Probability and Statistics
Soviet Essays

Selected and Translated by Oscar Sheynin
Berlin 2004

Contents

Foreword

I am presenting a collection of Russian essays on probability and statistics written by most eminent Soviet authors, mostly on occasions of official anniversaries and devoted to Soviet advances in these disciplines. Many essays also dwelt on events in pre-revolutionary Russia and briefly described the findings of Chebyshev, Markov and Liapunov and in a few cases mentioned Buniakovsky and Ostrogradsky. Regrettably, however, Daniel Bernoulli, who published seven pertinent memoirs in Petersburg, was passed over in silence as was Euler, the author of an important commentary on one of these memoirs.

Again, the information concerning foreign findings was scant and, in addition, some of the essays were politically influenced. Nevertheless, I hope that this collection will be extremely useful for those interested in the development of probability and statistics during the 20th century.

In some instances I changed the numeration of the formulas and I subdivided into sections those lengthy papers which were presented as a single whole; in such cases I denoted the sections in brackets, for example thus: **[2]**. My own comments are in curly brackets.

In the sequel, I am using abbreviations

L = Leningrad; M = Moscow; (R) = in Russian.

Almost all the translations provided below were published in microfiche collections by Hänsel-Hohenhausen (Egelsbach; now, Frankfurt/Main) in their series Deutsche Hochschulschriften NNo. 2514 and 2579 (1998); 2656 (1999), 2696 (2000); and 2799 (2004). The copyright to ordinary publications remained with me.

Acknowledgement. Dr. A.L. Dmitriev (Petersburg) sent me photostat copies of some of the Russian papers published in sources hardly available outside Russia.

It is not amiss to mention my earlier translations of two more Russian essays (only one of them dealing with the Russian scene):

1) Bernstein, S.N. The Petersburg school of the theory of probability (1940). In *Probability and Statistics. Russian papers*. Berlin, 2004, pp. 101 – 110.

2) Khinchin, A.Ya. Mises' frequentist theory and the modern concepts of the theory of probability (posthumous, 1961). *Science in Context*, vol. 17, 2004, pp. 391 – 422. The eager editorial staff somewhat tampered with Khinchin's text and on p. 42 wrongly rendered *restricted* postulate as *restricting*.

More important, however, is to list the recently published books of which I was either author or translator. All of them were put out in Berlin by NG Verlag; each was printed in 50 copies which were distributed to my colleagues and reputed libraries the world over. The same will hopefully happen with the proposed 50 copies of this book.

1. Sheynin, O. *History of the Theory of Probability to the Beginning of the 20th Century*, 2004.

2. Sheynin, O. [Translation of my] *Russian Papers on the History of Probability and Statistics*, 2004.

3. Chuprov, A. [Collected] *Statistical Papers and Memorial Publications*, 2004.

4. Chebyshev, P.L. *Theory of Probability. Definite Integrals; The Theory of Finite Differences; Theory of Probability*, 2004. (Translation of his lectures read 1879/1880 and published in Russian in 1936.)

5. Nekrasov, P.A. *The Theory of Probability. Central Limit Theorem; Method of Least Squares; Reactionary Views; Teaching of Probability Theory; Further Developments*, 2004.

6. *Probability and Statistics*. [Collected Translations of] *Russian Papers*, 2004.

7. *Probability and Statistics*. [Collected Translations of] *Russian Papers of the Soviet Period*, 2005.

1. S.N. Bernstein. The Present State of the Theory of Probability and Its Applications (1928)

Report at the All-Russian Congress of Mathematicians Moscow, 28 April 1927

Собрание сочинений (Coll. Works), vol. 4. N.p., 1964, pp. 217 – 232 …

The Organizing Committee invited me to deliver a general report on the present state of the theory of probability and its applications to the Congress.You know how vast is this subject and you should not therefore expect any comprehensive analysis of its entire field from me. Indeed, this would have demanded several speakers and much more time than I have at my disposal. Under such conditions my main difficulty lies in choosing the material and selecting the guiding principle of its exposition. I believe that in this case my main aim should be to try to present some synthesis methodologically uniting the general mathematical problems of the theory of probability and its most important applications. I shall therefore have to pass over in silence many special mathematical studies, and, when considering the applications, I shall pay more attention to problems of fundamental significance than to those playing an important part in some sphere.

[1] Not so long ago, even until the second half of the last {the 19th} century, the importance of the theory of probability as a method of scientific investigation was very restricted; isolated attempts at applying it for studying phenomena of nature connected with the names of Bernoulli, Laplace, Poisson, Quetelet and others were rather weakly justified and gave rise to well-deserved criticisms which found their most brilliant expression in Bertrand's generally known Préface to his course written only 40 years ago [1].

Bertrand's scepticism did not, however, arrest the further spontaneous, if I may say so, penetration of the theory into various domains of science. Already his contemporaries, Maxwell and Boltzmann, transform molecular statistics into an important and experimentally justified branch of physics; and, on the other hand, owing to the discovery of the elementary Mendelian law of heredity, the application of probability theory in biology becomes not only possible but necessary. By the same time, extensive statistical data revealing a mysterious connection between randomness and regularity had been collected in almost every field of knowledge (astronomy, meteorology, demography, etc). An analysis of this connection, the problem of classifying and characterizing statistical series on the basis of stochastic notions becomes topical.

At present, we may say quite definitively that without applying probability theory the further systematization of human knowledge or the development of science are impossible. In the first place, therefore, we should ascertain whether science, based on this theory, is not a second-rate ersatz. A formal logical substantiation of the theory of probability as a united mathematical discipline becomes therefore especially important. Only after purging it from paradoxes and axiomatically constructing it like geometry was, will it be possible to employ the theory as a rigorous cognitive means whose applicability in each particular case will both demand and allow experimental and mathematical checking.

[2] A purely mathematical theory of probability may be uninterested in whether or not the coefficient called mathematical probability possesses any practical meaning, subjective or objective. The only demand that must be observed is the lack of contradictions: when keeping to the admitted axioms, various methods of calculating this coefficient should lead under given conditions to one and the same value.

In addition, if we want the conclusions of the theory to admit of empirical checking rather than to remain a *jeu d'esprit*, we must only consider such sets of propositions or judgements about which it is possible to establish whether they are true or false. The process of cognition is intrinsically irreversible, and its very nature consists in that some propositions that we consider possible become true (*i.e.*, are realized) whereas their negation thus become false or impossible.

It follows that the construction of the theory of probability as a single method of cognition demands that the truth of a proposition be uniquely, without any exceptions, characterized by a certain maximal value of the mathematical probability which is assumed to be equal to unity, and that the falsity of an assertion be identical with its minimal probability set equal to zero. For the case of finite totalities of propositions it is not difficult to satisfy these demands; however, since their absolute necessity was not recognized clearly enough, paradoxes cropped up (and continue to appear) when considering infinite sets. I venture to dwell on one of them.

[3] I bear in mind the well-known problem about determining the probability that a fraction chosen at random will be irreducible. Markov, in his classical treatise, offered its solution following Kronecker and Chebyshev. It is based on the assumption that all the remainders, occurring when dividing a randomly chosen number N by an arbitrary number a ($a < N$), are equally probable no matter what is the remainder obtained by dividing N by b, a number coprime with a. It is not difficult to conclude that the required probability is $6/\pi^2$ which is the limit of the infinite product

$$\prod_{p=2}^{\infty} [1 - (1/p^2)]$$

where the p's are prime numbers. However, given the stated assumptions, the probability of N being a prime is zero which is the limit of the product

$$\prod_{p=2}^{\infty} [1 - (1/p)].$$

Just the same, the probability of $N = 2p, 3p, \dots$ is also zero. Therefore, the probability that N is a product of two primes, equal to the limit of a sum of a finite number of zeros, should be zero. In a similar way, the probability that N consists of three, of four, … factors is also zero. Therefore, again applying the addition theorem, we find that, in general, the probability that an arbitrary number N is a product of a finite number of prime factors is zero. On the other hand, it is doubtless that a finite number consists of a finite number of factors, so that we encounter a contradiction: $0 = 1$.

Thus, Markov's assumptions are obviously unacceptable. In essence they are tantamount to supposing that all the values of an integer chosen at random are equally possible. Then, however, the probability of a definite value of that number is $1/\infty = 0$ so that the probability that it will not exceed any number N given beforehand is also zero. However, since a given number cannot be infinite, we arrive at the same contradiction: $0 = 1$.

From the point of view of the theory of probability the result obtained by Chebyshev and Markov and the conclusion necessarily connected with it that the probability for any number to be prime is zero, should therefore be considered senseless. And, had we nevertheless wished to insist on its correctness, physicists and statisticians could have rightfully told us that then our law of large numbers with the Bernoulli theorem that only states that some probabilities are very close to zero, cannot claim to possess a serious experimental or practical meaning. Actually, the result obtained should have been formulated as follows: If all the values of an integer N lesser than a given number are equally probable, then n can be chosen to be so large that the probability that N is prime will be arbitrarily close to zero. This proposition is similar in form to the Bernoulli and the Laplace theorems whereas its short statement above would have quite corresponded with such an *inadmissible* form of the Laplace limit theorem: For any integer number N of independent trials

$$P(t_0 < \frac{m - Np}{\sqrt{2Npq}} < t_1) = (1/\sqrt{\pi}) \int_{t_0}^{t_1} \exp(-t^2)dt$$

where p is the probability of an event A in each trial, m, the number of its occurrences in these N trials, and $q = 1 - p$, the probability of the event contrary to A.

Psychologically, the mistake made by Markov and other mathematicians, who apply the term *probability* in number theory, is quite understandable. For them, probability is not a magnitude that, preserving one and the same meaning in all applications, always admits of being measured by uniform and objective statistical methods. In the number theory, we have to do with given regular number sequences, and we are interested not in probabilities, never to be determined experimentally, but rather in the limiting, asymptotic frequencies of numbers of some class, regularly distributed in these sequences. These limiting frequencies present some similarities with mathematical probabilities heuristically very valuable for the theory, but the confusion of these two concepts is an inadmissible misunderstanding.

[4] Analogous paradoxes occur also in some geometrical problems of probability theory. An investigation of this subject from the viewpoint of the theory's unity and possibility of an empirical checking of its deductions leads to the conclusion that the theory of probability is far from being able to consider all abstract sets. However, under appropriate restrictions and taking some precautions, on which I cannot dwell here, an arithmetization of infinite sets, *i.e.*, the determination of the probabilities of all their sensible propositions, is possible without contradiction. I shall only remark that the main source of the paradoxes was that the arithmetization of infinite sets had been carried out more or less intuitively rather than by distinctly formulating which of the two principles, *continuity* or *discontinuity*, was chosen as its foundation. The former corresponds to the assumption that it is senseless to state that two magnitudes are equal one to another because in an experiment equalities can only be realized with some error and not absolutely precisely. On the contrary, the latter is applicable to magnitudes whose precise equality admits of actual checking; their totality is always calculable and they cannot be equally probable (if the totality is infinite). In each particular case, only experiment can and should decide which of the two hypotheses is true.

Thus, the theory of radiation of a blackbody, issuing from the principle of continuity, arrives at the Rayleigh law of distribution of radiant energy. However, since this law did not conform to experiment, Planck was compelled to assume the principle of discontinuity, or of step-wise changes of energy, and to create his celebrated quantum theory that was entirely corroborated not only by all the phenomena of radiation, but also by the properties of heat capacity of bodies at all possible temperatures. Classical mechanics, whose laws were derived by observing the movement of bodies of finite dimensions, demands essential changes in order to interpret phenomena connected with radiation. A deeper reason for this consists in that this discipline leads to a uniform distribution of energy among an infinite number of degrees of freedom, *i.e.*, in essence, to the contradictory assumption of an infinite set of equally possible incompatible cases, see above.

Quantum mechanics encounters serious difficulties as well; these, however, are not in the plane of probability theory but in the sphere of our mechanical and geometric ideas which we desire to apply to such elements as the electron, never observed all by themselves. In any case, from the standpoint to be developed below, it is not necessary to seek for a definite mechanical-geometric model of the atom; we may be satisfied by a physical theory constructed on a pattern of a harmless play between all the electrons of a given body with its rules being dictated by observed macroscopic phenomena.

[5] Returning to the principles of probability theory, it may be thought, as it seems to me, that the main formal logical difficulties of its construction have by now been surmounted [3]. It is not enough, however, to acknowledge that, when keeping to the well-known rules of calculation accepted by the theory, we may without logical contradictions attribute definite probabilities to various facts. Is there any physical sense in saying that two different facts possess equal probabilities? Do not we sin against the law of causality when stating that in two trials, when throwing two identical dice, the probability of a six is the same in both cases whereas the actual outcomes were a six and a five?

Much attention, both previously and recently, was given to this question important both theoretically and practically. If we are now closer to some like-mindedness with respect to its solution than we were several decades ago, we owe it not to deeper philosophical reasoning as compared with the deliberations expressed by Laplace or Cournot, but to the experimental successes of physical statistics. Beginning with Galileo and Newton, mathematicians never held the principle of causality in special esteem. For us, much more important are the functional dependences or equations in several magnitudes allowing to determine any of them given the other ones and supposing that the elements not included in the equation do not influence the value of the magnitude sought. The so-called laws of nature, as for example the law of inertia or the Newton law of universal gravitation, whose cause he did not think fit to seek for, are expressed by dependences of such a type. Einstein united both these laws and now they comprise the highest synthesis of the general theory of relativity, although we remain as far as Newton was from knowing their cause.

[6] *The new contemporary stage in the development of scientific thought is characterized by the need to introduce the notion of probability into the statements of the elementary laws of nature.* And since we do not inquire into the cause of the law of inertia which is a property of the four-dimensional Minkowski space, I think that we may just as well assume, as a characteristic of the isotropy of space, the law expressing that, when occurring far from the attracting masses, inertial movement can with equal probabilities take place within each of two equal angles.

In a similar way, the postulate on the existence of independent magnitudes and phenomena, without which no general law of nature can be formulated, and lacking which the very category of causality would have become senseless, must be precisely stated in the language of the theory of probability. For example, the independence of two points moving under their own momenta is expressed by equal probabilities of all the values of the angle between their velocities [4].

Such are the initial assumptions of statistical mechanics, which, in the kinetic theory of gases, led Maxwell to his well-known law of distribution of molecular velocities. They are fully corroborated experimentally by their corollaries emerging in accord with the calculus of probability. The paradox of the irreversibility of heat processes which are caused by reversible molecular movements can be fully explained by two equivalent suppositions, – by the Jeans hypothesis of molecular chaos and by the Ehrenfest quasi-ergodic hypothesis, both of them consistently advancing the principle of continuity.

The study of the proper motions of stars revealed that the hypothesis of the isotropy of the stellar space in the above sense should be altered owing to the existence of an asymmetric field of attraction. Nevertheless, Eddington and Charlier, in developing Poincaré's idea, established interesting similarities between molecular and stellar motions. Indeed, the isotropic Maxwellian law of the distribution of velocities is replaced in the latter phenomenon by an analogous Schwarzschild ellipsoidal law of distribution of stellar velocities which rather well conforms to astronomical observations.

Thus, in general, along with the laws of nature according to which, under certain conditions α and arbitrary circumstances of any other kind, the occurrence of a definite result A in all trials is necessary, we also admit such laws which do not always lead to the occurrence of the event A given the corresponding conditions α. However, in this second case, no matter what are the other circumstances, all trials are characterized by some uniformity of the link between α and A that we express by stating that conditions α determine the probability of the event A.

The Mendelian law of heredity is of such a kind. It states that, when interbreeding hybrids of some definite species, for example, of lilac beans, with each other, the probability of the appearance of white beans (that is, of individuals belonging to one of the pure races) is 1/4. It is out of order to dwell here on the genetic foundations of the Mendelian theory, or, in general, on the various justifications that can guide a researcher when he assumes that in all trials of a certain type the probability of the occurrence of the event A is one and the same.

[7] For us, it was only important to ascertain that there exist various trials of such a kind that the probability of the occurrence of a given event has one and the same quite definite value; dice, urns or playing cards can serve for statistical experiments proving obvious and methodologically useful checking of some conclusions of the probability theory. The general postulate according to which such checks are made, consists in that the facts having probability close to zero occur very seldom; and if this probability is sufficiently low, they should be considered practically impossible, as in the well-known Borel's example about monkeys typing out a poem [5].

Issuing from this postulate, we must study how to check whether in a given concrete series of trials the probability of the occurrence of the event A has one and the same value p. To this end, it is first of all necessary for the ratio m/n of the number m of the occurrences of A to the number n of the trials, as n increases, to approach p in accord with the Bernoulli theorem.

Some, and especially English statisticians formulate this property more categorically and even consider it as the definition of probability that runs approximately thus: The probability of the event A in each of the trials whose number increases unboundedly is the limit of the frequency m/n if it exists under this condition [6]. In my opinion, this formula suffers from grave defects and is therefore unacceptable. First, the existence of a limit cannot be proved empirically; the fraction m/n can considerably fluctuate until n amounts to many millions and only then begin to approach slowly its limit. And, on the contrary, m/n can be very stable in the beginning, and then, owing to the appearance of some perturbational causes, considerably deviate from the value that we would have been apt to regard as its limit. Thus, the definition above, that attempts to avoid the main problem on the possibility of a fundamental uniformity of trials leading to contrary results, provides no grounds at all either for statistical experiments or for conclusions exceeding the bounds of crude empiricism.

In addition, when theoretically admitting the existence of a limit, we assume a more or less definite order in the series of the trials under consideration and thus implicitly introduce an obscure idea of some special regular dependence between consecutive and supposedly independent trials. All this resembles the concept that formed the basis for the considerations on probability theory stated by the philosopher Marbe [7]. We are therefore dealing not with probability that would have characterized some general property of all our trials independent of their order, but with something akin to limiting frequencies of the number theory; these, as we saw, cannot be without logical contradictions identified with mathematical probabilities. Therefore, from our point of view, we can only assume that the event A continues to have a constant probability in spite of some visible differences in organizing our trials, when, *separating all of them up* into appropriate groups, we note a certain definite stability of the corresponding ratios m/n. In particular, if we suppose that the trials are *independent* and if the number of the groups is sufficiently large, it is necessary that the so-called coefficient of dispersion be close to unity [8]. This indication introduced by Lexis is the first important step in scientifically treating statistical data.

Not only in special trials with playing cards, dice, urns, etc, but also in biological experiments of interbreeding, this coefficient had indeed been close to unity thus corroborating the correctness, or, more precisely, the admissibility of the Mendelian hypothesis and the direct applicability of the concept of mathematical probability in its simplest form to phenomena of heredity. On the contrary, in most of the series occurring in practical statistics the stability of the ratio m/n is less robust, the coefficient of dispersion is larger than unity, or, as it is said, the dispersion is not normal any longer, but supernormal. Examples such as the frequency of boys among all the newly-born, in which the dispersion under an appropriate choice of conditions is normal, are almost exceptional. Thus, apart from the cases such as the just mentioned instance, in which the sex of the newly-born is apparently determined by some biological law of probability almost independent of the economic conditions, we are unable, at least today, to isolate in most social phenomena such classes of independent facts whose probabilities are constant. This is because, as distinct from biological and especially physical experiments, where we can almost unboundedly increase the number of objects being in uniform conditions, in communal life we cannot observe any substantial populations of sufficiently uniform individuals all of them having quite the same probabilities with respect to some indication.

In general, the supernormal dispersion is therefore a corollary of the change of probability from one object to another one. Until we have no theoretical directions about the nature and the conditions of this variability, it is possible to suggest most various patterns of the laws of probability for interpreting the results of statistical observations. Previous adjoining investigations due to Poisson were recently specified and essentially supplemented.

Keeping to the assumption of independence of the trials it is not difficult to become certain that, as n increases, the frequency m/n considered above approaches the mean group probability. And, if this probability persists in the different groups, the dispersion must be even less than normal. This case, which happens very seldom, is of no practical consequence; on the contrary, the dispersion becomes supernormal if the group probabilities are not equal one to another. Depending on the law of variability of the mean probability from one group to another one, some definite more or less involved expressions for the dispersion are obtained; these were studied by Prof. Yastremsky [9]. In general, it is possible to say that, if the elements of a statistical population are independent, the more considerable is the mean variability of their probabilities, the more does the dispersion exceed normal. And, whatever is the law of this variability, the coefficient of dispersion increases infinitely with the number of the elements of the group if only the mean square deviation of the probabilities from their general mean does not tend to zero. Therefore, when dealing with vast groups, it is possible to state, if only the coefficient of dispersion is not too large, that the perturbational influence of the various collateral circumstances that change the probabilities of individuals, is small, and to measure the mean {values} of these perturbations.

Thus, in many practical applications where the dispersion is supernormal, it is still possible to apply simple patterns of probability theory considering them as some approximation to reality which is similar to what technicians do when employing theoretical mechanics. In this connection those statistical series for which the dispersion is more or less stable even if not normal are of special interest.

Statistical populations of such a kind, as shown by Markov in his well-known investigations of dependent trials, can be obtained also in the case of one and the same probability for all the individuals of a population if only these are not quite independent one of another. Before going on to study dependent trials it should be noted, however, that in the opposite case the normality of the dispersion is in itself only one of the necessary corollaries of the constancy of the probabilities. We can indicate an infinity of similar corollaries and, for example, instead of the sum of the squares of the deviations included in the dispersion it would have been possible to consider any powers of these deviations, *i.e.*, the so-called moments of consecutive orders whose ratios to the corresponding expectations, just as the coefficient of dispersion, should be close to unity.

In general, when the probability p is constant, and considering each of a very large number S of large and equal {equally strong} groups of n individuals into which all the statistical population is separated, we know that, because of the Laplace limit theorem, the values of the deviations $(m - np)$ for each group must be distributed according to the Gaussian normal curve the more precisely the larger are the numbers S and n.

[8] In cases of supernormal dispersion discussed just above the normal distribution is not theoretically necessary, however, once it is revealed, and the coeffcient of dispersion is stable, it may be interpreted *by keeping to the hypothesis of a constant probability only assuming that there exists a stochastically definite dependence between the individuals.*

Indeed, it can be shown by developing Markov's ideas that both the law of large numbers and the Laplace limit theorem persist under most various dependences between the trials. In particular, this is true whatever is the dependence between close trials or individuals if only it weakens sufficiently rapidly with the increase in the distance from one of these to the other one. Here, the coefficient of dispersion can take any value.

Markov himself studied in great detail the simplest case of dependent trials forming (in his terminology) a *chain*. The population of letters in some literary work can illustrate such a chain. Markov calculated the frequency of vowels in the sequences of 100,000 letters from Aksakov's *Детские годы Багрова-внука* (Childhood of Bagrov the Grandson) and of 20,000 letters from {Pushkin's} *Евгений Онегин* (Eugene Onegin). In both cases he discovered a good conformity of his statistical observations with the hypothesis of a constant probability for a randomly chosen letter to be a vowel under an additional condition that this probability appropriately lowers when the previous letter had already been a vowel. Because of this mutual repulsion of the vowels the coefficient of dispersion was considerably less than unity. On the contrary, if the presence of a certain indicator, for example of an infectious disease in an individual, increases the probability that the same indication will appear in his neighbor, then, if the probability of falling ill is constant for any individual, the coefficient of dispersion must be larger than unity. Such reasoning could provide, as it seems to me, an interpretation of many statistical series with supernormal but more or less constant dispersion, at least in the first approximation.

Of course, until we have a general theory uniting all the totality of data relating to a certain field of observations, as is the case with molecular physics, the choice between different interpretations of separate statistical series remains to some extent arbitrary. Suppose that we have discovered that a frequency of some indicator has a normal dispersion; and, furthermore, that, when studying more or less considerable groups one after another, we have ascertained that even the deviations also obey the Gaussian normal law. We may conclude then that the mean probability is the same for all the groups and, by varying the sizes of the groups, we may also admit that the probability of the indicator is, in general, constant for all the objects of the population; independence, however, is here not at all obligatory. If, for example, all the objects are collected in groups of three in such a way that the number of objects having indicator *A* in each of these is always odd (one or three), and that all four types of such groups (those having *A* in any of the three possible places or in all the places) are equally probable, then the same normality of the dispersion and of the distribution of the deviations will take place as though all the individuals were independent and the probability of their having indicator *A* were 1/2. Nevertheless, at the same time, in whatever random order our groups are arranged, at least one from among five consecutive individuals will obviously possess this indicator.

It is clear now that all the attempts to provide an exhaustive definition of probability when issuing from the properties of the corresponding statistical series are doomed to fail. The highest achievement of a statistical investigation is a statement that a certain simple theoretical pattern conforms to the observational data whereas any other interpretation, compatible with the principles of the theory of probability, would be much more involved.

[9] In connection with the just considered, and other similar points, the mathematical problem of generalizing the Laplace limit theorem and the allied study of the conditions for the applicability of the law of random errors, or the Gaussian normal distribution, is of special importance. Gauss was not satisfied by his own initial derivation of the law of randomness which he based on the rule of the arithmetic mean, and Markov and especially Poincaré subjected it to deep criticism [10]. At present, the Gauss law finds a more reliable justification in admitting that the error, or, in general, a variable having this distribution, is formed by a large number of more or less independent variables. Thus, the normal distribution is a corollary of the limit theorem extended onto the appropriate sums of small variables.

I shall not tire you with the precise formulation of the results concerning the necessary and sufficient conditions for the applicability of the limit theorem as obtained by various mathematicians. Research in this field distinguished by extreme subtlety and deepness is linked with the main problem in analysis and makes use of two externally different methods. One of these, the method of expectations of the consecutive powers, or of moments, whose idea is due to {Bienaymé and} Chebyshev, underlies Markov's fundamental works. It consists in solving a system of an infinite number of equations in an infinite number of unknowns by the algorithm of continued fractions; the solution is directly connected with the problem of summing everywhere divergent Taylor series. The second method applied by Liapunov, that of characteristic functions, is based on the Dirichlet discontinuity factor that connects the calculation of the limiting probability with the theory of improper integrals and trigonometric series [11].

The just mentioned scholars had investigated the case of sums of independent variables with an exhaustive completeness, and the later work of Lindeberg, Pólya and others, without introducing essentially new ideas, has only simplified some proofs and provided another, sometimes more general formulations for the results of Liapunov and Markov.

Here, I shall only note one corollary of the Liapunov theorem especially important for the statistical practice and, in particular, for justifying the method of sampling: For any distribution of the values of some main {parent} population, the arithmetic mean of these values, when the number of observations is sufficiently large, always obeys the Gauss law.

The investigation of sums of dependent variables, an example of which I have considered above, presents special difficulties. However, in this field rather considerable findings had also been already obtained. In particular, they allow to explain why most of the curves of distribution of indications occurring in more or less uniform biological populations, as already noticed by Quetelet, obey in the first approximation the Gauss law. By similar methods {?} it became also possible to substantiate mathematically the theory of normal correlation whose main formulas were indicated by Bravais and applied by Galton for studying the phenomena of heredity. I shall not expound here the Galtonian statistical theory of heredity which Pearson developed later in detail. Its essence consists in his law of hereditary regression according to which normal correlation exists between the sizes of some quantitatively measured indication in parents and offspring. At present, owing to the experiments connected with the Mendelian theory, it should be considered as experimentally established that the Galtonian theory is not as universal as Pearson, who based his opinion on his numerous statistical observations, thought it was.

However, the abovementioned mathematical investigations enable us to prove that, even if the Mendelian law is not the sole regulator of the inheritance of elementary indications, the Galtonian law of hereditary regression must be applicable to all the complicated indications (for example, to the stature of man) made up of a large number of elementary ones. The same theorems explain why Pearson and his students could have also statistically revealed, in many cases, the existence of normal correlation between the sizes of various organs in individuals of one and the same race. The investigations also show that both the Gaussian normal curve and the normal correlation are only the limiting cases of some general theoretical patterns so that the actually observed more or less considerable deviations from them are quite natural.

[10] Thus we approach a new cycle of problems in the theory of probability which comprises the theories of distribution and of the general non-normal correlation. From the practical viewpoint the Pearsonian British school is occupying the most considerable place in this field. Pearson fulfilled an enormous work in managing statistics; he also has great theoretical merits, especially since he introduced a large number of new concepts and opened up practically important paths of scientific research. The justification and the criticism of his ideas is one of the central problems of the current mathematical statistics. Charlier and Chuprov, for example, achieved considerable success here whereas many other statisticians are continuing Pearson's practical work definitively lousing touch with probability theory; uncritically applying his formulas, they are replacing science by technique of calculation.

The purely theoretical problem of analytically expressing any statistical curve, just as any problem in interpolation, can always be solved, and by infinitely many methods. And, owing to the more or less considerable discrepancies allowed by the theory of probability, it is quite possible, even when only having a small number of arbitrary parameters at our disposal, to obtain a satisfactory theoretical curve. Experience shows that in many cases this can be achieved by applying the Pearsonian curves which depend on four parameters; theoretically, however, in the sense of the corresponding stochastic pattern, these are only justified when the deviation from the normal curve is small. It would be interesting therefore to discover the cause of the conformity for those cases in which it actually exists given a large number of observations (Bernstein 1926).

On the other hand, Bruns' theory supplemented by Charlier that introduces a perturbational factor into the Gauss or the Poisson function provides a theoretical possibility for interpolating any statistical curve. However, for a curve considerably deviating from the normal curve, a large number of parameters can be necessary, and, moreover, in this case the theoretical meaning of the perturbational factor becomes unclear. Thus, excluding curves approaching in shape the Gauss curve, or the Poisson curve [12], interpolation of statistical distributions is of an empirical nature and provides little help in understanding the essence and regularities of the phenomena considered.

Of a certain interest is therefore the rarely applied method suggested by Fechner [13] and employed later by Kapteyn and some other authors. It consists in that, by an appropriate change of the variable, the given statistical curve is transformed into a normal curve. Indeed, we have seen that very diverse patterns of the theory of probability lead to the normal distribution so that it is natural to expect, and especially in biology, that in many cases when the measured variable does not obey the Gauss law, it can in one or another way be expressed as a function of one (or of a few) normal random variable(s). Without restricting our efforts to mechanical interpolation, but groping for, and empirically checking theoretical schemes corresponding to the statistical curves {I omit here a barely understandable phrase}, we should attempt to come gradually to an integral theory of the studied phenomena. In this connection, molecular physics is very instructive and it should serve as a specimen for theoretical constructions in other branches of statistics.

[11] The main causes simplifying the solution of the formulated problems in physics are, first, the hardly restricted possibility of experimentation under precisely determined conditions [14]. The second favorable circumstance is the enormous number of elements, molecules or electrons, with which physics is dealing. The law of large numbers, when applied to bodies of usual size, – that is, to tremendous statistical populations, – thus leads to those absolutely constant regularities which until recently were being regarded as the only possible forms of the laws of nature. Only after physicists had managed to study experimentally such phenomena where comparatively small populations of molecules or electrons were participating, as for example the Brownian motion, and to ascertain that the deviations foreseen by probability theory actually take place, the statement that physical bodies were statistical populations of some uniform elements was turned from a hypothesis into an obvious fact.

In addition, most complete are the studies of those phenomena of statistical physics that have a stationary nature. In other branches of theoretical statistics as well we should therefore examine in the first place the curves of distribution corresponding to established conditions which regrettably do not occur often. A specimen of a stationary distribution most often encountered in statistical physics is the simple law of geometric progression, or the linear exponential law of the distribution of energy among uniform elements with one degree of freedom. The latter corresponds to a given total amount of energy possessed by the whole population of the elements under consideration. The same exponential law also regulates the process of natural decay of the atoms of radioactive substances.

A similar problem also corresponds to the economic issue of the stationary distribution of wealth among the individuals of a given society. The law discovered here by Pareto can illustrate a methodologically correct approach to constructing the theoretical curves with which the appropriate curves of economic statistics should be compared when searching for an explanation of the deviations from these curves in the peculiarities of the social structure and in the dynamics {in the moving forces and trends} in the given society.

I shall not dwell on the fundamental difference in stating the problem of the distribution of energy depending on whether we assume the Boltzmann hypothesis of continuity or the Planck hypothesis of discontinuity. I only note that the latter leads to issues in finite combinatorial analysis whereas the mathematical problem corresponding to the former consists in determining the most probable distribution of probabilities of a positive variable with a given expectation; under some general assumptions the distribution sought is exponential. When dropping the condition of positivity but additionally assigning the value of the expectation of the square, we arrive at the exponential function of the second degree, *i.e.*, to the Maxwellian law of distribution of velocities to which the Gauss normal law corresponds.

Generalizing this result further, we find that, under the same overall assumptions, the most probable curve corresponding to given moments of the first k orders is expressed by an exponential function with a polynomial of the k-th degree as its exponent. Therefore, if it occurs that in some cases the method of moments applied in statistics is not only a technical trick applied for calculation, but that the moments of several lower orders are indeed substantially constant, then an exponential curve with an exponent of the corresponding power should be regarded as a typical distribution of such a statistical population.

A general mathematical theory of stationary statistical curves does not exist yet. Their determination in some cases, as in the just considered instance, is reduced to a problem in the calculus of variations, and, in other cases, to functional and integral equations. It is natural to apply the latter method {?} in biology where some law of heredity and selection plays the part of an iterative function or operator determining the transformation of the curve of distribution of one generation into the curve of the next one. Inversely, issuing from the statistical distribution of consecutive generations, it is possible to seek the simplest iterative laws compatible with the given dynamical process. In particular, in this way it became possible to establish that the Mendelian law of heredity is almost the only such elementary law that, when selection is lacking, realizes stationary conditions beginning already with the second generation.

[12] Finally, turning over to correlation theory, it should be indicated first of all, that, excluding biological applications, most of its practical usage is based on misunderstanding. The desire to express all non-functional dependences through correlation is natural. However, no technical improvements replacing the hypothesis of normal correlation by any curvilinear correlation are attaining this goal since in probability theory the concept of correlation, according to its meaning, assumes stationarity which consists in that every variable involved possesses some fixed law of distribution.

It is therefore senseless to consider, for example, the correlation between the amount of {paper} money in a country and the cost of a given product or the mean wholesale index. As it seems to me, in such cases we should study some approximate functional dependences between several magnitudes, x, y, z, and establish whether the hypothetical functions constructed on the basis of economic considerations are indeed sufficiently stable and little depend on time or place. The role of the theory of probability in such matters is far from simple and its formulas should be applied with great care. When comparing dynamical series, the very concept of correlation should be replaced, as some authors do, by the term covariation with a purely technical descriptive meaning attached to it. In any case, the numerous studies concerning covariations are until now of a purely empirical nature and do not belong to the province of the theory of probability. In restricting the field of application of the correlation theory by more or less stationary populations, we lessen its practical importance; however, its conclusions in this {smaller} domain possess indisputable value and in some cases the correlation dependences express the same regularities as the functional dependences.

The need to complete my report which has already dragged on for an extremely long time, makes it impossible to dwell on purely mathematical and not yet fully solved problems connected with correlation theory. I hope that I was able to show that the methods of probability theory have now attained a sufficient degree of flexibility and perfection so as not to be afraid of most severe scientific criticism and to serve as a solid foundation for the further development of science. It should only be remembered that the concept of probability is a precise mathematical idea and that it should not be abused in the absence of strict preconditions for its application. Because, as apparently Poincaré put it, the theory did not offer us a wonderful gift of deriving something out of nothing [15]; it only embodies a distinctive method of stating, combining and uniting our knowledge into an harmonious mathematical system.

Notes

1. {Bernstein hardly had much knowledge of the (then not yet studied) history of probability. Thus, he did not mention De Moivre at all. And Bertrand (1888) had indeed severely criticized the theory of probability, – not only in its Préface,– but in many cases he was mistaken, see Sheynin (1994).}

2. {Its application became necessary, above all, owing to Darwin's *Origin of Species*.}

3. {This statement seems too optimistic: the Kolmogorov axiomatics was yet to appear (in 1933).}

4. I think that it is unnecessary to repeat that, owing to the generally accepted continuity of space, the values of the angles are supposed here to be physically measured rather than absolutely precise and determined arithmetically.

5. {The Editors inserted here a reference to the Russian translation of Borel (1914).}

6. {Bernstein did not mention Mises.}

7. {Possibly Marbe (1899).}

8. {The reader will encounter this coefficient time and time again. Bernstein also devoted much attention to it much later, in his treatise (1946) but in either case he did not mention that Markov and Chuprov had all but rejected the coefficient of dispersion as a reliable tool. See Sheynin (1996, §§14.3 – 14.5).}

9. {Boris Sergeevich Yastremsky (1877 – 1962). See Yastremsky (1964) and Anonymous (1962).}

10. {The second reference is perhaps to Poincaré's remark (1912, p. 171) that he borrowed from G. Lippmann to the effect that experimenters believe that the normal law is a mathematical theorem whereas the latter think that it is an experimental fact.}

11. {On Liapunov's alleged use of the discontinuity factor see his note (1901).}

12. According to Bortkiewicz' terminology, the latter corresponds to the law of small numbers (to the terms of a binomial of an increasing degree with the probability tending to zero.

13. {On Fechner see Sheynin (2004).}

14. As compared with social statistics, biology also partly enjoys the same benefit.

15. {It was Ellis (1850, p. 57) rather than Poincaré: *Mere ignorance is no grounds for any inference whatever. … ex nihilo nihil.*

References

1. Anonymous (1962), B.S. Yastremsky. An obituary. *Vestnik Statistiki*, No. 12, pp. 77 – 78. (R)

2. Bernstein, S.N. (1926), Sur les courbes de distribution des probabilités. *Math. Z.*, Bd. 24, pp. 199 – 211.

3. --- (1946), *Теория вероятностей* (Theory of Probability). M., 4th edition.

4. Bertrand, J. (1888), *Calcul des probabilités*. Paris.

5. Borel, E. (1914), *Le hasard.* Paris.

6. Ellis, R.L. (1850), Remarks on the alleged proof of the method of least squares. In author's *Math. and Other Writings*. Cambridge, 1863, pp. 53 – 61.

7. Liapunov, A.M. (1901), Answer to P.A. Nekrasov. *Zapiski Kharkov Univ.*, vol. 3, pp. 51 – 63. Translated in Nekrasov, P.A. (2004), *Theory of Probability*. Berlin.

8. Marbe, K. (1899), *Naturphilosophische Untersuchungen zur Wahrschein-lichkeitsrechnung.* Leipzig.

9. Poincaré, H. (1912), *Calcul des probabilités*. Paris. First edition, 1896.

10. Sheynin, O. (1990, in Russian), *Chuprov*. Göttingen, 1996.

11. --- (1994), Bertrand's work on probability. *Arch. Hist. Ex. Sci.*, vol. 48, pp. 155 – 199.

12. --- (2004), Fechner as a statistician. *Brit. J. Math. Stat. Psychology*, vol. 57, pp. 53 – 72.

13. Yastremsky, B.S. (1964), *Избранные труды* (Sel. Works). M.

2. A.Ya. Khinchin. The Theory of Probability

. In *Наука в СССР за 15 лет. Математика* (Science in the Soviet Union during 15 Years. Mathematics). Editors, P.S.Aleksandrov et al. Moscow – Leningrad, 1932, pp. 165 – 169

[Introduction] The theory of probability is nearly the only branch of mathematics where, as also acknowledged abroad, even the pre-revolutionary Russian science from the time of Chebyshev had been occupying a leading position. The responsibility for the maintenance and strengthening of this leading part naturally lay on the Soviet mathematicians and became even greater and more honorable since, during the last 10 – 15 years, the European scientific thought in the sphere of probability theory (Italy is here on the first place, then come the Scandinavian countries, Germany and France) considerably advanced from its infantile stage and rapidly attained the level established in Russia by the contribution of Chebyshev, Markov and Liapunov. The foremost European schools at least qualitatively even exceeded that level by coming out at once on the wide road provided by the modern methods of mathematical analysis and free from the touch of provincialism from which (in spite of the greatness of its separate findings) our pre-revolutionary scientific school nevertheless suffered.

It seems that now, after 15 years of work [1], we, Soviet mathematicians, may state that we have with credit accomplished the goal that had historically fallen to our lot. Notwithstanding the already mentioned very considerable advances of the West-European scientific thought, today also the Soviet probability theory is occupying the first place if not in accord with the *number of publications*, but in any case by their *basic role* and *scientific level*. The present-day portrait of this discipline has no features, nor has its workshop a single range of problems whose origin was not initiated by Soviet mathematics. And the level of our science is best described by noting that European mathematicians are until now *discovering* laws known to, and published by us already five – ten years ago; and that they (sometimes including scientists of the very first rank) publish results which were long ago introduced into our lectures and seminars and which we have never made public only because of considering them shallow and immature if not altogether trivial.

It is interesting to note that after the October upheaval [2] our probability theory underwent a considerable geographical shift. In former times, it was chiefly cultivated in Leningrad {in Petersburg/Petrograd} but during these 15 years Leningrad {Petrograd until 1924} did not offer any considerable achievements. The main school was created in Moscow (Khinchin, Kolmogorov, Glivenko, Smirnov, Slutsky) and very substantial findings were made in Kharkov (Bernstein) and Tashkent (Romanovsky).

Before proceeding to our brief essay on those lines of development of the theory of probability which we consider most important, we ought to warn our readers that it is not our goal to provide even an incomplete list of Soviet accomplishments in this sphere; we aim at tracing the main issues which occupied our scientists and at indicating the essence which we contributed to the international development of our science, but we do not claim any comprehensiveness here. Then, we ought to add a reservation to the effect that we restrict our description to those lines of investigation which have purely theoretical importance, and we leave completely aside all research in the field of practical statistics and other applications of probability in which Soviet science may also take pride. We consider the following issues as having been basic for the development of the Soviet theory of probability.

1. Investigations connected with the so-called limit theorem of the theory of probability, *i.e.*, with the justification of the Gauss law as the limiting distribution for normed sums of random variables. Formerly, almost exclusively considered were the one-dimensional case and series of mutually *independent* random variables. Nowadays we are able to extend the limit theorem, on the one hand, to the many-dimensional case, and, on the other hand, to series of mutually *dependent* variables. Soviet mathematicians and Bernstein in the first place both initiated these generalizations and achieved the most important discoveries {here}. Bernstein accomplished fundamental results in both directions; Kolmogorov arrived at some extensions; Khinchin justified normal correlation by the direct Lindeberg method. This sphere of issues met with a most lively response in the European literature and continues to be developed both at home and abroad.

2. For a long time now, the theory of functions of a real variable, that was worked out at the beginning of our {of the 20^{th}} century, compelled mathematicians to feel a number of deep similarities connecting the stochastic concepts and methods with the main notions of the metric theory of sets and functions. The appropriate process of *modernizing* the methodology of probability theory, its notation and terminology, is one of the most important aspects of the modern issues of our scientific domain. This process is far from being completed, but we already feel with complete definiteness how much it is offering to probability theory, – how much its development fosters the unity and the clarity of the scientific method, the harmony and the visibility of the scientific building itself. With respect to their level, even the most powerful stochastic schools (for example, the Italian school) insufficiently mastering the methods of the function theory become noticeably lower than those (the French, the Moscow school) where these methods are entrenched. It is therefore quite natural that Moscow mathematicians, most of whom had developed in the Luzin school, have marched in the front line of this movement. Slutsky should be named here in the first instance, then Glivenko and Kolmogorov. In particular, the last-mentioned had revealed, with exhausting depth and generality, the similarities taking place between the main notions of the theory of probability and the problem of metrization. Work in this direction is continuing. Kolmogorov's yet unpublished research promises to advance considerably our understanding of the limiting stochastic regularities.

3. During these years, the estimation of the probabilities connected with an infinitely continued series of trials acquired both an essential theoretical importance and a considerable practical interest. The Italian, and the Moscow mathematical schools, independently of one another, advanced the so-called *strong* law of large numbers as one of the general and main laws of probability theory, which Borel discovered in the simplest cases already long ago. However, whereas the Italians (Cantelli) did not go further than its formulation, we have minutely worked out both its connection with the usual law of large numbers and the conditions for its applicability (Khinchin). We were also the first to discover *the law of iterated logarithm* that determined, in a certain sense, the precise upper bound of the deviations of sums of large numbers of random variables from their expectations (Khinchin); and to establish that it has a very wide field of applications (Kolmogorov). These issues are recently attracting considerable attention of the European (mostly French and Italian) scientists.

4. The investigation of more subtle limiting regularities under the conditions of the classical Bernoulli pattern, and especially the study of the behavior of the distribution function at large distances from the center, which are also of a considerable practical interest, had mostly been developed during these years in the Soviet Union. We derived here a number of findings exhausting the posed problems (Smirov. Khinchin).

5. During the last years, after some interruption, the interest towards the simplest case of a series of dependent random variables known as a *Markov chains* has again strengthened. The first pertinent works had appeared abroad (Hadamard; Hostinsky), but they also met with a rapid response in our country and were supplemented (Romanovsky) and considerably generalized (Kolmogorov). Important investigations on the applicability of the limiting theorem to Markov chains (Bernstein) constitute a separate entity.

6. Kolmogorov extended the theory of Markov chains to the continual case thus converting it into a *general theory of stochastic processes*. This is one of the most remarkable achievements of Soviet mathematics in general. The new theory covers any process where the instantaneous state of a system uniquely determines the probability of any of its states at any subsequent moment irrespective of its previous history. Mathematically speaking, this theory establishes for the first time the general principles connecting the problems of stochastic processes with differential equations of definite types. Kolmogorov especially considered processes where the distribution function of the increment of the random variable remained constant, independent either of time or of the value of the variable at a given moment. He derived the general analytical form of such processes. Parallel investigations abroad were only carried out in Italy (Finetti) where some particular results were attained.

7. Along with Markov chains attention during the latest years was attracted to *stationary* series of random variables, that is, to series in which all the terms have the same expectation and the same variance and the correlation coeffi-cient between two terms only depends on their mutual location in the series.

In some aspects, these series, of essential importance for various applications, present an extension of Markov chains. Soviet mathematicians have developed their theory (which, however, is yet far from being completed). In the first instance, we derived series whose terms are connected by a recurrent linear relation which takes place with a high probability. These terms are therefore situated in the vicinity of certain sinusoids (or of combinations of such curves) which constitutes the *limiting sinusoidal law* and very interesting models of such series were constructed (Slutsky, Romanovsky). Khinchin recently proved that each stationary series obeys the law of large numbers and this fact certainly considerably strengthened the interest in them. Gelfond and Khinchin, in yet unpublished contributions, studied the properties of the Gram determinants for stationary series.

8. The interest in the so-called *congestion problems*, that is, in stochastic investigations connected with the running of generally used plants, essentially increased mostly in connection with the development of automatic telephony. By now, these studies resulted in the creation of a special theoretical chapter of the doctrine of probability, and we are therefore mentioning them here. The Moscow school (Kolmogorov, Khinchin) published a number of pertinent writings which theoretically solved sufficiently general problems.

[9] Finally, we ought to say a few words about some isolated works. In spite of their applied nature, it is difficult to pass over in silence Bernstein's remarkable investigations of heredity possessing considerable theoretical interest {contradiction!}. Kolmogorov recently solved a number of separate, and, again, theoretically important related problems. He also studied the general forms of mean values satisfying definite natural demands. His work occasioned essential response from foreign scientific circles.

We are now concluding our essay, and we repeat that it is very incomplete. We did not give their due to all the works mentioned, but we still hope that we have attained our main aim by showing that Soviet mathematics, in spite of the tenfold efforts exerted by our European comrades in competition, is firmly holding that banner of championship in probability theory which the pre-revolutionary Russian science had already deserved.

Notes

1. {The civil war ended in 1920 and scarcely any serious work had begun until then, or even until several years later.}

2. {The official term was *Great October* [new style: November 7] *Socialist Revolution*. Contrary to Russian grammatical rules, all four words were capitalized.}

3. A.N. Kolmogorov. On Some Modern Currents in the Theory of Probability

Труды второго всесоюзного математического съезда 1934г. (Proc. Second All-Union Mathematical Conference 1934), vol. 1. Leningrad – Moscow, 1935, pp. 349 – 358

[Introduction] The first, classical period in the development of the theory of probability essentially ended with the investigations of Laplace and Poisson. Then, the theory was mostly engaged in the calculation of the probabilities of various combinations of a finite number of random events. Entirely in accord with the problems studied, its mathematical tools were mainly combinatorial analysis, difference equations, and, when solving these, the method of generating functions.

Owing to their fundamental research, Chebyshev, Markov and Liapunov initiated a new direction. During that {new} period the concept of random variable occupied the central position. New analytic machinery for studying these variables, substantially based on the notion of expectations, on the theory of moments and distribution functions was created. The main objects of examination were sums of an increasing (but always finite) number of random variables, at first independent, and later dependent. Mises (1919) developed a complete theory of n-dimensional distribution functions for n random variables depending one on another, as well as the corresponding tool of n-dimensional Stieltjes integrals, and he also essentially supplemented {the theory of} limit theorems.

In our time, the Chebyshev, Markov and Liapunov direction outlined above culminated in Bernstein's investigations (1925 – 1926) who was the first to prove the main multidimensional limit theorem and to provide the most thorough and deep study of sequences of dependent variables.

Recalling that mechanics does not restrict its objects by considering systems of a finite number of material points, it would have certainly been unnatural to suppose that the theory of probability will not go beyond the patterns that only study a finite number of random variables. Excluding Bachelier (1900), who remained misunderstood, not pure mathematicians but physicists (Smoluchowski, Fokker, Planck), biologists (Fisher) [1], actuaries (Lundberg) and those applying statistics to technology (Fry) originated wider investigations. All their studies may be regarded as particular cases of a *general theory of stochastic processes*, that is, of a general theory of random changes of the states of some system in time.

Already when the state of a studied system is determined at each given moment t by the corresponding value of a single parameter x, this latter, if understood as a function $x(t)$ of time, provides an example of a *random function*. Other applications of such functions are yet very little developed, but it seems likely that they ought to be very numerous and important, in particular for the theory of random oscillations, for the construction of a statistical theory of turbulence and in quantum physics. In all these applications, the state of a system at each given moment is described by some function of a certain number of arguments; and, since the very state of the system is random, we deal with some random function already at each fixed t.

The study of random functions and, therefore, of distributions in functional spaces, inevitably leads to a certain revision of the axiomatic basis of the theory of probability. A sufficiently general *axiomatic exposition* of the fundamentals of the theory, satisfying all the requirements of modern physics and other applied fields, was created during the last decade. The guiding principle of a considerable part of the pertinent studies was, however, not the desire to cover a wide range of new applications going beyond the old boundaries, but the wish to trace, in all their generality, the recently discovered deep *similarities between a number of notions of the theory of probability and the metric theory of functions of a real variable.*

The formulation of new problems led also to the creation of new analytic tools, such as, first, integral, differential and integro-differential *equations of stochastic processes* that originated as a generalization of the Smoluchowski integral equation and the Fokker – Planck differential equation. And the second side of the new machinery should be the still very little developed *theory of characteristic functions and moments for distributions in infinite-dimensional* (in particular, in functional) *spaces*. We note in concluding that the differential and integro-differential equations of stochastic processes led to the construction of a very powerful method for *proving limit theorems* which directly adjoin the studies of the Chebyshev direction.

