The Annals of Probability 1976, Vol. 4, No. 4, 509–546.

A CENTURY WITH PROBABILITY THEORY: SOME PERSONAL RECOLLECTIONS

By **HARALD CRAMÉR**
University of Stockholm

CONTENTS

We deeply appreciate Professor Cramér's gracious interest in presenting these personal recollections of the development of probability during the half-century, 1920–70. Professor Cramér has had a tremendous influence upon the development of both probability and statistics. His 1946 text alone, Mathematical Methods of Statistics [24], has so greatly influenced the studies and research of many during the past 30 years. During 1947–49, he served as Associate Editor of Ann. Math. Statist., the forerunner of this journal. He also served as a Council member of the IMS in 1953–56 and the Rietz Lecturer of the IMS in 1953.

AMS 1970 subject classification. Primary 60–03.
Key words and phrases. History, personal recollections.

1. Introduction

The previous Editor of this journal, Professor Ronald Pyke, has kindly suggested that I should write down some personal recollections from the period, approximately consisting of the half–century from 1920 to 1970, when I was actively engaged in work on mathematical probability theory. In attempting to follow this suggestion, I am anxious to make it quite clear that I shall not be concerned with writing a history of probability theory. I am only trying to give my own personal impressions of the work that was being done, and of some of the people who did it. I am going to deal mainly with the types of problems that attracted my personal interest, while many other lines of investigation, possibly even more interesting from a general point of view, will hardly be mentioned.

On account of this strongly subjective character of the paper, it may perhaps be convenient to begin with some brief notes of my personal background, and the reasons for my interest in probability theory.

I was born in Stockholm on September 25, 1893, as the second son of my parents, who were cousins. Our ancestors had for several centuries lived in the old town of Wisby, on the island of Gotland, but my parents lived in Stockholm, where my father was a banker.

When I started my academic studies at the University of Stockholm in 1912, my interest was fairly equally divided between chemistry and mathematics. My first University employment was as a research assistant in biochemistry, during the year immediately before World War I, and my first scientific publications belonged to that field. But I soon came to the conclusion that mathematics was the right subject for me. As my main mathematical teacher and friend, I was happy to have Marcel Riesz, a young Hungarian who had come to Sweden to work in the Mittag-Leffler Institute, and who stayed on in Sweden, and later became Professor at the University of Lund. Through him I was educated according to current standards of mathematical rigor, and was introduced to the modern subject of Lebesgue measure and integration theory. My personal research work was concerned with analytic number theory, where I became familiar with the technique of using Fourier integrals of a type closely similar to those I was to encounter later on when studying the relations between a probability distribution and its characteristic function. I obtained my Ph.D. in 1917, having written a thesis on Dirichlet series, one of the chief analytic tools used in prime number theory.

For a young Swedish mathematician of my generation, who wanted to find a job that would enable him to support a family, it was quite natural to turn to insurance. It was a tradition for Swedish insurance companies to employ highly qualified mathematicians as actuaries, and several of my University friends had jobs of this kind. After starting insurance work in 1918, I advanced in 1920 to a position as actuary of a life insurance company.

It was my actuarial work that brought me into contact with probability problems and gave a new turn to my mathematical interests. From 1918 on, I tried to get acquainted with the available literature on probability theory, and to do some work on certain problems connected with the mathematical aspect of insurance risk. These problems, although of a special kind, were in fact closely related to parts of probability theory that would come to occupy a central position in the future development, and I will say some introductory words about them already here.

The net result of the risk business of an insurance company during a period of, say, one year, could be regarded as the sum of the net results of all the individual insurances. If these were supposed to be mutually independent, the connection with the classical "central limit theorem" in probability theory was evident. But there was also the possibility of regarding the total risk business as an economic system developing in time, and in every instant subject to random fluctuations. Systems of this kind had been considered in some pioneering works which today appear as forerunners of the modern theory of stochastic processes.

Both these problems caught my interest already at this early stage. In the following chapter I shall try to give a brief account of the general situation in probability theory, and also of the history of those particular problems, up to 1920.

2. Probability theory before 1920

2.1. General remarks, foundations. It was rather a confused picture that met the eyes of a young man educated in pure mathematics, according to the standards of rigor current since the early part of the present century. There was the great classical treatise of probability theory written by Laplace and first published in 1812. It offered an interesting and stimulating reading, but was entirely nonrigorous from a modern mathematician's point of view. And it was surprisingly uncritical both with regard to the foundations and the applications of the theory. Incidentally, it may be interesting to mention that the Emperor Napoléon, who used Laplace as Minister and Senator, later criticized him, saying that Laplace had introduced "l'esprit des infiniments petits" into the practical administration.

And the French followers of Laplace, even such high class mathematicians as Poincare and Borel, did not seem to have produced a connected and well organized theory, built on satisfactory foundations. With very few exceptions, books and papers on probability problems were only too obviously lacking in mathematical rigor.

The work of the Russian school, with Tchebychev, Markov and Liapounov, which was perfectly rigorous and of high class, was at this stage still not very well known outside their own country.

The situation may be characterized by some quotations. The British economist Keynes, in his book [61], says about the classical probability theory that "there is about it for scientists a smack of astrology, of alchemy". And the German mathematician von Mises asserts in his paper [98] of 1919 that "today, probability theory is not a mathematical discipline", while the French probabilist Paul Lévy, talking in his autobiography [85] of his first acquaintance with probability theory as a young man, says that "in a certain sense this theory did not exist; it had to be created".

By Laplace and his followers, the classical definition of the probability of an event in terms of the famous "equally possible cases" was regarded as generally applicable, even when it seemed impossible to give a clear explanation of the nature of these "cases". There had been some attempts to overcome this difficulty, but they were not convincing.

Some authors had tried to work out a definition of probability based on the properties of statistical frequency ratios. In his paper [98] of 1919 and his book [99], von Mises gave a definition of this kind, founded on an axiomatic basis. He considered a series of independent trials, made under similar conditions, and postulated the existence of limiting values for the frequencies of various observed events, as well as the invariance of these limits for any appropriately selected subsequence of trials. He had enthusiastic followers and also severe critics. Paul Lévy later expressed the opinion [85, page 79] that it is "just as impossible as squaring the circle" to obtain a satisfactory definition in this way. Personally I was interested in von Mises' work, but took a critical attitude, looking forward to something more satisfactory. It was to come, but the time was not yet ripe for it about 1920.

2.2. The central limit theorem. The term "central limit theorem" was introduced by G. Polya in his paper [103] of 1920, to which I shall return in the following chapter. In present day terminology the theorem asserts that, under appropriate conditions, the probability distribution of the sum of a large number of independent random variables will be approximately normal (or Gaussian). In a very special case this had been proved already in 1733 by de Moivre [100]. The general theorem was given by Laplace, but his proof was incomplete. If the terms in the sum are considered to be small "elementary errors", the theorem was regarded as giving an explanation of the occurrence of the normal distribution in connection with errors of observation.

After a number of unsuccessful attempts by various authors to find a correct proof of the general theorem stated by Laplace, a promising approach was made byTchebychev[115, 116], who used the method of moments. The first complete proof was given in 1901 by Liapounov [86], who worked with the analytic tool known today under the name of characteristic functions. The work of Liapounov was very little known outside Russia, but I had the good luck to be allowed to see some notes on his work made by the German mathematician Hausdorff, and these had a great influence on my subsequent work in the field.

2.3. Pioneering work on stochastic processes. During the first ten years of the present century several works were published which, to a modern reader, appear as forerunners of the theory of stochastic processes created during the 1930's. All these works deal with the temporal development of some variable system exposed to random influences.

Bachelier, in his paper [2] of 1900, analysed the fluctuations of stock market prices, and was led to an important particular case of a stochastic process. In 1905, the same process was encountered by Einstein in his famous study [37] of the Brownian movement. This process is designed to give a mathematical description of the continuous, although extremely irregular path of a particle suspended in a fluid and subject to random molecular shocks.

Between these works of Bachelier and Einstein, the Uppsala Thesis [92] of Filip Lundberg appeared in 1903, written in Swedish. Lundberg studies the essentially discontinuous variations of the accumulated amount of claims incurred by an insurance company. He is thus led to a stochastic process of a type entirely different from the continuous Brownian movement process. The "Poisson process", well known today for its numerous applications, comes out as a very special case of Lundberg process, arising when all sums due under the claims are equal. But for his general case, developed also in some later works, he uses a functional equation which is a particular case of the famous "forward equation", introduced in 1931 by Kolmogorov [73] for a general class of continuous processes, and in 1940 by Feller [43] for the corresponding class of discontinuous processes, which includes the Lundberg process of 1903 as special case.

In 1909 the Poisson process was introduced by Eriang [39] in the study of telephone traffic problems, and by Rutherford and Geiger [113] in the analysis of radioactive disintegration.

All these pioneers used mathematical methods more or less lacking in rigor. But they developed a wonderful ability of dealing intuitively with concepts and methods that would have to wait until the 1930's before being placed on rigorous foundations.

3. A decade of preparation: 1920 to 1929

3.1. Stochastic processes and limit theorems. In the summer of 1920 I spent some time in Cambridge, England, working with number theory under the guidance of the great mathematician G. H. Hardy. There I met an American of my own age, Norbert Wiener, who says in his autobiography I Am a Mathematician [122, page 64] that it was quite a coincidence that he should have met both Paul Lévy in France and me on this European trip, since our work has always had close relations to his own. Already at this early stage, Norbert made the impression of being "quite a character". At this first meeting we had no opportunity to talk probability, but it was only three years afterwards in 1923 that he published his famous paper "Differential space" [118] where, several years before the basic works of Kolmogorov, he introduced a probability measure in a function space, thus giving a rigorous theory of the Brownian movement stochastic process, also known today as the Wiener process. Among other results he was able to show that almost all the trajectories of the process considered are continuous functions without a derivative. Unfortunately the paper was quite difficult to read, and many people interested in the field — including myself — did not see its true significance until the great works of the 1930's had paved the way.

Incidentally I may mention that my own first probabilistic paper was a brief note [12] of 1919 on the Poisson process, written in Swedish, where I deduced the expression for the relevant probability distribution from a simple set of necessary and sufficient conditions, well known today. In this connection I studied the Lundberg risk process, but my publications on that subject belong to a later period.

With respect to the central limit theorem, I was greatly interested in G. Polya's paper [103] of 1920 and J. W. Lindeberg's [87] of 1922. Polya, who was a Hungarian friend of Marcel Riesz and often visited Sweden, introduced in his paper the name "central limit theorem", as I have already mentioned. He referred to the work of Liapounov and indicated a proof based on the use of characteristic functions, pointing out the analogy with the methods currently used in prime number theory. He also discussed the method of moments used by Tchebychev and, in a more general setting, by Stieltjes.

Lindeberg, in his 1922 paper, gave a complete proof of the central limit theorem under more general conditions than Liapounov. He introduced the famous Lindeberg condition of which I shall have more to say below. I was happy to make his personal acquaintance at a mathematical congress in Helsingfors in the summer of 1922. He was Professor at the University of Helsingfors, and owned a beautiful farm in the eastern part of the country. When he was reproached for not being sufficiently active in his scientific work, he said, "Well, I am really a farmer". And if somebody happened to say that his farm was not properly cultivated, his answer was "Of course, my real job is to be a professor". I was very fond of him, and saw him often during the following years.

However, for the risk problems in which I was interested, it was not enough to know that a certain probability distribution was approximately normal; it was necessary to have an idea of the magnitude of the error involved in replacing the distribution under consideration by the normal one. Liapounov had given an upper limit for this error. I studied his method and gave in a paper [13] of 1923, a simplified version of his proof, together with a numerical estimation of the error, implying a slight improvement of Liapounov's result. For the particularly important case of identically distributed variables, Liapounov's theorem in my 1923 version runs as follows:

Let $x_1, x_2, \ldots, x_n$ be independent and identically distributed random variables, such that every x_i has zero mean, standard deviation σ and a third, absolute moment β_3. Let G_n be the distribution function of the normed sum

$$z_n = \frac{x_1 + x_2 + \cdots + x_n}{\sigma n^{1/2}}, \tag{1}$$

while Φ is the normal distribution function

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

Then for all $n > (\beta_3/\sigma^3)^2$

$$|G_n(x) - \Phi(x)| < \frac{3\beta_3}{\sigma^3} \cdot \frac{\ln n}{n^{1/2}}. \tag{2}$$

For the case when the x_i are not required to be identically distributed, there is a corresponding, slightly more complicated proposition.

