THE FUNDAMENTAL LIMIT THEOREMS IN PROBABILITY

W. FELLER

1. **Introduction.** The main purpose of this address is to explain the mathematical content and meaning of the two most important limit theorems in the modern theory of probability: the central limit theorem[1] and the recently discovered precise form of what was generally known as "Kolmogoroff's celebrated law of the iterated logarithm." The former traces its origin to the very beginnings of the theory of probability and is often called after Laplace and Ljapunov. For a long time it was clouded in mystery, and Poincaré once remarked that mathematicians regard it as a physical law, whereas physicists hold mathematicians responsible for it. A great many mathematicians have contributed to the gradual recognition of the mathematical content of the theorem and to the establishment of the precise conditions of its validity. The complete solution came finally in 1935 and was possible only by an elimination of all classical restrictions and a reconsideration of the problem in a new generality.

The central limit theorem (like its little brother, the weak law of large numbers) is a statement on distribution functions, and can be formulated, either as such or in terms of Fourier analysis, without any appeal to probability or measure. This is not true of the infinitely more delicate law of the iterated logarithm and its generalizations (or of the strong law of large numbers): these are essentially measure-theoretic. The starting point of the long series of papers which lead to the present form of the iterated logarithm was not a problem in probability but, surprisingly enough, a problem in Diophantine approximations treated by Hardy and Littlewood [1914].[2] Their original estimate has gradually been improved for their particular number-theoretical case and, as a matter of fact, even the precise form of the iterated logarithm has first been checked for this particular case. It is therefore instructive to realize that, from the point of view of the general theory, the Hardy-Littlewood problem constitutes an exceedingly special case comparable only to the role of the linear function within the domain of all real functions. Such special

cases are in many respects misleading, and usually do not lend them-
selves for generalizations. Thus in its number-theoretical application
our problem reduces to an evaluation of certain sequences of binomial
coefficients, and such special techniques are not applicable even to
the most trivial generalization. The history of probability shows that
our problems must be treated in their greatest generality: only in this
way can we hope to discover the most natural tools and to open
channels for new progress. This remark leads naturally to that char-
acteristic of our theory which makes it attractive beyond its impor-
tance for various applications: a combination of an amazing generality
with algebraic precision.

The analytical formulation of our limit theorems seems unfortu-
nately to obscure the fact that a great many individual problems can
be treated as special cases. This fact seems little appreciated and
often an unnecessary effort is spent on treating such problems. A few
illustrative mathematical applications will be found in §4. Better
examples are furnished by physical applications, but it would be too
time consuming to explain them. The applicability of the central
limit theorem to problems in number theory has been amply demon-
strated in papers by Erdös, Hartman, Kac, Wintner, and others.

Another point to be stressed concerns the abundance of open
problems. The fact that we now have necessary and sufficient condi-
tions both for the central limit theorem and the iterated logarithm,
and that we are in a position to make a series of statements of the
"best result" type, seems to have created the impression that "noth-
ing remains to be done." Actually we have just succeeded in pro-
ducing good working tools and in opening the gate to a multitude of
new problems both of theoretical interest and of practical impor-
tance. (This is true even for the classical field of so-called independent
variables. The much wider domain of dependent variables, excepting
only the theory of Markov chains, remains practically untouched
despite the excellent pioneer work by P. Lévy and S. Bernstein.)

It must be understood that the following exposition is concerned
only with *one* aspect of the limit theorems and is not intended as a
survey of modern tendencies in probability. This theory has devel-
oped rapidly (thanks in particular to the famous Moscow School in
probability) and many new channels have been opened which link
the theory to many branches of mathematics. Thus the true role of
the Gaussian distribution can be understood only in connection with
stochastic processes. The foundations of this new branch of probabil-
ity have been laid in a well known paper by Kolmogoroff [1931]. It
leads to partial integrodifferential equations of a special kind, but it

throws new light even on the classical equation of diffusion and puts new interesting problems concerning this and other parabolic equations [Fortet 1941; Feller 1936]. More generally, this theory seems to lead to a new type of functional equations which has not yet been investigated. Another aspect of the Gaussian distribution leads to the modern theory of infinitely divisible laws [Gnedenko, Khintchine, Kolmogoroff, P. Lévy] and to the so-called arithmetic of distribution functions, inaugurated by P. Lévy [Cramér, Khintchine, Raikov]. A third approach is that from the classical time series problem or, in modern language, from the measure theory in functional spaces [Wiener, Doob]: this approach would lead to the theory of random noises which now occupies so many minds [cf. Doob, 1944]. In order not to get lost in a jungle of general remarks we shall have to restrict the considerations to the well defined case of *independent variables*; we shall not even pause to consider S. Bernstein's well known generalization of the central limit theorem to certain classes of dependent variables or the important application, due to Kolmogoroff and W. Doblin, of that theorem to a more precise study of the ergodic properties of Markov chains.

2. **Random variables associated with the dyadic case.** For the convenience of the uninitiated reader we shall explain the modern terminology and the type of our problems on the trivial special case of random variables associated with "spinning a coin." It will be seen that this case is still considerably more general than the number-theoretical case to which we have alluded before.

It is simplest to consider only infinite sequences of tossings of a coin: each trial results in a symbol H (head) or T (tail), and the sequence of trials will be represented by an infinite sequence like $HHHTHT \cdots$. The aggregate of all such sequences (all thinkable results of our "experiment") forms the *label space* $\mathfrak{S}$ and each sequence is called a *point*. The mapping $H\rightarrow 1$, $T\rightarrow 0$ makes to each point of $\mathfrak{S}$ correspond a dyadic fraction like $.111010 \cdots$, that is to say, a number x $(0 \leqq x \leqq 1)$ in its dyadic representation; the label space $\mathfrak{S}$ is in this way mapped onto the unit interval, which will also be denoted by $\mathfrak{S}$. It is true that the mapping is not unique for numbers like $.011111 \cdots$ which contain only a finite number of zeros or of ones; but this ambiguity will be seen to be of no consequence. In the usual way we shall associate with the symbols H and T probabilities 1/2 each, which is equivalent to saying that we introduce the ordinary Lebesgue measure on the unit interval. In this manner the latter becomes the analytical model of a real "experiment"; every

picturesque expression concerning tossings of a coin is automatically translated into a statement concerning certain subsets of the unit interval: and the words "event" and "probability" become synonymous with "set" and "measure," respectively. The mathematician can from here on forget all about coins, while the classical probability student would from the beginning refuse to translate his picturesque statements into equivalent statements referring to the unit interval.

Consider now "a player who at the kth trial receives or loses an amount a_k, depending on whether the kth trial results in H or T." The kth trial stands for the kth dyadic digit of the number x representing our particular sequence of trials. Our picturesque description therefore associates with every x, given by the dyadic expansion

$$(1) \qquad\qquad x = \sum_{k=1}^{\infty} \frac{\epsilon_k}{2^k} \qquad\qquad (\epsilon_k = 0 \text{ or } 1),$$

a sequence of functions $X_k(x)$ defined by

$$(2) \qquad\qquad X_k(x) = \begin{cases} a_k \\ -a_k \end{cases} \qquad \text{if} \qquad \begin{aligned} \epsilon_k &= 1, \\ \epsilon_k &= 0. \end{aligned}$$

Accordingly, $X_k(x)$ is a real function defined in $\mathfrak{S}$ assuming only two values, each with probability $1/2$ (on a set of measure $1/2$). The word *random variable* is synonymous with "measurable real function defined in the label space." Having defined the individual gains $X_k(x)$, "the total gain in n trials" is a new random variable given by

$$(3) \qquad\qquad S_n(x) = \sum_{k=1}^{n} X_k(x).$$

The number-theoretical case is included herein as the simplest special case

$$(4) \qquad\qquad a_k \equiv 1.$$

It concerns a player who loses or wins always the same amount, and can also be interpreted physically as a random walk in a one-dimensional lattice. Of course, $S_n(x)$ is simply the excess (positive or negative) of the number of occurrences of the digit 1 over the number of occurrences of the digit 0 among the first n digits in the dyadic expansion of x.

We shall put

$$(5) \qquad\qquad s_n^2 = \sum_{k=1}^{n} a_k^2.$$

In the sequel we shall assume that $\{a_k\}$ is an arbitrarily prescribed sequence such that

$$(6) \qquad\qquad s_n \to \infty \quad \text{and} \quad a_n = o(s_n);$$

if one of the conditions (6) is dropped, definite conclusions can still be stated, but the considerations become rather trivial.

The classical or Laplace's problem may be formulated as follows. Let n be large, but fixed, and consider $S_n(x)$ as a function of x (which means essentially that we compare the total gain after n trials for all thinkable results of the experiment). The problem is to determine the distribution of values of $S_n(x)$. Now the central limit theorem states in our special case that asymptotically for all real ξ and η $(\xi < \eta)$

$$(7) \qquad\qquad \Pr\left\{\xi s_n < S_n(x) < \eta s_n\right\} \sim \Phi(\eta) - \Phi(\xi);$$

here the left side stands for the measure of those x for which the inequality within the braces is satisfied, and $\Phi(\xi)$ for the Gaussian distribution

$$(8) \qquad\qquad \Phi(\xi) = \int_{-\infty}^{\xi} \exp\left(-\frac{v^2}{2}\right) dv.$$

The analytic form of (8) is here of no interest: what matters is that for the "reduced" variable

$$(9) \qquad\qquad S_n^*(x) = S_n(x)/s_n$$

the measure of those x for which $S_n(x)$ lies in the interval (ξ, η) is, asymptotically, independent of the structure of the particular sequence $\{a_k\}$. We mention only in passim that much more precise estimates of the asymptotic behavior (7) are available; they will be discussed together with the central limit theorem for arbitrary random variables in arbitrary spaces. The importance of the relation (7) for many applications is so well known that it needs no amplification. Less known is that the central limit theorem is sometimes wrongly used in situations where it can not be applied. To this category belong all cases of so-called optional stopping, where the number n of trials is not a constant but itself a random variable. Thus a player (or the subject in psychological card-guessing experiments) does not necessarily decide in advance on the number of trials but will stop at an opportune moment. He embarks on *one* (potentially infinite) sequence of trials which is represented by *one* number x. He is not interested in comparing his gain after n trials with other sequences but rather in the fluctuation of his gain $S_n(x)$ for one particular point x and as a function of n.

