Probability and Statistics

Russian papers
Selected and Translated by Oscar Sheynin
Berlin 2004

Contents

19. Sheynin, O. On the probability-theoretic heritage of Cournot. *Istoriko- Matematich. Issledovania*, vol. 7 (42), 2002, pp. 301 – 316.
20. Gnedenko, B.V. A review of the current state of the theory of limiting laws of independent terms. *Uchenye Zapiski Tomsk. Gosudarstven. Pedagogich. Inst.*, No. 1, 1939, pp. 5 – 28.

0. Foreword

I am presenting a collection of translations of Russian papers on probability and statistics important at least for the history of these disciplines. One of Gnedenko's papers included here treats comparatively new developments, which, however, might be considered as a continuation of the works of the 19th century. I also included two my own papers one of which (on Cournot) might seem to be out of the general context, but it describes the work of an outstanding Western scientist who at the very least influenced both contemporaneous (Davidov) and later (Chuprov) Russian scholars. Along with the abovementioned work of Gnedenko, it forms the *Supplement* to this collection whereas the first few items bearing on the early period in Russian probability constitute *Part 1*. The main body of this book is given over to *Part 2*, describing some aspects of the Petersburg school of probability whose main representatives were Chebyshev, Markov and Liapunov.

In many 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 by numbers in brackets, for example thus: **[2]**. My own comments are in curly brackets and I often referred to my articles in the *Archive for History of Exact Sciences* although I have recently generalized my contributions in a single book, *History of the Theory of Probability to the 20th century.* Berlin, NG Verlag, 2004.

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 publication remained with me. Note that those collections also contain not reprinted material.

I am rendering the terms Central limit theorem, Law of large numbers, and Method of least squares as CLT, LLN, and MLSq, respectively, and in the lists of references I am using the following abbreviation:

Akad. Nauk SSSR (Union Soviet Socialist Republics) = AN SU
Istoriko-Matematich. Issledovania = IMI
Imp. Akad. Nauk (or *Acad. Imp. Sci.*) = AN Psb
Matematich. Sbornik = MS
L, M, Psb, (R) = Leningrad, Moscow, Petersburg, in Russian.
Note: During 1914 – 1924 Petersburg was called Petrograd.

It is hardly amiss to add that, first, the same publisher (NG Verlag) has put out, in 2004, in my translation,

1. A collection of my papers initially published in Russian: *Russian Papers on the History of Probability and Statistics*.
2. Chebyshev, P.L. [Lectures in] *The Theory of Probability.*
3. *Chuprov, A. Statistical Papers and Memorial Publications.*
4. Nekrasov, P.A. *Theory of Probability* (collected papers, mostly of Nekrasov himself, Markov and Liapunov).

Second, a few of my translations have appeared/will appear in periodicals:

1. Bernstein, S.N. (1947), Chebyshev and his influence on the development of mathematics. *Math. Scientist*, vol. 26, 2001, pp. 63 – 73.
2. Kolmogorov, A.N. (1948), E.E. Slutsky, an obituary. Ibidem, vol. 27, 2002, pp. 67 – 74.

3. Bernstein, S.N. (1912), Demonstration of the Weierstrass theorem based on the theory of probability. Ibidem, to appear in 2004.
4. Sheynin, O. (1997), Markov and life insurance. Ibidem, to appear in 2005.
5. Khinchin, A.Ya. (1961), Mises' frequentist theory, etc. *Science in Context*. To appear in 2004.
6. Soloviev, A.D. (1997), Nekrasov and the central limit theorem. In Youshkevich: Memorial Volume. Paris. To appear in 2005.

Acknowledgement. Dr. A.L. Dmitriev (Petersburg) sent me photostat copies of papers NNo. 15 and 20 published in sources hardly available outside Russia.

Part 1. The Early Period

1. Paevsky, V.V. Euler's work in demography.

In collected papers *Leonard Euler*. Moscow – Leningrad, 1935, pp. 103 – 110 …

[1] It is known that by no means had Euler's amazing genius only manifested itself in the field of pure mathematics. Along with his most fundamental mathematical works contained in the surprisingly long list of his contributions, it is possible to find many writings on mechanics, on some sections of physics (optics, etc), astronomy, exterior ballistics, on the theory of music, the theory of some games, and on many other branches of science. Less widely known are his works in the field of mathematical statistics, or, more correctly, of one of its parts, mathematical demography. Nevertheless, Euler was in essence the first in this area to lay quite a distinct mathematical foundation for a number of the main demographic concepts such as the sequence of extinction (mortality tables), increase in population, the period of its doubling, etc. In addition, Euler formulated, as lucid as possible, the guiding rules for establishing the institution of private insurance (of life insurance in all its forms).

Here, I intend to touch on Euler's work in this field. It is known that, in spite of his great fame among his contemporaries, Euler's quite exceptional scientific fertility often led to his being unable to find publishers for his permanently appearing new works. As a result, his writings became scattered in a great number of various editions. In particular, those devoted to demographic problems, now collected together in his *Opera omnia* [1], had first appeared in the memoirs of the Berlin, or of our Academy, and in his {previous} collected works, whereas some of them remained unpublished and were only revealed to demographers in this source.

[2] His *Recherches* [2] and *Sur la multiplication* [3] deserve maximal attention from among his contributions. In the first one Euler begins by clearly formulating the concept of the *sequence of extinction or survival* that underpins the entire modern teaching of mortality tables. Then, in a series of *Questions* and answers to them, he expounds the entire methodology of applying them. Here, Euler in essence sets forth, with unusual clarity and logical completeness, the whole main part of the mathematical (or, of the so-called formal) theory of mortality. In addition, he introduces, for the first time ever, the new concept of probable duration of life [1].

