LIMIT DISTRIBUTIONS FOR SUMS OF INDEPENDENT RANDOM VARIABLES

LIMIT DISTRIBUTIONS FOR SUMS OF INDEPENDENT RANDOM VARIABLES

by

B. V. GNEDENKO and A. N. KOLMOGOROV

Translated from the Russian and annotated

by

K. L. CHUNG
Syracuse University

With an Appendix by

J. L. DOOB
University of Illinois

1954

ADDISON-WESLEY PUBLISHING COMPANY, INC.
CAMBRIDGE 42, MASS.

Printed in the United States of America

Library of Congress Catalog No. 53-10460

TRANSLATOR'S PREFACE

This is a translation of the Russian book **ПРЕДЕЛЬНЫЕ РАСПРЕДЕЛЕНИЯ ДЛЯ СУММ НЕЗАВИСИМЫХ СЛУЧАЙНЫХ ВЕЛИЧИН** (1949). There are various points of contact with the treatises by P. Lévy [76] and by H. Cramér [21], but much of the material in the book has been hitherto available only in periodical articles, many of which are in Russian. The systematic account presented here combines generality with simplicity, making some of the most important and difficult parts of the theory of probability easily accessible to the reader. Beyond a knowledge of the calculus on the level of, say, Hardy's *Pure Mathematics*, the book is formally self-contained. However, a certain amount of mathematical maturity, perhaps a touch of single-minded perfectionism, is needed to penetrate the depth and appreciate the classic beauty of this definitive work.

It is hoped that the English translation may serve both as a standard reference on the subject and as a text or supplementary reading for advanced courses in probability. Part of the book may also be used to suit other needs. For example, Chapters 1 and 2 may serve as the basis for any rigorous course in probability. Readers who are interested in learning the fundamental facts about stable laws and the more general infinitely divisible laws may then go on to §§ 16–18 and §§ 33–34. Those who are interested in the (weak) law of large numbers, the central limit theorem, and the analogous limit theorem leading to the Poisson law in their simpler formulations may find their needs met in § 21. Those who are interested in asymptotic expansions will need only Chapters 1, 2, 8, and 9; in particular, §§ 46–47, 49, and 51 are elementary and will be found useful for many applications.

Now a few words about the translation as compared with the original. There are two major textual changes in the English edition. The first occurs in § 32, where a mistake found in the original necessitated the deletion of several paragraphs there and thereafter. The details are explained in the second half of Appendix II. The second change occurs in §§ 46–47, where I have incorporated a substantial improvement from the 1951 Hungarian translation; see the Translator's Note to Theorem 1 of § 46.

Some minor corrections, including those of misprints, are made without mention; in a few places I have profited by the Hungarian edition which corrected some of the errors in the Russian edition. In other cases where I found fault with the Russian text, I have added a note in addition to, or instead of, changing the text. As a result, about fifty such notes are appended. These Translator's Notes are also used to supply references omitted by the authors and to add further explanatory remarks. In one case, namely in connection with Theorem 1 of § 32, where a rather long note would be needed, I have put the added material in the first part of Appendix II.

Appendix I was written by J. L. Doob and should be of interest to the reader who may be puzzled by the measure-theoretic complications in Chapter 1.

Of the many friends who have lent me assistance of one kind or another, the following persons deserve special mention: J. L. Doob, for a variety of advice and aid; F. J. Dyson, for consultations on the Russian language; G. A. Hunt, for critically reading the manuscript; J. V. Wehausen, for helping with the Bibliography; J. Wolfowitz, for encouragement in the rather thankless job of translating. Miss Madelyn M. Keady typed the manuscript expertly and tirelessly, and my only regret is that we did not fully utilize her flawless efforts, since the formula matter was reproduced directly from the Russian edition to reduce the cost of printing. The undertaking of the translation was part of a project at Cornell University in 1952–1953, under a contract with the Air Research and Development Command, whose support is gratefully acknowledged here.