The central limit theorem (7) teaches us only that, on the average, $S_n(x)$ will be of the order of magnitude of s_n. However, it leaves theoretically open the possibility that for every x the gain $S_n(x)$ will *occasionally* reach the magnitude s_n^{17}, and in many cases it is only such an occasional maximum that actually counts.[3] The *maxima* of $S_n(x)$ will indeed for almost all x be larger than the probable values given by (7). The *Khintchine-Kolmogoroff law of the iterated logarithm states in our case that with probability* 1 *(for almost all x)*

$$(10) \qquad \limsup_{n \to \infty} \frac{S_n(x)}{s_n \{ 2 \log \log s_n \}^{1/2}} = 1,$$

provided only that

$$(11) \qquad a_n = o(s_n \{ \log \log s_n \}^{-1/2}).$$

To explain the meaning of (10) consider the Hardy-Littlewood case (4) in which $S_n(x)$ is the excess of digits one over digits zero among the n first digits in the dyadic expansion (2). Condition (11) is here trivially satisfied, and the law of the iterated logarithm states in that particular case that for every positive ϵ and almost all x the following statements hold: (i) There are infinitely many n such that $S_n(x) > \{ (2 - \epsilon)n \log \log n \}^{1/2}$; (ii) for all n sufficiently large $S_n(x) < \{ (2+\epsilon)n \log \log n \}^{1/2}$. This is Khintchine's [1924] refinement of Borel's well known theorem that *"almost all numbers are normal,"* which in our notation means that $S_n(x) = o(n)$ for almost all x. Many intermediary steps have led from Borel's theorem to Khintchine's result. Hausdorff [1913] proved that $S_n(x) = o(n^{1/2+\epsilon})$, Hardy-Littlewood [1914] that $S_n(x) = O((n \log n)^{1/2})$, Steinhaus [1922] that $\limsup S_n(x)/(2n \log n)^{1/2} \leqq 1$ (note the log n instead of the iterated logarithm); finally Khintchine himself [1923] had proved that $S_n(x) = O((n \log \log n)^{1/2})$. It is well to remember the tremendous computational effort which was necessary for the investigation of such a simple special case: only against this background can one fully appreciate the strength and value of the general arguments which permitted Kolmogoroff [1929] to prove the law of the iterated logarithm (10) not only for all sequences $\{a_n\}$ but for perfectly arbi-

---

[3] In the theory of diffusion the central limit theorem corresponds to the statement that the random position of the particle is subject to a Gaussian distribution with variance propertional to the time parameter t. It leaves unanswered questions of the following type. What is the probability that a particle under diffusion, starting at $t = 0$ from $x = 0$, will forever remain within the domain, say, $|x| < t$? The significance of this, and similar more refined problems, stands to reason; they are of the category related to the law of the iterated logarithm.

trary random variables in arbitrary spaces (cf. §8). Moreover, the computational part of the argument in the general case is considerably simpler than in the special number-theoretical case.

Looking back at the modest beginnings of the theory, Kolmogoroff's result (10) would seem as complete as one could desire. That it nevertheless is not the final word was first made clear by an exciting discovery due to Marcinkiewicz and Zygmund [1937]. They constructed an example showing that the law of the iterated logarithm (10) does not necessarily hold if the condition (11) is replaced by the only slightly weaker hypothesis

$$(12) \qquad\qquad a_n < \epsilon s_n \{\log \log s_n\}^{-1/2}$$

with ϵ an arbitrarily small constant. To make things more puzzling, the equality sign in (10) is, in the Marcinkiewicz-Zygmund example, replaced by the sign "smaller than," contrary to all expectations. To understand the inner mechanism of the phenomenon we must embark on the more ambitious undertaking and investigate, not only the upper limit (10), but also the manner in which it is approached. Again, for a real understanding and in order to find natural tools, such a problem must be considered in its greatest generality and not only for our very special random variables. However, we shall here describe the results for our particular case (but for arbitrary $\{a_n\}$).

Our results can best be described by means of a convenient terminology due to P. Lévy: *A nondecreasing sequence $\{\phi_n\}$ of positive numbers will be said to belong to the lower class, or in symbols*

$$(13) \qquad\qquad \{\phi_n\} \in \mathcal{L},$$

if, for almost all x (that is, with probability 1), *there exist infinitely many n such that*

$$(14) \qquad\qquad S_n(x) > s_n\phi_n;$$

the sequence $\{\phi_n\}$ belongs to the upper class $(\in \mathcal{U})$, if for almost all x there exist at most finitely many n such that (14) *holds.* Every sequence $\{\phi_n\}$ belongs either to $\mathcal{L}$ or to $\mathcal{U}$.[4] With this notation Kolmogoroff's result (10) states that

$$(15) \qquad \phi_n = \{(2 + \epsilon) \log \log s_n\}^{1/2} \begin{matrix} \in \mathcal{L} & \text{if } \epsilon < 0, \\ \in \mathcal{U} & \text{if } \epsilon > 0. \end{matrix}$$

Thus here the gap between the two classes is of the same order of

---

[4] This is a special case of the "null-or-one-law." It will be noticed that the statement is by no means obvious: a priori one might expect that a sequence $\{\phi_n\}$ could satisfy the criterion for each class on a set of measure 1/2.

magnitude as ϕ_n. We shall not pause to describe certain improvements for special cases which are due to P. Lévy [4, 1931], Cantelli [1933] and Cramér [1934], but pass to the illustration of the complete result. To begin with, we shall replace (11) by the slightly stronger condition

$$(16) \qquad a_n = O(s_n \{\log \log s_n\}^{-3/2}).$$

Of course, this condition is certainly satisfied if the a_k remain bounded or increase slowly; it holds in particular for the number-theoretical case (4). Now the following *criterion holds* (Feller [7, 1943]): *if the constants $\{a_n\}$ satisfy* (16), *then the necessary and sufficient condition that $\{\phi_n\} \in \mathcal{U} \ (\mathcal{L})$ is that*[5]

$$(17) \qquad \sum \frac{a_n^2}{s_n^2} \phi_n \exp\{-\phi_n^2/2\} \in \mathcal{C}(\mathcal{D}).$$

The law of the iterated logarithm is, of course, contained in this criterion and follows from (17) and the Abel-Dini theorem on infinite series. More generally the latter theorem and the conventional logarithmic scales show that *the sequence*

$$(18) \qquad \phi_n = \{2 \log_2 s_n + 3 \log_3 s_n + 2 \log_4 s_n + \cdots$$
$$+ 2 \log_{r-1} s_n + (2 + \delta) \log_r s_n\}^{1/2}$$

belongs to $\mathcal{L} \ (\mathcal{U})$ *if, and only if,* $\delta \leq 0 \ (\delta > 0)$.

In the special case (4) we have $s_n^2 = n$, and (17) reduces to

$$(19) \qquad \sum \frac{\phi_n}{n} \exp\{-\phi_n^2/2\} \in \mathcal{C} \ (\mathcal{D}).$$

The special result (19) (in an equivalent integral form) has been stated by Kolmogoroff (communicated without proof in P. Lévy's book of 1937) and confirmed by Erdös [1942]. The most general conditions under which a similar result holds will be indicated later.

Several interesting corollaries can be deduced from (17). Thus it follows that *for any constant M the sequence*

$$(20) \qquad \phi_n^* = \phi_n + M/\phi_n$$

belongs to the same class as $\{\phi_n\}$. This is in a certain sense a "best" result and holds also in the general case [Feller, 1943]. Moreover, *if $\{\phi_n\} \in \mathcal{L}$ then there are* (for almost all x and every positive δ) *infinitely many n such that*

---

[5] Here and in the following $\mathcal{C}$ and $\mathcal{D}$ stand for "converges" and "diverges," respectively.

(21) $$\phi_n < S_n(x)/s_n < \phi_n + \delta/\phi_n.$$

The criterion (17) is only a special case of a more general theorem. The next simplest case arises when the condition (16) is replaced by the weaker one

(22) $$a_n = O(s_n\{\log\log s_n\}^{5/6}).$$

If (22) holds, we shall have $\{\phi_n\} \in \mathcal{U}\ (\mathcal{L})$ if, and only if,

(23) $$\sum \frac{a_n^2}{s_n^2} \phi_n \exp\left\{-\phi_n^2/2 + M_4\phi_n^4/12\right\}$$

where

(24) $$M_4 = \sum_{k=1}^{n} a_k^4.$$

It is readily seen that under the stronger condition (16) the second term in the exponent in (23) has no influence on the convergence of the series, so that (17) is, in the strict sense, a special case of (23). We can now proceed to relax the hypothesis of our criterion. If in (22) the exponent 5/6 is replaced by 7/10, an additional term containing ϕ_n^6 will appear in the exponent of (23), and its coefficient will depend on $\sum_{k=1}^{n} a_k^6$. Generally, if the exponent in (22) is replaced by $(2r+1)/(4r-2)$, the exponent in the criterion will contain r terms and be a polynomial of degree $2r$ in ϕ_n. Letting $r \to \infty$ we obtain the following final form of our criterion, which contains all others as a special case:

Define a number $\zeta = \zeta(n)$ by the identity

(25) $$\phi_n = s_n^{-1} \sum_{k=1}^{n} a_k \tanh(a_k\zeta).$$

There exists a numerical constant $\eta > 1/100$ such that for all sequences satisfying the condition

(26) $$a_n < \eta \frac{s_n}{\{\log\log s_n\}^{1/2}}$$

the criterion holds: $\{\phi_n\} \in \mathcal{U}\ (\mathcal{L})$ if, and only if,

(27) $$\sum \frac{a_n^2}{s_n^2} \phi_n \exp\left\{-\zeta\sum_{k=1}^{n} a_k \tanh(a_k\zeta) - \sum_{k=1}^{n} \log \operatorname{ch}(a_k\zeta)\right\}.$$

In this general form the criterion is somewhat unhandy, but it

explains completely the breaking down of the law of the iterated logarithm when the condition (11) is not satisfied.

