THEORY OF PROBABILITY AND ITS APPLICATIONS

Volume I Number 1

1956

CENTRAL LIMIT THEOREM FOR NONSTATIONARY MARKOV CHAINS. I

R. L. DOBRUSHIN

1. Introduction

1.1. In 1910 A. A. Markov began to study the conditions under which the central limit theorem is applicable to nonstationary Markov chains. He showed [26] that if for all n the variables $X_1^{(n)}, \cdots, X_n^{(n)}$ form a sequence of random variables which can take on the values 0 and 1 and are connected by a nonstationary chain (the general case of a chain of variables with $N = 2$ states is usually reduced to this particular case), and if $S^{(n)} = X_1^{(n)} + \cdots + X_n^{(n)}$, then a sufficient condition that the sums

$$\bar{S}^{(n)} = \frac{S^{(n)} - \mathbf{M}S^{(n)}}{\sqrt{\mathbf{D}S^{(n)}}} \tag{1.1}$$

be normal asymptotically is that for all n and k there is some $\alpha > 0$ such that

$$\begin{aligned} \alpha^{(n)} &\leqq \mathbf{P}\{X_k^{(n)} = 1 \mid X_{k-1}^{(n)} = 1\} \leqq 1 - \alpha^{(n)}, \\ \alpha^{(n)} &\leqq \mathbf{P}\{X_k^{(n)} = 0 \mid X_{k-1}^{(n)} = 0\} \leqq 1 - \alpha^{(n)}, \end{aligned} \tag{1.2}$$

where $\alpha^{(n)} \geqq \alpha$. Markov made use of this fact in developing the method of moments. He mentioned also that the difficult problem of reducing the number of restrictions he introduced must remain open for other investigators.

The next step in the study of this problem was made by S. N. Bernstein. In his work published in 1922, Bernstein [1] treated Markov's problem as a special case of a more general problem involving the conditions for the sums of weakly dependent variables to be normal asymptotically. He showed, in particular, that the normalized sums (1.1) are normal asymptotically if the variables $\alpha^{(n)}$ in (1.2) satisfy the condition that $\alpha^{(n)} n^{1/7-\varepsilon} \to \infty$ for some $\varepsilon > 0$. In his now classical work of 1926, Bernstein [2] strengthened his result and showed that this property is true also if $\alpha^{(n)} n^{1/5-\varepsilon} \to \infty$, where $\varepsilon > 0$. In addition, in the same work he gave an example which showed that if $\alpha^{(n)} \sim cn^{-1/3}$, where $c (0 < c < \infty)$ is a constant, then the sequence $\bar{S}^{(n)}$ can have a limit distribution which is not normal. (This example is dealt with in detail in Section 2 of the present article.) A further refinement of the conditions of validity of the central limit theorem required a deeper study of the Markov chain itself. This Bernstein did [3, 4] in 1926—1928. In these works he proved the sufficiency of the condition $\alpha^{(n)} n^{1/3-\varepsilon} \to \infty$. Finally, as late as in 1947, N. A. Sapogov [29] developed Bernstein's method to the point where he was able to obtain the best possible result. He showed that normality is obtained in the limit if $\alpha^{(n)} n^{1/3} \to \infty$.

The next problem was to study Markov chains with the number of states N greater than 2. This problem may be stated as follows. Assume that for fixed n the variables $X_1^{(n)}, \cdots, X_n^{(n)}$ are connected by a (nonstationary) Markov chain. Further, let each of the variables $X_i^{(n)}$ take on only a finite set N of numerical values $(a_{i1}^{(n)}, \cdots, a_{iN}^{(n)})$. (It is assumed that all $|a_{ij}^{(n)}| < K < \infty$.)

Finally, let

$$\alpha^{(n)} = \min \mathbf{P}\{X_i^{(n)} = a_{ik}^{(n)} | X_{i-1}^{(n)} = a_{i-1,j}^{(n)}\}, \tag{1.3}$$

where the minimum is taken over all $i = 2, \cdots, n$; $j = 1, \cdots, N$; $k = 1, \cdots, N$. Again, let us write $S^{(n)} = X_1^{(n)} + \cdots + X_n^{(n)}$ and define $S^{(n)}$ as in (1.1). We wish to establish the conditions which must be satisfied by the $\alpha^{(n)}$ in order that $\bar{S}^{(n)}$ have asymptotically a normal distribution in the limit as $n \to \infty$. The methods developed by Bernstein in the literature cited above depend in an essential way on the fact that only two states exist. One of the most difficult problems in the proof of the central limit theorem for nonstationary Markov chains is to obtain a lower limit for the variace, or to prove that $\mathbf{D}S^{(n)} \to \infty$ sufficiently rapidly. (The point is, that for dependent variables in a chain the variance of the sum is not equal to the sum of the variances, so that $\mathbf{D}S^{(n)}$ may increase much slower than n.) The most accurate evaluation of the lower bound of the variances was obtained by Bernstein [5] in 1936. To do this, he developed a new method for studying Markov chains, known as the "cross section method". The idea of this method is based on the fact that if $1 = i_1 < i_2 < \cdots < i_k = n$, then by fixing the values of $X_{i_1}^{(n)} = x_1, X_{i_2}^{(n)} = x_2, \cdots, X_{i_k}^{(n)} = x_k$ the sums $\tilde{S}_r = X_{i_r+1}^{(n)} + \cdots + X_{i_{r+1}}^{(n)}$ become independent random variables. Basing his work on Bernstein's evaluation of the variance and limit theorems for sums of weakly dependent variables, N. A. Sapagov [30] showed in 1947 that $\alpha^{(n)} n^{1/5-\varepsilon} \to \infty$ is a sufficient condition for $\bar{S}^{(n)}$ to be normal in the limit. Yu. V. Linnik [23, 24] developed a profound generalization of Bernstein's method in 1948, and applied it to the study of higher moments of the $S^{(n)}$. As a result he showed the normality of $\bar{S}^{(n)}$ in the limit if $\alpha^{(n)} n^{1/3-\varepsilon} \to \infty$.

Further, Sapogov [30] considered the case in which the $X_i^{(n)}$ take on values on a segment of the real axis. In this case $\alpha^{(n)}$ denotes the lower bound of the transition probability densities which will give this particular chain. Using Bernstein's evaluation of the variance, Sapogov proved the central limit theorem for the condition $\alpha^{(n)} n^{1/5-\varepsilon} \to \infty$. For convenience, we shall tabulate the above mentioned results.

Condition on the transition probability	number of states		
	2	$2 < N < \infty$	segment of the real axis
$\alpha^{(n)} > c > 0$	Markov 1910	Bernstein 1936 Sapogov 1947	
$\alpha^{(n)} n^{1/7-\varepsilon} \to \infty$	Bernstein 1922		
$\alpha^{(n)} n^{1/5-\varepsilon} \to \infty$	Bernstein 1926		
$\alpha^{(n)} n^{1/3-\varepsilon} \to \infty$	Bernstein 1928	Linnik 1948	—
$\alpha^{(n)} n^{1/3} \to \infty$	Sapogov 1947	—	—

1.2. We have mentioned all the works containing general results on the question of interest [1]. We now indicate some problems which still remain unanswered after these far-reaching investigations.

1) The study of the applicability conditions for the central limit theorem have not been carried to an end with respect to restrictions placed on the lower bound of the transition probability (dashes in the table).

2) It is clear that the validity of the central limit theorem is explained by the fact that distant terms are "almost independent", rather than by the specific nature of the state space of the system (a finite set, a segment, etc.) It would therefore seem desirable to establish conditions for asymptotic normality which would be valid for any state space.

