \to \infty$ satisfy the multivariate central limit theorem.

Theorem 2. *Let conditions A and* (1) *be fulfilled and $n \to \infty$. If there exists the number sequence $\{\sigma_n\}_{n=1}^{\infty}$: $\sigma_n > 0$ $\forall n \geqslant 1$, $\lim\limits_{n\to\infty} \sigma_n = \infty$, and*

$$\frac{1}{\sigma_n}\left(\varsigma_1 - \frac{n}{N}, \ldots, \varsigma_N - \frac{n}{N}\right) \xrightarrow{d} (\eta_1, \ldots, \eta_N),$$

where $\mathcal{L}(\eta_1, \ldots, \eta_N) = \mathcal{N}(\mathbf{0}; \Sigma)$ *with* $tr\Sigma > 0$, *then*

$$\xi_r^{**} = \frac{N^{r-2} r! \xi_r}{\sigma_n^2 n^{r-2} C_r^2} - \frac{n^2 N^{r-2}}{\sigma_n^2 C_r^2} \sum_{k=1}^{N} \left(\prod_{j=0}^{r-1} \left(\frac{1}{N} - \frac{j}{n} \right) \right) \xrightarrow{d} \sum_{k=1}^{N} \eta_k^2 .$$

The statements of Theorems 1 and 2 can be easily extended by to the sequences of events considered in [3]–[7]. The results presented in Theorems 1 and 2 can also be used to construct a statistical test for verifying the condition (1).

REFERENCES

1. *Ibragimov I. A.*, *Linnik Yu. V.* Independent and stationary sequences of random variables. Wolters-Noordhoff, Groningen, Netherlands, 1971.
2. *Kolchin V. F.*, *Sevast'yanov B. A.*, *Chistyakov V. P.* Random allocations. Scripta Series in Mathematics. Winston & Sons, Washington, DC, 1978.
3. *Zubkov A. M.*, *Mikhailov V. G.* Limit distributions of random variables connected with long duplications in a sequence of independent trials. — Theory Probab. Appl., 1974, v. 19, № 1, p. 172–179.

4. *Mikhailov V. G.* Limit distributions of random variables connected with multiple long duplications in a sequence of independent trials. — Theory Probab. Appl., 1974, v. 19, № 1, p. 180–184.
5. *Zubkov A. M.*, *Mikhailov V. G.* On the repetitions of s-tuples in a sequence of independent trials. — Theory Probab. Appl., 1979, v. 24, № 2, p. 269–282.
6. *Mikhailov V. G.* , *Shoitov A. M.* On multiple repetitions of long tuples in a Markov chain. — Mat. Vopr. Kriptogr., 2015, v. 6, № 3, p. 117–133.
7. *Mikhailov V. G.* On the estimation of the accuracy of Poisson approximation for distribution of number of series of repetitions of long tuples in a Markov chain. — Discrete Math. Appl., 2015, v. 27, № 4, p. 67–78.
8. *Tikhomirova M. I.*, *Chistyakov V. P.* The statistical test of the empty cells type. — Mat. Vopr. Kriptogr., 2010, v. 1, № 1, p. 101–108.
9. *Chistyakov V. P.* On the asymptotic normality of the number of empty cells in a scheme of group allocation of particles. — Discrete Math. Appl., 2014, v. 24, № 6, p. 319–325.

УДК 519.214

Mezhennaya N. M., ***Mikhailov V. G.*** (Moscow, Bauman Moscow State Technical University (BMSTU); Moscow, Steklov Mathematical Institute of Russian Academy of Sciences)**. On the distribution of multiple repetitions in stationary sequence satisfying uniform mixing condition.**

Abstract: The paper presents limit theorems for the number of repetitions of letters in a segment of a stationary random sequence of length n satisfying the uniform mixing condition for $n \to \infty$. The fulfillment of the condition that the stationary distribution is equiprobable essentially changes the form of the limit law, namely, in the unequiprobable case, the asymptotic distribution for the number of multiple repetitions is normal, and in the equiprobable case it converges to a quadratic form from the normal random vector.par

Keywords: multiple repetitions, limit theorem, uniform mixing, discrete random sequence.