We can now systematize the new currents in the theory of probability which is the aim of my report.

1) Investigations that originated owing to the analogy with the metric theory of functions of a real variable.

a) The general axiomatics of the theory of probability (Borel, Fréchet, Kolmogorov, Hopf).

b) Research connected with the law of large numbers (Borel, Cantelli, Slutsky, Fréchet, Khinchin, Kolmogorov, Glivenko, Lévy).

2) New patterns created owing to the physical and other applied issues.

a) The theory of stochastic processes (Finetti, Hostinsky, Hadamard, Mises, Kolmogorov, Fréchet, Khinchin, Lévy).

b) The theory of random functions (Wiener, Slutsky, Lévy).

3) New analytic tools

a) Equations of stochastic processes (Kolmogorov, Hostinsky, Fréchet, Bernstein, Pontriagin).

b) Characteristic functions and moments in infinite-dimensional and functional spaces (Khinchin).

c) New methods of proving limit theorems (Kolmogorov, Petrovsky, Bernstein, Khinchin, Bavli).

1a. As stated above, all the problems of probability theory considered up to the last decades can be reduced to the study of a finite number of random variables. All the probabilities encountered here will be determined if we provide an *n*-dimensional distribution function of these variables. In particular, Mises, in his course in the theory of probability (1931), systematically kept to the mentioned restriction. In his terminology, this was expressed by assuming that the set of indications (*Merkmalmenge*) of an arbitrary collective was always a set of points of an *n*-dimensional space.

It is the most natural to perceive a distribution function in an *n*-dimensional space as an additive function in a domain of that space. The mathematical interpretation of a problem in the theory of probability of the indicated classical type depends exclusively on the corresponding distribution function. Therefore, if the only aim of the *axiomatic theory of probability* is the most compact and clear enumeration of the logical assumptions of subsequent mathematical constructions, then, when dealing with problems of the classical type, it would be simplest to select directly as the axioms the characteristic properties of distribution functions (non-negativity; additivity; and their being equal to 1 for the complete space). According to my deep conviction, the axiomatics of probability theory cannot have any other goals because the question about the applicability of a given mathematical pattern to some concrete physical phenomena cannot in essence be solved by the axiomatic method. Mises' attempt clearly illustrates this idea. To confine his construction within the boundaries of fixed axioms, he is compelled only to postulate the approach of the frequencies to certain limits as the trials are unboundedly continued, without saying anything about when, beginning with what finite number of repetitions, may we conclude that the former practically coincide with the latter. Indeed, an answer to this question can only be provided after going beyond the boundaries of a rigorously formal mathematical thinking. Thus, the Mises axioms, irrespective of the connected intrinsic difficulties, are, one the one hand, not needed for justifying the mathematical theory, and insufficient for substantiating its applicability on the other hand.

My system of axioms [1] is a direct generalization of the properties of distribution functions listed above. This generalization allows us to cover all those new non-classical problems described in the Introduction. I only indicate here one point concerning the axiomatic construction of the main notions of the theory of probability, – that point, which, as it seems to me, still requires to be developed.

In the applications, researchers often consider conditional probabilities which are determined under the restriction that some random variable x took a definite particular value $x = a$. If x has a continuous law of distribution, the elementary method of determining the conditional probability $P_{x=a}(A)$ of event A given that $x = a$,

$$P_{x=a}(A) = P(A|x = a)/P(x = a),$$

is inapplicable because the right side is indefinite. Nevertheless, I was able to determine $P_{x=a}(A)$ in the most general case although my definition seems to be too complicated. In addition, there exist many physical problems, where, properly speaking, only conditional probabilities are studied so that the reduction of all the applied conditional probabilities to some definite system of unconditional probabilities is altogether impossible. Such, for example, is the case of *Brownian* motion along an infinite straight line. Here, the conditional probabilities of the position of a particle $x(t_2)$ at moment $t_2 > t_1$ are known if the position of the particle $x(t_1)$ at moment t_1 is given. To reduce this problem to the pattern developed in my book [1], it is necessary to choose some initial t_0 and to assume a corresponding unconditional $(n + 1)$-dimensional law of distribution of the variables $x(t_n)$, $x(t_{n-1})$, …, $x(t_1)$, $x(t_0)$ for any $t_n > t_{n-1} > \ldots > t_2 > t_1 > t_0$. Knowing these unconditional $(n + 1)$-dimensional distributions, it will be possible to calculate, for any $t_2 > t_1 > t_0$, the conditional law of distribution $x(t_2)$ given a fixed $x(t_1)$. Actually, however, the conditional laws of distribution of this last type are indeed initially given, and, for that matter, they are known for each pair $t_2 > t_1$ rather than only for $t_1 > t_0$. An important problem therefore originates: To construct a *direct axiomatics of conditional probabilities* instead of *determining* them by issuing from unconditional probabilities.

Introductory Literature

1. Kolmogoroff, A.N. (1933), *Grundbegriffe der Wahrscheinlichkeits-rechnung*. Berlin.

2. Hopf, E. (1934), On causality, statistics and probability. *J. Math. Phys*., vol. 13, pp. 51 – 102.

3. Lomnicki, Z., Ulam, S. (1934), Sur la théorie de la mesure. *Fund. Math*., Bd. 23, pp. 237 – 278.

1b. A full analogy between the notions of probability and measure of a set; between the definitions of expectation and the Lebesgue integral; and a partial analogy between the independence of random variables and orthogonality of functions led to the possibility of transferring the methods of the metric theory of functions of a real variable {to the theory of probability} and back. As the main result, the theory of probability gained a full solution of the problem concerning the conditions for the applicability of the law of large numbers to sequences of independent random variables and the creation of the concept of the strong law of large numbers. Even for independent random variables no necessary and sufficient conditions, clear and convenient to any extent, were yet discovered for the applicability of the latter law.

For a long time the Khinchin and Kolmogorov necessary and sufficient condition for the *convergence of a series* of independent random variables was considered as being more interesting for applications in the theory of functions of a real variable. Recently, however, it was established that this condition is of basic importance for the harmonic analysis of random functions (Slutsky, Kolmogorov) and for studying random functions with independent increments (Lévy).

Bibliographic indications can be found in the last chapter of [1], see §1a.

2a. The first systematically studied general pattern of a stochastic process was the scheme of Markov chains. Already Markov himself rather thoroughly developed the mathematical side of their theory. However, only after the works of Hostinsky, Hadamard and Mises (1931) it became clear that these chains are the simplest and in many aspects typical specimens of arbitrary *stochastic processes without aftereffect*, – that is, of such processes where the knowledge of the state of the system $x(t_0)$ at moment t_0 determines the law of distribution of the possible states $x(t)$ of the system at moment $t > t_0$ irrespective of its states at moments preceding t_0.

Markov chains correspond to the case in which the states of a system are only considered at integer values of t and the number of the possible states is finite. Here, all is reduced to the probabilities $P_{ik}(t)$ of the transition from state i to state k during the interval of time between the moments t and $(t + 1)$. The more general probabilities $P_{ik}(s; t)$ of the same transition between moments s and $t > s$ are expressed through $P_{ik}(t)$. At present, a large number of generalized Markov chains have been studied, and I discuss some of them in §3a. The following pattern covers all of them.

Let E be the set of the possible states of the system, F, – the distribution function of probabilities in E. Then, for each interval of time from moment s to moment $t > s$ there exists an operator $H_{st}(F) = F_1$ by whose means we can determine the distribution function F_1 at moment t given F at moment s. This operator is inevitably linear and unitary. In addition, it obeys the equation

$$H_{su} = H_{tu} \cdot H_{st} \quad (1)$$

for any $s < t < u$. Equation (1) is indeed the main equation of stochastic processes without aftereffect. For the case of Markov chains it takes the well-known form

$$P_{ik}(s; u) = \sum_j P_{jk}(t; u) \cdot P_{ij}(s; t).$$

The general solution of equation (1) seems to be very difficult even for separate particular cases. The most important instance is the time-homogeneous case in which $H_{st} = H_{t-s}$ and in which, accordingly, equation (1) becomes

$$H_{s+t} = H_s \cdot H_t. \quad (1')$$

Equation (1′) shows that the matter indeed concerns the determination of the general form of a one-parameter group of unitary operations H_t in the space of distribution functions F. The natural assumption that $H_t = e^{tU}$ always leads to wide and important cases of solutions of the equation (1′). It remains unknown whether a convenient method of forming a general solution, when the symbol U is understood in a sufficiently general sense, is possible.

In addition to stochastic processes without aftereffect, another class of such processes with aftereffect but *stationary* (where all the distributions persist under the change from t to $t' = t + a$) was also deeply studied. Khinchin proved a profound theorem for stationary processes generalizing the Birkhoff ergodic proposition. Hardly anything is known about non-stationary processes with aftereffect.

Introductory literature

1. Hostinsky, B. Méthodes générales du calcul des probabilités. *Mém. Sci. Math.*, t. 52, 1931, pp. 1 – 66.

2. Kolmogoroff, A. Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Math. Ann.*, Bd. 104, 1931, pp. 415 – 458.

Further bibliographic indications are in

3. Khintchine, A. *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung*. Berlin, 1933.

2b. A systematic study of random functions is just beginning. It is certain, however, that a deeper theory of stochastic processes will be essentially based on the notions connected with random functions. Owing to the incompleteness of the axiomatics of probability theory, many investigations (Wiener, Slutsky) were restricted to considering the values of functions at a finite number of points. In this case, even the formulation of the problem concerning the conditions of continuity, integrability, differentiability of a function, etc was impossible.

Instead, Slutsky introduced new notions of stochastic continuity, integrability and differentiability. However, given a sufficiently developed axiomatics, it will also be possible to raise and solve the problem of determining the probability of some differential or integral properties of a function in the usual sense. Appropriate research ought to be important for the solution of problems with boundary conditions for stochastic processes.

3a. Stochastic processes without aftereffect were especially studied in the particular case in which the state of the system is determined for each moment by one real parameter. For the sake of simplicity of writing we restrict our exposition to the case of time-homogeneity. Then

$$H_t(F) = \int K_t(x; y) dF(x),$$

and, for kernels $K_t(x; y)$, the equation (1′) will be written as

$$K_{s+t}(x; y) = \int K_t(z; y)\, d_z K_s(x; z). \quad (2)$$

This is indeed the Smoluchowski equation well known in physics. Two main types of its solutions are known. The first one corresponds to the case of *continuous* changes of the state $x(t)$ with time. Under some additional assumptions it can be proved here that the kernel $K_t(z; y)$ satisfies the Fokker – Planck partial differential equation

$$\frac{\partial K}{\partial t} = -\frac{\partial (AK)}{\partial y} + \frac{\partial^2 (BK)}{\partial y^2} \quad (3)$$

where A and B are some functions of y. The solutions of this type are the most widely applied (§3c).

Finetti (1929) was the first to examine the other type of solutions of equation (2). It corresponds to a discontinuous variation of $x(t)$ with time.

It is also possible to construct solutions corresponding to a mixed type of variations. Such solutions can be obtained, for example, by issuing from the integro-differential equation

$$\frac{\partial K(x;y)}{\partial t} = -\,U(y)K(x; y) - \frac{\partial\,[A(y)K(x;y)]}{\partial y} + \int U(y; z)K(x; z)dz + \frac{\partial^2 [B(y)K(x;y)]}{\partial y^2} \quad (4)$$

which I indicated in 1931. Hostinsky obtained solutions of the same type by other methods. By introducing Stieltjes integrals of a special type in equation (4) and thus generalizing it, it will perhaps be also possible to obtain the general solution of equation (2). Until now, this was achieved, owing to the works of Finetti, myself and Lévy, only in the particular case in which the kernels were restricted to the type $K(x; y) = K(y - x)$. A number of contributions was devoted to many-dimensional generalizations of the considered patterns.

3b. Only a part of the theory here supposed {to be constructed} was yet developed, – the theory of the second moments of random functions (Khinchin, Slutsky). The general definition of a characteristic function of a distribution law in any linear space (in particular, therefore, in any functional space) is this.

Let E be some linear space with elements x, and $P(A)$ – the probability that element x belongs to set A. We denote linear functionals of x by f. Then a characteristic function of distribution $P(A)$ is a function of the functional

$$\Phi(f) = \int_{E_x} e^{ifx} P(dE_x).$$

If $\Phi(f)$ can be expanded into a Taylor series

$$\Phi(f) = 1 + h_1(f_1) + h_2(f_1; f_2) + h_3(f_1; f_2; f_3) + \ldots,$$

the multilinear forms $h_n(f_1; f_2; \ldots; f_n)$ provide the moments of the distribution $P(A)$. Until now, only a very small number of corollaries were deduced from the indicated definitions. However, they {definitions or corollaries?} promise to be interesting, in particular for a number of physical problems.

3c. The Fokker – Planck differential equation and its conjugate underpin the derivation of a number of new limit theorems concerning sums of a large number of random variables. Several methods for passing from integral equations determining the laws of distribution of finite sums to the corresponding differential equations were offered. The most elegant, as it seems to me, is the Petrovsky method which is a modification of the Perron method of upper and lower functions. However, it was Bernstein who obtained the fullest results in some directions by applying another method for passing to the Fokker –Planck equations.

Another series of limit theorems can be connected with the equations of step-wise stochastic processes (§3a). Whereas the first series of these propositions generalizes the Laplace – Liapunov theorem, the second one does the same with respect to the Poisson limit theorem. The theorems of the second type were gotten by Khinchin and Bavli.

Introductory Literature

1. Khinchin, A. (1933), *Asymptotische Gesetze der Wahrscheinlichkeits-rechnung*. Berlin.

Note 1. {Fisher's *Statistical Methods for Research Workers* first appeared in 1925.}

4. A.Ya. Khinchin. The Theory of Probability in Pre-Revolutionary Russia and in the Soviet Union

Front Nauki i Tekhniki, No. 7, 1937, pp. 36 – 46

Foreword by Translator

Khinchin described the findings of the Moscow school of the theory of probability and argued that Soviet mathematics was far more advanced in these years (about 1937) than before 1917. And he attempted to show that this fact was due to the favorable atmosphere created in the Soviet Union for scientists. Khinchin's high standing is the only reason why such rubbish deserved to be translated. If proof is needed, see Kolman (1931) and Sheynin (1998). Kolman (a minor mathematician and a high-ranking *Parteigenosse* who much later escaped to the West) named Vernadsky, Sergei Vavilov, Ramsin et al and mathematicians Schmidt, Kagan and Egorov as the bad guys; Egorov, for example, was indeed declared a saboteur, exiled and died soon afterwards. The horrible situation existing then (and both earlier and later) in the Soviet Union is now widely known so that either Khinchin was forced to state nonsense or he had been completely blind. True, the *Luzin case* was suddenly abandoned in August 1936, but it is nevertheless remarkable that Khinchin had not condemned any *saboteur*. Then, his comparison of pre-1917 with the 1930s is not convincing also because, for example, British statisticians could have made similar conclusions in favor of the later period.

Khinchin made mistakes when describing the history of probability which once again proves that in those times hardly anyone knew it. He favored the Laplacean justification of the method of least squares at the expense of Gauss; he did not mention Bienaymé; did not explain Laplace's part in proving the De Moivre – Laplace limit theorem, etc. But the most disappointing error (that bears on one of his important conclusions) is his failure to notice that Chebyshev's main papers in probability had appeared in French in Europe; again, his *Oeuvres* were published in Petersburg, in 1899 and 1907, in Russian and French. A similar statement is true with regard to Markov and Liapunov. The second edition of Markov's treatise (1908) as well as three of his papers were translated into German in 1912. Liapunov published his memoirs in Petersburg, but in French, and both his preliminary notes appeared in the *C.r. Acad. Sci. Paris*. Moreover, Khinchin declares that Liapunov's works in mechanics also remained unknown, but this is what Liapunov wrote to Markov on 24.3.1901 (Archive, Russian Acad. Sci., *Fond* 173, Inventory 1, Item 11, p. 17), apparently in connection with his forthcoming election to full membership at the Petersburg Academy of Sciences:

> *You are asking me whether foreign scientists have referred to my works. If you need to know it, I ought to indicate* […]

He listed Poincaré, Picard, Appell, Tisserand, Levi-Cività and two lesser known figures and added that cannot say anything about Klein.

Still, I hesitate to deny Khinchin's overall conclusion that Western scientists had not been sufficiently acquainted with the work being done in Russia. An important related fact is that a similar inference about Russian scholars would have been wrong. Here is Liapunov's appropriate remark from his letter to Markov of 28.10.1895 (same Archive; Fond, Inventory and Item, p. 12):

Believing that it is very desirable that Chebyshev's contributions {*Oeuvres*} *be published as soon as possible, I am prepared to participate here*. […] *I ought to say, however, that, owing to my rather superficial knowledge of French, I am afraid to take up translations from Russian into French. As to the translations from French into Russian, I could undertake them although I do not quite sympathize with this business.* […] *I think that any* {Russian} *mathematician is able to read French.*

Days bygone!

[1] The theory of probability belongs to those very few branches of the mathematical science whose level of culture, even in pre-revolutionary Russia, was not lower than that in foreign countries. Even more can be stated: During the second half of the 19th century and the first years of this {the 20th} century, Russia was nearly the only country where the mathematical foundations of probability had been cultivated as earnestly as it deserved owing to its outstanding part in natural sciences, technology and social practice [1]. The Russian theory of probability completely owes its exceptional standing to the works of Chebyshev. Being in this respect considerably ahead of his time, this great scholar paved new ways to solving many problems dating back for a number of decades. Chebyshev also created in Russia the tradition, which, after being grasped by his followers, prompted many Russian scientists to devote their energy, at the turn of the 19th century, to the theory of probability; and which thus considerably fostered the development of this science.

The status of probability in Europe had then been unenviable. Already in the 18th century, the magnificent century of probability theory, Jakob Bernoulli and De Moivre discovered the two main laws of the doctrine of mass phenomena, the law of large numbers and the so-called limit theorem, for the simplest particular case, – for the Bernoulli trials. The Bernoulli theorem stated that the relative frequency of any event in a given series of homogeneous and mutually independent trials should, with an overwhelming probability, be close to that probability which the event had at each trial. The De Moivre theorem (which had until recently been attributed to Laplace) stated that, under the same conditions, the probabilities of the values of this frequency can fairly be approximated by a formula now called the normal distribution.

Already Laplace repeatedly stated his belief in that both these theorems were valid under much more general conditions; in particular, he thought it highly probable that under most general conditions a sum of a very large number of independent random variables should possess a distribution close to the normal law. He perceived here (and the further development of probability theory completely corroborated his viewpoint) the best way for mathematically justifying the theory of errors of observations and measurements. Introducing the hypotheses of *elementary errors*, that is, the assumption that the actual error is a sum of a large number of mutually independent and very small as compared with this sum *elementary* errors, we may, on the strength of the abovementioned principle, easily explain the generally known fact that the distribution of the errors of observation is in most cases close to the normal law. However, Laplace applied methods that did not allow to extend this principle beyond the narrow confines indicated by the De Moivre theorem and were absolutely inadequate for substantiating the theory of errors. Gauss is known to have chosen another way for attacking that goal, much less convincing in essence, but leading to it considerably easier [2].

In the interval between Laplace's classical treatise and the appearance, in the second half of the 19th century, of the works of Chebyshev, only one bright flash, Poisson's celebrated treatise, had illuminated the sky of probability theory. Poisson generalized the Bernoulli theorem to events possessing differing probabilities in different trials; he called this theorem the *law of large numbers* having thus been the first who put this term into scientific circulation. There also, Poisson offered his illustrious approximation for the probabilities of *seldom events*; and, finally, he made a new attempt at extending the De Moivre theorem beyond the boundaries of the Bernoulli trials. Like Laplace's efforts, his attempt proved unsuccessful.

And so, a twilight lasting all but a whole century fell over the European probability theory. Without exaggerating at all, it might be stated that, in those times, in spite of winning ever more regions of applied knowledge, European probability not only did not develop further as a mathematical science, – it literally degraded. The treatises written by Laplace and Poisson were on a higher scientific level than the overwhelming majority of those appearing during the second half of the 19th century. These latter reflect the period of decline when the encountered mathematical difficulties gradually compelled the minor scientists to follow the line of least resistance, to accept the theory of probability as a semi-empirical science only in a restricted measure demanding theoretical substantiation [3]. They usually inferred therefore that its theorems might be *proved not quire rigorously*; or, to put it bluntly, that wrong considerations might be substituted for proofs. And, if no theoretical justification could be found for some principle, it was declared an *empirically established* fact. This demobilization of theoretical thought, lasting even until now in some backward schools, has been to a considerable extent contributing to the compromising of the theory of probability as a mathematical science. Even today, after the theory attained enormous successes during the last decades, mathematicians more often than not somewhat distrust the rigor and irrevocability of its conclusions. During this very period there appeared Czuber's *celebrated* course which serves as a lively embodiment of the most ugly period in the life of the theory of probability and which was a model for many translations including Russian ones{?}.

[2] But then, in Russia, at the beginning of the second half of the 19th century, Chebyshev began destroying, one after another, the obstacles that for almost half a century had been arresting the development of the theory of probability. At first he discovered his majestically simple solution of the problem of extending the law of large numbers. The history of this problem is indeed remarkable and provides almost the only example in its way known in the entire evolution of the mathematical sciences. Until Chebyshev, the law was considered as a very complicated theorem; to prove those particular cases that Jakob Bernoulli and Poisson had established, transcendental and complicated methods of mathematical analysis were usually applied.

Chebyshev, however, proved his celebrated theorem, that extended the law of large numbers to any independent random variables with bounded variances, by the most elementary algebraic methods, – just like it could have been proved before the invention of the analysis of infinitesimals. And it certainly contained as its simplest particular cases the results of Bernoulli and Poisson. His proof is so simple that it can be explained during a lecture in 15 or 20 minutes. Its underlying idea is concentrated in the so-called Chebyshev lemma that allows to estimate the probabilities of large values of a random variable by means of its expectation [4] whose calculation or estimation is in most cases considerably simpler. Formally speaking, this lemma is so simple as to be trivial; however, neither Chebyshev's predecessors or contemporaries, nor his immediate followers were able to appreciate properly the extreme power and flexibility of its underlying idea. This power only fully manifested itself in the 20th century, and, moreover, mostly not in probability theory but in analysis and the theory of functions.

Chebyshev, however, did not restrict his attention to establishing the general form of the law of large numbers. Until the end of his life he continued to work also on the more difficult problem, on extending just as widely the De Moivre – Laplace limit theorem. He developed in detail for that purpose the remarkable *method of moments* which {Bienaymé and} he had created, and which still remains one of the most powerful tools of probability theory and is essentially important for other mathematical sciences. Chebyshev was unable to carry out his investigations to a complete proof of the general form of the limit theorem. However, he had correctly chosen the trail which he blazed to that goal, and his follower, Markov, completed Chebyshev's studies, although only at the beginning of this {the 20th} century, after the latter's death.

However, another follower of Chebyshev, Liapunov, published the first proof of the general form of the limit theorem somewhat earlier, in 1901. He based it on a completely different method, the method of Fourier transforms, not less powerful and nowadays developed into a vigorous theory of the so-called *characteristic functions* which are one of the most important tools of the modern probability theory. When applying the modern form of the theory of these functions, the Liapunov theorem is proved in a few lines, but at that time, when that theory was not yet developed, the proof was extremely cumbersome.

In a few years after Liapunov had published his results, Markov, as indicated above, showed that the limit theorem can be proved under the same conditions by the more elementary method created by {Bienaymé and} Chebyshev. And we ought to note that the later development of this issue confirmed that the conditions introduced by Liapunov and Markov for proving the limit theorem were very close to their natural boundaries: those discovered recently were only insignificantly wider.

Thus, whereas in Europe for more than half a century the problem posed by Laplace could not find worthy performers, and the theory of probability, lacking in refreshing scientific discoveries, certainly degraded into a semi-empirical science, only Russian mathematicians were maintaining the Chebyshev school's tradition of considering the theory as a serious mathematical discipline. Markov's course in the theory of probability, not dated even in our days, was then the only serious pertinent manual in the world whereas the contemporary European textbooks embodied either conglomerates of prescriptions pure and simple, unjustified theoretically (or, even worse, wrongly substantiated), or collections of separate problems, or even funny scientific stories.

Already before the Revolution, Russia was, as we see, rich in the most prominent creators in the field of the science of chance. However, the Russian mathematical science of that period, on the whole outdated and reactionary, did not induce European mathematicians to keep an eye on Russian periodicals. As a result, the achievements of Chebyshev and his nearest followers not only did not serve (for which they were fully qualified) as a banner for the revival of the theory of probability the world over; in most cases they simply remained absolutely unknown to scientists abroad. When, in 1919, the famous French mathematician Lévy discovered a proof of the Liapunov theorem, he was convinced, as he himself stated, that he was the first to justify it. Only later he was able to ascertain by chance that Liapunov had already proved this theorem in 1901 in all rigor (and, having ascertained this fact, he made it known to all the world). It is not amiss to note that the same fate befell not only the now celebrated Liapunov limit theorem; the international scientific world has only recently *discovered* his no less important investigations into various issues of mechanics. And Markov's works found themselves in much the same situation. Pre-revolutionary Russian mathematicians, for all their personal endowment and great achievements, were representatives of such a reactionary, in the scientific-managerial respect, academic routine, that already for this reason they had no possibility of influencing the development of the world science in any noticeable measure. And so it happened that the only country, that had for many decades actually been a worthy successor to the glorious deeds of Bernoulli, Laplace and Poisson, was, owing already to the reactionary nature of its political and academic regime, during all that time removed from any participation in the development of the international science of probability.

[3] The science of the Soviet period proved, above all, that it can perfectly well preserve and cultivate the best achievements of the old Russian science. At the same time, the situation and the atmosphere created for the Soviet scientists are such that their potential, their gifts and scientific-cultural skills can fittingly influence the development of the world science. The pre-revolutionary Russia and the atmosphere of the old academic regime neither wanted to, nor could create such a situation. A prominent work of a Soviet scientist cannot pass unnoticed as it happened with the contributions of Markov and Liapunov. On the one hand, our Academy of Sciences publishes the investigations of Soviet scientists in foreign languages and distributes the pertinent materials all over the world. On the other hand, and this is most important of all, the prestige of Soviet science is raised to such a level that neither do the writings published in Russian ever remain unnoticed. Foreign journals publish their abstracts, many scientists study Russian. Our science and its language may by right claim international importance.

Bernstein is a representative of the old academic science in our country, and a worthy successor to the deeds of Chebyshev, Markov and Liapunov. The fate of his researches is nevertheless incomparably happier than the mournful destiny of the works of his predecessors: they are known to the entire scientific world. Bernstein maintains personal contacts with a large number of foreign scientists and they hold him in great respect. He delivered a number of reports at international mathematical congresses and in Zurich he was charged with making the leading plenary report on the theory of probability. All this describes the forms of contacts and influences of which the pre-revolutionary mathematicians were completely deprived.

Bernstein's predecessors were almost exclusively examining sums of independent random variables thus continuing the traditions of the classics of probability theory [5]. However, practice poses problems that very often demand the study of series of random variables rather considerably depending one on another. In most cases this dependence is the stronger the nearer in the given series are the considered variables to each other; on the contrary, variables situated far apart occur to be independent or almost so (meteorological factors, chronologically ordered market prices). Bernstein was the first who successfully attempted to generalize the main principles of probability theory to these cases as well. His remarkable theorem on the law of large numbers extends its action to all series of dependent terms where the correlation (the measure of dependence) between the terms of the series unboundedly decreases with an infinite increase in the distance between them. His deep investigations devoted to the limit theorem showed that it also, under very wide assumptions, can be generalized to series of dependent random variables. He also was the first to formulate and solve the problem about many-dimensional generalizations of the limit theorem which is of a fundamental importance for mathematical physics (theory of diffusion, Brownian motion).

Along with effectively and fittingly continuing the most glorious work done before the Revolution, the Soviet period brought about many essentially new points concerning both the substance of the issues under development and the forms of scientific work. Here, the most significant example is the scientific school created within Moscow University, a collective of researchers whose like the probability theory in pre-revolutionary Russia could not have known. It possesses its own style, its scientific traditions, its rising generation; at the same time, it enjoys quite a deserved reputation as one of the leading and most advanced schools in the world. There is no important region of probability theory in whose development the Moscow school had not participated actively and influentially. It had initiated and attained the first achievements in a large number of modern issues whereas foreign scientists only joined in their investigations later.

[4] Let me attempt now to shed as much light as it is possible for an author of a paper not addressed to specialists, on the main achievements of the Moscow school. Until the very last years, the theory of probability only studied infinite patterns that were sequences of random variables. Most often we imagined such sequences as series of consecutive values of one and the same randomly changing (for example, with time) magnitude (consecutive readings on a thermometer; consecutive positions of a particle experiencing Brownian motion). However, for suchlike patterns a direct study of a variable *continuously* changing with time; an examination of a *continuous* interchange, of a continuous series of values where the change occurring between any two moments of time is subject to the action of chance, rather than of a sequence, would better conform to reality.

We thus arrive at the idea of a *stochastic* (a random) *process* where consecutive interchange is replaced by a continuous current conditioned by randomness. The same problems that occupied the theory of probability when studying sequences of random variables arise here, for these stochastic processes; and, in addition, a number of essentially new issues crop up. Along with direct theoretical interest, these stochastic processes are very important for a number of applied fields (mathematical physics). However, it was difficult to create mathematical tools which would enable to cover them. In 1912 the French scientist Bachelier had attempted to accomplish this [6], but he did not succeed, and only in 1930 Kolmogorov discovered a method based on the theory of differential equations which ensured an analytical formulation of the main problems arising in the theory of stochastic processes. From then onward, this theory has been actively developing and today it is one of the most urgent chapters of probability, rich in results and problems. Along with the Moscow school, a number of foreign mathematicians, with whom we have been in constant contact, were participating in the pertinent work.

During the last decades, the theory of probability encountered extremely diverse limiting processes. It is about time for a mature mathematical science {such as probability} to sort out this variety, to establish the connections and interrelations between these limiting formations. Only in 1936 the French scientist Fréchet systematically explicated this *topology* of random variables in a complete form that demanded a half of a lengthy treatise. However, as he repeatedly indicated, the founder of the theory, who had first developed it considerably, was the Moscow mathematician Slutsky. He also played a substantial part in working out the theory of stochastic processes (above); he highly successively studied the application of the main analytic operations (integration, differentiation, expansion into Fourier series, etc) on random variables and often connected his investigations with problems arising in applied natural sciences.

One of the objects initially studied by Markov but then pretty well forgotten was the series of random variables now usually called Markov chains. These are sequences of mutually dependent random variables connected by an especially simple dependence when the law of distribution of any of them is completely defined (in the most simple case) by the value of the immediately preceding variable so that the influence of the earlier *history* is eliminated. Markov rather considerably developed the study of such *chains* but after his death his results were forgotten whereas foreign scientists probably never heard of them at all. This happened partly because of the abovementioned causes, and partly because scientists in those days were yet unable to connect these theoretical investigations with current issues in natural sciences or practice. In 1928, when the need to study such chains became necessary for various reasons, and when their possibilities for applications were clearly outlined, they were *discovered* for the second time. Foreign scientists, believing that they were turning up virgin soil, proved a number of his findings anew. From then onwards, the theory of Markov chains became, and is remaining one of the most intensively developing chapters of the theory of probability. The Moscow school, mainly in the person of Kolmogorov, actively and very successively participated in this work. I ought to add right here, that one of our outstanding specialists, Romanovsky (Tashkent), who does not belong to the Moscow school, was and is also energetically taking part here, so that the total contribution of the Soviet science to the creation of this theory that originated in Russia seems very considerable.

Then, the Moscow school initiated and attained the main achievements in developing a sphere of issues directly adjoining the classical epoch of probability theory. It was known long ago that in the most important instances the deviation of the arithmetic mean of a long series of independent random variables from its expectation obey the normal law and that, in particular, the value of such a deviation is therefore, in a sense, and to a certain extent, bounded, but until recently the problem of determining the exact boundaries for these deviations did not arise even for the most elementary cases. In 1924, Khinchin first formulated and solved this problem for the Bernoulli trials; then, after he, somewhat later, had extended this solution (now known as the law of the iterated logarithm) to some more general cases, Kolmogorov showed that it persisted under considerably more general assumptions. In 1932 Khinchin extended this result to continuous stochastic processes. Then some foreign scientists (Lévy, Cantelli) refined the solution provided by the law of the iterated logarithm, but its most precise formulation, at least for random continuous processes, was again discovered in Moscow by Petrovsky by means of a remarkable method covering many various problems of the modern probability theory, see below.

If a magnitude changing under random influences is shown as a point moving along a straight line, on a plane, or in space, we will have a picture of a random motion, or *walk* of a point. Many most urgent issues of theoretical physics (problems of diffusion, Brownian motion and a number of others) are connected with this picture. Separate problems of this kind have been comparatively long ago solved by the theory of probability. Nevertheless, no satisfactory general method existed, and this led to the need for introducing very restrictive assumptions as well as for inventing a special method for each problem. In 1932 Petrovsky discovered a remarkable method connecting the most general problem of random walks with problems in the theory of differential equations and thus providing a possibility of a common approach to all such problems. At the same time, his method eliminates the need for almost all the restrictive assumptions so that the problems can be formulated in their natural generality.

Since the De Moivre limit theorem; the Liapunov theorem; and all their later extensions including those considering many-dimensional cases, are particular instances of the general problem of random walks, the Petrovsky method provides, in particular, a new, and, for that matter, a remarkable because of its generality proof of all these propositions. Moreover, the power of this analytic method is so great that in a number of cases (as, for example, in the abovementioned law of the iterated logarithm) it enabled to solve also such problems that did not yield to any other known method. The application of the Petrovsky method by other workers of the Moscow collective (Kolmogorov, Khinchin) had since already led to the solution of a large number of problems of both theoretical and directly applied nature. In an ad hoc monograph Khinchin showed that this method covered most various stochastic problems.

When speaking about the theoretical investigations of the Moscow school, it is also necessary to touch on its considerable part in the logical justification of the doctrine of probabilities. The construction of a robust logical foundation for the edifice of the theory became possible comparatively recently, after the main features of the building were sufficiently clearly outlined. And, although a large number of scientists from all quarters of the globe participated in constructing a modern, already clearly determined logical base for the theory, it is nevertheless necessary to note that it was Kolmogorov who first achieved this goal having done it completely and systematically.

[5] However, considering the development of a stochastic theory as its main aim, the Moscow school may nevertheless be reproached for not sufficiently or not altogether systematically studying the issues of practical statistics. The main problem next in turn, to which the Moscow collective is certainly equal, consists in revising and systematizing the statistical methods that still include many primitive and archaic elements. This subject is permanently included in the plans of the Moscow school and is just as invariably postponed. However, the collective has been accomplishing a number of separate studies of considerable worth in mathematical statistics. Here, statisticians have even created a special direction arousing ever more interest in the scientific world, viz, the systematic study of the connections and interrelations between the theoretical laws of distribution and their empirical realizations. If a long series of trials is made on a random variable obeying a given law of distribution, and a graph of the empirical distribution obtained is constructed, we shall see a line unboundedly approaching the graph of the given law as the number of trials increases. The investigation of the rapidity and other characteristics of this approach is a natural and important problem of mathematical statistics. The closeness, and, in general, the mutual location of these two lines can be described and estimated by most various systems of parameters; the limiting behavior of each of these parameters provides a special stochastic problem. Its solution often entails very considerable mathematical difficulties, and, as a rule, is obtained as some limit theorem. During the last years, Moscow mathematicians achieved a number of interesting results in this direction and Smirnov should be named here first and foremost. He had discovered a number of strikingly elegant and whole limiting relations and Glivenko and Kolmogorov followed suit. These works were met with lively response also by foreign scientists and are being continued. This year the Moscow collective began working {cf. below} on another urgent issue of modern mathematical statistics, on the so-called principle of maximum likelihood. If adequately developed, this promises to become a most powerful tool for testing hypotheses and thus to foster the solution of a most important problem of applied statistics all the previous approaches to which did not yet for various reasons satisfy the researchers. This work is, however, only planned.

[6] I have already mentioned a number of studies done by the Moscow school and connected with theoretical physics. For a long time now, a number of branches of biology (genetics, natural selection, struggle for existence) have been resting on stochastic methods. Even before the Revolution Bernstein developed a number of stochastic applications to genetics. Its mathematical requirements grown during the last years have demanded new, more subtle stochastic studies which the Moscow school did not take up. Glivenko, with Kolmogorov participating, developed a peculiar genetic algebra. They, as well as Petrovsky and Piskunov, investigated various issues in natural selection and struggle for existence.

In the sphere of technical applications of probability, the most complicated nowadays are the problems arising in connection with the running of systems designed for general use; and, from among these, the estimation of the {necessary} equipment of telephone exchanges and networks. In this direction the Moscow school had studied a number of problems both of general theoretic and directly applied nature. Khinchin constructed a general mathematical theory of stationary queues whose particular cases are both the telephony (see just above) and the estimation of the time passing between a machine tool, etc goes out of service and its repair. And he, together with Bavli, developed in the practical sense the urgent theory of shared telephone lines and made a number of calculations directly required by the Ministry of Communications. All this work was carried out while keeping in constant touch with practical specialists, engineers at the Ministry's research institute. Finally, a special commission of mathematicians and engineers headed by Slutsky aims at systematically developing statistical problems arising in technology. It is now working regularly, but I ought to remark that until now it is still restricting its efforts to gathering information and did not compile any plans for active operations.

This far from complete list of works on probability theory accomplished by the school of the Moscow University is sufficiently convincing and shows the range of the school's activities. An attentive collective discussion of all the works being carries out from their initiation onwards; regular ties with practical specialists and natural scientists with respect to all the applied issues; intimate contacts with all prominent scientists including foreigners concerning all the parallel and related studies; a speedy publication of results; and efforts directed at disseminating these among all the interested scientific circles, – none of these features of managing scientific work were, or could have been known to the pre-revolutionary theory of probability.

[7] The works of Bernstein and the Moscow school do not, however, exhaust the accomplishments of the Soviet theory of probability. The third prominent center of creative work in this field is Tashkent. The leader of mathematicians at Sredneaziatsky {Central Asian} University, Romanovsky, is a most oustanding world authority on mathematical statistics. Whereas Bernstein and his associates and the Moscow stochastic school mainly concentrated their efforts on the theory of probability, the entire scientific world of mathematical statistics is attentively following the work issuing from the Soviet Central Asia. It is rather difficult and unnecessary to draw a clear boundary line between the two abovementioned sciences, but the border is mainly determined by the fact that probability theory is mostly interested in theoretical regularities of mass phenomena whereas mathematical statistics creates practical methods for scientifically mastering these phenomena. It is self-evident that any antagonism between these two branches of the essentially indivisible science of mass phenomena is out of the question. On the contrary, they most indispensably supplement one another. Romanovsky is one of the most productive Soviet scientists and the remoteness of his city from the old scientific centers does not hinder his uninterrupted close ties with scientists the world over working in this sphere. It is difficult to name any considerable area in current mathematical statistics in whose development Romanovsky did not actively, and, moreover, weightily and authoritatively participate.

The Soviet theory of probability also includes separate workers in other scientific centers of the nation (Zhuravsky in Leningrad, Persidsky in Kazan, et al).

[8] In an overwhelming majority of the other branches of mathematics the Russian pre-revolutionary science (excepting its separate representatives who embodied those exceptions that confirm the rule) lagged considerably behind their European counterparts, lacked its own flavor and was even unable to follow the world science in a sufficiently civilized way. We are glad to observe a flourishing of nearly all of these branches, but skeptics will perhaps be apt to explain this away as a peculiar illusion: Since nothing was available before, and at least something is present now, it is easy to assume the something for very mush. In the theory of probability the matter is, however, different: here we had much already in the pre-revolutionary period and what we have now we cannot compare with a blank space. Nevertheless, here also, as we see, the Soviet science wins this comparison totally and undoubtedly. The scientific accomplishments of the Soviet period are incomparably wider and much more versatile, but we see the main and decisive progress in the management of science which certainly explains the successes of the Soviet theory of probability. The pre-revolutionary mathematics was unaware of scientific collectives working orderly and in concord; nowadays, we have them.

The pre-revolutionary mathematics stewed in its own juice, and, in spite of all its accomplishments, was barely able to influence the world science. The Soviet theory of probability rapidly secured one of the leading positions in the world science and achieved such an authority about which the pre-revolutionary science could not have even dreamt. Non-one can say that my statement is an exaggeration since it is completely based on facts. It is a fact that the most prominent scientists the world over publicly recognize the authority and the influence of the Soviet stochastic school. It is a fact that in a number of cases foreign publishers apply to Soviet authors when compilation of treatises and monographs on probability is required and that before the Revolution there were no such cases. It is a fact that Soviet scientists were charged with delivering the leading plenary reports on the theory of probability at the two latest international congresses of mathematicians (in 1932 and 1936) and that no such cases had happened at the congresses before the Revolution.

Thus, if even before the Revolution the Russian theory of probability, owing to its specific weight, might have by right claimed to be a leading force in world science, the Soviet theory of probability has not only totally confirmed this right and justified it even better. For us, it is no less important, however, that our branch of mathematics has exercised this right and continues to exercise it ever more persistently. The power necessary for this is certainly drawn exclusively from the inexhaustible source of cheerfulness contained in our new Socialist culture.

Notes

1. {It is possible that Khinchin did not dare to mention *sociology*, an exclusive domain of the Marxist dogmas.}

2. {The method of least squares is a peculiar field in that many leading mathematicians (for example, Chebyshev, Lévy, Fisher and Koomogorov) formulated unfounded or even wrong statements about it. here, Khinchin indirectly and wrongly declared that the method stood in need of the central limit theorem (in essence, of the ensuing normal distribution). It is certainly true that it is best of all, in a definite sense, to secure observations whose errors obey the normal law, but this is not necessary for justifying the method of least squares. It was Gauss, who, in 1823, provided the definitive substantiation of the method (perhaps *less convincing* for *pure* mathematicians).}

3. {Many publications of that period were aimed at popularizing the contributions of Laplace and Poisson which had been remaining scarcely understood.}

4. {Obviously, large deviations from the appropriate mean value.}

5. {Khinchin should have mentioned Markov right here, not only two pages later.}

6. {Bachelier's first publication appeared in 1900, cf. Kolmogorov's essay of 1959 (§4), also translated in this book.}

The Theory of 5. A.N. Kolmogorov. Probability and Its Applications

In *Математика и естествознание в СССР* (Mathematics and Natural Sciences in the Soviet Union). Moscow, 1938, pp. 51 – 61 …

[Introduction] The last two decades have been a period of rapid international growth and reconstruction of the theory of probability and of a further strengthening of its influence on the physical, biological and technical investigations. The heightened attention to the theory led to the establishment, for the first time ever, of a systematic international cooperation between specialists in this field. New ideas originating in one country therefore find a response in another one during the very next years, and sometimes even in a few months. Soviet mathematicians participated to a considerable extent in this lively and intensive work and in many directions they have formulated the main guiding ideas.

True, for the theory of probability the matter for the Soviet science consisted not in gaining, for the first time, an honorary place in the international scientific work, but rather in keeping for itself, in the situation of a sharply increased activity of foreign scientific schools, that first place which the works of Chebyshev, Markov and Liapunov firmly won for the Russian science. During that preceding period, when general theoretical research in probability somewhat fell into decay in Western Europe, the three Russian scholars deeply developed the classical heritage of Laplace, Poisson and Gauss [1].

Actually, their investigations created that complete system of the classical theory of probability which now constitutes the main substance of the textbooks. However, concerning the pre-revolutionary period, we may only speak about the leading position of the Russian school with respect to the world science with a serious reservation. Whereas Chebyshev's findings were appreciated abroad a long time ago, Liapunov's fundamental memoirs of 1900 – 1901 containing the proof of the main limit theorem of the theory of probability remained for many years hardly noticed and Mises (in 1909) [2] and Lindeberg (in 1923) discovered their results anew.

The fate of Markov's research devoted to the pattern of the course of random phenomena, now everywhere called the scheme of *Markov chains*, was still more sorrowful. He himself only dealt with his pattern theoretically, and, as an illustration, he considered the alternation of vowels and consonants in the text of {Pushkin's} *Евгений Онегин* (Eugene Onegin). Only about 1930 his results got widely known (and were partly rediscovered anew) and became the theoretical foundation for very general and important concepts of statistical physics. This example concerning *Markov chains* is also connected with another feature of the pre-revolutionary Russian school, with its being exclusively directed to the solution of classical problems and detached from the newly originating requirements to probability theory from other sciences.

Such new demands were formulated in the second half of the 19th century, first and foremost owing to the development of statistical methods in the social and biological sciences. The complex of theories that is usually understood and taught as mathematical statistics took shape on these very grounds. The theory of probability enters the sphere of issues mainly when determining whether a {given} restricted number of observations is sufficient for some deductions. Exactly here we may perceive a certain (naturally relative) specificity of the stochastic problems of mathematical statistics. In the 20th century, the supremacy of the British (Karl Pearson, Fisher) and partly American schools in mathematical statistics had definitely shown itself. An increasing number of tests for the compatibility of a given finite series of observations with some statistical hypothesis, which were offered on various special occasions, has led during the last years to an intensive search for general unifying principles of mathematical statistics considered from this very point of view of hypotheses testing (for example, in the works of Neyman and E.S. Pearson). Here, Soviet mathematicians accomplished a number of remarkable special investigations (§3), but, for us, the development and reappraisal of the general concepts of mathematical statistics, which are now going on abroad under a great influence of the idealistic philosophy, remains a matter for the future.

In statistical physics, the issue of the sufficiency of a restricted number of observations withdraws to the background. Here, the directly observed quantities are most often the results of a superposition of a great number of random phenomena (for example, on a molecular scale). The material for testing hypotheses on the course of separate random phenomena is thus collected independently of us in quite a sufficient measure. On the other hand, statistical physics makes especially great demands for developing new patterns, wider than the classical ones, of the course of random phenomena. The first decades of the 20th century had been especially peculiar in that the physicists themselves, not being satisfied with the possibilities provided by the classical theory of probability, began to create, on isolated particular occasions, new stochastic patterns. A number of biologists and technicians encountered the same necessity. For example, studies of the theory of diffusion had led Fokker and Planck to the construction of the apparatus of differential equations which were later discovered already as general differential equations of arbitrary continuous stochastic processes without aftereffect (§2). Fisher introduced the same equations quite independently when studying some biological issues. To such investigations, that anticipate the subsequent development of the general concepts of probability theory, belong also the works of Einstein and Smoluchowski on the theory of the Brownian motion, a number of studies accomplished by American technicians (T. Fry et al) on the issues of queuing connected in particular with problems of exploiting telephone networks, etc. Finally, the theory of probability encountered most serious problems in the direction of stochastically justifying the so-called ergodic hypothesis that underpins thermodynamics. Only around 1930 mathematicians earnestly undertook to systematize all this material. It is the most expedient to unite all the thus originated investigations under the name of general theory of stochastic processes.

After ascertaining the situation in which probability theory developed during the last twenty years, we go on to review, under three main heads, the pertinent accomplishments of Soviet scientists.

1) Extension of the classical investigation of limit theorems for sums of independent and weakly dependent random variables.

2) General theory of stochastic processes.

3) Issues in mathematical statistics.

Irrespective of this methodical division, we believe that considerable work on the theory of probability in the Soviet Union began roughly in 1924 – 1925. S.N. Bernstein's fundamental research (see below) appeared in 1925 – 1926, and it alone would have been sufficient for considering that the traditions of Chebyshev, Markov and Liapunov were worthily continued.

His *Теория вероятностей* (Theory of Probability), which already became classical, soon followed. At the same time, from 1923 -1925 onward, the works of the Moscow school (Khinchin, Kolmogorov) started developing. In the beginning, they were restricted to a rather narrow sphere of issues within the reach of the methods taken over from the theory of functions of a real variable [3].

1.The central issue of research done by Chebyshev, Markov and Lipaunov was the ascertaining of the conditions for the applicability of the *normal*, or Gaussian *law of distribution* to sums

$$s_n = x_1 + x_2 + \ldots + x_n \quad (1)$$

of a large number of independent or weakly dependent random terms x_i. Bernstein, in a fundamental memoir of 1926, completed the classical methods of studying this issue. He essentially widened these conditions for dependent terms and was the first who rigorously justified the application of the many-dimensional Gaussian law to sums of vectors in spaces of any number of dimensions. This last-mentioned result also theoretically substantiates the applicability of the formulas of normal correlation for the case in which correlated variables may be considered as sums of a very large number of terms with the connection between these variables being restricted by that between the corresponding, or those close to them, terms of these sums. Such, in particular, is the situation when quantitative indications, caused by an additive action of a very large number of genes, are inherited. Consequently, Bernstein was able to show that the Galton law of the inheritance of such indications was a corollary of the Mendelian laws (under the assumption that a large number of uncoupled genes act additively) and does not contradict them at all as it had been often stated.

Returning to the one-dimensional case, we may formulate the conditions for the applicability of the normal law to sums of independent terms in the following way: With probability close to 1, all the terms are much less than their sum s_n [4]. A natural question here is, What limit distributions can be obtained if we only require the same for each separate term x_i (the principle of *individual negligibility*)? G.M. Bavli and Khinchin have recently answered this question (as also did P. Lévy in France in a somewhat vague form). In the limit, we obtain the so-called *infinitely divisible* laws that include as particular cases the Gauss, the Poisson, and the Caushy laws. These undoubtedly deserve to be more systematically introduced into statistical practice as well.

In addition to this essential extension of the classical approach to the limit theorems concerning sums of random terms, we ought to indicate the following. In applications, we refer to the limit theorems when dealing with the distributions of finite sums. At the same time, however, the existing estimates are such that in many of the most important practical cases the guaranteed estimate of the remainder term exceeds the main term many times over [5]. Actually, quite a satisfactory estimate only exists for the simplest case of the Laplace – Bernstein theorem.

Another, even more venerable subject of classical investigations in the theory of probability, is the issue about the conditions for the applicability of the law of large numbers to sums of independent terms. In a number of studies, Khinchin, Kolmogorov and others also developed it further. In addition to obtaining necessary and sufficient conditions, new concepts of *strong* and *relative* stability of the sums were created and conditions for their application were formulated. Finally, Khinchin discovered an absolutely new remarkable asymptotic formula for the order of the maximal deviations of consecutive sums (1) from the mean, – the so-called law of the iterated logarithm, – for sequences $x_1, x_2, \ldots, x_n \ldots$ of independent terms x_i. Bernstein and Khinchin studied the conditions for the applicability of the law of large numbers to sums of dependent terms.

The addition of independent random terms also gives occasion for raising a number of questions about the *decomposability* of random variables into sums of such terms. Khinchin and a few of his students examined the problems of the *arithmetic of the laws of distribution* appearing here. The abovementioned far from exhausts the field of research adjoining the classical limit theorems but it is sufficient for appraising the significance of the findings in this area, which are now, owing to the latest works of Soviet and foreign authors [6], near to completion. From among the not yet solved problems, the one discussed above concerning practically effective estimates of remainder terms, is likely the most important.