However, it soon became clear that the estimation of the remainder given by (2) was far from sufficient for the numerical applications to risk problems that I had in view, and a great part of my work during the 1920's was concerned with attempts to find an improved estimation, preferably in the form of some asymptotic expansion of the remainder for large values of n.

In the literature on mathematical statistics available at the time, two kinds of expansion in series of a difference like $G_n(x) - \Phi(x)$ had been considered. As pointed out by Gnedenko and Kolmogorov in their book [48], they had both been introduced by Tchebychev, although they are often referred to respectively as the Edgeworth and the Charlier expansions. I shall here only give the explicit expression of the former, which was discussed by Edgeworth in [36], in a purely formal way. The Charlier series is a rearrangement of the terms of the Edgeworth one, which in some respects is simpler to deal with, but does not provide the best possible estimation. Charlier in his paper [11] of 1905 had claimed for his series certain asymptotic properties, but his proof was entirely wrong, being based on an incorrect use of what we now call the inversion formula for characteristic functions. Moreover, the assertion which Charlier claimed to have proved is true only in a modified form and under more restrictive conditions than those given by him (cf. Cramér, [33]). In fact, no rigorous discussion of the asymptotic properties of those series had been made, so that the problem was still open. In a preliminary paper [14] of 1925, and a more definitive one [16] of 1928 I attacked it, and gave a solution valid under certain conditions (cf. also Cramér [19] and Gnedenko–Kolmogorov [48] Chapter 8). For the particular case of identically distributed variables, my main theorem of 1928 runs as follows. (In the order of the error term, the $\underline{O}$ of my original statement is here replaced by $\bar{o}$, as given by Esseen [40] in 1944.)

Let the x_i of (1) be independent and identically distributed, each with zero mean, standard deviation a and finite moments $\alpha_1 = 0$, $\alpha_2 = \sigma^2$, $\alpha_3, \dots \alpha_k$, where $k \geqslant 3$. Let each x_i have the distribution function $F(x)$ and the characteristic function

$$f(t) = \int_{-\infty}^{\infty} e^{itx} dF(x). \tag{3}$$

Suppose that

$$\limsup_{t \to \infty} |f(t)| < 1. \tag{4}$$

Then we have the expansion

$$G_n(x) = \Phi(x) + (2\pi)^{-1/2} e^{-x^2/2} \sum_{j=1}^{k-2} \frac{p_j(x)}{n^{j/2}} + \bar{o}(n^{-(k-2)/2}) \tag{5}$$

uniformly in x, where $p_j(x)$ is a polynomial in x of degree $3j-1$, the coefficients of which depend only on the moments $\alpha_3, \dots, \alpha_{j+2}$. In particular we have

$$p_1(x) = \frac{\alpha_3}{3!}(1 - x^2).$$

Since $k \geqslant 3$, it will be seen that the condition (4) has enabled us to remove the factor $\ln n$ in the upper limit for the error term in Liapounov's theorem. An important particular case when the condition (4) is satisfied occurs when the given distribution function F contains an absolutely continuous component not identically zero.

In my 1928 paper [15] I also treated the case of nonidentically distributed variables, as well as the corresponding expansion for the probability density $G_n'(x)$. I further showed that, in a case when the condition (4) is not satisfied, a simple integrated average of the series in (5) over a small interval containing the point x still has the required asymptotic properties.

In this connection, I may mention that many years afterwards, I studied in a paper [29] of 1963 the asymptotic expansion of $G_n(x)$ in the case when the limiting distribution is nonnormal and stable. Under appropriate conditions there is still an expansion corresponding to (5), but of a more complicated structure.

3.2. Lévy's book of 1925; my own plans for a book. While I was deeply engaged in my work on the asymptotic expansions, the book "Calcul des probabilitiés" by Paul Lévy [80] appeared in 1925. Although I could not quite agree with him in respect of the foundations, I at once realized that this was a major event in the development of mathematical probability theory. It seemed clear that here was a first attempt to present the theory as a connected whole, using mathematically rigorous methods. It contained the first systematic exposition of the theory of random variables, their probability distributions and their characteristic functions. Although I had used these concepts in my own work for several years, Lévy's account of the theory brought much that was new to me. He also gave a discussion of the central limit theorem and of the stable probability distributions, as well as a very interesting chapter on kinetic gas theory.

During a visit to England in 1927, I saw my old teacher and friend G. H. Hardy. When I told him that I had become interested in probability theory, he said that there was no mathematically satisfactory book in English on this subject, and encouraged me to write one. I was more than willing to follow up his suggestion, but it seemed evident that it would take a fairly long time. In fact, it was not until ten years afterwards, in 1937, that my Cambridge Tract [19] was ready to appear.

3.3. Foundations. With respect to the foundations of probability theory, the attitudes of von Mises and Lévy were fundamentally different, and I did not feel able to agree with either of them. In the 1919 paper [98] of von Mises there was, however, one general statement with which I wholeheartedly agreed, although it seemed to me that he had not followed up the consequences of it when building his own collective theory.

On page 53 of that paper, von Mises expressed the opinion that probability theory is "a natural science of the same kind as geometry or theoretical mechanics". It is the object of this theory to describe certain observable phenomena, "not exactly, but with some abstraction and idealization". In other words, probability theory is to be regarded as a mathematical model of a certain class of observable phenomena.

In a Swedish paper [15] of 1926 I referred to this statement of von Mises. I expressed my agreement and made some further comments, from which I should like to quote here:

> "The probability concept should be introduced by a purely mathematical definition, from which its fundamental properties and the classical theorems are deduced by purely mathematical operations. ... Against such a mathematical theory, no objection can be valid except on mathematical grounds. On the other hand, it should be emphasized that the mathematical theory does not prove anything about the real events that will occur. Probability formulas are just as unable to dictate the behavior of real events as are the formulas of classical mechanics to prescribe that the stars must attract one another according to the Newton law. It is only experience that can guide us here and show if our mathematical model yields an acceptable approximation of our observations".

I still feel that these remarks are fairly reasonable, and am glad to have put them in print seven years before the definite formalization of probability theory by Kolmogorov.

3.4. The new Russian school. All through the latter part of the 1920's, it was clear that a strong new activity in mathematical probability was taking place in the Soviet Union. In a remarkable paper [7] of 1927, S.N. Bernstein discussed the extension of the central limit theorem to sums of random variables which are not required to be independent. He introduced an important method of dealing with such cases, to which I shall return later.

But the main Russian works of this period were written by two quite young mathematicians, A.Ya. Khintchine and A.N. Kolmogorov, who were to be leaders in the coming development of the field. In a joint paper [70] of 1925 they proved the famous "three series theorem", giving necessary and sufficient conditions for the convergence of a series, the terms of which are independent random variables. The probability that such a series is convergent can only be equal to zero or one, which is a particular case of the so called "zero-one law", discovered at the same time.

In his paper [71] of 1928, Kolmogorov proved his celebrated inequality for sums of independent random variables, which is a highly refined generalization of the well-known elementary inequality due to Tchebychev. Let $x_1, \ldots, x_n$ be independent random variables with zero means and finite (not necessarily equal) standard deviations. Then Kolmogorov's inequality asserts that

$$\mathsf{P}\Big(\max_{j=1,\ldots,n} \sum_{i=1}^{j} x_i \geqslant k\Big) \leqslant \frac{\mathsf{E}\Big(\sum_{i=1}^{n} x_i\Big)^2}{k^2}. \tag{6}$$

The proofs of this and other similar inequalities are based on an expert use of conditional probabilities and expectations, foreshadowing the general theory of this part of the subject that Kolmogorov was soon to give. The inequality (6) is an invaluable tool in all investigations concerning sums of independent random variables.

In 1929 Kolmogorov gave in [72] a proof of the so called law of the iterated logarithm, previously found by Khintchine [62] in a particular case. For the case of a normed sum z_n of n independent and identically distributed variables as defined by (1), it follows from the central limit theorem that, for any function $h(n)$ tending to infinity with n, the probability of the relation $z_n > h(n)$ will tend to zero as n tends to infinity. Still, if we consider the infinite sequence $z_1, z_2, \ldots$, we may expect sometimes to observe very large values. The law of the iterated logarithm gives a precise expression of this vague statement by asserting that the relation

$$\lim\sup_{n\to\infty} \frac{z_n}{(2\ln\ln n)^{1/2}} = 1 \tag{7}$$

will hold with a probability equal to one. It is interesting to observe that the particular case of this proposition proved by Khintchine in his paper [62] of 1924 was concerned with the frequency of digits in dyadic fractions. This was considered as a purely measure-theoretic problem, which had already attracted my interest before I had taken up probability theory. The general statement proved by Kolmogorov made a great impression, and prepared the way for the identification of probability with measure that he was soon to give.

4. The great changes: 1930 to 1939

4.1. The Stockholm group. In the present chapter it will be even more impossible than in the preceding one to follow a strictly chronological order. Among the great number of new ideas that were coming forward during the 1930's there are certain main groups, the development of which will have to be considered separately, always from my personal point of view. I will begin with a few words on my own activities during the beginning of this "heroic period" of mathematical probability theory.

On the initiative of the Swedish insurance companies, a professorship for "Actuarial Mathematics and Mathematical Statistics" had been founded at Stockholm University, and in the fall of 1929 I was nominated to be its first holder. From the beginning and all through the years, I was fortunate to work with a group of ambitious and well-qualified students. We followed with keen interest the important new works that were forthcoming abroad, and tried to give some contributions of our own to the development of the field. During the first years we were mainly following up the work of the 1920's with respect to limit theorems and risk processes, where important new results seemed to be within reach.

In the fall of 1934 our group had the good fortune to receive a new member from abroad. It was during the bad days of the Nazi regime in Germany, when so many outstanding scientists were leaving that country. Will Feller, who had been turned out from the University of Kiel, came to join our group, and stayed on in Stockholm for five years. He made a great number of Swedish friends, collaborating with economists and biologists as well as with the members of our probabilistic group. He had studied in Gottingen and was well initiated in the great traditions of this mathematical center. We tried hard to get a permanent position for him in Sweden, but in those years before the war this was next to impossible, and it was with great regret that we saw him leave for the United States, where an outstanding career was awaiting him. In the following sections I shall have more to say about his work during the years he spent with us.

4.2. Foundations. Looking back towards the beginning of a new era in mathematical probability theory, it seems evident that the real breakthrough came with the publication in 1933 of Kolmogorov's book [75] Grundbegriffe der Wahrscheinlichkeitsrechnung. In this book he laid the foundations of an abstract theory, designed to be used as a mathematical model for certain classes of observable events. The fundamental concept of the theory is the now familiar, classical concept of a probability space (Ω, A, P), where Ω is a space of points ω which are denoted as elementary events, while A is a σ-algebra of sets in Ω, and P is a probability measure defined for all A-measurable events, i.e. for all sets S belonging to A.