3. **Some unsolved problems.** What happens when not even (26) is satisfied? The answer is unknown. Admittedly the question is somewhat artificial, since all "reasonable" sequences will satisfy even the strongest of our conditions, namely (16). However, considered as a special case of the general theory, the answer to the question would be of interest, for it would lead into new and unexplored domains. From special cases treated by P. Lévy [1931], Marcinkiewicz [1939], Hartman [1941], and Feller (unpublished) we know that the asymptotic behavior of sums of independent random variables changes its nature completely when (26) (or its analog in the general case) is dropped. Our present tools are not sufficient to treat such cases. Natural methods would probably be applicable also to quite different problems linking the theory of probability to certain unexplored functional equations.

Our theorems give precise theoretical information concerning the probable *amplitude* of the oscillations of $S_n(x)$ as a function of n. It would be of considerable theoretical and practical interest to have more information as to the *frequency* or wave-length of these oscillations. What can be said concerning the frequency with which $S_n(x)$ changes sign? Many similar questions can be raised, but again very little is known in this direction. (However, in the special case (4) some interesting results were recently obtained by Erdös; they are not yet published.) These questions are related to the iterated logarithm, that is to say, they are of the measure-theoretic or "strong" type. However, there are many open questions of the "weak" type, which are really problems concerning distribution functions and can be formulated in terms of Fourier analysis.

The central limit theorem gives us information concerning the average of $S_n(x)$ for a fixed n. This information is useful, but in many practical cases (for example, in all cases with "optional stopping") we are interested not so much in $S_n(x)$ as in the function

$$(28) \qquad\qquad S_n^*(x) \;=\; \max_{k=1,\cdots,n}\; S_k(x).$$

It would be of great theoretical importance to have a theorem analogous to the central limit theorem and relative to $S_n^*(x)$. Once such a theorem is obtained, one will proceed to obtain estimates of the asymptotic error and statistical tests of significance of the same type as are now available for $S_n(x)$. These remarks apply to the general

theory of random variables even more than to the present special case.[6]

In the special case (4) the problem can be easily solved but, as frequently happens, the solution is of little use in the general theory. For the particular case (4) our problem is essentially equivalent to the classical problem of ruin, or physically, to the problem of the symmetric restricted random walk in one dimension with one absorbing wall. The simplest method here seems to be that of difference equations, which also permits us to treat the more interesting problem of a random walk with two absorbing walls, say at $k=0$ and $k=r$. The problem arises classically in connection with a player who wins or looses in each trial a unit amount and starts with a capital less than r: the game ends when the player has either lost his capital ($S_n = 0$) or if he succeeds in increasing it to r. Analytically one is led to a difference equation corresponding to the differential equation of diffusion

$$(29) \qquad\qquad u_t = u_{xx}$$

with a boundary value problem $u(0, t) = u(r, t) = 0$ and appropriate initial values. As in the case of the differential equation, our problem can be solved by two methods (method of images or trigonometric interpolation). Identifying the two solutions leads to curious identities between certain sums of binomial coefficients on one hand and trigonometric polynomials on the other. Passing to the limit one obtains the familiar transformation formula for theta functions which is usually proved operating in the same way on the equation of diffusion.[7] In the general case the difference equations are of no use, but Petrowski [1934] (cf. Khintchine's [1933] exposition) has shown that it is possible at least to obtain some limit theorems by the direct use of differential equations.

4. **The general notion of random variables.** In the simple example of §2, there were only two possible results of each individual trial, and therefore the random variables X_k could assume only two distinct values. In the general theory we shall consider perfectly arbitrary spaces. We consider first some empirical examples: throwing dice leads to a label space consisting of six points; in the case of the roulette each trial will result in a certain angle, and the label space consists of an interval $0 \leq \psi < 2\pi$; in the theory of diffusion the result

---

[6] For recent results connected with this problem, cf. Wald [1944].

[7] These facts are doubtless known in the classical literature. Related arguments have been used by P. Lévy [1931].

of an observation is, theoretically, a point, and the label space con-
sists of a portion of the ordinary space; in statistical mechanics each
observation is represented by a point in the phase-space, and the
latter is the label space. Frequently a certain function will be asso-
ciated with the points of the label space: in the case of games this
function may represent the gain associated with the possible results
of the experiment; in the theory of diffusion it may be the distance of
the particle from the origin; in statistical mechanics it may be the
kinetic energy or the entropy. This is the picturesque empirical back-
ground for our abstract definitions; purely mathematical illustrations
will be supplied in the next section.

A *label space will in the sequel mean an arbitrary space in which a
probability measure is defined* (that is to say an absolutely additive,
non-negative measure such that the label space itself has measure 1).
*A random variable is a real-valued measurable function in the label
space.* How the probability measure has been obtained (or defined) is
of no interest for our present purposes.

We have now to define the fundamental notion of *independent
random variables.* For that purpose we can again use the language of
"repeated trials," although the notion can actually be defined in a
more general way. As in §2 we shall consider an infinite sequence of
trials and suppose that to the kth trial there corresponds the label
space $\mathfrak{S}_k$. In §2 all these spaces were congruent, each consisting of
only two points. This, however, is by no means necessary, and the
simplest applications actually lead to *variable* sequences of label
spaces. The result of an infinite sequence of trials will now be repre-
sented by a sequence $P_1, P_2, \cdots$, where P_k is a point of $\mathfrak{S}_k$. Thus to
our infinitely repeated trials there corresponds a label space $\mathfrak{S}$ whose
points are represented by symbols $P = (P_1, P_2, \cdots)$ with $P_k \in \mathfrak{S}_k$.
In modern terminology, $\mathfrak{S}$ is simply the combinatorial product of the
spaces $\mathfrak{S}_k$,

$$(30) \qquad \mathfrak{S} = \mathfrak{S}_1 \times \mathfrak{S}_2 \times \mathfrak{S}_3 \times \cdots .$$

Since in each $\mathfrak{S}_k$ a probability-measure has been defined, we can de-
fine probabilities in the product space $\mathfrak{S}$ in the usual way by taking
the product measures. This definition of the product measure is the
mathematical equivalent of the empirical notion of *independent
trials.* We shall from now on suppose that the $\mathfrak{S}_k$ are arbitrary, and
that in $\mathfrak{S}$ (given by (30)) the product measure has been defined.
Every random variable $X_k(P_k)$ defined in $\mathfrak{S}_k$ then automatically
becomes a random variable in $\mathfrak{S}$; any such sequence of random vari-
ables will be called a sequence of *mutually independent random vari-*

ables. We shall again be concerned in particular with sums

(31) $$S_n(P) = \sum_{k=1}^{n} X_k(P_k).$$

The consideration of infinite product spaces (30) is in many cases most natural, but it means an unnecessary restriction. For example, the consideration of infinite product spaces led us in §2 to the unit interval in its dyadic representation. Clearly such a representation will not always be desirable, and it may be more convenient for many purposes to consider the unit interval itself as the directly given label space. A closer survey shows that the essential feature of statistical independence consists in a certain property of multiplicativity, and this consideration leads us to the following general definition (which contains the construction in product spaces as a special case). *A sequence of random variables X_k defined in an arbitrary label space $\mathfrak{S}$ will be called mutually independent, if for any choice of N, b_i and c_i*

(32) $$\Pr\left\{b_k < X_k < c_k; k = 1, 2, \cdots, N\right\} = \prod_{k=1}^{N} \Pr\left\{b_k < X_k < c_k\right\};$$

here the letters Pr denote the measure of the set in $\mathfrak{S}$ in which the inequalities within the braces are satisfied.

The *distribution function* $V_k(x)$ of X_k is defined by

(33) $$V_k(x) = \Pr\left\{X_k \leq x\right\}.$$

(In the case where $X_k = f(t)$ is a function defined on the unit interval, the inverse of the distribution function can be interpreted as a reordering of the values of $f(t)$ with preserved measure; in this way it has been used in real variable theory by Hardy, Littlewood, and other authors.)

Now let $F_n(x)$ denote the distribution function of the random variable $S_n(x)$ defined by (31). If the X_k are mutually independent, *the distribution function $F_n(x)$ of the sum $S_n(x)$ is given by the recurrence formulas*

(34) $$F_1(x) = V_1(x), \quad F_{n+1}(x) = \int_{-\infty}^{+\infty} F_n(x - y) dV_{n+1}(y).$$

The characteristic function (Fourier-Stieltjes transform) of a distribution function $V(x)$ is

$$v(s) = \int_{-\infty}^{+\infty} e^{isx} dV(x).$$

With this definition the characteristic function of $F_n(x)$ is simply the product of the characteristic functions of $V_1(x), \cdots, V_n(x)$.

