PROBLEMS IN PROBABILITY THEORY

By Harald Cramér

University of Stockholm

1. Introduction. The following survey of problems in probability theory has been written for the occasion of the Princeton Bicentennial Conference on "The Problems of Mathematics," Dec. 17–19, 1946. It is strictly confined to the purely mathematical aspects of the subject. Thus all questions concerned with the philosophical foundations of mathematical probability, or with its ever increasing fields of application, will be entirely left out.

No attempt to completeness has been made, and the choice of the problems considered is, of course, highly subjective. It is also necessary to point out explicitly that the literature of the war years has only recently—and still far from completely—been available in Sweden. Owing to this fact, it is almost unavoidable that this paper will be found incomplete in many respects.

I. FUNDAMENTAL NOTIONS

2. Probability distributions. From a purely mathematical point of view, probability theory may be regarded as the theory of certain classes of *additive set functions*, defined on spaces of more or less general types. The basic structure of the theory has been set out in a clear and concise way in the well-known treatise by Kolmogoroff [53]. We shall begin by recalling some of the main definitions. Note that the word *additive*, when used in connection with sets or set functions, will always refer to a *finite or enumerable* sequence of sets.

Let ω denote a variable point in an entirely arbitrary space Ω, and consider an additive class C of sets in Ω, such that the whole space Ω itself is a member of C. Further, let $P(S)$ be an additive set function, defined for all sets S belonging to the class C, and suppose that

$$P(S) \geqq 0 \text{ for all } S \text{ in } C,$$

$$P(\Omega) = 1.$$

We shall then say that $P(S)$ is a *probability measure*, which defines a *probability distribution in* Ω. For any set S in C, the quantity $P(S)$ is called the *probability* of the *event* expressed by the relation $\omega \subset S$, i.e. the event that the variable point ω takes a value belonging to S. Accordingly we write

$$P(S) = P(\omega \subset S).$$

Suppose now that $\omega' = g(\omega)$ is a function of the variable point ω, defined throughout the space Ω, the values ω' being points of another arbitrary space Ω'. Let S' be a set in Ω' and denote by S the set of all points ω such that $\omega' = g(\omega)$ belongs to S'. Whenever S belongs to C, we define a set function $P'(S')$ by writing

$$P'(S') = P(S).$$

It is then easy to see that $P'(S')$ is defined for all S' belonging to a certain additive class C' in the new space Ω', and that $P'(S')$ is a probability measure in Ω', such that $P'(S')$ signifies the probability of the event $\omega' \subset S'$ (which is equivalent to $\omega \subset S$). We shall say that $P'(S')$ is attached to the probability distribution in Ω' which is *induced* by the given distribution in Ω and the function $\omega' = g(\omega)$.

3. Random variables. Consider in particular the case when ω' is a real number ξ, such that $\xi = g(\omega)$ is a real-valued C-measurable function of the argument ω. Then C' includes the class B_1 of all Borel sets S' of the space $\Omega' = R_1$ of all real numbers, and we shall call ξ a *one-dimensional real random variable.* The probability of the event $\xi \subset S'$ is uniquely defined for any Borel set S' of R_1, as soon as the function

$$F(x) = P(\xi \leqq x)$$

is known for all real x. $F(x)$ is called the *distribution function* (*d.f.*) of the random variable ξ. If the function $\xi = g(\omega)$ is integrable over Ω with respect to the measure $P(S)$, we write

$$E\xi = \int_\Omega g(\omega)\, dP = \int_{-\infty}^{\infty} x\, dF(x),$$

and denote this expression as the *expectation* or *mean value* of the random variable ξ. Any real-valued B-measurable function $\eta = h(\xi)$ is also a random variable with the probability distribution induced by the original ω-distribution and the function $\eta = h(g(\omega))$. If η is integrable over Ω with respect to P, its mean value may be written in the form

$$E\eta = Eh(\xi) = \int_\Omega h(g(\omega))\, dP = \int_{-\infty}^{\infty} h(x)\, dF(x).$$

More generally, if $\omega' = (\xi_1, \cdots, \xi_n)$ is a point in an n-dimensional Euclidean space R_n, while C' includes the class B_n of all Borel sets of R_n, we are concerned with an *n-dimensional real random variable.* The distribution of this variable, which is also called the joint distribution of the n one-dimensional variables $\xi_1, \cdots, \xi_n$, is uniquely defined, as soon as the joint d.f.

$$F(x_1, \cdots, x_n) = P(\xi_1 \leqq x_1, \cdots, \xi_n \leqq x_n)$$

is known for all real $x_1, \cdots, x_n$.

The variables $\xi_1, \cdots, \xi_n$ are said to be *independent,* if $F(x_1, \cdots, x_n) = F_1(x_1) \cdots F_n(x_n)$, where $F_\nu(x_\nu)$ is the d.f. of the variable ξ_ν.

The extension to *complex random variables* is obvious. Suppose e.g. that $\xi = g(\omega)$ and $\eta = h(\omega)$ are two one-dimensional real variables, and consider the complex variable $\xi + i\eta = g(\omega) + ih(\omega)$. By definition, we identify the distribution of this variable with that of the two-dimensional real variable (ξ, η), and we put

$$E(\xi + i\eta) = E\xi + iE\eta.$$

Joint distributions of several complex variables are introduced in a corresponding way.

4. Characteristic functions. If ξ is a one-dimensional real random variable, the mean value

$$\varphi(z) = Ee^{iz\xi} = \int_{-\infty}^{\infty} e^{izx}\, dF(x)$$

exists for all real z, and we have

$$| \varphi(z) | \leqq 1, \qquad \varphi(0) = 1.$$

$\varphi(z)$ is called the *characteristic function* (*c.f.*) of the distribution corresponding to the variable ξ. The reciprocal formula (Lévy)

$$F(x) - F(y) = -\frac{1}{2\pi i} \lim_{Z\to\infty} \int_{-Z}^{Z} \frac{e^{-izx} - e^{-izy}}{z} \varphi(z)\, dz,$$

which holds for any continuity points x and y of F, shows that there is a one-one correspondence between the d.f. $F(x)$ and the c.f. $\varphi(z)$. As we shall see below, the c.f. provides a powerful analytical tool for operations with probability distributions.

When a complex-valued function $\varphi(z)$ of the real variable z is given, it is often important to be able to decide whether $\varphi(z)$ is or is not the c.f. of some distribution. If we assume a priori that $\varphi(0) = 1$, each of the following conditions is necessary and sufficient for $\varphi(z)$ to be a c.f.

A. $\varphi(z)$ should be bounded and continuous for all z, and such that the integral

$$\int_0^A \int_0^A \varphi(z - u) e^{ix(z-u)}\, dz\, du$$

is real and non-negative for all real x and all $A > 0$ (Cramér [11], in simplification of an earlier result due to Bochner, [4]).

B. There should exist a sequence of functions $\psi_1(z), \psi_2(z), \cdots$ such that

$$\varphi(z) = \lim_{n\to\infty} \int_{-\infty}^{\infty} \psi_n(x + z)\overline{\psi_n(x)}\, dx$$

holds uniformly in every finite z-interval (Khintchine, [45]).

These general theorems are not always easy to apply in practice. Among less general results which are more easily applicable, we mention the almost trivial fact that a function $\varphi(z)$ which near $z = 0$ is of the form $\varphi(z) = 1 + o(z^2)$ cannot be a c.f. unless $\varphi(z) = 1$ for all z, and the two following theorems:

1) An integral function $\varphi(z)$ of order $\gamma < 1$ can never be a c.f. (Lévy, [64]), and

2) an integral function $\varphi(z)$ of finite order $\gamma > 2$ cannot be a c.f. unless the convergence exponent of its zeros is equal to γ (Marcinkiewicz, [72]). The latter result shows e.g. that no function of the form $e^{g(z)}$, where $g(z)$ is a polynomial of degree > 2, can be a c.f.

It would be highly desirable to obtain further results in this direction.

The c.f. of the joint distribution of n real random variables $\xi_1, \cdots, \xi_n$ is the function $\varphi(z_1, \cdots, z_n)$ defined by the relation

$$\varphi(z_1, \cdots, z_n) = Ee^{i(z_1\xi_1+\cdots+z_n\xi_n)}.$$

Most of the above results for c.f. in one variable can be directly generalized to the multi-variable case.

5. Random sequences and random functions. Let t be a variable point in an arbitrary space $\mathbf{T}$, and consider the space Ω, where each point ω is a real-valued function $\omega = x(t)$ of the variable argument t. Let $t_1, \cdots, t_n$ be any finite set of distinct points t. The set of all functions $\omega = x(t)$ satisfying the inequalities

$$a_j < x(t_j) \leqq b_j, (j = 1, \cdots, n),$$

will be called an *interval* in the space Ω. The Borel sets in Ω will be defined as the smallest additive class B of sets in Ω containing all intervals.

Suppose now that, for any choice of n and the t_j, the variables $x(t_1), \cdots, x(t_n)$ are random variables having a known n-dimensional joint distribution. If the family of all distributions corresponding in this way to finite sequences t_1, $\cdots$, t_n satisfies certain obvious consistency conditions, a fundamental theorem due to Kolmogoroff asserts that this family determines a unique probability distribution in the space Ω of all functions $x(t)$. The corresponding probability

$$P(S) = P(x(t) \subset S)$$

is uniquely defined for all Borel sets S of Ω.

Consider in particular the case where $\mathbf{T}$ is the set of non-negative integers $t = 0, 1, 2, \cdots$. The space Ω then is the space of all sequences $(x_0, x_1, \cdots)$ of real numbers. As soon as the joint distribution of any finite number of variables $x_{\nu_1}, \cdots, x_{\nu_n}$ is defined, and these distributions are mutually consistent, it then follows that there is a unique probability distribution of the *random sequence* $(x_0, x_1, \cdots)$, the corresponding probability being defined for every Borel set of the space Ω of sequences. Similarly we may consider the doubly infinite sequence $(\cdots, x_{-1}, x_0, x_1, \cdots)$.