K. L. C.

CONTENTS

PREFACE

1

In the formal construction of a course in the theory of probability, limit theorems appear as a kind of superstructure over elementary chapters, in which all problems have finite, purely arithmetical character. In reality, however, the epistemological value of the theory of probability is revealed only by limit theorems. Moreover, without limit theorems it is impossible to understand the real content of the primary concept of all our sciences — the concept of probability. In fact, all epistemologic value of the theory of probability is based on this: that large-scale random phenomena in their collective action create strict, nonrandom regularity. The very concept of mathematical *probability* would be fruitless if it did not find its realization in the *frequency* of occurrence of events under large-scale repetition of uniform conditions (a realization which is always approximate and not wholly reliable, but that becomes, in principle, arbitrarily precise and reliable as the number of repetitions increases).

Therefore the elementary arithmetical calculations of probabilities relating to games of chance, in the works of Pascal and Fermat, can be considered only as the pre-history of the theory of probability, while its proper history began with the limit theorems of Bernoulli ([3], 1713) and de Moivre ([86], 1730). The fundamental importance of the result of de Moivre was completely revealed by Laplace ([72], 1812). To the limit theorems of Bernoulli and de Moivre-Laplace it is natural to add three more limit theorems of Poisson as the principal achievements of the theory of probability before Chebyshev. One of them generalizes the theorem of Bernoulli, another the theorem of de Moivre-Laplace, and the third leads to the so-called Poisson law of distribution. For a clear understanding of what follows it is useful to cite here somewhat modernized formulations of the five limit theorems enumerated above.

The first four deal with a sequence of *independent* events

$$\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \ldots$$

We shall denote the probabilities of these events by

$$p_n = \mathsf{P}(\mathcal{E}_n),$$

and the number of actually occurring events among the first n events

$$\mathfrak{E}_1, \mathfrak{E}_2, \ldots, \mathfrak{E}_n$$

by μ_n. In the first two theorems all p_n have the same value p ($p \neq 0$, $p \neq 1$).

1. BERNOULLI'S THEOREM. For every $\epsilon > 0$

$$\mathsf{P}\left(\left|\frac{\mu_n}{n} - p\right| > \epsilon\right) \to 0$$

as $n \to \infty$.

2. LAPLACE'S THEOREM.

$$\mathsf{P}\left\{z_1 < \frac{\mu_n - np}{\sqrt{np(1-p)}} < z_2\right\} \to \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{z^2}{2}} dz$$

as $n \to \infty$ uniformly with respect to z_1 and z_2.

In the next two theorems p_n may depend on n, but subject to the condition that the series

$$\sum p_n (1 - p_n)$$

diverges. We set

$$p_1 + p_2 + \ldots + p_n = A_n,$$

$$p_1(1-p_1) + p_2(1-p_2) + \ldots + p_n(1-p_n) = B_n^2.$$

3. LAW OF LARGE NUMBERS IN POISSON'S FORM. For every $\epsilon > 0$

$$\mathsf{P}\left(\left|\frac{\mu_n}{n} - \frac{A_n}{n}\right| > \epsilon\right) \to 0$$

as $n \to \infty$.

4. CENTRAL LIMIT THEOREM IN POISSON'S FORM.

$$\mathsf{P}\left\{z_1 < \frac{\mu_n - A_n}{B_n} < z_2\right\} \to \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{z^2}{2}} dz$$

as $n \to \infty$ uniformly with respect to z_1 and z_2.

The fifth of the theorems we are interested in deals with a scheme of events

$$\begin{array}{l} \mathfrak{E}_{11}, \\ \mathfrak{E}_{21}, \mathfrak{E}_{22}, \\ \mathfrak{E}_{31}, \mathfrak{E}_{32}, \mathfrak{E}_{33}, \\ \cdots\cdots\cdots\cdots \\ \mathfrak{E}_{n1}, \mathfrak{E}_{n2}, \mathfrak{E}_{n3}, \ldots, \mathfrak{E}_{nn}, \\ \cdots\cdots\cdots\cdots \end{array}$$

in which the events in the same row are mutually independent and have the same probability p_n, depending only on the index of the row. We denote by μ_n the number of events in the nth row which actually occur.