Along with a theoretical treatment of the measurement of mortality, Euler solves a number of practical problems belonging to insurance mathematics. His pertinent numerical calculations were based on the then famous Kersseboom [2] mortality table of 1742 that he reproduces here.

In the second part of the *Recherches* Euler considers the increase in population. After precisely defining the concept of increase, he solves, again by {methodologically} applying *Questions*, several problems on determining this increase, on the future population, on the number of deceased, etc. Here also he dwells on the issue of the so-called laws of mortality.

It should be noted that, whereas Euler, in the first part of his memoir, provided a quite complete theory of formally measuring mortality at different ages, to which all the subsequent science of demography had not added anything essentially new, he introduced simplified assumptions on birth-rate in its second part.

Having no results of censuses (in the modern sense of this word) at hand, Euler was unable to recognize all the importance of the problem of fertility as a function of age, or of the great variability of the coefficients of reproduction at different ages. As a result, from our point of view his proposition, on which he based his determination of the increase in population with time, proved too simplified.

In his calculations, Euler supposed that the birth-rate was constant. He had not noticed that, generally speaking, this was not true even for fertility invariable over different ages. Indeed, the age structure of the population is determined by past events and changes inevitably. Nevertheless, his investigation led Euler to the celebrated proposition on the increase in population (in the absence of perturbation factors) by a geometric progression. Even now, this principle (true, under a number of restrictive conditions) is usually taken as the basis of pertinent hypotheses. As stated above, Euler was unable to make use of censuses, but his great intellect prompted him to sense the importance of their data for solving problems in measuring mortality.

In the concluding sentences of the *Recherches*, Euler, without using the word *census*, nevertheless arrives at a quite modern idea of constructing mortality tables by means of data on the age distribution of the living and the deceased (for a given year).

The second mathematical-statistical memoir, *Sur la multiplication*, is closely connected with Euler's statistical treatment of some chapters of Pastor Süssmilch's famous book [4]. Being a son of his century, and, besides, having grown up and being reared in a pastor's family, Euler time and time again participated in argumentative defence of Christianity. It is ascertained that he played a great part in the revision of the second edition of the *Ordnung* [4], and that, apparently, he wrote to a considerable degree its Chapter 8, *Von der Geschwindigkeit der Vermehrung und von der Zeit der Verdoppelung* … And his memoir *Sur la multiplication* (written partly in French and partly in Latin) is an expanded mathematical version of the materials on which this chapter of Süssmilch's *Ordnung* was based.

Both these writings (*Von der Geschwindigkeit* and the French – Latin memoir) contain a number of extremely interesting and profound propositions belonging to mathematical demography. Their main subject-matter is the determination of the time needed for a closed (*i.e.*, not liable to migrations) population to double under various hypotheses on mortality and birth-rates. Without going into detail, I note that Euler had here intentionally simplified his main assumptions so as to make their mathematical corollaries as simple and as clear as possible. He had not considered the age composition of the population as a function of two variables (time-epoch and time-age) as it is done nowadays, but, instead, issued for example from the following artificially simplified pattern of the *structure* of the population

1) All marriages (Euler only dwelt on marital fertility) are contracted exactly at age 20 (at the same age for man and wife).

2) Each marriage yields six living children, three boys and three girls.

3) All children are twins: a boy and a girl are born each time when their parents are exactly 22, 24 and 26 years old.

4) Nobody dies until age 40, but everyone dies at 40.

Under these conditions, beginning from the time when the husband and the wife comprising the first pair were exactly 22 years old each, the number of births which occurred bi-annually form the following sequence:

2, 2, 2, 0, 0, 0, 0, 0, 0, 0, 0, 2, 4, 6, 4, 2,0, 0, 0, 0, 0, 0, 2, 6, 12, 14, 12, 6, 2,
0, 0, 0, 0, 2, 8, 20, 32, 38, 32, 20, 8, 2, 0, 0, 2, 10, 30, 60, 90, 102, 90, 60,

30, 10, 2,2, 12, 42, 100, 180, 252, 282, 252, 180, 100,42, 14, …

Euler states that the terms of such sequences are the coefficients of the expansion 8achieved by division) of some generating function and indicates the form of this function for another numerical example [3, p. 551]. He [4, Chapter 8, §161] also notes that, in spite of its extreme irregularity which becomes evident when considering only the first terms, the sequence, when continued sufficiently far, begins to *transform itself* into a geometric progression, and all its irregularity disappears:

However irregular these sequences appear initially, after being ever continued they will nevertheless finally turn into geometric progressions, and the irregularities noticed at the beginning will decrease the more, the further, until finally disappearing almost completely.

About 160 years later, the German {German – American} statistician and mathematician Gumbel [5] proved Euler's statement for the abovementioned sequence by indicating its generating function, $2/(1 - x^{11} - x^{12} - x^{13})$. He also showed that the common ratio of its limiting geometric progression was the greatest in absolute value root, $x = 1.0961$ **[3]**, of the appropriate equation,

$$x^{13} - x^2 - 1 = 0.$$

In each of his writings, under the most various hypotheses on birth-rates and mortality, Euler invariably concluded that for a closed population the number of births (and therefore the population itself) grows by a geometric progression. Of course, modern demographic science cannot entirely agree with such a statement. The growth of a given population depends on so many factors, most of them unyielding to interpretation and being of a social nature, that it is hardly possible to hope to represent it by a mathematical formula, let alone by such a simple expression. However, modern mathematical demography does not decline to study additionally abstract patterns of the changes in the population and its structure, although by no means attaching to them the rank of predictions and considering them only as an aid for determining the most perfect indicators of the changes observed in populations. And the patterns themselves are now remote from those primitive and simple propositions which Euler had to accept as his basis. This applies above all to the concept of the age structure of the population that modern demographers got used to consider as continuously changeable in time-epoch.