2. When studying, from a general viewpoint, the process of random changes in an arbitrary physical system, it is natural to isolate, first of all, processes without aftereffect; that is, processes in which the law of distribution for the future states of the studied system is completely determined by its state at the present moment irrespective of its previous history. If the number of possible states is here finite, and if they are only being recorded at moments constituting a discrete sequence, we have a pattern examined by Markov already about 30 years ago. In this case everything is determined by the conditional probabilities $p_{ik}^{(n)}$ that the system at moment n being in state i will find itself in state k at moment $(n + 1)$. During the last years, the homogeneous case, in which these probabilities do not depend on k, was the object of extraordinarily numerous studies, and in the Soviet Union Romanovsky's investigations occupy the first place with respect to completeness. From the point of view of statistical physics, the main issue here is the limiting law of probability for the frequencies of the system finding itself in various states over long periods of time. Under very wide (and now definitively ascertained) conditions the limiting values of the frequencies do not depend on the initial state of the system and the deviations of the frequencies from these values obey a many-dimensional Gaussian law. Much less is known about non-homogeneous Markov chains (*i.e.*, about the case of variable $p_{ik}^{(n)}$).

The transition to the infinite discrete (countable) set of states is connected with rather considerable mathematical difficulties but does not yet essentially change the method of study. Kolmogorov was recently able to obtain general results concerning such kind of Markov chains with an infinite number of states. N.M. Krylov and N.N. Bogoliubov, in a recent note that continues and generalizes the investigations of Fréchet and his students, examine the case of an arbitrary (uncountable) set of states.

In accordance with the established methods of the classical theory of probability, the just outlined investigations of usual and generalized Markov chains reduce the study of a stochastic process to a consideration of a discrete sequence of *trials*. The transition to studying stochastic processes, in which random changes of the state are possible during any arbitrary short interval of time, and especially to examining continuous stochastic processes, has required a considerable reconstruction of the analytical tools of probability theory. As mentioned above, physicists and technicians had earlier studied separate cases of such patterns *with continuous time*; Bachelier (1900) and de Finetti (1929) considered some instances from a purely mathematical angle.

Kolmogorov (1931) made the first attempt to systematize from a sufficiently general side all the possibilities occurring here (restricting his efforts to processes without aftereffect). Generalizing Smoluchowski, Fokker and Planck, he established the main integral and differential equations governing such processes under various assumptions about the sets of possible states of the system and about its continuous or step-wise changes.

Especially many studies were further devoted to the case of continuous finite-dimensional manifolds of possible states and continuity of the very process of random changes. In this case, under some natural assumptions, the process is governed by parabolic partial differential equations. In these, the coefficients of the first derivatives are connected with the mean direction of the change of state at a given moment of time, and those of the second derivatives express the intensity of the random deviations from this direction. For a physical theory of oscillations allowing weak random perturbations it is essential to study the limiting relations as the coefficients of the second derivatives tend to zero. Andronov, Pontriagin et al obtained a number of pertinent findings and some of their conclusions were unexpected and extremely interesting for physicists.

For inertial continuous random motion the main parabolical differential equations degenerate and Kolmogorov considered this case in one of his notes. N.S. Piskunov studied the problem of the existence and uniqueness of the solution of the appropriate differential equations. From among the other applications of the differential equations of continuous stochastic processes we note the work of Kolmogorov and M.A. Leontovich on Brownian motion and the former's writing that continued Fisher's study of the theory of natural selection in vast populations.

In addition to {allowing} a direct examination of continuous stochastic processes, their guiding differential equations possess another, not less important for the theory of probability, property. Indeed, their solutions provide asymptotic formulas for the laws of distribution also in the case of discrete processes consisting of a very large number of very small changes. A large number of pertinent investigations (originated by Kolmogorov and widely developed by Bernstein, I.G. Petrovsky and Khinchin) had appeared. Owing to them, the Liapunov classical limit theorem is now interpreted as a particular instance of some unified general theory. It is also possible to consider the limit theorems obtained here as {providing} a method, irrespective of the theory of continuous stochastic processes, for justifying the use of the corresponding differential equations. However, the ideas created by this theory too often guide the actual course of investigations so that a complete alienation of the limit theorems from the continuous theory is not expedient.

Processes without aftereffect allowing both continuous and step-wise changes are studied less. The main theoretical problem is here the search for the general solution of the so-called Smoluchowski integral equation that guides all such processes. De Finetti, Kolmogorov and Lévy have definitively examined the one-dimensional homogeneous case. Their investigations led, among other results, to the abovementioned infinitely divisible laws. Under some particular assumptions the problem {?} was recently solved by reducing {it} to integro-differential equations. This method is also promising for the general case.

Beyond the province of stochastic processes without aftereffect only *stationary processes* are well studied. The main works are here due to Khinchin and E.E. Slutsky. The former proved the main ergodic theorem on the existence of means over time for any quantities depending on the state of a stationary system and having finite expectations [7] and constructed a spectral theory of stationary processes [8] whereas the latter specifically studied stationary processes with discrete spectra. The sphere of applications for the theory of stationary stochastic processes is not yet sufficiently determined, but it ought to be extraordinarily wide. Apparently the most complete understanding of the nature of continuous spectra in acoustics and optics is possible within the boundaries of this very theory. And the study of the spectra of the stationary processes, without deciding in advance whether they will be discrete or continuous, should perhaps largely replace the so-called detection of latent periodicities (also see §3) which interests, for example, meteorologists so much. In particular, Slutsky's investigations of stationary stochastic processes originated from this sphere of meteorological problems. Finally, we mention a number of studies made by L.B. Keller on the theory of turbulent motion where he widely applies the ideas of the general theory of stationary stochastic processes and arrives at some findings valuable also from a general angle [9].

In concluding the review of the various directions of the study of stochastic processes we indicate two special fields which originated owing to particular applications and which were not yet reflected in the previous essay [10]. This is in the first place the study of *queuing* connected with the servicing of telephone and telegraph networks, the maintenance of automatic lathes, etc. Here, the matter concerns stochastic processes in which chance enters as the random distribution of the number of calls in a telephone network or of the interruption of the work of a lathe owing to some breakdown; or as the form of the law of distribution for the duration of telephone calls; or of the time needed for repairing the lathes. Only under greatly simplifying assumptions such processes conform to the pattern of processes without aftereffect having a discrete set of possible states and are {then} easily included in the general theory. Without these simplifications such a subordination of the processes without aftereffect to the general theory is {also} possible but too complicated for direct application. In a number of works Khinchin developed special methods for examining the phenomena of queuing. He himself and N.V. Smirnov used these methods for solving the problems raised by the Central Institute of Communications and pertaining to automatic telephony. Gnedenko applied the same methods to problems of the textile industry.

Peculiar problems of spatial queuing originate when studying the crystallization of metals and metallic alloys. Kolmogorov solved some of them as formulated by the Institute of Steel. M.A. Leontovich developed a statistical theory of bimolecular reactions which is a somewhat specialized and complicated version of the theory of stochastic processes without aftereffect having a finite number of states. He issued from differential equations corresponding to the pattern of continuous time. Smirnov andV.I. Glivenko showed that a similar specialization and complication of classical Markov chains (with a discrete sequence of trials) should be an essential tool of the theory of heredity.

We note finally that the extraordinary extension of the sphere of stochastic research outlined above would have been exceptionally difficult without a reconstruction of the logical foundation of the theory of probability, – of its axiomatics and the system of its main notions. From among the various directions in justifying the theory only one is yet developed to an extent that enables it to provide a formally irreproachable system of the main notions covering all the ramifications of the theory caused by the various requirements of physics and technology. This is the direction developing the axiomatics of probability theory by issuing from the definition of probability as an additive function of sets given on an adequate system of sets of *elementary events* [11]. In particular, the modern theory of probability cannot be satisfied only by considering finite-dimensional laws of distribution. The study of continuous stochastic processes and of a number of other physical problems inevitably leads to the consideration of *random functions*, or, in other words, of laws of distríbution in functional spaces. Slutsky and Kolmogorov engaged in a systematic construction of the theory of these functions.

3. From among the general problems of mathematical statistics, Soviet mathematicians developed with maximal success, first, those concerned with the statistical detection of latent periodicities and the establishment of the formulas of forecasting; and, second, problems of the statistical determination of distribution functions.

The classical theory of periodograms enables to analyze a series of random variables if we assume that it is made up by superimposing several periodic oscillations and additional perturbations, *independent* from one trial to another one. The last-mentioned hypothesis is, however, usually arbitrary and Slutsky and Romanovsky deeply studied all the circumstances occurring when we assume dependences between the random perturbations. Slutsky, in his concrete contributions, accomplished at the Geophysical Institute, offered many simple and elegant methods for analyzing series with a supposed latent periodicity.

Meteorologists were very interested in establishing formulas of forecasts by statistical means (we especially mention the works of V.Yu. Weese {Vise?}. The same problem especially concerned the hydrometeorological service; incidentally, Slutsky worked on its instructions. A number of investigations was devoted to the composition of formulas for determining the prospects for the harvest by issuing from meteorological data (Obukhov).

Slutsky; Obukhov; and others have been developing the appropriate general mathematical problems. The main issue here is to avoid fictitious dependences artificially introduced in a usually very restricted material (30 – 100 observations) **12**. This method can undoubtedly be useful when being adequately careful, but even from a purely mathematical side far from all the appropriate circumstances encountered when applying it are sufficiently ascertained.

The statistical determination of distribution functions consists in the following. A distribution function $F(x)$ is unknown; n independent observations corresponding to it are made and an *empirical* distribution step-function $F_n(x)$ is computed. It is required to ascertain, to what extent may we form an opinion about the type of the function $F(x)$ by issuing from $F_n(x)$. Glivenko subordinated the very problem about $F_n(x)$ tending to $F(x)$ to a general law of large numbers in functional spaces which he {also} established. For statistics, it was, however, important to have as precise estimates of the deviations of $F_n(x)$ from $F(x)$ as possible. Kolmogorov had provided the first asymptotic formulas for the law of distribution of these deviations **13** whereas Smirnov deeply and thoroughly studied all the relations between these functions by far exceeding all the previous results. His findings have most various applications in statistical investigations.

We have indicated two directions of research where the results of Soviet mathematicians offered a considerable contribution to the general development of mathematical statistics. In connection with the permanently occurring practical problems very many more special investigations were also made. Among these we point out for example the works of A.M. Zhuravsky on the statistical determination of the composition of minerals; and B.V. Yastremsky's studies of the application of sampling. It ought to be said, however, that the assistance rendered by mathematicians to the applications of mathematical statistics and the theory of probability was until now of a somewhat casual and amateurish nature and was mostly directed either by personal links or special interests of the individual researchers in applied issues connected with the essence of their general theoretical work. The further development of applied studies will undoubtedly require the creation of adequate computational tools, the compilation and publication of tables, and the design of {special} devices **14**.The creation of a scientific institution which would be able to shoulder all these duties and to attend systematically to the requirements of an applied nature is a problem for the near future.

Notes

1. {Gauss should not have been mentioned here.}

2. {Apparently, 1919.}

3. Beyond this special sphere these methods subsequently proved themselves essential for a rigorous formal justification of the theory of probability including extensions that have been required by its further development.

4. Feller has recently precisely formulated the necessary and sufficient conditions corresponding to this idea somewhat vaguely expressed by me.

5. We are usually interested in low probabilities of the order of 1/1,000 or 1/10,000. For reliably estimating them allowing for the existing expressions of the remainder terms the number of observations ought to be essentially larger than 10^6 or, respectively, 10^8.

6. Among these, in addition to the abovementioned Lévy, Cramér and Feller, is Mises who achieved fundamental results. {Kolmogorov did not mention Cramér.}

7. He generalized Birkhoff's ergodic theorem for dynamic systems.

8. In accordance with its intention the spectral analysis of stationary stochastic processes adjoins Wiener's generalized harmonic analysis in the theory of functions.

9. Keller's research was going on independently of Khinchin's and Slutsky's work as well as of that of Wiener (above).

10. {Kolmogorov apparently referred to his earlier essay also translated in this book.}

11. Kolmogorov, A.N. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin, 1933. {Russian translation, 1936.}

12.{The Russian phrase is not sufficiently clear, and its translation is tentative.}

13. Earlier Mises only furnished an estimate of their mean value and variance.

14. We note for example that some of our establishments have spent years for compiling tables of the coefficients of correlation between a large number of studied quantities. If adequate devices are available, such work can be done almost at once.

6. A.N. Kolmogorov. The Role of the Russian Science in the Development of the Theory of Probability

Uchenye Zapiski Moskovsk. Gosudarstven. Univ., No. 91, 1947, pp. 53 – 64 …

1. The theory of probability occupies a peculiar position among other sciences. Random phenomena admitting an estimate of their probabilities occur in mechanics, physics and chemistry as well as in biology and social domains. Accordingly, probability theory has no special and exclusive field, it is applicable to any sphere of the real world. At the same time, the theory is not a part of pure mathematics since the notions of causality, randomness, probability cannot be considered as belonging to the latter. This combination of greater specificity, and greater richness in concepts taken from concrete reality as compared with pure mathematics, on the one hand, with complete generality and applicability to most various fields of real phenomena, on the other hand, imparts special attraction to probability theory, but at the same time engenders peculiar difficulties in mastering it broad-mindedly and creatively.

In a certain sense, the theory of probability can be converted into pure mathematics, and this is accomplished by its axiomatization. According to the axiomatic exposition, and issuing for example from the system developed in my book on the main concepts of the theory of probability, *events* are replaced by sets whose elements are *elementary events*, and *probability* simply becomes an additive non-negative function of these sets. Formally speaking, the theory of probability is converted into a pure mathematical discipline, and, more precisely, into a special part of the abstract theory of measures of sets and *metric* theory of functions. However, from the viewpoint of such a formal reduction of probability to measure theory, the former's main specific problems become extremely artificial and special; the ideological orientation of the entire development of probability theory is obscured, and, finally, the possibility of a specifically stochastic intuitive prediction of results is lost.

In a formal sense, mechanics can be similarly considered a part of pure mathematics (mainly of the theory of differential equations). Mechanicians, however, hotly oppose this. And we, specialists in probability theory, also believe ourselves to be representatives of a special science possessing its own specific style of thinking. Cultivating a total mathematical formal rigor, also possible in many branches of mechanics, we direct all our investigations, even including the most general and abstract research, by our wish to understand the laws of real random phenomena and the origin of rigorous causal dependence resulting from the joint operation of a large number of independent or weakly connected random factors; and inversely, by our desire to comprehend the emergence of one or another probability distribution resulting from the superposition of small random perturbations on a rigorous causal dependence, etc. Just as the mechanicians, who especially appreciate researchers both mastering the analytical mechanical tools and having a mechanician *common sense* and intuition, – just the same, we make some distinction between pure analysts engaged in isolated problems posed by probability theory, and specialists in the theory proper, who, issuing from visual *stochastic* considerations, often perceive the solution of problems from the very beginning, before finding the appropriate analytic tools.

2. The history of probability theory may be tentatively separated into four portions of time. *The first period*, when the elements of our science were created, is connected with the names of Pascal (1623 – 1662), Fermat (1601 – 1665) and especially Jakob Bernoulli (1654 – 1705). *The second one* lasted throughout the 18th, and the beginning of the 19th century: De Moivre (1667 – 1754), Laplace (1749 – 1827), Gauss (1777 – 1855) and Poisson (1781 – 1840). *The third period, i.e.*, the second half of the 19th century, is largely connected with the names of Russian scientists, Chebyshev (1821 – 1894), Markov (1856 – 1922) and Liapunov (1857 – 1918). In Western Europe, general theoretical research in probability theory during this time remained somewhat in the background. With regard to its theoretical stochastic methods, the emerging mathematical statistics (Quetelet, Cournot, Galton, K. Pearson, Bruns, Bortkiewicz) mainly managed with the results of the previous period, whereas the new requirements made by statistical physics were not yet sufficiently expressed in general contributions on the theory of probability. In Russia, meanwhile, almost exclusively by the efforts of the three abovementioned celebrated mathematicians, the entire system of the theory was reconstructed, broadened and essentially deepened. Their work formed a solid basis for the development of probability theory during the *fourth period*, the beginning of the 20th century. This was the time of a general strengthening of interest in the theory as manifested in all countries, and of an extraordinary broadening of its field of application in various special branches of natural sciences, technology and social sciences. Although the Soviet {school of} probability does not possess such an exclusive place in this intensive international scientific work as the one that fell to the lot of the classical Russian research of the previous period, it seems to me that its rank is also very significant, and that, with regard to the general problems of the probability theory itself, it even occupies the first place.

3. Russian scientists did not participate in the work of the first period, when the main elementary concepts of our science, the elementary propositions such as the addition and the multiplication theorems, and the elementary arithmetical and combinatorial methods were established. The concrete material studied mostly amounted to problems in games of chance (dice, playing cards, etc). Paradoxically, however, this was mainly a *philosophical* period in the development of the theory of probability.

It was the time when mathematical natural science was created. The goal of the epoch was to comprehend the unusual broadness and flexibility (and, as it appeared then, omnipotence) of the mathematical method of studying causal ties. The idea of a differential equation as a law uniquely determining the forthcoming evolution of a system, given its present state, occupied an even more exclusive place in the mathematical natural science than it does nowadays.

For this branch of knowledge, the theory of probability is required when the deterministic pattern of differential equations is not effective anymore; at the same time, the concrete natural-scientific material for applying the theory in a calculating, or, so to say, business-like way, was yet lacking. Nevertheless, the inevitability of coarsening real phenomena when fitting them in with deterministic patterns of the type provided by systems of differential equations, was already sufficiently understood. It was also clear that quite discernible regularities may occur *in the mean* out of the chaos of an enormous number of phenomena defying individual account and unconnected one with another. Exactly here the fundamental role of probability theory in theoretical philosophy was foreseen. Of course, just this aspect rather than the servicing of the *applied* problems posed by Chevalier de Méré, so strongly attracted Pascal to probability, and (already explicitly) guided Jakob Bernoulli during the twenty years when he was searching for a proof of his limit theorem that also nowadays is the basis of all applications of probability theory. This proposition solved with sufficient completeness the main problem of theoretical philosophy encountered in the theory's first period [1] and remained, until the appearance of De Moivre's work, the only limit theorem of the theory of probability.

4. In the next, the second period according to my reckoning, separate fields had already appeared where quantitative probability-theoretic calculations were required. These fields were not yet numerous. The main spheres of application were the theory of errors and problems in the theory of artillery firing. The chief results obtained in the former theory were connected with Gauss, and the achievements in the latter subject, with Poisson [2]. Neither field was, however, alien for Laplace who was the main figure of that time. Here are the main pertinent theoretical results.

1) The De Moivre – Laplace limit theorem. It asymptotically estimates the probability

$$P_n(t) = P(\mu \le np + t\sqrt{np(1-p)}$$

that, in n independent trials, each having probability p of a *positive outcome*, the number of such outcomes μ will not exceed $np + t\sqrt{np(1-p)}$. The theorem states that, as $n \to \infty$, $P_n(t)$ tends to

$$P(t) = (1/\sqrt{2\pi}\,)\int_{-\infty}^{t} \exp(-x^2/2)dx.$$

From then onwards, the probability distribution $P(t)$, appearing here for the first time, is playing a large part in the entire further theory of probability and is {now} called *normal.*

2) The Poisson generalization of this theorem to the case of variable probabilities $p_1, p_2, \ldots, p_n$.

3) The substantiation of the method of arithmetic mean {of least squares} by Gauss.

4) The development of the method of characteristic functions by Laplace.

Thus, not only from the ideological and philosophical side, but in the regular everyday scientific work, the main attention was transferred from the elementary theorems about a finite number of events to limit theorems. Accordingly, non-elementary analytic methods were dominating.

Note that the maturity of the contemporary Russian science revealed itself in that Lobachevsky's probability-theoretic work, in spite of his remote peripheral scientific interest in the theory of probability, was quite on a level with international science and approvingly quoted by Gauss [3]. Ostrogradsky also left several works in probability, but the dominant influence of Russian science on the entire development of probability theory begins later.

5. The third period in the development of the theory of probability, *i.e.*, the second half of the 19th century, is especially interesting for us. In those times, a rapid development of mathematical statistics and statistical physics occurred in Western Europe. However, it took place on a rather primitive and dated theoretical basis with Petersburg becoming the center of studies in the main general problems of probability. The activity of academician Buniakovsky, who, in 1846, published an excellent for his time treatise, *Основания математической теории вероятностей* (Principles of the Mathematical Theory of Probability), and widely cultivated applications of probability to insurance, statistics, and, especially, demography, and paved the way for the flourishing of the Petersburg school of probability theory.

It was Pafnuty Lvovich Chebyshev, however, who brought the Russian theory of probability to the first place in the world. From a methodological aspect, the principal upheaval accomplished by him consisted not only in that he was the first to demand, with categorical insistence, absolute rigor in proving limit theorems [4]. The main point is that in each instance Chebyshev strove to determine exact estimates of the deviations from limit regularities taking place even in large but finite numbers of trials in the form of inequalities unconditionally true for any number of these.

Furthermore, Chebyshev was the first to clearly appreciate and use the full power of the concepts of *random variable* and its *expectation* (mean value) [5]. These notions were known earlier and are derivatives of the main concepts, *event* and *probability*. However, they are subordinated to a much more convenient and flexible algorithm. This is true to such an extent that we now invariably replace the examination of event A by considering its characteristic random variable ξ_A equal to unity when A occurs, and to zero otherwise. The probability $P(A)$ of event A is then nothing but the expectation $E\xi_A$ of ξ_A. Only much later the appropriate method of characteristic functions of sets came to be systematically used in the theory of functions of a real variable.

The celebrated {Bienaymé –} Chebyshev inequality

$$P(|\xi| \geq k) \leq E\xi/k$$

is also quite in the spirit of the later theory of functions. Nowadays such a method of estimating appears to us quite natural and goes without saying. In Chebyshev's time, however, when the similar way of thinking was alien to analysis or the theory of functions (the concept of measure did not exist!), this simple method was absolutely new.

Having given his main attention to the concept of random variable, Chebyshev was naturally led to consider limit theorems on the number of positive outcomes in a series of trials as subordinated to more general propositions on the sums of random variables [6]. The celebrated Chebyshev theorem appeared as a natural generalization of the Bernoulli proposition: If random variables $x_1, x_2, \ldots, x_n, \ldots$ are independent one from another, and bounded by the same constant, $|x_n| \leq N$, then, for any $\varepsilon > 0$, a limit relation

$$P(|(s_n/n) - E(s_n/n)| > \varepsilon) \to 0 \text{ as } n \to \infty$$

exists for the arithmetic means $(s_n/n) = (x_1 + x_2 + \ldots + x_n)/n$. Markov extremely widened the conditions of this limit relation.

The ascertaining of a proposition similar to the Laplace theorem for sums of random variables proved much more difficult. At the same time, this problem could not have failed to attract attention. Without solving it, the special role of the normal distribution in the theory of errors, in artillery and other technical and natural-scientific fields could not have been considered sufficiently cleared up. The problem was to ascertain under sufficiently wide conditions the limit relation

$$P[s_n \leq Es_n + t\sqrt{\text{var}(s_n)}\,] \to (1/\sqrt{2\pi}\,)\int_{-\infty}^{t} \exp(-x^2/2)dx \text{ as } n \to \infty$$

where var $(s_n) = E[s_n - Es_n)]^2$ is the so-called variance of the sum s_n, equal, as it is well known, to the sum of the variances of the terms x_i:

$$\text{var}(s_n) = \text{var}(x_1) + \text{var}(x_2) + \ldots + \text{var}(x_n).$$

Here, it was impossible to manage without some sufficiently complicated analytical tool. Chebyshev chose the method of moments, *i.e.*, the study of the quantitative characteristics of a random variable x of the type $m_k = E(x^k)$. He was unable to carry through the proof of the limit theorem by means of this method. The final success fell to Markov's lot, but the choice of the analytical tool rather than the essence of the problem led him to demand the existence of finite moments of any order for the random variables under his consideration.

Liapunov offered a formulation of this limit theorem free from the restriction just mentioned. His methods, and the generality of his result created such a great impression, that even to our day the proposition in his wording is often called the *main*, or the *central* limit theorem of probability theory [7].

From among the other directions of research followed by the Petersburg school the so-called *Markov chains* should be especially cited. This peculiar term screens one of the most general and fruitful patterns of natural processes. One of the main concepts for the entire modern natural science is the notion of *phase* space of the possible *states* of a studied system. A change of an isolated system is supposed to be deterministic and free from the so-called *aftereffect* if, during time interval τ, the system certainly transfers from state α to state $\beta = F(\alpha; \tau)$ where F is some definite single-valued function of the initial state and the interval of time. For random processes *without aftereffect*, given α and τ, we only have, instead of the function $F(\alpha; \tau)$, a definite probability distribution depending on α and τ for the state β to replace the state α after τ units of time.

Markov considered the simplest case of such processes in which the phase space only consisted of a finite number of states $\alpha_1, \alpha_2, \ldots, \alpha_n$. In addition, he restricted his attention to considering processes, as we say, *with discrete time, i.e.*, only observed after $\tau = 1, 2, 3, \ldots$ Under these conditions, the studied probability distributions are given by transition probabilities $p_{ij}^{(\tau)}$ for the time interval τ from state α_i to state α_j with the recurring formulas

$$p_{ij}^{(\tau+1)} = \sum_k p_{ik}^{(\tau)} p_{kj}^{(1)}$$

allowing the calculation of $p_{ik}^{(\tau)}$ for any $\tau = 1, 2, 3, \ldots$ given the matrix (p_{ij}) of the transition probabilities $p_{ij} = p_{ij}^{(1)}$ for an elementary *unit* time interval.

This simple pattern nevertheless allows to study all the main general properties of processes *without aftereffect*. In particular, Markov ascertained the first rigorously proved *ergodic theorem*: if all p_{ij} are positive, then $p_{ij}(\tau) \to p_j$ as $\tau \to \infty$ where p_j do not depend on i.

I have not in the least exhausted even the main achievements of the Petersburg school. I have mainly dwelt on its essentially new ideas, and sometimes explicated them in a somewhat modernized form so as to show more clearly their influence on the further development of the theory of probability, and their importance for mathematical natural science. It would have been more difficult to offer, in a popular article, a notion of the technical skill, elegance and wit of the school's exceptionally eminent analytical methods.

Only with considerable delay, in the 1920s or even the 1930s, the importance of the works of Chebyshev, Markov and Liapunov was quite appreciated in Western Europe. Nowadays they are everywhere perceived as the point of departure for the entire further development of the theory of probability. In particular, the main Liapunov limit theorem [8] and the theory of Markov chains were exactly what was most of all needed for a reliable substantiation of the developing statistical physics. That the West had slowly adopted the ideas of the Petersburg school may perhaps be indeed partly explained by the fact that the school was very remote from statistical physics, so that Markov only illustrated the application of his theory of *trials connected into a chain* (the application of Markov chains) by considering the distribution of vowels and consonants in the text of {Pushkin's} *Евгений Онегин* (Eugene Onegin) [9].

Hopefully, my last remark will not lead to an impression that the works of the Petersburg school lacked an animated feeling of connection with the requirements of mathematical natural science. A keen sense of reality in formulating mathematical problems was especially characteristic of Chebyshev. Issuing from comparatively special elementary, and sometimes rather old-fashioned applied problems, he elicited from them with exceptional insight such general mathematical concepts that potentially embraced an immeasurably wider circle of technical and natural-scientific problems.

6. The fourth period of the development of probability theory begins in Russia with the works of Sergei Natanovich Bernstein. With regard to their scope, only the writings of the German mathematician now living in the USA, Richard Mises, can be compared with them. They both posed the problems of

1) A rigorous logical substantiation of the theory of probability.

2) The completion of research into limit theorems of the type of Laplace and Liapunov propositions leading to the normal law of distribution.

3) The use of modern methods of investigation possessing full logical and mathematical value for covering, to the greatest possible extent, the new domains of application of probability theory.

In this last direction, the activity of Mises, who headed a well organized Institute of Applied Mathematics {in pre-Nazi Germany}, was perhaps even wider than Bernstein's research. The latter, however, offered many specimens of using stochastic methods in most various problems of physics, biology and statistics. And in the second, purely mathematical direction, Bernstein accomplished his investigations on a considerably higher methodological and technical level. He extended the conditions for applying the main limit theorem for independent random variables to such a degree of generality that proved to be essentially final. To him also belong the unsurpassed in generality conditions for applying the main limit theorem to dependent variables as well as the first rigorously proved bivariate limit theorem [10].

Finally, with respect to the logical substantiation of the theory of probability, Bernstein is the author of its first systematically developed axiomatics based on the concept of a qualitative comparison of events according to their higher or lower probabilities. The numerical representation itself of probability appears here as a derivative concept. The American mathematician Koopman has comparatively recently rigorously formalized a development of such a notion.

The work of the Moscow school of probability theory began in 1924 (Khinchin, Kolmogorov, Slutsky, with Glivenko, Smirnov, Gnedenko and others joining them later). Essential investigations belonging to the range of ideas of this school were also due to Petrovsky, who, with regard to his style, nevertheless remained a pure analyst. In its main part (Slutsky began his research independently), the Moscow school was founded by N.N. Luzin's pupils (Khinchin and Kolmogorov) who issued from transferring the methods of the metric theory of functions of a real variable to a new field. Exactly these methods have determined their success in the first two directions of the work done in Moscow:

1) The determination of the necessary and sufficient conditions for the applicability of the law of large numbers to sums of independent terms; the discovery of extremely general conditions for the applicability of the so-called *strong law of large numbers* to the same sums [11]; the necessary and sufficient conditions for the convergence of a series of independent terms; Khinchin's so-called *law of the iterated logarithm.*

2) The creation of an axiomatics of the theory of probability, very simple with regard to its formal structure and embracing its entire applications, both classical and most modern ones.

3) As stated above, limit theorems of the Liapunov type demand more special analytic tools. In this direction, the Moscow school applied the method of *characteristic functions*

$$\varphi_\xi(t) = \mathrm{E}e^{it\xi}.$$

Here, as a result of Khinchin's and Gnedenko's investigations, it was completely ascertained to which laws of distribution can the sums of independent terms tend if the size of each term, with probability approaching unity, becomes arbitrarily small as compared with the sums. The necessary and sufficient conditions under which such convergence takes place were also discovered. It turned out, that, in addition to the normal law appropriate for the classical limit theorems, all the other *stable laws*, found by the French mathematician Lévy, can also appear, whereas the entire class of *admissible* laws coincides with the so-called *infinitely divisible laws* whose study was begun by the Italian de Finetti.

A large part of the further work done by the Moscow school was connected with the concept of *random processes* (for the time being, in its classical, non-quantum understanding). Two large fields were here studied:

4) The theory of processes without aftereffect. Being a direct generalization of Markov chains, they are therefore called *Markov processes*. For them, the probabilities of transition $F(x; E; s; t)$ from state x at moment s to one of the states belonging to set E at moment t are connected by the so-called equation of Smoluchowski or Chapman [12].

5) The theory of stationary random processes with their spectral theory.

Kolmogorov originated the first of these directions. He discovered that, under some wide conditions and given transition probabilities, the non-linear Smoluchowski integral equation invariably leads to some linear partial differential equation called after Fokker and Planck [13]. Still wider are the conditions under which a variable Markov process depending on a parameter asymptotically approaches an ideal Markov process obeying the Fokker – Planck equations. In such relations we perceive now the common root of all the limit theorems of the Laplacean and Liapunov type. Only from this point of view the fact that the classical function of the normal density of probability

$$\varphi(x; D) = (1/\sqrt{2\pi D})\exp(-x^2/2D)$$

is the solution of the equation of heat conduction

$$\partial\varphi/\partial D = 2\partial^2\varphi/\partial x^2$$

ceased to appear accidental.

Mathematicians of the Moscow school (including, in particular, Petrovsky) and Bernstein are studying this new vast field, termed theory of *stochastic differential equations*, that opened up here. Most works of the Moscow school assume that the considered states of the system are represented by points of some compact part of space whereas Bernstein examined with special attention those new facts that appear when this restriction is abandoned. His generalization is all the more natural, since the special case of the classical limit theorems leading to the normal probability distribution should indeed be considered on the entire number axis, *i.e.*, on a non-compact set [14].

I am unable to dwell as minutely on the spectral and ergodic theories of stationary random processes created (as a general mathematical theory) by Khinchin. This direction of research occupies a prominent place in the work of other representatives of the Moscow school as well and many foreign mathematicians cultivated it. The remarkable investigations on statistical periodography independently started by Slutsky have also joined Khinchin's direction of work. In the field of statistics, the importance of all this research is widely recognized abroad. One of Slutsky's main contributions was reprinted {translated into English} in England on the initiative of the English statisticians. Wold's book on stationary time series published in Sweden was entirely based on the works of Khinchin, etc. For some reason the appreciation of the importance of the stochastic, statistical concept of oscillations with a continuous spectrum for physics and mechanics, as insisted on by Wiener in a somewhat different form even before the appearance of the Moscow works, is established to a lesser degree. Here, the contributions of the Moscow school sometimes become known only tardily. For example, Taylor, the celebrated specialist in the statistical theory of turbulence, published in 1938 a work on the connection between the distribution of energy over the spectrum and the coefficient of correlation for various distances, a work that contained nothing except for a particular case of Khinchin's formulas published in 1934. As to Soviet mechanicians, they only came to know about these relations from Taylor's work.

Such cases are, however, becoming atypical. With regard to those mathematical circles proper which are engaged in the theory of probability, the situation attained during the last 15 years before the war {before 1941 – 1945} was such that Soviet works began enjoying considerable response abroad almost immediately after publication. For our part, we also painstakingly follow everything going on in other countries. Studies in probability theory are very intensive everywhere, and it is often difficult to isolate the achievements made by scientists of separate nations. For example, the Italian de Finetti had originated the theory of the so-called infinitely divisible laws of distribution, I have vastly widened it, and Khinchin and the Frenchman Lévy developed it in persistent competition. And Gnedenko and Doeblin, a young Austrian mathematician who emigrated to France, accomplished, with variable success as to greater breadth and finality of the results obtained, an entire cycle of investigations connected with the application of these laws to the limit theorems of the classical type.

In wartime, intensive work abroad in the theory of probability was going on almost exclusively in the USA [15], where not only American, but likely the best European scientific specialists who fled from Germany, Italy and France, were concentrated. When examining the then arriving American scientific periodicals, it was possible to see how intensively and successfully they developed, in particular, the directions that originated here. For example, we have first perceived the importance of studying *random functions*, and Slutsky and Kolmogorov made the first relevant steps, but nowadays the most exhaustive works on this subject belong to Americans. To retain their place in this competition after the war, Soviet specialists in probability theory will undoubtedly have to work very intensively.

7. The modern period in the development of mathematical statistics began with the fundamental works of English statisticians (K. Pearson, Student, Fisher) that appeared in the 1910s, 1920s and 1930s. Only in the contributions of the English school did the application of probability theory to statistics ceased to be a collection of separate isolated problems and became a general theory of statistical testing of stochastic hypotheses (*i.e.*, of hypotheses about laws of distribution) and of statistical estimation of parameters of these laws [16].

The first popularizer of this wide current in the Soviet Union was Romanovsky (Tashkent) who is also the author of important investigations in pure probability theory, – in Markov chains and other topics. In addition to his own interesting results achieved in the direction of the English school, Romanovsky published a vast course in mathematical statistics where he collected with an exceptional completeness the new findings of this discipline most essential for applications.

The Moscow school only introduced into mathematical statistics one new chapter naturally following from its theoretical investigations. With the exception of one isolated work due to Mises, statisticians always assumed, when determining an unknown law of distribution by empirical data, that it belonged to some family depending on a finite number of parameters and reduced their problem to estimating these. Glivenko, Kolmogorov, and especially Smirnov systematically developed *direct methods* of solving this problem, and of testing the applicability of a certain law to a given series of observations. These methods are very simple and are gradually becoming customary.

Indirectly, the Moscow contributions have also played an essential role in developing mathematical statistics in another sense. The investigations made by Fisher, the founder of the modern English mathematical statistics, were not irreproachable from the standpoint of logic. The ensuing vagueness in his concepts was so considerable that their just criticism led many scientists (in the Soviet Union, Bernstein) to deny entirely the very direction of his research [17]. The Polish mathematician Neyman working in the USA solved with a greatest completeness the problem of impeccably substantiating the theory of statistical hypotheses testing and estimating parameters. His constructions, adopted in most of the new American publications, are based on the Kolmogorov system of explicating the theory of probability. I note in this connection that the fear of introducing the simplest notions of modern mathematics into applied treatises is gradually got rid of in the USA. It is now already possible to begin the exposition of the theory of probability in lectures on statistics for American agronomists with the concepts of *field of sets* and *additive function of a set* determined on such a field. These notions are in essence extremely simple and their introduction into an elementary course, if only there is no idiosyncrasy to the word *set,* makes the exposition considerably more transparent.

Notes

1. {Kolmogorov might have thought about Jakob Bernoulli's explanation of his aim which was to find out whether or not induction could provide results as reliable as deduction does.}

2. {In the 19th century, the (elements) of the theory of artillery firing had not yet made any serious demands on probability theory. And it is strange that Kolmogorov had not mentioned population statistics.}

3. {On this point see Gnedenko (1949). Lobachevsky derived the law of distribution of a finite sum of mutually independent uniformly distributed variables. He was unaware of the previous work of Simpson, Lagrange and Laplace. That Lobachevsky had not noticed Laplace's derivation is apparently explained by the entangled structure of the latter's classic. In addition, I do not understand Lobachevsky's aims: it is thought that he strove to check his geometry by astronomical observations which was then absolutely impossible. And I doubt that Gauss quoted Lobachevsky in this connection. See Sheynin (1973, p. 301).}

4. The derivations provided by De Moivre, Laplace and Poisson were not at all irreproachable from the formal-logical point of view, although Jakob Bernoulli proved his limit theorem with an exhaustive arithmetical rigor.

5. {Yes, Chebyshev introduced the method of moments for proving the limit theorem and used the inequality first offered by Bienaymé. However, it may be argued that the entire development of probability theory was connected with an ever fuller use of the concept of random variable (Sheynin 1994, pp. 337 – 338).}

6. If A, B, …, F are random events, the sum

$$\mu = \xi_A + \xi_B + \ldots + \xi_F$$

of their characteristic random variables ξ_A, ξ_B, … , ξ_F is a random variable equal to the number of those events that actually occur.

7. {It was Pólya (1920) who introduced this term.}

8. Markov extended this limit theorem to many cases of dependent variables, and, in addition, formulated its bivariate analogue, without, however, proving it.

9. {Markov was hardly prepared to go beyond mathematics, see his letters to Chuprov of 23 Nov. and 7 Dec. 1910 (Ondar 1977, pp. 38 and 52).}

10. The case of any larger number of dimensions does not present any new difficulties of principle.

11. The first results in this topic belonged to the French mathematician Borel.

12. After the physicists who considered important particular cases of the general problem.

13. See Note 12.

14. Kolmogorov studied this particular case in his first work by means of new methods. I think that another distinction strongly stressed by Bernstein is less essential. For the Moscow school, both the discrete pattern of gradually increasing sums of separate terms, and the limit scheme of a random variable continuously changing with a continuous change of a parameter (of *time*), are stochastic patterns of full value. According to Bernstein's concept, stochastic terminology is only used for pre-limit discrete schemes. He proved that, as the size of separate terms diminishes, the laws of distribution of their sums tend to laws subordinated to the Fokker – Planck equations, but he did not connect the idea of a continuous series of random variables depending on a parameter (of a *random function*) with these limit laws.

It is true that the limit patterns introduced by the Moscow school possess some paradoxical properties (infinite velocities and non-differentiability of the random functions considered), but it should be borne in mind that

1) These properties, paradoxical in their limit expression, represent, although sketchily, the quite real characteristics of many physical processes. Take for example particles up to those, sometimes extremely small sizes for which their inertia is felt {begins to be felt}, undergoing Brownian motion considered as a function of time. Their coordinates quite really behave as non-differentiable functions obeying the Lipschitz condition of the order $\alpha < 1/2$, but not satisfying it for $\alpha = 1/2$ in any point.

It can also be noted on this special occasion that the Brownian motion with the allowance for the forces of inertia can be studied, as long ago indicated by Kolmogorov, by means of more complicated patterns with degenerate Fokker – Planck equations. More essential than this last remark, however, is that

2) The Moscow school introduced continuous patterns of *random* processes, characterized by their compliance with physical reality only under the restriction of an adequate scale. This, however, is a general property of all the mathematical continuous schemes of natural phenomena. Once their right to exist is denied, it would be natural to declare war also on all the methods of continuum mechanics which admit that density or the components of velocity are continuous differentiable functions of the coordinates. Indeed, these assumptions become senseless on the atomic scale. Finally,

3) From the point of view of substantiating the limit theorems of the classical type on discrete sums of random variables, their connection with new patterns of random variables continuously depending on a parameter does not introduce any arbitrary assumptions of a vague nature into the proof of the theorems. After an axiomatic construction of the mathematical theory of probability is successfully accomplished, no stochastic considerations differ with regard to rigor in any way from deliberations in pure mathematics: from the logical point of view, all the relevant stochastic terms are nothing but names of quite definite and purely mathematical objects. It is therefore not at all more pernicious to turn to the theory of continuous random processes when proving classical limit theorems, than, for example, to apply the theory of characteristic functions or parabolic partial differential equations.

15. Interesting work is also being done in Sweden, from where it became recently possible to obtain a few latest publications of the abovementioned Wold.

16. {Kolmogorov did not mention the so-called Continental direction of statistics (Lexis, Bohlmann, Bortkiewicz, Chuprov, Markov).}

17. {This statement contradicts Bernstein's qualification remark (1941, p. 386) to the effect that he does not reject Fisher's work altogether.}

References

1. Bernstein, S.N. (1941), On the Fisherian "confidence" probabilities. *Собрание сочинений* (Coll. Works), vol. 4. N.p., 1964, pp. 386 – 393. Translated in *Probability and Statistics. Russian Papers of the Soviet Period*. Berlin, 2005.

2. Gnedenko, B.V. (1949), On Lobachevsky's work in the theory of probability. *Istoriko-Matematich. Issledovania*, vol. 2, pp. 129 – 136. (R)
3. Ondar, Kh.O., Editor (1977, in Russian), *The Correspondence between A.A. Markov and A.A. Chuprov*. New York, 1981.
4. Pólia, G. (1920), Über den zentralen Grenzwertsatz der Wahrschein-lichkeitsrechnung und das Momentproblem. *Math. Z.*, Bd. 8, pp. 171 – 181.
5. Sheynin, O. (1973), Finite random sums. *Arch. Hist. Ex. Sci.*, vol. 9, pp. 275 – 305.
6. --- (1990, in Russian), *Chuprov*. Göttingen, 1996.
7. --- (1994), Chebyshev's lectures on the theory of probability. *Arch. Hist. Ex. Sci.*, vol. 46, pp. 321 – 340.

7. B.V. Gnedenko, A.N. Kolmogorov. Theory of probability

Uchenye Zapiski Moskovsk. Gosudarstven. Univ., No. 91, 1947, pp. 53 – 64

[Introduction]

The history of probability theory may be tentatively separated into four portions of time. […]{The authors repeat here, almost word for word, Kolmogorov's previous account of 1947 also translated in this book.} The fourth period of the development of the theory begins in Russia with the works of Bernstein. […]

The activities of the Moscow probability-theoretic school began somewhat later. It is natural to consider that it originated with Khinchin's works on the law of the iterated logarithm (in 1924) and Slutsky's paper of 1925 on stochastic asymptotes and limits. The direction created by Bernstein along with the works of the Moscow school are still determining the development of the Soviet probability theory. A number of new directions have, however, appeared; the theory is successfully cultivated in an ever increasing number of mathematical centers (Moscow, Leningrad, Tashkent, Kiev, Kharkov, etc), and the works of the various directions become woven together ever more tightly [1].

Since the theory of probability has numerous diverse applications, scientific work often transforms here into solving separate and very special problems, sometimes demanding masterly mathematical technique but introducing little innovation into the development of its general dominating ideas. Following the Chebyshev traditions, Soviet specialists always attempted to isolate the main probability-theoretic patterns, deserving deep and exhaustive study out of this mess of separate applied problems. The limit theorems for sums of independent terms; Markov chains; general Markov processes; random functions and random vector fields having distributions invariant with respect to some transformation group, – all these classical or newer general subjects studied by Soviet mathematicians had originated as a result of thoroughly reasoning out the reduction of a large number of separate problems from most diverse fields of natural sciences and technology to typical theoretical patterns.

A simple formal classification is not at all sufficient for discovering these main theoretical patterns. Often only hard work on isolated problems makes it possible to reveal the fruitful general concept that enables to approach all of them by a single method. It is natural therefore that, at each stage of the development of science, which gradually appears out of a multitude of particular problems put forward for study from most various quarters, only a part of them is taken over by some established section of the general theory, whereas the solution of a large number of problems is left to the devices of isolated *amateurish* methods. Such isolated problems should not at all be considered with contempt, especially if their applied importance is great. However, their solution could have only been included in a general review of the achievements of Soviet mathematicians during thirty years by listing the titles of the appropriate contributions. We have therefore preferred to focus all our attention on a small number of main directions, each of these being united by a sufficiently clear dominating idea.

Smirnov elucidates the application of probability to mathematical statistics in a companion article {also translated below}. Applications to statistical physics would have been worthy of a special paper since the appropriate problems are specific; we only treat some of them.

1.Sums of Independent Terms

Chebyshev's and Liapunov's main research was almost entirely concentrated on studying the behavior of sums of a large number of such independent random variables that the influence of each of them on their sum was negligible. More special investigations connected with sequences of independent trials, which constituted the chief object of attention for Jakob Bernoulli, Laplace and Poisson, are reducible to studying such sums. This problem is of main importance for the probability-theoretic substantiation of the statistical methods of research (the sampling theory) and of the {Laplacean} theory of errors; the interest in it is therefore quite well founded.

In addition, as it usually happens in the history of mathematics, this problem, that occupied most considerable efforts of the scholars of the highest caliber belonging to the preceding generation, became important as a touch-stone for verifying the power of new methods of research. When only the laws of distribution of separate sums are involved, the method of characteristic functions proved to be the most powerful. It gradually swallowed up the classical Russian method of moments and superseded the *direct* methods which the new Moscow school had borrowed from the theory of functions of a real variable. Elementary *direct* methods of the Moscow school are still providing the most for problems in which an estimation of the probabilities of events depending on many sums is needed. In future, these methods will possibly be replaced by the method of stochastic differential or integro-differential equations.

1.1. The Law of Large Numbers. A sequence of random variables

$$\zeta_1, \zeta_2, \ldots, \zeta_n, \ldots$$

is called *stable* if there exists such a sequence of constants $C_1, C_2, \ldots, C_n, \ldots$ that, for any $\varepsilon > 0$,

$$\lim P(|\zeta_n - C_n| > \varepsilon) = 0 \text{ as } n \to \infty. \quad (1.1.1)$$

Practically this means that, as n increases, the dependence of the variables ζ_n on *randomness* becomes negligible. If ζ_n have finite expectations $E\zeta_n = A_n$, it will be the most natural to choose these as the constants C_n. We shall therefore say that the sequence of ζ_n with finite expectations A_n is *normally stable* if, for any $\varepsilon > 0$,

$$\lim P(|\zeta_n - A_n| > \varepsilon) = 0 \text{ as } n \to \infty. \quad (1.1.2)$$

When stating that the sequence ζ_n *obeys the law of large numbers*, we mean that it is stable. Classical contributions always had to do with normal stability but in many cases it is more logical and easier to consider stability in its general sense. If the variables ζ_n have finite variances $B_n = \text{var}\,\zeta_n = E(\zeta_n - E\zeta_n)^2$ the relation

$$\lim B_n = 0 \text{ as } n \to \infty \quad (1.1.3)$$

will be sufficient, although not necessary, for normal stability. This condition was the basis of Chebyshev's and Markov's classical theorems.

A remark made in 1918 by Bernstein [2] may be considered as a point of departure for the work of Soviet mathematicians. He noted that the equality

$$\lim \mathrm{E}\frac{(\zeta_n - C_n)^2}{1+(\zeta_n - C_n)^2} = 0 \text{ as } n \to \infty \quad (1.1.4)$$

can serve as a necessary and sufficient condition of stability for given constants C_n. In 1925, Slutsky [5] provided similar but more developed ideas.

If the sequence of ζ_n is stable, their medians $\mathrm{m}\zeta_n$ can always be chosen as the constants C_n. Therefore, owing to Bernstein's finding, the necessary and sufficient condition for stability of the sequence of ζ_n can be written down as

$$\lim \mathrm{E}\frac{(\zeta_n - \mathrm{m}\zeta_n)^2}{1+(\zeta_n - \mathrm{m}\zeta_n)^2} = 0 \text{ as } n \to \infty \quad (1.1.5).$$

It is easy to apply the classical condition (1.1.3) to sums of independent terms

$$\zeta_n = \xi_1^{(n)} + \xi_2^{(n)} + \ldots + \xi_n^{(n)} \quad (1.1.6)$$

having finite variances $\mathrm{var}\xi_k^{(n)} = b_k^{(n)}$; indeed, the variances of the sums ζ_n are the sums of the variances of the appropriate terms:

$$B_n = b_1^{(n)} + b_2^{(n)} + \ldots + b_n^{(n)}.$$

Unlike the variances B_n, the expectations included in conditions (1.1.4) and (1.1.5) cannot be expressed in any easy manner through magnitudes describing the separate terms $\xi_k^{(n)}$. In 1928 Kolmogorov [2; 5] discovered a necessary and sufficient condition of stability for sums of independent terms which was easily expressed through the properties of the separate terms. It can be written down as (Gnedenko[24])

$$\lim \sum_{k=1}^{n} \mathrm{E}\frac{(\xi_k^{(n)} - \mathrm{m}\xi_k^{(n)})^2}{1+(\xi_k^{(n)} - \mathrm{m}\xi_k^{(n)})^2} = 0 \text{ as } n \to \infty. \quad (1.1.7)$$

A necessary and sufficient condition of normal stability for the sums of independent terms is somewhat more complicated (Gnedenko [24]), but, once (1.1.7) is established, its derivation does not present any great difficulties. Of special interest is the case

$$\xi_k^{(n)} = \xi_k/n, \ \zeta_n = (\xi_1 + \xi_2 + \ldots + \xi_n)/n,$$

where $\xi_1, \xi_2, \ldots, \xi_n \ldots$ is a sequence of identically distributed independent variables. Here, the necessary and sufficient condition of stability becomes extremely simple:

$$\lim nP(|\xi| > n) = 0 \text{ as } n \to \infty. \quad (1.1.8)$$

This condition {the authors have not specified ξ} is satisfied if such variables ξ_k have finite expectations, and the stability is here certainly normal. This fact constitutes the essence of

Khinchin's theorem [17]: *The arithmetic means ζ_n of independent and identically distributed variables ξ_k having finite expectations are always normally stable.*

The results listed above elucidate sufficiently fully and definitively the conditions for the applicability of the law of large numbers to sums of independent variables. The authors cited above had made use of rather simple mathematical tools which differ, however, in essence from the classical method of moments. This is unavoidable. Khinchin [5] showed that, even if the terms have finite moments of all the orders, there exists no necessary and sufficient condition of stability expressed through them. However, if the problem concerns not the applicability of the law of large numbers on principle, but rather a sufficiently precise estimation of the probabilities $P(|\zeta_n - A_n| > \varepsilon)$, then the transition to higher moments is quite natural. The main findings in this direction belong to Bernstein [41, pt. 3, chapter 2].