Consider further the more general case when $\mathbf{T}$ is any set of real numbers. Then Ω is the space of all real-valued functions $\omega = x(t)$ defined on the set $\mathbf{T}$, and as before the knowledge of the distributions for all finite sets of variables $x(t_1), \cdots, x(t_n)$ permits us to determine a probability distribution in the space Ω of *random functions* $x(t)$, the probability $P(S) = P(x(t) \subset S)$ being always defined for all Borel sets S in Ω.

The generalization of the above considerations to *complex-valued* random sequences and functions is immediate.

6. Various modes of convergence. Consider a sequence $F_1(x)$, $F_2(x)$, $\ldots$ of d.f:s, and let the corresponding c.f:s be $\varphi_1(t)$, $\varphi_2(t)$, $\cdots$. In order that $F_n(x)$

converge to a d.f. $F(x)$, in every continuity point of the latter, it is necessary and sufficient[1] that $\varphi_n(t)$ converge for every real t to a limit $\varphi(t)$ which is continuous at $t = 0$. Then $\varphi(t)$ is the c.f. corresponding to the d.f. $F(x)$.

Further, let x and $x_1, x_2, \cdots$ be complex-valued random variables, such that the random sequence $(x, x_1, x_2, \cdots)$ has a well defined distribution. We shall be concerned with various modes of convergence of x_n to x.

A. When $P(|x_n - x| > \epsilon) \to 0$ as $n \to \infty$, for any $\epsilon > 0$, we shall say that x_n *converges to* x *in probability.*

B. When $E|x_n - x|^\gamma \to 0$, as $n \to \infty$, where $\gamma > 0$ is fixed, we shall say that x_n converges to x in the mean of order γ. Unless otherwise stated we shall in the sequel always consider the case $\gamma = 2$, and in this case we shall use the notation

$$\underset{n\to\infty}{\text{l.i.m.}}\, x_n = x.$$

C. When $P(\lim_{n\to\infty} x_n = x) = 1$, we shall say that x_n converges with probability one, or *converges almost certainly to* x.

With respect to the last definition, we may remark that the set defined by the relation $\lim x_n = x$ is always a Borel set in the space of our random sequence, so that the probability of this relation is well defined. In fact, this probability is given by the expression

$$\lim_{m\to\infty} \lim_{n\to\infty} \lim_{p\to\infty} P\left(|x_\nu - x| < \frac{1}{m} \quad \text{for} \quad \nu = n, n+1, \cdots, n+p\right)$$

where the limit process applies to a probability attached to a Borel set in a finite number of dimensions. The case of almost certain convergence is precisely the case when this expression takes the value 1.

Convergence in the mean of any positive order, as well as almost certain convergence, both imply convergence in probability, which may be written symbolically $B \to A$ and $C \to A$. Between B and C, there is no simple relation of this kind. Further, A and B both imply almost certain convergence for any partial sequence $x_{n_1}, x_{n_2}, \cdots$ such that the subscripts n_k increase sufficiently rapidly with k.

II. PROBLEMS CONNECTED WITH THE ADDITION OF INDEPENDENT VARIABLES

7. During the early development of the theory of probability, the majority of problems considered were connected with gambling. The gain of a player in a certain game may be regarded as a random variable, and his total gain in a

[1] As I have already stated in a paper published in 1938, there is an error in the statement of this theorem given in my Cambridge Tract [9] *Random Variables and Probability Distributions*. For the truth of the theorem, it is essential that $\varphi_n(t)$ should be supposed to converge to $\varphi(t)$ *for every real t*. However, in the particular case when the limit $\varphi(t)$ is analytic and regular in the vicinity of $t = 0$, it can be proved that it is sufficient to assume convergence in some interval $|t| < a$.

sequence of repetitions of the game is the sum of a number of independent variables, each of which represents the gain in a single performance of the game. Accordingly a great amount of work was devoted to the study of the probability distributions of such sums. A little later, problems of a similar type appeared in connection with the theory of errors of observation, when the total error was considered as the sum of a certain number of partial errors due to mutually independent causes. At first only particular cases were considered, but gradually general types of problems began to arise, and in the classical work of Laplace several results are given concerning the general problem to study the distribution of a sum

$$z_n = x_1 + \cdots + x_n$$

of independent variables, when the distributions of the x_j are given. This problem may be regarded as the very starting point of a large number of those investigations by which the modern Theory of Probability was created. The efforts to prove certain statements of Laplace, and to extend his results further in various directions, have largely contributed to the introduction of rigorous foundations of the subject, and to the development of the analytical methods. At the same time, more general types of problems have developed from the original problem, and the number and importance of practical applications have been steadily increasing.

8. Composition of distributions. Let x_1 and x_2 be two independent variables, with the d.f.'s F_1 and F_2, and the c.f.'s φ_1 and φ_2, and let the sum $x_1 + x_2$ have the d.f. F and the c.f. φ. Then

$$F(x) = \int_{-\infty}^{\infty} F_1(x - y)\, dF_2(y) = \int_{-\infty}^{\infty} F_2(x - y)\, dF_1(y).$$

We shall say that F is the *composition* of F_1 and F_2, and write this as a symbolical multiplication:

$$F = F_1 * F_2 = F_2 * F_1 .$$

To this symbolical multiplication of the d.f:s corresponds a real multiplication of the c.f.'s:

$$\varphi(z) = \varphi_1(z)\varphi_2(z).$$

The operation of composition is both commutative and associative, so that any symbolical product $F = F_1 * F_2 \cdots * F_n$ is uniquely defined and independent of the order of the components. When at least one of the components is continuous (absolutely continuous), the same holds for the composite, and in many cases it is true that the composite is at least as regular as the most regular of the components (Lévy, [58], [63], etc.). However, this general statement does not hold generally, as is shown by an interesting example due to Raikov, [77], where F_1 and F_2 are integral analytic functions, while the composite $F = F_1 * F_2$ is not regular at the origin.

It seems to be an important unsolved problem to find convenient restrictions ensuring the validity of the above statements of the "smoothing effect" of the operation of composition.

When $F = F_1 * F_2$, we may say that F is "divisible" by each component F_1 and F_2, and it seems natural to try to develop a theory of symbolical factorization for d.f.'s. In this connection, it is important to note that symbolical division is not unique. In fact, Khintchine has shown by an example that it is possible to find the d.f.'s F, F_1, F_2, and F_3 such that

$$F = F_1 * F_2 = F_1 * F_3,$$

while $F_2 \neq F_3$. Another fundamental problem belonging to this order of ideas is to decide whether a given d.f. F is decomposable or not. F is called decomposable, if there is at least one representation of the form $F = F_1 * F_2$, where each component F_ν has more than one point of increase. So far, this problem has only been solved in very special cases, and the general problem still remains open for research. A particular case of some interest would be to know if there exists an absolutely continuous and indecomposable d.f., such that $F(a) = 0$ and $F(b) = 1$ for some finite a and b.

As soon as we restrict ourselves to certain special classes of distributions, it is possible to reach results of a more definite character concerning the factorization problems. Some results of this type will be considered below.

9. Closed families of distributions. The fact that certain families of distributions are closed with respect to the operation of composition has played an important part in many applications. If F_1 and F_2 belong to a family of this character, so does the symbolical product $F = F_1 * F_2$. We first give some simple examples of such families.

The normal distribution. The d.f. F has the form $F = \phi\left(\frac{x - m}{\sigma}\right)$, where $\sigma > 0$, and

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(t^2/2)}\, dt.$$

The c.f. corresponding to F is $e^{miz - \frac{1}{2}\sigma^2 z^2}$, and it follows that for any real m_1, m_2 and any positive σ_1, σ_2 we have

$$\phi\left(\frac{x - m_1}{\sigma_1}\right) * \phi\left(\frac{x - m_2}{\sigma_2}\right) = \phi\left(\frac{x - m}{\sigma}\right),$$

where

$$m = m_1 + m_2, \qquad \sigma^2 = \sigma_1^2 + \sigma_2^2.$$

The Poisson distribution. Here the d.f. is $F = F(x; \lambda, m, a)$ where $\lambda > 0$, $a \neq 0$, and F is a step-function with a jump equal to $\frac{\lambda^\nu}{\nu!} e^{-\lambda}$ in the point $x = m + \nu a$, where $\nu = 0, 1, \cdots$. The corresponding c.f. is $e^{miz + \lambda(e^{aiz} - 1)}$, and it follows that for any fixed a we have

$$F(x; \lambda_1, m_1, a) * F(x; \lambda_2, m_2, a) = F(x; \lambda_1 + \lambda_2, m_1 + m_2, a).$$

The Pearson Type III distribution. $F = F(x;\ \alpha,\ \lambda) = \dfrac{\alpha^\lambda}{\Gamma(\lambda)} \displaystyle\int_0^x t^{\lambda-1} e^{-\alpha t}\, dt$, $(x > 0)$. The corresponding c.f. is $\left(1 - \dfrac{iz}{\alpha}\right)^{-\lambda}$, and for any fixed $\alpha > 0$ and any positive λ_1 and λ_2 we have

$$F(x;\ \alpha,\ \lambda_1) * F(x;\ \alpha,\ \lambda_2) = F(x;\ \alpha,\ \lambda_1 + \lambda_2).$$

Stable distributions. We shall say that a closed family is stable, when all its members are of the form $F(ax + b)$, where F is a d.f., while $a > 0$ and b are constants. Obviously the normal family is an example of a stable family. It has been shown by Lévy and Khintchine [49], that a d.f. $F(x)$ generates a stable family when and only when the logarithm of its c.f. is of the form

$$\log \varphi(z) = \beta iz - \gamma \, |z|^\alpha \left(1 + i\delta \frac{z}{|z|} \omega\right), \tag{9.1}$$

where α, β, γ, δ are real constants such that

$$0 < \alpha \leqq 2, \qquad \gamma > 0, \qquad |\delta| \leqq 1,$$

while

$$\omega = \begin{cases} tg \dfrac{\alpha\pi}{2} & \text{for} \quad \alpha \neq 1 \\ \dfrac{2}{\pi} \log |z| & \text{for} \quad \alpha = 1. \end{cases}$$

For $\alpha = 2$ we obtain the normal family.