3) For the same reason, the condition that none of the transition probabilities vanish would seem to be too stringent. We note one interesting result in this connection. Sapogov [30] has treated the case in which there exists a finite number of states and each transition matrix has a column all of whose elements are greater than $\beta^{(n)}$. He proved the central limit theorem with the condition $\beta^{(n)}n^{1/7-\varepsilon} \to \infty$. This result is quite special, however. It is clear that asymptotic normality is related to certain conditions which are the same for any state space whose chains are "ergodic" and this does not prevent many of the transition probabilities from vanishing. It would be interesting to state such conditions.

4) As yet there exist no significant results concerning unbounded $X_i^{(n)}$. It would be interesting, in particular, to establish whether or not in this case the number 1/3 plays the same extremal role as in the case of bounded terms.

The present work is devoted to eliminating the above gaps in the theory. Certain of the results have already been published [13].

1.3. We first give an exact and general definition of a sequence of random variables connected by a Markov chain [2].

Consider two arbitrary sets Ψ and $\bar{\Psi}$, and let there be defined on Ψ the σ-algebra [3] $\mathfrak{R}$ of its subsets, and on $\bar{\Psi}$ a σ-algebra $\bar{\mathfrak{R}}$. Consider the function $P(\psi, \bar{R})$, with $\psi \in \Psi$, $\bar{R} \in \bar{\mathfrak{R}}$, which has the following two properties.

a) For fixed $\psi \in \Psi$, the function $P(\psi, \bar{R})$ treated as a function of $\bar{R}$ is a probability measure on $\bar{\mathfrak{R}}$.

b) For fixed $\bar{R} \in \mathfrak{R}$, the function $P(\psi, \bar{R})$ treated as a function of ψ is measurable with respect to $\mathfrak{R}$.

Such a function $P(\psi, \bar{R})$ is called a transition probability function whose domain of definition is $(\Psi, \mathfrak{R}, \bar{\Psi}, \bar{\mathfrak{R}})$.

Consider arbitrarily given sets Ω_i, $i = 1, \cdots, n$. The set Ω_i is called the state space of the system at the i-th instant of time. In each Ω_i let there be defined a σ-algebra $\mathfrak{A}_i$ of its subsets, which are called measurable. Further, let there be given a sequence of transition probability functions $P_i(\omega, A)$,

[1] The theorem contained in Section 79 of V. I. Romanovskii's book [27] is in error. It is disproved, for instance, by the results of Shirokorad [35]. Another counterexample will be given in the next part of the present work. See also footnote 6.

[2] A reader interested only in Markov chains with finite or denumerable numbers of states may proceed directly to paragraph 1.5, using eq. (1.19) for the definition of the ergodic coefficient.

[3] In other words, a Borel family. The measure theory terminology we shall use is that of Halmos [34].

$i = 1, \cdots, n$, such that the domain of definition of $P_i(\omega, A)$ is $(\Omega_i, \mathfrak{A}_i, \Omega_{i+1}, \mathfrak{A}_{i+1})$. Then $P_i(\omega, A)$ is called the i-th transition function of the Markov process. Finally, let there be given a probability measure $P(A)$ for $A \in \mathfrak{A}_1$ called the initial probability distribution.

As the fundamental space of elementary events of our process, we shall take the direct product $(\Omega_1 \times \Omega_2 \times \cdots \times \Omega_n, \mathfrak{A}_1 \times \mathfrak{A}_2 \times \cdots \times \mathfrak{A}_n)$. This is the set Ω of sequences $(\omega_1, \omega_2, \cdots, \omega_n)$ on which is defined the σ-algebra of its subsets, this algebra being generated by sets of the form $\{\omega_1 \in A_1, \omega_2 \in A_2, \cdots, \omega_n \in A_n\}$, where $A_i \in \mathfrak{A}_i$. The probabilities of sets $A \in \mathfrak{A}$ are defined as

$$P(A) = \int_A P(d\omega_1) \int_A P_1(\omega_1, d\omega_2) \cdots \int_A P_{n-1}(\omega_{n-1}, d\omega_n). \tag{1.4}$$

This space Ω together with its σ-algebra of subsets and its probability measure $P(A)$ is called a *Markov chain* with n instants of time, with a state space $(\Omega_i, \mathfrak{A}_i)$ with transition functions $P_i(\omega, A)$ and initial probability distribution $P(A)$.

We shall call the σ-algebra of the states of the system at the k-th instant of time the subalgebra $\overline{\mathfrak{A}}_k$ of $\mathfrak{A}$ which consists of sets of the form

$$\tilde{A}_k = \{\omega_1 \in \Omega_1, \cdots, \omega_{k-1} \in \Omega_{k-1}, \omega_k \in A_k, \omega_{k+1} \in \Omega_{k+1}, \cdots, \omega_n \in \Omega_n\},$$

where $A_k \in \mathfrak{A}_k$. We shall then say that the set $\tilde{A}_k$ corresponds to or is associated with A_k. A real function $X(\omega)$ is called a function depending on the state of the system at the k-th instant of time if it is measurable with respect to the σ-algebra $\tilde{\mathfrak{A}}_k$. A set of functions $X_1(\omega), X_2(\omega), \cdots, X_n(\omega)$, where $X_k(\omega)$ depends on the state of the system at the k-th instant of time, is called a sequence of random variables connected by a Markov chain. This abstract definition is dealt with in more detail in Doob's book [14].

We shall usually assume that we are given a sequence of Markov chains where the n-th chain is defined for n time units. Therefore everything defined in the present paragraph will depend also on another index n, which we shall write as a superscript in parenthesis (for instance $\Omega_i^{(n)}$). We shall assume further that for each n there is given a sequence of random variables $X_1^{(n)}, \cdots, X_n^{(n)}$ connected by an n-element Markov chain. We shall write $S^{(n)} = X_1^{(n)} + \cdots + X_n^{(n)}$. In all cases we shall assume that all the $X_i^{(n)}$ have finite mathematical expectations $\mathbf{M}X_i^{(n)}$ and variances $\mathbf{D}X_i^{(n)}$. Then the mathematical expectations $\mathbf{M}S^{(n)}$ and variances $\mathbf{D}S^{(n)}$ are also finite.

1.4. Let us now discuss a quantity which is characteristic of the transition function $P_i^{(n)}$ and which generalizes the quantity $\alpha^{(n)}$ used in previous papers. Consider some transition probability function $P(\psi, \bar{R})$ whose domain of definition is $(\Psi, \mathfrak{R}, \overline{\Psi}, \overline{\mathfrak{R}})$. We write

$$\alpha(P) = 1 - \sup |P(\psi, \bar{R}) - P(\varphi, \bar{R})|, \tag{1.5}$$

where the upper bound is taken over all $\psi \in \Psi$, $\varphi \in \Psi$, $\bar{R} \in \overline{\mathfrak{R}}$. We shall call the quantity $\alpha(P)$ the ergodic coefficient of the transition function $P(\psi, \bar{R})$. In proving the central limit theorem for a stationary Markov chain, E. B. Dynkin [15] made use of the condition that the ergodic coefficient be less than unity.