The concepts of random variable $x = x(\omega)$ and stochastic process $x(t) = x(t, \omega)$, where t belongs to some parameter space T, are introduced in the way well known today, which in 1933 represented a remarkable innovation. It was made clear that, viewed in this way, a stochastic process defines a probability distribution in the space X of all functions $x(t)$ of the variable t. For any finite set of points $t_1, \ldots, t_n$ the n-dimensional joint distribution of the random variables $x(t_1), \ldots, x(t_n)$ is called a finite-dimensional distribution of the $x(t)$ process. The family of all these distributions satisfies certain evident conditions of consistency. One of the main propositions given in Kolmogorov's book states that, if a family of finite-dimensional distributions is given and satisfies the consistency conditions, there exists a stochastic process corresponding to the given distributions. Further, the probability that the function $x(t)$ belongs to a set S in the function space X is uniquely determined by the finite-dimensional distributions for all Borel sets in X, i.e. for all sets S belonging to the smallest σ-algebra of sets in X containing all sets of functions satisfying a finite set of inequalities of the form $a_i < x(t_i) < b_i$ (This applies to the case of a real-valued $x(t)$; the extension to the complex-valued case being obvious.) In this way, rigorous foundations were established for the investigations of stochastic processes, which were rapidly increasing in number and importance.

However, it soon appeared that in many cases one encounters sets of functions interesting from the point of view of the applications, which are not Borel sets. Thus for the Wiener process and for the Lundberg risk process, both mentioned above in 2.3 and 3.1, it would be desirable to find the probability of the set of all functions $x(t)$ such that $x(t) < a$ for all t with $0 < t < b$, but these sets are not Borel sets, and their probabilities are not uniquely defined by the finite-dimensional distributions. For such cases it is necessary to modify the general definitions. In the book [120] by Paley and Wiener, a possible modification is shown for the Wiener process, and a similar method can be used for the risk process. For the general case Doob has analyzed the question in a series of penetrating works, summed up in his great book [35] of 1953. But all these works rest ultimately on the foundations laid down by Kolmogorov.

Kolmogorov's book of 1933 also contains a chapter on the subject of conditional probabilities and expectations, where these concepts are introduced and treated by a radically new method.

Kolmogorov's book still ranks as the basic document of modern probability theory. If in 1920 it might be said (cf. above, 2.1) that this theory was not a mathematical subject, it was impossible to express such an opinion after the publication of this book in 1933.

4.3. Markov processes. In his paper [73] published in 1931, two years before the "Grundlagen" book, Kolmogorov investigated a general class of stochastic processes, which later received the name of Markov processes, owing to the fact that they form a natural generalization of the classical concept of Markov chains. Consider a process $x(t)$, where t is a real parameter representing time. If, for any $t_0 < t_1$, the conditional distribution of $x(t_1)$, relative to the hypothesis $x(t_0) = a$, is independent of any additional information about the values assumed by $x(t)$ for $t < t_0$, then $x(t)$ is called a Markov process. Kolmogorov showed that the probability distributions associated with a Markov process satisfy certain functional equations, which under appropriate continuity conditions reduce to partial differential equations of parabolic type, and that these equations uniquely determine the corresponding distributions.

In our Stockholm group, it was particularly Feller who followed up this general theory of Kolmogorov with two papers ([42] and [43], the latter written during his first year in the United States) dealing with the partial differential equations of a continuous process and the integro-differential equations encountered in the discontinuous case. It is well known that the subject of Markov processes has since this time developed into a very extensive field of research.

Personally, while admiring the 1931 paper by Kolmogorov and its continuation by Feller, I did not take up any research on the general Markov process theory, perhaps owing to the fact that I have never learned to feel quite at home with partial differential equations. But there was a particular class of Markov processes that, already in the early 1930's, presented a direct interest to those of us who were working with the stochastic processes of risk theory. I am referring to the processes with independent increments, which will be discussed in the following section.

4.4. Processes with independent increments. In 1932, Kolmogorov [74] gave an expression for the characteristic function of a random variable $x(t)$ associated with a stochastic process satisfying the following conditions:

(1) x(t) has zero mean and a finite second order moment; $x(0) = 0$.

(2) For $0 \leqslant t_0 < t_1 < \cdots < t_n$ the differences $x(t_i) - x(t_{i-1})$ are independent random variables.

(3) The probability distribution of $x(t_i) - x(t_{i-1})$ depends only on the difference $t_i - t_{i-1}$.

Such a process is known as a process with stationary and independent increments. Under more general conditions, without assuming the existence of any finite moments, these processes were investigated in an admirable paper by Paul Lévy [81], published in 1934. He gave the classical general expression for the characteristic function of $x(t)$. An alternative expression which is sometimes more easily manageable was given in 1937 by Khintchine [67].

For our Stockholm group these works appeared as revelations and were eagerly studied. Already from the 1932 formula of Kolmogorov it followed that the Wiener process and the Lundberg risk process come out as opposite particular cases of the general expression. In fact, under the above conditions (1)-(3), the characteristic function

$$f(z,t) = \mathsf{E}(e^{izx(t)})$$

is given by the Kolmogorov formula

$$\ln f(z,t) = t\Big(-\frac{1}{2}\sigma^2 z^2 + \int_{-\infty}^{\infty} \frac{e^{izu} - 1 - izu}{u^2} dK(u)\Big),$$

where $\sigma^2 \geqslant 0$ is a constant, while $K(u)$ is a bounded and never decreasing function which is continuous at $u = 0$.

It is clear that if $K(u)$ is identically zero we obtain the Wiener process. Lévy made in [81] the interesting remark that if $x(t)$ is a function of t that is almost certainly continuous, we necessarily have this case.

On the other hand, if $\sigma^2 = 0$ and $K(u) = \lambda \int_{-\infty}^{u} v^2 dG(v)$, where λ is a positive constant and $G(u)$ is a distribution function, we have a risk process where the claims occur according to a Poisson process with the parameter λ, while the amounts of the claims are independent random variables, each with the distribution function G. The variation of $x(t)$ is here essentially discontinuous. It was this case that we had particularly studied in Stockholm. $x(t)$ is here the accumulated amount of claims up to the time t, and the "ruin problem" which is particularly important for the insurance applications, is concerned with the probability of having $x(t) < a+bt$ during the whole interval $0 < t < T$. In a thesis [114] of 1939, C.O. Segerdahl, a member of our group, studied this problem and proved certain important inequalities for the probability of ruin.

4.5. Infinitely divisible distributions, arithmetic of distributions. If a random variable x can be represented in the form $x = x_1 + x_2$, where x_1 and x_2 are independent, Lévy proposed in 1934 to say that the distribution of x is the product of the distributions of x_1 and x_2, and contains each of these as a factor. If there is no nontrivial representation of this form, the distribution of x is said to be indecomposable.

If $x(t)$ is the random variable associated with a process with stationary and independent increments, it was shown in the 1934 paper [81] of Lévy that $x(t)$ can be represented as the sum of an arbitrary number of independent variables, all having the same distribution. The distribution of $x(t)$ is then said to be infinitely divisible, and the Lévy formula gives the general expression for a distribution of this kind. The normal and Poisson distributions, as well as the stable distributions, all belong to this class.

In his 1934 paper Lévy had expressed the conjecture that any factor of a normal distribution must itself be normal. He had repeated this conjecture in some subsequent papers, saying that he regarded this as very probably true, but had not been able to prove it. In the beginning of 1936 I had the good luck of finding a proof of this conjecture, which I published in [18]. In his autobiography of 1970, Lévy says [85, page 111] that he regretted not having found the proof himself, since it was a direct application of the theory of characteristic functions which he had systematically used in his own work. Soon afterwards Raikov [105] proved the corresponding statement for the Poisson distribution.

In 1937 Khintchine [68] proved the general theorem that any distribution can be represented as the product of an infinitely divisible distribution and an at most enumerable number of indecomposable distributions. In general this factorization is not unique.

4.6. Limit theorems. The works of Liapounov and Lindeberg mentioned above in 3.1 give sufficient conditions for the validity of the central limit theorem for sums of independent random variables. In an important book of 1933, Khintchine [64] gave an account of the main results so far known in this field. However, the problem of finding conditions which are both necessary and sufficient still remained open, as well as the case when the existence of finite moments of the variables in the sum is not assumed. The general problem attacked during the 1930's by several authors can be thus expressed:

If $x_1, x_2, \ldots,$ are independent random variables, it is required to find conditions under which there exist constants a_n and b_n such that the probability distribution of the normed sum

$$(x_1 + \cdots + x_n - a_n)/b_n$$

tends to the normal distribution as n tends to infinity.

Important contributions to the investigation of this problem were given by Lévy and Khintchine, but it was Feller who first published the complete solution, in a paper [41] consisting of two parts, written in 1935 and 1937 while he was working as a member of our Stockholm probabilistic group.

Feller showed that the required necessary and sufficient conditions can be expressed by appropriate modifications of the sufficient condition given in 1922 by Lindeberg. In the particular case when the x_i have distribution functions F_i with zero means and finite standard deviations σ_i, such that $\sigma_i/s_n \to 0$ as $n \to \infty$, uniformly for $i = 1, 2, \dots, n$, where as usual $s_n^2 = \sigma_1^2 + \cdots + \sigma_n^2$, we can take $a_n = 0$ and $b_n = s_n$, and Feller proved that the original Lindeberg condition

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{n} \int_{|x|>\varepsilon s_n} x^2 dF_i(x) = 0$$

for any given $\varepsilon > 0$ is both necessary and sufficient for the convergence to the normal distribution. He also gave more elaborate conditions for the case when the x_i have no finite moments.

The outstanding Italian mathematician F. P. Cantelli was engaged in actuarial work on similar lines to those of several members of our Stockholm group. In a paper [10] of 1933, he had given an improved form of the iterated logarithm law mentioned above in 3.4. In the course of a correspondence with him, I considered a modification of the problem, where a "most possibly precise" result could be obtained. This was published in my paper [17] in 1934. For the classical law of the iterated logarithm as expressed by (7), a statement of this character was to be given by Feller [44] in 1943. I will quote here for comparison first the Feller theorem of 1943, and then my own of 1934, in both eases without giving detailed statements of the conditions under which our results are shown to hold.

Using the above notation, Feller proved that the probability $\mathsf{P}(|x_1 + \cdots + x_n| < \mu_n s_n)$ for all sufficiently large n is equal to 1 or 0 according as the series

$$\sum_{n=1}^{\infty} \frac{\sigma_n^2}{s_n^2} \mu_n e^{-\frac{1}{2}\mu_n^2}$$

is convergent or divergent.

On the other hand, in my 1934 paper I had somewhat modified the problem. Instead of a single sequence of independent random variables $x_1, x_2, \dots,$ I considered the double sequence of variables

$$\begin{array}{l} x_{11}, \\ x_{21}, x_{22} \\ \cdots \\ x_{n1}, x_{n2}, \dots, x_{nn} \end{array}$$

all supposed to be independent, with zero means and finite standard deviations σ_{ij}. Writing $s_n^2 = \sigma_{n1}^2 + \cdots + \sigma_{nn}^2$, I showed that the probability $\mathsf{P}(|x_{n1} + \cdots + x_{nn}| < \lambda_n s_n)$ for all sufficiently large n) is equal to 1 or 0 according as the series

$$\sum_{n=1}^{\infty} \frac{\lambda_n}{e}^{-\frac{1}{2}\lambda_n^2}$$

is convergent or divergent.

A similar, but much more far-reaching generalization of the whole problem of limiting distributions for sums of independent random variables was given by Khintchine [67] in 1937, followed by works due to Gnedenko [47] and other Russian authors. I shall here only indicate the main lines of their work, referring for more details to the important book [48] by Gnedenko and Kolmogorov. Consider a double sequence of random variables

$$x_{11}, x_{12}, \ldots, x_{1k_1},$$
$$\ldots$$
$$x_{n1}, x_{n2}, \ldots, x_{nk_n}$$

the variables in each line being supposed to be mutually independent, and let z_n denote the sum of the variables in the nth line. Suppose that we have for any $\varepsilon > 0$

$$\lim_{n\to\infty} \mathsf{P}(|x_{nk}| > \varepsilon) = 0$$

uniformly for $k = 1, \ldots, k_n$. Then Khintchine in 1937 proved the following fundamental result: In order that F should be the limiting distribution function for the variables $z_n - b_n$ associated with such a double sequence, where the b_n are appropriately chosen constants, it is necessary and sufficient that F be infinitely divisible. Thus the class of all possible limit laws associated with sums of the type here considered coincides with the class of all infinitely divisible laws. It is hardly necessary to say that in our Stockholm group we followed these new ideas of our Russian colleagues with the greatest interest.