The central limit theorem, the ordinary (weak) law of large numbers, and some other limit theorems are of the *weak type*, that is to say, they describe only the asymptotic behavior of the distribution functions $F_n(x)$. They can therefore be formulated in terms of an arbitrarily given sequence $V_k(x)$ of distribution functions and the formulas (34); *the notion of random variable and measure in the label space is, for the weak theorems, irrelevant*. However, the *strong* theorems, like the law of the iterated logarithm and the strong law of large numbers, are strictly speaking measure theoretic.

Before passing to examples, let us introduce a few definitions. The *expectation* of the random variable X_k is, by definition,

$$(35) \qquad \mu_k = \int_{-\infty}^{+\infty} x dV_k(x),$$

provided the integral exists (is absolutely convergent). A simple computation shows that the expectation of the sum S_n is given by

$$(36) \qquad m_n = \mu_1 + \cdots + \mu_n.$$

The *variance* of X_k is defined by

$$(37) \qquad \sigma_k^2 = \int_{-\infty}^{+\infty} (x - \mu_k)^2 dV_k(x),$$

again provided that the integral converges. It is readily seen that the variance

$$(38) \qquad s_n^2 = \int_{-\infty}^{+\infty} (x - m_n)^2 dF_n(x)$$

of S_n is given by

$$(39) \qquad s_n^2 = \sigma_1^2 + \cdots + \sigma_n^2.$$

5. **Examples.** The following examples are intended to illustrate the notion of independent random variables and their sums by means of some of the simplest purely mathematical applications. At the same time they will show that our limit theorems can sometimes be applied with greatest ease to problems which are frequently treated in a much more complicated manner; in such cases the refinements of the limit theorems are apt to give much more precise results than other methods.

(a) *Inversions. Among the $n!$ permutations of the elements* 1, 2, $\cdots$, n, *how many will exhibit exactly r inversions?*

Here the label space $\mathfrak{S}$ consists of the $n!$ permutations; with each point of $\mathfrak{S}$ we associate probability $1/n!$. Consider now an arbitrary, but fixed, permutation P. For $k = 1, 2, \cdots , n$ we define the value of $X_k(P)$ to be the number of elements among 1, 2, $\cdots$, $k-1$ which *succeed* the element k; in other words, $X_k(P)$ is the number of inversions in P which are produced by the element k with regard to the smaller elements. The sum

$$(40) \qquad\qquad S_n(P) = X_1(P) + \cdots + X_n(P)$$

is the *total number of inversions* exhibited by P.

It is also possible to describe the situation in the terminology of independent trials; for that purpose we have to represent the label space $\mathfrak{S}$ as an n-tuple product space, which means that we have to build the permutations step by step. We first write down the element 1; for the next element, 2, we have two possibilities, namely to put it ahead or behind: accordingly, the label space $\mathfrak{S}_2$ corresponding to the second trial consists of two points, and we have to attribute to each the probability $1/2$. For the third element we have three possibilities (places), and in general, the label space $\mathfrak{S}_k$ ($k = 1, 2, \cdots , n$) will consist of k points, each having probability $1/k$. The space $\mathfrak{S}$ of all permutations is clearly the combinatorial product of these $\mathfrak{S}_k$, and the original measure ($1/n!$ for each point) is the corresponding product measure. It is also seen that the random variable $X_k(P)$, the number of inversions produced by k, is defined in $\mathfrak{S}_k$ without regard to the other trials. Our X_k are therefore independent random variables, but this could also be checked directly in $\mathfrak{S}$ using the definition (32).

According to construction, X_k assumes the values 0, 1, $\cdots$, $k-1$, each with probability $1/k$. Therefore (cf. (35)–(39))

$$(41) \qquad\qquad \mu_k = (k-1)/2, \qquad \sigma_k^2 = (k^2-1)/12$$

whence

$$(42) \quad m_n = n(n-1)/4 \sim n^2/4, \quad s_n^2 = (2n^3 + 3n^2 - 5n)/72 \sim n^3/36.$$

The central limit theorem (cf. the next section) shows *that the average number of inversions is m_n; that the number of permutations for which the number r of inversions satisfies*

$$(43) \qquad\qquad (n^2/4) + (x_1/6)n^{3/2} < r < (n^2/4) + (x_2/6)n^{3/2}$$

is asymptotically given by n! multiplied by $\Phi(x_2) - \Phi(x_1)$ (cf. (8)). Incidentally, known estimates of the asymptotic error in the central limit theorem permit us to give also more precise estimates for (43). Here we wish only to point out that the result is a trivial consequence of the central limit theorem.

(b) *Cycles. How many permutations of the elements* 1, 2, $\cdots$, n *exhibit exactly r cycles?*

In a self-explanatory way we begin to write the permutations in the form $1 \rightarrow e_1 \rightarrow e_2 \rightarrow \cdots$; the first cycle is completed when $e_s = 1$; then we start with the smallest of the remaining elements and continue in the same way. Thus the permutation $1 \rightarrow 3 \rightarrow 4 \rightarrow 1$; $2 \rightarrow 2$; $5 \rightarrow 7 \rightarrow 8 \rightarrow 6 \rightarrow 5$ is a permutation of the elements 1, $\cdots$, 8 with three cycles. Now for any permutation P we define $X_k(P)$ to equal 1 if a cycle is completed at the kth place, and $X_k(P) = 0$ otherwise $(k = 1, \cdots, n)$. Again the $X_k(P)$ form a set of mutually independent random variables, and their sum (40) now gives the total number of cycles in P. From the way in which the permutations are built it follows trivially that

$$(44) \qquad \begin{aligned} \Pr \{X_k = 1\} &= 1/(n - k + 1), \\ \Pr \{X_k = 0\} &= (n - k)/(n - k + 1), \end{aligned}$$

so that

$$(45) \qquad \mu_k = 1/(n - k + 1), \qquad \sigma_k^2 = (n - k)/(n - k + 1)^2$$

and

$$(46) \qquad \begin{aligned} m_n &= 1 + 1/2 + 1/3 + \cdots + 1/n \sim \log n, \\ s_n^2 &= \sum_{k=1}^{n} \frac{n - k}{(n - k + 1)^2} \sim \log n. \end{aligned}$$

Therefore, according to the central limit theorem, *the average number of cycles is* m_n; *asymptotically, the number of permutations with r cycles, where*

$$(47) \qquad \log n + x_1(\log n)^{1/2} < r < \log n + x_2(\log n)^{1/2},$$

tends to $\Phi(x_2) - \Phi(x_1)$. Again more precise estimates are readily available.

If, instead of considering *all* cycles, only cycles of a prescribed length are taken into account, the method does not apply without modification: it is then necessary to introduce certain dependent random variables.

(c) *Runs.* We shall now consider what is known as Bernoulli trials. The label space is the same as in §2, that is to say each point is represented by an infinite sequence of symbols H or T; only this time we shall associate with H an arbitrary probability p (not necessarily $1/2$), and with T the complementary probability $q = 1 - p$. For example, we may consider the ordinary decimal representation of the unit interval with its natural measure, calling the digit zero H; then $p = 1/10$. By definition, *an H-run of length r is an uninterrupted sequence of at least r symbols H*. A run of length $r+1$ is then automatically a run of length r; and we shall agree to say that a run of length $2r, 3r, \cdots$ contains $2, 3, \cdots$ runs of length r. This is not strictly the usual nomenclature, but it is for our purposes by far the most convenient one. Moreover, the numerical differences arising from the change in the usual definition are perfectly negligible. Also, it requires only trivial modifications to adapt our developments to the standard definitions.

For any point P of the label space let $X_1(P)$ denote the number of trials before the completion of the first run of length r. Then X_1 is a random variable which may assume any of the values $r, r+1, \cdots$ (the probability that no run of length r occurs is zero). The distribution function $V_1(x)$ of X_1 is well known.[8] Next let $X_2(P)$ denote the number of trials from the completion of the first run to that of the second run. Clearly X_2 is a random variable which is independent of X_1 but has the same distribution function. Proceeding in the same way, we may define n independent random variables $X_1, \cdots, X_n$, all having the same distribution function and such that their sum S_n (cf. (40)) will give the number of trials from the beginning to the completion of the nth run.

The usual problem (gaining new importance in applied statistics) is to find *the probability of having in N trials k runs or more* (each of length r). In our notations this probability is

$$(48) \qquad\qquad \Pr\left\{ S_k \leqq N \right\},$$

and the central limit theorem permits to evaluate this probability in the most trivial manner and with all desirable acuracy. Thus we have obtained, as a perfectly trivial corollary, the famous theorem to the effect that *the number of runs is approximately normally distributed.*[9]

(d) *Partial sums of the exponential series.* This example is, even

---

[8] Cf., for example, Uspensky [1937].

[9] More interesting and much deeper results concerning the asymptotic distribution of runs have been obtained by Levene and Wolfowitz [1944].

formally, independent of the notion of random variable. Let $a > 0$ be a constant, and define a distribution function $V(x; a)$ depending on the parameter a by

$$(49) \qquad V(x; a) = e^{-a} \sum_{k=0}^{[x]} \frac{a^k}{k!}.$$

This is the familiar *Poisson distribution*. Let, for every n, $V_n(x; a) = V(x; a)$ and define new distribution functions $F_n(x; a)$ by (34). A trivial computation shows that $F_n(x; a) = V(x; na)$. Accordingly, by the central limit theorem,

$$(50) \qquad F_n(x; a) \sim \Phi\left(\frac{x - na}{(na)^{1/2}}\right).$$

With a slight change in notation this result can be restated by saying that, *for large values of* λ,

$$(51) \qquad e^{-\lambda} \sum_{[\lambda + t_1 \lambda^{1/2}]}^{[\lambda + t_2 \lambda^{1/2}]} \frac{\lambda^k}{k!} \sim \Phi(t_2) - \Phi(t_1).$$

Again, known asymptotic estimates for the central limit theorem permit us to give, without computation, more precise estimates (cf. Kac [1942]).