1.2. Attraction to the Gauss Law. Keeping to the notation of §1.1, we shall say that a sequence of random variables ζ_n is attracted to the *Gauss law*, if, after appropriately choosing the constants C_n and $H_n > 0$, as $n \to \infty$,

$$\lim P([(\zeta_n - C_n)/H_n] < t) = (1/\sqrt{2\pi}) \int_{-\infty}^{t} \exp(-t^2/2)dt \quad (1.2.1)$$

for any real t. According to the main classical case, the variables ζ_n have finite expectations A_n and variances B_n, and (1.2.1) takes place for $C_n = A_n$ and $H_n = \sqrt{B_n}$. We shall say here that the variables are *normally attracted to the Gauss law.*

The derivation of extremely general sufficient conditions for the normal attraction of the sums (1.1.6) of an increasing number of independent terms to the Gauss law is an immortal merit of Chebyshev, Markov and Liapunov. Their investigations were developed by Bernstein [13; 40]. With respect to the problem now concerning us, he offered conditions essentially equivalent to those which later occurred to be, in a sense explicated below, necessary and sufficient.

The search for unrestricted necessary conditions for attraction to the Gauss law can apparently only lead to barely interesting results formulated in a rather difficult way. This is because the very idea of limiting laws for sums of an increasing number of terms is only natural when at the same time the separate influence of each of these terms decreases. This demand can be precisely expressed by stating that, in addition to (1.2.1)

$$\lim[\sup_k P(|\xi_k^{(n)} - \mathrm{m}\xi_k^{(n)}| > \varepsilon H_n)] = 0 \text{ as } n \to \infty \quad (1.2.2)$$

should be satisfied for any $\varepsilon > 0$. This is the so-called demand of *limiting negligibility* of the separate terms. Note that, if the laws of their distributions are identical for a given sum, it is an inevitable conclusion from relation (1.2.1). The ascertaining of the necessary and sufficient conditions for the attraction of sums of independent terms to the Gauss law under the additional requirement (1.2.2) was the result of the investigations made by Khinchin, Lévy and Feller; see their reviews by Khinchin [42] and Gnedenko [24].We adduce the formulation of one of Khinchin's theorems that reveals the essence of the matter with an especial transparency: *If condition* (1.2.2) *is fulfilled, and, as* $n \to \infty$,

$$\lim P([(\zeta_n - C_n)/H_n] < t) = F(t)$$

where $F(t)$ *is a non-singular* [2] *distribution function, then the validity of the condition*

$$\lim \sum_{k=1}^{n} P(|\xi_k^{(n)} - \mathrm{m}\xi_k^{(n)}| > \varepsilon H_n) = 0 \text{ as } n \to \infty \quad (1.2.3)$$

for any $\varepsilon > 0$ *is necessary and sufficient for*

$$F(t) = (1/\sigma\sqrt{2\pi}) \int_{-\infty}^{t} \exp[-(t-a)^2/2\sigma^2]\, dt.$$

Feller was the first to publish the necessary and sufficient conditions themselves in an explicit form for attraction to the Gauss law under restriction (1.2.2). They can be expressed in the following way: *For the existence of constants* C_n *and* $H_n > 0$ *such that conditions* (1.2.1) *and* (1.2.2) *are satisfied, it is necessary and sufficient that such constants* K_n *exist for which*

$$\lim\inf P(|\zeta_n - \mathrm{m}\zeta_n| > K_n) > 0 \text{ as } n \to \infty, \quad (1.2.4a)$$

$$\lim \sum_{k=1}^{n} P(|\xi_k^{(n)} - \mathrm{m}\xi_k^{(n)}| > K_n) = 0 \text{ as } n \to \infty. \quad (1.2.4b)$$

The former relation concerns properties of the sums rather than that of separate terms and it can be written down in an equivalent form

$$\lim\inf K_n^2 \sum_{k=1}^{n} \frac{(\xi_k^{(n)} - \mathrm{m}\xi_k^{(n)})^2}{K_n^2(\xi_k^{(n)} - \mathrm{m}\xi_k^{(n)})^2} > 0 \text{ as } n \to \infty, \quad (1.2.4'a)$$

free, as (1.2.4b) also is, from the indicated shortcomings but somewhat less obvious.

Bernstein [13; 40] originated a similar investigation of the conditions for the attraction of sums of independent vectors to n-dimensional Gauss laws. In the Soviet Union, Khinchin, Romanovsky and Gnedenko followed up this direction; the last-mentioned author offered the most polished formulation [25].

1.3. Specifying the Classical Limit Theorem. Even now, normal attraction

$$\lim F_n(t) = \Phi(t) \text{ as } n \to \infty,$$

$$F_n(t) = P[(\zeta_n - A_n)/\sqrt{B_n} < t],\ \Phi(t) = (1/\sqrt{2\pi}) \int_{-\infty}^{t} \exp[-(t^2/2)\, dt,$$

remains the most important case of attraction to the Gauss law. The desire to estimate as precisely as possible the difference $F_n(t) - \Phi(t)$ is natural, both theoretically and practically. If this difference is not sufficiently small, it is also natural to add to $\Phi(t)$ some correction terms expressed simply enough through the distributions of the terms $\xi_k^{(n)}$ so that $F_n(t)$ will then be estimated sufficiently precisely.

In 1911, Bernstein discovered the most effective method of precisely estimating $F_n(t)$ for the particular instance of the Laplace limit theorem; later he [39] somewhat strengthened his findings. The foundation of most of the subsequent investigations of the general case is Chebyshev's method. It approximately represents $F_n(t)$ as a series of the type of

$$\Phi(t) + C_3^{(n)}\Phi^{(3)}(t) + \ldots + C_s^{(n)}\Phi^{(s)}(t) + \ldots$$

where $\Phi^{(s)}$ are consecutive derivatives of the Gauss function Φ and the coefficients $C_s^{(n)}$ are expressed through the moments of the terms $\xi_k^{(n)}$. The Swedish mathematician Cramér most fully developed this idea of Chebyshev.

Any domain of mathematics having to do with determining successful approximate expressions or with improving estimates, becomes more theoretically interesting when such formulations of its problems are discovered that allow us to speak about best approximations and best estimates. In the field under our consideration, such a stage of research is only beginning. In the foreign literature, remarkable results of this type adjoining Cramér's findings are due to Esseen (1945). He had studied consecutive sums

$$\zeta_n = \xi_1 + \xi_2 + \ldots + \xi_n \quad (1.3.1)$$

of identically distributed terms $\xi_1, \xi_2, \ldots, \xi_n$ and Linnik [1] published a deep investigation of the general case of differing {pertinent} laws of distribution.

1.4. Limiting Laws of Distribution for Sums of Identically Distributed Terms. The findings of Khinchin, Lévy and Feller on the conditions of attraction to the Gauss law (above) were obtained while studying a more general, although not less natural problem which is the subject of this, and of the next subsections. It consists in discovering all the laws of distribution for sums of an increasing number of independent terms negligible as compared with their sum.

Let us begin with the simplest case of consecutive sums (1.3.1) of identically distributed terms assuming that there exist constants C_n and $H_n > 0$ such that, as $n \to \infty$,

$$\lim P([(\zeta_n - C_n)/H_n] < t) = F(t) \quad (1.4.1)$$

where $F(t)$ is a non-singular law of distribution. Khinchin [39], also see Khinchin & Lévy [1], discovered the necessary and sufficient conditions which $F(t)$ must obey to appear here as a limiting distribution. It occurred that the logarithm of the characteristic function

$$\varphi(\lambda) = \int_{-\infty}^{\infty} e^{i\lambda t} dF(t)$$

must be expressed as

$$\log \varphi(\lambda) = i\gamma\lambda - \mu|\lambda|^{\alpha}[1 + i\beta(\lambda/|\lambda|)\, \omega(\lambda; \alpha)]$$

where α, β, γ and μ are real constants, $0 < \alpha \le 2$, $|\beta| \le 1$, $\mu > 0$, γ is arbitrary and

$$\omega(\lambda; \alpha) = \text{tg}[(\pi/2)\alpha] \text{ if } \alpha \ne 1 \text{ and } = (2/\pi)\log|\lambda| \text{ otherwise.}$$

These are the so-called *stable* laws. If $\beta = 0$ they are symmetric; Lévy considered such laws before Khinchin did. A linear transformation of t can lead to $\gamma = 0$ and $\mu = 1$. On the contrary, essentially differing stable laws of distribution $F(t)$ correspond to different values of the parameters α and β. The case of $\alpha = 2$ is the Gauss law.

Gnedenko [13] ascertained quite transparent necessary and sufficient conditions for the attraction of the sums now discussed to each of the stable laws. It is generally thought that the relevant limit theorems are only academic since they relate to sums of random variables with infinite variances (finite variances lead to the Gauss law). In spite of its prevalence, this opinion is not quite understandable because sums of independent terms with infinite variances and even infinite expectations appear naturally indeed, for example in such a favorite classical issue as the *gambler's ruin* when considering a series of games each of them being continued until reaching a given loss. Kolmogorov & Sevastianov [1] indicated more topical applications.

Let us go over now to a more general pattern of sums (1.1.6) where the terms of one and the same sum are identically distributed but can possess different distributions for different sums. Khinchin [41] proved that here the limiting laws $F(t)$ that can appear in relation (1.4.1) are *infinitely divisible*; that is, they can be expressed as laws of distribution of a sum of any number of independent and identically distributed terms. Such laws of distribution appeared in connection with studying random processes with continuous time (Finetti). Kolmogorov [14; 15] discovered the general expression for infinitely divisible laws with a finite variance: the logarithms of their characteristic functions are written down as

$$\log \varphi(\lambda) = i\gamma\lambda + \mu \int_{-\infty}^{\infty} S(\lambda; x)\, dg(x),$$

$$S(\lambda; x) = [(e^{i\lambda x} - 1 - i\lambda x)/x^2] \text{ if } x \neq 0 \text{ and } = -\lambda^2/2 \text{ otherwise.}$$

Here, γ and μ are constants and $g(x)$ is a supplementary distribution function. The Gauss law emerges if $g(x)$ is reduced to the unity function $E(x) = 0$ at $x \leq 0$ and $= 1$ otherwise. Later Lévy discovered the general formula for the $\log \varphi(\lambda)$ of an infinitely divisible distribution without demanding the finiteness of the variance.

1.5. Limit Theorems for Sums of Terms of Limiting Negligibility. In 1936 Bavli [2] ascertained that in case of finite variances the limiting law of distribution for sums of arbitrarily (not necessarily identically) distributed independent terms of limiting negligibility is always infinitely divisible. A year later Khinchin [41] proved this proposition abandoning the demand of finiteness of the variances. This general proposition includes the finding (§1.4) relating to the sums of terms identically distributed within each sum since here the *limiting negligibility* of the terms is inevitable.

Gnedenko [10] exhaustively studied the conditions for attraction to infinitely divisible laws. The essence of his method consists in that a quite definite *accompanying* infinitely divisible law of distribution is constructed for each sum of independent terms. The limiting negligibility of the terms leads to the true law of probability coming closer to the accompanying law. The problem about the existence of limiting laws is therefore reduced to a simpler problem concerning the existence of limiting expressions for the accompanying laws. From the viewpoint of applications, all this concept would have deserved to be studied anew with regard to effectively estimating the closeness of the laws of distribution of the sums to some infinitely divisible law and discovering methods of determining the best (even if only in the asymptotic sense) approximation of the true law of distribution of the sums by an infinitely divisible law. Such work did not even start yet.

1.6. New Problems about the Limiting Behavior of the Sums. The conditions for the applicability of the law of large numbers; the discovery of the possible types of limiting laws of distribution for sums of a large number of terms; and of the conditions for attraction to each of them, – all these problems inevitably appeared out of the desire for carrying the works of Chebyshev, Markov and Liapunov to its logical conclusion.

At the same time, much attention is recently being paid to new problems on the limiting behavior of sums of a large number of independent terms. Those new problems that only involve the laws of distribution of the sums are not too peculiar and their solution demands not more than a small modification of the methods considered above. Such, for example, is the problem concerning the conditions for the *relative stability* of sums of positive terms (Khinchin; Bobrov [1]; Gnedenko & Raikov [3]).

Much more peculiar are the problems connected with estimation of the probabilities of events depending on the values taken by several consecutive sums [1.3.1] of one and the same series of their independent terms. It was Borel who first drew the attention of mathematicians to these problems which we shall consider in the sequel. Borel formulated the question on the conditions for the applicability of *the strong law of large numbers*, *i.e.*, of the relation

$$P\,[\lim(|\zeta_n - \mathrm{E}\zeta_n|/n) = 0] = 1 \text{ as } n \to \infty. \quad (1.6.1)$$

He also posed the problem of determining the asymptotic estimate of the range of oscillations of the deviations $|\zeta_n - \mathrm{E}\zeta_n|$ which led Khinchin to his celebrated law of the iterated logarithm (below).

The methods of solving such problems can be separated into two groups. The first group of *elementary* methods is based on applying inequalities similar to the {Bienaymé –} Chebyshev inequality but enabling to estimate the behavior of some finite sequence of sums rather than of one sum. In essence, already Khinchin made use of such methods in his first work [2] on the law of the iterated logarithm. However, it was an absolutely elementary inequality which Kolmogorov [2] established in 1929 that was later applied most of all. In our notation, it can be written down as

$$P(\sup|\zeta_n - \mathrm{E}\zeta_n| \geq a) \leq \mathrm{var}\,\zeta_n/a^2,\ 1 \leq k \leq n. \quad (1.6.2)$$

Bernstein [30; 41] remarkably strengthened it.

The other group of methods should have been based on asymptotic formulas for the probabilities of some behavior of a long series of consecutive sums, small with regard to the sums of the terms. Kolmogorov [13; 18] offered such formulas in 1931. His method of *stochastic differential equations* is also applicable to sums of certain dependent variables, and we shall therefore consider it below.

Let us go now to the findings obtained while solving problems of the described type. Khinchin & Kolmogorov [1] exhaustively determined the conditions for the series [1.3.1] with $n \to \infty$ of independent random variables to converge with probability 1. The results concerning the strong law of large numbers are less full. All that can be elicited from considering the variances of the terms is Kolmogorov's sufficient (but not necessary) condition [8] about the convergence of the series

$$(b_1/1^2) + (b_2/2^2) + \ldots + (b_n/n^2) + \ldots$$

It is natural to say that, similar to the case of the usual law of large numbers, the demand (1.6.1) determines the *strong normal stability* of the arithmetic means ζ_n/n, and to call these means *strongly stable* if there exist such constants C_n that

$$P[\lim |(\zeta_n/n) - C_n| = 0] = 1 \text{ as } n \to \infty. \quad (1.6.3)$$

Kolmogorov [24] ascertained that for identically distributed ξ_n the existence of finite expectations was necessary and sufficient for strong stability which is here always normal. Bobrov [4] discovered interesting results in this direction, but the problem of determining the necessary and sufficient conditions for strong stability remains open.

1.7. The Law of the Iterated Logarithm and Related Problems. Let us consider a sequence of random variables

$$\delta_1, \delta_2, \ldots, \delta_n, \ldots \quad (1.7.1)$$

and some quite definite (not dependent on randomness) function $f(n)$. We shall determine the set of such values of n for which $\delta_n > f(n)$. If this set is finite with probability 1, we shall call $f(n)$ the *upper function* of the sequence (1.7.1); if it is infinite, again with probability 1, $f(n)$ will be the *lower function*. According to the formulation of §1.6, the strong law of large numbers means nothing else but that for any $\varepsilon > 0$ the function $f(n) = \varepsilon n$ is the upper function for the absolute values $\delta_n = |\zeta_n - E\zeta_n|$ of the deviations of the sums ζ_n from their expectations. The desire to ascertain for each sequence of independent random variables $\xi_1, \xi_2, \ldots, \xi_n, \ldots$, the class of the upper functions for the corresponding absolute deviations δ_n is quite natural.

In 1924 Khinchin [2] provided an almost exhausting solution of this problem for the case of the Bernoulli trials (when the variables ξ_n were identically distributed and only took two values). His finding can be formulated thus: For any $\varepsilon > 0$

$$f(n) = (1 + \varepsilon)\sqrt{2B_n \log\log B_n} \quad \text{and} \quad g(n) = (1 - \varepsilon)\sqrt{2B_n \log\log B_n}$$

are the upper and lower functions respectively. Here, as above, we assume that $B_n = \text{var}\,\zeta_n$.

Later Kolmogorov [4] discovered that the same law also takes place under considerably more general conditions. After that the problem was mostly considered by foreign authors. Among the easily formulated findings we should indicate that the law of the iterated logarithm is in any case valid for identically distributed terms ξ_n under a single (unavoidable according to the very formulation of the law) restriction, the finiteness of their variances.

The conditions for the application of the law of the iterated logarithm are not yet exhaustively studied. Various foreign authors have expended much efforts on widening these conditions as well as on a more precise separation of the upper functions from the lower ones. The main fundamental progress was, however, achieved here in the Soviet Union by Petrovsky [2]. He himself only solved a problem to which Kolmogorov's abovementioned methods led and which belonged to the theory of differential equations. Petrovsky's research enables to formulate the following test: *Supposing that* $\Phi(t)$ *is a monotone function*, $f(n) = \Phi(B_n)$ *will be the upper function if the integral*

$$\int_0^\infty (1/t)\,\Phi(t)\exp[-\Phi^2(t)/2]dt$$

converges, and the lower if it diverges. Certain conditions justifying the transition to Petrovsky's problem are still required for applying his criterion; Erdös and Feller, see Feller (1943), ascertained them.

2. Sums of Weakly Dependent Random Variables

The main properties of sums of independent terms persist if the dependence between the terms is sufficiently weak, or sufficiently rapidly weakens when the difference between their serial numbers increases. In the latter case, the terms are supposed to be enumerated in some definite *natural* order. This idea, first put forward and worked out by Markov, was further developed in a number of Soviet investigations.

2.1. The Law of Large Numbers. If all the terms of the sum (1.1.6) have finite variances, then the variance $B_n = \text{var}\,\zeta_n$ can be written down as

$$B_n = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}^{(n)} \quad (2.1.1)$$

or as

$$B_n = \sum_{i=1}^{n} \sum_{j=1}^{n} \sqrt{b_i^{(n)} b_j^{(n)}}\; r_{ij}^{(n)} \quad (2.1.2)$$

where

$$c_{ij}^{(n)} = \mathrm{E}[(\xi_i^{(n)} - \mathrm{E}\xi_i^{(n)})\,(\xi_j^{(n)} - \mathrm{E}\xi_j^{(n)})]$$

are the mixed moments of the second order, and

$$r_{ij}^{(n)} = [c_{ij}^{(n)}/\sqrt{b_i^{(n)} b_j^{(n)}}\,] = R(\xi_i^{(n)};\ \xi_j^{(n)})$$

are the correlation coefficients between $\xi_i^{(n)}$ and $\xi_j^{(n)}$.

Formula (2.1.2) enables to elicit from the classical condition

$$\lim B_n = 0 \text{ as } n \to \infty \quad (2.1.3)$$

a number of other sufficient conditions for the normal stability of the sums ζ_n expressed through the variances of the terms $b_i^{(n)}$ and the coefficients $r_{ij}^{(n)}$. Bernstein and Khinchin engaged in this subject; thus, Bernstein [41] provided the following sufficient condition for the arithmetic means of the sequence of variables ξ_i:

$$\mathrm{var}\xi_i \le C,\ R(\xi_i;\ \xi_j) \le \varphi(|j - i|). \quad (2.1.4)$$

Here, C is a constant, and the function $\varphi(m)$ tends to zero as $m \to \infty$. It is natural, however, that, as we already noted in the case of independent terms, the problem cannot be solved in a definitive way only by considering the second moments.

In a few of his works Khinchin studied the applicability of the strong law of large numbers to sequences of dependent variables. Here, however, even the question of what can the second moments ensure is not ascertained in full [3]. Fuller findings are obtained for the cases in which the terms appear as a result of a random process of some special (Markov or stationary) type, see below.

2.2. The Classical Limit Theorem. Bernstein [11; 13; 18; 40] continued Markov's research of the conditions for the normal attraction to the Gauss law of sums of an increasing number of weakly dependent terms. His results relate to the pattern of Markov chains, see below. The formulations of his subtle findings directly expressed by demanding a sufficiently rapid weakening of the dependence of the terms when the difference between their serial numbers increases are rather complicated. We shall only note that for this problem the estimation of the dependence by the correlation coefficients is too crude. Instead of applying them we have to demand either complete independence for terms with sufficiently large serial numbers or to require that the conditional moments of the first and the second order of the terms, when the values of the previous terms are fixed, should little differ from the unconditional moments. The meaning of the conditions of the second kind became quite clear in the context of the theory of stochastic differential equations that emerged later.

3. The Ideas of the Metric Theory of Functions in Probability Theory

As a science devoted to a quantitative study of the specific domain of the *random*, the theory of probability is not a part of pure mathematics. Its relation to the latter is similar to that of mechanics or geometry, if geometry is understood as a science of the properties of the real space. Nevertheless, a purely mathematical part can be isolated from it just as from geometry. For the latter, this was done at the turn of the 19th century, when it, considered as a part of pure mathematics, was transformed into a science of a system of objects called *points, straight lines, planes*, and satisfying certain axioms.

A similar full axiomatization of the theory of probability can be carried out by various methods. During the last years, the development of concrete branches of the theory was especially strongly influenced by a construction that assumes as initial objects of study left without formal definition the set $U = \{u\}$ of *elementary events*, and the function $P(A)$ called *probability* having as its domain some system F of subsets of the main set U. In 1933 Kolmogorov [24] offered the appropriate axiomatics [4] in a complete form although the French school (Borel), and, from its very beginning, the Soviet Moscow school had begun to develop a related range of ideas much earlier.

From the viewpoint of logic and philosophy, this system of constructing the theory of probability is not either the only possible, or preferable to other systems, see our last section. Its great success is apparently due to the following circumstances:

1) It is the simplest system of full axiomatization of the theory from among those offered until now [5].

2) It enabled to cover, by a single simple pattern, not only the classical branches of the theory, but also those new chapters that were put forth by the requirements of natural sciences and are connected with distributions of probabilities for *random functions*.

3) It connected the theory with the theory of measure and the metric theory of functions which boast a rich collection of subtle methods of research.

We shall indeed concentrate on the two last points. According to the concept which we are now discussing, the probability $P(A)$ is nothing else but an *abstract measure* obeying the norming condition $P(U) = 1$; the random variable ξ is a function $\xi(u)$ measurable with respect to this measure; the expectation $\mathrm{E}\xi$ is the Lebesgue integral

$$\mathrm{E}\xi = \int_u \xi(u)dP, \text{ etc.}$$

3.1. The Joint Distribution of Probabilities of an Infinite System of Random Variables. A number of findings of the Moscow school discussed in §2 (and especially those related to the convergence of series with random terms; to the strong law of large numbers; and to the law of the iterated logarithm) is based on considering a sequence of random variables

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots \quad (3.1.1)$$

as a sequence of functions $\xi_n(u)$ of one and the same argument u with probability $P(A)$ being defined as measure on its domain U. We still have to indicate some more general results.

Problems of estimating the probabilities of events depending on the values of a finite number of variables (3.1.1) are solved by means of appropriate finite-dimensional distributions

$$F_{\xi_1,\xi_2,\ldots,\xi_n}(x_1; x_2; \ldots; x_n). \quad (3.1.2)$$

If, however, the occurrence of some events depends on values taken by an infinite number of variables from (3.1.1), then the law of distribution is naturally considered in the space of number sequences $x_1, x_2, \ldots, x_n, \ldots$ of the possible values of variables (3.1.1). This law is uniquely determined by the totality of distributions (3.1.2).

Kolmogorov [24] established a theorem which states that, for the existence of random variables (and, consequently, of their infinite-dimensional law) with given laws of distribution (3.1.2), the compatibility of these laws, in its usual elementary sense, is not only necessary but also sufficient.

The solution of many separate problems repeatedly led to the result that, under certain general conditions, the probability of some limiting relations concerning sequences of random variables can only be equal to 0 or 1. Analogies with the theory of measure prompted Kolmogorov [24] to establish the following general theorem: *If* $f(\xi_1, \xi_2, \ldots, \xi_n, \ldots)$ *is a Baire function of independent random variables* ξ_n *whose value persists when the values of a finite number of its arguments are changed, then the probability of the equality* $f(\xi_1, \xi_2, \ldots, \xi_n, \ldots) = a$ *can only be equal to* 0 *or* 1.

3.2. Random Functions. Suppose that a function $\xi_u(t)$ belonging to some functional space E corresponds to each *elementary event*, and that for any Borel measurable subset A of space E the set of those n for which $\xi_u(t) \in$ A, belongs to system F. Then we say that a *random function* $\xi(t)$ of type E is given. Following Kolmogorov's first findings [24], a number of American authors (especially

Doob) have developed this logical pattern. Independently of polishing its formal logical side, Slutsky systematically studied the most interesting problems naturally emerging when considering random functions.

An n-dimensional law of distribution $F_{t_1, t_2, \ldots, t_n}(x_1; x_2; \ldots; x_n)$ of the random variables $\xi(t_1)$, $\xi(t_2)$, …, $\xi(t_n)$ corresponds to each finite group of values $t_1, t_2, \ldots, t_n$ of the argument t. One of the main problems of the theory of random functions is, to determine the conditions to be imposed on the functions F so that they can correspond to a random function of some type E. In any case, the functions should be consistent in the elementary sense; additional conditions differ, however, for differing functional spaces. Slutsky and Kolmogorov (especially Slutsky [28]) offered sufficiently general conditions of this type for the most important instances. These conditions are expressed through the laws of distribution of the differences $\xi(t') - \xi(t)$, that is, through the two-dimensional distributions $F_{t, t'}(x; x')$. According to Slutsky, *stochastic continuity* of $\xi(t)$, *i.e.*, the condition that, for any $\varepsilon > 0$,

$$\lim P(|\xi(t+h) - \xi(t)| > \varepsilon) = 0 \text{ as } h \to \infty$$

for all (or almost for all) values of t, is sufficient for this space of measurable functions. For the same case, Kolmogorov established a somewhat more involved necessary and sufficient condition (Ambrose 1940). For the space of continuous functions Kolmogorov's sufficient condition is that there exist such $m > 0$ and $\alpha > 1$ that

$$\mathrm{E}|\xi(t+h) - \xi(t)|^m = \mathrm{O}|h|^\alpha.$$

The most essential from among the more special results concerning random functions of a real variable were obtained in connection with the concepts of the theory of random processes (where the argument is treated as *time*), see the next section. For statistical mechanics of continuous mediums both scalar and vectorial random functions of several variables (points in space) are, however, essential. Findings in this direction are as yet scarce (Obukhov, Kolmogorov).

4. The Theory of Random Processes

The direction of research now united under the heading *General theory of random processes* originated from two sources. One of these is Markov's work on *trials connected into a chain*; the other one is Bachelier's investigations of *continuous probabilities* which he began in accordance with Poincaré's thoughts. The latter source only acquired a solid logical base after the set-theoretic system of constructing the foundations of probability theory (§3) had been created.

In its further development the theory of random processes is closely interwoven with the theory of dynamic systems. Both conform to the ideas of the classical, pre-quantum physics. The fascination of both of them consists in that, issuing from the general notions of *determinate process* and *random process*, they are able to arrive at sufficiently rich findings by delimiting, absolutely naturally free from the logical viewpoint, the various types of phase spaces (the sets of possible states) and of regularities in the changes of the states (the absence, or the presence of aftereffect, stationarity, etc). A similar logical-mathematical treatment of the concepts of the modern quantum physics remains to a considerable extent a problem for the future.

4.1. Markov Chains. Suppose that the system under study can be in one of the finite or countable states $E_1, E_2, \ldots, E_n, \ldots$ and that its development is thought to occur in steps numbered by integers t (discrete time). Suppose in addition that the conditional probability of the transition $P(E_i \to E_j) = p_{ij}^{(t)}$ during step t does not depend on the earlier history of the system (absence of aftereffect). Such random processes are called *Markov chains*.

It is easy to see that the probability of the transition $E_i \to E_j$ during t steps having numbers 1, 2, …, t which we shall denote by $P_{ij}^{(t)}$ can be calculated by means of recurrent formulas

$$P_{ij}^{(1)} = p_{ij}^{(1)},\ P_{ij}^{(t)} = \sum_k P_{ij}^{(t-1)} p_{kj}^{(t)},\ t > 1. \quad (4.1.1)$$

Especially important is the time-homogeneous case $p_{ij}^{(t)} = p_{ij}$. Here, the matrix $(P_{ij}^{(t)})$ is equal to matrix (p_{ij}) raised to the power of t.

A natural question of the limiting behavior of the probabilities $P_{ij}^{(t)}$ as $t \to \infty$ presents itself. In case of a finite number of states and time-homogeneity Markov proved that when all the p_{ij} were positive there existed limits

$$\lim P_{ij}^{(t)} = P_j \text{ as } t \to \infty \quad (4.1.2)$$

which did not depend on the initial state E_1. Beginning with 1929, this result (Markov's ergodic theorem) became the point of departure for a long cycle of works done by Romanovsky and a number of foreign authors (Mises, Hadamard, Fréchet, et al). Romanovsky made use of algebraic methods dating back not only to Markov but also to Frobenius whereas Mises initiated *direct* probability-theoretic methods. Romanovsky, in his fundamental memoir [35], fully studied by algebraic methods the limiting behavior of the transition probabilities $P_{ij}^{(t)}$ under conditions of time-homogeneity and finiteness of the number of states. An exhaustive exposition of the solution by means of a *direct* method can be found in Bernstein's treatise [41]. Kolmogorov [29] was able completely to ascertain by direct methods the limiting behavior of these probabilities under time-homogeneity for a countable number of states.

Already Markov had studied non-homogeneous chains. For them, a natural generalization of the relation (4.1.2) is

$$\lim |P_{ik}^{(t)} - P_{jk}^{(t)}| = 0 \text{ as } t \to \infty \quad (4.1.3)$$

which expresses the vanishing of the dependence of state E_k after t steps (as $t \to \infty$) from the initial state (E_i or E_j). On related problems see Kolmogorov [22; 32] and Bernstein [41].

Markov's typical problem was the study of the sums of a sequence of random variables $\xi_1, \xi_2, \ldots, \xi_t, \ldots$ such that ξ_t takes value a_j if, after t steps, the system, having initially been in state E_i, finds itself in state E_j. The sums

$$\eta_t = \xi_1 + \xi_2 + \ldots + \xi_t$$

can obviously be written down as

$$\eta_t = \sum_j a_j \mu_{ij}^{(t)}$$

where $\mu_{ij}^{(t)}$ is the number of times that the system finds itself in state E_j during the first t steps given that its initial state was E_i. Therefore, from the modern point of view, the main natural problem is the study of the limiting behavior of the random variables $\mu_{ij}^{(t)}$ as $t \to \infty$.

For the case of time-homogeneity, all the problems connected with the usual, and the strong law of large numbers are easily solved. Sarymsakov devoted his article [9] to the law of the iterated logarithm. The problem of the limiting laws of distribution are deeper. For a finite number n of states the main problem consists in studying the limiting behavior as $t \to \infty$ of probabilities

$$Q_i^{(t)}(m_1; m_2; \ldots; m_n) = P(\mu_{i1}^{(t)} = m_1; \mu_{i2}^{(t)} = m_2; \ldots; \mu_{in}^{(t)} = m_n).$$

A number of authors have quite prepared a full solution of this problem; more precisely, they discovered the necessary and sufficient conditions for the applicability of the appropriate local limit theorem concerning the $(n-1)$-dimensional Gauss law (indeed, $m_1 + m_2 + \ldots + m_n = n$) and even for an exhausting study of special cases. However, definite and simple formulations of such results, from which would have inevitably followed the local and integral one-dimensional limit theorems for the sums η_t, are still lacking (see the works of Romanovsky and Sarymsakov, but, above all, Romanovsky [35]).

For Markov and Bernstein, the case of non-homogeneity with respect to time served as a subject for a subtle study of the boundaries of the applicability of the classical limit theorem for sequences of dependent random variables. They especially investigated the case of two states and a matrix

$$\begin{pmatrix} p_{11}^{(t)} & p_{12}^{(t)} \\ p_{21}^{(t)} & p_{22}^{(t)} \end{pmatrix}$$

reducible to a unit matrix. If, asymptotically, $p_{12}^{(t)} \approx c/n^{\alpha}$, $p_{21}^{(t)} \approx c'/n^{\alpha}$, then the limit theorem is applicable for $\alpha < 1/3$; generally, however, this is not the case anymore when $\alpha \geq 1/3$ (Bernstein [11]; [18]).

4.2. General Markov Processes. The main results of the theory of Markov chains essentially depend on assuming the absence of aftereffect. When keeping to this restriction, but going over to an arbitrary phase space $\Omega = \{\omega\}$ of possible states and abandoning the demand that the values of *time t* be integers, we shall arrive at the concept of general Markov process governed by the probabilities $P(t_1; t_2; \omega; \Delta)$ of transition from state ω to the set of states $\Delta \subseteq \Omega$ during time interval $(t_1; t_2)$. For any $t_1 < t_2 < t_3$ these probabilities obey the *Smoluchowski equation* [6]

$$P(t_1; t_3; \omega_0; \Delta) = \int_{\Omega} P(t_2; t_3; \omega; \Delta)\, P(t_1; t_2; \omega_0; d\omega). \quad (4.2.1)$$

Kolmogorov [10] developed the general theory of Markov processes and offered their classification. The special cases, important for applications, can be isolated on various ways:

1) The cases in which the phase space Ω is a finite or a countable set (as it was for Markov chains) or a differentiable n-dimensional manifold, etc are considered separately.

2) The case of discrete or continuous change of *time t* is studied.

3) The case of time-homogeneity in which the transition probabilities $P(t_1; t_2; \omega; \Delta)$ only depend on the time difference $(t_2 - t_1)$ is isolated.

4) The demand that the change of $\omega(t)$ with time is continuous is included; or, on the contrary, the number of moments when one state is step-wisely changed to another one is restricted.

5) Some differentiability is demanded on the distribution of probabilities P; the first such requirement is the condition that they be expressed through the corresponding densities of the probabilities of transition.

The emerging vast program is far from being fulfilled. Only some cases are studied in detail.

4.2a. The set Ω is finite, the time changes continuously, and the transition probabilities P are differentiable with respect to t_1 and t_2. These probabilities are here obeying linear differential equations which were made use of long before the general theory originated. For this case, the general theory is simpler, and leads to more simple and more polished formulations than the theory of Markov chains with discrete time does.

4.2b. The case of a countable set of states also leads, under some restrictions, to systems of linear differential equations which, however, include here an infinite set of unknown functions. Nevertheless, these are being successfully solved in a number of instances taking place in applications. The case of *branching processes* where the matter can be reduced to a finite system of non-linear differential equations (Kolmogorov & Dmitriev [1]; Kolmogorov & Sevastianov [1]) had been especially studied. It covers important patterns of branching chain reactions.

4.2c. Step-wise processes with any sets of states and continuous time were the subject of V.M. Dubrovsky's numerous studies.

4.2d. Bebutov [1; 2] as well as Krylov & Bogoliubov [1; 2; 3] investigated processes with discrete time under various restrictions superimposed on the topological or differential-geometric nature of the phase space and on the differentiability of the transition probabilities (at least concerning the existence of their densities with respect to some measure).

4.2e. Especially numerous studies are devoted to the case of continuous time with the phase space being a differential manifold when $\omega(t)$ is continuous and the transition probabilities are adequately differentiable. Kolmogorov showed that these probabilities obeyed here parabolic or hyperbolic partial differential equations that had first appeared in Fokker's and Planck's works on the Brownian motion. Kolmogorov [19], also see Yaglom [2], offered the general expressions for these equations.

The investigations of the behavior of $\omega(t)$ are reduced here to various boundary value problems for the appropriate partial differential equations. For examples of applying this method see Petrovsky [2] or Kolmogorov & Leontovich [1]. Bogoliubov & Krylov developed important applications of this tool in statistical physics, but their works are beyond the scope of our review.

4.2f. For one-dimensional *processes with independent increments* Kolmogorov and Lévy ascertained in all generality the analytical form of the transition probabilities. These are represented by means of *infinitely divisible* laws (§1). In a number of works Gnedenko and Khinchin studied more subtle problems about processes with independent increments.

The success of Kolmogorov and Lévy was assured by the applicability here of the method of characteristic functions. However, their result is not inseparably connected with it, and the decomposition of the process into a *step-wise* part and a *continuous* (for independent increments, inevitably a Gaussian) part, accomplished by these authors, can probably be generalized to a considerably more general case of Markov processes. It seems that under some very general conditions the Smoluchowski equation can be replaced by a mixed integro-differential equation of the type indicated at the end of Kolmogorov's contribution [32] on the general theory of Markov processes.

4.3. Stationary Processes. For Markov processes, the transition probabilities $P(t_1; t_2; \omega; \Delta)$ at a given state ω and moment t_1 play the main role. Generally speaking, the very existence of unconditional probabilities for some course of the process is not here assumed. We shall now adopt another viewpoint: The distribution of probabilities in the space of functions $\omega(t)$ is given with t taking all real values and the values of ω belonging to the phase space Ω.

The demand of time-homogeneity is here transformed into a demand for stationarity: the probability $P[\omega(t) \in A]$ that the function $\omega(t)$, describing the course of the process in time, belongs to some set A does not change if A is replaced by set $H_\tau A$ appearing after each $\omega(t)$ from A is replaced by $\omega(t - \tau)$.

The study of stationary processes with phase space Ω is tantamount to investigating dynamic systems with invariant measure in space Ω^R of functions $\omega(t)$ of a real variable with values from Ω. From among the general results the first place is occupied here by the celebrated Birkhoff – Khinchin ergodic theorem (Khinchin [25; 27; 52]).

4.4. The Correlation Theory of Processes Stationary in the Broad Sense. The stationarity of the process of the variation of $\omega(t)$ leads, for any functions $f(\omega)$ and $g(\omega)$ having finite expectations $E\{f[\omega(t)\}^2$ and $E\{g[\omega(t)\}^2$, to the expectations

$$E\{f[\omega(t)\, g[\omega(t+\tau)\} = B_{fg}(\tau)$$

being functions only of τ, and to the constancy of the expectations $E\{f[\omega(t)\} = A_f$, $E\{g[\omega(t)\} = A_g$.

When only turning our attention to the results that can be expressed through these first and second moments for a finite number of functions $f(\omega)$, $g(\omega)$, …, it is natural to pass on to the following concept of processes stationary in the broad sense: The phase space Ω is an n-dimensional space of vectors $x = (x_1, x_2, \ldots, x_n)$; the random process is given by the distribution of probabilities in the space of vector-functions $x(t) = [x_1(t), x_2(t), \ldots, x_n(t)]$, that is, by the joint law of distribution of n random functions $\xi_1(t)$, $\xi_2(t)$, …, $\xi_n(t)$, and stationarity is restricted by demanding that $E\xi_k(t) = A_k$ do not depend on time and

$$E[\xi_i(t)\, \xi_j(t + \tau)] = B_{ij}(\tau)$$

only depend on τ.

Without loss of generality it is possible to assume that $A_k = 0$ and $B_{kk} = 1$, $k = 1, 2, …, n$. Then $B_{ij}(\tau)$ is nothing but the correlation coefficient between $\xi_i(t)$ and $\xi_j(t + \tau)$. Khinchin [32] outlined the general patterns emerging from these propositions of the correlation theory of random processes. Kolmogorov [47] reviewed its further development in which Slutsky, he himself, V.N. Zasukhin and M.G. Krein had participated.

4.5. Other Types of Random Processes. A number of studies is devoted to various special types of random processes that do not belong to the patterns described above. These include the so-called generalized chains introduced already by Markov which can be reduced (as the number of states increases) to usual *simple* Markov chains. *Stochastic bridges* [6] introduced by Bernstein [21], which generalize the notion of Markov process, would have deserved a more detailed investigation. It is natural that the study of stationary processes leads to *processes with stationary increments*. Bogoliubov carried out important investigations of the limiting transitions from some types of random processes to other ones (for example, he studied the conditions for a limiting transition of a non-Markov process depending on a parameter into a Markov process).

5. A New Approach to Limiting Theorems for Sums of a Large Number of Terms

It is well known from the history of the classical analysis that a limiting transition from relations between finite differences to the corresponding differential expressions often leads to considerably simpler results as compared with those accomplished by a direct investigation of the initial relations. In a similar way, most asymptotic formulas derived by long reasoning in the classical studies of sums of a large number of random terms, or of results of a large number of trials, appear as precise solutions of naturally and simply formulated problems in the theory of random processes with continuous time. This idea that goes back to Poincaré and Bachelier is convincingly developed by means of simplest illustrations, see for example Khinchin [33]. It offers a leading principle for constructing new proofs and for discovering new formulations of the limit theorems of the classical type. Some foreign authors from Courant's school (Pólya, Lüneberg) were first to apply such a method and Kolmogorov [32] provided its general outline. In his abovementioned works [13; 18] he also applied it for proving the classical Liapunov limit theorem and limit theorems of a new type for sums of independent terms.

The method was further developed according to two methodologically different viewpoints. Bernstein systematically develops the *method of stochastic differential equations* but does not connect {it with} the limiting differential equations of the limiting probability-theoretic pattern of a random process with continuous time. On the contrary, Petrovsky and Khinchin always bear in mind this limiting random process.

As to actual results, Bernstein [22; 27; 31; 41] examined considerably more fully the difficulties that occur in cases of a non-compact phase space (such, for example, as even the usual real straight line with which he indeed has to do), whereas Petrovsky [1] and Khinchin [33] more fully studied *problems with boundaries*.

We note that the totality of the available theoretically rigorous contributions of this type does not at all cover the cases in which similar limit transitions are made use of in applications. Further intensive work in this direction is therefore desirable.

The abovementioned works of the three authors are mostly concerned with limit theorems connected with continuous random processes governed by the Fokker – Planck equations. Only Khinchin [33] generalizes the Poisson limit theorem for step-wise processes with continuous time. This theorem was the foundation of all the investigations of limit theorems for sums leading to infinitely divisible laws and expounded in detail in §1. It should be noted in this connection that there are yet no limit theorems corresponding to more general step-wise or mixed processes with continuous time governed by integro-differential equations indicated above.

6. The Logical Principles of the Theory of Probability

From the viewpoint of justifying the theory it is natural to separate it into two parts. The first, *elementary* component is only concerned with patterns involving a finite number of events. The *non-elementary* part begins when some random variable is assumed to be able to take an infinite number of values (for example, when it is being strictly subject to the Gauss law) and extends up to the most modern constructions with distributions of probabilities in various functional spaces. It is evident in advance that the entire second part connected with mathematical infinity cannot claim to have a simpler relation to reality than the mathematics of the infinite in general, *i.e.*, than, for example, the theory of irrational numbers or the differential calculus.

6.1. The Logical Principles of the Elementary Theory of Probability. The events of any finite system can be compiled from a finite number of pairwise incompatible events. The matter is therefore completely reduced to

1) Ascertaining the real sense of the following pattern: Under some conditions S one and only one of the events $A_1, A_2, \ldots, A_n$ necessarily occurs with each of them having, given these conditions, probability $P(A_i)$;

2) Justifying that the pattern in Item 1) always leads to

$$P(A_i) \geq 0, P(A_1) + P(A_2) + \ldots + P(A_n) = 1.$$

This problem is known to be solved in different ways. Some believe that the determination of the probabilities $P(A_i)$ is only scientifically sensible if the conditions S can be repeated indefinitely many times and they substantiate, by some method, the concept of probability by the frequency of the occurrence of the event. Other authors, however, consider that the concept of *equiprobability* (which, generally speaking, can be introduced without demanding an indefinite repetition of the conditions S) is primary and assume it as the base for defining the numerical value of probability. Both these approaches can be subjectively and idealistically colored; apparently, however, they can also be worked out from the viewpoint of objective materialism. The pertinent Soviet literature is scarce; in addition to introductions in textbooks and popular literature, we only indicate Khinchin's critical paper [13].

6.2. The Substantiation of the Non-Elementary Chapters of the Theory of Probability. When retracting the finiteness of the system $F = \{A\}$ of events to which definite probability {probabilities} $P(A)$ is {are} assigned, it is natural to demand that these events constitute a *Boolean algebra* with the probability being a non-negative function of its element equal to 1 for a *unit* of this algebra (Glivenko [6]). It is also possible to demand that $P(A) = 0$ only for the *zero* of the algebra (that is, for the one and only *impossible* event).

The question about *countable additivity* of the probability does not arise here since for a Boolean algebra a sum of a countable number of elements has no sense. It is natural, however, to define the sum $\bigcup A_k$, $1 \leq k < \infty$, as an event A for which

$$\lim P[(\bigcup A_k - A) \bigcup (A - \bigcup A_k)] = 0 \text{ as } n \to \infty.$$

Then countable additivity will certainly take place.

The Boolean algebra of events can be *incomplete* in the sense that a countable sum of its elements does not always exist. Then, however, it can be *replenished* and this operation is uniquely defined. It is as natural as the introduction of irrational numbers. On principle, Glivenko's concept just described is the most natural. However,

1) The axiomatics of the theory of probability understood as the theory of complete normed Boolean algebras is rather complicated.

2) In this theory, the definition of the concept of random variable is too complicated.

It is therefore necessary to make use of the main Boolean algebra as a Boolean algebra of the sets of *elementary events*. Such events can be introduced as ideals of the main algebra (Glivenko [6]) which again leads to the Kolmogorov system (§3).

The shortcoming of the transition to the set-theoretic concept with elementary events is that it invariably leads to non-empty events having zero probability. That the sacrifice of one of the natural principles of the elementary theory of probability is unavoidable is already seen in the simplest examples of random variables with continuous distributions.

Notes

1. Our further exposition is centered around scientific problems rather than scientific schools.

2. A *singular* distribution function describes a constant magnitude.

3. The problem can be precisely posed in the following way: Under what conditions imposed on the numbers c_{ij} can we guarantee the validity of the relation

$$P\{\lim [|\zeta_n - E\zeta_n|/n] = 0\} = 1$$

for the sums (1.3.1) of random variables ξ_n having second moments

$$c_{ij} = E[(\xi_i - E\,\xi_i)(\xi_j - E\,\xi_j)]?$$

4. For the theory of probability, Borel's countable additivity of probability was new here. On the justification of this axiom see our last section. Right now, we only note that all the interesting concrete results based on this axiom allow also a *pre-limiting* interpretation independent from it; see for example Bernstein's interpretation [41, pp. 155 – 156] of the strong law of large numbers.

5. That is, from among those systems where any appeal to the obvious meaning of such notions as *event, incompatible events, event A is a corollary of event B*, etc is ruled out.

A**6.** {Later Gnedenko began calling the equality (4.2.1) the *generalized Markov equation*, see his *Курс теории вероятностей* (Course in Theory of Probability). M., 1954, p. 387, and *Lehrbuch der Wahrscheinlichkeitsrechnung*. Berlin, 1968, p. 287.}

7. Boiarsky [3] indicated their interesting two-dimensional analogue.

8. N.V. Smirnov. Mathematical statistics: new directions.

Vestnik Akademii Nauk SSSR, No. 7, vol. 31, 1961, pp. 53 – 58

[Introduction] For a long time *mathematical* or *variational* statistics was understood as a special discipline that justified various methods of studying the biological phenomena of variability and heredity. This comparatively narrow range of problems constituted the main subject of researches of the British Biometric school headed by Karl Pearson. Soon, however, the methods that he had advanced found fruitful applications in a number of other fields, – in meteorology, geophysics, hydrology, agronomy, animal science, forestry, in problems of checking and testing mass production, etc. Under the influence of the increasing demands this discipline developed during the last decades in a considerably wider channel. At present, we already see a sufficiently shaped outline of a new branch of mathematics aiming at developing rational methods of studying mass processes.

For the last 30 years, the works of Soviet mathematicians were playing a sufficiently important progressive part in developing mathematical statistics. The splendid achievements of our mathematicians in the directly adjoining field of probability theory (of Kolmogorov, in axiomatics and the theory of stochastic processes; of Bernstein, in limit theorems; of Khinchin, et al) obviously influenced the progress in this science to a very considerable extent.

The usual theoretical pattern (that does not, however, claim to be exhaustively general) with which various formulations of the problems in mathematical statistics are connected, is known well enough. Here, the object of investigation is some system whose states are described, from the viewpoint that interests us, by a definite number of parameters. In the simplest cases this set is finite. To illustrate: the set of the studied characteristics of a biological individual (stature, weight, volume, etc); the set of the coordinates and impulses of a certain number of the particles of a physical system. In more complicated cases the set of the parameters is infinite, as it occurs for example for the field of the velocities of a turbulent current of liquid; for the field of pressure or temperature of the Earth's atmosphere, etc. The exceptional complexity of the processes taking part in such systems compels us to apply statistical methods of research. In following the statistical approach, we consider each observed state of the system as a random *representative*, or specimen selected by chance from an abstract *general* population of the states possible under identical general conditions. We assume that over this general population the random parameters can possess some distribution of probabilities corresponding to certain conditions usually formulated as a hypothesis. In the simplest cases this will be a multivariate distribution; and, for an infinite number of parameters, a distribution of a random function or of a random field in a functional space.

The observed data can be either the registered states of a more or less vast population of specimens of the given system (the states comprising a *sample* from the general population), or only some mean (space or temporal) characteristics of the states of the system. The interrelations between the empirical material and the theoretically allowed distribution of the general population constitute the main subject of mathematical statistics. Included problems are, for example, the fullest and most precise reconstruction of the law of distribution of the general population, given the sample; an adequate check of various hypotheses concerning this population; an approximate estimation of the parameters and of the theoretical means characterizing the theoretical distribution; an interpretation of various relations and dependences observed in samples; and many other practically and theoretically vital points originating in the applications of the statistical method.

I shall now go over to characterize separate prominent achievements of Soviet scholars in solving the most important problems of mathematical statistics.