4.7. Characteristic functions. The use of an analytic tool more or less equivalent to that known to us under the name of characteristic functions goes back to Lagrange, Laplace and Cauchy. As mentioned above, it was used by Liapounov for his proof of the central limit theorem. But the first systematic account of the theory was given in the 1925 book of Lévy [80] referred to in 3.2. I had used this method for my work on asymptotic expansions, and throughout the 1930's we tried in Stockholm to develop the subject further. My proof of Lévy's conjecture for the normal distribution mentioned in 4.5 was an outcome of this work. H. Wold and I published in 1936 a joint paper [34] on multidimensional distributions, where generalizations of characteristic functions and the central limit theorem were discussed.

In the same year 1936 I had concluded the work on a book on mathematical probability, which G. H. Hardy in 1927 had encouraged me to write. It took the form of a Cambridge Tract [19] entitled "Random Variables and Probability Distributions", published in the early part of 1937. It was based on Kolmogorov's foundations, and its main contents were a full account of the theory of probability distributions in finite-dimensional spaces and their characteristic functions, with applications to subjects such as the central limit theorem, the allied asymptotic expansions, and stochastic processes with independent increments.

With respect to the general theory of characteristic functions I was able to give some complements to Lévy's work. But in this connection I had made a regrettable error which had to be corrected in later editions of the book, of 1963 and 1970. As the fact is of some interest, I shall give a brief account of it here.

The well-known "continuity theorem" for characteristic functions asserts that a sequence of distribution functions F_n converges to a distribution function F in every continuity point of the latter, if and only if the corresponding characteristic functions $f_n(t)$ converge for every t to a limit which is continuous for $t = 0$. In the statement of this theorem given in the first edition of my Tract, I required only convergence of the $f_n(t)$ in some finite interval $|t| < c$. However, a letter from Khintchine informed me of my error. It is sufficient to point out that it is possible to find two characteristic functions $f_1(t)$ and $f_2(t)$ which are equal for $|t| < 1$, but not identically equal. Take, in fact,

$$f_1(t) = f_2(t) = 1 - |t|$$

for $|t| \leqslant 1$, and let $f_1(t)$ be periodic with period 2, while $f_2(t) = 0$ for $|t| > 1$. It is easily seen that both of these are characteristic functions. A sequence the members of which are alternatively equal to f_1 and f_2, thus converges for $|t| < 1$, but not for all t, which proves the falsehood of the statement in my first edition.

In this connection I may add that in a paper [21] of 1939 I gave some theorems on the representation of functions by Fourier integrals which, among other results, include a simplification of an earlier theorem by Bochner, giving a necessary and sufficient condition for a function to be the characteristic function of some probability distribution.

4.8. Stationary processes. I must now go some years back in time, and take up an important new thread of the development during the 1930's. In 1934 Khintchine published a basic paper [65] where he introduced the class of stationary stochastic processes. He pointed out that a Markov process cannot be used in a case when the past history of the system under consideration has an essential influence on any prediction with respect to its future development, as e.g. in statistical mechanics. As a convenient tool for the study of such systems he introduced the class of stationary processes.

Khintchine gave definitions both for the class of strictly stationary processes, and for the class which I shall here simply call stationary. A process $x(t)$ with a continuous time parameter t is called strictly stationary if the associated finite-dimensional distributions are all invariant under a translation in time. It is called stationary, if the same invariance holds for its moments of the first and second order. When reviewing the results of Khintchine's basic paper [65] and of the subsequent developments, I shall in the sequel in general consider a complex-valued process $x(t)$ which is continuous in mean square, and such that

$$\mathsf{E}x(t) = 0, \quad \mathsf{E}x(t)\overline{x(u)} = r(t-u).$$

The covariance function $r(t)$ is then continuous for all real t, and Khintchine showed that it admits the spectral representation

$$r(t) = \int_{-\infty}^{\infty} e^{itu} dF(u), \tag{8}$$

where the spectral function $F(u)$ is real, never decreasing and bounded. From this representation he deduced various properties of $r(t)$. He also proved that the time average

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

converges in quadratic mean as $T \to \infty$, and pointed out that this is equivalent to the mean ergodic theorem of von Neumann.

In a somewhat earlier paper [63], Khintchine had considered the case of a strictly stationary process with discrete time, or a sequence of random variables $\dots, x_{-1}, x_0, x_1, \dots$, satisfying the condition of strict stationarity. For this case he proved that the time average

$$\frac{1}{N}\sum_{n=1}^{N} x_n$$

converges even with probability one, which for this process is the equivalent of Birkhoff's ergodic theorem. The corresponding result for continuous time was proved somewhat later by Kolmogorov [76].

There are interesting relations between Khintchine's theory of stationary processes and the earlier work by Norbert Wiener on generalized harmonic analysis.

Wiener published in 1930 an extensive paper [119] on this subject, where he dealt with complex-valued functions f such that the limit

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} f(u+t)\overline{f(u)}du = r(t) \tag{9}$$

exists for all real t. The limiting function r has properties quite similar to those of the covariance function of a stationary process. In particular it is continuous for all t if it is continuous for

$t = 0$, and in this case it admits a spectral representation of the form (8). It seems reasonable to expect that the Wiener relation (9) should, in general, be satisfied by the sample functions of a stationary process.

Wiener's paper, quite like his 1923 paper on differential space, is not easy to read, but Masani has given in [95] a very readable account of it from a modern point of view. He shows, in particular, that Wiener gives an example of what was later to be called a normal stationary process such that, with probability one, any sample function of the process satisfies the Wiener relation (9). However, I should like to point out that it is also easy to find an example where this property does not hold. Let, in fact, a real-valued and normal stationary process have the covariance function $r(t) = 1 - |t|$ for $|t| \leqslant 1$, and let $r(t)$ be periodic with period 2. Then it can be shown that, for any given t, there is a set of sample functions of this process having positive measure, and such that the first member of the Wiener relation (9) does not converge to any finite limit as $T \to \infty$.

Khintchine's paper [65] on stationary processes had a very significant impact. It seemed clear that this new type of stochastic process would provide a convenient mathematical model not only for statistical mechanics, but also in other fields, such as meteorology and economics. In particular it opened up new possibilities for the investigation of phenomena showing a tendency to periodic behavior, and for applications to information theory.

In our Stockholm group the subject was taken up by Herman Wold, who in his thesis [123] of 1938 dealt with stationary processes with discrete time, i.e. stationary sequences of random variables x_n. He established their covariance and spectral properties, and gave extensive applications to various statistical problems. His most remarkable result was the proof, for the class of processes considered, of an important decomposition theorem which has later been shown to hold also, with due modifications, for more general types of processes. He proved that, for a stationary sequence x_n, there is a unique decomposition

$$x_n = u_n + v_n,$$

where u_n and v_n are mutually uncorrelated stationary sequences such that, in present-day terminology, u_n is purely nondeterministic, while v_n is deterministic. Further, u_n can be represented in the form

$$u_n = \sum_{i=-\infty}^{n} c_{n-i} z_i,$$

where the z_i are mutually uncorrelated random variables, while the c_i are non-random constants.

In a paper [22] of 1940, I investigated the covariance properties of a stationary vector process $\mathbf{x}(t) = (x_1(t), \ldots, x_q(t))$ with continuous time t. I proved that there is a spectral representation of the mutual covariance functions

$$r_{mn}(t) = \mathsf{E} x_m(t+u)\overline{x_n(u)} = \int_{-\infty}^{\infty} e^{tu} dF_{mn}(u), \tag{10}$$

where the F_{mn} are complex-valued functions of bounded variation, such that the increments ΔF_{mn} over any interval form a nonnegative Hermitian matrix. This implies that the matrix function $\mathbf{F}$ with the entries F_{mn} is never decreasing, and I proved that, like a never decreasing function of one variable, it is the sum of three components: one absolutely continuous, one purely discontinuous, and one singular. I also proved the corresponding properties for a vector process with discrete time.

4.9. Paris, London and Geneva, 1937-1939. In the spring of 1937 I was invited to Paris, is order to give some lectures at the Sorbonne. Of the French probabilists I had met Frechet before, but this was my first acquaintance with Paul Lévy, and with a number of the younger generation, such as Doeblin, Dugue, Fortet and Loève. In earlier years Frechet had been an outstanding mathematician, doing pathbreaking work in functional analysis. He had taken up probabilistic work at a fairly advanced age, and I am bound to say that his work in this field did not seem very impressive to me. On the other hand, already in 1937 it seemed clear that Lévy was to be one of the leaders in the development of probability theory, particularly after the publication of his book [83] "L'Addition des Variables Aléatoires". Incidentally, he mentions in the preface to this book that it was on receiving my letter in the beginning of 1936 containing my proof of his conjecture about the normal distribution that he made the decision to write the book. Among the younger men, Doeblin had already done outstanding work, and it was a great loss to science when he was killed during the first months of the war.

Later in the same year 1937 there was a conference for probability theory in Geneva. Feller and I attended from Stockholm, and it was quite exciting to see such a large group of eminent probabilists assembled. Among new acquaintances, there were Steinhaus from Poland, Hopf from Germany, and Jerzy Neyman, at that time still working in England. Several colleagues from the Soviet Union had accepted the invitation and announced lectures, but to our great disappointment none of them turned up. Neyman gave a lecture on his theory of confidence intervals, which was then something quite new. At this early stage his ideas had not yet received their final expression, and both Frechet and Lévy took a very critical attitude. I had already read his paper [101] on the subject, and had come to the conclusion that his basic ideas were quite sound, as, indeed, the later development showed them to be. Feller lectured about axiomatics, and I gave a talk [20] on the "large deviations" connected with the central limit theorem. If, using the notation of 3.1, we allow x to tend to infinity with n, both $G_n(x)$ and $\Phi(x)$ tend to unity, so that the assertion of Liapounov's theorem becomes trivial. In my paper, I considered the ratios

$$\frac{1 - G_n(x)}{1 - \Phi(x)} \quad \text{and} \quad \frac{G_n(x)}{\Phi(x)}$$

where x and n both tend to $+\infty$ in the first case, and to $-\infty$ in the second. I showed that, subject to appropriate conditions, it is possible to find an asymptotic expansion for each case. A first generalization of my results was given by Feller in 1943 who used it for his improved form of the iterated logarithm law [44] mentioned above in 4.6. Very important further generalizations were given by Yu.V. Linnik and his Leningrad group. An account of their work is given in the excellent 1966 book [55] by Ibragimov and Linnik.

In October 1938 I made a visit to England. It was shortly after the Munich conference, and the question of peace or war was already on everybody's mind. In Cambridge I saw my old teacher and friend G. H. Hardy, who lived in Newton's rooms in Trinity College. He expressed his satisfaction with my tract, which was written on his initiative. In London I was received by R. A. Fisher, William Elderton and Egon Pearson, and saw something of the work in mathematical statistics that was being done there. Neyman was already in California.

In the summer of 1939 there was again a conference in Geneva, this time for mathematical statistics. I was happy to make the acquaintance of Sam Wilks and Maurice Bartlett, who both gave interesting lectures. Neyman did not appear, but had contributed a basic paper [102] on the theory of testing statistical hypotheses, recently founded by him in collaboration with Egon Pearson. From these days I remember a conversation with R. A. Fisher. I had expressed my admiration for his geometrical intuition in dealing with probability distributions in multidimensional spaces, and received the somewhat acid reply: "I am sometimes accused of intuition as a crime!"

When leaving Geneva in July 1939, it seemed fairly clear that the war was rapidly approaching.

5. The war years: 1940 to 1945

5.1. Isolation in Sweden. During the war our international contacts were reduced to a minimum. Sweden remained neutral, but surrounded by war: Denmark and Norway were occupied by the Nazis, and Finland was at war with Russia. There were very few opportunities to exchange literature or correspondence with colleagues in England, France and the United States, and none at all with those in the Soviet Union. But we tried to keep our research work going as far as possible.