A more general and very important closed family is the family I of *infinitely divisible distributions.* A d.f. F belongs to I if to every $n = 1, 2, \cdots$ there exists a d.f. G such that $F = G^{[n]}$, where $G^{[n]}$ denotes the symbolical nth power of G. Obviously the family I is a closed family which contains all the families mentioned above. Lévy [60], [63], has shown that F is infinitely divisible when and only when the logarithm of its c.f. is of the form

$$\begin{aligned} \log \varphi(z) = \beta iz - \gamma z^2 &+ \int_{-\infty}^0 \left(e^{izu} - 1 - \frac{izu}{1 + u^2}\right) dM(u) \\ &+ \int_0^\infty \left(e^{izu} - 1 - \frac{izu}{1 + u^2}\right) dN(u), \end{aligned} \tag{9.2}$$

where β and $\gamma > 0$ are real constants, while $M(u)$ and $N(u)$ are non-decreasing functions such that

$$M(-\infty) = N(+\infty) = 0,$$

$$\int_{-a}^0 u^2\, dM(u) < \infty \quad \text{and} \quad \int_0^a u^2\, dN(u) < \infty$$

for any finite $a > 0$. When M and N reduce to zero, we obtain the normal family. When $\gamma = 0$ and one of the functions M and N reduces to zero, while the other is a step-function with a single jump equal to λ at the point $x = a$, we obtain a Poisson family. Generally, it follows from (9.2) that any infinitely divisible distribution may be regarded as a product of a normal distribution and a finite, enumerable or continuous set of Poisson distributions.

The representation of $\log \varphi(z)$ in the form (9.2) is unique. It follows that the problem of finding all possible factorizations of an infinitely divisible d.f. F can be completely solved, as long as we restrict ourselves to factors which are themselves infinitely divisible. In fact, in order that

$$F = F_1 * F_2,$$

where all three d.f.'s belong to I, it is necessary and sufficient that the logarithms of the corresponding c.f.'s should be of the form (9.2), with

$$\beta = \beta_1 + \beta_2, \qquad \gamma = \gamma_1 + \gamma_2,$$
$$M = M_1 + M_2, \qquad N = N_1 + N_2.$$

In the two simple cases of the normal and the Poisson distributions, the decompositions obtained in this way remain the only possible, even if we remove the restriction that the factors should belong to I. Thus in any factorization of a normal distribution, all factors are normal (Cramér, [8]), while in any factorization of a Poisson distribution, all factors belong to the Poisson family (Raikov, [75]). For the type III distribution, and the non-normal stable distributions, however, the corresponding property does not hold.

In some cases, an infinitely divisible distribution may be represented as a product of indecomposable distributions, or as a product of an indecomposable distribution and another infinitely divisible distribution. The results so far obtained in this direction (Lévy, [63], [64], Khintchine, [46], [47]; Raikov, [76]) are all concerned with more or less particular cases, and the general factorization problem for infinitely divisible distributions still remains unsolved. A particular case of some interest would be the case when the functions M and N are both absolutely continuous. There does not seem to have been given any example of this type, where a factor not belonging to I may occur.[2]

Finally we mention a general theorem due to Khintchine, [46], which asserts that an arbitrary d.f. F may be represented in one of the forms

$$F = G, \qquad F = H \text{ or } F = G * H,$$

where G is infinitely divisible, while H is a finite or infinite product of indecomposable factors. This seems to be practically the only result so far known concerning the factorization of a general distribution.

A certain number of the results mentioned above have been generalized to multi-dimensional distributions.

[2] While the present paper was being printed, I have proved that such factors do occur, as soon as at least one of the derivatives M' and N' is bounded away from zero in some interval $(-a, 0)$ or $(0, a)$.

10. The Laws of large numbers. In modern terminology, the classical Bernoulli theorem may be expressed in the following way. Let $x_1, x_2, \cdots$ be a sequence of independent variables, such that each x_ν may only assume the values 1 and 0, the corresponding probabilities being p and $q = 1 - p$. Then the arithmetic mean

$$\frac{z_n}{n} = \frac{x_1 + \cdots + x_n}{n} \cdot \tag{10.1}$$

converges in probability to p, as $n \to \infty$.

Both classical and modern authors have laid down much work on the generalization of this simple result in various directions. Generally, we shall say that a sequence of random variables $x_1, x_2, \cdots$ satisfies the *Weak Law of Large Numbers* if there exist two sequences of constants $a_1, a_2, \cdots$ and $b_1, b_2, \cdots$, such that $a_n > 0$, and

$$\frac{z_n - b_n}{a_n} = \frac{x_1 + \cdots + x_n - b_n}{a_n}$$

converges in probability to zero.

Let $x_1, x_2 \cdots$ be independent variables, such that x_ν has the d.f. $F_\nu(x)$. It has been shown by Feller [27] that *for any given sequence* $a_1, a_2, \cdots$, *the conditions*

$$\begin{aligned} \sum_{\nu=1}^{n} \int_{|x| > a_n} dF_\nu(x) &= o(1), \\ \sum_{\nu=1}^{n} \int_{|x| < a_n} x^2\, dF_\nu(x) &= o(a_n^2), \end{aligned} \tag{10.2}$$

are sufficient for the validity of the weak law of large numbers, and that the corresponding sequence $b_1, b_2, \cdots$ *can be defined by*

$$b_n = \sum_{\nu=1}^{n} \int_{|x| < a_n} x\, dF_\nu(x).$$

When there is a constant $c > 0$ *such that for all* ν

$$F_\nu(+0) > c, \qquad F_\nu(-0) < 1 - c, \tag{10.3}$$

the conditions are also necessary. This theorem contains as particular cases all previously known results in this direction. A simple NS condition for the existence of at least one sequence $a_1, a_2, \cdots$ such that 10.2 holds does not seem to be known.

When the weak law is satisfied, this means that, for any given $\epsilon > 0$ and *for any fixed large* n, there is a probability very near to 1 that the sum $z_n = x_1 + \cdots + x_n$ will fall between the limits $b_n \pm \epsilon a_n$. The more stringent condition that, with a probability tending to 1 as $n \to \infty$, z_ν will fall between the limits $b_\nu \pm \epsilon a_\nu$ *for all values of* $\nu \geqq n$ is equivalent to the condition that $\frac{z_n - b_n}{a_n}$ converges *almost certainly* to zero. When this holds, we shall say that the variables x_ν satisfy the *Strong Law of Large Numbers.* The most important result so far known in this connection is concerned with the case $a_n = n$, and is expressed by the following theorem (Kolmogoroff, [52], [55]):

When the x_ν are independent and (10.3) holds, a sufficient condition for the validity of the strong law with $a_n = n$ consists in the simultaneous convergence of the two series

$$\Sigma \int_{|x|>n} dF_n(x) \qquad \textit{and} \qquad \Sigma \frac{1}{n^2} \int_{|x|<n} x^2 \, dF_n(x).$$

Some improved conditions of this type have been given by Marcinkiewicz and Zygmund, [73], but the problem of finding a NS condition for the strong law is still unsolved, even in the case $a_n = n$.

Important generalizations of the laws of large numbers to cases when the x_ν are not assumed to be independent have been given i.a. by Khintchine [44], Lévy [62], [63] and Loève [67].

11. The central limit theorem and allied theorems. It was already known to De Moivre that, in the case 10.1 of the Bernoulli distribution, the d.f. of the normalized sum

$$\frac{x_1 + \cdots + x_n - np}{\sqrt{npq}}$$

tends, as $n \to \infty$, to the normal d.f. $\phi(x)$. Considerably more general results in this direction were stated by Laplace. After a long series of more or less successful attempts, a rigorous proof of the main statements of Laplace was given in 1901 by Liapounoff, [65]. More general cases were later considered i.a. by Lindeberg [66], Lévy [61], [63], Khintchine [43] and Feller, [25]. The following final form of the *Central Limit Theorem* is due to Feller.

Consider the expression

$$u_n = \frac{z_n - b_n}{a_n} = \frac{x_1 + \cdots + x_n - b_n}{a_n}, \tag{11.1}$$

where the x_ν are independent variables. We shall say that the x_ν obey the central limit law, if the sequences $\{a_\nu\}$ and $\{b_\nu\}$ can be found such that the d.f. of u_n tends to $\phi(x)$ as $n \to \infty$. In order to avoid unnecessary complications, we shall restrict ourselves to sequences $\{a_\nu\}$ such that

$$a_\nu \to + \infty, \qquad \frac{a_{\nu+1}}{a_\nu} \to 1,$$

and we shall assume that the conditions (10.3) are satisfied. Then Feller's theorem runs as follows:

The independent variables $x_1, x_2, \cdots$ obey the central limit law if, and only if, there exists a sequence $q_n \to \infty$ such that simultaneously

$$
\begin{aligned}
&\sum_{\nu=1}^{n} \int_{|x|>q_n} dF_\nu(x) \to 0, \\
(11.2) \qquad & \\
&\frac{1}{q_n^2} \sum_{\nu=1}^{n} \int_{|x|<q_n} x^2 \, dF_\nu(x) \to \infty.
\end{aligned}
$$

When these conditions are satisfied, explicit expressions for the a_n and b_n can be obtained.

Feller's theorem gives a complete solution of the problem. However, we might still try to express in a more direct way the condition that the q_n should exist. We may also ask what happens when the conditions (11.2) are not satisfied. Some particular cases of the latter question will be considered below. However, very few general results are known in this direction.

The central limit theorem has been extended in various directions. Bernstein [3], Lévy [62], [63], Loève [67] and others have considered cases where the x_ν are not assumed to be independent. Important results have been reached but still much remains to be done.