Before discussing this definition from various points of view, let us formulate our fundamental theorem. To do this, we write $\alpha_i^{(n)} = \alpha(P_i^n)$. We then have the following

Theorem 1. *Let all the $X_i^{(n)}$ be uniformly bounded $|X_i^{(n)}| \leqq C < \infty$ and assume that for all i and n*

$$\mathbf{D}X_i^{(n)} \geqq c > 0. \tag{1.6}$$

Further, let

$$\alpha^{(n)} = \min_{i=1,\cdots,n-1} \alpha_i^{(n)} \tag{1.7}$$

and as $n \to \infty$, assume that

$$\alpha^{(n)} n^{1/3} \to \infty, \tag{1.8}$$

so that uniformly in t as $n \to \infty$ we have

$$\mathbf{P}\left(\frac{S^{(n)} - \mathbf{M}S^{(n)}}{\sqrt{\mathbf{D}S^{(n)}}} < x\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt, \tag{1.9}$$

which is equivalent to stating the central limit theorem for $S^{(n)}$.

Consider some set W on which the σ-algebra $\mathfrak{A}$ of its subsets is defined. Further, consider the linear space $M_W = \{\mu\}$ of all denumerably additive finite functions given on the σ-algebra $\mathfrak{A}$. In this space we introduce the norm

$$||\mu|| = \tfrac{1}{2}(\sup_{A \in \mathfrak{A}} \mu(A) - \inf_{A \in \mathfrak{A}} \mu(A)). \tag{1.10}$$

This norm $||\mu||$ is equal to half the total variation of the generalized measure μ [4]. (See the definition by Halmos [34], p. 123.)

For any probability measure μ, the norm $||\mu|| = \frac{1}{2}$. Let us consider further the sub-space $L_W \subset M_W$ of functions $\lambda \in M_W$ such that $\lambda(W) = 0$. If $\lambda \in L_W$ for all $A \in \mathfrak{A}$, then $\lambda(A) = -\lambda(W - A)$. Therefore $\sup_{A \in \mathfrak{A}} \lambda(A) = -\inf_{A \in \mathfrak{A}} \lambda(A)$ and

$$||\lambda|| = \sup_{A \in \mathfrak{A}} |\lambda(A)|. \tag{1.10'}$$

It is seen from this that if, for fixed ψ, $P(\psi, \bar{R})$ is considered an element of $M_{\bar{\Psi}}$, we have

$$\alpha(P) = 1 - \sup_{(\psi,\varphi)} ||P(\psi, \bar{R}) - P(\varphi, \bar{R})||, \tag{1.5'}$$

where $\psi \in \Psi$, $\varphi \in \Psi$.

We define the operator P which corresponds to the function $P(\psi, \bar{R})$ and maps the space M_Ψ onto $M_{\bar{\Psi}}$ by the equation $P_\mu = \tilde{\mu}$, where $\tilde{\mu} \in M_{\bar{\Psi}}$ and

$$\tilde{\mu}(\bar{R}) = \int_\Psi P(d\psi, \bar{R})\mu(d\psi), \ \bar{R} \in \bar{\mathfrak{R}}. \tag{1.11}$$

This operator P is clearly linear and continuous. It maps probability measures into probability measures and its norm is always 1. A general Markov transition function was first treated as an operator in distribution space by Kolmogorov

[4] It is not difficult to show that M_W is a Banach space (see, for instance, Birkhoff]7], p. 351). We shall not, however, make use of this fact.

[18]. Yosida and Kakutani [16] have begun a systematic study of transition functions treated as operators in the Banach space constructed above.

This operator P maps the sub-space L_Ψ into the sub-space $L_{\bar{\Psi}}$. Let $N(P)$ be the norm of P considered as an operator acting only on L_Ψ or

$$N(P) = \sup_{\lambda \in L_\Psi} \frac{||P\lambda||}{||\lambda||}. \tag{1.12}$$

The sub-space L_Ψ coincides with the sub-space generated by the differences $\mu_1-\mu_2$, where μ_1 and μ_2 are probability measures. Therefore

$$N(P) = \sup_{(\mu_1,\mu_2)} = \frac{||P\mu_1-P\mu_2||}{||\mu_1-\mu_2||}, \tag{1.12'}$$

where $\mu_1 \in M_\Psi$, $\mu_2 \in M_\Psi$. The quantity $N(P)$ is a measure of contraction of the distance of probability measures induced by the operator P. If we write δ_x, $x \in \Psi$, for a probability measure concentrated at the point x (that is to say, $\delta_x(R) = 1$ for $x \in R$ and 0 for $x \notin R$, $R \in \mathfrak{R}$), we have $\delta_x \in M_\Psi$. It is clear that $P\delta_x = P(x, R)$. If there is some $R \in \mathfrak{R}$ such that the point $x \in R$ but the point $y \notin R$, then $||\delta_x-\delta_y|| = 1$; if, on the other hand, for all $R \in \mathfrak{R}$ it follows from $x \in R$ that $y \in R$, then the measurability of the function $P(\psi, \bar{R})$ implies that $P(x, \bar{R}) \equiv P(y, \bar{R})$. We then immediately arrive at the fact that $\alpha(P) \geqq 1-N(P)$. In the next part of this paper we shall show without any difficulty that

$$\alpha(P) = 1-N(P). \tag{1.5''}$$

In all cases we have $0 \leqq \alpha(P) \leqq 1$. It is easily seen that $\alpha(P) = 1$ if and only if $P(\psi, \bar{R}) \equiv P(\bar{R})$, which means that this function is independent of $\psi \in \Psi$. Further $\alpha(P) = 0$ if and only if for any x there exist $x \in \Psi$ and $y \in \Psi$ such that the measures [5] $P(x, \cdot)$ and $P(y, \cdot)$ are concentrated on sets A_x and A_y, respectively, for which $P(x, A_xA_y) < \varepsilon$ and $P(y, A_xA_y) < \varepsilon$.

We give finally one more interpretation of the definition of the ergodic coefficient, which is somewhat more useful than the previous one. Consider a set W with the σ-algebra $\mathfrak{A}$ of its subsets, and let there be given measures $\mu_1 \in M_W$, $\mu_2 \in M_W$. We shall say that the measure $\nu \in M_W$ minorizes the measure μ if for all $A \in \mathfrak{A}$ we have

$$\mu(A) \geqq \nu(A). \tag{1.13}$$

It is not difficult to show (see Section 3.2 of the next part of the work) that if we write

$$\tilde{\alpha}(\mu_1, \mu_2) = \sup \nu(W), \tag{1.14}$$

where the upper bound is taken over all measures ν which simultaneously minorize μ_1 and μ_2, then

$$||\mu_1-\mu_2|| = 1-\tilde{\alpha}(\mu_1, \mu_2). \tag{1.15}$$

As has been mentioned by Kolmogorov (the proof will be presented in the

[5] The notation $P(x, \cdot)$ denotes $P(x, A)$ treated as a function of A for fixed x.

second part of our work), $\tilde{\alpha}(\mu_1, \mu_2)$ can be defined also as

$$\tilde{\alpha}(\mu_1, \mu_2) = \inf_{\{S_i\}} \sum_{i=1}^{m} \min(\mu_1(S_1), \mu_2(S_i)), \tag{1.16}$$

where the lower bound is taken over all possible resolutions of W into pairs of non-intersecting subsets S_i, $i = 1, \cdots, m$. Applying (1.15) to the transition probabilities $P(\psi, \bar{R})$ for fixed ψ and taking account of (1.5′), we obtain

$$\alpha(P) = \inf_{(\psi,\varphi)} \tilde{\alpha}(P(\psi, \cdot), P(\varphi, \cdot)). \tag{1.5‴}$$

In an earlier paper [13] $\beta(P)$ instead of $\alpha(P)$ was used to characterize the transition probability function. This quantity is defined as

$$\beta(P) = \sup \nu(\Psi), \tag{1.17}$$

where the upper bound is taken over the set of all measures $\nu(\bar{R})$, $\bar{R} \in \bar{\mathfrak{R}}$, which minorize simultaneously all measures $P(\psi, \bar{R})$, $\psi \in \Psi$. On the basis of this definition, results were formulated which are in all other ways analogous to the present results. From (1.14) and (1.5‴) it follows that

$$\beta(P) \leqq \alpha(P), \tag{1.18}$$

in all cases, so that the results of the present work include those of the previous one. Speaking quite loosely, the conditions which give the lower bound of $\alpha(P)$ assure the existence of a common part for every pair of measures $P(\psi, \bar{R})$, $P(\varphi, \bar{R})$, whereas those conditions which give the lower bound for $\beta(P)$ assure the existence of a common part for all measures $P(\psi, \bar{R})$, $\psi \in \Psi$. This remark becomes clear when we deal with a denumerable chain (see paragraph 1.5).