5. POISSON'S LIMIT THEOREM FOR RARE EVENTS. If

$$np_n \to a$$

as $n \to \infty$, then

$$\mathsf{P}(\mu_n = m) \to \frac{a^m}{m!} e^{-a}.$$

By introducing the random variables

$$\xi_{\mathcal{E}} = \begin{cases} 1 \text{ if } \mathcal{E} \text{ occurs,} \\ 0 \text{ if } \mathcal{E} \text{ does not occur,} \end{cases}$$

we can write

$$\mu_n = \xi_{\mathcal{E}_1} + \xi_{\mathcal{E}_2} + \dots + \xi_{\mathcal{E}_n},$$

in Theorems 1, 2, 3, 4, and

$$\mu_n = \xi_{\mathcal{E}_{n1}} + \xi_{\mathcal{E}_{n2}} + \dots + \xi_{\mathcal{E}_{nn}}.$$

in Theorem 5.

This makes it possible to include all five limit theorems enumerated above as very special cases of limit theorems concerning *sums of independent random variables.*

The idea that the normal probability distribution

$$\mathsf{P}(\zeta < z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{z^2}{2}} dz,$$

which turned out to be the limit in Theorems 2 and 4, must also appear in a more general problem about the limit distribution of the sum of a large number of individually negligible independent summands is one of the essential ideas of the theory of errors developed by Gauss. However, in the matter of rigorous proofs Gauss did not reach results equivalent to the theorem of de Moivre-Laplace.

Effective methods for the rigorous proof of limit theorems concerning sums of arbitrarily distributed independent variables were created in the second half of the nineteenth century by Chebyshev. His classical work opened a new period of development of the entire theory of probability.

2

All of Chebyshev's efforts were devoted to the solution of two problems. Consider a sequence of independent random variables

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots$$

having finite mathematical expectations

$$a_n \doteq \mathsf{M}\xi_n$$

and finite variances

$$b_n^2 = \mathsf{D}^2\xi_n = \mathsf{M}\,(\xi_n - a_n)^2.$$

Put

$$\zeta_n = \xi_1 + \xi_2 + \ldots + \xi_n,$$
$$A_n = a_1 + a_2 + \ldots + a_n,$$
$$B_n^2 = b_1^2 + b_2^2 + \ldots + b_n^2.$$

FIRST PROBLEM. What additional conditions ensure the *law of large numbers: for every* $\epsilon > 0$

$$\mathsf{P}\left(\left|\frac{\zeta_n}{n} - \frac{A_n}{n}\right| > \varepsilon\right) \to 0$$

as $n \to \infty$?

SECOND PROBLEM. What additional conditions ensure the *central limit theorem:*

$$\mathsf{P}\left(\frac{\zeta_n - A_n}{B_n} < z\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{z^2}{2}}\, dz$$

as $n \to \infty$ uniformly with respect to z?

For application to the first problem the method developed by Chebyshev in his work ([16], 1867) requires only the condition

$$B_n = o(n),$$

This is usually called Markov's condition, since Markov first pointed out clearly the degree of generality of Chebyshev's reasoning. The law of large numbers under Markov's condition not only includes Theorems 1 and 2 of Bernoulli and Poisson, but in the great majority of applications more or less completely settles the question for sums of independent summands.

The solution of the second problem was considerably harder. For it Chebyshev created the method of moments, which is one of his most important achievements in mathematics. The solution given by Chebyshev in his paper ([17], 1887) is based on a lemma which was proved only later by Markov ([82], 1898). Soon afterwards the second problem of Chebyshev was solved by Lyapunov under considerably more general conditions by another method ([79], 1900; [80], 1901). Subsequently Markov succeeded in proving that the method of moments is capable of giving as general a result as that obtained by Lyapunov. However, the method of Lyapunov turned out in its further development to be much simpler and more powerful in application to the entire circle of problems concerning limit theorems for sums of independent variables. This is the method of characteristic functions, which is the principal method employed in our book.

The solution given by Lyapunov satisfies all the needs of the great majority of applications. Nevertheless, we shall give instead of Lyapunov's theorem the solution of Chebyshev's second problem in the form of Theorem 4 of § 21. The condition used there, namely Lindeberg's condition that for every $\epsilon > 0$

$$\lim_{n \to \infty} \sum_{k=1}^{n} \mathsf{P}\{|\xi_k - a_k| \geqslant \varepsilon B_n\} = 0,$$

is somewhat broader than Lyapunov's condition. In its logical structure it is even simpler than Lyapunov's condition

$$\lim_{n \to \infty} \frac{C_n}{B_n^{2+\delta}} = 0,$$

where

$$C_n = c_1 + \ldots + c_n,$$
$$c_k = \mathsf{M}|\xi_k - a_k|^{2+\delta}.$$

3

Let us turn to the simpler special case of a sequence

$$\xi_1, \ \xi_2, \ \ldots, \ \xi_n, \ \ldots$$

of independent *identically distributed* variables. In this case, the central limit theorem is applicable without any additional conditions other than the mere existence of the mathematical expectations

$$a_n = a$$

and variances

$$b_n = b$$

(see Theorem 4 of § 35). However, it is erroneous to conclude, even for the case of identically distributed summands, that there exist no really interesting limit theorems in which the limit laws are different from the normal law.†