In his 1940 thesis [93] Ove Lundberg, a member of our Stockholm group, investigated a new class of stochastic processes which he applied to certain statistical problems of non-life insurance. On the foundations laid by him, this subject has had a strong international development among actuaries.

In February 1941 I organized a conference for mathematical probability in Stockholm. We were happy to have some guests from Denmark and Finland; in Norway the German occupation was harder, and none of our colleagues had been allowed to come. Harald Bohr and B0rge Jessen of Denmark talked about the foundation of probability and its relations to general mathematical analysis, and Gustav Elfving of Finland about Markov processes. Several members of our Stockholm group gave accounts of their theses which I have mentioned above, and I presented on this occasion a theorem about the spectral representation of a stationary stochastic process which I had just found. If $x(t)$ is the stationary process considered above in 4.8, the covariance function of which has the spectral representation (8), there is a process with orthogonal increments $z(u)$ such that, with an appropriate definition of the stochastic integral,

$$x(t) = \int_{-\infty}^{\infty} e^{itu} dz(u), \tag{11}$$

where in the usual symbolism

$$\mathsf{E} dz(u) = 0, \quad \mathsf{E}|dz(u)|^2 = dF(u).$$

This shows how $x(t)$ is additively built up by elementary harmonic oscillations $e^{itu}dz(u)$, each of which has an angular frequency u, while the amplitude and the phase are random variables determined by $dz(u)$.

In 1942 I published a paper [23] on the spectral representation (11), explicitly pointing out the relations to Wiener's generalized harmonic analysis. But it was not yet clear to me that what I had done was really to give a probabilistic version of Stone's theorem on the spectral representation of a unitary group in Hilbert space. The method I had used for my proof employed stochastic Fourier integrals, but no Hilbert space theory. The fundamental importance of this theory for the study of stochastic processes did not become known to us until after the war.

A young Danish physicist, N. Arley, had written a thesis on the application of stochastic processes to the theory of cosmic radiation [I], and to my great surprise I was allowed to come to Copenhagen in the spring of 1943 as a member of the examination committee. The thesis was a good piece of work, and I was happy to see my Danish colleagues again, but it was painful to see German troops marching through the streets of Copenhagen.

In 1944 the Uppsala thesis [40] on Fourier analysis of distribution functions by C. G. Esseen appeared. This was a very important work, based on a penetrating investigation of the properties of characteristic functions. In particular Esseen showed that the factor $\ln n$ in the above evaluation (2) of the remainder in Liapounov's theorem can always be omitted, thus generalizing my result mentioned in connection with (5), which holds only subject to the condition (4). He also gave estimations of the remainder depending both on n and x, as well as an improvement of my asymptotic expansion of 1928.

While it still seemed clear that the end of the war was far away, I decided to use the years of undesired isolation to write a book. This book [24] was to be ready in 1946, and was entitled Mathematical Methods of Statistics. From the preface I quote the following lines:

> "During the last 25 years, statistical science has made great progress, thanks to the brilliant schools of British and American statisticians, among whom the name of Professor R. A. Fisher should be mentioned in the foremost place. During the same time, largely owing to the work of French and Russian mathematicians, the classical calculus of probability has developed into a purely mathematical theory satisfying modern standards with respect to rigor. The purpose of the present work is to join these two lines of development in an exposition of the mathematical theory of modern statistical methods, in so far as these are based on the concept of probability".

Thus the book is not to be regarded as a contribution to mathematical probability theory, but rather as a treatise of its applications to modern statistical methods. The book is dedicated to my wife, who had supported and encouraged me all through my work, as she had always done. While I was writing, I sometimes said to her that I hoped this book would be my entrance card to the new world after the war. Perhaps there was something in it: today there are editions of the book in English, Russian, Spanish, Polish and Japanese.

5.2. International development during the war years. On the few occasions when there was mail from the United States, we received letters and reprints from Feller, and so were able to follow his work on Markov processes and on the perfection of the iterated logarithm law mentioned above in 4.3 and 4.6.

Mathematicians in war-making countries became often engaged in work with antiaircraft fire control and noise filtration in radar. It appeared that stationary stochastic processes provided an efficient tool for this kind of work. In particular the possibility of making predictions for the future course of such a process, based on observations during the past, was of vital importance. Independently of one another, Kolmogorov in the Soviet Union and Wiener in the United States made important contributions to this subject. They do not seem to have been aware of one another's work until after the war.

In two papers [77, 78] of 1944, Kolmogorov investigated a stationary stochastic process with discrete time. He pointed out that the class of all random variables with finite second order moments constitutes a Hilbert space, if the inner product of two points is defined as the covariance moment of the corresponding random variables. A stochastic process with discrete time can thus be regarded as a sequence of points in Hilbert space, so that the theory of this space becomes available for the study of the process.

For stationary sequences of random variables, Kolmogorov showed that the application of Hilbert space theory makes it possible to deduce in a simple way all results previously known, such as the Wold decomposition and the covariance properties of a vector process with discrete time given in my paper [22] mentioned above in 4.8. Moreover, he used powerful methods of complex function theory to give for the first time a necessary and sufficient condition for a stationary sequence to be purely nondeterministic (regular in Kolmogorov's terminology), and deduced a complete solution of the linear least squares prediction problem. His work was followed up by Zasuchin [127], who considered a stationary vector process with discrete time, and gave a number of important results for such a process.

The fundamental importance of this work by Kolmogorov lies in the fact that he showed how the abstract theory of Hilbert space (as well, of course, as of other types of spaces) could be applied to the theory of random variables and stochastic processes. This had a powerful influence on the whole subsequent development of the theory.

Wiener's war work was related to the problems of linear prediction and filtering for stationary processes, both with discrete and continuous time. For the prediction of the value of a stationary process $x(t)$ with continuous time at $t = h > 0$, based on the observation of the past of the process up to $t = 0$, he introduced a predictor of the form

$$x^*(h) = \int_{-\infty}^{0} x(t)dK(t),$$

where K is of bounded variation, and showed how K could be determined so as to minimize the mean square error of prediction

$$\mathsf{E}(x^*(h) - x(h))^2.$$

This is not a complete solution of the mathematical problem of linear prediction, but it has important applications to various engineering problems. Wiener's work on this and allied subjects was completed in 1942, and during the following years circulated in mimeographic copies bearing the nickname "the yellow peril", owing to the difficult mathematics involved. It did not become publicly available until 1949, in the book [121].

6. After the war: 1946 to 1970

6.1. Introductory remarks. After the great turmoil of the war years, it gradually became possible again to take up scientific research work on an international basis, and to renew contacts with colleagues in other countries. It was now clear to everybody concerned that mathematical probability theory had passed through a radical change during the twenty-five years between 1920 and 1945. While in 1920 it had hardly deserved the name of a mathematical theory, in 1945 it entered into the postwar world as a well-organized part of pure mathematics with problems and methods of its own, and with an ever growing field of applications to other sciences, as well as to many practical activities. There were intimate mutual relations between the applications and the purely mathematical theory, and already it was hardly possible for an individual research worker to survey the whole field.

The years after the war brought a further powerful development in many different directions. In a review of personal recollections like the present one, it is evidently even more impossible than for the time up to 1945 to give a full account of the development. I shall have to confine myself strictly to those parts of the field in which I was able to be more or less actively engaged, and to those people with whom I had personal contacts.

I will begin by giving a brief account of my contacts with scientific colleagues during the years immediately following the war, and then proceed to a survey of the subsequent development, always from my personal point of view.

6.2. Paris, Princeton, Yale, Berkeley, 1946-1947. Soon after the end of the war, I received invitations to the universities named above. In Paris I was to give a series of five lectures, two on statistical estimation and three on stationary processes, in the spring of 1946. Then I was appointed as Visiting Professor in Princeton for the fall term of 1946, at Yale for the spring term of 1947, and in Berkeley for the following summer term.

In Paris I was happy to see again my old friends from 1937, with the exception of the young Doeblin, who had been killed in the beginning of the war. Lévy had had his apartment sacked by the Nazis, who had destroyed his books and papers, but he was already starting new work on stochastic processes, which were soon to give rise to a new important book [84] of 1948.

What I had to say in Paris about statistical estimation was taken from my book on mathematical methods, which was now in the course of being printed. I planned to say something about the controversial subjects of Fisher's fiducial probability and Neyman's confidence intervals, where I definitely sided with Neyman. It was a not altogether pleasant surprise that R. A. Fisher was in Paris and attended my lecture on these topics. Afterwards we had a private discussion which ended better than I had feared, perhaps partly owing to the fact that I happened to know a good eating place, of which there were not so many in Paris less than a year after the end of the war.

In my Paris lectures on stationary processes I gave, among other things, an account of Khintchine's spectral representation (8) of the covariance function and my own representation (11) of the process variable itself. It appeared that Loève had independently found the latter representation. In an appendix [90] to Lévy's book of 1948, he gave an account of his important work on this and other related topics, which somewhat later formed a chapter in his own great book [91] of 1955.

In September 1946 I made my first journey to the United States, where I was received in Princeton by Sam Wilks. I gave a course on stochastic processes, and among my audience were K. L. Chung, Ted Harris, G. A. Hunt and Sam Karlin. It was a great pleasure to work with these intelligent young men. It was the year of the bicentennial of Princeton University, and among the scientific conferences forming part of the programme was one entitled "Problems of Mathematics". Among the great number of outstanding mathematicians attending were my old friends Einar Hille, Will Feller, Jerzy Neyman and Norbert Wiener, whom I was happy to see again. There were also many new acquaintances, including outstanding probabilists and statisticians, such as J. L. Doob, Harold Hotelling and Mark Kac. There was a section on mathematical probability, where all these mathematicians made interesting contributions. I was particularly glad to meet Doob, whose work on stochastic processes I had read and admired. In a paper [24] on "Problems in probability theory", I gave a survey of old and new problems in the field. Following an invitation from Gertrude Cox and Harold Hotelling, I made my first visit to Chapel Hill, where this time I only spent a few days, and met P. L. Hsu and Herbert Robbins.

During the spring term at Yale and the summer term in Berkeley I continued my work and my lectures on stochastic processes. In Berkeley, Neyman was making preparations for the celebrated series of Berkeley Symposia on Probability and Mathematical Statistics. Among the group engaged in this work I was happy to meet people like J. L. Hodges, Erich Lehmann, Henry Scheffe and Betty Scott. They are now all well known for their outstanding scientific work.

6.3. Work in the Stockholm group. During my absence from Sweden in 1946 and 1947, Gustav Elfving of Helsingfors acted as my substitute at the Stockholm University. Through him and his fellow countryman Kari Karhunen, our group had valuable contacts with probabilistic research in Finland. In his Helsingfors thesis [59] of 1947, Karhunen gave a systematic treatment of the application of Hilbert space theory to stochastic processes, thus following up the work of Kolmogorov mentioned above in 5.2. After completing his thesis, which contains remarkable new results on stochastic integrals and stationary processes, Karhunen for some time worked in Stockholm as a member of our group. In a paper [60] of 1949 he treated a stationary process with continuous time, and obtained for this case results corresponding to those given by Kolmogorov for the case of discrete time, including the Wold decomposition as well as a necessary and sufficient condition for such a process to be purely nondeterministic. Some similar results were given in a paper [53] by 0. Hanner, another member of our group.

In a paper [26] written for the Berkeley Symposium in 1950, I considered some more general classes of stochastic processes, using Hilbert space theory. I gave a derivation of a general type of spectral representation which, in the particular case of a stationary process, seems to be the simplest so far known (cf. Doob [35], page 483).