On the other hand, several authors have considered symmetrical functions, other than sums, of n independent random variables. The problem of investigating the asymptotic behaviour of the distributions of such functions, as n tends to infinity, is of great importance in the theory of statistical sampling distributions. It is known (c.f. e.g. Cramér, [15]) that under certain general regularity conditions there exists a normal limiting distribution. However, it is also known that it is possible to give examples of particular functions (such as e.g. the function which is equal to the largest of the n variables), where there exist limiting distributions which are non-normal. The conditions under which this phenomenon may occur seem to deserve further study.

A further problem belonging to the same order of ideas is to find a closer asymptotic representation of the d.f. of the standardized sum z_n than that provided by the normal function $\phi(x)$. Consider e.g. the simple case when the x_ν are independent variables all having the same d.f. $F(x)$ with a finite mean m, a finite variance σ^2, and finite moments up to a certain order $k \geqq 3$. Let $G_n(x)$ be the d.f. of the variable

$$\frac{x_1 + \cdots + x_n - nm}{\sigma\sqrt{n}}.$$

It then follows from a theorem of Cramér [5], [9] that, as soon as the d.f. $F(x)$ contains an absolutely continuous component, there is an asymptotic expansion

$$G_n(x) = \phi(x) + \sum_{\nu=1}^{k-3} \frac{p_\nu(x)}{n^{\nu/2}} e^{-x^2/2} + O(n^{-(k-2)/2}), \tag{11.3}$$

where the constant implied by the O is independent of n and x. Cramér has also given similar expansions in more general cases, and his results have been further extended by P. L. Hsu [39], who deduces analogous expansions also for other functions than sums. The most general conditions under which expansions of this type exist are still unknown.

It follows from (11.3) that the difference $G_n(x) - \phi(x)$ is, for any fixed x, of the order $n^{-\frac{1}{2}}$ as $n \to \infty$. It is often important to know the asymptotic behaviour of $G_n(x)$ when n and x increase simultaneously, and in that case (11.3) yields only a trivial result. This case has been investigated by Cramér [10], and Feller [29], and the results so far obtained permit important applications to the so called law of the iterated logarithm (cf. below). However, it seems likely that similar results may be obtained in considerably more general cases than those hitherto investigated.

A further interesting type of problems belonging to this order of ideas may be approached in the following way. Consider the variables (11.1) in the particular case when $x_1, x_2, \cdots$ are independent variables all having the same d.f. $F(x)$. When the a_n and b_n can be found such that the d.f. of the normalized sum u_n tends to $\phi(x)$, we shall say that F belongs to the *domain of attraction* of the normal law. Feller's theorem gives a NS condition that this should be so. Now when this condition is not satisfied, it may still occur that the a_n and b_n can be so chosen that the d.f. of u_n tends to a limiting d.f. $\Psi(x)$, which is necessarily different from $\phi(x)$. Then it is easily seen that $\Psi(x)$ must be a stable distribution, with its c.f. defined by (9.1), and it is natural to say that F belongs to the domain of attraction of Ψ. NS and sufficient conditions that this should hold have been given by Doeblin [16], and Gnedenko [34]. When the a_n and b_n cannot be found such that the d.f. of the normal sum u_n converges to a limit, it may still be possible to obtain a limiting d.f. by considering only a partial sequence $u_{n_1}, u_{n_2}, \cdots$. Khintchine [47] has proved the interesting theorem that the totality of limiting d.f.'s that may be obtained in this way coincides with the class of infinitely divisible d.f.'s defined by (9.2). There are also further results in the same direction given by Bawly [2], Khintchine [44], Lévy, [61]-[63], and Gnedenko, [35].

12. The law of the iterated logarithm. Consider a sequence of independent variables $x_1, x_2, \cdots$, such that the mean $Ex_n = 0$ for all n, while the variances $Ex_n^2 = \sigma_n^2$ are finite. Put $s_n^2 = \sigma_1^2 + \cdots + \sigma_n^2$, and suppose that the variables obey the central limit law with $a_n = s_n$, $b_n = 0$. (In particular this will be the case when all x_n have the same distribution.) For any function $\psi(n)$ tending to infinity with n we then have

$$\lim_{n\to\infty} P(|z_n| > s_n\psi(n)) = 0. \tag{12.1}$$

On the other hand, if $\psi(n)$ tends to a finite limit > 0, the same probability has a positive limit.

It seems natural to consider the relation within the brackets in (12.1) not only for a single large value of n, but to require the probability that this relation holds simultaneously for *an infinite number of values of n*. The development of this problem has led to the so called law of the iterated logarithm.

We shall in this respect use the following terminology due to Lévy. A non-decreasing positive function $\psi(n)$ will be said to belong to the *lower class* with respect to the variables x_n if, with a probability equal to one, there are infinitely many n such that

$$| z_n | > s_n \psi(n).$$

On the other hand, $\psi(n)$ will be said to belong to the *upper class* if the probability of the same property is equal to zero.

Every $\psi(n)$ belongs to one of these two classes. This is a special case of the so called *null-or-one law*: if S is a Borel set in the space of the independent random variables $x_1, x_2, \cdots$, such that any two points differing at most in a finite number of coordinates either both belong to S or both belong to the complementary set, then $P(S)$ can only assume the values 0 or 1.

It was proved by Kolmogoroff [51] that, subject to certain restrictions, the function

$$\psi(n) = \sqrt{c \log \log s_n}$$

belongs to the lower class for any $c < 2$, and to the upper class for any $c > 2$, which may be expressed by the relation

$$P\left(\limsup \frac{z_n}{s_n \sqrt{2 \log \log s_n}} = 1\right) = 1. \tag{12.1}$$

More general results were proved by Feller [30], who proved i.a. that, subject to certain restrictions, $\psi(n)$ belongs to the lower or upper class according as

$$\Sigma \frac{\sigma_n^2}{s_n^2} \psi(n) e^{-(\psi^2(n)/2)} \tag{12.2}$$

is divergent or convergent (in certain special cases, this had been previously found by Kolmogoroff and Erdös [24]. Feller also proved a more complicated result, which contains the above as a particular case, and from which it follows that the simple criterion (12.2) no longer holds when the restrictions imposed in its proof are removed.

13. Convergence of series. For any sequence of random variables x_n, the probability

$$P\left(\sum_1^\infty x_n \text{ converges}\right)$$

has a uniquely determined value. When the x_n are independent, it follows from the null-or-one law that this probability is either 0 or 1. By a theorem of Khintchine and Kolmogoroff [48], the value 1 is assumed when and only when the three series

$$\sum_n \int_{|x_n|>1} dF_n, \qquad \sum_n Ey_n, \qquad \sum_n \sigma^2 y_n$$

are convergent, where

$$y_n = \begin{matrix} x_n & \text{when} & |x_n| \leqq 1. \\ 0 & \text{when} & |x_n| > 1. \end{matrix}$$

For the case when the x_n are not assumed to be independent, various results have been given by Lévy [63] and others, but our knowledge of the properties of these series is still not very advanced.

14. Generalizations. In several instances it has been pointed out above that the results concerning sums of independent variables may, to a certain extent, be extended to cases when the variables are not independent. Generally the independence condition has then to be replaced by some condition restricting the degree of dependence. Results of this type were first give by Bernstein [3], and then in more general cases by Lévy [62], [63], and Loève [67]. However, this field has so far only been very incompletely explored.

Similar remarks apply to the generalization of the various theorems quoted above to cases of variables and distributions in more than one dimension.

III. STOCHASTIC PROCESSES

15. The theory of random variables in a finite number of dimensions is able to deal adequately with practically all problems considered in classical probability theory. However, during the early years of the present century, there appeared in the applications various problems, where it proved necessary to consider probability relations bearing on infinite sequences of numbers, or even on functions of a continuous variable.

The mathematical set-up required for the study of such problems involves the introduction of probability distributions in spaces of random sequences or random functions (cf. 5 above). Generally, any process in nature which can be analyzed in terms of probability distributions in spaces of these types will be called a *stochastic process*. It is convenient to apply this name also to the probability distribution used for the study of the process. We shall thus say, e.g., that a certain random function $x(t)$ is attached to the stochastic process which is defined by the probability distribution of $x(t)$. In the majority of applications, the variable t will represent the time, and we shall often use a terminology directly referring to this case. However, there are also other types of problems in the applications (t may e.g. be a spatial variable in an arbitrary number of dimensions), and it is obvious that the purely mathematical problems connected with these classes of probability distributions will have to be considered quite independently of any concrete interpretation of the variable t or the funcion $x(t)$.

A well-known example of this type of problems is afforded by the Brownian movement. Let $x(t)$ be the abscissa at the time t of a small particle immersed in a liquid, and subject to molecular impacts. In every instant, the quantity $x(t)$ receives a random impulse, and the problem arises to study the behaviour of $x(t)$. According as we are content to consider $x(t)$ for a discrete sequence of t-points, say for $t = 0, 1, 2, \cdots$, or we wish to consider all positive values of t, we shall then have to introduce a probability distribution in the space of the random sequence $x(0)$, $x(1)$, $\cdots$, or in the space of the random function $x(t)$, where $t > 0$. We may then discuss such questions as the distribution of $x(t)$ for a given value of t, the joint and conditional distributions of $x(t)$ for two or more values of t, and, in the case of a continuous variable t, continuity, differentiability and other similar properties of the random function $x(t)$.

Wiener [82], [83] (cf. also Paley and Wiener [74]) was the first to give a rigorous treatment of this process. He proved in 1923 that it is possible to define a probability distribution in a suitably restricted functional space, such that the increment $\Delta x(t) = x(t + \Delta t) - x(t)$ is independent of $x(t)$ for any $\Delta t > 0$. With a probability equal to 1, the function $x(t)$ is continuous for all $t > 0$, and for any fixed $t > 0$, the random variable $x(t)$ is normally distributed.

Another example of stochastic processes studied at this stage occurs in the theory of risk of an insurance company. Let $x(t)$ denote the total amount of claims up to the time t in a certain insurance company. As in the case of the Brownian movement, it may seem natural to assume that the increment $\Delta x(t)$ is independent of $x(t)$. On the other hand, $x(t)$ is in this case an essentially discontinuous function, which is never decreasing, and increases only by jumps of varying magnitudes occurring for certain discrete values of t, which are not a priori known. Processes of this type were studied by F. Lundberg [69], [70], H. Cramér [6] and others.