It is interesting, however, that, assuming that some demographic factors (as mortality and fertility at different ages) are constant, the most subtle methods of constructing abstract patterns of the changes in population largely lead to the same Eulerian conclusions about its limiting growth by geometric progressions Such, for example, was the result attained by the most eminent American demographer Lotka who made use of integral equations of the Volterra type, see [7; 8], or by the German researchers Bonz & Hilburg [9] who determined the growth as the solution of a difference equation.

[3] In addition to Euler's purely demographic works, as they might be termed, see above, he left a number of memoirs devoted to practical issues of insurance:

1. *On life annuities* [10]. After defining the main notions (*i.e.*, mortality and discount) that form the basis of any insurance operation, Euler offers recurrence formulas determining the current cost of life (*postnumerando*) annuities and provides a number of practical tables which he constructed on the basis of the Kersseboom mortality table.

2. *On widow funds* [11]. Here, Euler considers problems connected with insurance depending on the survival or death of two persons (on *two lives*), and with life insurance in

general. He comes close to the concept of *commutative numbers* later introduced into the science of insurance by Tetens and very important for modern actuaries.

3. *On public institutions for widows and in case of death* [12]. This is Euler's most extensive writing on insurance mathematics. It consists of three parts

a) On widow pensions.

b) On mutual insurance against death; and

c) On the plan of a new tontine (which provided a possibility of new members being admitted continuously).

This is a vast treatise on insurance combining much written by Euler previously and supplied with a large number of tables not devoid of interest even nowadays.

4. *On insuring orphans* [13]. Here, Euler determines the cost of insurance to be received by the heirs upon the death of both their parents. Once more making use of the Kersseboom mortality table, he solves a number of separate pertinent problems by means of one general equation.

5. A memoir (more precisely, a draft fragment first published in the *Opera omnia*) on calculating tontine annuities [14]. Here. Euler considers various modifications of tontines. The fragment apparently represented an examination of a tontine proposal formulated by him on the request of Friedrich II.

[4] Èuler had always been interested in problems of mortality, birth-rates and growth of population; he repeatedly turned to them in ad hoc memoirs and they also served him as examples in his purely mathematical writings. Thus, in Chapter 6 of his *Introductio* [15] Euler supplied four demographic (although extremely simple) problems as illustrations for calculations. The following one is characteristic for his time:

> *After the Flood, mankind was reduced to six people. Supposing that in 200 years the population had increased up to 1 000 000, determine the yearly rate of its growth.*

It might be thought that, being a remarkable representative of a purely mathematical genius, Euler will restrict his efforts with a formal mathematical treatment of his problems. Actually, however, the situation was much different. His profound intellect compelled him to penetrate into the very essence of the problems under his study. He understood perfectly well the relation between his patterns and the facts of concrete reality and made a number of extremely interesting remarks on this subject. Thus, concerning mortality tables he warned against the desire to believe that the determined order of extinction was universal. He noted that each table was only applicable to that city, or province, for which it was constructed. He also stated that mortality in towns must be lower than in the large cities, but higher than in the villages, and this proposition was apparently true for his epoch.

Euler then remarked that the order of extinction calculated by means of mortality tables ceased to be applicable in any cases of epidemics, famine or war. He understood perfectly well the importance of migratory factors in constructing mortality tables, correctly allowed for the difference in mortality of men and women, and repeatedly indicated that the application of mortality tables constructed for both sexes taken together could lead to large errors,– he thus criticized the Kersseboom table.

Not only did Euler, time and time again, return to the problem of the comparatively lower mortality of women, but, what is especially remarkable, he had a perfect understanding of purely demographic problems of mortality in early childhood (during the first months of life). On this subject he formulated well-thought out ideas often escaping the notice even of modern specialists in public hygiene when they compare the mortality of all children and of those constituting isolated groups (in nursery schools and crèches).

Euler noted that a mortality table based on observing life annuitants did not express the actual mortality of the general population, the less so for infants younger than one year. The observations, as he stated, undoubtedly only covered children who had escaped the dangers of the first months of life, so that the table represented the mortality of a *selected collective of children* rather than of the entire children population. Indeed [2, pp. 86 – 87]:

In this respect one should not apply the records of the life annuities which begin with infants older than a year. Because, first of all, these infants cannot be considered as newly born, and most of them have undoubtedly escaped the dangers of the first months; and, further, infants of feeble frame often are not included at all, so that one should regard the infants, for whom life annuities were taken out, as selected …

And, considering the numbers of survivors according to the Kersseboom table, Euler added:

Therefore, because the table is constructed for selected infants who had already survived several months after birth – when wishing to apply it to all the newly born infants of a city or province, all its numbers should be decreased by a certain part to allow for the high mortality, to which the infants are at once subjected after being born.

How remote are these considerations from dry and formal mathematical patterns! Euler, the mathematical genius, shows here his features of a real demographer, deeply penetrating into the concrete essence of the problem under study.

Naturally, not everything created by Euler in the field of demography retains absolute importance for our time. Something became obsolete and is only of a historic interest. However, he by right occupies a distinguished place in the cohort of remarkable mathematicians who participated in studying demographic and related problems, in the cohort that had begun with De Moivre and included himself, Fourier and Laplace, and to which Gauss was not alien either; to which Volterra, von Mises, Fréchet and many others have recently joined.