The 1950 thesis [49] of Ulf Grenander, a member of our Stockholm group, was entitled "Stochastic processes and statistical inference", and turned out to be a pathbreaking work in this difficult and important field. It is well known that Grenander later followed up this work, e.g. in the excellent joint book [50] Statistical Analysis of Stationary Time Series of 1957, written in collaboration with Murray Rosenblatt. He has also extended his outstanding work to other related fields, in the joint book [51] Toeplitz Forms and Their Applications of 1958, written together with Gabor Szego, and in numerous other works. It was Ulf Grenander who took the initiative of publishing, in 1958, a "Harald Cramér Volume" [52], containing articles in probability and mathematical statistics written by a great number of my friends. I was very happy to receive this token of friendship. Ulf Grenander became my immediate successor as Professor in the University of Stockholm. He has since then left Sweden, and now belongs to Brown University, in Providence, R. I.

From 1950 on, I became engaged in administrative University work, which absorbed a great part of my time until I was able to leave it in 1961. Still, I managed to write a monograph on risk theory [27], published in 1955, the most important part of which is a study of the ruin problem for the Lundberg risk process (cf. above, 4.4). In the most general case, this problem leads to an integral equation of the Wiener-Hopf type, and the discussion of this equation makes it possible to obtain a number of asymptotic results valuable in the practical applications. By means of the new methods later introduced by Feller, some of these results can be obtained without the use of the Wiener-Hopf equation.

6.4. Moscow, 1955. In May 1955 the University of Moscow celebrated its bicentennial, and I was invited to attend, representing the Stockholm University. It was a great event, and it gave me an opportunity to make the personal acquaintance of the Soviet mathematicians, whose work had meant so much for the advancement of probability theory. Unfortunately, Khintchine was ill — he died shortly afterwards — but I met Kolmogorov, who gave the impression of being a great scientific personality, and I had some interesting conversations with him. I was also happy to meet other members of their probabilistic group. There was Dynkin, who was beginning his great work on Markov processes, Gnedenko who in collaboration with Kolmogorov had written the book on limit problems referred to above, Linnik who was beginning his work on large deviations so closely connected with my own work of 1937, Yaglom and Rozanov who were to do outstanding work on stationary processes, and many others. They formed a group of wonderful scientific activity, and were preparing to start their new Journal of Probability Theory and its Applications, which soon acquired an internationally leading position in the field.

In connection with this Bicentennial there was also a mathematical conference, and I gave a lecture on my recent work on the ruin problem. Later in the same year, Kolmogorov spent some time in Stockholm as the guest of our university, and gave a series of lectures on limit theorems in probability to our group.

6.5. Books on probability. Before the war there were only a small number of books on mathematical probability theory built on modern foundations. Some of these have been mentioned in previous chapters.

After the war the situation radically changed, and there has been a stream of general treatises as well as of monographs on special parts of the field. I am going to give a very brief survey, strictly limited to those books which have had direct influence on the development of my own research work. This means, of course, that many important contributions to the field will not be mentioned. In particular, the extensive literature on Markov processes and on martingales will be passed over in silence, even though I am well aware of the great importance of these subjects. In all cases I give only the year of first publication of a book in its original language.

The first general postwar treatise of probability theory was Feller's [45] which appeared in 1950, and was followed by a second volume in 1966. For the young generation of the 1950's this book was an excellent introduction into an important new field of scientific research. It contained the basic theory as well as a great number of applications, all written in the fascinating personal style of its author. The second volume gives many new results, and simplified proofs of some already known.

Loève's treatise [91] of 1955 entered deeply into the background of mathematical analysis, with chapters on measure and integration. The classical limit theorems and their modern extensions to interdependent variables were fully treated, as well as important classes of stochastic processes.

The Russian treatise [104] of 1967 by Prochorov and Rozanov, is a wellwritten and important work, taking account of the most recent developments, and showing the high class of the work in our field performed in the Soviet Union.

The early postwar monographs by Gnedenko and Kolmogorov on limit problems, and by Wiener on prediction and filtering for time series, as well as the books by Grenander-Rosenblatt and Grenander-Szego have already been referred to above.

In 1948 and 1953 there appeared three monographs on the general theory of stochastic processes, of which those written by Lévy and Doob have been briefly mentioned above. The third is a joint work by Blanc-Lapierre and Fortet. They are all three classics in the field. Lévy's book contains in particular a detailed study of the Brownian movement on the line and in the plane. Doob gives a complete account of the theory, including difficult basic measure-theoretic investigations as well as extensive chapters on the most important classes of processes. The book by Blanc-Lapierre and Fortet offers a particular emphasis in its treatment of the applications of stochastic processes to many important physical problems.

Yaglom gave in 1952 a useful introduction to the theory of stationary processes [124]. A complete monograph on this important class of processes was published by Rozanov [110] in 1963. A joint book of 1965 by Linnik and Ibragimov [55] gives in its first part a full account of the work of Linnik and his group on large deviations for sums of independent variables, while the second part deals with stationary processes. Both the Rozanov and the Linnik-Ibragimov book contain important new results, e.g. about the generalization of the central limit theorem to stationary processes. Another monograph by Rozanov [111] of 1968 on infinite-dimensional Gaussian distributions contains an excellent chapter on the problem of equivalence or perpendicularity of normal distributions in function space. In the 4th volume of the great work by Gelfand and Viljenkin on generalized functions there is an interesting chapter on generalized stochastic processes.

Finally, the autobiographies of two leading probabilists should be mentioned here. They are written by Norbert Wiener [122] of 1956, and by Paul Lévy [85] of 1970. Both contain a wealth of scientific material of the highest interest.

6.6. Stationary and related stochastic processes. My own research work in probability after 1960 followed mainly two different lines, and I now propose to give a brief account of the work of others and myself on these two lines, without paying too much regard to chronological order. I will begin by talking of the work done during the 1950's and 1960's on the subject of stationary and related stochastic processes.

The development that started with Kolmogorov's and Zasuchin's work during the war, and was continued by Karhunen, Hanner and others during the late 1940's soon led to remarkable further results. By means ofHilbert space theory the properties, of the univariate stationary process were generalized to vector processes with discrete and continuous time, and to homogeneous random fields, i.e. processes with several independent parameters, such as the two coordinates of a plane, or the four coordinates of space-time, which satisfy some condition analogous to stationarity in the one-parameter case.

The spectral properties of a stationary vector process have been investigated by the Russian authors whose books have been mentioned in the preceding section, and in a number of papers by Wiener and Masani [96]. For a nondeterministic univariate process with the spectral function F it had been shown by Kolmogorov and Karhunen that the spectral functions of the purely nondeterministic and the deterministic components are respectively identical with the absolutely continuous and the jump-singular parts of F. In the investigation of the corresponding properties for a vector process one encounters difficulties connected with the rank of the never decreasing spectral matrix $\mathbf{F}$ mentioned above in 4.8. If $\mathbf{F}$ has (appropriately defined) maximum rank, the decomposition property for the univariate case can be directly generalized, as shown by Masani [94], but otherwise the situation is more complicated.

In our Stockholm group, B. Matern wrote a thesis [97] in 1960 on "spatial variation", dealing with homogeneous random fields in the two-dimensional plane, with important applications to estimation problems in forestry.

An important line of investigations started from the attempt to generalize the central limit theorem to stationary processes. In 3.4 above I mentioned the 1927 paper of Bernstein. He showed that the central limit theorem can be extended to sums of weakly dependent variables, and introduced a technique for dealing with such cases. In a paper [109] of 1956, Rosenblatt gave a remarkable definition of weak dependence, leading to the important concept of "strong mixing" for stochastic processes. Consider a process x_n with discrete time, and let $\mathbf{M}_{a,b}$ be the smallest complete σ-algebra of events (ω sets) relative to which all x_n with $a \leqslant n \leqslant b$ are measurable. As a measure of dependence between $\mathbf{M}_{-\infty,m}$ and $\mathbf{M}_{m+n,\infty}$ Rosenblatt introduces the upper bound of $|\mathsf{P}(A \cap B) - \mathsf{P}(A)\mathsf{P}(B)|$ for all events $A \in \mathbf{M}_{-\infty,m}$ and $B \in \mathbf{M}_{m+b,\infty}$. If the upper bound of this quantity for all m tends to zero as n tends to infinity, this implies that the degree of dependence between any two sections of the x_n sequence is always small when the sections lie far apart, and the process is then said to be strongly mixing. This is a stronger form of the ordinary mixing condition used in ergodic theory. The generalization to the case of continuous time is immediate. If this condition is satisfied, and some further conditions relative to the behavior of moments of the form $\mathsf{E}\Big(\sum_m^{m+n} x_i\Big)^2$ and $\mathsf{E}\Big|\sum_m^{m+n} x_i\Big|^3$ for large n are imposed, it can be shown that the sum $\sum_m^{m+n} x_i$ for any fixed m, asymptotically normal as $n \to \infty$. For the particular case when the x_n sequence is strictly stationary, it has proved convenient to introduce a "uniformly strong mixing" condition due to Ibragimov, which simplifies the proof of the central limit theorem and makes it possible to generalize the law of the iterated logarithm. These questions are discussed in the book [55] by Linnik and Ibragimov and in papers by Volkonski and Rozanov [117] and by Reznik [106]. In lectures at the Universities ofAarhus in 1967 and Copenhagen in 1969, I gave surveys of these works, and was able to complete them in some places.

All these results are obtained by using the estimation technique originally introduced by Bernstein in his paper [7] of 1927. For the particular case of normal processes it is possible to deduce more precise results, as shown e.g. in an important paper [79] by Kolmogorov and Rozanov.

When taking up work at the Research Triangle Institute of North Carolina in 1962, I became engaged in research on certain problems connected with the extreme values of stationary processes, which had applications in spacecraft navigation. As my assistant I had Ross Leadbetter, a young New Zealander, with whom I found it very stimulating to work. Our collaboration led to some promising results, and it was suggested that we should develop it and write a joint monograph on the problems involved. In planning the book we soon came to the conclusion that it would be desirable to include a fairly broad account of the general theory of stationary processes, with special emphasis on the properties of their sample functions. In particular we wanted to discuss the analytic properties of the sample functions, such as continuity and differentiability, and study the random variables expressing the number of crossings in a given time interval between a sample function and a fixed level or curve. We were able to build our work on previous research by a large number of authors, among which I may quote here Beljaev [3-6], Bulinskaja [9], Kac and Slepian [56], Rice [107, 108], Volkonski and Rozanov [117] and Ylvisaker [125, 126]. In particular the work of Rice had a basic importance for the whole field of problems concerned. For a normal stationary process which satisfies the strong mixing condition, Volkonski and Rozanov had proved the important theorem that the crossings between a sample function and a very high level approximately form a Poisson process. I had been able to show that this property holds already under simpler conditions, and that it leads to interesting conclusions concerning the magnitude and distribution of the extreme values of the sample functions. The corresponding propositions were included in our book, which appeared in 1967 under the title Stationary and Related Stochastic Processes. Many of our results have since been improved by other authors, e.g. by Beljaev, who wrote an introduction to the Russian edition of the book, by S. M. Berman in a series of significant papers on extreme values, and by G. Lindgren, who in a 1972 thesis [88] at the University of Lund gave a number of results concerning the configuration and distribution of maxima and minima of sample functions.

6.7. Structure problems for a general class of stochastic processes. From 1958 on, I had tried to generalize some of the results obtained for stationary vector processes by Zasuchin, Wiener-Masani and others, as referred to above. It seemed clear that the possibilities of a generalization would be very restricted in the case of all results connected with spectral representations. On the other hand, that part of the field which Wiener and Masani have called the "time domain" analysis of the processes seemed to be open for a fairly wide generalization to nonstationary cases.

In a paper [28] presented to the Berkeley Symposium in 1960, I discussed these questions, both for the time domain and the spectral analysis. In the latter case, I could only give some tentative results for special classes of processes. With regard to the time domain analysis, however, it was possible to proceed much further. It turns out that the situation is quite different for processes with discrete and continuous time.