The quantity $\alpha(P)$ is a measure of the ergodic properties of the transition function P. That it is a natural quantity to use in this connection is affirmed to some extent by our Theorem 11.

1.5. Let us consider Markov chains with denumerable sets of states $E_1, E_2, \cdots$. A nonstationary Markov chain with n instants of time is then given by an initial probability distribution $\{p_i\}$ and matrices P_k whose elements $\{{}_{ij}P_k\}$, $i = 1, 2, \cdots$; $j = 1, 2, \cdots$; $k = 1, 2, \cdots, n-1$, give the probability for undergoing a transition from the state E_i in the k-th instant of time to the state E_j in the $(k+1)$-st instant. Random variables connected by this Markov chain are defined simply as arbitrary functions $X_k(E_i)$. We shall assume again that for every n we have a sequence of variables connected by a Markov chain with n instants of time. Quantities referring to the n-th chain shall be indicated by a superscript n.

For an arbitrary stochastic matrix Q whose elements are (q_{ij}) we shall set the ergodic coefficient equal to

$$\alpha(Q) = \inf_{k,l} \sum_{m=1}^{\infty} \min(q_{km}, q_{lm}), \tag{1.19}$$

where the lower bound is taken over all pairs k and l, and the minimum is taken for just the two probabilities indicated. It is not difficult to show (and this will be done in one of the future parts of our work) that this definition agrees with the ergodic coefficient $\alpha(Q)$ defined by equation (1.5)

Let us write

$$\alpha_i^{(n)} = \alpha(P_i^{(n)}). \tag{1.20}$$

Let $S^{(n)} = X_1^{(n)} + \cdots + X_n^{(n)}$. With this notation, Theorem 1 is valid.

In a previous work [13] instead of $\alpha_i^{(n)}$ we used

$$\beta_i^n = \sum_{j=1}^{\infty} \min_k {}_{kj}P_i^n. \tag{1.21}$$

Obviously in all cases $\alpha_i^{(n)} \geqq \beta_i^{(n)}$, and in many cases $\alpha_i^{(n)} > \beta_i^{(n)}$. Therefore the results of the present work are more inclusive than those of the previous one.

It is clear that if for some i and n and for some state E_m we have

$$_{km}P_i^{(n)} \geqq \alpha, \tag{1.22}$$

then we also have $\alpha_i^{(n)} \geqq \alpha$. Therefore from Theorem 1 we obtain all the results enumerated in paragraph 1.1. as well as Sapogov's result mentioned in paragraph 1.2.

1.6. Let us replace the condition that the variables $X_i^{(n)}$ are uniformly bounded by the condition that their variances are uniformly bounded. This gives us the following theorem.

Theorem 2. *Let the variances of all the variables $X_i^{(n)}$ be uniformly bounded above and below, which means that there exist constants c and C such that*

$$0 < c \leqq \mathbf{D}X_i^{(n)} \leqq C < \infty. \tag{1.23}$$

Further, let requirements (1.7) *and* (1.8) *be fulfilled. Then if the sequence of distributions of normalized sums*

$$\frac{S^{(n)} - \mathbf{M}S^{(n)}}{\sqrt{\mathbf{D}S^{(n)}}} \tag{1.24}$$

converges as $n \to \infty$ to some proper distribution, this limit distribution will be infinitely divisible.

An analysis of Bernstein's example (see Section 2 in the next part of this work) shows that the limit distribution obtained there for the sequence in equation (1.24) is not infinitely divisible, so that condition (1.8) cannot be replaced by a more stringent one in Theorem 2. It turns out that by "restricting the asymptotic growth" of all the $X_i^{(n)}$ uniformly, a restriction which is less stringent than by requiring that they be uniformly bounded, one is unable to guarantee that the sums of (1.24) will be normal asymptotically, so long as the only condition imposed on the transition probabilities is (1.8). In order to obtain normality in the limit with weaker restrictions on the growth of the $X_i^{(n)}$ one would have to use stronger restrictions on the decrease of the $\alpha_i^{(n)}$.

We shall start formulating the results with the following theorem.

Theorem 3. *For any $m > 0$ let there exist a constant γ_m such that for all i and n*

$$\mathbf{M}|X_i^{(n)} - \mathbf{M}X_i^{(n)}|^m \leqq \gamma_m < \infty \tag{1.25}$$

and assume that (1.6) *is fulfilled. If for any $\varepsilon > 0$ we have*

$$\alpha^{(n)} n^{1/3 - \varepsilon} \to \infty, \tag{1.26}$$

then (1.9) *is fulfilled for the sequence of sums $S^{(n)}$.*

As is known, the condition that the variances be uniformly bounded is insufficient for asymptotic normality even for independent $X_i^{(n)}$. If $\alpha_i^n \geqq \alpha > 0$ we can formulate a sufficient condition for asymptotic normality, this condition being formally similar to the known necessary and sufficient condition for independent terms.

Theorem 4. *Assume that* (1.6) *is fulfilled and that for any* $r > 0$ *we have*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \int_{|t| \geqq r\sqrt{n}} t^2 dF_i^{(n)}(t) = 0, \tag{1.27}$$

where $F_i^{(n)}(t) = \mathbf{P}\{X_i^{(n)} < t\}$. *Finally, for all* n *let*

$$\alpha^{(n)} \geqq \alpha > 0. \tag{1.28}$$

Then (1.9) *is asymptotically normal.*

Now we shall indicate "with the precision of the power scale of growth" the general relation between the restrictions which must be imposed on the distribution of the $X_i^{(n)}$ and the behavior of the $\alpha^{(n)}$.

Theorem 5. *Let there be given some number* β, $0 < \beta < \frac{1}{3}$, *and let* (1.6) *and*

$$\alpha^{(n)} n^{\beta} \to \infty \tag{1.29}$$

be fulfilled. If for

$$m = 2\,\frac{1-\beta}{1-3\beta} \tag{1.30}$$

the moments

$$\mathbf{M}|X_i^{(n)} - \mathbf{M}X_i^{(n)}|^m \leqq \gamma_m < \infty \tag{1.31}$$

are bounded above uniformly in i *and* n, *then* (1.9) *is valid.*

It is possible, on the other hand, to construct an example of a sequence of variables connected by a Markov chain for which conditions (1.6) and (1.29) are fulfilled, and for which for all $\tilde{m} < m$ and i we have

$$\mathbf{M}|X_i^{(n)} - \mathbf{M}X_i^{(n)}|^{\tilde{m}} \leqq \gamma_{\tilde{m}} < \infty \tag{1.31$'$}$$

but for which the sequence (1.24) has a limit distribution which is not normal. It is clear that Theorem 3 is a corollary of Theorem 5. The case in which the terms $X_i^{(n)}$ themselves have a normal distribution is of some interest.