Further examples of particular processes were discussed in connection with various applications, but no general theory of the subject existed until 1931, when Kolmogoroff published a basic paper [53] dealing with the class of stochastic processes which will here be denoted as Markoff processes (Kolmogoroff uses the term "stochastically definite processes"), of which the two examples mentioned above form particular cases. The theory of this class of processes was further developed by Feller [26], [28]. In 1934, Khintchine [42] introduced another important class of processes known as stationary processes. From 1937, the general theory of the subject was subjected to a penetrating analysis in a series of important works by Doob [18]–[22].[3]

16. Probability distributions in functional spaces. We have seen in 5 above how a probability distribution in the space of all functions $x(t)$ may be defined, when t varies in an arbitrary space T. Generally, we shall here content ourselves to consider the cases when T is the set of all real numbers, or the set of all non-negative real numbers. Most results obtained for these cases will be readily generalized to cases when t varies in a Euclidean space of a finite number of dimensions. On the other hand, when T is enumerable, say consisting of the points $t = 0, \pm 1, \pm 2, \cdots$, so that we are concerned with a random sequence $x(0)$, $x(\pm 1)$, $\cdots$, the results for the continuous case will generally hold and assume a simpler form which will not be particularly stated here.

[3] A further interesting paper by Doob has appeared while the present paper was being printed: "Probability in function space", *Bull. Amer. Math. Soc.*, Vol. 53 (1947), pp. 15–30.

The case when T is a space of an infinite number of dimensions does not seem to have been considered so far.

In the present paragraph, it will be convenient to assume the function $x(t)$ to be real-valued, but the generalization to a complex-valued $x(t)$ requires only obvious modifications. In the sequel we shall sometimes consider the real-valued and sometimes the complex-valued case, according as the occasion requires.

Let now X be the space of all real-valued functions $x(t)$ of the real variable t, where $-\infty < t < \infty$. According to 5, a probability measure $P(S)$ is uniquely defined for all Borel sets S in X by means of the family of joint distributions of all finite sequences $x(t_1), \cdots, x(t_n)$. In fact, $P(S)$ can be defined for a more general class of sets than the Borel sets. For any set S in X, we may define an outer P-measure $\overline{P}(S)$ as the lower bound of $P(Z)$ for all sums Z of finite or enumerable sequences of intervals, such that $S \subset Z$. Further, the inner P-measure $\underline{P}(S)$is defined by the relation $\underline{P}(S) = 1 - \overline{P}(X - S)$. When the outer and inner measures are equal, S is called P-measurable, and $P(S)$ is defined as their common value. Any P-measurable set differs from a Borelset by a set of P-measure zero.

In many cases, this definition will be sufficient for an adequate treatment of the problems that we wish to consider. However, in other cases we encounter certain characteristic difficulties, which make it desirable to consider the possibility of amending the basic definition. Thus it often occurs that we are interested in the probability that the random function $x(t)$ satisfies certain regularity conditions in a non-enumerable set of points t. We may, e.g., wish to consider the probability that $x(t)$ is continuous for all t, that $x(t)$ should be Lebesque-measurable for all t, that $x(t) \leqq k$ for all t, etc. Let S denote the set of all functions satisfying a condition of this type. It can then be shown that the inner measure $\underline{P}(S)$ is always equal to zero so that S is never measurable, except in the (usually trivial) case when $P(S) = 0$.

Consequently many interesting probabilities are left undetermined by the general definition of a probability distribution in X given above. The possibility of modifying the definition so as to enable us to study probabilities of this type has been thoroughly investigated by Doob [18]. He considers a subspace X_0 of the general functional space X, where X_0 is chosen so as to contain only, or almost only, "desirable" functions, i.e. functions satisfying such regularity conditions as seem natural with respect to the problem under investigation. We start from a given probability measure $P(S)$ in X, and ask if it is possible to define a probability measure in the restricted space X_0, which corresponds in some natural way to the given distribution in X. Let S_0 be a set in X_0, and suppose that it is possible to find a P-measurable set S in X such that $SX_0 = S_0$. According to Doob, a probability measure P_0 in X_0 is then uniquely defined by the relation

$$P_0(S_0) = P(S)$$

if and only if the condition

$$\overline{P}(X_0) = 1$$

is satisfied.

The problem is thus reduced to finding a subspace X_0 of outer P-measure 1, such that X_0 contains only functions of sufficiently regular behaviour. When this can be done, we can restrict ourselves to consider only functions $x(t)$ belonging to X_0, the probability distribution in this space being defined by the measure P_0. We shall then say that $x(t)$ is a random function, attached to a stochastic process with the restricted space X_0. Doob has obtained a great number of interesting results in this connection, e.g. with respect to the problem of choosing X_0 such that it contains almost only Lebesque-measurable functions, or such that the probability of the relation $x(t) \leqq k$ has a well-defined value for all k. In particular he has shown that the last problem can be solved for any given P-measure. However, our knowledge of the various possibilities which exist with respect to the choice of X_0 is still very incomplete, and it seems likely that further important results may be reached along this line of research.

An alternative method of introducing probability distributions in functional spaces has been used by Wiener [82], [83], (cf. also Paley and Wiener, [74]). Consider a given probability measure Π in an arbitrary space Ω, defined for all sets Σ of an additive class C. Let $x(t, \omega)$ denote a function (real- or complex-valued, as the case may be) of the arguments t (real) and ω (point in Ω), such that $x(t, \omega)$ for every fixed t becomes a C-measurable function of ω. On the other hand, when ω is fixed, $x(t, \omega) = x(t)$ reduces to a function of the real variable t. Let X_0 denote the set of all functions $x(t)$ corresponding in this way to points of Ω. Further, let $S_0 = SX_0$, where S is a Borel set in X, and let Σ denote the set of all points ω such that $x(t, \omega) \subset S_0$. Then Σ belongs to C, and a probability measure P_0 in the functional space X_0 is uniquely defined by the relation

$$P_0(S_0) = \Pi(\Sigma). \tag{16.1}$$

The relations between the two modes of definition have been discussed by Doob and Ambrose [23] who have shown that they are largely equivalent. However, it seems likely that in particular problems the one or the other procedure may sometimes be the more advantageous, and further investigations on this subject seem desirable.

17. Processes with a finite mean square. Consider a stochastic process defined by a probability measure $P(S)$ in the space X of all complex-valued functions $x(t)$ of the real variable t. For any fixed t_0, the random variable $x(t_0)$ is then a complex-valued function of the variable point $x(t)$ in the space X, i.e. a point Q_{t_0} in the space Ω of all complex-valued functions defined on X. When t_0 varies, the point Q_{t_0} describes a "curve" in Ω, which then corresponds to our stochastic process.

Suppose, in particular, that the mean square

$$E \mid x(t) \mid^2 = \int_X \mid x(t) \mid^2 dP$$

is finite for any fixed value of t. This implies that for fixed t the function $x(t)$ belongs to L_2 over X, relative to the probability measure P. The random variable $x(t)$ may then be regarded as an element of the Hilbert space H of all complex-valued functions f belonging to L_2 over X, the inner product (f, g) of two elements f and g being defined by the relation

$$(f, g) = \int_X f\bar{g}\, dP = E(f\bar{g}).$$

The stochastic process to which $x(t)$ is attached then corresponds to a "curve" in H (Kolmogoroff, [56], [57]), so that the well-known theory of Hilbert space is available for the study of the process. In particular, convergence in the usual metric of Hilbert space is equivalent to convergence in the mean of order 2 for random variables.

Let H_x be the smallest closed linear subspace of H which contains all elements of the form $a_1x(t_1) + \cdots + a_nx(t_n)$. If the *covariance function*

$$r(t, u) = (x(t), x(u)) = E(x(t)\overline{x(u)})$$

is continuous for all real values of t and u, then $x(t) \rightarrow x(t_0)$ in the mean, as $t \rightarrow t_0$, and we shall say that the process $x(t)$ is *continuous*. For any continuous process, H_x is separable. When $g(t)$ is a continuous non-random function of t, and $x(t)$ is attached to a continuous stochastic process, the Riemann-Darboux sums formally associated with the integral

$$\int_a^b g(t)x(t)\, dt$$

are easily shown to tend to a limit y, which is an element of H_x, i.e. a random variable. By definition, we may identify the integral with this variable y, and this integral will possess the essential properties of the ordinary Riemann integral (Cramér, [12]).

The application of the theory of Hilbert space to stochastic processes seems to open very interesting possibilities. Some applications to particular classes of stochastic processes will be mentioned below. Futher important results belonging to this order of ideas will be given in a work by K. Karhunen [40], which is in course of publication.

18. Relations to ergodic theory. There is a close connection between the theory of stochastic processes and ergodic theory. In ergodic theory, as summarized e.g. in the treatise of E. Hopf [38], we consider an arbitrary space Ω, and a probability measure Π, defined for all sets Σ belonging to the additive class C. We further consider a one-parameter group of one-one transformations of Ω into itself (a "flow" in Ω) such that the transformation corresponding to the parameter value t takes the point $\omega = \omega_0$ into ω_t, while $(\omega_t)_u = \omega_{t+u}$. Let $f(\omega)$ be a given function, defined throughout Ω, and such that $f(\omega_t)$ is C-measurable for every fixed t. The well-known ergodic theorems due to von Neumann, Birkhoff, Khintchine and others are then concerned with the asymptotic behaviours of mean values, which in the classical cases are of the types

$$\frac{f(\omega_0) + f(\omega_1) + \cdots + f(\omega_{n-1})}{n}$$

or

$$\frac{1}{T}\int_0^T f(\omega_t)\,dt,$$

as n or T tends to infinity. (In the case of the latter expression, it is necessary to introduce some additional condition implying measurability in t.)