Notes

1. {Huygens effectively applied this notion in his correspondence of 1669, see my paper in the *Arch.Hist. Ex. Sci*., vol. 17, 1977, pp. 247 – 248.

2. The Dutch scientist Wiliem Kersseboom (1691 – 1771) held financial posts and constructed his mortality table, common for both sexes, by basing his calculations on observations of the longevity of persons who had bought annuities for themselves from the government.

3. Gumbel omitted the numerical factor (2) in the generating function [6,p. XL].

References

1. Euler, L. (1923), *Opera omnia*, ser. 1, t. 7. Leipzig – Berlin.

2. --- (1767), Recherches générales sur la mortalité et la multiplication du genre humain [1, pp. 79 – 100].

3. --- (Manuscript), Sur la multiplication du genre humain [1, pp. 545 – 552].

4. Süssmilch, J.P. (1761 – 1762), *Die Göttliche Ordnung in den Veränderungen des menschlichen Geschlechts*, etc,Bde 1 – 2. Second edition. Göttingen – Augsburg, 1988.

5. Gumbel, E.J. (1916), Eine Darstellung statistischer Reihen durch Euler. *Jahresber. Deutsch. Mathematikervereinigung*, Bd. 25, pp. 251 – 264.
6. Du Pasquier, L.G. (1923), Préface to [1].
7. Dublin, L.I., Lotka, A.J. (1925), On the true rate of natural increase. *J. Amer. Stat. Assoc.*, new ser., vol. 20, pp. 305 – 339.
8. Sharpe, F.R., Lotka, A.J. (1911), A problem in age distribution. *London, Edinb. and Dublin Phil. Mag.*, 6th ser., vol. 21, p. 435ff.
9. Bonz, F., Hilburg, F. (1931), Die voraussichtliche Bevölkerungs-entwicklung in Deutschland. *Z. angew. Math. Mech.*, Bd. 11, pp. 237 – 243.
10. Euler, L. (1767), Sur les rentes viagères [1, pp. 101 – 112].
11. --- (1770), Nöthige Berechnung zur Einrichtung einer Witwencasse [1, pp. 153 – 161].
12. Fuss, N., sous la direction de Euler (1776), Eclaircissements sur les établissements publics en faveur tant des veuves que les morts, avec la description d'une nouvelle espèce de Tontine aussi favorable au Public qu'utile à l'Etat [1, pp. 181 – 245].
13. Euler, L. (1785), Solution questionis ad calculum probabilitatis : Quantum duo persolvere debeant, etc. [1, pp. 393 – 407].
14. --- (Manuscript), Sur le calcul des rentes tontinières [1, pp. 553 – 577].
15. --- (1748), *Introductio in analysin infinitorum*, t. 1. German translation: Berlin, 1885.

2. Daniel Bernoulli. On the Mean Duration of Marriages for Any Age of the Spouses and on Other Related Issues

Translated from Russian translation in Ptukha, M.V. *История статистики* (History of Statistics), vol. 1. Moscow, 1955, pp. 453 – 464...

Translator's Foreword

I have translated Bernoulli's memoir [3] from its Russian translation (see *Contents*) retaining the several Latin phrases that were left in its Russian text. With respect to the mathematical background needed here, Bernoulli referred to his previous work [2] calling it *preliminary* (§16). There, he attacked the same issue formulating it, however, as an urn problem and solved it by two methods,– by a combinatorial approach and by applying differential equations [4]. For that matter, he was the first to use such equations in probability in his memoir of 1738 generally known owing to the introduction there of moral expectation.

I have not fully reproduced the tables; readers can find them in the original Latin version of the memoir.

* * *

1. Since a perpetual substitution [*commutatio*] of a person who has to die by one born is going on, very many years ago various nations had compiled tables both of those born, and of the deceased, by whose means the laws of variations and reciprocities were observed and finally established with a surprising success. Because, although the fate of each separate man is utterly unknown, it cannot nevertheless be denied that, for a large arbitrarily taken number of things, the mean state corresponds to nearly invariable laws regardless of what this state concerns or of the issue at hand whatsoever.

Thus, it was observed that the number of the yearly born boys was invariably greater than that of the yearly born girls, and, what was more surprising, in each country the inequality itself between these numbers almost invariably retained one and the same proportion with respect to the total number of newly born. However, this last {property} can only be noticed for very large numbers in which the random fate of one or another phenomenon occurs to be hardly noticeable if only bearing in mind its ratio to the total quantity. Even these very proportions for one and for the other sex little differ from each other in different countries.

The most thoroughly compiled tables indicate, among other facts, the causes of death wherefrom we perfectly well acquaint ourselves with the distinctive peculiarities of deadly diseases. It is known now that in our days smallpox alone destroys a twelfth or a thirteenth part of each generation – either more or less depending on the way of life and the differences between the nations. It is also known that during the first year of life diseases of the first childrens' age carry off almost three tenth of the entire generation.

Then, what is closer to our goal, it was observed that in general women enjoy a longer life than men; regarding this, we have a table compiled by the illustrious Wargentin in Sweden which quite confirms that observation. And this last-mentioned circumstance cannot be attributed to the different ways of life led by men and women because the women's mentioned privilege most invariably reveals itself from the very first nappies and holds good with them during all their life. 1623 boys and 1438 girls die during the first year of life from among one and the same number of persons of each sex. The mean duration of life taken from the very day of birth is 24 years and 2 months for boys and 28 years and 10 months for girls; 2008 souls of the male sex live to the twentieth year and 2337 of the female sex. Hence, it is seen that the number of living women invariably exceeds the number of living men.

Whether we consider such observations from the point of view of politics, or medicine, they will not lose their importance; on the contrary, if the authors will understand all their value and properly use them, their application will become even wider.