Theorem 6. *Assume that all the* $X_i^{(n)}$ *have normal distributions, and that their variances* $d_i^{(n)}$ *satisfy the condition*

$$0 < c \leqq d_i^{(n)} \leqq C < \infty. \tag{1.32}$$

Then a sufficient condition for the validity of the statement (1.9) *concerning the asymptotic normality of the* $\bar{S}^{(n)}$ *is that*

$$\lim_{n\to\infty} \alpha^{(n)} n^{1/3} (\log n)^{-1/3} = \infty. \tag{1.33}$$

Conversely, for any $a < \infty$ *it is possible to give an example such that*

$$\lim_{n\to\infty} \alpha^{(n)} n^{1/3} (\log n)^{-1/3} = a \tag{1.33$'$}$$

and the limit distribution is not normal.

Similar results can be obtained with other conditions imposed on the behavior of the $X_i^{(n)}$ at infinity. For instance, if all the $X_i^{(n)}$ have exponential distributions, the role of the "bounding function" is played by $(\log n)^{-2/3} n^{1/3}$ etc.

Let us now consider the case in which $|X_i^{(n)}| \leqq C^{(n)}$ but in which the constants $C^{(n)}$ themselves may be unbounded in the limit as $n \to \infty$.

Theorem 7. *Let condition* (1.23) *be fulfilled, and for all i and n let the probability that*

$$|X_i^{(n)}| \leqq C^{(n)} \tag{1.34}$$

holds be 1.

Then a sufficient condition for the validity of (1.9) *is that in the limit as $n \to \infty$ we have*

$$C^{(n)}(\alpha^{(n)})^{-3/2} n^{-1/2} \to 0. \tag{1.35}$$

Conversely, if there is some sequence of constants $C^{(n)}$ and ergodic coefficients $\alpha^{(n)}$ such that

$$\lim_{n\to\infty} C^{(n)}(\alpha^{(n)})^{-3/2} n^{-1/2} = a > 0, \tag{1.35$'$}$$

it is possible to construct an example in which the limit distribution is not normal.

Finally, let us formulate one general criterion for normality from which, as we shall show in Section 8 of the next part of the work, it is easy to obtain the sufficient conditions for normality which are contained in Theorems 3-7. Since these sufficient conditions are optimum conditions for the particular special cases, it is natural to suppose that this general criterion is in some sense optimal.

Theorem 8. *Assume that conditions* (1.23) *are fulfilled. Let us set*

$$F_i^{(n)}(t) = \mathbf{P}\{X_i^{(n)} - \mathbf{M}X_i^{(n)} \leqq t\}.$$

Then a sufficient condition for the validity of (1.9) *is that for all $r > 0$ we have*

$$\lim_{n\to\infty} \frac{1}{n(\alpha^{(n)})^2} \sum_{i=1}^{n} \int_{|t| \geqq rn^{1/2}(\alpha^{(n)})^{3/2}} t^2 dF_i^{(n)}(t) = 0. \tag{1.36}$$

From the point of view of our fundamental Theorem 1, the definition given in (1.7) would seem somewhat restrictive in that it assumes that for no i do any of the $\alpha_i^{(n)}$ vanish. It would seem that "sufficient ergodicity" of the chain, and therefore asymptotic normality of the sum of variables connected by a chain, could be assured merely be assuming that a sufficiently large number of the $\alpha_i^{(n)}$ never vanish. In one of the later parts of our work we shall present an example which forces one to treat such conclusions with great care. This example is one in which for all even i the ergodic coefficients satisfy the condition $\alpha_i^{(n)} \geqq \alpha > 0$ uniformly in i and n, whereas the limit distribution is not normal. One can, however, prove the following generalization of Theorem 1.

Theorem 9. *Assume that all the $X_i^{(n)}$ are uniformly bounded, satisfying $|X_i^{(n)}| \leqq C < \infty$ and that* (1.6) *is fulfilled. Further, let*

$$(1.37) \qquad \lim_{n\to\infty} n^{-2/3} \sum_{i=1}^{n-2} \min\,(\alpha_i^{(n)}, \alpha_{i+1}^{(n)}) = \infty$$

and assume that for all k and l such that $k-l \geqq n^{1/3}$ there exists a $d(n)$ such that $d(n) \to \infty$ as $n \to \infty$ for which

$$(1.38) \qquad \sum_{i=1}^{k} \alpha_i^{(n)} \geqq d(n).$$

Then (1.9) *is valid.*

It is also of some interest to make condition (1.6) less restrictive, replacing it by the assumption that a sufficient number of the variances $\mathbf{D}X_i^n$ fail to vanish.

Theorem 10. *Theorem* 1 *remains valid if* (1.6) *and* (1.8) *are replaced by the condition that as $n \to \infty$ we have*

$$(1.39) \qquad n^{-2/3}\alpha^{(n)}\,\Big(\sum_{i=1}^{n} \mathbf{D}X_i^{(n)}\Big) \to \infty.$$

It would be possible to formulate the results by combining the generalizations of Theorem 1 which are given by Theorems 2-10 into one proposition. In view of the already great extent of this work, however, we shall refrain from such a generalization.

1.7. N. A. Sapogov [29, 30] gave a simple method for obtaining many-dimensional limit theorems from simple limit theorems. Thus if one assumes in our Theorem 1 that the $X_i^{(n)}$ are h-dimensional vectors (the definition of vector variables connected by a chain is given by the obvious analogy with the scalar case; then $|X_i^{(n)}|$ is understood as the length of the vector) and replaces (1.6) by the condition

$$(1.40) \qquad \mathbf{D}(X_i^{(n)}, U) \geqq c > 0$$

for any h-dimensional vector U with $|U| = 1$ (here $(X_i^{(n)}, U)$ is understood as the scalar product), then the vector sequence $S^{(n)}$ will be normal asymptotically. Analogous generalizations can be made of Theorems 2-10.

1.8. A single infinite sequence of variables connected by a Markov chain deserves special consideration. In our language of a sequence of series this means that the transition functions $P_i(\omega, A)$ and their domains of definition are independent of n, and that the $X_i^{(n)}$ are also n-independent. In this case let us write $\alpha_i^{(n)} = \alpha_i$. Shirokorad [35] has investigated a sequence of variables connected by Markov chains with two states, and has proved in one special case that their sums are asymptotically normal if $\alpha_i i^{1-\varepsilon} \to \infty$ for $\varepsilon > 0$. In this connection A. N. Kolmogorov has raised the question of whether or not it is possible to replace the Bernstein condition $\alpha_i i^{1/3} \to \infty$ by the stronger condition $\alpha_i i \to \infty$. Bernstein's example refers to a more general type of series. In Section 10 we shall show, by generalizing Bernstein's construction, that for a sequence of series with $\alpha_i \sim c i^{-1/3}$, the central limit theorem is not in general fulfilled. The fact that Shirokorad was successful was due to the fact that he chose a special case in which the transition probabilities behave particularly smoothly [6].

1.9. We shall formulate, finally, a single ergodicity condition for nonstationary Markov chains, which is in some sense optimal. We include this result in the article, since it shows why it is natural in the theory of Markov chains to make use of the ergodic coefficient $\alpha(P)$ which we have introduced.