1. The Theory of the Curves of Distribution. Correlation Theory

The limit theorems of the theory of probability which determine the applicability, under very general conditions, of the normal law to sums of independent or almost independent variables, ensure the suitability of the theoretical model of a normally distributed population to many concrete problems. Already the early statistical investigations made by Quetelet and then widely developed by the British Galton – Pearson school ascertained that the normal law was rather broadly applicable to biological populations. At the same time, however, it was also established that considerable deviations from the usual picture of the normal distribution, viz., an appreciable skewness and an excess of some empirically observed distributions, were also possible. To describe mathematically the distributions of such a type, Pearson introduced a system of distribution functions which were solutions of the differential equation

$$(1/y)\ \frac{dy}{dx} = \frac{x-a}{b_0 + b_1 x + b_2 x^2}$$

and worked out in detail the methods of determining the parameters of the appropriate curves given the empirical data. It occurred that the Pearsonian curves, very diverse in form, were applicable for interpolation in a broad class of cases. However, for a long time their stochastic nature was left unascertained; Pearson's own substantiation that he provided in some of his writings was patently unsound and led to just criticisms (Yastremsky [1]). The problem remained unsolved until Markov [1] showed how it was possible to obtain limiting distributions of some of the Pearsonian types by considering an urn pattern of dependent trials (that of an *added* ball). […] Pólya (1930), who apparently did not know Markov's findings, minutely studied this scheme of *contagion*, as he called it. Bernstein [34], Savkevich [1] and Shepelevsky [2] considered some of its generalizations.

Kolmogorov [32] outlined another approach to theoretically justifying the Pearsonian curves. He obtained their different types as stationary distributions that set in after a long time in a temporal stochastic process under some assumptions about the mean velocity and variance of the alteration of the evolving system's random parameter. Bernstein [27] proved that under certain conditions such a stationary distribution exists. Ambartsumian [1; 2] investigated in detail particular cases of stochastic processes leading to the main Pearsonian curves.

Romanovsky [20] generalized the Pearsonian curves to orthogonal series similar to the well-known Gram – Charlier series, Bernstein [12; 41] also studied another stochastic pattern admitting in may practically important cases a very concrete interpretation and leading to some transformations of the normal distribution.

Still more considerable are the achievements of Soviet mathematicians in the domain of correlation theory which already has vast practical applications. The works of Bernstein [13] and Khinchin [8] on the limit theorems for sums of random variables ensured a solid theoretical foundation for the theory of normal correlation. Bernstein [41] discovered interesting applications of these propositions to the case of hereditary transmission of polymeric indications (depending on a large number of genes). His work led to a theoretical explanation of the law of hereditary regression empirically established by Galton.

Bernstein's research [15; 16] into the geometric foundations of correlation theory are of paramount importance. He classified various surfaces of correlation according to simple geometric principles. If the change of one of the random variables only results in a translation of the conditional law of the distribution (of the density), the correlation is called *firm* {French original: *dure*}. Normal correlation is obviously firm, and a firm and perfect correlation is always normal. If all the conditional laws of one variable corresponding to various values of the other one can be obtained by contracting (or expanding) one and the same curve, the correlation is *elastic*. A more general type of *isogeneous* correlation is such that the elastic deformation of the conditional law is at the same time accompanied by a translation. Bernstein derived a differential equation that enabled him to determine all the types of the firm correlation and some particular cases of the isogeneous type. Sarmanov [1; 2] definitively completed this extremely elegant theory. The surfaces of isogeneous correlation are represented as

$$F(x; y) = \delta[Dx^2y^2 + 2Gx^2 y + 2Exy^2 + Ax^2 + By^2 + 2Hxy + I]^c.$$

In some cases the conditional laws are expressed by the Pearson curves. In the general case isogeneous correlation is heteroscedastic (with a variable conditional variance). The regression curve of y on x has equation

$$y = -\frac{Gx^2 + Hx + I}{Dx^2 + Ex + F}$$

and a similar equation exists for the regression of x on y.

Obukhov [1; 2] developed the theory of correlation for random vectors first considered by Hotelling (1936). It is widely applied in meteorological and geophysical problems, in the theory of turbulent currents and in other fields. Making use of tensor methods, he was the first to offer an exposition in an invariant form. He introduced tensors of regression and of conditional variance and determined various correlation invariants. For the normally distributed n-dimensional random vector the normal law is completely defined by the vector of the expectation and the tensor of variance. Obukhov's proposition on the canonical expansion of the correlation density that he proved reduces the study of multidimensional vectorial correlations to the case of one-dimensional vectors.

Romanovsky [21] studied various problems concerned with the connection of qualitative indications (the theory of association).

When determining the number and the size of the samples needed for establishing the mean value of some quantitative indication distributed over a certain area (for example, of the harvest, or the content of metal in ore), we have to allow for correlation between different points of the area. Boiarsky [3] made an interesting attempt to isolate continuous isotropic random fields of the Markov type and Obukhov [5] obtained more general results. Romanovsky [21], Nemchinov [1], Mitropolsky [2; 3; 4] and Lagunov [1] studied problems connected with the determination of the equations of regression.

2. Distribution of Sample Statistics. Estimating the Parameters of the Laws of Distribution

The problems of sampling, that is, of the methods of approximately determining various characteristics of the general population given the empirical material, can be very diverse depending on the nature of the theoretical law of distribution and the organization of the observations. A rational choice of such functions of the observations (the choice of *statistics*, as Fisher called them) which provide, under given conditions, the best (in a certain sense) approximation of (*information* on) the estimated theoretical magnitudes (for example, of/on the parameters of the law of distribution) is a complicated problem. The precision of the approximation can be estimated in full if the law of distribution of the sample *statistic* is known. In this case, it is also possible to evaluate the greater or lesser suitability of the chosen statistic as compared with other possible functions to serve as an approximate measure of the estimated parameter. The investigation of the laws of distribution of the various kinds of empirical means (means, variances, correlation coefficients, etc) is therefore one of the most important problems of mathematical statistics. Those mostly studied, naturally occurred to be samples from normal populations, and many contributions of English and American statisticians headed by Fisher were devoted to this subject. The possibility of entirely describing the normal distribution by a small number of parameters and the comparative simplicity of calculations clear the way for a deep analysis of the various relations between the general population and the sample that represents it.

Prominent and universally recognized achievements in this domain are due to Romanovsky. Issuing from the notions of the British school, his writings [8; 10; 11; 13; 18; 19; 21] nevertheless advantageously differ since they are based on rigorous methodological lines and are free from a rather considerable jumble of the main assumptions; indeed, he overcame the confusion of empirical and stochastic elements so characteristic of the English statisticians. With considerable analytic mastery Romanovsky applies the method of generating functions which leads to peculiar inversion problems in the theory of integral equations of the first kind.

He was the first to derive rigorously the laws of distribution of the well-known Student – Fisher t- and z-criteria, of empirical coefficients of regression and of a number of other statistics. A summary of his main results can be found in his well-known treatise [37] that played a fundamental part in the heightening of the mathematical level of the statistical thought.

Kuznetsov [2] studied the distributions of the length and the argument of a radius vector by the normal distribution of its components. Kuzmin [2] investigated the asymptotic behavior of the law of distribution of the empirical correlation coefficient (derived by Fisher). Smirnov [10] discovered the distribution of the maximal deviation (normed by the empirical variance) of observations from their empirical mean. This enabled him to make more precise the well-known Chauvenet rule for rejecting outliers [1].

Smirnov [2; 6] studied the terms of the variational series, *i.e.*, of the observed values of a random variable arranged in order of ascending magnitude, and established the appropriate limit laws under rather general assumptions. Gnedenko [15; 23] obtained interesting results about the distribution of the extreme terms of such series. There are three and only three limiting distributions of these terms (Fisher & Tippett and Mises); for the maximal term these are

$\Phi_\alpha(x) = 0$ if $x \le 0$ and $= \exp(-x^{-\alpha})$ otherwise;
$\Psi_\alpha(x) = \exp[-(-x)^\alpha]$ if $x \le 0$ and $= 1$ otherwise, $\alpha > 0$;
$\Gamma(x) = \exp(-e^{-x})$, $|x| < +\infty$.

By very subtle methods Gnedenko ascertained necessary and sufficient conditions for the occurrence of each of these and delimited in full the domains of their attraction. The works of Gumbel show that this theory finds applications in hydrological calculations (volumes of reservoirs), investigations of extreme age brackets, civil engineering, etc. Making use of Gnedenko's method, Smirnov, in a not yet published paper, presents an exhaustive to a certain extent classification of the limit laws for the central terms of the variational series and of the domains of their attraction.

Problems connected with a rational construction of *statistics* most effectively estimating the parameters of a theoretical law of distribution for a given size of the sample are urgent for modern science. Here, the classical approach, when the estimated parameter is considered as a random variable with some prior distribution is in most cases fruitless and the very assumption that a prior distribution exits is often unjustified. Fisher and Neyman put forward a new broad concept. It sees the main problem of the statistical method in establishing substantiated rules aiming at selecting hypotheses compatible with the observed data from among those admissible in the given concrete area of research. These rules should, first, be sufficiently reliable, so that, when used regularly, they would practically seldom lead to mistaken results; and, second, they should be the most effective, so that, after accounting for the observational data, their use would narrow the set of admissible hypotheses as much as possible. The measure of the good quality of a statistical rule is the *confidence coeffcient* defined as the lower bound of the probabilities of a correct conclusion resulting from the rule. Fisher and Neyman developed methods that allow, when only issuing from the sample data (without introducing prior probabilities), to indicate *confidence boundaries* that correspond to the assumed confidence coefficient and cover the estimated parameter of the general population. The revision of the already established methodology and the development of new ideas is the main channel of modern scientific work for those engaged in this domain.

Kolmogorov [43] presented an original interpretation of these ideas which specifies some subtle logical points as applied to the simplest problem of estimating the parameters of the Gauss law by a restricted number of observations. Bernstein [37] indicated the difficulties connected with Fisher's concept which restrict the applicability of his methods by conditions justified within the boundaries of the classical theorems. In the final analysis, the estimation of the efficiency of statistical rules is inseparable from an accurate notion of the aim of the statistical inmvestigation. The peculiarity of the logical situation and the uncommonness of the introduced concepts led to a number of mistaken interpretations (Fisher himself was also guilty here), but the fruitfulness of the new way is obvious. For our science, the further development of the appropriate problems is therefore an urgent necessity. Romanovky [44] and Brodovitsky [2] described and worked out a number of pertinent problems. Again

Romanovsky [45] and Kolmogorov [46] revised the canonical explication of the method of least squares and the theory of errors in the spirit of the new ideas [2].

3. Statistical Estimation of the Laws of Distribution. Testing the Goodness of Fit. Hypothesis Testing

As stated in the Introduction, one of the main problems of mathematical statistics is to reconstruct the theoretical law of distribution of the general population, given the empirical distribution of the sample. For a one-dimensional population this problem is reduced to establishing the distribution function $F(x)$ of some random variable by issuing from independent observations. If the variable ξ is discrete and takes s values $x_1, x_2, \ldots, x_s$ the data are separated in a natural way into s groups of $m_1, m_2, \ldots, m_s$ observa-tions ($m_1 + m_2 + \ldots + m_s = n$) corresponding to each of the possible values of ξ. The probable measure of approximation attained when the size of the sample is not too small can be estimated by asymptotic formulas; inversely, they also enable us to determine the adequate number of observations for guaranteeing, with a given degree of probability, the necessary closeness of the empirical and the theoretical distributions of probability. If the theoretical distribution is assumed to be known, and the admissibility of the observed deviation is questioned, the answer is rather fully obtained by means of the well-known Pearsonian chi-squared test whose rigorous theory was presented by Romanovsky [22]. For a continuous distribution function $F(x)$ all such problems become considerably more difficult; only recently the Moscow mathematicians Glivenko, Kolmogorov and Smirnov rigorously solved them.

The relative frequency $S_n(x)$ of those observations that do not exceed the given number x is called the empirical distribution function; it is depicted by a step-curve. Glivenko [4] offered the first rigorous proof that the empirical curve uniformly converges to the continuous theoretical law with probability 1. This gifted mathematician, who died prematurely, based his reasoning on a very abstract notion of the law of large numbers in functional spaces and developed it in a number of interesting writings. He [4] discovered the following simple estimate (from which the abovementioned theorem was immediately derived):

Let

$$D_n = \sup |S_n(x) - F(x)|, \ |x| < +\infty,$$

then, for any continuous $F(x)$ *and* $\varepsilon < 0$,

$$P(D_n < \varepsilon) > 1 - [1/(\varepsilon^6 n^2)].$$

Kolmogorov [16] proved a more precise proposition:

Let

$$\Phi_n(\lambda) = P(D_n \le \lambda/\sqrt{n}), \ \lambda > \lambda_0 > 0.$$

Then, for any continuous law of distribution $F(x)$, *the sequence of functions* $\Phi_n(\lambda)$ *tends to the limiting law*

$$\Phi(\lambda) = \sum_{n=-\infty}^{\infty} (-1)^k \exp(-2k^2\lambda^2)$$

as n increases.

The independence of the limiting distribution from $F(x)$ leads to a remarkable corollary: *Given a certain level of probability* (*confidence coefficient*) α *and having determined* λ_α *from the equation* $\Phi(\lambda) = \alpha$, *it is possible to state, for any sufficiently large n with probability* α, *that for any x the deviation* $|S_n(x) - F(x)|$ *will not exceed* $\lambda_\alpha / \sqrt{n}$.

The same Kolmogorov theorem can easily be applied for checking the degree of agreement between the theoretically admissible law and the empirical distribution. Smirnov [11] specified Glivenko's and Kolmogorov's results. One of his most general findings, of which the Kolmogorov theorem is a particular case, is formulated thus:

Let the curves

$$y = y_1(x) = F(x) + \lambda/\sqrt{n} \text{ and } y = y_2(x) = F(x) - \lambda/\sqrt{n}$$

delimit a band covering the theoretical curve $y(x) = F(x)$; *let* $\nu_n(\lambda)$ *be the number of times when the empirical curve exceeds the band, or, the number of points located on the vertical steps of* $S_n(x)$ *where this step-curve intersects*
$y = y_1(x)$ *or* $y = y_2(x)$; *introduce also*

$$\Phi^+_n(t; \lambda) = P(\nu_n(\lambda) < t\sqrt{n}) \text{ for } t > 0.$$

Then, as $n \to \infty$, *the sequence of* $\Phi^+_n(t; \lambda)$ *converges to the limiting function*

$$\Phi^+(t; \lambda) = 1 - 2\sum_{m=0}^{\infty} \frac{(-1)^m d^m}{m!dt^m} \{t^m \exp[-(1/2)(t + 2\sqrt{m+1}\lambda)^2]\}$$

for any continuous law $F(x)$.

This theorem provides a precise estimate of random oscillations of the empirical curve as it approaches the theoretical function $F(x)$. I also note a simple and most precise estimate

$$\Phi^+_n(\lambda) = P\{\text{Sup}[S_n(x) - F(x)] < \lambda/\sqrt{n}\} =$$
$$1 - \exp(-2\lambda^2)[1 - (2/3)\lambda/\sqrt{n} + O(\lambda^2/n)].$$

None of these results are restricted by assumptions about the types of the theoretical law $F(x)$: the set of admissible hypotheses is here unusually broad whereas classical problems considered only more or less special families depending on a finite number of parameters (*the parametric case*) so that everything was only reduced to estimating them in the best way.

Mises provided another test of goodness of fit that estimates the closeness of distributions in the sense of the weighted mean square measure. Smirnov [4] worked out the full theory of this test (of the Mises ω^2-test): when assigning an appropriate weight, its law of distribution also becomes independent of the type of the theoretical distribution $F(x)$. Under the same broad assumptions about the law $F(x)$ it occurred possible to interpret the problem about two independent samples belonging to one and the same general population. If $S_m(x)$ and $S_n(x)$ are empirical curves corresponding to independent samples of large sizes m and n respectively, and the hypothesis about the constancy of the law $F(x)$ is correct, then (Smirnov [7]) the test

$$D_{mn} = \sqrt{mn/(m+n)} \sup |S_m(x) - S_n(x)|, \; |x| < +\infty,$$

is distributed according to the Kolmogorov law.

Romanovsky [44] and Sarymsakov proved a number of interesting propositions adjoining these results that comply with the general trends of the modern statistical theory.

Let S be the general population with a discrete argument ξ *taking values* $x_1, x_2, \ldots, x_s$ *with probabilities* $P_1, P_2, \ldots, P_s$ ($P_i > 0$ *and* $P_1 + P_2 + \ldots + P_s = 1$); *let also the function* $\theta = \varphi(x_1, x_2, \ldots, x_s; P_1; P_2; \ldots; P_s)$ *represent some characteristic of the general population and* $T = \varphi(x_1, x_2, \ldots, x_s; m_1/n; m_2/n; \ldots; m_s/n)$ *be the corresponding characteristic of the empirical distribution obtained by replacing the* P_i*'s by the frequencies* m_i/n *of the sample values of* ξ. *If* φ *is continuous with respect to all the* P_i*'s, then, as* $n \to \infty$, $P(|T - \theta| < \varepsilon)$ *converges to* 1 *uniformly with respect to the* P_i*'s for any* $\varepsilon > 0$.

Sarymsakov specified this result by discovering that, *as* $n \to \infty$, a stronger relation $P(T \to \theta) = 1$ takes place.

Romanovsky [19] also considered a number of problems, this time of the parametric type, being connected with testing hypotheses of whether two independent samples belonged to one and the same normal population. Of special importance is his derivation of the distribution of the θ-test (also introduced by him) which found application in the so-called analysis of variance in agronomy and other similar fields. He [42] also studied statistical problems relating to series of events connected into a Markov chain, indicated methods for empirically ascertaining the law of the chain and for testing the hypothesis of its simplicity.

4. Problems of Prediction. Discovering Periodicities. Some Applications of Statistical Methods

In concluding my not at all comprehensive review, I indicate a number of writings devoted to quite concrete problems but at the same time having a considerable general methodical interest. Slutsky's studies [8; 16; 17; 18; 20; 21; 22; 24; 25] occupy a prominent place among these. They were devoted to connected time series and prediction and extrapolation, came to be widely known and enjoyed considerable response in the world literature. Slutsky, Khinchin and Kolmogorov largely created the theory of continuous stochastic processes. This led Slutsky to extremely interesting conclusions about pseudoperiodic properties of some classes of stationary random series (the limiting sinusoidal law). His further studies in this field were followed by the reconstruction of the Schuster theory of periodograms on incomparably broader foundation as well as by the revision of all the usual methods of estimating statistical constants only valid when the consecutive observations were independent. Slutsky threw new light on problems of correlation, prediction and extrapolation of connected series which had already for a long time attracted the attention of the most prominent representatives of our national statistics (Obukhov, N.S. Chetverikov, B.S. Yastremsky). With regard to the deepness of theoretical penetration into the nature of these difficult problems still awaiting to be scientifically solved, and to the wit of the methods applied, his investigations leave similar attempts made by West European and American statisticians far behind.

Peculiar *problems of congestion* that occur when providing mass service (automatic telephony) were the subject of deep research accomplished by Khinchin [19; 24] and Kolmogorov and continued by Volberg [1] and Bukhman [1; 2]. Obukhov [5], while studying the theory of turbulence by developing the profound ideas introduced there by Kolmogorov, worked out very complicated problems in statistical description of continuous random fields. A new development occurred in the problem of the general structure of mean values (Kolmogorov [9], Konius [1], Boiarsky et al [1]).

The penetration of statistical methods into most various fields of research is characterized by the abundance of new problems demanding special approaches and methods for solving them. Here I only mention the very interesting works of Kolmogorov on the theory of crystallization of metals [28] and on the law of distribution of the sizes of crushed particles

[40] and on the writings of a number of hydrologists on economic calculations (S.I. Kritsky, M.F. Menkel, A.D. Savarensky et al).

We can be justly satisfied in that the Soviet statistical thought had undoubtedly played a fundamental part in a number of paramount problems and provided specimens not yet surpassed with respect to deepness or ideological richness.

Notes

1. {With respect to this, nowadays hardly well-known rule, see Sheynin, O. (1994), Bertrand's work on probability. *Arch. Hist. Ex. Sci.*, vol. 48, pp. 155 – 199 (p. 190, Note 57).}

2. {Romanovsky [45] contained no *new ideas* worthy of mention. As many other most eminent mathematicians, he had not really studied the theory of errors.}

9. Joint Bibliography to the Papers by Gnedenko & Kolmogorov and Smirnov

Foreword by Translator

The Bibliography was definitely corrupted by mistakes, and the page numbers were sometimes missing. When possible, I corrected/supplemented it by consulting M.G. Kendall & Alison G. Doigt, *Bibliography of Statistical Literature Pre-1940*. Edinburgh, 1968. The authors referred to a few foreign sources in separate footnotes; I collected these and now they constitute a very short second part of the Bibliography. I have additionally listed the collected works of a few scholars published after 1948; in such cases, I did not mention lesser known periodicals in which their pertinent papers had initially appeared. I regret to add that the Bibliography in Bogoliubov, A.N. & Matvievskaia, G.P. *Romanovsky*. Moscow, 1997, was compiled carelessly. The DAN (see Abbreviations) were also published as the *C.r. Acad. Sci.* of the Soviet Union with the contributions appearing there in one of the three main European languages. If not stated (or not immediately apparent), the contributions are in Russian.

Abbreviations

AN = Akademia Nauk
C.r. = *C.r. Acad. Sci. Paris*
DAN = *Doklady AN USSR*
FAN = Filial AN
GIIA = *G. Ist. Ital. Attuari*
IAN = *Izvestia AN USSR*. If not indicated, the series is Matematika
IMM = Inst. Matematiki i Mekhaniki
L = Leningrad
LGU = Leningradsk. Gosudarstvenny Univ.
M = Moscow
MGU = Moskovsky Gosudarstvenny Univ.
MS = *Matematich. Sbornik*
NI = Nauchno-Issledovatelsk.
SAGU = Sredneaziatsk (Central Asian) Gosudarstvenny Univ.
SSR = Soviet Socialist Republic (e.g., Uzbekistan or Ukraine)
U = in Ukrainian
Uch. Zap. = Uchenye Zapiski
UMN = *Uspekhi Matematich. Nauk*
Uz = Uzbek
VS = *Vestnik Statistiki*

Part 1. Russian/Soviet Authors

Ambartsumian, G.A. 1. Curves of probability distributions leading in the limit to the Pearson curves. DAN 16, 1937, 259 – 262.

2. Considering a particular type of a continuous stochastic process. *Uch. Zap. LGU*, ser. Math., 10, 1940, 120 – 138.

Andronov, A.A., Vitt, A.G., Pontriagin, L.S. 1. Statistiche Auffassung dynamischer Systeme. *Fizich. Zhurnal*, 6, 1934, 1 – 24.

Anuchin, S.A. 1. The treatment of laboratory trials by mathematical statistics. *Izvestia Tekstiln. Promyshlennost* 8, 1931, 25 – 27.

Bakhurin, I.M. 1. A study of the method of least squares. *Zapiski L. Gorn. Inst.* 7, 1926, 13 – 25.

Barkalaia, A. 1. On Markov chains. *Uch. Zap. MGU* 73, ser. Math., 1944, 33 – 36.

Bavli, G.M. 1. Generalization of the Poisson limit theorem. DAN 2, 1935, 508 – 511.

2. Über einige Verallgemeinerungen der Grenzwertsätze der Wahrscheinlichkeitsrechnung. MS 1(43), 1936, 917 – 930.

3. Same title as **1.** *Uch. Zap. Sverdlovsk. Univ.* 2, 1937, 3 – 6.

4. On the local limit theorem. Ibidem, 7 – 24.

5. Über den lokalen Grenzwertsatz. *Rev. Fac. Sci. Univ. Istanbul*, 2:2, 1937, 1 – 14.

Bebutov, M.V. 1. Markov chains with a compact space of states. DAN 30, 1941, 480 – 481.

2. Same title. MS 10 (52), 1942, 213 – 238. (In English)

Bernstein, S.N. 1. Essay on axiomatic justification of the theory of probability (1917).

2. On the law of large numbers. (1918).

3. On the interrelation between the qualitative mark of the harvest and its actual weight, etc. *Statistich. Bull. Tsentraln. Statistich. Upravlenie Ukrainy* 4, 1921, 2 – 5.

4. On the application of mathematics in biology. *Nauka na Ukraine* 1, 1922, 14 – 19.

5. Sur le théorème limite du calcul des probabilités. *Math. Ann.* 85, 1922, 237 – 241.

6. Démonstration mathématique de la loi d'hérédite de Mendel. *C.r.* 177, 1923, 528 – 531.

7. Principe de stationarité et généralisation de la loi de Mendel. Ibidem, 581 – 584.

8. On a modification of the Chebyshev inequality and on the error of the Laplace formula. (1924).

9. Solution of a mathematical problem connected with the theory of heredity (1924).

10. On the economic barometer of the Conjuncture Institute. *Khoziastvo Ukrainy* 4, 1925, 12 – 22.

11. Sur les sommes de quantités dépendantes. IAN, 20, 1926, 1459 – 1478.

12. Sur les courbes de distribution des probabilités. *Math. Z.* 24, 1926, 199 – 211.

13. Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Math. Ann.* 97, 1926, 1 – 59. Cf. **40.**

14. The present state of the theory of probability and its applications (1927). Translated in this book.

15. On the application of a geometric principle to the correlation theory. *In Mem. Lobachevsky*. Kazan 2, 1927, 137 – 150.

16. Fondements géométriques de la théorie des corrélations. *Metron* 7:2, 1927, 3 – 27.

17. The concept of correlation between statistical magnitudes. *VS Ukrainy* 1, 1928, 111 – 113. (U)

18. Sur les sommes des quantités dépendantes. DAN, ser. A, 1928, 55 – 60. (Supplement to **11**).

19. Sur une propriété élémentaire du coefficient de corrélation. *Zapiski Kharkov Matematich. Tovarishchestvo* 5, 1932, 65 – 66.

20. Sur les liaisons entre les grandeurs aléatoires. *Verh. Math. Kongr. Zürich* 1, 1932, 288 – 309.

21. *Современное состояние теории вероятностей* (Present State of the Theory of Probability). M. – L., 1933.
22. Sur l'équation différentielle de Fokker – Planck. *C.r.* 196, 1933, 1062 – 1064.
23. Russian translation of **20** (1933).
24. On linear quasicontinuous Markov chains. DAN 1934.
25. On diffusion with absorption. Ibidem.
26. This is part 2 of **24.** Ibidem.
27. Principes de la théorie des équations différentielles stochastiques. *Trudy Fiziko-Matematich. Steklov Inst.* 5, 1934.
28. On the expected demurrage in complicated industrial processes (1935).
29. Détermination d'une limite inférieure de la dispersion des sommes de grandeurs liées en chaîne singulière. MS 1(43), 1936, 29 – 38.
30. On some modifications of the Chebyshev inequality. DAN 1937.
31. Équations différentielles stochastiques. *Acta Sci. Ind. 738. Conf. Intern. Sci. Math. Univ. Genève.* Théor. Prob, V. Paris, 1938, 5 – 31.
32. Some notes on the Liapunov limit theorem. DAN 1939.
33. Correcting a proof. DAN 25, 1939, 705 – 707.
34. A problem concerning urns with added balls. DAN 1940.
35. New applications of almost independent variables. IAN 1940.
36. On sums of dependent variables with an almost zero regression. DAN 1941.
37. On the Fisherian "fiducial" probabilities (1941).
38. On a property characterizing the Gauss law (1941).
39. Returning to the precision of the Laplace limit theorem. IAN 1943.
40. Russian translation of **13** (1944).
41. *Теория вероятностей* (Theory of probability). M. – L., 1946. Fourth edition.
42. On a limit theorem (1946).
Bernstein's *Собрание сочинений* (Coll. Works), vol. 4. N.p., 1964 contains reprints/translations of items 1, 2, 5, 8, 9, 11 – 14, 16, 18 – 20, 22, 24 – 32, 34 – 39 and 42. Items 1, 4, 9, and 37 are translated in *Probability and Statistics. Russian Papers of the Soviet Period.* Berlin, 2005.
Bobrov, A.A. 1. On the relative stability of sums of positive random variables. DAN 15, 1937, 239 – 240.
2. On the relative stability of sums in a particular case. MS 4 (46), 1938, 99 – 104.
3. Same title as **1.** *Uch. Zap. MGU* 15, 1939, 191 – 202.
4. Conditions of applicability of the strong law of large numbers. *Duke Math. J.* 12, 1945, 43 – 46.
Bogoiavlensky, I. 1. Patterns of balls in the theory of probability. *Zapiski Belorussk. AN* 12, 1930, 204 – 208.
Bogoliubov, N.N. 1. Method of functional data in statistical mechanics. *Zbornik Trudov Inst. Matematiki AN Ukrainian SSR* 8, 1947, 177 – 199. (U)
Boiarsky, A.Ya. 1. Correlation theory and differential calculus. *Trudy NI Sektora Inst. Narodno-Khoziastvenn. Ucheta* 1:2, 1934, 23 – 35.
2. On the probability of discovering a figure during mechanical trials on a plane. *Problemy Ucheta i Statistiki* 11 (5), 1937, 86 – 119.
3. On geometric correlation. IAN 5, 1941, 159 – 164.
Boiarsky, A.Ya., Starovsky, V.N., Khotimsky, V.I., Yastremsky, B.S.
1. *Теория математической статистики* (Theory of Math. Statistics). M., 1931, 2nd edition.
Brodovitsky, K.V. 1. Calculating probabilities of unobserved maxima (with brief indication on possible application in construction calculations). *Izvestia Dalnevostochn. Geofizich. Inst.* 2 (9), 1932, 213 – 239.

2. On necessary and sufficient conditions for prior probabilities to have sense. *Trudy SAGU*, ser. Math., 19, 1940, 1 – 8.
3. Formulating the problem of similarity in the theory of sampling. Ibidem, 20, 1940, 1 – 38.
Bukhman, E.N. 1. On the problem of congestion. *Prikladn. Matematika i Mekhanika* 3 – 4, 1939.
2. Problems of conjection in telephony. Ibidem 6, 1942, 247 – 256.
Chebanov,S.G. 1. On the subordination of the vocal structures of the Indo- European group of languages to the Poisson law. DAN 55, 1947, 103 – 106.
Chebotarev, A.S. 1. *Способ наименьших квадратов* и т.д. (Method of Least Squares with Elements of the Theory of Probability). M. – L., 1936. {Fourth edition, 1958. Unworthy.}
Chuprov, A.A. 1. *Основные проблемы теории корреляции* {*Principles of the Mathematical theory of Correlation*. London, 1939. In English. German edition, 1925. 2nd Russian edition, 1960.}
Davidov, B. 1. The Fokker – Planck equation in a phase space and the relaxation time of the Maxwell distribution. DAN 2, 1934, 212 – 219.
Derevitsky, N. 1. Rejection of some data in field experiments and their subsequent treatment. *Trudy SAGU*, ser. Math., 22, 1940, 1 – 21.
Dimanstein, Ya. 1. On the mathematical description of the movement of population. *Nauchn Zapiski*. (Kiev), 2, 1926, 115 – 131.
Dmitriev, N.A., Dynkin, E.B. 1. On the characteristic numbers of stochastic matrixes. DAN 49, 1945, 159 – 162.
2. Characteristic roots of stochastic matrixes. IAN 10, 1946, 167 – 194.
Dubrovsky, V.M. 1. A generalization of the theory of the Feller totally {rein} disconnected random processes. DAN 19, 1938, 439 – 444.
2. On a boundary value problem of the theory of probability. IAN 4, 1940, 411 – 416.
3. A study of totally disconnected random processes by the method of integro-differential equations. Ibidem 8, 1944, 107 – 128.
4. On totally disconnected random processes with aftereffect. DAN 47, 1945, 83 – 85.
5. On a problem linked with totally disconnected random processes. Ibidem, 475 – 477.
Egudin, G.I. 1. Parameters of the distribution of random variables invariant under transformation of translation, and algebraic semiinvariants. DAN 48, 1945, 643 – 645.
2. On an efficient method of calculating the expectations of central sample moments. DAN 53, 1946, 491 – 494.
3. On the stability of a very wide class of statistics. DAN 57, 1947, 115 – 117.
4. Some relations between the moments of the distribution of the extreme values in a random sample. DAN 58, 1947, 1581 – 1584.
Gelfond, A.O., Khinchin, A.Ya. 1. The Gram determinants for stationary series. *Uch. Zap. MGU*, 1, 1933, 3 – 5.
Gelvikh, P.A. 1. On a wrong derivation in the theory of probability. *Izvestia L. Voennotekhnich. Akad.* 1, 1927, 128 – 140.
Gerasimovich, B. 1. Probability problems connected with the discovery of variable stars in a photographic way. DAN (A), 1931, 93 – 100. In English.
Giliarovsky, N.I. 1. Justification of the theory of observational errors. *Trudy L. Inst. Inzhenerov Zheleznodorozhn. Transport*, 9, 1934, 113.
Gikhman, G.I. 1. On a pattern in the formation of random processes. DAN 58, 1947, 961 – 964.
Ginsburg, G.M. 1. On the limiting laws of distribution in stochastic processes. *Zapiski Kharkov Matematich. Tovarishchestvo* (4) 17, 1940, 65 – 74.

2. On the sufficient conditions for the uniqueness of limiting distributions. DAN 30, 1941, 293 – 295.
Glivenko, V.I. 1. Sur les valeurs probables de fonctions. *Atti Accad. Naz. Linzei* 8, 1928, 480 – 483.
2. Sur la loi des grands nombres dans les espaces fonctionels. Ibidem, 673 – 676.
3. Sur quelques formes générales de la loi des grands nombres. Ibidem, 9, 1929, 466 – 469.
4. Sulla determinazione empirica della leggi di probabilità. GIIA 4, 1933, 1 – 10.
5. *Теория вероятностей* (Theory of Probability).M., 1937.
6. *Курс теории вероятностей* (Course in Theory of Probability). M. – L., 1939.
Gnedenko, B.V. 1. Calculation of the mean time of servicing a loom before an adjustment of another loom becomes necessary. *Bull. Ivanovsk. NI Tekstiln. Inst.* 1 – 2, 1934, 117 – 122.
2. Same title. Ibidem, 11 – 12, 1934, 118 – 122.
3. On the mean demurrage of lathes when each worker is servicing a large number of them. *Izvestia Khlopchatobumazhn. Promyshlennost* 11, 1934, 15 – 18.
4. On a characteristic property of infinitely divisible laws of distribution. *Bull. MGU* (A), 1:5, 1937, 10 – 16.
5. On characteristic functions. Ibidem, 17 – 18.
6. On the convergence of the laws of distribution of sums of independent terms. DAN 18, 1938, 231 – 234.
7. On limit theorems for sums of independent terms. DAN 22, 1939, 61 – 64.
8. On limit theorems of the theory of probability. DAN 23, 1939, 868 – 871.
9. On domains of attraction of stable laws. DAN 24, 1939, 640 – 642.
10. On the theory of limit theorems for sums of independent random variables. IAN 1939, 181 – 232.
11. Corrections to **10.** Ibidem, 643 – 647.
12. Review of the contemporary state of the theory of limit theorems for sums of independent terms. *Uch. Zap. Tiumensk. Pedagogich. Inst.* 1, 1939, 5 – 28. Translated in *Probability and Statistics. Russian Papers.* Berlin, 2004.
13. On the theory of the domains of attraction of stable laws. *Uch. Zap. MGU* 30, 1939, 61 – 82.
14. A few theorems on the powers of distribution functions. Ibidem 45, 1940, 61 – 72.
15. Limit theorems for the maximal term of a variational series. DAN 32, 1941, 7 – 9.
16. On the theory of Geiger – Müller counters. *Zhurnal Eksperimentaln. i Teoretich. Fiziki* 11, 1941, 101 – 106.
17. On locally stable laws of distribution. DAN 35, 1942, 295 – 298.
18. A study of the growth of homogeneous random processes with independent increments. DAN 36, 1942, 3 – 4.
19. Locally stable laws of distribution. IAN 6, 1942, 291 – 308.
20. On the growth of homogeneous random processes with independent increments of the same kind. DAN 40, 1943, 103 – 107.
21. On the law of the iterated logarithm for homogeneous random processes with independent increments. Ibidem, 291 – 293.
22. On the growth of homogeneous random processes with independent increments. IAN 7, 1943, 89 – 110.
23. Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* 44, 1943, 423 – 453.
24. Limit theorems for sums of independent random variables. UMN 10, 1944, 115 – 165.
25. Elements of the theory of distribution functions of random vectors. Ibidem, 230 – 244.
26. On ellipsoids of concentration. *Uch. Zap. Lvov Univ.* 5:2, 1947, 116 – 120. (U)
27. On functions of random variables. Ibidem, pp. 121 – 128. (U)

Gnedenko, B.V., Groshev, A.V. 1. On the convergence of the laws of distribution of normed sums of independent random variables. MS 6(48), 1939, 521 – 541.

Godnev, I.N. 1. On the substantiation of the classical quantum statistics. *Uch. Zap. Ivanovsk. Pedagogich. Inst.*, Fiziko-Matematich. Fak. 1:1, 1941, 42 – 46.

Goncharov, V.L. 1. *Теория вероятностей* (Theory of Probability). M. – L., 1939.

2. On the distribution of cycles in permutations. DAN 35, 1942, 299 – 301.

3. On the alternation of events in a series of independent Bernoulli trials. DAN 38, 1943, 295 – 297.

4. From combinatorial analysis. IAN 8, 1944, 3 – 48.

Gorin, N.P. 1. Some remarks on geometric probabilities. *Izvestia Uralsk. Lesotekhnich. Inst*. 3, 1934, 58 – 64.

Groshev, A.V. 1. The domain of attraction of the Poisson law. IAN 5, 1941, 165 – 172.

Idelson, N.I. 1. *Способ наименьших квадратов* (Method of Least squares). M., 1947. Third edition.

Ivanov, I.I. 1. *Способ наименьших квадратов* и т.д. (Method of Least Squares and Elements of the Theory of probability). M., 1927. Third edition.

Kan, V.L. 1. On the application of the Laplace theorem for the case of a finite number of trials. *Trudy L. Politekhnich. Inst*. 3, 1941, 39 – 43.

Kantorovich, L.V. 1. Sur un problème de Steinhaus. *Fund. Math.* 14, 1929, 266 – 270.

2. Some considerations about the layout of minefields in connection with calculating the probabilities of hits. Trudy VVMISU {?} 6, 1944.

3. A method of calculating the vitality of systems of objects. Report at conf. at VITU VMS, abstract, 1946. {VMS likely means Voenno-Morskie Sily, Navy.}

4. *Теория вероятностей* (Theory of Probability). L., 1946.

Kazansky, A. 1. On a particular case of a limit theorem of the theory of probability. *Zapiski L. Gorn. Inst*. 8, 1934, 236 – 244.

Keller, L.V. 1. On generalizing the limit theorems of the theory of probability to integrals and mean values of functions of continuous arguments. *Trudy Glavn. Geofizich. Observatorii* 4, 1935, 5 – 20.

Khinchin, A.Ya. 1. Über dyadische Bruche. *Math. Z.* 18, 1923, 109.

2. Über einen Satz der Wahrscheinlichkeitsrechnung. *Fund. Math.* 6, 1924, 9 – 20.

3. Sur un théorème général relatif aux probabilités dénombrables. *C.r.* 178, 1924, 617 – 618.

4. On the Petersburg game. MS 32, 1925, 330 – 341.

5. Über die Anwendbarkeitsgrenzen des Tschebycheffschen Satzes in der Wahrscheinlichkeitsrechnung. Ibidem, 678 – 688.

6. *Основные законы теории вероятностей* (Main Laws of the Theory of Probability). M., 1927.

7. Über das Gesetz der großen Zahlen. *Math. Ann.* 96, 1927, 152 – 168.

8. Begründung der Normalkorrelation nach der Lindebergschen Methode. *Izvestia Assoziatsia Inst. MGU* 1 – 2, 1928, 37 – 45.

9. Über die Stabilität zweidimensionaler Verteilungsgesetze. MS 35, 1928, 19 – 23.

10. Sur la loi des grands nombres. *C.r.* 186, 1928, 285 – 287.

11. The strong law of large numbers and its meaning for mathematical statistics. VS No. 1, 1928, 123 – 128.

12. Über ein Problem der Wahrscheinlichkeitsrechnung. *Math. Z.* 29, 1928 – 1929, 746 – 751.

13. The Mises doctrine of probability and the principles of physical statistics. *Uspekhi Fizichesk. Nauk* 9, 1929, 141 – 166.

14. Über die positiven und negativen Abweichungen des arithmetischen Mittels. *Math. Ann.* 101, 1929, 381 – 385.

15. Über einen neuen Grenzwertsatz der Wahrscheinlichkeitsrechnung. Ibidem, 745 – 752.
16. Über Anwendungskriterien für das Gesetz der großen Zahlen. MS 36, 1929, 78 – 80.
17. Sur la loi des grands nombres. *C.r.* 188, 1929, 477 – 479.
18. Sur une généralisation de quelques formules classiques. Ibidem, 532 – 534.
19. On the mean demurrage of lathes. MS 40, 1930 {1933}, 119 – 123.
20. Die Maxwell – Boltzmannsche Energieverteilung als Grenzwertsatz der Wahrscheinlichkeitsrechnung. *Trudy Seminara MGU Veroiatn. i Statistiki* 1, 1930, 1 – 11.
21. Second edition of **6.** M. – L., 1932.
22. Sur les classes d'événements équivalents. MS 39:3, 1932, 40 – 43.
23. Remarques sur les suites d'événements obéissant à la loi des grands nombres. Ibidem, 115 – 119.
24. Mathematical theory of a stationary queue. Ibidem, No. 4, 73 – 84.
25. Zu Birkhoffs Lösung des Ergodenproblems. *Math. Ann.* 107, 1932, 485 – 488.
26. Sulla successioni stazionari di eventi. GIIA 4, 1933.
27. Zur mathematischen Begründung der statistischen Mechanik. *Z. angew. Math. Mech.* 13, 1933, 101 – 103.
28. Über stationäre Reihen zufälliger Variablen. MS 40, 1933, 124 – 128.
29. The method of spectral reduction in classical dynamics. *Trans. Amer. Math. Soc.* 19, 1933, 567 – 573.
30. Eine Verschärfung des Poincaréschen Wiederkehrsatzes. *Comp. Math.* 1, 1934, 177 – 179.
31. Zur mathematischen Begründung der Maxwell – Boltzmannschen Energieverteilung. *Uch. Zap. MGU* 2:2, 1934, 35 – 38.
32. Korrelationstheorie der stationären stochastischen Prozesse. *Math. Ann.* 109, 1934, 604 – 615.
33. *Асимптотические законы теории вероятностей* (Asymptotic Laws of the Theory of Probability). M. – L., 1936. {Originally published in German (Berlin, 1933).}
34. Su una legge dei grand numeri generalizzata. GIIA 7, 1936, 365 – 377.
35. Über Klassenkonvergenz von Verteilungsgesetzen. *Izvestia NI IMM Tomsk. Univ.* 1, 1937, 261 – 265.
36. A new derivation of a formula due to Lévy. *Bull. MGU* (A), 1:1, 1937, 1 – 5.
37. On the arithmetic of the laws of distribution. Ibidem, 6 – 17.
38. On a test for characteristic functions. Ibidem No. 5, 1 – 3.
39. Invariant classes of laws of distribution. Ibidem, 4 – 5.
40. Examples of random variables obeying stable laws of distribution. Ibidem, 6 – 9.
41. Zur Theorie der unbeschränkt teilbaren Verteilungsgesetze. MS 2(44), 1937, 79 – 120.
42. *Предельные законы для сумм независимых случайных величин* (Limit Laws for Sums of Independent Random Variables). M. – L., 1938.
43. Theory of fading spontaneous effects. IAN 1938, 313 – 322.
44. Two theorems on stochastic processes with increments of same type. MS 3 (45), 1938, 577 – 584.
45. Zur Methode der willkürlichen Funktionen. Ibidem, 585 – 590.
46. On unimodal distributions. *Izvestia NI IMM Tomsk. Univ.* 2:2, 1938, 1 – 7.
47. Theory of correlation of stationary stochastic processes. UMN 5, 1938, 42 – 51.
48. On the local growth of homogeneous stochastic processes without aftereffect. IAN 1939, 487 – 508.
49. The mean values of sum functions in statistical mechanics. DAN 31, 1941, p. 442.
50. On the analytical methods of statistical mechanics. DAN 33, 1941, p. 438.
51. The laws of distribution of sum functions in statistical mechanics. DAN 34, 1942, 61 – 62.

52. *Математические основания статистической механики* (Mathematical Principles of statistical mechanics). M. – L., 1943. {English translation: New York, 1949.}

53. Convex functions and evolutionary theorems of statistical mechanics. IAN 7, 1943, 111 – 112.

54. On an ergodic problem of quantum mechanics. Ibidem, p. 167.

55. Sur une cas de corrélation à posteriori. MS 12 (54), 1943, 185 – 196.

Khinchin, A.Ya., Kolmogorov, A.N. 1. Über Konvergenz von Reihen, deren Glieder durch den Zufall bestimmt werden. MS 32, 1925, 668 – 677.

Khinchin, A.Ya., Lévy, P. 1. Sur les lois stables. *C.r.* 202, 1936, 374 – 376.

Khotimsky, V.I. 1. *Выравнивание статистических рядов* и т.д. (Adjusting Statistical Series by the Method of Least Squares and Tables for Determining the Equations of Parabolic Curves). M. – L., 1925. {Second edition, M., 1959.}

Kletenik, D.I. 1. On a generalization of the De Moivre – Laplace theorem to events with three incompatible outcomes. *Юбилейный сборник научн. работ* (Jubilee Coll. Scient. Papers) Rostov/Don Mashinostroiteln. Inst. 1, 1940, 77 – 86.

Kolmogorov, A.N. 1. Sur la loi des grands nombres. *C.r.* 185, 1927, 919 – 921.

2. Über die Summen durch den Zufall bestimmter unabhängiger Großen. *Math. Ann.* 99, 1928, 309 – 319.

3. Sur une formule limite de A. Khintchine. *C.r.* 186, 1928, 824 – 825.

4. Über das Gesetz des iterierten Logarithmus. *Math. Ann.* 101, 1929, 126 – 135.

5. Bemerkungen zu **2.** Ibidem, 102, 1929, 484 – 488.

6. Same title as **1.** *Atti Accad. Naz. Lincei* 9, 1929, 470 – 474.

7. The general measure theory and the calculus of probability (1929).

8. Sur la loi forte des grands nombres. *C.r.* 191, 1930, 910 – 912.

9. Sur la notion de la moyenne. *Atti Accad. Naz. Lincei* 12, 1930, 388 – 391.

10. Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Math. Ann* 104, 1930, 415 – 458.

11. Sur la problème d'attente. MS 38, 1 – 2, 1931, 101 – 106.

12. The method of median in the theory of errors. Ibidem 3 – 4, 47 – 50.

13. Eine Verallgemeinerung des Laplace – Liapunoffschen Satzes. IAN, ser. Fiz. – Mat, 1931, 959 – 962.

14. Sulla forma generale di un processo stocastico omogeneo. *Atti Accad. Naz. Lincei* 15, 1931, 805 – 808.

15. Ancora sulla forma generale di un processo stocastico omogeneo. Ibidem, 866 – 869.

16. Sulla determinazione empirico di una legge di distribuzione. GIIA 4, 1933, 83 – 91.

17. Sur la détermination empirique d'une loi de distribution. *Uch. Zap. MGU* 1, 1933, 9 – 10.

18. Über die Grenzwertsätze der Wahrscheinlichkeitsrechnung. IAN, ser. Fiz – Mat, 1933, 363 – 372.

19. Zur Theorie der stetigen zufälligen Prozesse. *Math. Ann.* 108, 1933, 149 – 160.

20. On the suitability of statistically determined formulas of prediction (1933).

21. Zufällige Bewegungen. *Ann. Math.* 35, 1934, 116 – 117.

22. Zur Theorie der Markoffschen Ketten. *Math. Ann.* 112, 1935, 155 – 160.

23. Deviation from the Hardy formula under partial isolation. DAN 3, 1935, 129 – 132.

24. *Основные понятия теории вероятностей* (Main Notions of the Theory of Probability). M. – L., 1936. Translated from German (1933).

25. On the Plesner condition of the law of large numbers. MS 1(43), 1936, 847 – 850.

26. Anfangsgründe der Theorie der Markoffschen Ketten mit unendlich vielen möglichen Zuständen. Ibidem, 607 – 610.

27. On some contemporary trends in the theory of probability (1936). Translated in this book.

28. On the statistical theory of crystallization of metals. IAN 1937, 355 – 360.
29. Markov chains with a countable number of possible states (1937).
30. Ein vereinfachter Beweis des Birkhoff – Khintchinschen Ergodensatzes. MS 2 (44), 1937, 367 – 368.
31. Zur Umkehrbarkeit der statistischen Naturgesetze. *Math. Ann*. 113, 1937, 766 – 772.
32. This is a Russian translation of **10** (1938).
33. Same title as **30**. UMN 5, 1938, 52 – 56.
34. On the solution of a biological problem (1938).
35. Sur l'interpolation et extrapolation des suites stationaires. *C.r.* 208, 1937, 2043 – 2045.
36. On a new confirmation of Mendel's laws. DAN 27, 1940, 38 – 42.
37. Curves in a Hilbert space invariant under a one-parameter group of motions. DAN 26, 1940, 6 – 9.
38. Wiener spirals and some other interesting curves in a Hilbert space. Ibidem, 115 – 118.
39. Stationary sequences in a Hilbert space (1941).
40. On a lognormal distribution of the sizes of particles under grinding. DAN 31, 1941, 99 – 101.
41. Interpolation and extrapolation of stationary random sequences. IAN 5, 1941, 3 – 14. {Cf. **35.**}
42. Confidence limits for an unknown distribution function. *Ann. Math*. Stat. 12, 1941, 461 – 463. In English.
43. Determining the center of scattering and the measures of precision by a restricted number of observations. IAN 6, 1942, 3 – 32.
44. Number of hits after several shots and the general principles of estimating the efficiency of firing. *Trudy Steklov Matematich. Inst*. 12, 1945, 7 – 25.
45. Artificial scattering in the case of one hit and scattering in one dimension. Ibidem, 26 – 45.
46. On the substantiation of the method of least squares (1946).
47. Statistical theory of oscillations with a continuous spectrum. *Юбилейный сборник посвященный 30-летию Великой октябрьской социалистической революции* (Jubilee Coll. Articles Devoted to the 30th Anniversary of Great October Socialist Revolution), vol. 1, 1947, 242 – 254.
Kolmogorov's *Selected Works*, vols 1 – 2. Dordrecht, 1991 – 1992, contain (translations of) Items 9, 33, 37, 38 (vol. 1) and 1 – 8, 10 – 16, 18 – 23, 28, 29, 31, 34, 36, 39 – 41 and 46 (vol. 2). All of his joint works listed below are also contained there, in *Sel.* 1 or *Sel.* 2. Items 43 and 44 are translated in *Probability and Statistics. Russian Papers of the Soviet Period.* Berlin, 2005.
Kolmogorov, A.N:, Dmitriev, N.A. 1. Branching random processes. DAN 56, 1947, 7 – 10. (*Sel.* 2.)
Kolmogorov, A.N., Leontovich, M.A. 1. Zur Berechnung der mittleren Brownschen Fläche. *Fizich. Zhurnal* 4, 1933, 1 – 13. (Sel. 2.)
Kolmogorov, A.N., Sevastianov, B.A. 1. Calculation of the final probabilities for branching random processes. DAN 56, 1947, 783 – 786. (*Sel.* 2.)
Kolmogorov, A.N., Petrov, A.A., Smirnov, Yu.M. 1. A Gauss formula from the method of least squares. IAN 11, 1947, 561 – 566. (*Sel.* 2.)
Kolmogorov, A.N., Petrovsky, I.G., Piskunov, N.S. 1. A study of the equation of diffusion with an increasing quantity of the matter; its application to a biological problem (1937). (*Sel.* 1.)
Komarov, V.N. 1. On the second and the third central statistical moments as measures of the variance and the skewness. *Uch. Zap. L. Pedagogich. Inst*. 5, 1937, 49 – 56.
Konius, A.A. 1. On the theory of mean magnitudes. *Trudy SAGU*, ser. Math., 24, 1940, 1 – 10.