6. **The central limit theorem.** Returning to the general theory we shall now consider an arbitrary sequence of mutually independent random variables X_k and their partial sums

$$(52) \qquad S_n = X_1 + \cdots + X_n.$$

Using the notations defined in (36), (39) and (8), we may say that *the problem of the central limit theorem in its classical version is to establish the conditions under which the "reduced" random variable*

$$(53) \qquad S_n^* = (S_n - m_n)/s_n$$

will, in the limit, be normally distributed,[10] by which is meant that

$$(54) \qquad \Pr\{\xi < S_n^* < \eta\} \to \Phi(\eta) - \Phi(\xi).$$

Needless to say that the purely analytical content of the problem has not been understood from the beginning. For more than one hundred years a great many mathematicians have been working on the problem discovering many special cases to which the theorem applies and gradually establishing, and relaxing step for step, sufficient

---

[10] This formulation presupposes the existence of second moments (34).

conditions under which the theorem holds. To less critical minds the law appeared as a universal law or, occasionally, as a law of nature. The first special case where the law *does not* apply has been discovered by Cauchy, but less than ten years ago a respectable mathematical journal contained a proof that the central limit theorem applies without restrictions. Great analytical progress has been made by Ljapunov, but the modern era in the theory may be said to date from the discovery by the Finnish mathematician Lindeberg [1922] that *the central limit theorem certainly holds if $s_n \to \infty$ and if for every* $\epsilon > 0$

$$(55) \qquad \lim_{n \to \infty} s_n^{-2} \sum_{k=1}^{n} \int_{|x-\mu_k| > \epsilon s_n} (x - \mu_k)^2 dV_k(x) = 0.$$

This is the famous *Lindeberg condition*; it is of importance not only because it is of remarkable generality, but still more because it turned out to be a useful tool for many purposes. For example, it has been used by Kolmogoroff in the theory of stochastic processes [3, 1931]. As for its generality it may be stated that, somewhat surprisingly, it turned out later (Feller, [1935])[11] that *Lindeberg's condition* (55) *is not only sufficient, but also necessary for the validity of the central limit theorem in its classical version* (which was the only one considered at that time).

At first sight this theorem would appear to state that Lindeberg's condition completely solves the problem. This, however, is far from being true. Why are the random variables S_n in (53) normed just by means of the numerical sequences $\{m_n\}$ and $\{s_n\}$? This is a tradition which goes back to pre-Laplacian times, and has gradually taken root so firmly that it has tacitly been looked upon as the only possibility. In reality this convention has done much harm. Admittedly it is natural and useful in most standard applications. Nevertheless its indiscriminate use has obscured the true content of the laws of large numbers. Worse than that, the failure to understand the arbitrariness of the special norming (53), and of the use of moments in general, has lead to many misunderstandings and to lamentable useless discussions (like the endless controversy connected with the so-called St. Petersbourg paradox). For real progress it is necessary to develop the most natural tools, which in turn is possible only by finding the most general formulation of the problem and freeing it from all artificial restrictions

Accordingly, we shall say that *a sequence* $\{V_k(x)\}$ *of distribution*

---

[11] The proof will be found also in Cramér's booklet [1937].

*functions obeys the (generalized) central limit law if there exist two se-
quences of constants* $\{a_n\}$ *and* $\{b_n\}$ *such that the convolutions* $F_n(x)$
defined by (34) *satisfy the relation*

$$(56) \qquad \lim_{n\to\infty} F_n(a_n x + b_n) = \Phi(x);$$

in terms of the corresponding random variables X_k this means that
the random variable

$$(57) \qquad S_n^* = (\dot S_n - b_n)/a_n$$

is, in the limit, normally distributed (satisfies (54)). In this way we
have freed ourselves not only from the arbitrary classical norming
(53), but also from the restriction that the moments (37) should
exist: in the present formulation the definition applies to an arbitrary
distribution function. Only the case where

$$(58) \qquad a_n \to \infty, \qquad a_n/a_{n+1} \to 1$$

presents interest,[12] and to avoid trivialities and clumsy formulations,
we shall restrict our considerations to that case. The generalized cen-
tral limit problem has been completely solved by the following

THEOREM. *The sequence* $\{V_k(x)\}$ *obeys the central limit law if,
and only if, there exists a sequence* $q_n \to \infty$ *such that simultaneously*

$$(59) \qquad q_n^{-2} \sum_{k=1}^{n} \int_{|x| < q_n} x^2 dV_k(x) \to \infty, \qquad \sum_{k=1}^{n} \int_{|x| < q_n} dV_k(x) \to 0.$$

In that case one may put

$$(60) \qquad a_n^2 = \sum_{k=1}^{n} \left\{ \int_{|x| < q_n} x^2 dV_k(x) - \left\{ \int_{|x| < q_n} x dV_k(x) \right\}^2 \right\}$$

and define b_n *by*

$$(61) \qquad \int_{|x| < a_k} (x - \beta_n) dV_k(x) = 0; \qquad b_n = \sum_{k=1}^{n} \beta_k.$$

With the constants so defined (56) *will hold* (Feller [1935]).[13]

----

[12] According to a theorem of Cramér [1936], the central limit theorem can not
hold with any *bounded* sequence $\{a_n\}$ unless all the variables $\{X_n\}$ are themselves nor-
mally distributed. The case where the other relation in (58) does not hold similarly
leads to a trivial situation.

[13] Here and in the sequel it is supposed that the origin has been chosen in such a
way that $V_k(+0) \geq q$, $V_k(-0) \leq 1-q$ where q is some constant (arbitrarily small).
This will almost always be the case and represents no restriction whatsoever. If, in
particular, $q = 1/2$, the origin is the so-called median.

Whether or not a sequence $\{q_n\}$ satisfying (58) and (59) exists can be decided by means of a simple criterion, and the constants q_n, if they exist, can be computed explicitly in a trivial way.

It follows, in particular, that *for given sequences* $\{V_n(x)\}$, $\{a_n\}$, $\{b_n \equiv 0\}$ *the relation* (56) *will hold if, and only if, simultaneously*[14] *for every* $\epsilon > 0$

$$(62) \quad \lim_{n \to \infty} \sum_{k=1}^{n} \int_{|x| > \epsilon a_n} dV_k(x) = 0,$$

$$(63) \quad \lim_{n \to \infty} a_n^{-2} \sum_{k=1}^{n} \left\{ \int_{|x| < a_n} x^2 dV_k(x) - \left\{ \int_{|x| < a_n} x dV_k(x) \right\}^2 \right\} = 1,$$

$$(64) \quad \lim_{n \to \infty} a_n^{-1} \sum_{k=1}^{n} \int_{|x| < a_n} x dV_k(x) = 0.$$

(In most practical cases the quadratic term in (63) is too small to be of influence.)

As the simplest example where the generalized, but not the classical, central limit theorem applies we may consider the special case where all $V_k(x)$ are identical:

$$(65) \quad\quad\quad\quad\quad\quad V_k(x) \equiv V(x).$$

In this case the central limit theorem applies if, and only if,

$$(66) \quad\quad\quad \int_{|x| > z} dV(x) = o\left\{ z^{-2} \int_{|x| \leq z} x^2 dV(x) \right\}$$

as $z \to \infty$.[15] This criterion shows in particular that the central limit

---

[14] Cf. Feller [1935]. Alternative proofs have subsequently been given by Marcinkiewicz [1938], Gnedenko [1, 1939] and Doblin [2, 1939]. If the terms of the series in (64) are replaced by their absolute values, the quadratic terms in (63) may be omitted. The proof of the *sufficiency* of this set of (slightly stronger) conditions will be found in Cramér's booklet [1937]. Simultaneously with Feller and in more probabilistic terms, P. Lévy [6, 1935, and chap. 5 of his book of 1937] has given the following solution of the central limit problem which, in a sense, should be equivalent to the second of Feller's theorems of the text: *Si chacune des variables* X_k *est individuellement négligeable devant la dispersion de la somme* S_n, *la condition nécessaire et suffisante pour que* S_n *dépende d'une loi d'un type généralisé peu different de celui de Gauss est que le plus grand de* $|X_k|$ *soit négligeable (négligeable veut dire négligeable en probabilité, c'est-à-dire très petit, sauf dans des cas très peu probable)*. For further results cf. Raikov [2, 1938].

[15] This particular case has been discovered simultaneously by Khintchine [6, 1935], P. Lévy [6, 1935] and Feller [example (a) in 1, 1935].

theorem may apply even if the second moment does not exist (so that the classical formulation of the theorem would break down). In such cases the norms a_n^2 will increase more rapidly than n. It was the failure to observe the possibility of similar phenomena in connection with the law of large numbers which led to the discussions of the St. Petersbourg "paradox"; within the analytical theory of limit theorems the latter does not present the slightest difficulty. In the example (65) the Lindeberg condition (55) loses sense if the moments σ^2 are infinite. However, it is easy to construct examples which look perfectly classical in the sense that moments of arbitrarily high orders exist, and for which the classical central limit theorem breaks down simply because the norming (53) is unnatural and must be replaced by another one. (For examples cf. Feller, loc. cit.)