We shall assume for simplicity that the state space Ω and the σ-algebra $\mathfrak{A}$ of measurable sets are common to all moments of time. The Markov chain is given by a sequence of transition functions $P_i(\omega, A)$ $i = 1, 2, \cdots$. These transition functions can be used to express the probabilities $P_{k,l}(\omega, A)$ for going from the state ω at time k to the set of states A at time l. With Kolmogorov [7] we shall say that a Markov chain is ergodic if for all k and $\omega_1 \in \Omega$, $\omega_2 \in \Omega$ we have uniformly in $A \in \mathfrak{A}$

$$\lim_{n\to\infty} |P_{k,n}(\omega_1, A) - P_{k,n}(\omega_2, A)| = 0. \tag{1.41}$$

We shall say that a sequence of transition functions $\tilde{P}_i$ includes the sequence P_i if for some $i_1, i_2, \cdots, i_n, \cdots$, ($i_k \neq i_j$ if $k \neq j$) $\tilde{P}_{i_k}$ coincides with P_k. We shall say that the sequence P_i is strongly ergodic if for any sequence $\tilde{P}_i$ which includes P_i the Markov chain is ergodic. We shall write α_i for the ergodic coefficient $\alpha(P_i)$.

Theorem 11. *A necessary and sufficient condition for the sequence P_i to be strongly ergodic is*

$$\sum_{i=1}^{\infty} \alpha_i = \infty. \tag{1.42}$$

In particular, (1.42) is a sufficient condition for the chain given by the transition functions P_i to be ergodic. In spite of its simplicity, this sufficient condition for ergodicity has not been given in the literature, even for a finite chain. Almost all the ergodicity criteria with which we are familiar for nonstationary chains are also sufficient conditions for strong ergodicity and are therefore contained as special cases in (1.42) [17, 16], p. 206, [7], p. 362, [33]. The only exceptions are the recent very interesting results of T. A. Sarymsakov [28].

1.10. Let us introduce the following notation, which we shall use from now on. For $1 < k < l \leqq n$ we set

$$S_{k,l}^{(n)} = X_{k+1}^{(n)} + \cdots + X_l^{(n)}, \tag{1.43}$$

$$S_{1,l}^{(n)} = X_l^{(n)} + \cdots + X_l^{(n)}. \tag{1.43'}$$

[6] In our review of previous results we did not mention the work of Doeblin [10, 11] (see also Linnik and Sapogov [25]) which stands somewhat apart. Doeblin proves the central limit theorem for a finite nonstationary chain in the case in which all the transition probability matrices are in some sense of the same type, and their positive elements do not approach zero. This result is not entirely contained in our Theorem 1, since in Doeblin's work the ergodic coefficients may vanish. (On the other hand, Doeblin's result can be deduced from our theorem in an indirect way.)

In this connection one may hypothesize that the central limit theorem is true in a more general case if, on the one hand, one assumes that the transition probabilities are in some sense of a single type, and on the other that the ergodic coefficients may vanish. It is also possible that in such an "almost stationary" case it would be possible to allow chosen transition probability characteristics to approach zero not like $n^{-1/3}$, but like n^{-1}. This kind of result would include Shirokorad's result, as well as the results for a sequence of series of stationary chains [12].

1.11 [7]. We shall describe very briefly the basic methods for proving Theorems 1 and 2. We shall assume that $\mathbf{M}X_i^{(n)} = 0$. It is not difficult to show that $1-\alpha(P_{kl}^{(n)}) \leqq \prod_{i=k}^{l-1} (1-\alpha_i^{(n)})$. It follows from this that if, as is supposed in Theorems 1 and 2, $\alpha_i^{(n)} n^{1/3} \to \infty$, then for any $\varepsilon < 1$ there exists an n sufficiently large such that for all segments (i, k) with $k-i \geqq n^{1/3}$ the ergodic coefficient $\alpha(P_{k,l}^{(n)}) \geqq \varepsilon$. But if $\alpha(P_{k,l}^{(n)}) \geqq \varepsilon$, as is easily shown, the variables $S_{1,k}^{(n)}$ and $S_{l,n}^{(n)}$ become independent if they vary on a set of probability $1-\varepsilon$. It follows from this that if for any n one can choose k_n segments $(i, i+[n^{1/3}+1])$ so that $k_n \to \infty$ and so that the sum $\sum_{i=1}^{k_n} S_{i,i+[n^{1/3}+1]}^{(n)}$ can be neglected asymptotically compared with $S^{(n)}$, then the asymptotic behavior of $S^{(n)}$ will be the same as that of a sum of independent random variables, and will therefore be infinitely divisible. (An auxiliary difficulty is in choosing the segments of length $[n^{1/3}+1]$ which are "dropped out" so that the remaining terms are infinitesimally small.) A similar reduction to sums of independent random variables was used by Bernstein [2] for general sequences of weakly dependent variables, and by Doeblin [9] (see also the author's [12]) for stationary Markov chains.

We shall show that $\mathbf{D}S^{(n)} n^{-2/3} \to \infty$. It is clear that $\mathbf{D}S_{i,i+[n^{1/3}+1]}^{(n)} \leqq C^2(n+1)^{2/3}$. It follows from this that in an asymptotic consideration one can neglect $\sigma_n S_{i,i+[n^{1/3}+1]}^{(n)}$ and therefore also the set k_n if $k_n \to \infty$ sufficiently slowly. (Here $\sigma_n = (\mathbf{D}S^{(n)})^{-1/2}$ is a normalizing factor.) The fundamental difficulty in obtaining the lower bound given for $\mathbf{D}S^{(n)}$ lies in the fact that the variance $\mathbf{D}S_{1,m}^{(n)}$ does not vary monotonically as m increases. In our case it is found that Bernstein's method for obtaining this lower bound, which has been used in previous works, is insufficient. Let $\bar{D}_m^n$ be the mathematical expectation of the conditional variance of $S_{1,m}^n$ on the condition that the state of the system is fixed at the m-th instant of time. It is easily seen that in all cases we have $\mathbf{D}S_{1m}^{(n)} - \bar{D}_m^{(n)} = \tilde{D}_m^{(n)} \geqq 0$. We shall show that $\bar{D}_{m+1}^{(n)} - \bar{D}_m^{(n)} \geqq \frac{1}{100}\tilde{D}_m^{(n)}\alpha_m^{(n)}$ (or that $\bar{D}_m^{(n)}$ increases monotonically) and that $(\tilde{D}_m^{(n)}, \tilde{D}_{m+1}^{(n)}) \geqq \frac{1}{10}\mathbf{D}X_m^{(n)}$. This leads to the necessary result.

The most difficult part of the problem is in verifying that the sum of independent variables to which we have reduced the study of $S^{(n)}$ satisfies the condition of asymptotic normality (see p. 110 of Gnedenko and Kolmogorov [8]). We shall make use of a lemma which follows simply from these conditions and which is proved in Section 6.