2.In addition to tables of the born and the deceased, tables of marriages are also usually compiled, which, above all, serve for illuminating government economy, and will serve this purpose still better if compiled more thoroughly. It would be desirable to designate in the records by additional Arabic numerals the age of each of the marrying persons, and, by another Roman numeral, what marriage does the bridegroom enter, and what marriage the bride,– the first, the second, or the third one. From tables of marriages compiled in this manner the customs governing this matter in different nations and different towns become very intelligible. A sensible attack against these customs, if they do not quite correspond to state intentions, could be launched.

3. The mentioned thorough tables can only be compiled on the basis of permanent observations collected with great care. The use of prior considerations does not help here at all, but how many new truths would it be possible to elicit therefrom by theoretical reasoning alone! Mortality tables, for example, show us the mean duration of the remaining life for persons of any age; true, they show it not directly but through deductions which somehow all by themselves can randomly cross the mind of anyone. However, the problem with which we shall now deal, the one concerning the mean duration of marriages, is much more difficult and demands a lot of deduction.

Since, on the other hand, it affects various groups of mankind, the rights of succession and mutual relations, I think that I shall do a valuable deed if I point out the way for ascertaining a subject which has much in common with an essay promulgated by me seven years ago in the Paris Académie Royale des Sciences and inserted in the Commentaries {*Mémoires*} of that Academy for 1760 [1]. There, applying a kind of analysis unusual for that matter, I spoke in detail about mortality from smallpox.

4. Since the problem concerns the mean duration of marriages for a given age of the spouses, it is necessary first of all to compile a table for some number of homogeneous marriages that will show how many such marriages remains untouched by death on the termination of each subsequent year of life. Nevertheless, a method by whose means such a table could have been compiled, is not yet discovered. True, mortality tables show the numbers of both men and women still living after each year of their life. However, either number consists of the widowed and of those married, and because each of these two numbers is absolutely unknown, the application of mathematical analysis, by whose means these numbers could have been calculated, is required. This analysis should be borrowed from the theory of probability.

I shall begin with a simpler issue. First of all we assume the same age for both sexes, *i.e.*, that all the marrying bridegrooms and brides are of one and the same age. Let us suppose further that men and women of the same age are in an equal measure running the danger of death although observations show that the case is somewhat different and that the weaker sex is leading a way of life safer from death. The latter proposition, however, should be understood not as though a smaller number of women die yearly in populous cities or other places than men do. If as many women were born as men, then as many of them would have been dying. However, as it is known, more women are living in any nation than men wherefrom it follows that the yearly numbers of those dying of both sexes can even be absolutely equal. But this does not mean that, with respect to the total number of the living of both sexes, women do not die in a lower proportion than men. I wanted to recall this point because I noted that some authors express a wrong opinion about it.

Thus, after examining considerations in favor of our general hypothesis, I shall show in how it can be verified altogether and with utmost precision.

5. Suppose that the initial number of all the marriages = n so that the number of all those married will be $2n$. Then let their part carried off by death after a known number of years be counted. The number of those still living we denote by r so that the number of all the deceased will be $(2n - r)$. We denote the number of the remaining marriages by x, then the number of all the widowed will be equal to $(r - 2x)$. Having determined all this, I say that $x = (rr - r)/(4n - 2)$.

I provided an obvious solution of this problem in the memoir [2, §2]. True, I have there applied terms usually accepted in the theory of probability, but everyone can see that exactly this is the essence of the matter.

Then, the number of all the widowed, or $(r - 2x)$, will be $(2nr - rr)/(2n - 1)$ whose half thus expresses the number of either the widowers or the widows because we suppose that death equally strikes down men and women.

6. Consequently, it is now easy to determine the number of marriages which will remain untouched by death from among any number of those still living. Since this last number is given in the mortality tables for each year of the age, the number of remaining marriages can be determined at the same time for each year. I select the mortality table for the city of Breslau compiled by the most illustrious Halley and reprinted in various works. In particular, it can be found in the excellent work of Deparcieux entitled *Essai sur les probabilités de la durée de la vie humaine* {1746}. I shall use it for compiling new tables suitable for our goal.

I may naturally begin with any number of marriages and any common age. Suppose that there are 500 initial marriages contracted between 1000 persons each of them having the exact age of 20 years. Since, in accord with the Halley mortality table, 598 persons {out of 1000} reach twenty years, whereas I assume 1000 persons, each of Halley's numbers ought to be increased in conformity with the proportion between the numbers 598 and 1000 {1000 and 598}. I discard the fractions and substitute instead of them integers to which they are closer or which can better reveal the uniformity in the movement ahead [*qui melius uniformitatem in progressu observat*].

The table appended below consists of four columns. The first one shows the age expressed in years; the second column is the number of persons still living; the third one is the number of remaining marriages not destroyed by death; the fourth column shows the number of the widowed without distinguishing between their sexes. These last numbers should be divided into two equal groups, those of widowers and widows; all those who became single after their first marriage {was broken}, irrespective of whether they contracted new marriages or remained widowed, should be entered into this column.

Table to §6

Age in Years	Remained Living	Remained Marriages	Widowed
20	1000	500	0
21	990	490	10
………	………	………	………
41	729	265	199

Note: Daniel Bernoulli compiled this table up to and including age 89.