Consider an arbitrary vector process with discrete time, say $\mathbf{x}_n = (x_{n1}, \ldots, x_{nq})$, where n runs through all negative and positive integers. Suppose that the components have zero mean values and finite second order moments, and define the concepts of deterministic and purely nondeterministic processes in the well-known way. Then, without imposing any further conditions, the generalization of the Wold decomposition goes through, and we have

$$\mathbf{x}_n = \mathbf{u}_n + \mathbf{v}_n,$$

where $\mathbf{u}_n$ is purely nondeterministic and $\mathbf{v}_n$ is deterministic, while $\mathbf{u}_n$ and $\mathbf{v}_n$ are mutually orthogonal. Further, the nondeterministic component $\mathbf{u}_n$ can be linearly represented in terms of innovations, in the form

$$\mathbf{u}_n = \sum_{i=-\infty}^{n} \mathbf{c}_{ni}\mathbf{z}_i,$$

where $\mathbf{z}_i = (z_{i,1}, \ldots, z_{ir+i})$ is a column vector of order $r_i \leqslant q$, while $\mathbf{c}_{ni}$ is a matrix of order $q \times r_i$. The innovation components z_{ij} are mutually orthogonal for all i and j.

In the case of a vector process $\mathbf{x}(t) = (x_1(t), \ldots, x_q(t))$ with continuous t, there is a similar decomposition

$$\mathbf{x} = \mathbf{u} + \mathbf{v},$$

with a purely nondeterministic $\mathbf{u}(t)$ and a deterministic $\mathbf{v}(t)$, which are mutually orthogonal. However, in this case the representation $\mathbf{u}(t)$ in terms of innovations turns out to be more complicated. If we assume that the mean square limits $\mathbf{u}(t+0)$ and $\mathbf{u}(t-0)$ exist for all t, it can be shown that the Hilbert space spanned by all the components of $\mathbf{u}(t)$ is separable. With the aid of Hilbert space geometry it is then possibe to show that we have for all t

$$\mathbf{u}(t) = \int_{-\infty}^{t} \mathbf{G}(t,v)d\mathbf{z}(v),$$

where $\mathbf{z}(v) = (z_1(v), \ldots, z_N(v))$ is a column vector of order N, the components of which are mutually orthogonal stochastic processes with orthogonal increments, while $\mathbf{G}(t,v)$ is a nonrandom matrix of order $q \times N$. The number N is uniquely determined by the given $\mathbf{x}(t)$ process, and may be any finite positive integer, or equal to $+\infty$. Thus, while in the discrete time case, the innovation vectors $\mathbf{z}_i$ are at most of the order q of the given vector process, the innovation process $\mathbf{z}(v)$ that occurs in the continuous case may have any order, even an infinite one.

Both in the discrete and the continuous time case, the representation of the nondeterministic component $\mathbf{u}_n$ or $\mathbf{u}(t)$ immediately gives an explicit solution of the linear least squares prediction problem. I have discussed the properties of this representation in some further papers [29, 31, 32]. It would be important to be able to determine the multiplicity N corresponding to a given $\mathbf{x}(t)$ process, and I have given some contributions to the study of this problem, but no complete solution seems to be known so far.

For the case of a normal process the representation of $u(t)$ was given by Hida [54] at the same time as my paper [28] was read before the 1960 Berkeley Symposium. Further interesting contributions have been given by Kallianpur and Mandrekar [57, 58], and quite recently by Rozanov [112]. A number of papers on this and allied subjects have been collected in a volume [39] edited by Ephremides and Thomas.

6.8. Travels and work, 1961-1970. In the early summer of 1961 I retired from my work in the university administration and became a free scientist. At the Paris meeting of the International Statistical Institute I had a conversation with Gertrude Cox, who invited me to come and work at the Research Triangle Institute in North Carolina, where I was to spend several months in each of the years 1962, 1963 and 1965. I also acted as Visiting Professor, in 1963 at Columbia, 1965 at Yale, and 1966 in Berkeley. In the two preceding sections I have already said something about the work on stochastic processes in which I was engaged during these years. In all these places I was happy to work with old and new friends.

In the summer of 1962 we had the International Mathematical Congress in Stockholm. My wife and I arranged a probabilistic lunch in our home for a number of distinguished friends, among whom were Doob, Hunt, Ito, Kappos, Kolmogorov, Linnik, Masani, Renyi, Rosenblatt, Takacs and Urbanik,

Having received an invitation to an All-Soviet Conference for Probability and Mathematical Statistics, to be held in October 1963, I came fairly directly from the United States to the Soviet Union. The meeting took place in Tbilisi (Tiflis), and had an excellent programme, including lectures by those Russian probabilists whose work I have referred to above in this paper, and also contributions by young members, who made a strong impression of scientific interest and vitality. Kolmogorov lectured about the application of probability methods to the analysis of poetical works, Yaglom, Ibragimov and Rozanov on different aspects of stochastic processes, and Linnik on statistical tests. I talked about stochastic processes as curves in Hilbert space, giving some contributions to the theory of multiplicity mentioned in the preceding section. After the meeting I also had the occasion to lecture in Moscow and Leningrad.

In 1967 I attended a conference for the applications of statistical extreme values at Faro in Portugal, with good representation both from theoretical and various practical fields. Afterwards I acted as Visiting Professor at the University of Aarhus in Denmark, and in 1969 at the University of Copenhagen. For 1970 I received an invitation from John Tukey to give the first S.S. Wilks Lecture [32], at the inauguration of the new Fine Hall building in Princeton. In this connection, my wife and I visited Chapel Hill, Princeton, Providence and Storrs, and were happy to be received by friends in all these places.

And here, at 1970, I will put an end to these scattered recollections from the development of mathematical probability theory during half a century.

REFERENCES

[1] ARLEY, N. (1943). Stochastic processes and cosmic radiation. Ph. D. Thesis, Copenhagen.

[2] BACHELIER, L. (1900). Théorie de la spéculation. Ann. Ecole Norm. Sup. 17 21–86.

[3] BELJAEV, Yu.K. (1959). Analytic random processes. Teor. Verojatnost. i Primenen 4 437–444.

[4] BELJAEV, Yu.K. (1960). Local properties of the sample functions of stationary Gaussian processes. Teor. Verojatnost. i Primenen S 128–131.

[5] BELJAEV, Yu.K. (1960). Continuity and Hölder's conditions for sample functions of stationary Gaussian processes. Proc. Fourth Berkeley Symp. Math. Statist. Prob. 23–33, Univ. of California Press.

[6] BELJAEV, Yu.K. (1966). On the number of intersections of a level by a Gaussian stochastic process. Tear. Verojatnost. i Primenen 11 120–128.

[7] BERNSTEIN, S.N. (1927). Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. Math. Ann. 97 1–59.

[8] BROCKMEYER, E., HALSTR0M, H.L. and JENSEN, ARNE. (1948). The life and works of A.K. Erlang. Trans. Danish Acad. Tech. Sci. no. 2, 277 pages.

[9] BULINSKAJA, E.V. (1961). On the mean number of crossings of a level by a stationary Gaussian process. Tear. Verojatnost. i Primenen 6 474–477.

[10] CANTELLI, F.P. (1933). Considerazione sulla legge uniforme dei grandi numeri. Giorn. 1st. Ital. Attuari 4 no. 3.

[11] CHARLIER, C.V.L. (1905). Ueber das Fehlergesetz. Ark. Mat. Astr. Fys. 2, no. 8.

[12] CRAMÉR, H. (1919). Bidrag till utjämningsförsäkringens teori. Insur. Inspect. Sthlrn.

[13] CRAMÉR, H. (1923). Das Gesetz von Gauss und die Theorie des Risikos. Skand. Aktuarietidskr. 209–237.

[14] CRAMÉR, H. (1925). On some classes of series used in mathematical statistics. Sixth Scand. Math. Congr. Copenhagen 399–425.

[15] CRAMÉR, H. (1926). Sannolikhetskalkylen i den vetenskapliga litteraturen. Nordisk. Stat. Tidskr. 5 1–32.

[16] CRAMÉR, H. (1928). On the composition of elementary errors. Skand. Aktuarietidskr. 13–74, 141–180.

[17] CRAMÉR, H. (1934). Su un teorema relativo alia legge uniforme dei grandi numeri. Giom. 1st. Ital. Attuari 5 no. 1.

[18] CRAMÉR, H. (1936). Ueber eine Eigenschaft der normalen Verteilungsfunktion. Math. Z. 41 405–414.

[19] CRAMÉR, H. (1937). Random Variables and Probability Distributions. Cambridge Tracts in Math. 36.

[20] CRAMÉR, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilités. Actualité Sci. Indust. 736 5–23.

[21] CRAMÉR, H. (1939). On the representation of a function by certain Fourier integrals. Trans. Amer. Math. Sac. 46 191–201.

[22] CRAMÉR, H. (1940). On the theory of stationary random processes. Ann. of Math. 41 215–230.

[23] CRAMÉR, H. (1942). On harmonic analysis in certain functional spaces. Ark. Mat. Astr. Fys. 28B no. 12.

[24] CRAMÉR, H. (1945). Mathematical Methods of Statistics. Almquist and Wiksells, Uppsala. ((1946), Princeton Univ. Press.)

[25] CRAMÉR, H. (1947). Problems in probability theory. Am. Math. Statist. 18 165–193.

[26] CRAMÉR, H. (1950). A contribution to the theory of stochastic processes. Proc. Second Berkeley Symp. Math. Statist. Prob. 329–339, Univ. of California Press.

[27] CRAMÉR, H. (1955). Collective risk theory. Skandia Insurance Co., Stockholm.

[28] CRAMÉR, H. (1960). On some classes of non-stationary stochastic processes. Proc. Fourth Berkeley Symp. Math. Statist. 2 57–77, Univ. of California Press.

[29] CRAMÉR, H. (1961). On the structure of purely nondeterministic stochastic processes. Ark. Mat. 4 249–266.

[30] CRAMÉR, H. (1963). On asymptotic expansions for sums of independent random variables with a limiting stable distribution. Sankhya Ser. A 25 13–24.

[31] CRAMÉR, H. (1964). Stochastic processes as curves in Hilbert space. Tear. Verojatnost.i Primenen 9 169–179.

[32] CRAMÉR, H. (1971). Structural and statistical problems for a class of stochastic processes. S. S. Wilks lecture, Princeton Univ. Press.

[33] CRAMÉR, H. (1972). On the history of certain expansions used in mathematical statistics. Biometrika 59 205–207.

[34] CRAMÉR, H. and LEADBETTER, M. R. (1967). Stationary and Related Stochastic Processes. Wiley, New York.

[35] CRAMÉR, H. and WOLD, H. (1936). Some theorems on distribution functions. J. London Math. Soc. 11 290–294.

[36] DOOB, J.L. (1953). Stochastic Processes. Wiley, New York.

[37] EDGEWORTH, F. Y. (1905). The law of error. Proc. Cambridge Philos. Soc. 20 36–141.

[38] BINSTEIN, A. (1906). Zur Theorie der Brownschen Bewegung. Ann. Physics IV 19 371–381.

[39] EPHREMIDES, A. and THOMAS, J. B. (eds.) (1973). Random Processes. Dowden, Hutchinson, Ross, Stroudsburg.

[40] ESSEEN, C. G. (1945). Fourier analysis of distribution functions. Ada Math. 77 1–125.

[41] FELLER, W. (1935, 1937). Ueber den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung. Math. Z. 40 521–559; 42 301–312.

[42] FELLER, W. (1936). Zur Theorie der stochastischen Prozesse. Math. Ann. 113 113–160.

[43] FELLER, W. (1940). On the integro-differential equations of purely discontinuous Markoff processes. Trans. Amer. Math. Soc. 48 488–575.

[44] FELLER, W, (1943). The general form of the so-called law of the iterated logarithm. Trans. Amer. Math. Soc. 54 373–402.

[45] FELLER, W. (1950, 1966). An Introduction to Probability Theory and Its Applications 1 and 2. Wiley, New York.

[46] GELFAND, I. M. and VILJENKIN, N. JA. (1961). Generalized Functions 4. Moskva.