Writing $x(t, \omega) = f(\omega_t)$, it is seen that to a given transformation group $\omega \rightarrow \omega_t$ and a given function $f(\omega)$, there corresponds a stochastic process in the sense of Wiener's definition (cf. 16). The space X_0 of this process consists of all functions $x(t)$ representable in the form $x(t) = f(\omega_t)$, when $\omega = \omega_0$ varies over Ω. The corresponding probability measure P_0 is defined by (16.1).

Thus any of the above-mentioned ergodic theorems may be expressed as a theorem concerning "temporal" mean values of the types

$$\frac{x(0) + x(1) + \cdots + x(n-1)}{n}$$

or

$$\frac{1}{T}\int_0^T x(t)\,dt.$$

If, according to some reasonable convergence definition, we may assign a limit to either of these expressions, as n or T tends to infinity, this limit will be a random variable, and it is important to find conditions which imply that this variable has a constant value for "almost all" functions $x(t)$, i.e. for all $x(t)$ except at most a set of P_0-measure zero.

In the particular case when $x(0)$, $x(1)$, $\cdots$ are independent variables all having the same distribution, the classical ergodic theorems yield simple cases of the laws of large numbers (cf. 10). The mean ergodic theorem of von Neumann gives the weak law, while the Birkhoff-Khintchine theorem gives the strong law. Some more general results belonging to this order of ideas will be mentioned in the sequel.

It will be seen that the two theories are largely equivalent, and it seems likely that further comparative studies of the methods will be of great value to both sides.

19. Markoff processes. Consider now a stochastic process, defined by a probability measure $P(S)$ in the space X of all real-valued functions $x(t)$ of the real variable t. For any $t_1 < t_2$, there is a certain conditional probability $P(x(t_2) \subset S \mid x(t_1) = a_1)$ of the relation $x(t_2) \subset S$, relative to the hypothesis that $x(t_1)$ assumes the given value a_1. Suppose now that this conditional probability is independent of any additional hypothesis concerning the behaviour of $x(t)$ for $t < t_1$, so that we have e.g. for any $t_0 < t_1 < t_2$ and for any a_0

$$P(x(t_2) \subset S \mid x(t_1) = a_1) = P(x(t_2) \subset S \mid x(t_1) = a_1, x(t_0) = a_0).$$

In this case the process is called a *Markoff process*.

The general theory of this type of processes, which forms a natural generalization of the classical concept of Markoff chains, has been studied in basic works by Kolmogoroff [53] and Feller [26], [28]. Writing

$$P(x(t) \leqq \xi \mid x(t_0) = a_0) = F(\xi;\, t,\, a_0,\, t_0),$$

where $t_0 < t$, F will be the distribution function of the random variable $x(t)$, relative to the hypothesis $x(t_0) = a_0$. Then F satisfies the Chapman-Kolmogoroff equation

$$F(\xi; t, a_0, t_0) = \int_{-\infty}^{\infty} F(\xi; t, \eta, t_1)\, d_\eta F(\eta; t_1, a_0, t_0), \tag{19.1}$$

which expresses that, starting from the state $x(t_0) = a_0$, the state $x(t) \leqq \xi$ must be reached by passing through some intermediate state $x(t_1) = \eta$, where $t_0 < t_1 < t$. Subject to certain general conditions, it is possible to show that any solution of this equation satisfies certain integro-differential equations, which in some important cases reduce to partial differential equations of parabolic type, and that the d.f. F is uniquely determined by these equations. However, the general conditions mentioned above are in many cases difficult to apply to particular classes of processes, and it would be important to have further investigations concerning these questions.

Markoff processes (not belonging to the subclass of differential processes, which will be considered in the following paragraph) appear in several important applications, e.g. in the theory of cosmic radiation, in certain genetical problems, in the theory of insurance risk etc. In these cases, we are often concerned with the class of *purely discontinuous* Markoff processes, where the function $x(t)$ only changes its value by jumps. If, in addition, there are only a finite or enumerable set of possible values for $x(t)$, the Chapman-Kolmogoroff equation (19.1) reduces to

$$\pi_{ik}(t_0, t) = \sum_j \pi_{ij}(t_0, t_1)\pi_{jk}(t_1, t), \tag{19.2}$$

where $\pi_{ik}(t_0, t)$ denotes the "transition probability", i.e. the probability that $x(t)$ will be in the kth state at the time t, when it is known to have been in the ith state at the time t_0. In matrix form, this equation may be written

$$\Pi(t_0, t) = \Pi(t_0, t_1)\Pi(t_1, t), \tag{19.3}$$

where Π denotes the matrix of the π_{ik}.

When only a sequence of discrete values of t are considered, we have here the classical case of Markoff chains, which has received a detailed treatment in the well-known book by Fréchet [32] (cf. also Doob, [19]). The case when t is a continuous variable has been treated by Feller [28], O. Lundberg [71], Arley [1], and other authors. Some of the most important problems of this branch of the subject are concerned with the existence of a unique system of solutions of (19.2) or (19.3), and with the asymptotic behaviour of the solutions for large values of $t - t_0$. Though important results have been reached, there still remains much to be done here, and the same thing holds a fortiori with respect to the analogous problems for general Markoff processes.

20. Differential processes. A particularly interesting case of a Markoff process arises when, for any $\Delta t > 0$, the increment $\Delta x(t) = x(t + \Delta t) - x(t)$ is independent of $x(\tau)$ for $\tau \leqq t$. The process is then called a *differential process*. Some of the earliest studied stochastic processes belong to this class, which contains in particular the two examples discussed above in 15. Further cases of such processes arise e.g. in the theory of radioactive disintegration and in telephone technique.

Let us suppose that $x(0)$ is identically equal to zero, and that the process is uniformly continuous in probability in every finite interval $0 \leqq t \leqq T$, i.e. that for any fixed positive ϵ

$$P(\,|\, x(t + \Delta t) - x(t) \,|\, > \epsilon) \rightarrow 0$$

as $\Delta t \rightarrow 0$, uniformly for $0 \leqq t \leqq T$. Then it follows from the works of Lévy, [60], [63], Khintchine [47] and Kolmogoroff [54] that, for any $t > 0$, the random variable $x(t)$ has an infinitely divisible distribution, with a characteristic function $\varphi(z; t)$ given by (9.2), where β, γ, $M(u)$ and $N(u)$ may depend on t.

In the particularly important case when the distribution of the increment $x(t + \Delta t) = x(t)$ does not involve t, but only depends on the length Δt of the interval, we say that the process is *temporally homogeneous*, and in this case we have

$$\log \varphi(z; t) = t \log \varphi(z; 1),$$

so that we obtain the general formula for $\varphi(z; t)$ simply by replacing in (9.2) β, γ, $M(u)$ and $N(u)$ by $t\beta$, $t\gamma$, $tM(u)$ and $tN(u)$ respectively.

When $t \rightarrow \infty$, or $t \rightarrow 0$, the appropriately normalized distribution of $x(t)$ tends, under certain conditions, to a stable distribution (Cramér [7], Gnedenko [36]). When this limiting distribution is normal, there are sometimes even asymptotic expansions analogous to (11.3). Still, the problem of the asymptotic behaviour of the distribution for large t does not seem to be definitely cleared up.

Khintchine [41] and Gnedenko [37] have given interesting generalizations of the law of the iterated logarithm (cf. 12) to processes of the type considered here.

The continuous process discussed in 15 in connection with the Brownian movement corresponds to the temporally homogeneous case when β, $M(u)$ and $N(u)$ all reduce to zero, so that

$$\varphi(z) = e^{-\gamma t z^2},$$

which shows that the distribution of $x(t)$ is normal, with mean zero and variance $2\gamma t$.

On the other hand, in the applications to the theory of insurance risk, γ is zero, while $M(u)$ and $N(u)$ are connected with the distribution of the various magnitudes of claims. In this type of applications, it is often very important to find the probability that $x(t)$ satisfies an inequality of the form

$$x(t) < a + bt$$

for all values of t. It follows from the discussion in 16 that the definition of a probability of this type is somewhat delicate. The problem, which can be regarded as an extended form of the classical problem of "the gambler's ruin," has been solved in certain particular cases. It leads to integral equations, which in the simplest case are of the Volterra, in other cases of the Wiener-Hopf type (Cramér [6], [13], Segerdahl [79], Täcklind [81]).

21. Orthogonal processes. Consider now the case of a complex-valued $x(t)$, and suppose that $E \mid x(t) \mid^2$ is finite for all t. Without restricting the generality, we may assume that $Ex(t) = 0$ for all t.

Suppose now that instead of requiring, as in the case of a differential process, that the variables $x(\tau)$ and $\Delta x(t)$ should be *independent* when $\tau \leqq t$, we only lay down the less stringent condition that these variables should be *non-correlated*, i.e. that

$$E(x(\tau)\overline{\Delta x(t)}) = 0.$$

We then obtain a process which is no longer necessarily of the Markoff type. The condition implies that, for any two disjoint intervals (t_1, t_2) and (t_3, t_4), we have

$$E[(x(t_2) - x(t_1))(\overline{x(t_4)} - \overline{x(t_3)})] = 0,$$

so that the "chords" corresponding to two disjoint "arcs" of the curve in Hilbert space representing the process are always orthogonal (Kolmogoroff [56], [57]). A process of this type may accordingly be called an *orthogonal process.*

For a process of this type we have, writing $E \mid x(t) \mid^2 = F(t)$, $F(t + \Delta t) - F(t) = E \mid x(t + \Delta t) - x(t) \mid^2$, so that $F(t)$ is a never decreasing function of t. If $F(t)$ is bounded for all t, we shall say that the orthogonal process is bounded. For a bounded orthogonal process, the Stieltjes integral

$$\int_{-\infty}^{\infty} g(t)\, dx(t),$$

where $g(t)$ is bounded and continuous, may be defined as the limit in the mean of sums of the form

$$\sum_{\nu} g(t_\nu)(x(t_\nu) - x(t_{\nu-1})).$$

22. Stationary processes. When we are concerned with a process representing the temporal development of a system governed by laws which are invariant under a translation in time, it seems natural to assume that the joint distribution of any group of variables of the form

$$x(t_1 + \tau), \cdots, x(t_n + \tau) \tag{22.1}$$

is independent of τ. A process satisfying this condition will be called a *stationary* process. If a stochastic process is defined by means of a "flow" $\omega \to \omega_t$ in a space Ω (cf. 18), the process will be stationary when and only when the corresponding flow is *measure-preserving*, i.e. if the transformation $\omega \to \omega_t$ changes any C-measurable set S into a set S_t of the same measure.