Lemma. *Let $\mu_1^{(n)}, \cdots, \mu_{r_m}^{(n)}$ be independent random variales for fixed n such that the sum $\gamma^{(n)} = \mu_1^{(n)} + \cdots + \mu_{r_n}^{(n)}$ has a variance which satisfies $\mathbf{D}\gamma^{(n)} = 1$. Assume that as $n \to \infty$ the variances $\mathbf{D}\mu_k^{(n)} \to 0$ uniformly in k. Finally, let $\mu_k^{(n)} = \nu_k^{(n)} = \tilde{\nu}_k^{(n)}$, where as $n \to \infty$ we have $\mathbf{D}\nu_k^{(n)} \to 0$ and $\mathbf{D}\tilde{\nu}_k^{(n)} \to 0$ uniformly in k, where for some $C < \infty$ and for all n we have $\sum_{k=1}^{r_n} (\mathbf{D}\nu_k^{(n)} + \mathbf{D}\tilde{\nu}_k^{(n)}) \leqq C$, and where the $\nu_\alpha^{(n)}$ are independent of the $\nu_\beta^{(n)}, \tilde{\nu}_\beta^{(n)}$ for $\beta \neq \alpha$. Assume that the sums $\nu_1^{(n)} + \cdots + \nu_{r_n}^{(n)}$ and $\tilde{\nu}_1^{(n)} + \cdots + \tilde{\nu}_{r_n}^{(n)}$ are normal in the limit as $n \to \infty$. Then the $\gamma^{(n)}$ are also normal in the limit.*

Each of the independent terms, to the sum of which the study of the $S^{(n)}$ may be reduced, has the same distribution as the random variables $\sigma_n S_{k,l}^{(n)}$ for some values of k and l. By properly breaking up each of the $\sigma_n S_{k,l}^{(n)}$ into a sum

[7] This paragraph can be omitted by the reader without detriment to the understanding of the further material.

of two terms ν and $\tilde{\nu}$ we may, according to the lemma, prove the normality in the limit of the sums of terms ν and the sums of terms $\tilde{\nu}$. The idea of thus breaking up the problem can be explained as follows. We break up the segment $[k, l]$ into a sum of contiguous segments each of which, with the exception of the last, is of length $[n^{1/3}+1]$, ordering them naturally. We denote this division by R. We now set ν equal to the product of the normalizing factor σ_n by the $X_i^{(n)}$ summed over all i belonging to the segments of R which are assigned even numbers, doing the same for $\tilde{\nu}$ with odd numbers. The meaning of this representation is that the term ν is itself a sum of infinitesimal and almost independent random variables ζ which are sums of the $X_i^{(n)}$ over separate segments of R. Indeed, two neighboring terms ζ are separated by segments of length $[n^{1/3}+1]$ and the ergodic coefficient for the transition probability in this segment of time, as has been mentioned, approaches 1 as $n \to \infty$. According to the conditions of Theorem 1 the terms are bounded by infinitesimal constants, since they are the sum of $[n^{1/3}+1]$ uniformly bounded quantities $X_i^{(n)}$ and the normalizing factor σ_n has the property that $\sigma_n n^{1/3} \to 0$. An additional difficulty consists in attaining the condition that the sums of the dispersions be uniformly bounded, as is necessary in the lemma. This requires complicating the construction just described (see Section 6). For simplicity, however, we shall not go into these complications at present.

Let us now assume that $\alpha^{(n)} n^{1/3} \to \infty$ where $\varepsilon > 0$. A simple calculation shows that then for any segment $[k, l]$ of length $[n^{1/3}+1]$ the ergodic coefficient for sufficiently large n is greater than $1-n^{-1/2\varepsilon}$. Since the number of terms grows as a power function, and the "ergodicity" grows exponentially, one can neglect entirely the interdependence of the terms which form the sum ν. The study of $\sum \nu$ reduces to the study of $\sum \zeta$ where all the ζ are independent and uniformly bounded by an infinitesimal constant, which means that $\sum \zeta$ is normal in the limit. A more exact calculation shows that these considerations are valid if $\alpha^{(n)} n^{1/3} (\log n)^{-1} \to \infty$.

The proof that the $\sum \nu$ is normal in the limit for the optimum condition $\alpha^{(n)} n^{1/3} \to \infty$ requires a more complicated construction. We will later present the proof based on the fact that the sum is presented in the form of an unbounded number of terms each of which is a martingale, and to these we may apply the evaluations obtained by Levy [22] in proving the central limit theorem for variables connected by a martingale.

The present work is a candidate's dissertation carried out under the direction of Andrei Nikolaevich Kolmogorov. The author expresses his great indebtedness to Andrei Nikolaevich for important comments in completing the work. The author also expresses his gratitude both for interest and for valuable advice to E. B. Dynkin, Iu. V. Linnik, N. A. Sapogov, B. V. Gnedenko, Iu. V. Prokhorov, and A. V. Skorokhod.

Received by the editorial office December 20, 1955.

REFERENCES

1. S. N. Bernstein, *Sur le théorème limite du calcul des probabilités*, Math. Ann., 85 (1922), pp. 237–241.
2. S. N. Bernstein, *Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes*, Math. Ann., 97 (1926), pp. 1–59 (Russian translation in Uspekhi Mat. Nauk, 10 (1944), pp. 65–144).

3. S. N. Bernstein, *Sur les sommes de quantités dépendantes*, Izv. Akad. Nauk SSSR., VI Ser., 15–17 (1926), pp. 1459–1478.
4. S. N. Bernstein, *Sur les sommes de quantités dépendantes*, Dokl. Akad. Nauk SSSR., A, 4 (1928), pp. 55–60.
5. S. N. Bernstein, *Détermination d'une limite intérieure de la dispersion des sommes de grandeurs liées en chaîne singulière*, Mat. Sb., 1: 1 (1936), pp. 29–38.
6. S. N. Bernstein, *A Course in Probability Theory*, 4th ed., Moscow-Leningrad, Gostekhizdat, 1946 (In Russian).
7. G. Birkhoff, *Lattice Theory*, GIIL, Moscow, 1952 (Russian translation).
8. B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*, Moscow-Leningrad, Gostekhizdat, 1949 (In Russian).
9. W. Doeblin, *Sur les propriétés asymptotiques de mouvement regis par certains types des chaînes simples*, Bull. Math. Soc. Rom. Sci., 39: 1 (1937), pp. 57–115; 39: 2, pp. 3–61.
10. W. Doeblin, *Sur les chaînes de Markoff*, Comp. Rend. Paris, 203 (1936), pp. 1210–1211.
11. W. Doeblin, *Le cas discontinu des probabilités en chaîne*, Publ. Fac. Sci. Univ. Masaryk, 236 (1937), pp. 3–13.
12. R. L. Dobrushin, *Limit theorems for Markov chains with two states*, Izv. Akad. Nauk SSSR. Ser. Mat., 17: 4 (1953), pp. 291–330.
13. R. L. Dobrushin, *Central limit theorem for nonstationary Markov chains*, Dokl. Akad. Nauk SSSR., 102: 1 (1955), pp. 5–8.
14. J. L. Doob, *Stochastic Processes*, John Wiley & Sons, Inc., New York, Chapman & Hall, Ltd., London, 1953.
15. E. B. Dynkin, *On certain limit theorems for Markov chains*, Ukrain. Mat. Zh., 6: 1 (1954), pp. 21–27.
16. Yosida Kosaku and Kakutani Shizuo, *Application of the mean ergodic theorem to the problem of Markov's process*, Proc. Imp. Acad. Tokyo, 14: 9 (1938), pp. 333–339.
17. A. N. Kolmogorov, *On analytic methods in probability theory*, Uspekhi Mat. Nauk, 5 (1938), pp. 5–41.
18. A. N. Kolmogorov, *On some developments in probability theory*, Work of the Second All-Soviet Mathematical Conference, Leningrad, June 24—30, 1934, vol. 1, Plenary Session and Reviews, Izd. Akad. Hauk SSSR., Leningrad-Moscow, 1935, pp. 349–358.
19. A. N. Kolmogorov, *Fundamental Concepts of Probability Theory*, ONTI, Moscow-Leningrad, 1936 (In Russian).
20. H. Cramer, *Random Variables and Probability Distributions*, GIIL, Moscow, 1947 (Russian translation).
21. H. Cramer, *Mathematical Methods of Statistics*, GIIL, Moscow, 1948 (Russian translation).
22. P. Levy, *Théorie de l'addition des aléatoires*, Paris, 1937.
23. Yu. V. Linnik, *On nonstationary Markov chains*, Dokl. Akad. Nauk SSSR., 60: 1 (1948), pp. 21–24.
24. Yu. V. Linnik, *Contribution to the theory of nonstationary Markov chains*, Izv. Akad. Nauk SSSR. Ser. Mat., 13: 1 (1949), pp. 65–94.
25. Yu. V. Linnik and N. A. Sapogov, *Many-dimensional integral and local laws for nonstationary Markov chains*, Izv. Akad. Nauk SSSR. Ser. Mat., 13: 6 (1949), pp. 533–566.
26. A. A. Markov, *Investigations of general experiments connected by a Markov chain*, Zap. Akad. Nauk Fiz. Mat. Otdel., VIII Ser., 25: 3 (1910); A. A. Markov, Collected Works, Izd. Akad. Nauk SSSR., 1951, pp. 465–509.
27. V. I. Romanovskii, *Discrete Markov Chains*, Gostekhizdat, Moscow-Leningrad, 1949 (In Russian).
28. T. A. Sarymsakov, *On the ergodic principle for nonstationary Markov chains*, Dokl. Akad. Nauk SSSR., 90: 1 (1953), pp. 25–28.
29. N. A. Sapogov, *On singular Markov chains*, Dokl. Akad. Nauk SSSR., 58: 2 (1947), pp. 193–196.
30. N. A. Sapogov, *The Laplace-Laypunov limit theorem for singular Markov chains*, Dokl. Akad. Nauk SSSR., 58: 9 (1947), pp. 1905–1908.
31. N. A. Sapogov, *Concerning a certain limit theorem*, Dokl. Akad. Nauk SSSR., 69: 1 (1949), pp. 15–18.
32. N. A. Sapogov, *On many-dimensional nonstationary Markov chains*, Dokl. Akad. Nauk SSSR., 69: 2 (1949), pp. 137–139.
33. S. Kh. Sirazhdinov, *The ergodic principle for nonstationary Markov chains*, Dokl. Akad. Nauk SSSR., 71: 5 (1950), pp. 829–830.
34. P. Halmos, *Measure Theory*, GIIL, Moscow, 1953 (Russian translation).
35. B. V. Shirokorad, *On the applicability of the central limit theorem to Markov chains*, Izv. Akad. Nauk SSSR. Ser. Mat., 18: 1 (1954), pp. 95–104.