[47] GNEDENKO, B. V. (1939). On the theory of limit theorems for sums of independent random variables. Izv. Akad. Nauk SSSR Ser. Mat. 181–232, 643–647.

[48] GNEDENKO, B. V. and KOLMOOOROV, A.N. (1949). Limit theorems for sums of independent random variables. Gosudarstv. Izdat. Tehn.-Teor. Lit., Moscow-Leningrad.

[49] GRENANDER, U. (1950). Stochastic processes and statistical inference. Ark. Mat. 1195–277.

[50] GRENANDER, U. and ROSENBLATT, M. (1956). Statistical Analysis of Stationary Time Series. Almquist and Wiksells, Stockholm.

[51] GRENANDER, U. and SZEGÖ, G. (1958). Toeplitz Forms and Their Applications. Univ. of California Press.

[52] GRENANDER, U. (ed.) (1959). Probability and Statistics, the Harold Cramér Volume. Almquist and Wiksells, Stockholm and Wiley, New York.

[53] HANNER, O. (1949). Deterministic and nondeterministic stationary random processes. Ark. Mat. 1161–177.

[54] HIDA, T. (1960). Canonical representations of Gaussian processes and their applications. Mem. Coil. Sci. Univ. Kyoto, A 33 109–155.

[55] IBRAGIMOV, I. A. and LINNIK, Yu. V. (1965). Independent and stationarily connected variables. Izad. “Nauka”, Moscow.

[56] KAC, M. and SLEPIAN, D. (1959). Large excursions of Gaussian processes. Ann. Math. Statist. 30 1215–1228.

[57] KALLIANPUR, G. and MANDREKAR, V. (1970). On the connections between multiplicity theory and O. Hanner’s time domain analysis of weakly stationary stochastic processes. Essays in Probability and Statistics 385–396. Univ. of North Carolina Press.

[58] KALLIANPUR, G. and MANDREKAR, V. (1965). Multiplicity and representation theory of purely nondeterministic stochastic processes. Tear. Verojatnost. i Primenen 10 553–581.

[59] KARHUNEN, K. (1947). Ueber lineare Methoden in der Wahrscheinlichkeitsrechnung. Ann. Acad. Sci. Fenn. Ser. A 37 1–79.

[60] KARHUNEN, K. (1949). Ueber die Struktur stationärer zufälliger Funktionen. Ark. Mat. 1 141–160.

[61] KEYNES, J.M. (1921). A Treatise on Probability. Macmillan, London.

[62] KHINTCHINE, A.YA. (1923). Ueber dyadische Brüche. Math. Z. 18 109–116.

[63] KHINTCHINE, A.YA. (1933). Zur mathematischen Begründung der statistischen Mechanik. Z. Angew. Math. Mech. 13 101–103.

[64] KHINTCHINE, A.YA. (1933). Asymptotische Gesetze der Wahrscheinlichkeitsrechnung. Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer, Berlin.

[65] KHINTCHINE, A.YA. (1934). Korrelationstheorie der stationaren stochastischen Prozesse. Math. Ann. 109 604–615.

[66] KHINTCHINE, A.YA. (1936). Sul dominio di attrazione della legge di Gauss. Giorn. 1st. Ital. Attuari 7 3–18.

[67] KHINTCHINE, A.YA. (1937). A new derivation of a formula of P. Lévy. Bull. Moskva Univ. 1 1–5.

[68] KHINTCHINE, A.YA. (1937). On the arithmetic of distribution laws. Bull. Moskva Univ. 1 6–17.

[69] KHINTCHINE, A.YA. (1937). Zur Theorie der unbeschränkt teilbaren Verteilungsgesetze. Mat.Sb.44 79–119.

[70] KHINTCHINE, A.YA. and KOLMOGOROV, A.N. (1925). Ueber Konvergenz von Reihen, deren Glieder durch den Zufall bestimmt werden. Mat. Sb. 32 668–677.

[71] KOLMOGOROV, A.N. (1928). Ueber die Summen durch den Zufall bestimmter unabhängiger Grössen. Math. Ann. 99 309–319.

[72] KOLMOGOROV, A.N. (1929). Ueber das Gesetz des iterierten Logarithmus. Math. Am. 101 126–135.

[73] KOLMOGOROV, A.N. (1931). Ueber die analytischen Methoden in der Wahrscheinlichkeitsrechnung. Math. Ann. 104 415–458.

[74] KOLMOGOROV, A.N. (1932). Sulla forma generate di un processo stocastico omogeneo (un problema di B. de Finetti). R. Accad. Lincei 15 805–808, 866–869.

[75] KOLMOGOROV, A.N. (1933). Grundbegriffe der Wahrscheinlichkeitsrechnung. Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer, Berlin.

[76] KOLMOGOROV, A.N. (1938). A simplified proof of the Birkhoff-Khintchine ergodic theorem. Uspehi. Mat. Nauk 5 52–56.

[77] KOLMOGOROV, A.N. (1941). Stationary sequences in Hilbert space. Bull. Moskva Univ. 2 no. 6.

[78] KOLMOGOROV, A.N. (1941). Interpolation and extrapolation of stationary random sequences. Izv. Akad. Nauk SSSR 5 3–14.

[79] KOLMOGOROV, A.N. and ROZANOV, Yu.A. (1960). On strong mixing conditions for stationary Gaussian processes. Tear. Verojatnost. i Primenen 5 204–208.

[80] LÉVY, P. (1925). Calcul des Probabilités. Gauthier-Villars, Paris.

[81] LÉVY, P. (1934). Sur les intégrates dont les éléments sont des variables aléatoires independantes. Ann. Ecole. Norm. Sup. Pisa (2) 3 337–366.

[82] LÉVY, P. (1935). Propriétés asymptotiques des sommes de variables aléatoires indépendantes ou enchainées. J. Math. Pares Appl. (9) 14 347–402.

[83] LÉVY, P. (1937). Théorie de L'Addition des Variables Aléatoires. Gauthier-Villars, Paris.

[84] LÉVY, P. (1948). Processus Stochastiques et Mouvement Brownien. Gauthier-Villars, Paris.

[85] LÉVY, P. (1970). Queleques aspects de la pensée d'un mathèmaticien. Blanchard, Paris.

[86] LIAPOUNOV, A.M. (1901). Nouvelle forme du theoreme sur la limite des probabilites. Mem. Acad. Sci. St. Petersbourg. 12 (no. 5) 1–24.

[87] LINDEBERG, J.W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. Math. Z. 15 211–225.

[88] LINDGREN, G. (1972). On wave-forms in normal random processes. Thesis, Lund.

[89] LINNIK, Yu.V. See Ibragimov and Linnik.

[90] LOÈVE, M. (1948). Fonctions aléatoires du second ordre. Note a Lévy, Processus Stochastiques et Mouvement Brownien [84] 299–352. Paris.

[91] LOÈVE, M. (1955). Probability Theory. Van Nostrand, Princeton. (3rd ed., 1963.)

[92] LUNDBERG, F. (1903). Approximerad framstallning av sannolikhetsfunktionen. Aterforsakring av kollektivrisker. Thesis, Uppsala.

[93] LUNDBERG, O. (1940). On random processes and their application to sickness and accident statistics. Thesis, Stockholm.

[94] MASANI, P. (1959). Cramér's theorem on monotone matrix-valued functions and the Wold decomposition. Probability and Statistics: The Harold Cramér Volume (UlfGrenander, ed.) 175–189. Almquist and Wiksell, Stockholm; Wiley, New York.

[95] MASANI, P. (1966). Wiener's contributions to generalized harmonic analysis, prediction theory and filter theory, Norbert Wiener 1894–1964. Bull. Amer. Math. Soc. 72 73–125.

[96] MASANI, P. and WIENER, N. (1957). The prediction theory of multivariate stochastic processes. Ada Math. 98 111–150, and (1958), 99 93–137.

[97] MATERN, B. (1960). Spatial variation. Swed. Forestry Res. Inst, 49 1–144.

[98] VON MISES, R. (1919). Grundlagen der Wahrscheinlichkeitsrechnung. Math. Z. 5 52–99.

[99] VON MISES, R. (1928). Wahrscheinlichkeit, Statistik und Wahrheit. Springer, Wien.

[100] DE MOIVRE, A. (1733). Miscellanea analytica. Second supplement, London.

[101] NEYMAN, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. Philos. Trans. Roy. Soc. London 236 333–380.

[102] NEYMAN, J. (1942). Basic ideas and some recent results of the theory of testing statistical hypotheses. J. Roy. Statist. Soc. (N.S.) 105 292–327.

[103] PÓLYA, G. (1920). Ueber den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem. Math. Z. 8 171–181.

[104] PROCHOROV, Yu.V. and ROZANOV, Yu.A. (1967). Probability Theory: Fundamental Concepts, Limit Theorems, Random Processes. Izdat. "Nauka", Moscow.

[105] RAIKOV, D. (1938). On the decomposition of Gauss and Poisson laws. Izv. Akad. Nauk. SSSR, Ser. Mat. 91–124.

[106] REZNIK, M.K. (1968). The law of the iterated logarithm for some classes of stationary processes. Tear. Verojatnost. i Primenen. 13 606–621.

[107] RICE, S.O. (1945). Mathematical analysis of random noise. Bell System Tech. J. 24 46–156.

[108] RICE, S.O. (1958). Distribution of the duration of fades in radio transmission. Bell System Tech. J. 37 581–635.

[109] ROSENBLATT, M. (1956). A central limit theorem and a strong mixing condition. Proc. Nat. Acad. Sci. 42 43–47.

[110] ROZANOV, Yu.A. (1963). Stationary random processes. Gosudarstv. Izdat. Fiz.-Mat. Lit., Moscow.

[111] ROZANOV, Yu.A. (1968). Infinite-dimensional Gaussian distributions. Proc. Steklov Inst. Math. 108 1–136.

[112] ROZANOV, Yu.A. (1974). Theory of'innovation processes 1–128. Izdat. "Nauka", Moskva.

[113] RUTHERFORD, E, and GEIGER, H. (1908). An electrical method of counting the number of particles from radioactive substances. Proc. Roy. Soc. A81 141–161.

[114] SEQERDAHL, C.O. (1939). On homogeneous random processes and collective risk theory. Thesis, Stockholm.

[115] TCHEBYCHEV, P.L. (1867). On mean values. J. Math. Pures Appl. 12 177–184.

[116] TCHEBYCHEV, P.L. (1890). Sur deux theoremes relatifs aux Probabilites. Ada Math. 14 305–315.

[117] VOLKONSKI, V.A. and ROZANOV, Yu.A. (1959, 1961). Some limit theorems for random functions. Tear. Verojatnost. 4 186–207 and 6 202–215.

[118] WIENER, N. (1923). Differential space. J. Math. and Phys. 2 131–174.

[119] WIENER, N. (1930). Generalized harmonic analysis. Ada Math. 55 117–258.

[120] WIENER, N. (1934). Fourier transforms in the complex domain (with R. Paley). Amer. Math. Soc. Coil. Publ. 19.

[121] WIENER, N. (1949). Extrapolation, Interpolation and Smoothing of Stationary Time Series. With engineering applications. Wiley, New York.

[122] WIENER, N. (1956). I Am a Mathematician. The Later Life of a Prodigy. Doubleday, New York.

[123] WOLD, H. (1938). A study in the analysis of stationary time series. Thesis, Stockholm.

[124] YAGLOM, A.M. (1952). An introduction to the theory of stationary random functions. Uspehi Matem. Nauk (N.S.) 7, no. 5 (51), 3–168.

[125] YLVISAKER, N.D. (1965). The expected number of zeros of a stationary Gaussian process. Ann. Math. Statist. 36 1043–1046.

[126] YLVISAKER, N.D. (1966). On a theorem of Cramér and Leadbetter. Ann. Math. Statist. 37, 682–685.

[127] ZASUCHIN, V. (1941). On the theory of multi-dimensional stationary random processes. Dokl. Akad. Nauk USSR 33 435–437.