Under appropriate conditions with respect to the measurability of $x(t)$, the Birkhoff-Khintchine ergodic theorem holds for a stationary process, i.e. there exists a random variable y such that we have

$$P_0\left(\lim_{T\to\infty} \frac{1}{T}\int_0^T x(t)\,dt = y\right) = 1, \tag{22.2}$$

where P_0 is the probability measure in a suitably restricted space in the sense of Doob. Further work seems to be required here, in order to make the situation quite clear, also with regard to metric transitivity.

For a stationary process, any finite moment of the joint distribution of the variables (22.1) is obviously independent of τ. Suppose now that we only require that this invariance under translations in time should hold for moments of the first and second order of the joint distributions, which are assumed to be finite. The wider class of processes obtained in this way may be called *stationary of the second order*. Processes of this type have been studied for the first time by Khintchine [42]. We shall assume that $x(t)$ is complex-valued. Without restricting the generality, we may further assume that $Ex(t) = 0$ for all t. The product moment $E(x(t)\overline{x(u)})$ will then be a function of the difference $t - u$:

$$E(x(t)\overline{x(u)}) = R(t - u). \tag{22.3}$$

Assuming, in addition, that $R(t)$ is continuous at $t = 0$, it follows that $R(t)$ is continuous for all t, and the process is continuous in the sense of 17. It was shown by Khintchine that a NS condition that a given function $R(t)$ should be associated with a second order stationary and continuous process by means of the relation (22.3) is that we should have

$$R(t) = \int_{-\infty}^{\infty} e^{itx}\,dF(x) \tag{22.4}$$

for all t, where the spectral function $F(x)$ is real, never decreasing and bounded. In particular, we have

$$F(+\infty) - F(-\infty) = R(0) = E \mid x(t) \mid^2 = \sigma^2.$$

Khintchine's condition for $R(t)$ was generalized by Cramér to the case of an arbitrary number of processes $x_1(t), \cdots, x_n(t)$, such that the product moments $E(x_i(t)\overline{x_j(u)})$ are functions of the difference $t - u$. The corresponding spectral functions $F_{ij}(x)$ are in general complex-valued and of bounded variation. Further, the expression (Cramér, [12])

$$\sum_{i,j=1}^{n} z_i \bar{z}_j \Delta F_{ij},$$

where $\Delta F_{ij} = F_{ij}(b) - F_{ij}(a)$ is, for any $a < b$, a non-negative Hermite form in the variables z_i. This result is closely connected with a theorem on Hilbert space considered by Kolmogoroff and Julia. It is further shown that, to any given functions $F_{ij}(x)$, $(i, j = 1, \cdots, n)$, satisfying these conditions, we can always find n processes $x_1(t), \cdots, x_n(t)$ such that the joint distribution of any set of variables $x_i(t_j)$ is always *normal*, while the covariance functions $R_{ij}(t - u) = E(x_i(t)\overline{x_j(u)})$ are given by the expression

$$R_{ij}(t) = \int_{-\infty}^{\infty} e^{itx}\, dF_{ij}(x).$$

For a process $x(t)$ which is continuous and stationary of the second order, with $Ex(t) = 0$ for all t, we have the mean ergodic theorem

$$\text{l.i.m.}_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-\lambda it} x(t)\, dt = y \tag{22.5}$$

for any real λ. The random variable y has the mean 0 and the variance $F(\lambda + 0) - F(\lambda - 0)$, where F is the spectral function appearing in (22.4). If λ is a point of continuity for F, it thus follows that $y = 0$ with a probability equal to 1. On the other hand, if λ is a discontinuity, y has a positive variance. Let $\lambda_1, \lambda_2, \cdots$ be all the discontinuities of $F(x)$, and let $\sigma_1^2, \sigma_2^2, \cdots$ be the corresponding saltuses, while $y_1, y_2, \cdots$ are the limits in the mean obtained from (22.5) for $\lambda = \lambda_1, \lambda_2, \cdots$. Then two different y_j are always orthogonal: $E(y_j\bar{y}_k) = 0$ for $j \neq k$, and we have

$$x(t) = \sum_{\nu} y_\nu e^{\lambda_\nu it} + \xi(t), \tag{22.6}$$

where $E\xi(t) = 0$ and

$$E \mid \xi(t) \mid^2 = \sigma^2 - \sum_{\nu} \sigma_\nu^2.$$

If $F(x)$ is a step-function, we have $\sigma^2 = \sum_\nu \sigma_\nu^2$, and it follows that $\xi(t) = 0$ with a probability equal to 1, so that (22.6) gives a "stochastic Fourier expansion" of $x(t)$ (Slutsky, [80]).

Even when $F(x)$ is arbitrary, we can obtain a spectral representation of $x(t)$ generalizing (22.6). In fact, it can be shown (Cramér, [14]) that $x(t)$ can always be represented by a Fourier-Stieltjes integral

$$x(t) = \int_{-\infty}^{\infty} e^{itu} \, dz(u), \tag{22.7}$$

where $z(u)$ is a random function attached to a bounded orthogonal process (cf. 21), such that

$$E \mid z(u + \Delta u) - z(u) \mid^2 = F(u + \Delta u) - F(u).$$

Conversely, we have

$$z(u + \Delta u) - z(u) = -\int_{-\infty}^{\infty} \frac{e^{-it(u+\Delta u)} - e^{-itu}}{2\pi it} x(t) \, dt, \tag{22.8}$$

so that there is a one-one correspondence between $x(t)$ and $\Delta z(u)$. The integrals (22.7) and (22.8) are defined as limits in the mean, as shown above in 17 and 21. These results are in close correspondence with generalized harmonic analysis for an arbitrary function, as developed by Wiener [83] and Bochner [4]. The spectral representation of a stochastic process has important applications, some of which will be considered in a forthcoming paper by Karhunen [40]. An extension of the spectral representation to a more general class of processes has been given by Loève [68].

When, in particular, the $x(t)$ process is such that the joint distribution of any group of variables $x(t_1), \cdots, x(t_n)$ is normal, it follows that any increment $\Delta z(u)$ is normally distributed. Since two uncorrelated normally distributed variables are always independent, it follows that in this case the $z(u)$ process is a differential process with normally distributed increments. Important results for this case have recently been given by Doob [22].

The properties of continuity, differentiability etc. for processes of the type here considered are still incompletely known, and further work is required. A further group of important unsolved problems are connected with an interesting decomposition theorem by Wold [84], which holds for processes with a discrete time variable. The generalization of this theorem to the continuous case does not seem to have so far been given in a final form.

REFERENCES

[1] N. Arley, "On the theory of stochastic processes and their applications to the theory of cosmic radiation," Thesis, Copenhagen, 1943.

[2] G. M. Bawly, "Ueber eine Verallgemeinerung der Grenzwertsätze der Wahrscheinlichkeitsrechnung," *Rec. Math.* (*Mat. Sbornik*), *N. S.*, Vol. 1 (1936), pp. 917–929.

[3] S. N. Bernstein, "Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes," *Math. Ann.*, Vol. 97 (1927), pp. 1–59.

[4] S. Bochner, "Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse," *Math. Ann.*, Vol. 108 (1933), pp. 378–410.

[5] H. Cramér, "On the composition of elementary errors," *Skand. Aktuarietidskr.*, Vol. 11 (1928), pp. 13–74, 141–180.

[6] ———, "On the mathematical theory of risk." Published by the Insurance Company Skandia, Stockholm, 1930.
[7] ———, "Sur les propriétés asymptotiques d'une classe de variables aléatoires," *C. R. Acad. Sci. Paris*, Vol. 201 (1935), pp. 441–443.
[8] ———, "Ueber eine Eigenschaft der normalen Verteilungsfunktion," *Math. Zeit.*, Vol. 41 (1936), pp. 405–414.
[9] ———, *Random variables and probability distributions*, Cambridge Tracts in Math., Cambridge, 1937.
[10] ———, "Sur un nouveau théorème—limite de la théorie des probabilités," *Actualités Scientifiques*, Paris, No. 736 (1938), pp. 5–23.
[11] ———, "On the representation of a function by certain Fourier integrals," *Trans. Amer. Math. Soc.*, Vol. 46 (1939), pp. 191–201.
[12] ———, "On the theory of stationary random processes," *Ann. of Math.*, Vol. 41 (1940), pp. 215–230.
[13] ———, "Deux conférences sur la theorie des probabilités," *Skand. Aktuarietidskr.*, 1941, pp. 34–69.
[14] ———, "On harmonic analysis in certain functional spaces," *Ark. Mat. Astr. Fys.*, Vol. 28B (1942), pp. 1–7.
[15] ———, *Mathematical Methods of Statistics*. Princeton Univ. Press, Princeton, 1946.
[16] W. Doeblin, "Premiers éléments d'une étude systématique de l'ensemble de puissances d'une loi de probabilités," *C. R. Acad. Sci. Paris*, Vol. 206 (1938), pp. 306–308.
[17] ———, "Sur un théorème du calcul des probabilités," *C. R. Acad. Sci. Paris*, Vol. 209 (1939), pp. 742–743.
[18] J. L. Doob, "Stochastic processes depending on a continuous parameter," *Trans. Amer. Math. Soc.*, Vol. 42 (1937), pp. 107–140.
[19] ———, "Stochastic processes with an integral-valued parameter," *Trans. Amer. Math. Soc.*, Vol. 44 (1938), pp. 87–150.
[20] ———, "Regularity properties of certain families of chance variables," *Trans. Amer. Math. Soc.*, Vol. 47 (1940), pp. 455–486.
[21] ———, "The law of large numbers for continuous stochastic processes," *Duke Math. Jour.*, Vol. 6 (1940), pp. 290–306.
[22] ———, "The elementary Gaussian processes," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 229–282.
[23] J. L. Doob and M. Ambrose, "On the two formulations of the theory of stochastic processes depending upon a continuous parameter," *Ann. of Math.*, Vol. 41 (1940), pp. 737–745.
[24] P. Erdös, "On the law of the iterated logarithm," *Ann. of Math.*, Vol. 43 (1942), pp. 419–436.
[25] W. Feller, "Ueber den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung," *Math. Zeit.*, Vol. 40 (1935), pp. 521–559.
[26] ———, "Zur Theorie der stochastischen Processe (Existenz- und Eindeutigkeitssatze)," *Math. Ann.*, Vol. 113 (1936), pp. 113–160.
[27] ———, "Ueber das Gesetz der Grossen Zahlen," *Acta. Univ. Szeged.*, Vol. 8 (1937), pp. 191–201.
[28] ———, "On the integro-differential equations of purely discontinuous Markoff processes," *Trans. Amer. Math. Soc.*, Vol. 48 (1940), pp. 488–575.
[29] ———, "Generalization of a probability limit theorem of Cramér," *Trans. Amer. Math. Soc.*, Vol. 54 (1943), pp. 361–372.
[30] ———, "The general form of the so-called law of the iterated logarithm," *Trans. Amer. Math. Soc.*, Vol. 54 (1943), pp. 373–402.
[31] ———, "The fundamental limit theorems in probability," *Bull. Amer. Math. Soc.*, Vol. 51 (1945), pp. 800–832.