CENTRAL LIMIT THEOREM FOR NONSTATIONARY MARKOV CHAINS

R. DOBRUSHIN (*MOSCOW*)

(*Summary*)

This paper is a continuation of investigations on the central limit theorem for nonstationary Markov chains, carried out by Markov (1910), Bernstein (1922—1936), Sapogov (1947—1949) and Linnik (1948—1949).

Let Ω_i, $i = 1, 2$, be sets of states of the chain, and $\mathfrak{A}_i$ be σ-algebras of measurable subsets of these sets. A function $\mathbf{P}(x, B)$ $(x \in \Omega_1, B \in \mathfrak{A}_2)$ is called a stochastic transition function if it has the following properties:

1) if $\mathbf{P}(x, B)$ for fixed B is a measurable function of x;

2) if $\mathbf{P}(x, B)$ for fixed x is a probability measure in B. Let Λ_i, $i = 1, 2$, be Banach spaces of completely additive functions of sets $\lambda(B_i)$ $(B_i \in \mathfrak{A}_i)$ with $\lambda(\Omega_i) = 0$ and with norm

$$||\lambda|| = \sup_{B \in \mathfrak{A}_i} |\lambda(B)|.$$

By means of the equation

$$\mathbf{P}\lambda(B) = \int_{\Omega_1} \mathbf{P}(x, B)\lambda(dx), \qquad B \in \Omega_2,$$

any stochastic transition function defines the operator $\mathbf{P}$ taking Λ_1 into Λ_2. Let $N(\mathbf{P})$ be the norm of this operator $\mathbf{P}$.

The number $\alpha(\mathbf{P}) = 1 - N(\mathbf{P})$ is called the ergodic coefficient of this stochastic transition function. If we have a denumerable chain, then $\mathbf{P}$ is a stochastic matrix (p_{ij}) and

$$\alpha(\mathbf{P}) = \inf_{k,l} \sum_{m=1} \min(p_{km}, p_{lm}).$$

A nonstationary chain with n moments of time is defined by a sequence of pairs $(\Omega_i^{(n)} \cdot \mathfrak{B}_i^{(n)})$, $i = 1, \ldots, n$, transition functions $\mathbf{P}_i^{(n)}(x, B)$, $(x \in \Omega_i^{(n)}, B \in \Omega_{i+1}^{(n)}, i = 1, \ldots, n-1)$ and the initial probability distribution $\mathbf{P}^{(n)}(B)$ $(B \in \mathfrak{A}_1)$. In a natural manner the probability measure is given in the space $\Omega^{(n)} = \Omega_1^{(n)} \times \ldots \times \Omega_n^{(n)}$. Let $X_i^{(n)}$ be a measurable function defined on $\Omega_i^{(n)}$ and continued onto $\Omega^{(n)}$, and let

$$S^{(n)} = X_1^{(n)} + X_2^{(n)} + \ldots + X_n^{(n)}.$$

We shall consider the limit distribution for a sequence of random variables

$$\bar{S}^{(n)} = \frac{S^{(n)} - \mathbf{E}S^{(n)}}{\sqrt{\mathbf{D}S^{(n)}}}.$$

By definition

$$\alpha^{(n)} = \min_i \alpha(\mathbf{P}_i^{(n)}).$$

Theorem 1: *If*

1) $\mathbf{D}X_i^{(n)} > c > 0$,

2) $|X_i^{(n)}| < C < \infty$,

3) $n^{1/3} \min_i \alpha(\mathbf{P}_i^{(n)}) \to \infty \quad (n \to \infty)$,

then $\bar{S}^{(n)}$ *is an asymptotically normal sequence of random variables.*

If we replace condition (2) *by*

(2′) $\mathbf{D}X_i^{(n)} < C < \infty$,

then the limit distribution for $\bar{S}^{(n)}$ *will become an infinitely divisible distribution if it exists.*

This theorem is not valid if we replace condition (3) by

(3′) $n^{1/3} \min_i \alpha(\mathbf{P}_i^{(n)}) \to k < 0 \quad (n \to \infty)$.

Theorem 8: *If conditions* (1), (2′) *and* (3) *are satisfied,*

$$F_i^{(n)}(t) = \mathbf{P}\{X_i^{(n)} - \mathbf{E}X_i^{(n)} < t\}$$

and

(4) $$\lim_{n\to\infty} \frac{1}{n(\alpha^{(n)})^2} \sum_{i=1}^{n} \int_{|t| \geq rn^{1/2}(\alpha^{(n)})^{3/2}} t^2 dF_i^{(n)}(t) = 0,$$

then $\bar{S}^{(n)}$ *is an asymptotically normal sequence of random variables.*

If we replace condition (4) by a less stringent condition, Theorem 8 is not valid.

A criterion for the ergodic properties of nonstationary Markov chains is also given.