7. **Generalizations.** The central limit problem was the starting point of many investigations. To begin with the simplest, the conditions can be generalized to various cases of convergence to distributions other than the normal (Bawly [1936], Gnedenko [1, 2, 1939], Gnedenko and Groshev [1939], P. Lévy [8, 1936], Marcinkiewicz [1, 1938]); such questions are related to the nature of boundary values of analytic functions. More interesting and deeper results concern cases in which the central limit theorem does *not* hold. The classical example is furnished by the *Cauchy-distribution*

$$(67) \qquad V_k(x) = V(x) = 1/2 + \pi^{-1} \arctan x;$$

here

$$(68) \qquad F_n(nx) = V(x).$$

The Gaussian and the Cauchy-distributions are the oldest examples of a *stable distribution*, that is to say of a distribution function satisfying a functional relation of the form $V(x) * V(x) = V(ax)$,[16] where a is a constant. The systematic study of stable distributions has been initiated by Pólya [1923]; the most general form of the Fourier-transform of stable distributions has been obtained by Khintchine and P. Lévy [1936] (cf. also P. Lévy's book of 1937). More generally, Khintchine has obtained the totality of solutions of the functional equation $V(x) * V(x) = V(ax+b)$. Such distribution functions are the only ones which occur as limits of distribution functions of random variables of the form (57). The stable distributions are a subclass of a much wider and much

---

[16] The star denotes the operation of convolution, defined in (34).

more important class of distribution functions, the so-called *infinitely divisible laws*. A distribution function $F(x)$ is said to be infinitely divisible if it is the convolution of an arbitrarily large number, n, of distribution functions: $F(x) = F_1(x) * \cdots * F_r(x)$ where with increasing n the components $F_k(x)$ $(k=1, \cdots, n)$ tend to the unitary distribution (which equals one for $x > 0$ and zero for $x < 0$).[17] Roughly speaking, the infinitely divisible distributions represent integrals of random variables and are therefore intimately connected with the theory of stochastic processes. The most general form of infinitely divisible distributions has been established by Kolmogoroff [1932] and P. Lévy [1934].[18] It is unfortunately impossible to describe the ties which link the infinitely divisible laws to many other topics: partial differential and more general functional equations, the theory of semi-groups, the arithmetic of distribution functions (inaugurated by P. Lévy and studied by Khintchine, Raikov, Gnedenko and others), and so on. Here it must suffice to mention the most direct connection with the central limit theorem which is furnished by the following beautiful theorem of Khintchine [8, 1937]: In order that $V(x)$ be the limiting distribution of a *subsequence* S_{n_k} of a sequence of form (57) it is necessary and sufficient that $V(x)$ be infinitely divisible.[19]

Returning to the central limit theorem itself, we may briefly touch on the important question concerning more precise estimates of the asymptotic error involved in (54). That this question is by no means simple is seen from the great number of papers treating the accuracy of the Gaussian distribution in the trivial special case of the binomial distribution.[20] The most general and most satisfactory asymptotic estimates now available are furnished by the well known asymptotic expansion due to Cramér [1928, 1937]. Quite recently P. L. Hsu [1945] has shown that Cramér's proof can be considerably simplified by introducing a Cesàro type kernel instead of M. Riesz' singular kernels.[21] At the same time it is seen that the same method can be used to obtain asymptotic expansions in much more general cases. It follows from these theorems that the difference between the

---

[17] For an account of the theory of infinitely divisible laws cf. P. Lévy's book [1937].

[18] Alternative proofs by Khintchine [9, 1937] and Feller [5, 1937].

[19] For further results concerning subsequences cf. Doblin [1, 3, 1938–1939], Gnedenko [1, 2, 1938].

[20] For quite recent results in this special case cf. S. Bernstein [1943].

[21] The use of Cesàro kernels has actually been suggested by Berry [1941], who has used this method to obtain very precise numerical estimates for the maximum of the difference between the distribution F_n and the corresponding Gaussian distribution.

two sides in (54) is usually of the order of magnitude of $n^{-1/2}$. Clearly such a statement has practical meaning only if ξ and η are of moderate magnitude. For large values of ξ obviously $\Phi(\xi) \sim 1$, and the relation (54) expresses a triviality. Actually the case of large ξ (or of ξ's increasing with n) is of great importance in statistics; moreover, more precise estimates for this case are essential for many theoretical investigations, for example, in connection with the iterated logarithm. Several special sequences $V_k(x)$ have been investigated from this point of view. General results for the case of equal components (65) have been obtained by Cramér [1938]; they have been generalized by Feller [6, 1943], but much more remains to be done in that direction.

A quite different type of open question connected with the central limit theorem has been discussed in §3.

8. The iterated logarithm. We consider again an arbitrary infinite sequence of mutually independent random variables X_k defined in arbitrary space. For example, the space may be the unit interval, and[22]

$$(69) \qquad\qquad X_k = \text{sign} \sin(2^k \pi x);$$

in that case the distribution function $V_k(x)$ defined by (33) is a step function with jumps of magnitude $1/2$ at $x = \pm 1$. For simplicity we shall consider only individually bounded variables X_k; standard methods of truncation permit us to generalize all results, but the essence of our theorems will become clearer in the more restricted formulation. For bounded variables the moment (35) exists. Now the variable $X_k - \mu_k$ has a vanishing first moment, and therefore we can always by a simple change of notation achieve that

$$(70) \qquad\qquad \mu_k \equiv 0.$$

Accordingly, we do not lose any generality by assuming from now on that (70) holds.

As before, the study of the partial sums

$$(71) \qquad\qquad S_n = X_1 + \cdots + X_n$$

depends on the numbers σ_k and s_n defined by (37) and (39), respectively. With these notations, *Kolmogoroff's law of the iterated logarithm states in the most general case that with probability* 1

$$(72) \qquad\qquad \limsup \frac{S_n}{s_n \{2 \log \log s_n\}^{1/2}} = 1,$$

provided only that

---
[22] Cf. Rademacher [1922].

(73) $$\text{L.U.B. } |X_k| = o(s_n\{\log\log s_n\}^{-1/2}).$$

Again we can obtain more general and more precise results using the terminology of *upper and lower classes* introduced in §2. The definition given there applies without change if the numbers s_n in (14) are interpreted according to the general definition (39). To obtain the first generalization of Kolmogoroff's law we have simply to replace (16) by[23]

(74) $$\text{L.U.B. } |X_k| = O(s_n\{\log\log s_n\}^{-3/2}).$$

If the variables X_k satisfy (74), *then the necessary and sufficient condition that $\phi_n \in \mathcal{U}$ ($\mathcal{L}$) is that*

(75) $$\sum \frac{\sigma_n^2}{s_n^2} \phi_n \exp\{-\phi_n^2/2\} \in \mathcal{C}\ (\mathcal{D}).$$

In particular, if all X_k have the same distribution function (as in the case (69)), the criterion (75) assumes the form (19). It can be shown that condition (74) is the *best possible* of its type. As in §2 we can proceed to relax it step for step: the exponent in (75) will again be replaced by increasingly more complex expressions. Now the exponents encountered in §2 were *even polynomials* in ϕ_n, but this is not true in general. It remains true only in the case of symmetric variables, that is to say in the case where the function $-X_k$ has the same distribution function as X_k. In §2 we passed directly from (16) to (22). In the general case, if the exponent $3/2$ in (74) is replaced by 1, the exponent in (75) must be replaced by a certain polynomial of third degree. In the same way, if the exponent in (74) is replaced by $(m+2)/2m$, we shall have a polynomial of degree $(m+1)$ in ϕ_n figuring in the criterion analogous to (75). Passing to the limit we obtain the most general and most precise statement in the following form:

    There exists a universal contant $\eta > 1/100$ such that for all sequences $\{X_k\}$ *with*

(76) $$\text{L.U.B. } |X_k| < \eta s_n\{\log\log s_n\}^{-1/2}$$

the criterion holds: $\{\phi_n\} \in \mathcal{U}$ ($\mathcal{L}$) *if, and only if,*

(77) $$\sum \frac{\sigma_n^2}{s_n^2} \phi_n \exp \mathfrak{P}_n(\phi_n) \in \mathcal{C}\ (\mathcal{D});$$

---

[23] The iterated logarithms in conditions (74) and (76) can be replaced by other functions, and the conditions restated in a slightly more general form; the above special form has been chosen only for comparison with Kolmogoroff's condition.

here $\mathfrak{P}_n(x)$ is a power series whose mth coefficient depends only on the moments of $X_1, \cdots, X_n$ up to the order m.

It can be shown that $\mathfrak{P}_n(x)$ is majorated by a certain geometric series, and other more precise estimates are available; we refer for such further results to Feller [7, 1943].

It seems that only quite artificial sequences $\{X_n\}$ will satisfy (76) but not (74). *Thus the rather complicated criterion (77) can, for all practical purposes, be replaced by the exceedingly simple and elegant criterion (75).* The theoretical importance of the general criterion lies in the fact that it helps us understand the actual mechanism which makes so amazingly general classes of functions exhibit an asymptotic behavior in accordance with the simple scheme given by (75). The individual terms in the power series in (77) are closely related to the asymptotic error terms for the *tails* of the Gaussian in the central limit theorem. In this sense (77) reveals the increasing complexity as we approach the outer boundaries of the domain in which the central limit theorem holds.

We are naturally led to the question of what happens if (76) does not hold. This problem is not answered and presents a challenge for the very reason that it leads beyond the central limit theorem into a domain where we still lack natural tools. Its solution would automatically give necessary and sufficient conditions for the strong law of large numbers (cf. §9), an elusive problem which has been many times attacked without success. Also, our problem would apply to several interesting stochastic processes exactly as our generalization applies to ordinary diffusion. Some interesting results leading beyond (76) have been obtained by Hartman [1941] for the case where all the random variables X_k are normally distributed (not necessarily with the same variance). A very special case studied both by P. Lévy [1, 1931] and Marcinkiewicz [1939] shows that the asymptotic behavior of sequences not obeying (76) is very different from that which we have considered so far. This is also borne out by the precise criterion obtained by the writer for the case where all the X_k have the same distribution function (not yet published).