[32] M. FRÉCHET, *Recherches théoriques modernes sur la théorie des probabilités*, Vol. 2, Paris, 1937.

[33] B. V. GNEDENKO, "Sur les fonctions caractéristiques," *Bull. Math. Univ. Moscou*, Vol. 1, (1937), pp. 16–17.

[34] ———, "On the theory of the domains of attraction of stable laws," *Učenye Zpiski, Moskovskoga gosudaistvenogo Univaziteta*, Vol. 30 (1939), pp. 61–81.

[35] ———, "On the theory of limit theorems for sums of independent random variables," *Bull. Acad. Sci. URSS*, Vol. 30 (1939), pp. 181–232, 643–647.

[36] ———, "On locally stable probability distributions," *C. R. Acad. Sci. URSS*, Vol. 35 (1942), pp. 263–266.

[37] ———, "Investigation on the growth of homogeneous random processes," *C. R. Acad. Sci. URSS*, Vol. 36 (1942), pp. 3–41.

[38] E. HOPF, *Ergodentheorie*, Ergebnisse der Mathematik, Vol. 5, No. 2, Berlin, 1937.

[39] P. L. HSU, "The approximate distribution of the mean and of the variance of independent variates," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 1–29.

[40] K. KARHUNEN, Paper on stochastic processes, to appear in the *Acta Soc. Sci. Fennicae.*

[41] A. KHINTCHINE. *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, Vol. 2, No. 4, Berlin, 1933.

[42] ———, "Korrelationstheorie der stationären stochastischen Prozesse," *Math. Ann.*, Vol. 109 (1934), pp. 604–615.

[43] ———, "Sul dominio di attrazione della legge di Gauss," *Giorn. Ist. Ital. Attuari*, Vol. 6 (1935), pp. 378–393.

[44] ———, "Su una legge dei grandi numeri generalizzata," *Giorn. Ist. Ital. Attuari*, Vol. 7 (1936), pp. 365–377.

[45] ———, "Zur Kennzeichnung der characteristischen Funktionen," *Bull. Math. Univ. Moscou*, Vol. 1 (1937), pp. 1–31

[46] ———, "Contribution à l'arithmetique des lois de distribution," *Bull. Math. Univ. Moscou*, Vol. 1 (1937), pp. 6–17.

[47] ———, "Zur Theorie der unbeschränkt teilbaren Verteilungsgesetze," *Rec. Math. N. S.*, Vol. 2 (1937), pp. 79–117.

[48] A. KHINTCHINE AND A. KOLMOGOROFF, "Ueber Konvergenz von Reihen, deren Glieder durch den Zufall bestimmt werden," *Rec. Math.*, Vol. 32 (1925), pp. 668–677.

[49] A. KHINTCHINE AND P. LÉVY, "Sur les lois stables," *C. R. Acad. Sci. Paris*, Vol. 202 (1936), pp. 374–376.

[50] A. KOLMOGOROFF, "Ueber die Summen durch den Zufall bestimmter unabhängiger Grössen," *Math. Ann.*, Vol. 99 (1928) and Vol. 102 (1929), pp. 484–489.

[51] ———, "Ueber das Gesetz des iterierten Logarithmus," *Math. Ann.*, Vol. 101 (1929), pp. 126–135.

[52] ———, "Sur la loi forte des grands nombres," *C. R. Acad. Sci. Paris*, Vol. 191 (1930), pp. 910–911.

[53] ———, "Ueber die analytischen Methoden der Wahrscheinlichkeitsrechnung," *Math. Ann.*, Vol. 104 (1931), pp. 415–458.

[54] ———, "Sulla forma generale di un processo stocastico omogeneo," *Atti. Acad. naz. Lincei, Rend.*, Vol. 15 (1932), pp. 805–828, 866–869.

[55] ———, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, Vol. 2, No. 3, Berlin, 1933.

[56] ———, "Wiener Spiralen und einige andere interessante Kurven im Hilbertschen Raum," *C. R. Acad. Sci. URSS*, Vol. 26 (1940), pp. 115–118.

[57] ———, "Kurven im Hilbertschen Raum, die gegenüber einer ein parametrigen Gruppe von Bewegungen invariant sind," *C. R. Acad. Sci. URSS*, Vol. 26 (1940), pp. 6–9.

[58] P. LÉVY, *Calcul des probabilités*, Paris, 1925.

[59] ———, "Sur les séries dont les termes sont des variables éventuelles indépendantes," *Studia Math.*, Vol. 3 (1931), pp. 117–155.

[60] ———, "Sur les intégrales dont les éléments sont des variables aléatoires indépendantes," *Ann. Scuola norm. super. Pisa, (2)*, Vol. 3 (1934), pp. 337–366.

[61] ———, "Propriétés asymptotiques des sommes de variables aléatoires indépendantes," *Ann. Scuola norm. super. Pisa (2)*, Vol. 3 (1934), pp. 347–402.

[62] ———, "La loi forte des grands nombres pour les variables aléatoires enchainées," *Jour. Math. Pures Appl.*, Vol. 15 (1936), pp. 11–24.

[63] ———, *Théorie de l'addition des variables aléatoires*, Paris, 1937.

[64] ———, "L'arithmetique des lois de probabilités," *Jour. Math. Pures Appl.*, Vol. 17 (1938), pp. 17–39.

[65] A. M. Liapounoff, "Nouvelle forme du théorème sur la limite de la probabilité," *Mémoires Acad. Saint-Petersbourg s. 8*, Vol. 12 (1901).

[66] J. W. Lindeberg, "Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung," *Math. Zeft.*, Vol. 15 (1922), pp. 211–225.

[67] W. Loéve, "Étude asymptotique des sommes de variables aléatoires liées," *Jour. Math. Pures Appl.*, Vol. (9) 24 (1945), pp. 249–318.

[68] ———, "Analyse harmonique générale d'une fonction aléatoire," *C. R. Acad. Sci. Paris*, Vol. 220 (1945), pp. 380–382.

[69] F. Lundberg, "Zur Theorie der Rückversicherung," *Verhandlungen Kongr. f. Versicherungsmathematik*, Wien, 1909.

[70] ———, *Försäkringsteknisk riskutjämning*, Stockholm, 1927.

[71] O. Lundberg, "On random processes and their application to sickness and accident statistics," Thesis, Stockholm, 1940.

[72] J. Marcinkiewicz, "Sur une propriété de la loi de Gauss," *Math. Zeit.*, Vol. 4 (1938), pp. 612–618.

[73] J. Marcinkiewicz and A. Zygmund, "Sur les fonctions indépendantes," *Fund. Math.*, Vol. 29 (1937), pp. 60–90.

[74] R. E. A. C. Paley and N. Wiener, *Fourier Transforms in the Complex Domain*, Amer. Math. Soc. Colloquium Publ., Vol. 19, New York, 1934.

[75] D. Raikov, "On the composition of Poisson laws," *C. R. Acad. Sci. URSS*, Vol. 14 (1937), pp. 9–11.

[76] ———, "On the decomposition of Gauss and Poisson laws," *Bull. Acad. Sci. URSS, Ser. Math.*, 1938, pp. 91–120.

[77] ———, "On the composition of analytic distribution functions," *C. R. Acad. Sci. URSS*, Vol. 23 (1939), pp. 511–514.

[78] I. J. Schoenberg, "Metric spaces and positive definite functions," *Trans. Amer. Math. Soc.*, Vol. 44 (1938), pp. 522–536.

[79] C. O. Segerdahl, "On homogeneous random processes and collective risk theory," Thesis, Stockholm, 1939.

[80] E. Slutsky, "Sur les fonctions aléatoires presque périodiques et sur la décomposition des fonctions aléatoires stationaires en composantes," *Actualités Scientifiques*, No. 738, Paris, 1938, pp. 33–55.

[81] S. Täcklind, "Sur le risque de ruine dans des jeux inéquitables," *Sknad. Aktuarietidskr.*, 1942, pp. 1–42.

[82] N. Wiener, "Differential space," *Jour. Math. Phys. M. I. T.*, Vol. 2 (1923), pp. 131–172.

[83] ———, "Generalized harmonic analysis," *Acta Math.*, Vol. 55 (1930), pp. 117–258.

[84] H. Wold, "A study in the analysis of stationary time series," Thesis, Stockholm, 1938.