7. The formula that we made use of for finding the values in the third column by calculation, to wit, $(rr - r)/(4n - 2)$, clearly shows that its numbers are not directly proportional to the first arbitrarily taken number if only this initial number is not very large. However, even then the subsequent numbers will ever more move away from the unity of proportion. Then, it should be noted that since we have discarded all the fractions, small mistakes larger than 1 and less than 2 could have crept into the numbers of the fourth column. In general, these numbers correspond to the formula $(2nr - rr)/(2n - 1)$ (§5). Consequently, the number of the widowed reaches its maximum when the initial number of those marrying decreases to its half, *i.e.*, when $x = n$. In this case the maximal number of the still living widowed will be equal, in accord with the formula, to $nn/(2n - 1)$, or, in our example, 250 ¼, although the table shows 252. I recall this so that someone will not attribute the committed small mistakes, made so as to facilitate calculations, to our method itself.

8. The table that we just adduced begins with 20 years. It is not difficult to compile a new table beginning with any other age. Suppose that it is desired to study the peculiarities [*indolem*] of the marriages for the age of the bridegroom and bride who are exactly 30 years old. The previous table provides the number of marriages remaining at that age as being equal to 395. Since the number of those still living is always equal to twice the number of the marriages, the second column should begin with 790 persons. Consequently, the number of those living, or 888, should be replaced by 790, and all the subsequent numbers given in the second column should be decreased in the same proportion. As to the numbers in the third column, all of them should be the same as they were previously. Finally, if the doubled numbers of the third column be subtracted from the numbers of the second column decreased in the indicated way, the numbers of the fourth column will be obtained. Tables for any common (for men and women) age of those contracting marriages are compiled in the same way.

9. If it is further asked, when will the initial marriages contracted at the age of 20 decrease by one half owing to deaths, this is clearly seen at once when looking at the table. Since 500 is assumed as the initial number of marriages, we should only look in the third column at what age 250 marriages remain intact. Then we shall indeed see that this number corresponds to the age between 42 and 43 years of life of the spouses, or, more precisely, to 42 years and 4 ½ months. Consequently, we can bet, with equal risks, on whether or not a marriage, contracted at the age of 20 years common for both spouses, will exist after 22 years and 4 ½ months. The same question for any other age of those marrying, if only this age is the same for both spouses, is solved in exactly the same way. Let for example the initial age of the spouses be 40 years. Our table shows that at this age 276 of the initial marriages remain in existence and that this number is decreased by one half after 12 years and 10 months.

In this way I compiled the appended short table for the beginning and the end of each five-year period. The first column shows the common age of both spouses; the second one indicates the time expressed in years and months after which a half of the marriages will probably be destroyed. By the method of interpolation these quantities can be determined rather precisely for any intermediate age.

{Two tables combined} Table to §§9 and 11

Age in Years	Duration of Marriages, Years & Months	
	Probable	Mean
20	22 4 ½	23 10
25	19 7	21 3
……………………………………………………………..		
80	2 0	3 0

Note: Daniel Bernoulli compiled both his tables with interval 5 years.

10. The question about the time when half the marriages will be destroyed at whatever age they began, should not be confused with our fundamental question on the mean duration of marriages expected at any age of those marrying if only it is the same for both spouses. However, it is easy to foresee that these questions will not very considerably differ from each other. The method for finding the mean duration of marriages resembles that which we apply for determining the remaining duration of life expected at some age. But this method requires that the values of each yearly wastage be known. Since such losses were not yet determined by any observations, it crossed my mind whether this can be done by calculations with respect to any age, and I have indeed accomplished this work in accord with my desire [*ex sententia*]. And so, I shall show now how, by means of the table appended to §6, the remaining duration of marriages for any age of the spouses should be looked for.

11. It is necessary to add together all the numbers indicated in the third column of our table beginning with the given age to the end, and to divide the sum by the number concerning the given age. The quotient obtained will show, as it is known, the sought mean duration even if the separate marriages from among those destroyed during each given year were actually only destroyed at the end of the year [*simul in fine anni*]. Indeed, since the destruction is going on almost uniformly during the whole year, there will not be a perceptible mistake in attributing the obtained sum to the mean yearly number of {the remaining} marriages.

It follows therefrom that the said quotient should be decreased by half a year, or by six months. Thus, if, for example, the matter will concern the mean duration of marriages between spouses 55 years of age, the sum of the numbers of the third column should be taken beginning with 118 and to the end; this sum equals 1208. It should be divided by 118,– by the number corresponding to the initial age. The quotient will approximately be 10 ¼ designating the same number of years. Subtracting half a year from this, we obtain 9 ¾ or nine years and nine months. Consequently, the mean duration of marriages between 55-year-old spouses equals nine years and the same number of months. […]

12. So that it would somehow become clear how much the results of our theoretical computations fit in with observations; I choose one example that I found recently in the *Bern Acts* [1] where it is said that 1053 remaining marriages are recorded in Lausanne and that 49 new marriages are contracted there yearly. It follows therefrom that the mean duration of these recently contracted marriages is 21 years and six months. Such a mean duration of marriages is longer than that which corresponds to the age of 25 years in the table. However, it is not at all likely that the mean age was about 25 years for each of the contracted marriages so that our calculations quite well fit in with observations. At present, I have no time [*non vacat*] for looking up other examples offered by other authors. If others will desire to take upon themselves this matter, it would be easy for them to draw up their own opinion about it, let them only remember that mortality in different countries is only somewhat [*paululum*] different and that depending on the per cent of mortality the duration of marriages increases or decreases.

13. Now, if we desire to find out the total number of all the existing marriages in a country where 500 new marriages are contracted yearly, if these, according to our supposition are only contracted between persons having 20 years of age, then, to this end, the mean duration of marriages should be multiplied by their yearly number, *i.e.*, 23 5/6 multiplied by 500. The obtained product, or 11 917, will provide the required sum: it will be the number of all the existing marriages. And the total sum of all the people living in conjugal union is 23 834. If we take some other initial age for persons contracting a marriage, then the table of §11 can help {to solve} this problem. The number of all the preserved marriages is everywhere equal to the product of the marriages contracted during one year by the mean duration which is recorded in the second column of this short table.