Time and space unfortunately do not permit more than brief reference to Gnedenko's recent investigations related to our problems but pertaining to the case of continuous stochastic processes (like homogeneous diffusion).

9. **The laws of large numbers.** The implications of the so-called laws of large numbers are considerably weaker than the statements of the central limit theorem or the law of the iterated logarithm, but

these weaker implications naturally hold under much more general conditions. The distinction between *weak* and *strong* laws of large numbers is the same as between the central limit theorem and the iterated logarithm: the weak laws concern only the asymptotic behavior of convolutions of distribution functions and can be formulated without appeal to random variables. The strong laws, on the contrary, are of measure-theoretical nature. In classical textbooks only the weak law (in a special form) is proved; nevertheless the strong laws are of much greater importance in statistics, games, and other applications. In many books the weak law is proved, the strong one used.

Let $\{X_k\}$ be a sequence of mutually independent random variables with distribution functions $V_k(x)$ and expectations μ_k (cf. (33) and (35)). Let again S_n denote the nth partial sum of $\{X_k\}$, and $m_n = \mu_1 + \cdots + \mu_n$ its expectation (if it exists). *In the restricted classical form one would say that the weak law of large numbers holds if for every positive ϵ*

$$(78) \qquad\qquad \Pr\left\{ \, |\, S_n - m_n\,| > \epsilon n \right\} \to 0.$$

In terms of the distribution functions (34) this means that $F_n(nx + m_n)$ tends to the unitary distribution function. Necessary and sufficient conditions for (78) to hold have been established by Kolmogoroff [1929]. It must be understood that the formulation (78) originates with the classical theory of games, where all X_k have the same distribution function. In that particular case the special norming by the factor n makes sense, since then m_n and "the total amount at stake" are proportional to n. In general, however, the factor n is perfectly arbitrary. The failure to understand this arbitrariness has led to many unnecessary discussions and "paradoxes." Also, the law of large numbers has frequently been regarded either as a law of nature or as a consequence of the definition of probability. In this sense its universal validity has been assumed and it has been applied to cases where it can easily be shown not to hold. For example, in the so-called St. Petersbourg problem the moments μ_k do not exist: by a series of misinterpretations and treating μ_k as a number (assuming, in particular, that $\infty - \infty = 0$ and that all passages to limits may be interchanged), the classical theory managed to "prove" a mathematical theorem which contradicts common sense. Actually there is no problem to analyze the asymptotic behavior of the sums S_n in the St. Petersbourg case and thus to determine what used to be called "fair price." To analyze the real mathematical problem we must proceed

as in the case of the central limit theorem and free the theory from all artificial restrictions. Accordingly we shall say that *the sequence* $\{X_k\}$ *obeys the (generalized) weak law of large numbers if there exist two sequences of constants* $\{c_n\}$ *and* $\{p_n\}$ *such that for every positive* ϵ

$$(79) \qquad \Pr\left\{\,|\,S_n - c_n\,| > \epsilon p_n\right\} \to 0.$$

The necessary[24] and sufficient condition for (79) *to hold is that*

$$(80) \qquad \sum_{k=1}^{n} \int_{|x|>p_n} dV_k(x) = o(1),$$

$$(81) \qquad \sum_{k=1}^{n} \int_{|x|<p_n} x^2 dV_k(x) = o(p_n^2);$$

the constants c_n *can then be defined by*

$$(82) \qquad c_n = \sum_{k=1}^{n} \int_{|x|<p_n} x dV_k(x).$$

This theorem[25] completely solves the problem of the weak law of large numbers. Despite several attempts the problem of the strong law still remains open. We say that *the sequence* $\{X_k\}$ *obeys the strong law of large numbers if there exists a sequence of constants* $\{c_n\}$ *such that with probability one*

$$(83) \qquad \Pr\left\{\,|\,S_n - c_n\,|\,/n\right\} \to 0.$$

There exists a famous *sufficient* condition for this law which is due to Kolmogoroff [1928]. For its formulation we shall suppose that the origin has been chosen as described in footnote 13. *Kolmogoroff's condition then consists in the simultaneous covergence of the two series*

$$(84) \qquad \sum_{k=1}^{n} \int_{|x|>k} dV_k(x)$$

and

$$(85) \qquad \sum \frac{b_n}{n^2}$$

[24] The condition is necessary only if the origin has been chosen as described in footnote 13, but sufficient even without this restriction.

[25] Feller [3, 1937]. For alternative proofs cf. Marcinkiewicz [1938], Gnedenko [1939], Doblin [3, 1939]. Several special cases (in particular for non-negative random variables) have been treated previously by Khintchine [7, 1936], Bawly [1936], Plessner [1936], and, perhaps, others.

where

$$(86) \qquad b_n = \int_{|x|<n} x^2 dV_n(x).$$

The condition (84) is clearly also necessary. Other conditions, in part necessary and in part sufficient, can easily be obtained from the law of the iterated logarithm. However, up to now all attempts have failed to replace (85) by a weaker condition which is also necessary.[26]

REFERENCES

G. M. BAWLY
  1. *Ueber eine Verallgemeinerung der Grenzwertsätze der Wahrscheinlichkeitsrechnung*, Rec. Math. (Mat. Sbornik) N.S. vol. 1 (1936) pp. 917–929.
S. N. BERNSTEIN
  1. *Retour au problème de l'évaluation de l'approximation de la formule limite de Laplace*, Bull. Acad. Sci. URSS. Sér. Math. vol. 7 (1943) pp. 3–16.
A. C. BERRY
  1. *The accuracy of the Gaussian approximation to the sum of independent variates*, Trans. Amer. Math. Soc. vol. 49 (1941) pp. 122–136.
A. A. BOBROFF
  1. *Ueber relative Stabilität von Summen zufälliger Grössen*, C. R. (Doklady) Acad. Sci. URSS. vol. 15 (1937) pp. 239–240.
  2. *Conditions of applicability of the strong law of large numbers*, Duke Math. J. vol. 12 (1945) pp. 43–46.
F. P. CANTELLI
  1. *Considerazioni sulla legge uniforme dei grandi numeri e sulla generalizzazione di un fondamentale teorema del sig. Paul Lévy*, Giornale dell'Istituto Italiano degli Attuari vol. 4 (1933) pp. 327–350.
H. CRAMÉR
  1. *On the composition of elementary errors*, Skandinavisk Aktuarietidskrift vol. 11 (1928) pp. 13–74, 141–180.
  2. *Su un teorema relativo alla legge uniforme dei grandi numeri*, Giornale dell'Istituto Italiano degli Attuari vol. 5 (1934) pp. 1–15.
  3. *Ueber eine Eigenschaft der normalen Verteilungsfunktion*, Math. Zeit. vol. 41 (1936) pp. 405–414.
  4. *Random variables and probability distributions*, Cambridge Tracts in Mathematics, No. 36, 1937.
W. DOBLIN (DOEBLIN)
  1. *Premiers éléments d'une étude systématique de l'ensemble de puissances d'une loi de probabilité*, C. R. Acad. Sci. Paris vol. 206 (1938) pp. 306–308.

[26] Generalizing a result of Halmos [1939], Bobroff has just published the following theorem: If ψ_n and $\sum\psi_n^{-1}$ converges, then the convergence of $\sum b_n \psi_n^{-2}$ is a necessary condition for the strong law of large numbers. Unfortunately this theorem contains nothing new and is misleading. Actually Bobroff's condition is much weaker than the condition (81) (with $p_n = n$) which is necessary for the *weak law*. It is trivial that not only Bobroff's series, but also its majorating series $\sum b_n | (n\psi_n)$ must converge, even in the case of the weak law.

2. *Sur les sommes d'un grand nombre de variables aléatoires indépendantes*, Bull. Sci. Math. vol. 63 (1939) pp. 1–40.

3. *Sur un problème du calcul des probabilités*, C. R. Acad. Sci. Paris vol. 209 (1939) pp. 742–743.

J. L. Doob
1. *The elementary Gaussian processes*, Ann. Math. Statist. vol. 15 (1944) pp. 229–282.

W. Dubrovski
1. *Eine Verallgemeinerung der Theorie der rein unstetigen stochastischen Prozesse von W. Feller*, C. R. Acad. Sci. URSS vol. 19 (1938) pp. 439–446.

P. Erdös
1. *On the law of the iterated logarithm*, Ann. of Math. vol. 43 (1942) pp. 419–436.

W. Feller
1. *Ueber den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung*, Math. Zeit. vol. 40 (1935) pp. 521–559.

2. *Zur Theorie der stochastischen Prozesse (Existenz- und Eindeutigkeitssätze)*, Math. Ann. vol. 113 (1936) pp. 113–160.

3. *Ueber das Gesetz der grossen Zahlen*, Acta Univ. Szeged. vol. 8 (1937) pp. 191–201.

4. *Ueber den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechung II*, Math. Zeit. vol. 42 (1937) pp. 301–312.

5. *On the Kolmogoroff-P. Lévy formula for infinitely divisible distribution functions*, Proceedings of the Yugoslav Academy of Sciences vol. 82 (1937) pp. 95–112 (in Yugoslav with summaries).

6. *Generalization of a probability limit theorem of Cramér*, Trans. Amer. Math. Soc. vol. 54 (1943) pp. 361–372.

7. *The general form of the so-called law of the iterated logarithm*, Trans. Amer. Math. Soc. vol. 54 (1943) pp. 373–402.