Kozuliaev, P.A. 1. Sur la répartition de la partie fractionnaire d'une variable aléatoire. MS 2 (44), 1937, 1017 – 1019.
2. An asymptotical analysis of a main formula of the theory of probability. *Uch. Zap. MGU* 15, 1939, 179 – 182
3. On the problems of interpolating and extrapolating stationary sequences. DAN 30, 1941, 13 – 17.
4. On extrapolating stationary random processes. DAN 56, 1947, 903 – 906.
Kravchuk, M.F. 1. On the method of moments in mathematical statistics. *Zapiski S.*{Selsko ?}*-Gospodarsk. Inst. Kiev* 2, 1927, 83 – 95. (U)
Kravchuk, M.F., Okonenko, A.A. 1. On the normal law of distribution for two variable indications. Ibidem, 1, 1926, 95 – 99. (U)
Krein, M.G. 1. On Kolmogorov's extrapolational problem. DAN 46, 1945, 339 – 342.
Krutkov, D. 1. Expectation of an entire function. *Uch. Zap. Kazan Univ*. 2, 1930, 147 – 149.
Krylov, N.M., Bogoliubov, N.N. 1. Sur les propriétés ergodiques le l'équation de Smoluchowski. *Bull. Soc. Math. France* 64, 1936, 49 – 56.
2. Sur les probabilités en chaîne. *C.r.* 204, 1937, 1386 – 1388.
3. Les propriétés ergodiques des suites des probabilités en chaîne. Ibidem, 1454 – 1456.
4. Consequences of the statistical {statical?} change of parameters with regard to the ergodic properties of dynamical non-conservative systems. *Zapiski Kafedry matematich. fiziki AN Ukr SSR* 3, 1937, 153 – 189. (U)
5. On the Fokker – Planck equations derived in the theory of perturbations by a method based on the spectral properties of the perturbational Hamiltonian. Ibidem, 4, 1939, 5 – 158. (U)
6. On some problems of the ergodic theory of stochastic systems. Ibidem, 243 – 287. (U)
Krylov, V. 1. Serial samples. *Trudy SAGU*, ser. Math., 25, 1940, 1 – 24.
Kuzmin, R.O. 1. On the Liapunov method in the theory of probability. *Trudy L. Inst. Inzhenerov Promyshlenn. Stroitelstva* 2, 1934, 49 – 64.
2. On the law of distribution of the correlation coefficient in samples from a normal population. DAN 22, 1939, 302 – 305.
Kuznetsov, E. 1. On the coefficients of Prof. Egiz' parallel correlation. *Zhurnal Opytn. Agronomii* 6, 1928, 175 – 185.
2. The law of distribution of a random vector. DAN 2, 1935, 187 – 193.
Labutin, D.N. 1. A continuous Markov chain. *Uch. Zap. L. Pedagogich. Inst.* 1937, 67 – 72.
2. On a simple Markov chain. Ibidem, 28, 1939, 171 – 178.
3. On the generating function. *Uch. Zap. LGU*, ser. Math., 10, 1940, 139 – 147.
Lagunov, B.I. 1. On the practice of adjusting statistical series. *Kiev, Okruzhnoe Statistich. Buro*, 1927, 22. German version: *Metron* 6, 1926, 3 – 23.
2. Orthogonal linear functions in the method of least squares. *Zhurnal Matematich. Zikly AN Ukr. SSR* 1, 1932, 79 – 82. (U)
Lakhtin, L.K. 1. *Кривые распределения* и т. д. (Curves of Distribution and the Construction of Interpolational Formulas for Them by the Methods of Pearson and Bruns).M., 1922.
2. *Курс теории вероятностей* (Course in the Theory of Probability). Petrograd, 1924.
Latysheva, K.Ya. 1. On the Gauss law of distribution for functions of two variables. *Zhurnal Inst. Matematiki AN Ukr. SSR* 2, 1934, 113 – 119. (U)
Levinsky, V.P. 1. *Краткий курс вариационной статистики* (Short Course in Math. Statistics). M., 1935.
2. On summary characteristics of statistical populations. *Trudy SAGU*, ser. Math., 26, 1940, 1 – 17.

Liapin, N.M. 1. Deriving a formula for the mean random variation of the diurnal rate of the chronometer. *Uch. Zapiski Odessa Vyssh. Shkoly* 1, 1921, 9 – 14.
2. On a main property of random errors. Ibidem, 15 – 18.
Linnik, Yu.V. 1. On the precision of the approximation of sums of independent random variables to the Gauss distribution. DAN 55, 1947, 575 – 578.
2. Same title. IAN 11, 1947, 111 – 138.
Loshchinin, P.E. 1. A variety of an urn problem. *Sbornik Rabot Bukhara Pedagogich. Inst*. 1938, 69 – 72.
Lukomsky, Ya.I. 1. Application of the theory of means to a random sample. *Problemy Ucheta i Statistiki* 11 (5), 1937, 146 – 153.
Lurie, A.L. 1. The direct, the inverse, and the unconditional laws of large numbers. DAN 49, 1945, 566 – 569.
2. On the inverse Bernoulli theorem. DAN 50, 1945, 45 – 48.
Lvov, N.N. 1. On the probable error of a curve constructed by means of scattered points. *Zapiski Saratov Plan. Inst*. 7, 1940, 197 – 203.
Maltsev, A.I. 1. A remark to Kolmogorov, Petrov & Smirnov **1.** IAN 11, 1947, 567 – 568.
Markov, A.A. 1. On some limiting formulas of the calculus of probability. IAN 11, 1917, 177 – 186.
2. Generalization of the problem of a consecutive exchange of balls. IAN 12, 1918, 261 – 266.
3. A few problems from the calculus of probability. Ibidem, 2101 – 2116.
4. On the coefficient of dispersion, second note. IAN 14, 1920, 191 – 198.
5. The difficulty of the method of moments and two examples of incompletely surmounting it. IAN, 16, 1922, 281 – 286.
6. *Исчисление вероятностей* (Calculus of Probability). M., 1924. Fourth edition.
7. On ellipsoids (ellipses) of concentration and correlation. IAN, 18, 1924, 117 – 126.
Markov's *Избранные труды* (Sel. Works). N.p., 1951, contain reprints of Items 1, 2 and 4.
Matusevich, N.N. 1. On a formula of the theory of errors. *Zapiski po Gidrografii* 2, 1932, 43 – 48.
Mechnikov, V.V. 1. On the probability of hits when shooting at a moving target. *Izvestia L. Voennotekhnich. Akad.* 1, 1927, 114 – 127.
Mitropolsky, A.K. 1. *Основы статистики* (Fundamentals of Statistics), vol. 1. L., 1925.
2. On establishing correlation equations by the Chebyshev method. IAN 1937, 125 – 138.
3. On working out correlation equations having a small number of trials. *Trudy L. Lesotekhnich. Akad*. 48, 1937, 3 – 48.
4. On multiple non-linear correlation equations. IAN 1939, 399 – 406.
5. On working out correlation equations by the method of sums. *Trudy M. Lesotekhnich. Akad*. 60, 1947, 63 – 72.
Mordukhai-Boltovkoy, D.D. 1. On Cesaro's problem concerning the calculus of probability. *Izvestia Rostov/Don Pedagogich. Inst*. 9, 1938, 24 – 26.
Nalbaldian, Ya.A. 1. A generalization of the Buffon problem. *Uch. Zap. Rostov/Don Univ*. 8, 1936, 1644 – 1669.
Natanson, I.P. 1. On a limit theorem of the theory of probability. *Trudy L. Inst. Tochn. Mekhaniki i Optiki* 1:2, 1940, 109 – 111.
Nemchinov, V.S. 1. *Полиномы Чебышева* и т.д. (Chebyshev Polynomials and Mathematical Statistics). M., 1946.
Novikov, V.S. 1. On some properties of the formula for the aggregate index. *Problemy Ucheta i Statistiki* 11 (5), 1937, 140 – 145.
Nuvariev, V.S. 1. On substantiating the method of least squares. *Izvestia Tomsk. Politekhnich. Inst*. 62:2, 1945, 201 – 212.

Obukhov, A.M. 1. Normal correlation of vectors. IAN 1938, 339 – 370.
2. Theory of correlation of vectors. *Uch. Zap. MGU* 45, 1940, 73 – 92.
3. On dissipation of sound in a turbulent current. DAN 30, 1941, 611 – 614.
4. On distribution of energy in a spectrum of a turbulent current. DAN 32, 1941, 22 – 24.
5. *Приложение методов* и т. д. (Application of Methods of Statistical Description of Continuous Fields to the Theory of Atmospheric Turbulence). M., 1947. A thesis.
Ogorodnikov, K.F. 1. Sur le théorème limite de théorie des erreurs d'observation. IAN, ser. Fiz.-Math., 1931, 1 – 21.
Omshansky, M.A. 1. On the variance of the singularities of a random series without dependences. *Trudy Glavn. Geofizich. Observatorii* 10, 1936, 112 – 116.
Pankin, A.V. 1. Application of the theory of random errors to the treatment of time-studies. *Vestnik Metallopromyshlennosti* 5, 1931, 69 – 78.
Parfentiev, N.N. 1. Application of the theory of probability for obtaining a precise comparative estimate of observed phenomena when the scale for measuring is lacking. *Izvestia Kazan Fiz.-Matematich. Obshchestvo* (2), 23, 1923, 21 – 23.
2. La déduction asymptotique de la loi de Bernoulli dans la théorie des probabilités. *Uch. Zap. Kazan Univ*. 87, 1927, 90 – 91.
Persidsky, K.P. 1. On Markov's theorem. *Izvestia Kazan Fiz.-Matematich. Obshchestvo* 4, 1929 – 1930, 37 – 40.
2. On the main theorem of the theory of probability. IAN, ser. Fiz.-Mat., 1932, 639 – 656.
3. On a theorem of the theory of probability. *Uch. Zap. Kazan Univ*., Math., 4:1, 1932, 37 – 42.
4. On limit theorems. Ibidem, 43 – 48.
5. A few remarks on the law of large numbers. Ibidem, 49 – 54.
6. On the law of large numbers. DAN 18, 1938, 81 – 84.
Petrenko, A.I. 1. On checking the solution of normal equations. *Nauchn. Zapiski Voronezh Selskokhoziastven Inst*. 3 (18), 1936, 139 – 149.
Petrovsky, I.G. 1. Über das Irrfahrtproblem. *Math. Ann.* 109, 1934, 425 – 444.
2. Zur ersten Randwertaufgabe der Wärmeleitungsgleichung. *Comp. Math.* 1, 1935, 383 – 419.
Plesner, A.I. 1. Über das Gesetz der großen Zahlen. MS 1 (43), 1936, 165 – 168.
Pomerantseva-Iliinskaia, E.N. 1. On estimating the error of the mean having a small number of observations. *Zhurnal Geofiziki* 6:1 (19), 1936, 34 – 49.
Popov, V.S. 1. On determining a straight line given the measured coordinates of its points. *Zapiski Belorussk. AN* 2, 1930, 79 – 80.
Raikov, D.A. 1. On the decomposition of the Poisson law. DAN 14, 1937, 9 – 12.
2. On the decomposition of the Poisson and Gauss laws. IAN 1938, 91 – 124.
3. On the connection between the central limit theorem and the law of large numbers. Ibidem, 323 – 338.
4. A theorem from the theory of analytic characteristic functions. *Izvestia NI IMM Tomsk Univ.* 2:2, 1938, 8 – 11.
5. On the arranging of analytic distribution functions. DAN 23, 1939,

511 – 514.
Romanov, N.A. 1. On the possibility of a connection between the theory of probability and the Pavlov doctrine of conditioned reflexes. DAN 1, 1935, 193 – 201.
Romanovsky, V.I. 1. Statistical Weltanschauung. *Voenn. Mysl* 1, 1921, 59 – 76.
2. On the correlation ratio. VS 9 – 12, 1922, 29 – 33.
3. On the probabilities of connected indications and on their application in statistics. Ibidem, 34 – 41.
4. On a generalization of the Markov inequality. *Bull. SAGU* 8, 1925, 107 – 111.
5. On the distribution of sums of normally distributed variables. Ibidem, 9, 1925, 89 – 94.
6. A new proof of the Poisson theorem. Ibidem, 95 – 101.
7. On statistical tests of the given individual belonging to some of close species. *Trudy Turkestan Nauchn. Obshchestvo SAGU* 2, 1925, 173 – 184.
8. Sur la distribution des écarts quadratiques moyens dans les observations sur les quantités à distribution normale. *C.r.* 180, 1925, 1320 – 1323.
9. Généralisations d'un inégalité de Markoff. Ibidem, 1468 – 1470.
10. Sur certaines espérances mathématiques et sur l'erreur moyenne du coefficient de corrélation. Ibidem, 1897 – 1899.
11. On the distribution of the regression coefficient in samples from normal populations. IAN, 20, 1926, 643 – 648.
12. On the distribution of the arithmetic mean in series of independent trials. Ibidem, 1087 – 1106.
13. On the moments of standard deviations and correlation coefficient in samples from a normal population. *Metron* 5, 1926, 3 – 46.
14. A brief proof of the Pearson formula for the moments of the hypergeometric series. *Bull. SAGU* 12, 1926, 127 – 129.
15. On the central moments of two normally distributed random variables. Ibidem 15, 1927, 307 – 312.
16. On a problem due to Fisher. Ibidem, 313 – 317.
17. *Элементы теории корреляции* (Elements of the Correlation Theory). Tashkent, 1928. Second edition.
18. On statistical tests of belonging to a group. DAN, A, 1928, 347 – 352.
19. On the criteria that two given samples belong to the same normal population. *Metron* 7, 1928, 3 – 46.
20. Sur la généralisation des courbes de Pearson. *Atti del Congr. Intern. dei Matem. Bologna 1928*. Bologna, 1932, t. 6, 107 – 110.
21. On the moments of means of functions of one and more random variables. *Metron* 8, 1929, 251 – 290.
22. Sur la loi de probabilité de frequences assujetties aux conditions linéares et le criterium χ^2 de Pearson. DAN (A), No. 4 1929, 63 – 105.
23. Sur les chaînes de Markoff. Ibidem, No. 9, 203 – 208.
24. Sur une extension du théorème de Liapunoff sur la limite de probabilité. IAN, fiz.-mat.,1929, 209 – 225.
25. Sur un théorème limite du calcul des probabilités. MS 36, 1929, 36 – 64.
26. Sur les probabilités à posteriori. *C.r.* 189, 1929, 515 – 517.
27. Sur les chaînes biconnexes continues de Markoff. Ibidem 190, 1930, 695 – 697.
28. Sur les chaînes discrètes de Markoff. Ibidem 191, 1930, 450 – 452.
29. Sur les zéros de matrices stochastiques. Ibidem 192, 1931, 266 – 269.

30. Généralisations d'un théorème de E. Slutsky. Ibidem, 718 – 721.
31. Sulle regressioni multiple. GIIA 2, 1931, 161 – 171.
32. Sulle probabilità à posteriori. Ibidem, 493 – 511.
33. Sur la loi sinusoidale limite. *Rend. Circ. Mat. Palermo* 56, 1932, 82 – 111.
34. Sur une généralisation de la loi sinusoidale limite. Ibidem 57, 1933, 130 – 136.
35. Recherches sur les chaînes de Markoff. *Acta Math.* 66, 1936, 147 – 251.
36. Analytical inequalities and statistical tests. IAN 38, 1938, 457 – 474.
37. *Математическая статистика* (Mathematical Statistics). M. – L., 1938.
38. *Элементарный курс* и т.д. (Elementary Course in Mathematical Statistics). M. – L., 1939.
39. On inductive conclusions in statistics. DAN 27, 1940, 419 – 421.
40. Chain connections and tests of randomness. *Izvestia Uzbek FAN* 7, 1940, 38 – 50.
41. On discovering an unknown probability. Ibidem 8, 1940, 22 – 23.
42. Statistical problems connected with Markov chains. *Trudy FAN*, 1940.
43. On the Markov method of establishing limit theorems for chain processes. *Izvestia FAN* 3, 1941, 3 – 8.
44. Same title as **39.** *Trudy Uzbek FAN*, ser. Math., 3, 1942, 3 – 22.
45. On some theorems concerning the method of least squares. DAN 51, 1946, 259 – 262.
46. On the Pearson curves. *Trudy IMM AN Uzbek SSR*, No. 1, 1946, 3 – 13.
47. *Применение математической статистики* и т.д. (Application of Mathematical Statistics in Experimentation). M. – L., 1947.
48. *Основные задачи* и т.д. (Main Problems of the Theory of Errors). M. – L., 1947.
Romanovsky's *Избранные труды* (Sel. Works), vol. 2. Tashkent, 1964, contain reprints/translations of Items 8, 11, 13, 21 and 33.
Sapogov, N.A. 1. On singular Markov chains. DAN 58, 1947, 193 – 196.
2. The Laplace – Liapunov limit theorem for singular Markov chains. Ibidem, 1905 – 1908.
Sarmanov, O.V. 1. On isogeneous correlation. DAN 32, 1941, 16 – 18.
2. Same title. IAN 9, 1945, 169 – 200.
3. On monotone solutions of correlation integral equations. DAN 53, 1946, 781 – 784.
4. On rectifying symmetric correlation. DAN 58, 1947, 745 – 748.
5. A generalization of the limit theorem of the theory of probability to sums of almost independent variables satisfying the Lindeberg condition. IAN 11, 1947, 569 – 575.
Sarymasakov, T.A. 1. On the asymptotic laws of distribution of real roots of oscillating integrals of a linear differential equation of the second order. DAN 24, 1939, 322 – 324.
2. On elementary arithmetical properties of the laws of a chain. *Izvestia Uzbek FAN* 6, 1940, 46 – 49.
3. Application of the methods of the theory of probability to investigating some functions. *Trudy FAN* 1, 1941.
4. On sequences of stochastic matrixes. DAN 47, 1945, 331 – 333.
5. On Markov chains with a countable number of possible states. Ibidem, 641 – 644.
6. On a synthesis of the two systems of exposition in the theory of discrete Markov chains. DAN 48, 1945, 167 – 169.
7. A new necessary and sufficient test of regularity of Markov chains with a continuous set of possible states. DAN 49, 1945, 85 – 88.
8. On Bruns' chains. Ibidem, 246 – 248.
9. The law of the iterated logarithm for the Markov scheme. *Юбилейный бюллетень* (Jubilee Bull.) *SAGU*, 1945.
10. A necessary and sufficient test of the fulfilment of the ergodic principle for homogeneous stochastic processes. *Bull. AN Uzbek SSR* 8, 1946.
Savkevich, V.P: 1. On the pattern of an urn with added balls. DAN 28, 1940, 8 – 12.

Shchigolev, B.M. 1. On the expansion of asymmetric curves of distribution into two Gaussian curves. *Astron. Zhurnal* 1: 3 – 4, 1924, 76 – 89.
2. Expansion of a distribution function of three variables into two spherical distributions. Ibidem 2:1, 1925, 1 – 6.
3. Expansion of a distribution of three variables into two reduced normal distributions. Ibidem 3:3 – 4, 1926, 145 – 256.
Shepelevsky, A. 1. On the criticism of the Pearsonian theory of frequency curves. *Zhurnal Matematich. Zikla AN Ukr. SSR* 2, 1932, 1 – 24. (U)
Shillo, P.I. 1. *Теория ошибок* и т.д. (Theory of Errors and the Method of Least Squares). M., 1925.
Schmidt, O.Yu. 1. On the Bertrand paradox in the theory of probability. MS 33, 1926, 33 – 40.
Slutsky, E.E. 1. On the logical foundation of the calculus of probability. VS 9 – 12, 1922, 13 – 21. Translated in *Probability and Statistics. Russian Papers of the Soviet Period.* Berlin, 2005.
2. On some patterns of correlation connections and on the systematic error of the empirical coefficient of correlation. VS 1 – 3, 1923, 31 – 50.
3. On a new coefficient of the mean density of population. *Zapiski Sozialno-Ekonomich. Otdelenie AN Ukr. SSR* 1, 1923. (U)
4. Same title. VS 4 – 6, 1923, 5 – 19. Transalted: Deutsche Hochschulschriften 2799, 2004.
5. Über stochastische Asymptoten und Grenzwerte. *Metron* 5, No. 3, 1925, 3 – 89.
6. On a trial construction of the theory of means without applying the theory of probability. VS 1 – 3, 1925, 103 – 124.
7. On the law of large numbers. VS 7 – 9, 1925, 1 – 55.
8. Summation of random causes as a source of cyclic processes. *Voprosy Koniunktury* 3, No. 1, 1927, 34 – 64, English summary 156 – 160. English translation: **29.**
9. Über die zufällige zyklische Anordnung paarweise gleicher Elemente. *Z. angew. Math. Mech.* 6, 1926, 150 – 159.
10. Über stochastische Asymptoten und Grenzewerte. *Tôh. Math. J.* 27, 1926. Cf. **5.**
11. Sur un théorème limite rélatif aux séries des quantités éventuelles. *C.r.* 185, 1927, 169 – 171.
12. Über einen Fall wo das Gesetz der großen Zahlen auf von einander abhängige Größen Anwendung findet. *Tôh. Math. J.* 28, No. 1 – 2, 1927, 26 – 32.
13. Sur un critérium de la convergence stochastique des ensembles des valeurs éventuelles. *C.r.* 187, 1928, p. 370.
14. Sur les fonctions éventuelles continues, intégrables et dérivables dans les sens stochastiques. Ibidem, 878 – 880.
15. Quelques propositions sur les limites stochastiques éventuelles. Ibidem, 189, 1929, 384 – 386.
16. Sur l'erreur quadratique moyenne du coefficient de corrélation dans le cas des suites des épreuves non indépendantes. Ibidem, 612 – 614.
17. Sur l'extension de la théorie de périodogrammes aux suites de quantités dépendantes. Ibidem, 722 – 723.
18. On the standard error of the correlation coefficient for homogeneous coherent series. *Trudy Koniunkturn. Inst.* 2, 1929 (1930), 64 – 101. English summary 150 – 154.
19. Sur les functions éventuelles compactes. *Atti Congr. Intern. Mat. Bologna 1928.* Bologna, 1932, t. 6, 101 – 115.
20. On the distribution of the errors of the correlation coefficient in homogeneous coherent series. *Zhurnal Geofiziki* 2, No. 1, 1932, 66 – 98.
21. On the existence of a connection between the solar constant and temperature. Ibidem, vol. 3, No. 3, 1933, 263 – 281.

22. Alcune applicazioni dei coefficienti di Fourier all'analisi delle funzioni aleatorie stazionarie. GIIA 5, No. 4, 1934, 3 – 50.
23. On the solar constant. *Zhurnal Geofiziki* 4, No. 3, 1934, 392 – 399.
24. Statistical experiment as a method of research. Ibidem, 5, No. 1, 1935, 18 – 38.
25. On extrapolation in connection with the problem of forecasting. Ibidem, No. 3, 263 – 279.
26. On the 11-year periodicity of solar spots. DAN 4 (9), 1935, 35 – 38.
27. On connected random functions of one independent variable. *Труды первого всесоюзного съезда математиков* (Proc. First All-Union Congr. Mathematicians). M. – L., 1936, 347 – 357.
28. Qualche proposizione relativa alla teoria delle funzioni aleatorie. GIIA 8, No. 2, 1937, 3 – 19.
29. English translation of **8.** *Econometrica* 5, No. 2, 1937, 105 – 146.
30. Sur les fonctions aléatoires presque périodiques et sur la décomposition des fonctions aléatoires stationnaires en composantes. *Actualités Sci. Industr. No. 738, Colloque consacré à la théorie des probabilités*, 5me pt. Paris, 1938, 33 – 55.
31. A few proposals on the theory of random functions. *Trudy SAGU*, Math., No. 31, this being a collection of articles dedicated to V.I. Romanovsky. Tashkent, 1939, 3 – 15.
32. On the tables of the inverse incomplete B-function. *Trudy IMM AN Uzbek SSR* No. 1, 1946, 41 – 49.
Slutsky's *Избранные труды* (Sel. Works). M., 1960, contain reprints/translations of most of the items above.
Smirnov, N.V. 1. On the probabilities of large deviations. MS 40, 1933, 443 – 454.
2. Über die Verteilung des allgemeines Gliedes in der Variationsreihe. *Metron* 12, 1935, 59 – 81.
3. Sur la distribution de ω^2 (criterium de Mises). *C.r.* 202, 1936, 449 – 452.
4. Same title. MS 2 (44), 1937, 973 – 994.
5. On the number of changes of signs in a sequence of deviations. IAN 1937, 361 – 372.
6. On the dependence between terms of a variational series. *Bull. MGU* (A), 1:4, 1937, 1 – 12.
7. Estimating the discrepancies between empirical curves of distribution for two independent samples. Ibidem, 2:2, 1939, 3 – 14.
8. On the deviations of the empirical curve of distribution. MS 6 (48), 1939, 3 – 26.
9. On a limit theorem for a pattern of independent trials. IAN 1939, 319 – 328.
10. On estimating the maximal term in a series of observations. DAN 33, 1941, 346 – 349.
11. Approximating the laws of distribution of random variables by empirical data. UMN 10, 1944, 179 – 206.
12. On a test of symmetry of the law of distribution of a random variable. DAN 56, 1947, 13 – 15.
Smirnov's *Теория вероятностей и математическая статистика. Избранные труды* (Theory of Probability and Mathematical Statistics. Sel. Works). M., 1970, contains reprints/translations of all the Items above excepting **3.**
Smogorzhevsky, A.S. 1. On orthogonal polynoms connected with the Bernstein stochastic pattern. *Zhurnal Inst. Matematiki AN Ukr. SSR* 3 – 4, 1935, 119 – 125 (U)
Sokolov, I.G. 1. On the problem of moments. *Izvestia Donskoi Politekhnich. Inst*. 14, 1930, 29 – 37.
Sorkin, R.E. 1. On the most probable values of the measure of precision derived from trials. *Izvestia Artill. Akad. Raboche-Krestiansk. Krasnoi Armii* 20, 1936, 81 – 92.
Sretensky, L.N. 1. On the influence of added observations on the correlation coefficient. *Geofizich. Bull*. 14, 1926, 50 – 53.
Subbotin, M.F. 1. On the law of frequency of error. MS 31, 1924, 296 – 301. In English

2. Same title, in French. *C.r.* 180, 1925, 1716 – 1719.
Sverzhensky, S.B. 1. The function E(x) and its application in the theory of probability. *Izvestia Kazan Fiz.-Matematich. Obshchestvo* (2), 25, 1925, 41 – 43.
Syzrantsev, P.I. 1. The main coefficients of the mathematical treatment of experimental data when applying the method of means. *Trudy Saratov Selsko-Khoziastven. Inst*. 2, 1940, 123 – 128.
Tikhomirov, E. 1. On applying the method of correlation in meteorology. *Meteorologich. Vestnik* 9 – 12, 1929, 286 – 293.
Vashakidze, D.R. 1. On the maximal deviation of the theoretical law of distribution from the corresponding empirical curve. *Trudy Matematich. Inst. Gruzinsk. FAN* 4, 1938, 101 – 122.
2. Same title. *Uch. Zap. MGU* 15, 1939, 183 – 190.
Vasiliev, A.S. 1. The most probable result of measuring one and the same magnitude by several instruments. IAN, 21, 1927, 373 – 384.
2. Same title. Ibidem, 543 – 566.
Velmin, V.P. 1. Some properties of the correlation connections between magnitudes whose difference is a strict linear function. *Trudy Inst. Matematiki i Estestvoznania Severokavkazsk. Univ.* 16, 1930, 95 – 98.
Vikhliaev, G.A. 1. *Очерки теоретической статистики* (Essays on Theoretical Statistics). M., 1926. Second edition.
Volberg, O.A. 1. The problem about stationary and non-stationary queues. DAN 24, 1939, 656 – 661.
2. On the distribution of delays. *Uch. Zap. L. Pedagogich. Inst*. 28, 1939, 81 – 102.
3. The limiting case of the stationary problem of delays. Ibidem, 103 – 110.
4. The problem of delays. *Izvestia L. Voenno-Elektrotekhnich. Akad.* 17, 1939, 93 – 126.
Yaglom, A.M. 1. The ergodic principle for Markov processes with a stationary distribution. DAN 56, 1947, 347 – 349.
2. On the statistical reversibility of the Brownian motion. Ibidem, 691 – 695.
3. Some limit theorems of the theory of branching random processes. Ibidem, 795 – 797.
Yastremsky, B.S. 1. Naked empiricism and the Pearsonian curves of distribution. VS No. 1, 1927, 173 – 197.
2. On the possible values of the coefficient of total correlation in cases of linear connections. *Trudy Seminara MGU Teoria Veroiatnostei i Statistika* 1, 1930, 12 – 20.
3. *Избранные труды* (Sel. Works). M., 1964.
Yudin, M.I. 1. The general case of determining a {surveying} station on a plane by intersection. DAN 49, 1945, 485 – 488.
Zasukhin, V.N. 1. On the theory of multivariate stationary random processes. DAN 33, 1941, 435 – 437.
Zhuravsky, A.M. 1. On a limit theorem of the theory of probability. *Trudy Fiz.-Matematich. Steklov Inst.* 4, 1933, 9 – 36.

Part 2. Foreign Authors
Ambrose, W. (1940), On measurable stochastic processes. *Trans. Amer. Math. Soc.* 47, 66 – 79.
Feller, W. (1943), The general form of the so-called law of the iterated logarithm. *Trans. Amer. Math. Soc.* 54, 373 – 402.
Hotteling, H. (1936), Relations between two sets of variates. *Biometrika* 28, 321 – 377.
Koopman, B.O. (1940), The axioms and algebra of intuitive probability. *Ann. Math.* 41, 269 – 292.
Pólya, G. (1930), Sur quelques points de la théorie des probabilités. *Ann. Inst. H. Poincaré* 1, 117 – 161.

10. A.N. Kolmogorov. The Theory of Probability

In *Математика в СССР за 40 лет* (Mathematics in the Soviet Union during 40 Years), vol. 1. Moscow, 1959, pp. 781 – 795 …

Foreword by Translator

The references to all the essays comprising vol. 1 of the memorial publication which included this essay by Kolmogorov and the essay by Gikhman & Gnedenko, also translated in this book, were collected in vol. 2 of the same source. I extracted the bibliography pertinent for both essays just mentioned above in a Joint Bibliography appended below, after the second essay. In other words, my Joint Bibliography is a small portion of that vol. 2.

An example is necessary. Kolmogorov cited Khinchin [133], but since he only referred to 15 of the latter's contributions, and since Gikhman & Gnedenko had not mentioned Khinchin at all, my Joint Bibliography only includes 15 Khinchin's writings (out of the 149 listed in vol. 2 of the memorial publication).

[Introduction] The essay on the Soviet work in probability theory during 1917 – 1948 {written by Gnedenko & Kolmogorov; translated in this book and called in the sequence G&K} appeared in the period when the general methods of the theory of random functions and the theory of stochastic processes with continuous time still required popularization and proof of their power and importance. Nowadays their place in science is sufficiently ascertained and the danger is rather felt of underestimating the work directed at obtaining precise and effective results when solving concrete problems both remaining from the previous periods of the development of probability theory and occurring in connection with new practical requirements or in the main body of our science.

Unlike G&K, this essay is being tentatively composed not chronologically (from classical problems to new concepts and methods) but logically (from general concepts to special problems, classical included). We follow the classification of the central theoretical problems of the theory of probability itself and do not aim at explicating the use of stochastic methods. Instead, we indicate here some spheres of applications where mathematicians have been working systematically. These are

1) Mathematical foundations of statistical physics (Khinchin [129; 131; 134; 136; 137]; Yaglom [24]; Sragovich [2] and others).

2) Stochastic foundations of the theory of information (Khinchin [141; 147]; Youshkevich [1]; Pinsker [3; 5]; Kolmogorov [160]; Faddeev [29]; Gelfand, Kolmogorov & Yaglom [75]; Gelfand & Yaglom [79] and others).

3) Queuing theory (Khinchin's monograph [143] and others).

The works of Linnik [74; 83], Kubilius [10; 14], Postnikov [10] were devoted to *intramathematical* applications of stochastic methods to the theory of numbers. This purely theoretical line of research will possibly also acquire practical interest, e.g., for comparing the real stochastic *Monte-Carlo* method with its number-theoretic imitations.

My citing or non-citing of a certain contribution should not be considered as an attempt of appraisal. Works of comparatively little importance can be mentioned here even at the expense of deeper but more isolated writings if this seems opportune for illustrating the nature of some investigations of an apparently essential *direction* requiring wide development.

1. Distributions. Random Functions and Stochastic Processes

G&K (§§3 and 6) described the works of sufficiently general nature published during the previous period on the logical foundations of the theory of probability and on the tools applied for studying the distributions of *random objects*. From among the new Soviet contributions again raising the issue about the possibility of assuming the normed Boolean algebra (without introducing the set of *elementary events*) as the main initial concept of the system of events, it is only possible to indicate Kolmogorov [105]. Neither did Soviet mathematicians systematically study the problem concerning a sensible introduction of the notion of conditional probability; or of the possibility of a general construction of the entire probability theory without unconditional probabilities at all by adequately axiomatizing the concept of conditional probability. The definitions of a process as a system of concordant conditional distributions (of the process's future course given its present state) are only introduced in works of Markov processes (§3).

On the contrary, studies of distributions in functional and abstract linear spaces were numerous. In the theory of stochastic processes it is natural to consider, in addition to the space of continuous functions, the space of functions of one variable with gaps of only the first kind (processes with jumps). Chentsov [1] offered a simple condition for a random function to belong to this class: such constants $p \geq 0$, $q \geq 0$, $r > 0$ and C should exist that, for $t_1 < t_2 < t_3$,

$$\mathrm{E}|\xi(t_1) - \xi(t_2)|^p\, |\xi(t_2) - \xi(t_3)|^q < C|t_1 - t_3|^{1+r}.$$

He established that if this condition is fulfilled uniformly for a sequence of functions $\xi_n(t)$ for which the finite-dimensional distributions $[\xi_n(t_1); \xi_n(t_2); \dots; \xi_n(t_s)]$ weakly converge to finite-dimensional distributions $[\xi(t_1); \xi(t_2); \dots; \xi(t_s)]$, then $\xi(t)$ also belongs to the space of functions with gaps of only the first kind and in some natural sense possesses in that space a distribution limiting with respect to the distributions of the functions $\xi_n(t)$.

Especially Skorokhod [4; 7; 8; 10] (also see Kolmogorov [151] and Prokhorov [16, Chapt. 2, §3]) had widely studied topologies in the space of functions with gaps of only the first kind in connection with limiting theorems (see §4). Distributions in functional spaces are usually given by finite-dimensional distributions of $\xi(t_1); \xi(t_2); \dots; \xi(t_s)$; by a characteristic functional; or in some special way. The convergence of these characteristics usually leads to a weak convergence of the distributions themselves only after the additional condition of compactness. Thus occurred the problem of determining the conditions for the compactness of the families of distributions in functional spaces which Prokhorov [16] (also see the essay Kolmogorov & Prokhorov [158]) solved in a sufficiently general way (for metric complete separable spaces).

Characteristic functionals introduced already in 1935 (Kolmogorov [53]) remained for a long time without application, but they became an essential tool for studying distributions in linear spaces in the works of the French (Fortret, Edith Mourier) and the Soviet (Prokhorov [16]) schools. A characteristic functional

$$H(f) = \mathrm{E}e^{if(\xi)} = \sum_{n=0}^{\infty} (i^n/n!)\mathrm{E}[f(\xi)]^n = \sum_{n=0}^{\infty} (i^n/n!)A_n(f) = \exp[\sum_{n=1}^{\infty} (i^n/n!)B_n(f)]$$

naturally leads to the moments $A_n(f)$ and the semi-invariants $B_n(f)$ of a distribution in a linear space. For concrete functional spaces [when ξ is a function $\xi(t)$] these are written down as

$$A_n(f) = \int \dots \int a_n(t_1; \dots; t_n)\, f(t_1) \dots f(t_n)\, dt_1 \dots dt_n,$$
$$B_n(f) = \int \dots \int b_n(t_1; \dots; t_n)\, f(t_1) \dots f(t_n)\, dt_1 \dots dt_n.$$

Kuznetsov, Stratonovich & Tikhonov [18; 20; 21] called the factors a_n and b_n *momentous* and *correlational* functions respectively. They began to realize a wide program of applying this tool to solving concrete problems (also see Cherenkov [1]).

A more elementary theory only using the first moments $A_1(f)$ and $B_1(f)$ and the second central moments $B_2(f)$ is widely applied in contributions to technical sciences. The quadratic functional $B_2(f)$ is reduced to a sum of squares by an adequate choice of the coordinates; this is simply Pugachev's *canonical expansion* [13]. Pugachev's book [21] summarizes the effective methods and the experience in applying the theory of random functions in engineering [1].

The theory of distributions in infinite-dimensional linear spaces, when compared with the theory of finite-dimensional distributions, suffers from some defects. Thus, not any continuous positive-definite functional $H(f)$ having $H(0) = 1$ is a characteristic functional of some distribution. Even for the Hilbert space the additional conditions which should have ensured the existence of a corresponding distribution remain scarcely effective. This fact does not, however, persist in the theory of generalized random functions also having many direct applications. A number of findings pertaining to the theory of generalized functions were contained in Gelfand's short note [65].

2. Stationary Processes and Homogeneous Random Fields

We (G&K, §§4.3 – 4.4) briefly mentioned the works of Khinchin, Kolmogorov, Zasukhin and Krein on the spectral theory of stationary processes. During that previous period the main achievements were as follows. The contributions of Slutsky on stationary sequences and Wiener's general harmonic analysis brought about the understanding that stationarity of a process automatically leads to the possibility of a spectral representation. Khinchin [72] provided an appropriate harmonious and very simple general spectral theory of stationary (in the wide sense) processes that initiated a large section of the modern theory of probability. Kolmogorov remarked that from a formal mathematical point of view this theory was a direct corollary of the spectral theory of one-parameter groups of unitary operators. This enabled him to offer a simple exposition of the findings of Khinchin and Cramér, and to construct, issuing from the works of Wold, a spectral theory of extrapolation and interpolation of stationary sequences [84; 90; 92].After its continuous analogue was discovered; and after it was supplemented with the somewhat later Wiener theory of *filtration*, it acquired essential importance for radio engineering and the theory of regulation. Kolmogorov [85; 86] provided the theory of processes with stationary increments in a geometric form. Zasukhin [1] solved some problems of the many-dimensional spectral theory (see a modern explication of his findings in Rosanov (1958)). Finally, Krein [85] offered the abovementioned continuous version of the theory of extrapolation. In particular, he established that the divergence of the integral

$$\int_{-\infty}^{+\infty} \frac{\log f(\lambda)}{1+\lambda^2} d\lambda$$

with $f(\lambda)$ being the spectral density was necessary and sufficient for singularity (*i.e.*, for the possibility of precise extrapolation). Yaglom's essay [12] described the further development of the spectral theory of stationary processes up to 1952. Krein [134] and Yaglom [11; 15; 19] devoted their writings to issues in extrapolation and filtration for stationary one-dimensional processes. Rosanov [2; 3] studied the many-dimensional case for sequences [2]; in particular, he discovered an effective necessary and sufficient condition for the possibility of a precise interpolation of a stationary sequence of vectors. The corresponding extrapolational problem is still not quite effectively solved.

Pinsker & Yaglom [1] constructed a systematic theory of processes *with n-th stationary increments* (also see Yaglom [17; 18]). In essence, this theory ought to fit in with the theory of stationary generalized functions as created according to Ito and Gelfand (§1) as a particular case. Pinsker, whose work [3] I have already mentioned in the Introduction, provided a simple spectral representation of the *amount of information per unit time in process* $\xi(t)$ *with respect to process* $\eta(t)$ for Gaussian processes. Kolmogorov, in his essay [160], explained how to apply this formula to calculate *the velocity of creating messages for a given precision of transmission* and the *channel capacity*. Pinsker [5] derived one-sided estimates of the same magnitudes also for non-Gaussian processes.

Obukhov [7; 8] and Yaglom [28], in connection with the development of the theory of homogeneous and locally homogeneous (in particular, of isotropic and locally isotropic) turbulence, worked out the spectral theory of homogeneous vector fields having homogeneous increments. It can have many other applications as well. Chiang Tse-pei (1957) initiated the study of the corresponding extrapolation problems (from semi-space to the entire space).

The main difficulty in the engineering applications of the theory of stationary processes (see the books of Bunimovich [1] and Levin [7]) is encountered when considering non-linear problems. If the sought process is connected with the given processes by linear differential or integral equations, its spectrum is calculated by issuing from the spectra of these latter. This, however, is not so if the connections are non-linear, and even when calculating the spectrum of a given process we have to apply more subtle characteristics of the given and the intermediate processes. Theoretically, we can use the characteristic functional, and, in most cases, moments of the higher order. However, in spite of numerous works published in this sphere, any harmonious general non-linear theory of stationary processes is still lacking. In applications, a prominent part is played by considering a real stationary process

$$\xi_1(t) = \int e^{i\lambda t}\Phi(d\lambda)$$

together with an adjoint process $\xi_2(t)$ which leads to the concepts of envelope $v(t) = \sqrt{\xi_1^2(t)+\xi_2^2(t)}$ and phase. Approximate methods for calculating the spectrum of the envelope are worked out.

In applications, much attention is given to the calculation of such magnitudes as the mean number of ejections (of $\xi(t)$ passing beyond pre-assigned limits) and the distribution of their durations. The first problem is easily and definitively solved, but the methods offered for solving the second one are still very imperfect (see Bunimovich [1] and Kuznetsov, Stratonovich & Tikhonov [18]).

3. Markov Processes with Continuous Time

The transition probabilities

$$P_s^t(x; M) = P[\xi(t) \in M|\xi(s) = x],\ s \le t$$

describing the transition from state x into the set of states M during the interval of time from s to t generate a system of operators $F^t = H_s^t F^s$ which transform the distribution of probabilities F^s at time s into distribution F^t at time t. These operators are connected by the relation $H_s^t = H_u^t H_s^u$, $s \le u \le t$.

In the homogeneous case $H_s^t = H^{t-s}$ and the operators H^τ form a semi-group: $H^{\tau+\sigma} = H^\tau H^\sigma$. This makes it natural to assume the existence of an *infinitesimal operator U* by whose means the operators H^τ are expressed in accordance with the formula $H^\tau = e^{\tau u}$ (Kolmogorov [46]). It is natural to suppose that, also among processes non-homogeneous in time, an important (in applications, the main) part should be played by processes for which the infinitesimal operator

$$U(t) = \lim \frac{H_t^{t+\Delta} - E}{\Delta}, \ \Delta \to 0,$$

exists and the operators H_s^t are expressed through $U(t)$ by means of the *multiplicative integration* according to Volterra [3]

$$H_s^t = \int\int_s^t [1 + U(\tau)]d\tau.$$

The concrete realization of this program (Iosida, Feller, Dynkin) revealed the expediency of considering both the operators adjoining to H_s^t, $f_s = T_s^t f_t$, which transform the *martingales*, *i.e.*, the functions of the state $f_t(x)$ obeying the relation

$$\mathrm{E}\{f_t[\xi(t)]|\xi(s) = x\} = f_s(x),$$

and the adjoining infinitesimal operators

$$A_t = \lim \frac{T_{t-\Delta}^{t} - E}{\Delta}, \ \Delta \to 0.$$

For non-homogeneous cases it is inevitable to *postulate* the existence of infinitesimal operators. Otherwise, we may only reckon on proving under wide conditions the existence of families of operators $V(s; t)$, $B(s; t)$ additively depending on the interval $(s; t)$ of the time axis which will allow to express H_s^t and T_s^t as multiplicative Stieltjes integrals

$$H_s^t = \int\int_s^t [1 + V(\tau; \tau + d\tau)], \ T_s^t = \int\int_s^t [1 + B(\tau; \tau + d\tau)]. \ (*)$$

When, however, $U(t)$ and $A(t)$ exist, the operators $V(s; t)$ and $B(s; t)$ themselves are expressed through these by usual integrals

$$V(s; t) = \int_s^t U(\tau)\, d\tau, \ B(s; t) = \int_s^t A(\tau)\, d\tau.$$

Dobrushin [4] fulfilled this program for a finite number of states. If the infinitesimal operators are given, the reconstruction of the process by issuing from them is possible by the Ito method of *stochastic differential equations*. In the Soviet literature the work of Martynov [1] is devoted to this issue.

For the homogeneous case, a finite number of states and stochastic independence of the process, the existence of infinitesimal operators uniquely determining the process is simple to prove and was known long ago. Only Dynkin [37; 42] obtained a similar result for a continuous manifold of states (a straight line, an n-dimensional Euclidean space or an arbitrary metric space). He thus completed a number of investigations made by other authors (Iosida, Feller).

3.1. Strictly Markov Processes. When studying Markov processes by direct stochastic methods it is often assumed that independence of the future course of a process given $\xi(\tau)$ from its course at $t < \tau$ exists not only for a constant τ but also for a random τ determined, in a certain sense, by the previous course *without looking ahead*. Hunt, in 1956, indicated the need to justify this assumption. Dynkin [43] and he & Youshkevich [7] defined the appropriate concept of a *strict Markov process* in a general way. They showed that not all the Markov processes were strictly Markovian, and, at the same time, that the class of the latter processes was sufficiently wide. In future, exactly the theory of strictly Markov processes will apparently be considered as the theory of *real* Markovian processes.

3.2. The Nature of the Paths of the Process. For a finite or a countable set of states the processes, where, during finite intervals of time, there only occur (with probability 1) a finite number of transitions from one state to another one, are of an unquestionably real interest. Such processes are here called regular. For homogeneous processes with a countable number of states, and for general (non-homogeneous in time) processes with a finite number of states Dobrushin [2; 5] discovered the necessary and sufficient conditions of regularity. Processes continuous on the right constitute an important class of processes with a finite or countable number of states. In such cases, issuing from any given state, the process will pass discretely over a sequence of states until the sequence of the moments of transition condenses near the limit point. Thus, for example, behave explosive processes of propagation. When considering such processes only on a random interval of time, up to the first condensation of the moments of transition, it is natural to call them *semi-regular*. Youshkevich [3] studied the conditions for continuity on the right.

Problems on reaching the boundaries play a similar part for continuous processes of wandering of the diffuse type. In particular, *non-attainment* of the boundaries here corresponds to regularity. Already Bernstein (for example, [114]), although applying another terminology, studied the attainment or non-attainment of the boundaries. In a modern formulation the problem is easily solved in a definitive way for one-dimensional diffuse processes (Khasminsky [1]).

Dynkin [21] determined the conditions of continuity of the path and existence of only jumps of the first kind in the path for general Markov processes with a set of states constituting an arbitrary metric space. Later L.V. Seregin derived a sufficient condition for the continuity of the paths which is also necessary for a wide class of processes.

3.3. The Concrete Form of the Infinitesimal Operators. Infinitesimal operators for a finite or countable number of states in simple problems having real meaning are prescribed by densities of the transition probabilities $a_{ij}(t)$ from state i to state j ($i \neq j$). Kolmogorov [124] somewhat strengthened a finding achieved by Doob: he established that in a homogeneous and stochastically continuous case these densities always exist and are finite. However, they only determine the process if it is regular (for *semi-regularity*, they determine the course up to the first condensation of the jumps).

Infinitesimal differential operators of the second order appeared already in the classical works of Fokker and Planck on continuous processes with sets of states being differentiable manifolds. Kolmogorov, in his well-known work [27], ascertained their statistical meaning and indicated some arguments for considering the case when a process was determined by such a differential operator as being *general* in some sense. However, it was clear from the very beginning that an exact sense could have only been attached to this assumption by adequately generalizing the concept of differential operators. Feller outlined the approaches to such a generalization. He only studied the one-dimensional case. Dynkin, in a cycle of important contributions [34; 38 40; 42], concluded this work by applying essentially new means. In particular, he was able to construct a generalization conforming to the problem at hand of the concept of an elliptical differential operator also for the many-dimensional case.

Adjoining Feller, Dynkin [42] and Ventsel [1] also exhaustively studied the problem of the form of differential operators for such one-dimensional processes which were continuous on some interval, but, after attaining its boundaries, could return within the interval either continuously or jumping to its inner point.

3.4. Some Special Problems for Continuous Processes. Continuous processes directed by the Fokker – Planck differential equations have many applications. Yaglom [10] investigated by an interesting tensor tool the degenerative Fokker – Planck equation of the type that appears when studying the Brownian motion with inertia (see Kolmogorov [45]). In particular, he was able, in this degenerative case, to solve the problem that Kolmogorov [76] had solved for the non-singular case by following Schrödinger. The study of many-dimensional diffuse processes by means of tensor differential geometry probably has a great future.

The calculation of the distributions of functionals of paths is of essential interest, see Gelfand & Yaglom [69]. From among the Soviet contributions the works of Dynkin [36] and Khasminsky [1] should also be indicated.

Yaglom [2] and Ginsburg [1 – 4] devoted their writings to the problems of existence and the nature of limiting distributions as $t \to \infty$ and of ergodicity for diffusive processes.

3.5. Branching Processes. Already in 1947 Dmitriev & Kolmogorov [3] established the main differential equations for branching processes with an arbitrary number of types of the particles. Sevastianov [2; 3] essentially supplemented the theory, and, in his essay [4], collected everything done by Soviet and foreign authors up to 1951.

Already in 1947 Kolmogorov & Sevastianov [102] indicated peculiar *transitional phenomena* occurring when the expected number of *descendants* passed through 1. Sevastianov [8] returned to this subject in 1957. Chistiakov [1] derived local limit theorems connected with the study of branching processes.

4. Limit Theorems

Issues connected with the *central limit theorem* on the attraction of the distributions of sums of a large number of independent or weakly dependent scalar or vector summands to the normal Gaussian distribution multiplied in a few directions.

1) The classical limit theorems on the attraction to the Gauss distribution were made more precise (G&K, §1.3).

2) After the infinitely divisible distributions had been discovered; and after Khinchin [86], in 1937, had proved the main theorem on whose strength the limiting distribution of a sum of independent and *individually negligible* summands can only be infinitely divisible, the issue concerning the conditions for attraction to the Gauss distribution definitively became a particular problem of attraction to arbitrary infinitely divisible distributions (G&K, §§1.4 – 1.5).

From the very beginning, the foundation of this second direction has been the idea of comparing the process of the formation of consecutive sums of independent summands and a limiting process with independent increments that strictly obeys the infinitely divisible laws of distribution. Bachelier, already in 1900, had been developing this idea with respect to the Gauss case.