In order to show by an example that such an opinion is only deep-rooted prejudice, we now consider the simple, classical scheme of random *motion* on a straight line, corresponding to the game of "heads or tails":

$$\eta(0) = 0,$$

$$\eta(t+1) = \begin{cases} \eta(t) + 1 \text{ with probability } \frac{1}{2} \\ \eta(t) - 1 \text{ with probability } \frac{1}{2} \end{cases}$$

independently of what

$$\eta(1),\ \eta(2),\ \ldots,\ \eta(t)$$

are.

It is well known that this scheme is the simplest of a long series of random motion schemes which have great importance in the most varied applications of the theory of probability, very remote from games of chance.

We number in an increasing sequence all the values of t for which

$$\eta(t) = 0.$$

We obtain (with probability one) an infinite sequence

$$0 = \tau_0 < \tau_1 < \tau_2 < \ldots < \tau_n < \ldots$$

The differences

$$\xi_n = \tau_n - \tau_{n-1}$$

form a sequence of independent and identically distributed random variables. Each of the variables ξ_n takes only positive even values with probabilities

$$p_m = \mathsf{P}(\xi_n = 2m) = \frac{2m\,(2m-2)!}{2^{2m}\,(m!)^2}.$$

Since

$$p_m \sim \frac{1}{2\sqrt{\pi}\, m^{3/2}},$$

asymptotically as $n \to \infty$, the mathematical expectation

$$\mathsf{M}\xi_n = 2 \sum_{m=1}^{\infty} m p_m$$

is infinite. Nevertheless, the sums

$$\tau_n = \xi_1 + \xi_2 + \ldots + \xi_n,$$

† *Translator's note.* The word "law" is taken to be synonymous with "distribution" in such contexts. In "the normal (or Poisson) law" often the corresponding type (see § 10) is meant.

with suitable normalization, are subject in the limit to a completely determined law of distribution:

$$\lim_{n\to\infty} \mathsf{P}\left(\frac{2\tau_n}{\pi n^2} < z\right) = \begin{cases} 0 & \text{for} \quad z \leqslant 0, \\ \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^z e^{-\frac{1}{2z}} z^{-\frac{3}{2}} dz & \text{for} \quad z > 0 \end{cases}$$

(see in this connection Theorem 5 of § 35 and the end of § 34).

The reader should turn his attention to n^2 in the denominator of the expression

$$\frac{2\tau_n}{\pi n^2}.$$

In the case of the sum ζ_n of identically distributed independent variables with finite variances the denominator of the expression

$$\frac{\zeta_n - A_n}{B_n}$$

in the central limit theorem would have the order $\sqrt{n}$. Comparison of these two special cases compels us to pose this general problem: Under what conditions on identically distributed independent variables

$$\xi_1, \ \xi_2, \ \ldots, \ \xi_n, \ \ldots$$

can a limit relation

$$\mathsf{P}\left\{ \frac{\zeta_n - A_n}{B_n} < z \right\} \to V(z)$$

hold, where A_n and B_n are constants, and what kind of limit laws $V(z)$ can appear?

The question about the class ~~of limit~~ laws which can possibly appear in the situation indicated above was completely settled by A. Ya. Khintchine. It turned out that up to linear transformations this class consists only of the normal law, occupying a special position; the unitary law

$$\varepsilon(x) = \begin{cases} 0 & \text{for} \quad x \leqslant 0, \\ 1 & \text{for} \quad x > 0; \end{cases}$$

and a family of distribution laws with infinite variances, depending on two parameters (α and β in the notations of Ch. 7). All these distribution laws, called "stable" because of circumstances which are explained in § 33, deserve the most serious attention. It is probable that the scope of applied problems in which they play an essential role will become in due course rather wide.