R. Fortet
1. *Sur le calcul de certaines probabilités d'absorbtion*, C. R. Acad. Sci. Paris vol. 212 (1941) pp. 1118–1120.

2. *Sur la résolution des équations paraboliques linéaires*, C. R. Acad. Sci. Paris vol. 213 (1941) pp. 553–556.

B. V. Gnedenko
1. *On the theory of limit theorems for sums of independent random variables*, Bull. Acad. Sci. URSS. (1939) pp. 181–232 and 643–647 (Russian with English summary).

2. *On the theory of the domains of attraction of stable laws*, Učenye Zapiski Moskovskoga Gosudarstvenogo Univerziteta vol. 30 (1939) pp. 61–81 (Russian with English summary).

3. *Locally stable distributions*, Bull. Acad. Sci. URSS. Sér. Math. vol. 6 (1942) pp. 291–308 (Russian with English summary).

4. *Sur la croissance des processus stochastiques homogènes à accroissements indépendants*, Bull. Acad. Sci. URSS. Sér. Math. vol. 7 (1943) pp. 89–110 (Russian with French summary).

B. V. Gnedenko and A. V. Groshev
1. *On the convergence of distribution laws of normalized sums of independent random variables*, Rec. Math. (Mat. Sbornik) N.S. vol. 6 (1939) pp. 521–541 (Russian with English summary).

P. R. Halmos
1. *On a necessary condition for the strong law of large numbers*, Ann. of Math. vol. 40 (1939) pp. 800–804.

G. H. Hardy and J. E. Littlewood
1. *Some problems of Diophantine approximation*, Acta Math. vol. 37 (1914) pp. 155–239.

P. Hartman
1. *Normal distributions and the law of the iterated logarithm*, Amer. J. Math. vol. 63 (1941) pp. 584–588.

F. Hausdorff
1. *Grundzüge der Mengenlehre*, Leipzig, 1913.

P. L. Hsu
1. *The approximate distribution of the mean and of the variance of independent variates*. Ann. Math. Statist. vol. 16 (1945) pp. 1–29.

M. Kac
1. *Sur les fonctions $2^n t - (2^n t) - 1/2$*, J. London Math. Soc. vol. 13 (1938) pp. 131–134.
2. *Note on the partial sums of the exponential series*, Revista de Matematica y Fisica Teorica Ser. A. vol. 3 (1942) pp. 151–153.

A. Khintchine
1. *Ueber dyadische Brüche*, Math. Zeit. vol. 18 (1923) pp. 109–116.
2. *Ueber einen Satz der Wahrscheinlichkeitsrechnung*, Fund. Math. vol. 6 (1924) pp. 9–20.
3. *Ueber das Gesetz der grossen Zahlen*, Math. Ann. vol. 96 (1926) pp. 156–168.
4. *Remarques sur les suites d'événements obeissants à la loi des grands nombres*, Rec. Math. (Mat. Sbornik) N.S. vol. 39 (1932) pp. 115–119.
5. *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, vol. 2, no. 4, Berlin, 1933.
6. *Sul dominio di attrazione della legge di Gauss*, Giornale dell'Istituto Italiano degli Attuari vol. 6 (1935) pp. 378–393.
7. *Su una legge dei grandi numeri generalizzata*, Giornale dell'Istituto Italiano degli Attuari vol. 7 (1936) pp. 365–377.
8. *Zur Theorie der unbeschränkt teilbaren Verteilungsgesetze*, Rec. Math. (Mat. Sbornik) N.S. vol. 2 (1937) pp. 79–117.
9. *Déduction nouvelle d'une formule de M. Paul Lévy*, Bull. Math. Univ. Moscou vol. 1 (1937) pp. 1–5.
10. *Contribution à l'arithmétique des lois de distribution*, Bull. Math. Univ. Moscou vol. 1 (1937) pp. 6–17.
11. *Invariante Klassen von Verteilungsgesetzen*, Bull. Math. Univ. Moscou vol. 1 (1937) pp. 4–5.
12. *Zwei Sätze über stochastische Prozesse mit stabilen Verteilungen*, Rec. Math. (Mat. Sbornik) N.S. vol. 3 (1938) pp. 577–583.
13. *Sur la croissance locale des processus stochastiques homogènes à accroissements indépendants*, Bull. Acad. Sci. URSS. Sér. Math. (1939) pp. 487–508.

A. Khintchine and P. Lévy
1. *Sur les lois stables*, C. R. Acad. Sci. Paris vol. 202 (1936) pp. 374–376.

A. Kolmogoroff
1. *Ueber die Summen durch den Zufall bestimmter unabhängiger Grössen*, Math. Ann. vol. 99 (1928) pp. 309–319 and corrections vol. 102 (1929) pp. 484–489.

2. *Ueber das Gesetz des iterierten Logarithmus*, Math. Ann. vol. 101 (1929) pp. 126–135.
3. *Ueber die analytischen Methoden der Wahrscheinlichkeitsrechnung*, Math. Ann. vol. 104 (1931) pp. 415–458.
4. *Sulla forma generale di un processo stocastico omogeneo*, Atti della Reale Accademia Nazionale dei Lincei: Rendiconti (6) vol. 15 (1932) pp. 805–808, 866–869.
5. *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, vol. 2, no. 3, Berlin, 1933.

H. LEVENE AND J. WOLFOWITZ
    1. *The covariance matrix of runs up and down*, Ann. Math. Statist. vol. 15 (1944) pp. 58–69.

P. LÉVY
    1. *Sur les séries dont les termes sont des variables éventuelles indépendantes*, Studia Mathematica vol. 3 (1931) pp. 117–155.
    2. *Sulla legge forte dei grandi numeri*, Giornale dell'Istituto Italiano degli Attuari vol. 2 (1931) pp. 1–21.
    3. *Nuove formule relative al giuco di testa e croce*, Giornale dell'Istituto Italiano degli Attuari vol. 2 (1931) pp. 127–160.
    4. *Sur un théorème de M. Khintchine*, Bull. Sci. Math. vol. 55 (1931) pp. 145–160.
    5. *Sur les intégrales dont les éléments sont des variables aléatoires indépendantes*, Annali della Scuola Normale Superiore di Pisa (2) vol. 3 (1934) pp. 337–366.
    6. *Propriétés asymptotiques des sommes de variables aléatoires indépendantes ou enchaînées*, J. Math. Pures Appl. vol. 14 (1935) pp. 347–402.
    7. *La loi forte des grands nombres pour les variables aléatoires enchaînées*, J. Math. Pures Appl. vol. 15 (1936) pp. 11–24.
    8. *Détermination générale des lois limites*, C. R. Acad. Sci. Paris vol. 203 (1936) pp. 698–700.
    9. *Théorie de l'addition de variables aléatoires*, Paris, 1937.
    10. *Compléments à théorème sur la loi de Gauss*, Bull. Sci. Math. vol. 61 (1937) pp. 115–128.
    11. *Sur les exponentielles des polynomes et sur l'arithmétique des produits de lois de Poisson*, Ann. École Norm. vol. 54 (1937) pp. 231–292.
    12. *L'arithmétique des lois de probabilité*, J. Math. Pures Appl. vol. 17 (1938) pp. 17–39.

J. W. LINDEBERG
    1. *Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung*, Math. Zeit. vol. 15 (1922) pp. 211–225.

J. MARCINKIEWICZ
    1. *Sur les fonctions indépendantes* I–III, Fund. Math. vol. 30 (1938) pp. 202–214, 349–364, vol. 31 (1938) pp. 86–102.
    2. *Quelques théorèmes de la théorie des probabilités*, Travaux de la Société des Sciences et des Lettres de Wilno, Classe des Sciences Mathématiques et Naturelles vol. 13 (1939) pp. 1–13.

J. MARCINKIEWICZ AND A. ZYGMUND
    1. *Remarque sur la loi du logarithme itéré*, Fund. Math. vol. 29 (1937) pp. 215–222.

I. PETROWSKI
    1. *Ueber das Irrfahrtproblem*, Math. Ann. vol. 109 (1934) pp. 425–444.

A. PLESSNER
  1. *Ueber das Gesetz der grossen Zahlen*, Rec. Math. (Mat. Sbornik) N.S. vol. 1
     (1936) pp. 165–168.
G. PÓLYA
  1. *Ueber den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das
     Momentenproblem*, Math. Zeit. vol. 8 (1920) pp. 171–181.
  2. *Herleitung des Gausschen Fehlergesetzes aus einer Funtionalgleichung*, Math.
     Zeit. vol. 18 (1923) pp. 96–108.
H. RADEMACHER
  1. *Einige Sätze über Reihen von Allgemeinen Orthogonalfunktionen*, Math. Ann.
     vol. 87 (1922) pp. 112–138.
D. RAIKOV
  1. *On the decomposition of Gauss and Poisson laws*, Bull. Acad. Sci. URSS. Sér.
     Math. (1938) pp. 91–120 (Russian, English summary pp. 120–124).
  2. *On a connection between the central limit theorem of the theory of probability
     and the law of large numbers*, Bull. Acad. Sci. URSS. Sér. Math. (1938) pp.
     323–336 (Russian, English summary pp. 337–338).
H. STEINHAUS
  1. *Les probabilités dénombrables et leur rapport à la théorie de la mesure*, Fund.
     Math. vol. 4 (1922) pp. 286–310.
J. V. USPENSKY
  1. *Introduction to mathematical probability*, New York, McGraw-Hill, 1937.
A. WALD
  1. *On cumulative sums of random variables*, Ann. Math. Statist. vol. 15 (1944)
     pp. 283–296.

BROWN UNIVERSITY