3) After the work of Borel (the strong law of large numbers) and Khinchin (the law of the iterated logarithm), the problems of estimating various probabilities connected with the behavior of consecutive sums have ´been becoming ever more prominent (G&K, §§1.6 – 1.7).

During the new period (1948 – 1957), the main attention was concentrated on studying the paths of discrete Markov processes by the tools of the analytic theory of these processes with continuous time (G&K, §5).

Limit theorems on the behavior of discrete Markov processes with a large number of small jumps are derived by approximating these processes by continuous Markov processes obeying the Fokker – Planck equation. When the discrete process has a small number of large jumps, the approximation is achieved by a process with continuous time having paths with gaps of only the first kind and directed by integro-differential equations (Kolmogorov [77, §19] or by their various generalizations. Theoretically, the classical limit theorems and the propositions of the second direction for independent summands are included into this pattern as particular cases.

We (G&K, end of §5) noted that up to 1947 the actual implementation of the second part of this program (application of processes continuous in time and having jumps) did not yet begin if cases of the second direction only connected with processes with independent increments are excluded.

I postpone until the next section the review of the new contributions devoted to strengthening the classical limit theorems on attraction to the Gauss distribution and on the issues of Item 2. Such a unification of the second direction with all the classical issues is caused by the fact that, although this direction had ideologically grown out of the notions of processes with continuous time, during its practical realization it followed along the lines unconnected with its origin.

As to the studying of the behavior of consecutive sums, I should indicate first of all that from among the difficult problems of the previous period Prokhorov [2; 3] essentially advanced the determination of sufficient conditions for the applicability of the strong law of large numbers to sums of independent terms. His conditions are very close to necessary restrictions and they are both necessary and sufficient for the case of Gaussian terms. A number of works is devoted to the law of the iterated logarithm (Diveev [6], Sapogov [6; 15], Sarymsakov [30]) and to the strong law of large numbers for dependent variables (Bobrov [9]).

Some findings in the contributions of Prokhorov and Skorokhod which are described below are still formulated and proved only for sums of independent terms but in essence their methods lack anything inseparably linked with such a restriction.

Gikhman, whose works [4 – 6] appeared in 1950 – 1951, studied processes $\xi(t)$ of a rather complicated nature which pass over in the limit into simpler Markov processes with continuous time $\xi^*(t)$. Prokhorov [8; 16] and Skorokhod [4; 7; 8] investigated processes of accumulating sums of independent terms or discrete Markov processes, supplemented so as to apply their method, to those of continuous time. Their (and Gikhman's) $\xi^*(t)$ was a Markov process with continuous time. The transition to $\xi^*(t)$ was made because its analytic nature is simpler (probabilities connected with it admit of a strict analytical expression) [4].

Here is the general pattern of all such investigations that explicitly appeared in American writings (Kac, Erdös et al): A process $\xi(t)$ is considered as being dependent on parameter n (the number of terms in the sum; or some indicator of their *smallness*; etc); it is required to determine the conditions under which the functional $F(x)$, defined in the space X of the realization of the processes $\xi(t)$ and $\xi^*(t)$ (by some method the space is made common), obey the relation

$$\mathrm{E}F(\xi) \to \mathrm{E}F(\xi^*) \text{ as } n \to \infty. \quad (1)$$

In the particular case of a characteristic functional

$F(\xi) = 1$ if $\xi \in A$ and $= 0$ otherwise

of the set $A \subseteq X$ the problem is reduced to that of the convergence of probabilities

$$P(\xi \in A) \to P(\xi^* \to A). \quad (2)$$

By definition of weak convergence $P_\xi \Rightarrow P_{\xi^*}$ (see §1) of distribution ξ to distribution ξ* in *X*, this is the convergence (1) for all bounded continuous functionals *F* which indeed Gikhman, Prokhorov and Skorokhod proved under sufficiently wide assumptions. If the functional *F* is not only continuous but also *smooth*, then, in some cases, even good quantitative estimations of the rapidity of the convergence can be derived (Gikhman [8]). Regrettably, in the most interesting problem of convergence (2) the functional *F* is discontinuous. Nevertheless, weak convergence (1) leads to (2) if the boundary of *A* has probability *zero* in the limiting distribution P_{ξ^*}. But good quantitative estimates of the convergence are difficult to make. Prokhorov [16] had by far overcome his predecessors by obtaining in some simple problems a remainder term of the order of $n^{-1/8}$ when it was expected to be of the order $n^{-1/2}$.

Only in such simple problems as the estimation of the probability $P(\zeta \le a)$ for the maximum

$$\zeta = \max \frac{\mu_n - np}{\sqrt{Np(1-p)}}, \; 0 \le n \le N$$

in the Bernoulli trials Koroliuk [10] was able to arrive at asymptotic expressions with a remainder term of the order of $1/N$. With similar refinements Smirnov [12] and Koroliuk [5] solved some problems on the probability of passing beyond assigned boundaries for wanderings connected with issues in mathematical statistics [5]. Gikhman [16; 19] had also applied methods explicated in this section to problems in the same discipline. Restricting my attention to this very general description of the pertinent writings, I note that much depends there on a successful choice of a topology of the space of realizations *X* (see §1).

5. Distributions of Sums of Independent and Weakly Dependent Terms and Infinitely Divisible Distributions

5.1. Owing to the insignificant success in estimating the remainder terms in limit theorems of the general type (§4), the study of the limiting behavior of the distributions of the sums of a large number of independent and weakly dependent terms remains important in itself. Many works of the last decade belong here. Kolmogorov, in his essay [136], attempted to systematize the trends in the works done during the latest period and to list the problems to be solved next in this already greatly studied sphere. The *convergence* to a limiting distribution of adequately *normed* and *centered* sums usually is not the real aim of investigation. It is usually required to have a good approximation to the distribution F_n of

$$\zeta_n = \xi_1 + \xi_2 + \ldots + \xi_n \quad (3)$$

in the form of a distribution *g* belonging to some class of distributions *G*. Thus, for independent terms ξ_k the smallness of the *Liapunov ratio*

$$L = \frac{\sum_k E|\xi_k - E\xi_k|^3}{(\sum_k \operatorname{var}\xi_k)^{3/2}} \quad (4)$$

ensures the closeness of $F_n(x) = P(\zeta_n < x)$ to the class G of normal distributions g in the sense that, in a uniform metric

$$\rho_1(F; g) = \sup |F(x) - g(x)|, \quad (5)$$

the estimation

$$\rho(F_n; G) \le CL \quad (6)$$

takes place. Here, C is an absolute constant whose best possible value is not yet determined. We (G&K, §1.3) have spoken about Linnik's works adjoining this problem.

If $\xi_1, \xi_2, \dots, \xi_n, \dots$ is a sequence of independent identically distributed summands, then, owing to the abovementioned Khinchin's theorem, the distribution of the variables

$$\eta_k = \frac{\xi_{n_k} - A_k}{B_k}$$

which correspond to any subsequence of $n_k \to \infty$ can only converge to an infinitely divisible distribution. In the general case this fact is, however, hardly interesting since the initial distribution of each of the ξ_n's can be such that, whichever subsequence and values of A_k and B_k be chosen, the limiting distribution (naturally, in the sense of weak convergence) will only be degenerate. Prokhorov [11], however, proved that under these conditions and without any norming a relation $\rho(F_n; D) \to 0$ as $n \to \infty$ with D being the class of all the infinitely divisible distributions took place for a uniform metric (5) and distribution F_n of the sum (3) whereas Kolmogorov [152] showed that $\rho(F_n; D) \le Cn^{-1/5}$ where C was an absolute constant. It is unknown whether the order $n^{-1/5}$ is definitive.

It is natural to formulate the problem of deriving approximations g to the distributions F_n of sums (3) in as wide as possible boundaries and being uniform in the sense of some metric in the space of distributions. Dobrushin's work [3] provides an example of solving a rather complicated problem where this requirement of uniformity of the estimation is met. He provided a system G of limiting distributions approximating F_n in the metric *with respect to variation* $\rho_2(F; g) = \operatorname{var}(F - g)$ so that

$$\rho_2(F_n; G) \le \frac{C \ln^{3/2} n}{n^{1/13}}$$

where C was an absolute constant. F_n was the distribution of the number of the occurrences of the separate states during n steps for a homogeneous Markov chain with two states.

The third general desire, yet rarely accomplished, is the derivation of best estimates of the approximation of distributions F_n by distributions from some class G under various natural conditions with respect to the construction of the sum ζ_n, *i.e.*, the precise or asymptotic calculation of expressions

$$E(F; G) = \sup \inf \rho(F_n; g), \; F_n; \in F, \; g \in G,$$

as it is usual for the modern theory of approximation of functions by polynomials or other analytic expressions. Dobrushin (above) was still a long way from attaining this ideal, even the order of the remainder term is apparently not definitive.

Prokhorov [7], in a considerably more elementary problem of approximating the binomial distribution by a combination of a *localized* normal and a Poisson distribution, obtained an asymptotically precise estimate with an order of $n^{-1/3}$. A more complete realization of the described program of passing on, under wide assumptions, to uniform estimates with calculation of their best possible values is mostly remaining a problem for the future. I only indicate here an interesting work by Tumanian [4] who did not offer any quantitative estimates but showed that the distribution of

$$\chi^2 = \sum_{i=0}^{s} (n/p_i)\,[(m_i/n) - p_i]^2$$

tended to its known approximate expression

$$P(\chi^2 < h^2) \approx \frac{1}{2^{s/2}\Gamma(s/2)} \int_0^{h^2} x^{(s-1)/2} e^{-x/2} dx$$

uniformly with respect to the number of classes s if only $n \min p_i \to \infty$.

5.2. Much attention was given to *local limit theorems* in both their classical versions, *i.e.*, for densities in the case of continuous distributions, and for probabilities in the case of *lattice* distributions. Some authors considered the convergence of densities in the metric

$$\rho(f; g) = \int |f(x) - g(x)| dx$$

and the convergence of probabilities, in the metric

$$\rho(P; P') = \sum_i |P_i - P_i'|$$

as particular instances of convergence with respect to variation. This seems reasonable since stronger types of convergence (for example, a uniform convergence of densities) lack direct stochastic sense. Definitive findings were obtained for identically distributed independent summands (wide conditions of applicability; estimation of the order of the remainder term; asymptotic expansions) (Gnedenko [80; 83], Meisler, Parasiuk & Rvacheva [1]; Prokhorov [6]). For non-identically distributed summands the problem is much more complicated and the findings obtained, sometimes very subtle (Prokhorov [9]; Rosanov [1]; Petrov [6; 7]), should rather be considered as points of departure in searching for more definitive results.

5.3. Persistently studied were limit theorems for sums (3) of scalars or vectors $\xi_k = f(\varepsilon_k)$ depending on the state ε_k in a Markov chain subject to the regularity of transitions $P(\varepsilon_{k+1} \in E | \varepsilon_k = e) = P_k(E; e)$. If the number s of states $e^1, e^2, \ldots, e^s$ is finite, the problem is entirely reduced to studying the (naturally, $(s-1)$-dimensional) distribution of the vector $\mu = (\mu^1, \mu^2, \ldots, \mu^s)$ composed of the number of times of the occurrences in separate states. For the homogeneous case (in which P_k do not depend on k) and the pattern of sequences Kolmogorov [115] obtained, by the *Doeblin* direct method, an absolutely general local limit theorem for the distribution of the vector μ. With respect to one point Rosenknop [2] and Chulanovsky [2] supplemented this work. For the same homogeneous case and applying an algebraic tool earlier developed in the Soviet Union by Romanovsky, and expansions of the Chebyshev – Hermite type, Sirazhdinov [10 – 12] obtained a refined local limit theorem.

For the pattern of homogeneous series the problem is more complicated; above, I mentioned Dobrushin's work where all the distributions for the case of two states were found. Iliashenko (1958) studied the case of any finite number of states.

The works of Markov and Bernstein on the non-homogeneous case found a remarkable sequel. In new formulations the part played by the transition of the exponent α through the boundary α = 1/3 and discovered by Bernstein (G&K, end of §4.1) occurred to be valid under very general conditions [6].

After the works of Linnik and Sapogov (Linnik [43]; Sapogov [9]; Linnik & Sapogov [44]), Dobrushin [9; 10; 12; 13] obtained especially wide formulations. Statuljavicus [3] arrived at the most definitive results in refining the non-homogeneous case by expansions of the Chebyshev – Hermite type.

5.4. Especially important both for the practical application of mathematical statistics and for some theoretical investigations is the refinement of asymptotic expressions and estimates of their remainder terms for low probabilities of *large* deviations (*i.e.*, of those whose order is higher than the mean square deviation).

Along with the expansions of the Chebyshev – Hermite type, the analytical tool here is a special method of approximately expressing the probabilities of large deviations introduced by Khinchin [43] already in 1929 for a special occasion of studying the binomial distribution but having a more general applicability (Cramér, in 1938). During the period under discussion, Khinchin [130; 133] examined the probabilities of large deviations for sums of positive variables. The Linnik Leningrad school (Petrov [1; 3]; Richter, 1957) systematically studied large deviations for sums of scalar and vector variables, independent or connected into a Markov chain.

5.5. Sums of the elements of a commutative or non-commutative group are a natural generalization of sums of scalar or vector variables. The works of Vorobiev [9] and Kloss [1] were devoted to them. Kloss obtained quite a definitive result by completely ascertaining the limiting behavior of sums of independent and identically distributed elements of an arbitrary bicompact group.

5.6. Closely linked with limit theorems are the analytical investigations of the properties of the limiting distributions which occur in these propositions. Linnik [101], Zolotarev [1; 3], Skorokhod [2; 4], Ibragimov [1], Rvacheva [11] studied the analytical properties of infinitely divisible distributions. A number of works were devoted to the *arithmetic of the laws of distribution*, as Khinchin called it, *i.e.*, to the decomposability of distributions into convolutions of distributions.

In 1917 – 1947 the works of Khinchin [82; 91], Raikov [2; 8], Gnedenko [6] were devoted to the arithmetic of distributions. During the last decade Sapogov [21] proved the *stability of the Cramér theorem*; that is, he established that a distribution close to the normal law can only be expanded into distributions close to the normal. Linnik [93; 101] showed that a convolution of the Gauss distribution with the Poisson law can only be expanded into distributions of the same type and obtained exhaustive results formulated in a more complicated way regarding the conditions for the possibility of expanding infinitely divisible distributions into infinitely divisible components.

Notes

1. {The second reference to Pugachev is obviously wrong; I replaced it in the Bibliography by Pugachev [27].}
2. {Rosanov [3] was not listed in the Bibliography.}
a**3.** {Here, in addition to the appearing double integral, Kolmogorov applied the symbol of another integral placing it above and across it. This additional symbol was also inserted below, in both cases, in line (*).}
4. The entire construction is quite parallel to the old Bernstein theory of *stochastic differential equations* (differing in that Bernstein forbade to apply the limiting process) and to the works of Kolmogorov, Khinchin and Petrovsky (differing in that these old contributions only considered simplest problems without applying to the functional space of the realization of the processes).
5. Some of Koroliuk's formulations were mistaken and Czan Li-Cyan (1956) corrected them.
6. Bernstein's result concerns the *pattern of series* with the *n*-th series being determined by the finite sequence of matrices

$$\begin{pmatrix} p_{11}^{(1)} & p_{12}^{(1)} \\ p_{21}^{(1)} & p_{22}^{(1)} \end{pmatrix}, \begin{pmatrix} p_{11}^{(2)} & p_{12}^{(2)} \\ p_{21}^{(2)} & p_{22}^{(2)} \end{pmatrix}, \ldots, \begin{pmatrix} p_{11}^{(n)} & p_{12}^{(n)} \\ p_{21}^{(n)} & p_{22}^{(n)} \end{pmatrix}.$$

The results for the *pattern of sequences* are different (Skorokhod [1]). {With respect to Bernstein Kolmogorov referred to G&K; there, two exact references were given: Bernstein (1926; 1928).}

11. I.I. Gikhman, B.V. Gnedenko. Mathematical Statistics*

In *Математика в СССР за 40 лет* (Mathematics in the Soviet Union during 40 Years), vol. 1. Moscow, 1959, pp. 797 – 808 …
{*See my Foreword to the preceding essay by Kolmogorov.}

[Introduction] During the last years, investigations in the field of mathematical statistics acquired a noticeably relative weight. This fact was called forth, above all, by two causes, namely, by the exceptional variety and importance of statistical applications, and by the width and the depth of the formulations of general statistical problems. Until recently, traditional issues in demography, the theory of firing and biology provided almost the sole object of statistical applications and the only source for formulating its problems; nowadays, in spite of all their importance, they are only occupying a part of the wide field of applications. Along with these issues, a considerable part is now being played by the problems raised by modern engineering, physics and economics as well as by other fields of human activities.

Soviet scientists considerably assisted in this progress by formulating general problems, obtaining findings of general theoretical nature, and by purely practical investigations. Nevertheless, we ought to indicate that, in spite of its very considerable advances, Soviet mathematical statistics is far from being equal, either in its development or in the width of its practical applications, to the increasing requirements of our country. We ought to admit therefore that, along with a further broadening of research in all the most important theoretical directions of statistics, it is necessary to speed up concrete statistical investigations of the urgent problems caused by the development of natural sciences, engineering, economics, etc.

We shall now pass on to a very short essay on the advances of Soviet mathematical statistics for the 40 years of Soviet power, remembering, however, that Smirnov (1948) had already compiled a similar review for the first 30 years. We have to leave aside applied research, which, according to our opinion, deserves to be described separately.

1. The largest part of the most important issues in mathematical statistics is reduced to estimating unknown parameters and testing statistical hypotheses. Concerning these topics, Fisher, Neyman, E.S. Pearson and others worked out the general viewpoint adopted at present. During recent years, Kolmogorov's paper [117] was occupying a prominent place from among those devoted to estimating parameters. He explicated some general ideas about the part played by unbiased estimators and their connection with sufficient statistics (somewhat generalizing Blackwell's findings (1947)) and provided examples of obtaining and investigating unbiased estimates in some concrete instances. He gave special attention to unbiased estimation of the relative number of defective articles either contained in a batch or mistakenly let through. Kolmogorov believes that unbiased estimators are not yet sufficiently developed.

Girchik et al (1946) preceded Kolmogorov in this direction. They obtained some estimates for sampling from a binomial population whereas Kolmogorov [129] and Sirazhdinov [13; 17] continued the former's earlier investigation. Dynkin [19] published a considerable study of sufficient statistics. It occurred that, along with this notion, his own concept of necessary statistics was also useful. A sufficient statistic contains all that possibly can be elicited from observations for estimating the unknown parameters, but a statistic should {also} be necessary for preventing a substantial loss of information. Dynkin derived (under some additional conditions) all those families of one-dimensional distributions that, for any sample of size $\geq r$, have an r-dimensional sufficient and necessary statistic. He especially examined the necessary and sufficient statistics for families of one-dimensional distributions of the type $F(x - \theta)$, $F(x/\sigma)$ and $F[(x - \theta)/\sigma]$. A number of researchers, and Romanovsky in the first place, studied the estimation of the values of unknown parameters.

2. An interesting work of A.A. Liapunov [35] adjoins the Neyman and E.S. Pearson investigation of the choice between two simple hypotheses. Let the sample space R be separated into n parts E_j, $j = 1, 2, \ldots, n$, in such a way that if $x \in E_j$ then hypothesis F_j is adopted. Liapunov proved that, under definite assumptions, there exists a unique statistical rule, that is, a separation of R into sets E_j possessing the greatest possible degree of reliability; in other words, such sets that min $P(E_j | F_j)$ = Max.

Petrov [2] studied the testing of statistical hypotheses on the type of distribution through small samples. He considered s samples, generally of small size n. It was required to test whether the random variables in each sample had distributions $F(a_ix+ b_i)$ whose parameters could have depended on the number of the sample. Drawing on each sample separately, the author constructed new variables having distribution $\Phi(x)$ expressed through $F(x)$ and independent of a_i and b_i. The agreement between empirical data and $\Phi(x)$ was then determined by usual methods. Petrov's main result was in some sense negative. He showed that for given errors of both kind a large number of series of trials was necessary for distinguishing between the hypotheses of belonging to types $F_0(x)$ and $F_1(x)$ even when these considerably differed from each other.

Linnik [91; 95] recently considered a new problem. Suppose that a sequence of independent observations $X = (x_1; x_2; \ldots; x_n)$ is made on random variable ξ with distribution function $F(x)$. The observations are not recorded and only some statistic $Q(x)$ becomes known. Given the distribution of $Q(x)$, it is required to estimate $F(x)$. Linnik only considered the analytical aspect of the problem: If the distribution of $Q(x)$ is precisely known, what conditions should this statistic obey so that $F(x)$ can be uniquely reconstructed?

Smirnov, Romanovsky and others made a number of interesting investigations concerning the estimation of statistical hypotheses. In §4 devoted to distribution-free methods we shall describe some studies of the tests of goodness-of-fit and homogeneity.

3. The variational series is known to be the ordered sequence $\xi_1^{(n)} < \xi_2^{(n)} < \ldots < \xi_n^{(n)}$ of observations of some random variable with continuous distribution $F(x)$. Many scientists examined the regularities obeyed by the terms of a variational series. Already in 1935 and 1937 Smirnov [3; 7] systematically studied its central terms. Gnedenko's examination [17; 25] of the limiting distributions for the maximal terms appeared somewhat later. These contributions served as points of departure for further research.

Smirnov [15] devoted a considerable study to the limiting distributions both for the central and the extreme terms with a constant rank number. A sequence of terms $\xi_k^{(n)}$ of a variational series is called central if $k/n \to \lambda$, $0 < \lambda < 1$, as $n \to \infty$. Magnitudes $\xi_k^{(n)}$ are called extreme terms of a variational series if either the subscript k, or the difference $(n - k)$ are constant. Denote the distribution of $\xi_k^{(n)}$ by $\Phi_{nk}(x)$ and call $\Phi(x)$ the limiting distribution of $\xi_k^{(n)}$ (k = Const), if, after adequately choosing the constants a_n and b_n, $\Phi_{nk}(a_nx + b_n) \to \Phi(x)$ as $n \to \infty$. The following Smirnov theorem describes the class of the limiting distributions. The proper limiting distributions for a sequence with a constant number k can only be of three types:

$$\psi_\alpha^{(k)}(x) = [1/(k-1)!]\int_0^{x^\alpha} e^{-x}x^{k-1}dx,\ x,\ \alpha > 0;$$
$$\varphi_\alpha^{(k)}(x) = [1/(k-1)!]\int_0^{|x|^{-\alpha}} e^{-x}x^{k-1}dx,\ x < 0,\ \alpha > 0; \quad (1)$$
$$\lambda^{(k)}(x) = [1/(k-1)!]\int_0^{e^x} e^{-x}x^{k-1}dx,\ -\infty < x < \infty.$$

Gnedenko earlier derived the limiting distributions for the minimal term; they can certainly be obtained from (1) by taking $k = 1$. The conditions for attraction to each of these three limiting distributions exactly coincide with those determined by Gnedenko [25] for the case of the minimal (maximal) term.

Smirnov established a number of interesting regularities for the central terms. We shall speak about a normal λ-attraction if there exists such a distribution $\Phi(x)$ that, if

$$[(k/n) - \lambda]\sqrt{n} \to 0 \text{ as } n \to \infty$$

and the constants a_n and b_n (which generally depend on λ) are adequately chosen,

$$P\{[(\xi_k^{(n)} - b_n)/a_n] < x\} \to \Phi(x) \text{ as } n \to \infty.$$

The following four types exhaust the class of the limiting distributions having domains of normal λ-attraction:

$$\Phi_\alpha^{(1)}(x) = (1/\sqrt{2\pi})\int_{-\infty}^{cx^\alpha} \exp(-x^2/2)dx,\ x \ge 0,\ c,\ \alpha > 0;\ \Phi_\alpha^{(1)}(x) = 0 \text{ if } y < 0;$$
$$\Phi_\alpha^{(2)}(x) = (1/\sqrt{2\pi})\int_{-\infty}^{-c|x|^\alpha} \exp(-x^2/2)dx,\ x < 0;\ \Phi_\alpha^{(2)}(x) = 1 \text{ if } x,\ c,\ \alpha > 0;$$

$$\Phi_\alpha^{(3)}(x) = \begin{cases} \dfrac{(1/\sqrt{2\pi})\int_{-\infty}^{-c_1|x|^\alpha} \exp(-x^2/2)dx, x<0;}{(1/2)+(1/\sqrt{2\pi})\int_0^{c_2x^\alpha} \exp(-x^2/2)dx, x, c_1, c_2, \alpha > 0;} \end{cases};$$

$\Phi_4(x) = (1/2)$ if $-1 < x \le 1$ and $= 1$ if $x > 1$.

These domains are here indicated.

Gartstein [1] examined the limiting distributions for the range $\rho_n = \xi_n^{(n)} - \xi_k^{(1)}$. In particular, she proved that the class of these distributions consists of laws of the following six types:

$\psi_\alpha^{(1)}(x)$; $\varphi_\alpha^{(1)}(x)$; $\lambda^{(1)}(x)$; $\varphi_\alpha^{(1)}(x)*\ \varphi_\alpha^{(1)}(ax)$; $\varphi_\alpha^{(1)}(x)*\ \psi_\alpha^{(1)}(ax)$; $\lambda^{(1)}(x)*\ \lambda^{(1)}(ax)$.

The notation is here the same as in formulas (1). She [2] extended these results to the case of an arbitrary extreme range (rank), *i.e.*, to the difference $\rho_{rk}^{(n)} = \xi_{n-k}^{(n)} - \xi_r^{(n)}$ where r and $(n - k)$ remained constant, as well as to mixed ranks when r (say) remained constant but $k/n \to \lambda$ and, at the same time, $[(k/n) - \lambda]\sqrt{n} \to t$. In the first case the class of limiting distributions consisted of nine types, and in the second instance, of eleven types.

Meisler [5 – 7] examined the maximal term of the variational series for independent observations when the distributions depended on the number of the trial. Here is his main finding. The distribution $\Phi(x)$ can be a limiting law for the maximal term of a series when and only when (with an adequate norming) either 1) For any $\beta > 0$ there existed such a non-decreasing function $\varphi_\beta(x)$ that the equality $\Phi(x) = \Phi(x + \beta)\ \varphi_\beta(x)$ persisted for all values of x; or 2) For any α $(0 < \alpha < 1)$ there existed such a non-decreasing, continuous at point $x = 0$ function $\varphi_\alpha(x)$ that the equality $\Phi(x) = \Phi(\alpha x)\ \varphi_\alpha(x)$ persisted for all values of x. Meisler [6] also studied the properties of the distributions of this class. Note that his theory developed parallel to the theory of the laws of class L in the limiting distributions for sums of independent terms [1]. And he [3] indicated conditions for attraction to the law $\lambda(x)$ differing from those of Gnedenko [25]. His condition is {conditions are?} of a sufficiently definitive nature.

Gnedenko [73] considered the distributions of the maximal term of a variational series in a somewhat different aspect. He indicated some interrelations between the limiting distributions for sums of independent random variables and for the *maximal summand*. Loeve (e.g., 1956) essentially developed these similarities.

Finkelstein [1] examined the limiting distributions for the extreme terms of a variational series for a two-dimensional random variable. He [2; 3] also studied limiting distributions of the terms of a variational series for such random variables $\xi_1, \xi_2, \ldots$, that for any a the sequence of events $\xi_i < a$ and $\xi_i \ge a$ formed a homogeneous stationary Markov chain. Under these conditions a theory generalizing the findings for independent trials is being developed. Note that the class of limiting distributions for the maximal term includes, in addition to the three earlier discovered types, a fourth type

$\mu_{\alpha,q}(x) = qe^{-|x|}$ for $x < 0$ and $= 1$ for $x > 0$.

Here, q is constant and $0 < q < 1$.

4. It follows from the Bernoulli theorem that for a fixed x the empirical distribution function $F_N(x)$, constructed by drawing on N independent observations $x_1, x_2, \ldots, x_N$ of some random variable ξ, converges in probability to its distribution function $F(x)$. The works of Glivenko [13] and Kolmogorov [40] initiated deeper investigations. Glivenko proved that, with probability 1, the empirical distribution *uniformly* converges to $F(x)$ and Kolmogorov determined the precise asymptotic characteristic of the maximal deviation of $F_N(x)$ from $F(x)$ (for a continuous $F(x)$). Denote

$$D_N = \sup |F_N(x) - F(x)|, \ |x| < +\infty.$$

Then Kolmogorov's finding means that, as $N \to \infty$,

$$\lim P[D_N < (\lambda/\sqrt{N})] = 1 - 2\sum_{k=1}^{\infty} (-1)^{k-1}\exp(-2k^2\lambda^2), \ \lambda > 0. \quad (2)$$

This fact is being used as a test of goodness-of-fit for checking whether, given a large N, $F(x)$ is the true distribution.

The problem solved by Kolmogorov served Smirnov as a point of departure for a number of wide and deep studies. In short, his results are as follows. He [3; 7] determined the limiting distribution of the ω^2 goodness-of-fit (Cramér – Mises – Smirnov) test

$$\omega^2 = N\int_{-\infty}^{\infty} [F_N(x) - F(x)]^2 g[F(x)]dF(x).$$

Later he [17] essentially simplified his initial proof, also see Anderson & Darling (1952).

Smirnov [8] examined the limiting distribution of the variable

$$\sqrt{\frac{N_1N_2}{N_1+N_2}}\,D_{N_1,N_2} = \sqrt{\frac{N_1N_2}{N_1+N_2}}\sup|F_{N_1}(x) - F_{N_2}(x)| \quad (3)$$

as $N_1, N_2 \to \infty$. Here, the empirical distribution functions corresponded to two independent sequences of observations of a random variable with a continuous distribution. The limiting distribution coincided with the Kolmogorov law (2). Smirnov's finding is widely used for checking the homogeneity of two samples of large sizes.

Smirnov [9] also extended the just described Kolmogorov theorem. His main result is this. Let us consider the curves

$$y_1(x) = F(x) + \lambda/\sqrt{N}, \ y_2(x) = F(x) - \lambda/\sqrt{N}, \ |x| < +\infty,$$

and denote the number of times that the empirical distribution function passes beyond the zone situated between them by $\nu_N(\lambda)$. Suppose also that

$$\Phi_N(t; \lambda) = P\{[\nu_N(\lambda)/\sqrt{N}] < t\}, \ t > 0, \ \lambda > 0.$$

Then, as $N \to \infty$,

$$\Phi_N(t; \lambda) \to \Phi(t; \lambda) = 1 - 2\sum_{k=1}^{\infty} [(-1)^k/k!]\,\frac{d^k}{dt^k}\{t^k \exp\frac{-[t+2(k+1)\lambda]^2}{2}\}. \quad (4)$$

In his later investigations Smirnov [18; 19] considered the convergence of the histograms (of empirical densities) to the density of the random variable. One of his relevant theorems established that, if the density $f(x)$ possessing a bounded second derivative on $(a; b)$ satisfied the conditions

$$\min f(x) = \mu > 0,\ a \le x \le b, \int_a^b f(x)dx = 1 - \alpha < 1,$$

and, as N and s increased,

$$\overline{\lim} \frac{s^3(\ln s)^3}{N} < \infty,\ \lim \frac{N \ln s}{s^5} = 0, \text{ as } N \to \infty,$$

then, for $|x| < +\infty$ and $N \to \infty$,

$$P\{\max \frac{|\varphi_N(x) - f(x)|}{\sqrt{f(x)}} \le \frac{l_s + \lambda / l_s}{\sqrt{Nh}}\} \to \exp(-2e^{-\lambda}).$$

Here, $\varphi_N(x)$ is the frequency polygon constructed by dividing the segment $(a; b)$ into s equal parts, $h = (b - a)/s$, and l_s is the root of the equation

$$(1/\sqrt{2\pi}) \int_{l_s}^{\infty} \exp(-x^2/2)dx = (1/s).$$

In the mid-1940s, when the interest in distribution-free tests of goodness-of-fit, and, in the first place, in the Kolmogorov test, increased under the influence of practical requirements, new methods of proving the limiting relation (2) were provided. We indicate here the Feller method (1948) based on the application of the Laplace transform and Doob's heuristic approach (1949) connecting these statistical problems with the theory of stochastic processes. Note that in principle this approach was already present in Kolmogorov's earlier works [30; 40].

Mania [1; 2] and Kvit [1] (also see Berliand & Kvit [1]) applied Feller's method for deriving the limiting distribution of the maximal deviation of the empirical distribution function from the true function, and also of the maximal discrepancy between two empirical distributions on a curtailed interval [on an interval of the type of $(x;\ \alpha < F(x) < \beta;\ 0 < \alpha < \beta < 1)$] and on the complementary interval.

Gikhman [11] derived more general formulas and some new asymptotic properties of the empirical distribution functions by justifying a limiting transition to boundary value problems for differential equations of the parabolic type. He [17] proved a general theorem on the number of points of passing beyond the boundaries of a given zone during a limiting transition from a process with discrete time (or from a totally disconnected process with continuous time) to a continuous Markov process. This result includes as particular cases the Smirnov theorem; his early finding [10] on the fluctuations of the empirical frequency in the Bernoulli trials about the probability; and some other facts.

Gnedenko and his students and collaborators devoted a large number of contributions to studying statistics

$$D_N^+ = \begin{matrix}\sup\\ \inf\end{matrix}\ [F_N(x) - F(x)],\ D_{M,R}^+ = \begin{matrix}\sup\\ \inf\end{matrix}\ [F_M(x) - F_R(x)]$$

for finite values of N, M and R [2]. For $M = R$ precise and rather simple formulas for the distributions $D_{N,N}$, $D_{N,N}^+$ and for their joint distribution were {also} derived (Gnedenko & Koroliuk [60], Gnedenko & Rvacheva [64]). Koroliuk [6] determined the distribution functions of $D_{M,R}$ and $D_{M,R}^+$. For $M = pR$, assuming that $p \to \infty$, he derived the distribution of the statistics D_N^+ and D_N. Rvacheva [8], also see Gnedenko [79], considered the maximal discrepancy between two empirical distributions not on all the axis, but on an assigned stochastic {random?} interval. The works of Ozols [1; 2] adjoin these investigations.

In addition to the listed issues, problems concerning the mutual location of two empirical distribution functions were also considered. Such, for example, was the problem about the number of jumps experienced by the function $F_M(x)$ and occurring above $F_R(x)$. Gnedenko & Mikhalevich [68; 69] studied the case $M = pR$, and Paivin (Smirnov's student) considered the general case ({although} only its limiting outcome). Mikhalevich [3] derived the distribution of the number of intersections of the function $F_M(x)$ with the broken line

$$y = F_R(x) + z\sqrt{\frac{M+R}{MR}}$$

for $M = R$; also see Gnedenko [79].

The test of a hypothesis that a distribution belongs to a given class of distributions presents a more general problem than that of testing the agreement between empirical data and a precisely known distribution function. The mentioned class of distributions can, for example, depend in a definite way on a finite number of parameters. Gikhman [13; 16; 17] initiated such investigations by studying these problems for the Kolmogorov and the ω^2 tests. At about the same time Darling (1955) examined similar problems for the latter test.

The investigation of the χ^2 goodness-of-fit test for continuous distributions and an unbounded increase both in the number of observations and in the intervals of the grouping is related to the issues under discussion. Tumnanian's [1] and Gikhman's [19] findings belong to this direction. We only note that this problem is akin to estimating densities.

5. During the last years, a considerable number of studies were devoted to developing statistical methods of quality control of mass manufactured goods. These investigations followed three directions: empirical studies; development of methods of routine control; and the same for acceptance inspection. Drawing on these works, Tashkent mathematicians worked out a draft State standard for acceptance inspection based on single sampling. Such issues are important for practice, and several conferences held in Moscow, Leningrad, Kiev and other cities were devoted to them.

Contributions on routine statistical control were mostly of an applied nature; as a rule, they did not consider general theoretic propositions. The investigations of Gnedenko, Koroliuk, Rvacheva and others (§4) nevertheless assumed these very contributions as a point of departure. In most cases the numerous methods of routine statistical quality control offered by different researchers regrettably remained without sufficient theoretical foundation. Thus, an analysis of their comparative advantages and economic preferences is still lacking. From the mathematical point of view, the method of grouping proposed by Fein, Gostev & Model [1] is perhaps developed most of all. Romanovsky [117] and Egudin [7] worked out its theory and Bolshev [1; 2] suggested a simple nomogram for the pertinent calculations. Egudin provided vast tables adapted for practical use and Baiburov [1] with a number of collaborators constructed several appropriate devices {?}.

A number of scientists investigated problems of acceptance inspection, but we only dwell on some findings. As stated above, Kolmogorov [129] applied the idea of unbiased estimates for such inspection. He assumed that one qualitative indicator was inspected after which the article was considered either good enough or not. He then restricted his attention to the case in which the sample size was assigned and the batch accepted or not depending on whether all the articles in the sample were good enough or at least one of them was not. As a result, a number of accepted batches will then include defective articles. The main problem here was to estimate the number of defective articles in the accepted, and in all the inspected batches. Kolmogorov provided an unbiased estimate of the accepted defective articles and concluded his contribution by outlining how to apply his findings. In particular, he indicated the considerations for determining the sample size.

Sirazhdinov [13; 17] methodologically followed Kolmogorov, but he considered a more complicated case in which a batch was rejected if the sample contained more than c rejected articles. Interesting here, in this version of the problem, is not only the estimate of an advisable sample size, but also of an optimal, in some sense, choice of the number c. The author offered reasoned recommendations for tackling both these questions.

From among other contributions on acceptance inspection, we indicate the papers of Romanovsky [112; 122], Eidelnant [12] and Bektaev & Eidelnant [1]. These authors were also influenced by Kolmogorov.

Mikhalevich [4; 5] studied sequential sampling plans. He also restricted his investigation by considering qualitative inspection when the acceptance/rejection of a batch depended on the number of defective articles in the sample but the quantitative information on the extent of overstepping the limits of the technical tolerance or on other data important for manufacturing were not taken into account. Mikhalevich's method of studying was based on Wald's idea of decision functions (1949).

In our context, this idea is as follows. To be practical, the method of inspection should be optimal in a number of directions which are to some extent contradictory. First of all, the inspection should ensure, with a sufficiently high reliability, the quality of the accepted batches. The cost of the inspection should be as low as possible. Then, the choice of the most economical method of inspection should certainly take into account the peculiar features of the manufacturing and the nature of the inspected articles. It is therefore reasonable to assume as the initial data the cost of inspecting one article (c); the loss incurred when accepting a defective article (a); and the same, when a batch is rejected (B).

Suppose that a batch has N articles, X of them defective. Then the mean loss incurred by its acceptance is

$$U_X = \sum_m a(X-m)P[d_1; m|X] + BP[d_2|X] + c\sum_k kP[\nu = k|X].$$

Here, m is the number of recorded defective articles, $P[d_1; m|X]$, the probability that the batch is accepted and m defective articles were revealed from among those inspected; $P[d_2|X]$, the probability that the batch is rejected; and $P[\nu = k|X]$ is the probability that the decision is made after inspecting k articles. If the probability of X defective articles in a batch is $\pi(X)$, then

$$u = \sum_{X=1}^{N} U_X\,\pi(X)$$

should be considered as the mean (unconditional) loss. The optimal method of inspection is such for which this is minimal.

Mikhalevich studied optimal methods of inspection assuming that the size of the batch was large and that consequently the hypergeometric distribution might be replaced by the binomial law. Optimal here were certain repeated curtailed samples. A number of

Mikhalevich's findings by far exceeded the boundaries of the problems of acceptance inspection.

6. Statistical practice often has to decide whether observations considerably diverging from the others or from the mean are *suitable*. Numerous pertinent rules established in the literature are often groundless and are only being applied out of tradition. Authors, who treat statistical data, not infrequently reject outlying observations without applying any rules and are therefore usually led to wrong conclusions.

Assuming that the initial distribution was normal, Smirnov [20] derived the distribution of the deviation of the maximal term of the variational series from the mean normed by the empirical variance. He illustrated his results by a short table of the distribution obtained. Later Grubbs (1950) explicated this result as well; obviously, it remained unnoticed.

Bendersky [1], who followed Smirnov, determined the distribution of the absolute value of the same deviation, again normed by the empirical variance. Bendersky & Shor [2] published a monograph devoted to estimating the *anormality* of observations complete with examples worked out in detail, vast tables and criticism of wrong rules. Incidentally, wrong recommendations had even slipped into the texts of widely used manuals written by eminent authors. Thus, Romanovsky [106, pp. 25 – 29] advanced a rule based on a misunderstanding: he assumed that the maximal term of a variational series and the mean of the other observations were independent.

In this section, we shall also touch on some isolated directions of research; in the nearest future a few of them will undoubtedly attract considerably more attention. We would like to indicate first of all that the contributions on statistics of dependent trials were uncoordinated. Smirnov [21] recommended a statistic similar to the χ^2 and derived its limiting distribution for testing the hypothesis on the constancy of transition probabilities p_{ij} ($p_{ij} > 0$) in a finite Markov chain with $(s + 1)$ states.

Linnik [51] offered a method of constructing confidence intervals for the correlation coefficient in a normal stationary Markov chain under various hypotheses concerning the parameters of a one-dimensional normal distribution.

Kolmogorov [114], Boiarsky [10] and others studied the analysis of variance. Bernstein's investigations of the correlation theory, that he explicated in a number of papers, led him to introducing motions of *firm*, isogeneous and elastic correlations [35]which proved very useful. In a series of papers Sarmanov [7 – 13] developed these ideas. Working on another aspect of correlation theory, Mitropolsky [16; 18] fulfilled a number of studies mostly devoted to correlation equations.

7. Statistical practice widely uses tables of the main distributions. In some cases, however, tabulation encounters not only practical difficulties, it often leads to more fundamental complications. In the first place we ought to indicate here the tabulation of functions depending on several parameters. Of unquestionable interest is therefore the widespread use of successfully compiled nomograms. Some not numerous attempts of such kind were made in the beginning of the 1930s when A.I. Nekrasov compiled a nomogram for the Student distribution function. Three nomograms pertaining to correlation theory were included in vol. 1 of the Pearson tables. Elementary nomograms for the normal distribution served as illustrations in the well-known books of Glagolev [24] and Frank [12]. However, only Ermilov initiated a systematic study of nomographic representation of the formulas of mathematical statistics. In a lengthy paper he [1] provided nomograms of the density and the distribution function of the Student law, the χ^2 and the Fisher distribution {the F-distribution?}. Later he published nomograms of the confidence intervals for estimating both unknown probabilities by observations and the expectation of the normal law.

At least two more authors offered nomograms for the Student distribution constructed by other methods: M.V. Pentkovsky (doctoral dissertation) and James-Levi [6]. Mitropolsky included a number of nomograms in his thorough many-volume course [20, 22].

During the 40 years (1917 – 1947) a large number of tables was calculated in the Soviet Union. They were mostly published as natural appendices to appropriate articles or monographs. A comparatively small number of contributions were devoted to tables as such. From among these we indicate Slutsky's fundamental work [34] where the author, by applying a number of clever computational tricks, was able to compile faultless five-place tables of the incomplete Γ function admitting a fair interpolation throughout.

Smirnov [8] compiled a table for the Kolmogorov distribution; it was reprinted in the USA and included in a number of educational manuals. We also mention tables of the expectation of the correlation coefficient (Dikovskaia & Sultanova [4]) and numerous useful tables concerning the practical use of the various methods of statistical quality control.

Notes

1. {Notation not explained.}

2. {Notation used here as well as in the next few lines insufficiently explained. Neither were the three contributions mentioned there, in these lines, helpful. I was only able to perceive that D, unlike D^+, was concerned with absolute values of some differences.}

12. Joint Bibliography to the Two Preceding Contributions

Foreword by Translator

In addition to what I noted in my *Foreword* to the previous Joint Bibliography, I mention several more points. First, in both essays references to joint papers were made in an extraordinary way. Thus, Kolmogorov cited Gelfand [75], himself [157] and Yaglom [25] bearing in mind a single contribution. In this particular case I wrote Gelfand, Kolmogorov & Yaglom [75] (arranging the authors alphabetically) and excluded Kolmogorov [157] and Yaglom [25] from this Joint Bibliography. Second, Kolmogorov cited some foreign authors without providing an exact reference. Also, in a few instances he referred to Soviet authors in the text itself and I distinguished these cases by mentioning them in a different way, both in the translation of his paper and here. Example: Rosanov (1958). Third and last, this time, I only provide the titles of books; in other cases, I indicate the pertinent periodical, volume number, etc.

Abbreviations

AN = Akademia Nauk

C.r. = C.r. Acad. Sci. Paris

DAN = *Doklady AN* (of the Soviet Union if not stated otherwise)

IAN = *Izvestia AN* of the Soviet Union (ser. Math. if not stated otherwise, or no series at all)

IMM = Inst. Math. & Mekh.

L = Leningrad

LGU = Leningrad State Univ.

M = Moscow

MGU = Moscow State Univ.

MIAN = Steklov Math. Inst.

MS = *Matematich. Sbornik*

SAGU = Srendeaziatsk. (Central Asian) Gosudarstven. Univ. (Tashkent)

SSR = Soviet Socialist Republic

TV = *Teoria Veroiatnostei i Ee Primenenia*

Uch. Zap. = Uchenye Zapiski
Ukr = Ukrainian
UMN = *Uspekhi Matematich. Nauk*
UMZh = *Ukr. Matematich. Zhurnal*
Uz = Uzbek

Baiburov, B.S. *Приборы и автоматы* … (Instruments and Automatic Devices for Statistical Analysis and Quality Inspection in Machine Production). M., 1952.

Bektaev, K.B., Eidelnant, M.I. 1. *Trudy IMM AN UzSSR* 20, 1957, 101 – 126.

Bendersky, A.M. 1. DAN 85, 1952, 5 – 8.

Bendersky, A.M., Shor, Ya.B. *Методы оценки* ... (Methods of Estimating Non-Normality of Observations). M., 1953.

Berliand, M.E., Kvit, I.D. 1. DAN Ukr SSR 1, 1952, 13 – 15.

Bernstein, S.N. 11. IAN 20, 1926, 1459 – 1478.

18. DAN, ser. A, 1928, 55 – 60.

35. *Metron* 7, No. 2, 1927, 3 – 27.

114. *Equations Différentielles Stochastiques*. Paris, 1938.

Bobrov, A.A. 9. *Trudy IMM AN UzSSR* 10, No. 1, 1952, 19 – 25.

Boiarsky, A.Ya. 10. In *Труды второго всесоюзного совещания по математической статистике* (Proc. 2nd All-Union Conf. Mathematical Statistics). Tashkent, 1948, 307 – 312.

Bolshev, L.N. 1. *Inzhenerny Sbornik* 21, 1955, 212 – 214.

2. TV 2, 1957, 124 – 126.

Bunimovich, V.I. 1. *Флюктуационные процессы* ... (Fluctuating Processes in Radio Receivers). M., 1951.

Chentsov, N.N. 1. TV 1, 1956, 155 – 161.

Cherenkov, A.P. 1. *Vestnik MGU*, ser. Math. & Mekh. 3, 1957, 117 – 124.

Chiang Tse-pei (1957), TV 2, 1957, 60 – 91.

Chistiakov, V.P. 1. TV 2, 1957, 360 – 374.

Chulanovsky, I.V. 2. DAN 69, 1949, 301 – 304.

Dikovskaia, E.I., Sultanova, M.M. 4. *Trudy SAGU* 37 (Matematich. Nauki 8), 1954, 79 – 92.

Diveev, R.K. 16. *Trudy IMM AN UzSSR* 10, No. 1, 1952, 44 – 47.

Dmitriev, N.A., Kolmogorov, A.N. 3. DAN 56, 1947, 7 – 10.

Dobrushin, R.L. 2. UMN 7, No. 6 (52), 1952, 185 – 191.

3. IAN 17, 1953, 291 – 330.

4. MS 33 (75), 1953, 567 – 596.

5. MS 34 (76), 1954, 541 – 556

9. DAN 102, 1955, 5 – 8.

10. TV 1, 1956, 72 – 89.

12. DAN 108, 1956, 1004 – 1006.

13. TV 1, 1956, 365 – 425.

Dynkin, E.B. 7. UMN 3, No. 3 (25), 1948, 157 – 158.

19. UMN 6, No. 1 (41), 1951, 68 – 90.

21. IAN 16, 1952, 563 – 572.

34. *Vestnik LGU* 11, 1955, 69 – 74.

36. DAN 104, 1955, 691 – 694.

37. DAN 105, 1955, 206 – 209.

38. Ibidem, 405 – 408.

40. *Труды третьего всесоюзного математического съезда* (Proc. 3rd All-Union Math. Congress), vol. 2. M., 1956, 44 – 47.

42. TV 1, 1956, 38 – 60.
Dynkin, E.B., Youshkevich, A.A. 43. TV 1, 1956, 149 – 155.
Egudin, G.I. 7. *Trudy IMM AN Uz SSR* 10, No. 1, 1952, 48 – 61.
Eidelnant, M.I. 12. *Trudy IMM AN UzSSR* 15, 1955, 57 – 68.
Ermilov, N.D. 1. *Uch. Zap. Magnitogorsk. Pedagogich. Inst.* 1, 1947, 32 – 96.
Faddeev, D.K. 29. UMN 11, No. 1 (67), 1956, 227 – 231.
Fein, F.A., Gostev, V.N., Model, G.A. *Статистический контроль* … (Statistical Control by the Method of Grouping). L., 1949.
Finkelstein, B.V. 1. DAN 91, 1953, 209 – 211.
2. *Vestnik LGU* 11, 1955, 50.
3. UMZh 7, 1955, 313 – 332.
Frank, M.L. 12. *Номографический справочник* ... (Nomographic Handbook in Mathematics, Mechanics, Physics and Strength of Materials). M. – L., 1933.
Gartstein, B.N. 1. DAN 60, 1948, 1119 – 1121.
2. DAN Ukr SSR 1, 1951, 25 – 30.
Gelfand, I.M. 65. DAN 100, 1955, 853 – 856.
Gelfand, I.M., Kolmogorov, A.N., Yaglom, A.M. 75. DAN 111, 1956, 745 – 748.
Gelfand, I.M., Yaglom, A.M. 69. UMN 11, No. 1 (67), 1956, 77 – 114.
79. UMN 12, No. 1 (73), 1957, 3 – 52.
Gikhman, I.I. 4. UMZh 2, No. 3, 1950, 45 – 69.
5. Ibidem, No. 4, 37 – 63.
6. UMZh 3, 1951, 317 – 339.
8. *Trudy IMM AN UzSSR* 10, No. 1, 1952, 36 – 43.
11. DAN 82, 1952, 837 – 840.
13. DAN 91, 1953, 715 – 718.
16. UMZh 6, 1954, 28 – 36.
17. *Nauchn. Zap. Kievsk. Univ*., Math., 5, 1954, 51 – 59.
19. TV 1, 1956, 344 – 348.
Ginsburg, G.M. 1. *Zap. Kharkov Matematich. Obshchestvo* 17, 1940. 65 – 73.
2. DAN 30, 1941, 293 – 295
3. IAN 15, 1951, 563 – 580.
4. DAN 102, 1955, 441 – 444.
Glagolev, N.A. *Курс номографии* (Course in Nomography). M. – L., 1943.
Glivenko, V.I. 13. *G. Ist. Ital. Attuari* 4, 1933, 1 – 10.
Gnedenko, B.V. 6. *Bull. MGU*, A, 1, No. 5, 1937, 17 – 18.
17. DAN 32, 1941, 7 – 9.
25. *Annals Math.* 44, 1943, 423 – 453.
73. UMZh 5, 1953, 291 – 298.
79. *Math. Nachr.* 12, No. 1 – 2, 1954, 29 – 66.
80. *Wiss. Z. Humb. Univ*., Math.-Naturwiss. Reihe, 3, No. 4, 1954, 287 – 293.
83. DAN 95, 1954, 5 – 7.
Gnedenko, B.V., Kolmogorov, A.N. (1948), Theory of probability. Translated in this book.
Gnedenko, B.V., Koroluik, V.S. 60. DAN 80, 1951, 525 – 528.
Gnedenko, B.V., Mikhalevich, V.S. 68. DAN 82, 1952, 841 – 843.
69. DAN 85, 1952, 25 – 27.
Gnedenko, B.V., Rvacheva, E.L. 64. DAN 82, 1952, 513 – 516.
Ibragimov, I.A. 1. IAN 1939, 553 – 568.
Iliashenko, A.A. UMZh 10, 1958, 23 – 36.
James-Levi G.E. 6. TV 1, 1956, 272 – 274.
Khasminsky, R.Z. 1. DAN 104, 1955, 22 – 25.