Suppose that the conclusion of each marriage is attributed to the end of the 30th year [*in finem trigesimi anni*]. Then, it will be necessary to multiply the number of the yearly marriages by 18 5/6. The number of yearly marriages itself will now be less because it depends on the number of those still living; that is, on the numbers 1000 and 888 which correspond to the ages of 20 and 30 years. Consequently, the number of yearly marriages will be equal to (888/1000)·500, or 444, which, being multiplied by 18 5/6, gives 8362 whereas it was 11 917 for the age of 20 years. It can be said already in advance that the marriages of the first kind will be more fruitful than those only contracted after the end of the 30th year [*finito trigesimo anni*] of age.

If some state will gain 2300 children yearly from 500 yearly marriages each of them contracted before the end [*sub finem*] of the age of 20, then this same state will be hardly capable of obtaining more than 1000 children if all the marriages are, and each of them separately is contracted only after the lapse [*elapso trigesimo anno*] of the 30th year of age [2].

14. I shall add some words about the number of all the widowed in the same state about which we spoke just now; suppose that they invariably remain widowed or that all men and women once becoming widowed are attributed to this category. If, in the table of §6, the values of the fourth column are added together, the number 10 326 will be obtained which is the approximate expression for the number of all the persons widowed until the end [*usque ad finem*] of age 89. About 24 souls outliving this age may be added here so that the number of all the widowed in that state may be reckoned as 10 350. Consequently, all the people connected into a conjugal union will be to all the widowed as approximately 23 834 to 10 350 or almost as 23:10. The matter is going on in such a way if we suppose that each marriage is contracted at the age of 20 for both spouses. Another age will have another table with other numbers which can be easily obtained after appropriate changes in the calculations. It would be too bold to speak about everything. There are very many issues that are resolved by calculations alone and certainly better than by the most diligent observations [*et quidem melius guam observationitas utcunque assiduis*].

15. I pass on to another kind of marriages contracted between spouses of differing ages. The common feature for marriages of this kind is that the husbands are mostly older than the wives. It is known, however, that the danger of death heightens with age, and, besides, that under equal conditions men are subjected to greater dangers than women. Thus, if we assume a higher age for the husbands, then they will both easier and sooner [*facilius et citius*] pass away and leave behind them far more widows than there will be widowers. True, I thought that this inequality between the numbers of widowers and widows will present too much difficulties for our subject. However, the very nature of things helped in that a differential equality {equation} between three unknown quantities fortunately admits not only differentiation but integration as well. I applied such an approach in the memoir [1] mentioned above in §3. And so, it is understandable that I made use of the method of the algorithm of infinitesimals which requires that the most greater number, or as though an infinite number be taken as the initial number of marriages. I have expounded this method in full with only a change of names in a preliminary essay [2, from §7 to the end]. Here, we have husbands and wives,– there, we had black and white cards; here, marriages,– there, pairs of cards; here, finally, a higher mortality of one of the two sexes,– there, a greater facility for being drawn from an urn, or a greater inclination for leaving it.

16. Let us again assume n as the initial number of marriages. Suppose that each marriage is contracted in such a way that all the husbands are of the same age one with another as are all the wives, but that for either sex the ages are different. Then, let the number of the still living husbands after a known number of years be s, and the number of living wives be t; both these numbers are determined by mortality tables. It is asked now, how many remaining or untouched by death marriages will there be between all these living people; and, further, how many will there be widowers, or at least those who once became a widower; and how many widows or women who once became a widow. Again, assume that the number of the remaining marriages is x. then the number of the remaining widowers will be $(s - x)$, and the number of the remaining widows, $(t - x)$. Under all these assumptions, as we proved in §7 of our preliminary memoir [2], we find that x is equal to st/n. And so, if we find the numbers s and t, all the rest will become known as well.

It is apparently not necessary for a direct ratio to exist between the different mortalities of men and women caused by the difference in age. This, however, is clear to anyone on account of a certain interrelation between the numbers s and t which necessarily includes different mortalities being also presumed. For all its simplicity, our formula is so general that it can express at the same time another distinction between the measures of mortality caused by the difference between the sexes if mortality tables for men and women separately will be available.

Then, to be able to use the Halley tables, we shall only consider the difference between the measures of mortality caused by the difference in ages. We did the same when speaking about the marriages between spouses of equal ages. It will be thus possible the better to compare the various kinds of marriages with each other.

17. I consider it needless to say how another table similar to the one appended to §6 can be compiled when the ages of the spouses are unequal. True, any difference between the ages requires new calculations, but it is sufficient to provide the data for five-year periods because the intermediate values will be rather precisely reckoned by interpolation. The example that I shall now present will be a specimen for all cases.

Suppose that new marriages are contracted between 500 girls each of them being 20 years old and the same number of men, 40 years old each. Thus, the number of marriages is again 500, but now the numbers s and t for each year, *i.e.*, the numbers of both men and women still living, have to be found. Both numbers can be precisely and, besides, easily determined through calculations by means of the second column of our table appended to §6. Indeed, the number t will be obtained when taking a half of each number indicated in the second column; if this half is then multiplied by 1000/744, we obtain the number s. This multiplication is easy for the reason that the initial number of men is assumed to be 500. In that way the second column is compiled; it provides the number of still living men and women after each year of life. Then the third column is compiled. It shows the number of the remaining marriages by assuming, for each year, the magnitude *st/n*. Finally, the fourth and the fifth columns showing the number of widowers and widows are compiled by subtracting the number of the remaining marriages from the number of the still living men or women. The calculation of the last-mentioned numbers is of no independent significance. I considered it necessary to adduce them by issuing from the other columns of the table.