4

Poisson's limit theorem for rare events should long ago have suggested that even in the case of finite variances there can exist interesting and useful limit theorems concerning sums of independent variables and leading to distribution laws essentially different from the normal. To obtain them in a systematic way, it is natural to turn to the scheme of a double sequence of random variables

$$(\xi_{n1}, \xi_{n2}, \ldots, \xi_{nm_n}), \quad n = 1, 2, 3, \ldots,$$

where the random variables of the same row are independent, and to consider the sums

$$\zeta_n = \xi_{n1} + \xi_{n2} + \ldots + \xi_{nm_n}.$$

The simplest and most important case is that in which all variables ξ_{nk} in the same row are identically distributed. The problem consists as before in classifying the conditions under which a limit relation

$$\mathsf{P}\left\{\frac{\zeta_n - A_n}{B_n} < z\right\} \to V(z)$$

can hold and what kind of laws $V(z)$ can appear. Here, of course, it is natural to consider only the case where

$$m_n \to \infty.$$

It is curious that if all random variables ξ_{nk} can take only two values x' and x'' independent of the indices n and k, then the only possible limit laws (up to a linear transformation) will be the normal law, the improper law $\epsilon(x)$ and the family of Poisson laws with one parameter a (see Kozulyaev [70]).

The class of possible limit laws in such a formulation of the problem coincides with the class of infinitely divisible laws, to which Chapter 3 is devoted. Naturally, it contains all the stable laws and Poisson's law. The corresponding limit theorems are proved in Chapter 4. Here we only mention that for a better understanding of their intuitive meaning it may be useful for the reader to become acquainted with a special case treated in the book of A. Ya. Khintchine [53] under the name of "generalized limit theorem of Poisson." This elementary limit theorem leads only to those infinitely divisible laws with characteristic functions of the form

$$f(t) = \exp\left\{c \int (e^{iut} - 1)\, dF(u)\right\}$$

(see § 16). Distribution laws of this type have as many finite moments as does their generating distribution $F(u)$. It is possible to indicate many physical and technical problems leading to them.

Among infinitely divisible laws, belonging neither to the class of stable laws nor to that of laws of the special type just mentioned, we mention also a family of distributions well known in mathematical statistics. They are given by the incomplete gamma functions

$$V(z)=\begin{cases} 0 & \text{for} \quad z \leqslant 0, \\ \dfrac{1}{\Gamma(\alpha)}\displaystyle\int_0^z z^{\alpha-1} e^{-z}\,dz & \text{for} \quad z > 0, \end{cases}$$

depending on the parameter $\alpha > 0$ (see Example 4, § 17). To this family belongs in particular (for $\alpha = 1$), the exponential distribution

$$V(z)=\begin{cases} 0 & \text{for} \quad z \leqslant 0, \\ 1-e^{-z} & \text{for} \quad z > 0. \end{cases}$$

If we renounce the assumption that all the random variables in the same row have the same law of distribution, then the problem of determining all possible laws $V(z)$, in its exact formulation above, becomes meaningless. The limit law $V(z)$ can be absolutely arbitrary. This is indeed natural, since now the requirement $m_n \to \infty$ is illusory. It does not prevent, for example, that in each row one single summand ξ_{nk} plays the dominating role. Meaningful results, comformable to the original lofty conception of the classical limit theorems in the theory of probability, are obtained only under the following additional requirement: for every $\epsilon > 0$ there should exist constants a_{nk} such that

$$\sup_{1 \leqslant k \leqslant m_n} \mathsf{P}\{|\xi_{nk} - a_{nk}| \geqslant \epsilon B_n\} \to 0.$$

This requirement of the "asymptotic negligibility" of the variation of each individual summand in comparison with the chosen scale B_n for the sum ζ_n is quite natural. In § 20 it is introduced in the particular case $B_n = 1$ under the name "asymptotic constancy."

A. Ya. Khintchine proved that with this restriction the only possible limit laws in the case of arbitrarily distributed terms are the same infinitely divisible laws as in the identically distributed case (§ 24). Therefore it is quite natural that the infinitely divisible laws turn out to be the central concept throughout the first part of this book. It seems to us that the theory of these laws and the general limit theorems connected with them