Khinchin, A.Ya. 43. *Math. Ann.* 101, 1929, 745 – 752.
72. Ibidem, 109, 1934, 604 – 615.
82. *Bull. MGU* A, 1, No. 1, 1937, 6 – 17.
86. MS 2 (44), 1937, 79 – 120.
91. *Предельные законы* … (Limiting Laws for Sums of Independent Random Variables). M. – L., 1938.
129. *Trudy MIAN* 33, 1950, 1 – 56.
130. UMZh 4, 1950, 3 – 17.
131. UMN 5, No. 3 (37), 1950, 3 – 46.
133. DAN 71, 1950, 1037 – 1039.
134. *Математические основания* …(Math. Principles of Quantum Statistics). M. – L., 1951.
136. *Trudy MIAN* 38, 1951, 345 – 365.
137. DAN 78, 1951, 461 – 463.
141. UMN 8, No. 3 (55), 1953, 3 – 20.
143. *Trudy MIAN* 49, 1955, 1 – 123.
147. UMN 11, No. 1 (67), 1956, 17 – 75.
Khinchin, A.Ya., Lévy, P. *C.r.* 202, 1937, 374 – 376.
Kloss, B.M. 1. DAN 109, 1956, 453 – 455.
Kolmogorov, A.N. 25. *C.r.* 191, 1930, 910 – 912.
27. *Math. Ann.* 104, 1931, 415 – 458.
30. IAN, Math. i Estestven. Nauki 1931, 959 – 962.
40. *G. Ist. Ital. Attuari* 4, 1933, 83 – 91.
45. *Ann. Math.* 35, 1934, 116 – 117.
46. *Труды второго всесоюзного математического съезда* (Proc. 2nd All-Union Math. Congress), vol. 1. L., 1934, 349 – 358.
53. *C.r.* 200, 1935, 1717.
76. *Math. Ann.* 113, 1937, 766 – 772.
77. This is a Russian translation of **27.**
84. *Uch. Zap. MGU* 30, 1939, 3 – 16.
85. DAN 26, 1940, 6 – 9.
86. Ibidem, 115 – 118.
90. DAN 31, 1941, 99 – 101.
92. *Annals Math. Stat.* 12, 1941, 461 – 463.
105. *VI Zjazd Mat. Polskich*, 1948, 22 – 30.
114. Same source as Boiarsky, 240 – 268.
115. IAN 13, 1949, 281 – 300.
117. IAN 14, 1950, 303 – 326.
124. *Uch. Zap. MGU*, Matematika No. 4, 1951, 53 – 59.
129. *Статистический прием* … (Statistical Quality Inspection When the Admissible Number of Defective Articles Is Zero). L., 1951.
136. *Vestnik MGU* 10, 1953, 29 – 38.
151. TV 1, 1956, 239 – 247.
152. Ibidem, 426 – 436.
160. *Сессия АН СССР* ... (Session of AN Soviet Union. Scientific Problems of Automation of Production. Plenary Sittings). M., 1957, 66 – 99.
Kolmogorov, A.N., Prokhorov, Yu.V. 158. *Bericht über die Taging Wahrscheinlichkeitsrechnung und math. Statistik*. Berlin, 1956, 113 – 126.
Kolmogorov, A.N., Sevastianov, B.A. 102. DAN 56, 1947, 783 – 786.
Koroliuk, V.S. 5. DAN 95, 1954, 443 – 446
6. IAN 19, 1955, 81 – 96.

10. DAN 108, 1956, 183 – 186.
Krein, M.G. 85. DAN 46, 1945, 339 – 342.
134. DAN 94, 1954, 13 – 16.
Kubilius, I.P. 10. DAN 100, 1955, 623 – 626.
14. UMN 11, No. 2 (68), 1956, 31 – 66.
Kuznetsov, P.I., Stratonovich, R.L., Tikhonov, V.I. 18. *Zhurnal Eksperimentaln. i Teoretich. Fiziki* 24, No. 1, 1954, 103 – 112.
20. DAN 94, 1954, 615 – 618.
21. DAN 97, 1954, 639 – 642.
Kvit, I.D. 1. DAN 71, 1950, 229 – 231.
Levin, B.R. 7. *Теория случайных* ... (Theory of Stochastic Processes and Its Application in Radio Engineering). M., 1957.
Liapunov, A.A. 35. UMN 6, No. 1 (11), 1951, 178 – 186.
Linnik, Yu.V. 43. IAN 13, 1949, 65 – 94.
51. DAN UzSSR 2, 1950, 3 – 6.
74. UMN 9, No. 4 (62), 1954, 203 – 210.
83. *Vestnik LGU* 13, 1956, 63 – 68.
91. TV 1, 1956, 466 – 478.
93. TV 2, 1957, 34 – 59.
95. DAN 113, 1957, 974 – 976.
101. DAN 116, 1957, 735 – 737.
Linnik, Yu.V., Sapogov, N.A. 44. IAN 13, 1949, 533 – 566.
Linnik, Yu.V., Sapogov, N.A., Sarmanov, O.V. Commentary on Markov's papers on probability theory. In Markov, A.A. *Избранные труды* (Sel. Works). N.p., 1951, 614 – 642. Translated in *Probability and Statistics. Russian papers.* Berlin, 2004. {Sarmanov only commented on some of Markov's memoirs (Ibidem, pp. 672 – 678) and he was not a coauthor of the general conmmentary.}
Mania, G.M. 1. DAN 69, 1949, 495 – 497.
2. *Uch. Zasp. Moskovsk. Gorodsk. Pedagogich. Inst.*, Fiz.-Matematich. Fak. No. 16 (3), 1951, 17 – 63.
Martynov, A.V. 1. DAN 113, 1957, 752 – 755.
Meisler, D.G. 3. *Sbornik Trudov Inst. Matematiki AN Ukr. SSR* 12, 1949, 31 – 35.
5. DAN Ukr. SSR 1, 1950, 3 – 10.
6. *Trudy IMM UzSSR* 10, No. 1, 1952, 96 – 105.
7. *Nauchn. Zapiski Lvov Politekhnich. Inst.*, ser. Fiz.-Matematich. No. 1, 1955, 24 – 44.
Meisler, D.G., Parasiuk, O.S., Rvacheva, E.L. 1. DAN 60, 1948, 1127 – 1128.
Mikhalevich, V.S. 3. DAN 85, 1952, 485 – 488.
4. UMZh 8, 1956, 454.
5. TV 1, 1956, 437 – 465.
Mitropolsky, A.K. 16. *Trudy L. Lesotekhnich. Akad.* 65, 1949, 73 – 92.
18. UMN 4, No. 5 (33), 1949, 142 – 175.
20. *Статистическое исчисление* (Statistical calculus), vol. 1. L., 1952.
22. Same, vol. 3. L., 1953.
Obukhov, A.M. 7. UMZh 6, 1954, 37 – 42.
8. *Trudy Geofizich. Inst.* 24 (151), 1954, 3 – 42.
Ozols, V.V. 1. IAN Latv. SSR 8 (109), 1956, 153 – 158.
2. Ibidem 10 (111), 141 – 152. (In Lettish)
Petrov, V.V. 1. *Vestnik LGU* 8, 1953, 13 – 25.
2. UMN 9, No. 1 (59), 1954, 41 – 62; 11, No. 2 (68), 1956, 250 – 251.
3. UMN 9, No. 4 (62), 1954, 195 – 202.
6. TV 1, 1956, 349 – 357.

7. DAN 115, 1957, 49 – 52.
Pinsker, M.S. 3. DAN 99, 1954, 213 – 216.
5. DAN 111, 1956, 753 – 756.
Pinsker, M.S., Yaglom, A.M. 1. DAN 90, 1953, 731 – 734.
Postnikov, A.G. 10. UMN 10, No. 1 (63), 1955, 147 – 149.
Prokhorov, Yu.V. 2. DAN 69, 1949, 607 – 610.
3. IAN 14, 1950, 523 – 536.
6. DAN 83, 1952, 797 – 800.
7. UMN 8, No. 3 (55), 1953, 135 – 142.
8. Ibidem, 165 – 167.
9. DAN 98, 1954, 535 – 538.
11. DAN 105, 1955, 645 – 647.
16. TV 1, 1956, 177 – 238.
Pugachev, V.S. 13. IAN 17, 1953, 401 – 420.
27. *Теория случайных* … (Theory of Random Functions and Its Application to Automatic Management). M., 1957.
Raikov, D.A. 2. DAN 14, 1937, 9 – 12.
8. IAN 2, 1938, 91 – 124.
Richter, V. (1957), DAN 115, No. 1.
--- (1957), TV 2, No. 2.
--- (1958), TV 3.
Romanovsky, V.I. 106. *Основные задачи* ... (Main Problems of the Theory of Errors). M. – L., 1947.
112. *Труды второго* … Same source as Boiarsky, 24 – 45.
117. *Bull SAGU* 30, 1949, 57 – 60.
122. On a method of double-level control. In *Взаимозаменяемость* … (Interchangeability and Control in Machine-Building). M., 1950.
Rosanov, Yu.A. 1. TV 2, 1957, 275 – 281.
2. DAN 116, 1957, 923 – 926.
--- (1958), UMN 13, No. 2 (80), 93 – 142.
Rosenknop, I.Z. 2. IAN 14, 1950, 95 – 100.
Rvacheva, E.L. 8. UMZh 4, 1952, 373 – 392.
11. *Uch. Zap. Lvov Univ*., ser. Mekh.-Math., 29 (6), 1954, 5 – 44.
Sapogov, N.A. 6. DAN 63, 1948, 487 – 490.
9. DAN 69, 1949, 133 – 135.
15. *Uch. Zap. LGU*, ser. Math., 19, 1950, 160 – 179.
21. IAN 15, 1951, 205 – 218.
Sarmanov, O.V. 7. DAN 59, 1948, 861 – 863.
8. Ibidem, 1061 – 1064.
9. DAN 60, 1948, 545 – 548.
10. Same source as Boiarsky, 90 – 91.
12. DAN 84, 1952, 887 – 890.
13. Ibidem, 1139 – 1142.
Sarymsakov, T.A. 30. *Trudy IMI AN UzSSR*, 10, No. 1, 1952, 13 – 18.
Sevastianov, B.A. 2. *Vestnik MGU* 3, 1948, 13 – 34.
34. DAN 59, 1948, 1407 – 1410.
4. UMN 6, No. 6 (46), 1951, 47 – 99.
8. TV2, 1957, 339 – 348.
Shirikorad, B.V. 1. IAN 18, 1954, 95 – 104.
Sirazhdinov, S.Kh. 10. DAN 84, 1952, 1143 – 1146.
11. DAN 98, 1954, 905 – 908.

12. *Предельные теоремы* … (Limit Theorems for Homogeneous Markov Chains). Tashkent, 1955.
13. *Trudy IMM AN UzSSR* 15, 1955, 41 – 56.
17. Ibidem 20, 1957, 89 – 100.
Skorokhod, A.V. 2. UMN 9, No. 2 (60), 1954, 189 – 190.
4. DAN 104, 1955, 364 – 367.
7. TV 1, 1956, 289 – 319.
8. DAN 106, 1956, 781 – 784.
10. TV 2, 1957, 145 – 177.
Skorokhod, A.V., Zolotarev, V.M. 8. UMN 10, No. 4, 1956, 181.
Slutsky, E.E. 34. *Таблицы* … (Tables for Calculating the Incomplete Γ-Function and the Probability Function χ^2). M. – L., 1950.
Smirnov, N.V. 3. *Metron* 12, No. 2, 1935, 59 – 81.
7. *Bull. MGU*, A, 1, No. 4, 1937, 1 – 12.
8. Ibidem, 2, No. 2, 1939, 3 – 14.
9. MS 6 (48), 1939, 3 – 24.
10. IAN 1939, 319 – 328.
12. UMN 10, 1944, 179 – 206.
--- (1948), Mathematical statistics. Translated in this book.
15. *Trudy MIAN* 25, 1949, 5 – 59.
17. UMN 4, No. 4 (32), 1949, 196 – 197.
18. DAN 74, 1950, 189 – 191.
19. *Uch. Zap. Moskovsk. Gorodsk. Pedagogich. Inst.*, Fiz.-Mat. Fak. 16 (3), 1951, 69 – 96.
20. *Trudy IMM ANUz SSR* 10, No. 1, 1953, 122 – 130.
21. *Vestnik LGU* 11, 1955, 45 – 48.
Smirnov's *Теория вероятностей и математическая статистика* … (Theory of Probability and Mathematical statistics. Sel. Works). M., 1970, contains reprints/translations of all the Items listed above excepting [21].
Sragovich, V.G. 2. DAN 111, 1956, 768 – 770.
Statuljavicus, V.A. 3. DAN 107, 1956, 516 – 519.
Tumanian, S.Kh. 1. DAN 94, 1954, 1011 – 1012.
4. TV 1, 1956, 131 – 145.
Ventsel, A.D. 1. DAN 111, 1956, 269 – 272.
Vorobiev, N.N. 9. *Trudy IMM AN UzSSR* 10, No. 1, 1952, 19 – 25.
Yaglom, A.M. 2. DAN 56, 1947, 347 – 349.
10. MS 24 (66), 1949, 457 – 462.
11. UMN 4, No. 4 (32), 1949, 173 – 178.
12. UMN 7, No. 5 (51), 1952, 3 – 168.
15. UMZh 6, 1954, 43 – 57.
17. DAN 98, 1954, 189 – 192.
18. MS 37 (79), 1955, 141 – 196.
19. *Trudy Moskovsk. Matematich. Obshchestvo* 4, 1955, 333 – 374.
24. TV 1, 1956, 161 – 167.
28. TV 2, 1957, 292 – 338.
Youshkevich, A.A. 1. UMN 8, No. 5 (57), 1953, 177 – 180.
3. TV 2, 1957, 187 – 213.
Zasukhin, V.N. 1. DAN 33, 1941, 435 – 437.
Zolotarev, V.M. 1. DAN 98, 1954, 735 – 738.
2. UMN 9, No. 2 (60), 1954, 147 – 156.
3. *Vestnik LGU* 1, 1956, 49 – 52.

Anderson, T.W., Darling, D.A. (1952), *Ann. Math. Stat*. 23, 193 – 212.
Blackwell, D. (1947), *Ann. Math. Stat.* 18, 105 – 110.
Czan, Li-Cyan (1956), *Acta Math. Sinica* 6, 55 – 81.
Doob, J.L. (1949), *Ann. Math. Stat.* 20, 393 – 403.
Darling, D.A. (1955), *Ann. Math. Stat*. 26, 1 – 20.
Girchik, M.A., Mosteller, F., Savage, L.J. (1946), *Ann. Math. Stat.* 17, 13 – 23.
Feller, W. (1948), *Ann. Math. Stat.* 19, 177 – 189.
Grubbs, F.E. (1950), *Ann. Math. Stat*. 21, 27 – 58.
Loeve, M. (1956), *Proc. Third Berkeley Symp. Math. Stat. & Prob.*, vol. 2, 177 – 194.
Wald, A. (1949), *Ann. Math. Stat*. 20, 195 – 205.

13. N.V. Smirnov. Mathematical Statistics: New Directions

Vestnik Akademii Nauk SSSR, No. 7, vol. 31, 1961, pp. 53 – 58 …

Foreword by Translator

The author did not mention demography which was either not considered important enough or thought to be too dangerous for allowing mathematicians to study it. The population between censuses was estimated, first and foremost, through lists of voters and police registration of residence, and the main characteristic of the population had been its class structure, see the anonymous article “Statistics of population” in *Bolshaia Sovetskaia Enziklopedia*, 2nd edition, vol. 40, 1957. A different description is in the third edition of this source (A.G. Volkov, Demographic statistics. *Great Sov. Enc*., vol. 8, 1975, this being a translation of the same volume (1972).

[1] The period of the 1920s – 1930s should be considered as the beginning of the modern stage in the development of mathematical statistics. Until then, mathematical, or variational statistics (and biometry) was usually understood as a narrow special discipline justifying the methods of studying biological phenomena of variability and heredity and of correlation of indicators; and substantiating the methodology of treating observations in agronomy, selection and forestry [1].

During the last decades a considerable widening and deepening of the subject-matter of research in mathematical statistics has been taking place and the pertinent investigations have already acquired an appreciable share in the general mathematical output. This fact is mainly due to the ever increasing demand for mathematical-statistical methods from almost all the branches of experimental science, from technical disciplines as well as from light and heavy industry. Indeed, in a very wide field of problems in natural sciences, technology and industry we encounter mass processes more or less influenced by random factors which cause the scattering of the results of experiments, measurements, trials or operations repeated many times under invariable conditions.

Given such a situation, an objective judgement on the regularities of the occurring processes, and, at the same time, the choice of a rational direction of practical activities (for example, when designing various buildings) are only possible on the basis of a statistical analysis of the pertinent data, trials or measurements. And this is exactly why statistical methods had been firmly established and became as though a fundamental part of modern studies in biology, anthropology {anthropometry}, meteorology, agronomy and similar sciences. They also begin to be instilled in medicine and psychology [2]; to play an appreciable part in industrial chemistry, in mechanical engineering and the instrument-making industry; in problems concerning the control of the quality of production and of technological processes.

In spite of the variety in the concrete conditions of the origin of issues and problems, the noticeable widening of their field still admitted a single mathematical interpretation. However, in a number of cases new investigations caused by direct practical requirements were also of fundamental importance since they fostered a better understanding of the cognitive aims and methods of mathematical statistics.

It ought to be noted that exactly the investigations concerning the inspection and rejection of expensive articles led Wald to the creation of a remarkable teaching of sequential analysis that did not conform to the previous understanding of the statistical science as a theory of purely cognitive estimates made by issuing from an already given data. It occurred that to avoid the loss of a considerable share of information, the compilation of statistical material (or the inspection of the objects of a given batch of produced articles, etc) should be planned and carried out after accounting for the results already achieved during each {previous} stage of the work. The compilation of the material (the sampling) is discontinued when the data obtained allow to make a decision ensuring that the probabilities of the possible errors of the first and of the second kind when testing a hypothesis, – of the errors measuring the risk of a wrong rejection of a correct hypothesis and of an acceptance of a hypothesis that does not really take place, – do not exceed certain boundaries established beforehand or securing that the greatest possible damage in case of a wrong decision be minimal (the minimax principle).

[2] The new approach to solving such problems led to another formulation of the main aims of mathematical statistics that stresses its active part characteristic of the theory of the most beneficial direction of practical activity under conditions of incomplete information on an occurring random process; of the theory of a rational choice from among those possible ensuring the least (in the mean) damage and the best use of the information available. This point of view, that Wald was the first to put forward, proved fruitful and allowed to unite into a single whole the previously developed sections of mathematical statistics (such as the theory of estimating parameters and of testing hypotheses) and the new ones, – the theory of statistical decisions and sequential analysis.

It is interesting that, in essence, this new understanding of the main aims of statistics makes use, in a more perfect and prudent way, of the old Bayesian concept [1] that issued from prior distributions. This characteristic feature of the new theory sharply separates it from the earlier concepts of Fisher and Neyman who resolutely (although not always with sufficient justification) had cut themselves off from any prior estimates and only issued from observational material.

In our national literature, along with fruitful investigations of acceptance inspection and of the estimation of the relative number of wrongly admitted defective articles, carried out in the spirit of the new ideas put forward by Kolmogorov [2; 3] and Sirazhdinov [4], very valuable findings concerning the optimal methods of quality inspection and ensuring the best economic results were due to Mikhalevich [5]. Aivazian [6] showed that the Wald sequential analysis allowed to reduce by two or three times the volume of observation as compared with the Neyman – Pearson optimal classical methodology.

Along with research into quality inspection, sequential analysis and the theory of decision functions, studies connected with automatic regulation and various problems in radio engineering constitute a substantial part of the modern statistical literature. Statistical and stochastic methods are being assumed as the basis for solving problems in analyzing and synthesizing various systems of automatic regulation. The regulation of the process of automatic manufacturing, of the work of automatic radars and computers demands an allowance for the continuously originating random perturbations and, consequently, calls for applying the modern theory of stochastic processes and of the statistical methodology of treating empirical materials based on that theory.

Statistical problems in radio engineering concerned with revealing signals against the background of interferences and noise constituting a stochastic process gave rise to a vast literature. The study of these and of many other problems caused by the requirements of modern technology (for example, by communication techniques as well as by problems in queuing theory; in studies of microroughness on the surfaces of articles; in designing reservoirs, etc), leads to the statistics of stochastic processes. This is certainly one of the most urgent and fruitful, but also of the most difficult areas of modern research. Until now, only separate statistical problems connected with testing the most simple hypotheses for Markov processes are more or less thoroughly studied. Thus, tests are constructed for checking simple hypotheses about transition probabilities or for studying the order of complication {generalization?} of a chain for some alternative, etc.

[3] The Scandinavian mathematicians Grenander and Rosenblatt [7] studied a number of statistical problems originating when examining stationary processes; from among these the estimation of the spectral density of a series by means of periodograms should be mentioned in the first instance.

The entire field of statistics of stochastic processes requires involved mathematical tools, and research often leads to results unexpected from the viewpoint of *usual* statistics of independent series of observations. Already Slutsky (1880 – 1948), the remarkable Soviet scholar, indicated this fact in his fundamental works devoted both to theoretical problems of studying stationary series with a discrete spectrum and especially in his investigations of concrete geophysical and geological issues.

The main channel of studies stimulated by the requirements of physics and technology nowadays lies exactly in the field of statistics of stochastic processes. The following fact testifies that this field excites interest. In September 1960 a conference on the theory of probability, mathematical statistics and their applications, organized by the Soviet and Lithuanian academies of sciences and the Vilnius University, was held in Vilnius. And, from among 88 reports and communications read out there, 26 were devoted to the applications and to problems connected with stochastic processes of various types (for example, the study of the roughness of the sea and the pitching {rolling? The author did not specify} of ships, design of spectral instruments, design and exploitation of power systems, issues in cybernetics, etc).

[4] Without attempting to offer any comprehensive idea about the entire variety of the ways of development of the modern statistical theory and its numerous applications in this note, we restrict our attention in the sequel to considering one direction that took shape during the last decades and is known in science as non-parametric statistics. Over the last years, distribution-free or non-parametric methods of testing statistical hypotheses were being actively developed by Soviet and foreign scientists and today they constitute a section of the statistical theory peculiar in its subject-matter and the methods applied.

The non-parametric treatment of the issues of hypotheses testing by sampling radically differs from the appropriate classical formulations where it was invariably assumed that the laws of distribution of the random variables under consideration belonged to some definite family of laws depending on a finite number of unknown parameters. Since the functional nature of these laws was assumed to be known from the very beginning, the aims of statistics were reduced to determining the most precise and reliable estimates of the parameters given the sample, and hypotheses under testing were formulated as some conditions which the unknown parameters had to obey. And the investigations carried out by the British Pearson – Fisher school assumed, often without due justification, strict normality of the random variables, Such an assumption essentially simplified mathematical calculations.

Exactly in such a way were determined the various *confidence* limits estimating the parameters by sampling and the criteria for testing hypotheses now constituting the main statistical tool described in all pertinent courses.

In our time, when statistical methods are being applied under conditions very unlike in nature one to another, the assumptions of the classical parametric statistics are unable to cover all the field of issues that we encounter. In practice, examining the distributions of random variables, we ought in many cases to restrict the problem only by very general suppositions (only assuming, for example, continuity, differentiability, etc). Tests or confidence estimates determined by issuing from these general premises were indeed designated non-parametric which stressed their distinction from their counterparts in classical statistics.

Practitioners have been applying some non-parametric methods for a long time. Thus, it was known how to obtain confidence limits for the theoretical quantiles of an unknown distribution function (under the sole assumption of continuity) given the terms of the variational series; and, in particular, how to estimate the position of the theoretical median. The application of the coefficients of rank correlation and of various tests of *randomness* based on the theory of runs were also known long ago.

Already during the 1930s – 1940s Soviet mathematicians achieved considerably deeper findings in the area of non-parametric statistics. Here, we only mention the remarkable test of the agreement between an empirical function of distribution $F_n(x)$ and the hypothetically admitted theoretical law $F(x)$. The appropriate theorem provides an asymptotic distribution of the criterion

$$D_n = \sup|F_n(x) - F(x)|$$

whose complete theory is based on a theorem due to Kolmogorov. Only continuity of $F(x)$ is here demanded and D obeys a universal law distribution independent of the type of $F(x)$. Similar tests independent of the type of the theoretical distribution function were later obtained in various forms and for various cases of testing hypotheses.

[5] The new direction attracted the attention of many eminent mathematicians and is now one of the most productive for the general development of statistical science. The independence from the kind of distribution enables to apply much more justifiably non-parametric tests in the most various situations. Such tests also possess a property very important for applications: they allow the treatment of data admitting either no quantitative expression at all (although capable of being ordered by magnitude) or only a quantitative estimate on a nominal scale. And the calculations demanded here are considerably simpler. True, the transition to non-parametric methods, especially for small samples, is connected with a rather essential loss of information and the efficiency of the new methods as compared with the classical methodology is sometimes low. However, the latest investigations (Pitman, Lehmann, Z. Birnbaum, Wolfowitz, van der Waerden, Smirnov, Chibisov and many others) show that there exist non-parametric tests which are hardly inferior in this respect to, and sometimes even better than optimal tests for certain alternatives.

Comparative efficiency is understood here as the ratio of the sample sizes for which the compared tests possess equal power for given alternatives (assuming of course that their significance levels are also equal). Thus, the Wilcoxon test concerning the shift of the location parameter under normality has a limiting efficiency of $e = 3/\pi \approx 0.95$ and is thus hardly inferior to the well-known Student criterion. And it was shown that under the same conditions the sequential non-parametric sign tests possesses a considerable advantage over the Student criterion: its efficiency is 1.3. A special investigation revealed the following important circumstance: supposing that the well-known classical χ^2 test demands n

observations for its power with regard to a certain class of alternatives to become not less than 1/2, the Kolmogorov criterion under the same conditions only demands $n^{4/5}$ observations. In other words, the limiting efficiency of the former test as compared with the latter is zero.

Non-parametric methods are now in the stage of intense development. The problems connected with the estimation of their comparative power and efficiency in different situations occurring in practice are far from being solved. A number of important statistical problems is not yet covered by non-parametric criteria (for example, the estimation of the agreement between a hypothetically given multidimensional law of distribution and the corresponding empirically observed distribution; of the discrepancy between two independent samples from a multidimensional general population, etc).

For a number of practical applications some criteria of the non-parametric type nevertheless demand a preliminary estimate of the parameters and new problems insufficiently studied even in the simplest cases (for example, already when checking the normality of the observed distribution of a given indicator) present themselves here. In addition, applications demand specification of asymptotic formulas and compilation of tables of the distribution functions of tests adapted for finite samples.

Many new problems whose solution is very difficult have appeared before non-parametric statistics in connection with the theory of stochastic processes which is the main channel of investigations stimulated by ever increasing demands made by physics and technology. In this comparatively young field of science the possibility of a fruitful application of statistical methods, possessing a wider scientific foundation and not restricted by the narrow assumptions of the classical methodology, appears as a significant and progressive phenomenon.

Notes

1. {The author said nothing about the studies of population, the main field of work of the Continental direction of statistics since the end of the 19th century (Lexis, Bortkiewicz, Chuprov, Markov).}

2. {Already in the 19th century, statistics became essential for several branches of medicine and psychology, see my papers in *Arch. Hist. Ex. Sci.*, vol. 26, 1982, and *Brit. J. Math., Stat. Psychology*, vol. 57, 2004.}

References

1. Kolmogorov, A.N. (1942), Determining the center of scattering and the measure of precision given a restricted number of observations. *Izvestia Akad. Nauk SSSR,* ser. Math., vol. 6, pp. 3 – 32. Translated in *Probability and Statistics. Russian papers of the Soviet Period.* Berlin, 2005.

2. Kolmogorov, A.N. (1950), Unbiased estimators. Translated in his *Sel. Works*, vol. 2. Dordrecht, 1992.

3. Kolmogorov, A.N. (1951), *Статистическтий приемочный контроль* ... (Statistical Acceptance Inspection When the Admissible Number of the Accepted Defective Articles Is Zero). L.

4. Sirazhdinov, S.Kh. (1957), Unbiased estimators of the relative number of accepted defective articles when applying the method of single sampling. *Trudy Inst. Matematiki i Mekhaniki Akad. Nauk Uzbek Soviet Socialist Republic,* vol. 20, pp. 89 – 100.

5. Mikhalevich, V.S. (1956), Sequential Bayesian solutions and the optimal methods of statistical acceptance inspection. *Teoria Veroiatnostei i Ee Primenenia*, vol. 1, 1956, pp. 437 – 465.

6. Aivazian, S.A. (1959), Comparison of the optimal properties of the Neyman – Pearson and the Wald criteria. Ibidem, vol. 4, 1959, pp. 86 – 93.

7. Grenander, U., Rosenblatt, M. (1955), *Statistical Analysis of Stationary Time Series.* New York.

14. A.N. Kolmogorov. Issues in the Theory of Probability and Mathematical Statistics

Report Made at the Branch of Mathematics, Academy of Sciences of the Soviet Union

Vestnik Akademii Nauk SSSR, No. 5, vol. 35, 1965, pp. 94 – 96 …

[1] I would like to begin the review of the present state and the main directions of the development of probability theory and mathematical statistics by mentioning that vol. 4 of S.N. Bernstein's *Собрание сочинений* (Coll. Works. N.p., 1964) containing {almost} all his writings on the theory of probability and statistics has appeared. Something is there nowadays interesting mainly for the history of science since it had been included, in a more clear form, in textbooks, but the store of ideas, far from being exhausted, and sometimes insufficiently known to young researchers, is also vast.

Issues belonging to the domain of limit theorems going back to Chebyshev and Liapunov, and essentially developed by Markov and Bernstein in the direction of studying dependent variables, which seemed for some time to be exhausted, experiences a period of new flourishing. V.A. Statuljavicus' report "Limit theorems in boundary value problems and some of their applications", read out on 29 Oct. 1964 at the General meeting of the Academy's section of mathematics on the theory of probability and mathematical statistics, was devoted to some pertinent issues.

[2] At the same meeting, A.A. Borovkov reported on a cycle of works which goes back to another current in the field of limit theorems apparently originated by Cramér, – to the so-called *theorems on large deviations*. Borovkov's works are beginning to show importance for applications in mathematical statistics and it is worthwhile to dwell somewhat on this point. The simplest typical problems of mathematical statistics contain two parameters, *significance level* α (the admissible probability of a mistaken judgement) and *n*, the number of observations. The approach based on limit theorems of the Chebyshev type corresponds to a limiting transition ($n \to \infty$) with a constant α. However, in practice *n* is often of the order of only a few hundred, or even a few dozens, and the significance level is usually chosen in the interval from 0.05 to 0.001. The number of problems demanding the guarantee of a very high *reliability*, *i.e.*, of a very small α, will probably increase ever more. Therefore, the formulas of the *theory of large deviations* corresponding to the asymptotic case at $\alpha \to 0$ are more often applicable.

[3] Markov originated the study of a vast class of stochastic processes now everywhere called Markov processes. After his time, our country continued to play a very large part in developing this direction, especially owing to the Dynkin school. Problems of obtaining the widest possible general conditions for the applicability of the main theorems of the theory of Markov processes; of ridding the theory of superfluous assumptions are still unsolved. However, I believe that the most essential work is here the search for new issues even if these do not demand the use of excessively refined mathematical tools but cover a wider field of applications.

In particular, an urgent issue is the study of only partly observable Markov processes, *i.e.*, processes of the type of $x(t) = \{x_1(t); x_2(t)\}$ where only the first component, $x_1(t)$, is observable. R.L. Stratonovich, in his theory of *conditional Markov processes*, formulated extremely interesting ideas about the approaches to solving the problems here encountered. Regrettably, his works sometimes lack not only any special mathematical refinement, but are often carried out on a level that does not guarantee a reasonable rigor not absolute, but securing against mistakes (A.N. Krylov's expression) [1]. A.D. Ventsel described in his report how a certain part of the theory of *conditional Markov processes* can be constructed with due rigor.

[4] The spectral theory of stationary stochastic processes whose rigorous foundation was laid in our country by {the late} Khinchin, is being intensively developed. Here, special attention, perhaps under the influence of Wiener's ideas, is paid now to the attempts at creating a spectral *non-linear* theory. This is indeed essential since the specialists in the fields of radio engineering, transmission of information, etc are inclined to apply spectral notions whereas only the linear theory, absolutely inadequate for many practically important applications, is yet mathematically worked out for the continuous spectra typical for the stochastic processes.

[5] In the area of information theory our scientists had to catch up with science abroad. We may assume that now this delay is made up for, and the works of Khinchin and of the representative of our younger generation, R.L. Dobrushin, have already occupied a prominent place in international science.

By its nature, information is not an exclusively stochastic notion. The initial idea of information as the number of binary symbols needed for isolating a certain object from among a finite number of objects has nothing in common with the theory of probability. Stochastic methods now only dominate the higher sections of the theory of information. It is possible, however, that the relationship between the two theories will radically change. I do not want to dwell here on this viewpoint (I am personally ever more attracted to it) according to which these relations may be reversed as compared with the present situation so that not probability theory will serve as a basis of the higher sections of the theory of information, but the concepts of the latter will form the foundation of the former.

[6] I only note the origin of the new branch of the *theory of dynamic systems*, *i.e.*, of the general theory of non-stochastic rigorously determinated processes where the ideas of the theory of information (beginning with the informational idea of *entropy*) play the main part. Extensive analogies between dynamic systems possessing the property of *intermixing* with stochastic processes were understood long ago. Now, however, in the works which I had begun and which V.A. Rokhlin and especially Ya.G. Sinai have continued, these similarities were essentially deepened. In particular, Sinai proved, under broad assumptions, and for some quite classical models (elastic balls in a box), the long-standing hypothesis on the asymptotically normal distribution of the *sojourn periods* for different sections of the phase space. For classical dynamic systems, defined by vector fields on compact manifolds, the two extreme instances, the *almost-periodic* case being studied by me and V.I. Arnold, and the case of *K-systems* with intermixing, are apparently the main ones in some sense.

[7] In mathematical statistics, in spite of many splendid investigations accomplished in the schools of N.V. Smirnov and Yu.V. Linnik, the work of Soviet mathematicians is yet far from being sufficient. As it seems, this situation is caused by the fact that the development of mathematical statistics is closely connected with the experience of direct contact with actual statistical material, whereas, for qualified Soviet mathematicians, such work with real data still remains although not rare, yet incidental and somewhat casual. Linnik reported on his remarkable accomplishments in solving difficult analytical problems appearing in mathematical statistics. Work on publishing mathematical tables required in statistical practice and on compiling a number of new tables is going on on a vast scale at the Steklov Mathematical Institute under Smirnov and L.N. Bolshev [2].

Some groups of mathematicians in Moscow, Leningrad and other cities are enthusiastically helping scientists of other specialities in solving practical problems in biology, geology, etc by statistical methods. But I have already mentioned that this work is somewhat casual, uncoordinated and sometimes amateurish. At a future conference, our branch ought to pay attention to the problem of organizing such work more rationally and wider.

Notes

1. {This is a paraphrase rather than a quotation from Krylov's Foreword to Chebyshev's lectures on probability theory published in 1936; translation: Berlin, 2004.}
2. {See Bolshev, L.N., Smirnov, N.V. (1968), *Таблицы математической статистики* (Tables of Mathematical Statistics). M.}

15. B.V. Gnedenko. Theory of Probability and Mathematical Statistics. Introduction
. In *История отечественной математики* (History of National Mathematics), vol. 4/2. Editor, I.Z. Stokalo. Kiev, 1970, pp. 7 – 13 …

Foreword by Translator

The following is a translation of the author's Introduction to the chapter on probability and statistics from a monograph on Soviet mathematics during 1917 – 1967. The main body of that chapter written by other authors was devoted to limit theorems and the theory of random processes. Concerning Lobachevsky whom Gnedenko mentioned see my Note 3 to Kolmogorov's paper of 1947 translated in this book.

[1] In Russia, the first investigations pertaining to probability theory date back to the beginning of the 19^{th} century when Lobachevsky, Ostrogradsky and Buniakovsky, on different occasions, had to solve a number of particular problems. Lobachevsky attempted to check by observations the geometric system that exercised dominion over the universe. Ostrogradsky examined some applied issues including acceptance inspection of goods delivered by providers. Buniakovsky also issued from the need to solve practical problems and he published a fundamental treatise (1846). This initial acquaintance with the theory of probability was a necessary and important period in developing an interest in this branch of mathematics in Russia.

The formulation and solution of general problems in the theory, and its initial formation as a vast mathematical science, characterized by a specific formulation of issues playing the main part in the entire domain of natural sciences, are connected with Chebyshev, Liapunov and Markov. By proving the law of large numbers {in a general setting} Chebyshev not only opened a general and important scientific regularity; he also provided an exceptionally simple and powerful method for the theory of probability and the entire field of mathematics. Later Markov perceived that the Chebyshev method allowed to establish {still} wider conditions for the applicability of the law of large numbers. The estimation of the probability, that the deviations of arithmetic means of independent random variables from the {appropriate} constants will not exceed the boundaries given beforehand, was a natural extension of investigating the conditions for the means to approximate a sequence of these constants.

By Chebyshev's time the classical findings of De Moivre and Laplace concerning the Bernoulli pattern were only generalized to sequences of independent trials with a variable probability of success. However, the theory of observational errors insistently demanded wider generalizations [1]. Laplace and Bessel surmised that, if the observational error was a sum of a very large number of errors, each of them being small as compared with the sum of the others, then its distribution should be close to normal. I do not know any {rigorous} mathematical findings made in this direction before Chebyshev. And, although his proof of the {central limit} theorem has logical flaws and the formulation of his theorem lacks the necessary restrictions, Chebyshev's merit in solving this issue is everlasting. It consists in that he was able, first, to develop a method of proof (the method of moments) [2]; second, to formulate the problem of establishing the rapidity of approximation and to discover asymptotic expansions; and third, to stress the importance of the theorem.

Incidentally, I note that soon after Chebyshev's work had appeared, Markov published two memoirs where he rigorously proved more general propositions. He applied the same method of investigation, – the method of moments. After Liapunov had made public two remarkable writings on the same subject, this method apparently lost its importance; indeed, whereas Chebyshev and Markov demanded that the terms {of the studied sum} possessed finite moments of all orders, Liapunov was able to establish such conditions of the theorem that only required a restricted number of moments (up to the third order and even somewhat lower). Liapunov's method was actually a prototype of the modern method of characteristic functions. Not only had he managed to prove the sufficiency of his conditions for the convergence of the distribution functions of the appropriately normed and centered sums of independent variables to the normal law; he also estimated the rapidity of the convergence.

Markov exerted great efforts to restore the *honor* of the method of moments. He succeeded by employing a very clever trick which is not infrequently used nowadays as well. Its essence consists in that, instead of a sequence of given random variables $\xi_1, \xi_2, \ldots$, we consider curtailed variables

$$\xi^*_n = \xi_n, \text{ if } |\xi_n| \le N_n, \text{ and } = 0 \text{ otherwise.}$$

The number N_n remains at our disposal, and, when it is sufficiently large, the equality $\xi^*_n = \xi_n$ holds with an overwhelming probability. Unlike the initial variables, the new ones possess moments of all orders, and Markov's previous results are applicable to them. An appropriate choice of the numbers N_n ensures that the sums of ξ^*_n and of the initial variables have approaching distribution functions. Markov was thus able to show that the method of moments allowed to derive all the Liapunov findings.

[2] In 1906 Markov initiated a cycle of investigations and thus opened up a new object of research in probability and its applications to natural sciences and technology. He began considering sequences of peculiarly dependent random variables (or trials) ξ_n. The dependence was such that the distribution of ξ_n, given the value taken by ξ_{n-1}, does not change once the values of ξ_k, $k < n - 1$, become known.

Markov only illustrated the idea of these chainwise dependences, which in our time enjoy various applications, by examples of the interchange of the vowels and consonants in long extracts from Russian poetry (Pushkin) and prose (S.T. Aksakov). In the new context of random variables connected in chain dependences he encountered the problems formulated by Chebyshev for sums of independent terms. The extension of the law of large numbers to such dependent variables proved not excessively difficult, but the justification of the central limit theorem was much more troublesome. The method of moments that Markov employed required the calculation of the central moments of all the integral orders for the sums

$$(1/B_n)\sum_{k=1}^{n} (\xi_k - \mathrm{E}\xi_k),\ B_n^2 = \mathrm{var}\sum_{k=1}^{n} \xi_k$$

and the proof of their convergence as $n \to \infty$ to the respective moments of the normal distribution.

In a number of cases Markov surmounted great calculational difficulties. On principle, even more important was that he substantiated new limit propositions, prototypes of the so-called ergodic theorems. For the Markov chains, the distribution of ξ_n as n increases ever less depends on the value taken by ξ_1: a remote state of the system ever less depends on its initial state.

The next direction of the theory of probability developed by Markov and other researchers before the Great October Socialist Revolution {before the Nov. 7, 1917, new style, Bolshevist coup} is connected with the construction of the theory of errors. Astronomers paid much attention to this subject, and their contribution was not restricted to methodologically improving the exposition of already known results **3**.

During the 19th, and the beginning of the 20th century, Buniakovsky {1846}, Tikhomandritsky {1898}, Ermakov {1878}, Markov (1900) and Bernstein {1911} compiled textbooks on probability theory on the level corresponding to the contemporary state of that science. Markov's textbook played a considerable part in developing probability theory in our country. He explicated a number of findings in sufficient detail, and, in the same time, in an elementary way **4** which fostered the readers' interest not only in passive learning, but in active reasoning as well. Already in its first edition, Bernstein's book, distinguished by many peculiar traits, for a long time exerted considerable influence. Then, Slutsky (1912) acquainted his Russian readers with the new issues in mathematical statistics that had originated in England in the first decade of the 20th century.

The works of the two mathematicians, Bernstein and Slutsky, who played an important part in building up new directions of research in probability theory and mathematical statistics in our country **5**, began to appear in the years immediately preceding the Revolution. During the first period of his work, Bernstein examined such important issues as the refinement of the {De Moivre –} Laplace theorem, the logical justification of probability theory, and the transfer of its peculiar methods to problems in the theory of functions. It was in this very period that he was able to discover a remarkable proof of the Weierstrass theorem (1912). {At the time,} Slutsky studied problems in mathematical statistics chiefly connected with correlation theory.

[3] Thus, already before the Revolution, scientific pre-requisites for the development of probability theory were created in our country. And the establishment, after the Revolution, of a vast network of academic and research institutes and of academies of sciences in the Union Republics **6** fostered the growth of scientific investigations in many cities as well as the creation of considerable mathematical bodies and the initiation of new directions of research.

In the then young Central Asian University {Tashkent} Romanovsky established a prominent school of mathematical statistics and the theory of Markov chains. In Moscow, in the nation's oldest university, the well-known school of the theory of probability was created on the basis of the school of the theory of functions of a real variable. It is difficult to overestimate its influence on the development of probability theory during the latest decades. The construction of the foundation of the theory; a vast development of the classical issues concerning limit theorems for sums of independent variables; the concept of stochastic processes (without aftereffect; stationary and with stationary increments, branching processes); the development of methods of statistical physics; of queuing, reliability and information theories; and many other issues are the subject of research done by Moscow specialists. The beginning of stochastic investigations in Moscow was connected with two outstanding mathematicians, Khinchin and Kolmogorov.

In Kiev, in the 1930s, N.M. Krylov and N.N. Bogoliubov began their study of ergodic theorems for Markov chains. They issued from the theory of dynamic systems and the direct cause of their research was the desire to justify the ergodic hypothesis formulated already by Boltzmann. Later, their problems widened and adjoined the work of the Moscow mathematicians.

[4] After the Great Patriotic War {1941 – 1945} Linnik in Leningrad and his students in Vilnius built up new powerful collectives working in various directions of the theory of probability and mathematical statistics. During the Soviet years a large number of monographs and textbooks in probability theory were published. Some of them won international recognition, have went and are going through many editions abroad [7]. Collected translations of papers of Soviet authors on probability theory and mathematical statistics regularly appear in the USA. The periodical *Теория вероятностей и ее применения* (Theory of Probability and Its Applications) is being translated and published there in its entirety. In mathematical statistics, the books Smirnov & Dunin-Barkovsky (1955; 1959) [8] and Linnik (1966) should be mentioned. Bolshev & Smirnov (1968) compiled excellent tables, and a number of tables were due to Slutsky, Smirnov and others.

The specialized periodical mentioned above was established owing to the considerable increase in the amount of investigations in probability theory and mathematical statistics. In addition, writings on probability continue to appear in general mathematical and various special editions. During the latest years, the number of articles devoted to the theory of probability and published in engineering journals greatly increased which undoubtedly testifies that the demands of modern technology and theoretical studies accomplished by mathematicians are connected with each other and that the engineers' level of stochastic education has risen.

[5] The contribution of Soviet scientists to the development of the theory of probability deserves highest appraisal. By creating an axiomatics of the theory on the basis of the theories of measure and of functions of a real variable, they fundamentally transformed the science and laid a robust foundation for the development of its new branches. The first works initiating the creation of the theory of stationary processes and random functions led to the focus of scientific attention in probability theory shifting to this very field. In addition, owing to the development of general methods of the theory of random processes and fields, great possibilities for studying phenomena in nature and economics as well as technological operations have opened up. Already the first steps in this direction yielded appreciable results. The classical issue of summing independent random variables was essentially promoted, and, in some aspects, settled. At the same time, new problems imparted freshness and fascination to this venerable subject.

The intensive progress in technology and physics advanced many unexpected problems and initiated absolutely new directions of research, in the first place information theory (that originated in the USA), the theory of stochastic automatic machines, theory of optimal control of stochastic processes, reliability theory. Soviet mathematicians have seriously contributed to the development of these new domains as well.

Notes

1. {Only the Laplacean theory of errors.}

2. {The method of moments is due to Bienaymé and Chebyshev, and Gnedenko himself said so later (Gnedenko & Sheynin 1978, p. 262).}

3. {I believe that the theory of errors is the statistical method as applied to the treatment of observations, and (unlike Gnedenko) that it had been essentially completed by the end of the 19th century. True, the spread of triangulation over great territories as well as the new technology demanded the solution of many practical problems.}

4. {Markov's textbook does not make easy reading. Recall his own words (Ltter to Chuprov of 1910; Ondar 1977, p. 21): *I have often heard that my presentation is not sufficiently clear*.}

5. {Gnedenko had greatly enhanced his appraisal of Slutsky by inserting his portrait, the only one except for Kolmogorov's.}

6. {Before ca. 1989 no academy of sciences ever existed in Soviet Russia, by far the largest union republic.}

7. {Gnedenko mentioned books of several eminent mathematicians (Kolmogorov, Khinchin, Linnik, Skorokhod, Dynkin and himself). Since then, many books, notably Kolmogorov's *Selected Works*, have appeared in translation abroad. One of these was the book mentioned in [7].}

8. {The authors of the latter book were mentioned on the title-page in the opposite order, see References.}

References

1. Bernstein, S.N. (1911), *Теория вероятностей* (Theory of Probability). Kharkov. Mimeographed edition. Four ordinary editions (1927 – 1946).

2. --- (1912, in French), Demonstration of the Weierstrass theorem based on the theory of probability. Russian translation: *Собрание сочинений* (Coll. Works), vol. 1. N.p., 1952, pp. 105 – 107. English translation: *Math. Scientist*, vol. 29, 2004, pp. 127 – 128.

3. Bolshev, L.N., Smirnov, N.V. (1968), *Таблицы математической статистики* (Tables of Mathematical Statistics). M.

4. Buniakovsky, V.Ya. (1846), *Основания математической теории вероятностей* (Principles of the Mathematical Theory of Probability). Petersburg.

5. Dunin-Barkovsky, I.V., Smirnov, N.V. (1955), *Теория вероятностей и математическая статистика в технике* (Theory of Probability and Mathematical Statistics in Technical Sciences). M.

6. Ermakov, V.P. (1878; 1879), *Элементарный курс теории вероятностей* (Elementary Course in the Theory of probability). Kiev.

7. Gnedenko, B.V., Sheynin, O. (1978, in Russian), Theory of probability, this being a chapter in *Mathematics of the 19 th Century*, vol. 1. Editors, A.N. Kolmogorov. A.P. Youshkevich. Basel, 1992 and 2001, pp. 211 – 288.

8. Kolmogorov, A.N. (1948, in Russian), Slutsky, an obituary. *Math. Scientist*, vol. 27, 2002, pp. 67 – 74.

9. Linnik, Yu.V. (1966), *Статистические задачи с мешающими параметрами* (Statistical problems with Nuisance Parameters). M.

10. Markov, A.A. (1900), *Исчисление вероятностей* (Calculus of Probability). Later Russian editions: 1908, 1913, 1924. German translation: 1912.

11. Ondar, Kh.O., Editor (1977, in Russian), *Correspondence between A.A. Markov and A.A. Chuprov*. New York, 1981.

12. Slutsky, E.E. (1912), *Теория корреляции и элементы учения о кривых распределения* (Theory of Correlation and the Elements of the Theory of Distribution Curves). Kiev.

13. Smirnov, N.V., Dunin-Barkovsky, I.V. (1959), in Russian). *Mathematische Statistik in der Technik*. Berlin, 1969, 1973.

14. Tikhomandritsky, M.A. (1898), *Курс теории вероятностей* (Course in Theory of Probability Theory). Kharkov.