Table to §17

Age in Years		Remained Living		Remained	Widowed	Widows
Men	Women	Men	Women	Marriages		
40	20	500	500	500	0	0
41	21	490	495	485	5	10
…	…	…	…	…	…	…
60	40	271	372	202	69	170

Note: Daniel Bernoulli compiled this table up to and including ages 99/79 of husbands/wives

18. The use of this second table is the same as of the previous one. Consequently, all about which we spoke in §§7 – 14 might also be applied for the present {the second} kind of marriages. Thus, for example, when in §9 I formulated the problem about determining the time when half the marriages contracted between persons of the age of 20 years will be destroyed, I found that this will happen after 22 years and 4 ½ months. For the marriages of the present kind, contracted between women having 20, and men of 40 years of age, this time is decreased to 16 years and 4 months as is seen just by considering the third column of the second table. Had a similar problem about marriages between women of 40, and men of 60 years of age be raised, the same third column would have shown that 202 marriages will remain and that they {that this number} will decrease to 101 in 9 years and one month.

Thus, it offers no difficulty to compile a short table similar to the one appended to §9, and it only suffices to remind about this. However, since the matter mainly concerns the mean duration of marriages for differing ages of the spouses, I do not shirk this work either.

19. This problem, when the husbands are 20 years older than the wives, is solved by an absolutely the same method which we made use of in §11where we assumed that the age of both spouses was the same. And so, for the initial age of men being 40 years and of women, 20 years, it will be necessary to take the sum of the numbers of the third column in the first table, which is equal to 9207, and to divide it by the number of the initial marriages, or by 500. The quotient will be 18 41/100 years, or roughly 18 years and 5 months. Subtracting 6 months from this we shall obtain 17 years and 11 months, the sought duration of a marriage. I have indeed compiled the appended table by this method. It can now be compared with the short table appended to §11 so that it would be seen how different are the two kinds of marriages with respect to their duration expected for persons of some age. The difference is not insignificant, and it would have been still greater had we possessed statistical data about the higher mortality of men and had at our disposal mortality tables where people of either sex were separated from each other with the numbers of men and women shown separately.

Table to §19

Age in Years		Mean Duration of Marriages, Years & Months
Men	Women	
40	20	17 11
45	25	15 8
………………………………………………..		
85	65	2 6

Note: Daniel Bernoulli compiled this table with interval 5 years.

20. Finally, some will desire to compare both kinds of marriages, *i.e.*, on the one hand, marriages where both spouses are of the same age, and, on the other hand, of those where the husband is 20 years older than the wife, so that even those who quail before new calculations were capable of forming, without a large error, a judgement about any mean state.

a) We have already spoken about the different mean durations of marriages of both kinds. Thus, for example, not without reason a woman 55 years old marrying a man of the same age promises herself a twice greater duration of her marriage than if she would have married a 75-year-old man.

b) If we imagine ourselves two nations the first of them having 500 yearly marriages of the first kind, and the second one, the same number of marriages of the second kind, then, under the assumption that each of the two nations will remain in invariable conditions, the first state will have {in all} 11 917 marriages and the second one, only 8958 (§13). We consider marriages contracted for the first time by either spouse and in any case such marriages make up the greatest number {the greatest part}.

c) Then, one and the same state, where 500 marriages of the first kind can be contracted yearly, will be unable to have more than 372 marriages of the second kind because certainly from among 500 young men only 372 will reach 40 years. Thus, a quarter of the girls will have to remain unmarried or marry widowers. This is why much more widowers than widows get married although the number of the former is smaller than that of the latter.

d) Let us consider now those widowed of either sex. To these we attribute all those who once became widowed irrespective of whether they remain widowed or contract a second marriage. We saw in §14 that for 500 yearly marriages of the first kind the number of all the widowed is, in accord with the indicated reckoning, 10 350, or 5175 widowers and 5175 widows. But the matter is far different when it concerns marriages of the second kind because then the number of widowers is considerably less than the number of widows. Then our table in §17 shows only 2154 widowers whereas the sum of all the numbers in the fifth column is 7949, and this number of widows should still be increased by 197,– by their number yet living after the termination of the last year in the table so that the total number of widows is 8146. However, the sum of all the widowed is almost the same in both cases. When comparing the number of widows with the number of widowers, we see that in the second case (where the bridegroom is 40, and the bride is 20 years old) the number of widows is almost four times greater than the number of widowers whereas in the first case (in which the spouses are of the same age) these numbers are equal to each other. Consequently, in a state where 500 marriages of the first kind are yearly contracted, there will be 11 917 husbands and as many wives, 5175 widowers and as many widows, the total number of all of them being 34 184. If, however, 500 marriages of the second kind will be contracted in that state, there will be 9207 husbands and as many wives, 2154 widowers and 8146 widows with the total number of them being 28 714.

e) In a state, which, let us suppose, acquired stable conditions, if 500 marriages of the first kind are contracted yearly, the same number of them will be destroyed during the same time. 250 husbands will die leaving as many widows and 250 wives as well who will make widowers out of 250 husbands. However, from the 500 marriages of the second kind, death will yearly carry off 345 husbands and 155 wives; and there will thus remain 155 widowers and 345 widows. Since the number of the remaining husbands is s, the number of the remaining wives, t, and the number of the remaining marriages, st/n (§16), the wastage of husbands will be expressed by the ratio of the number of the former to the number of the latter, that is, of s to st/n, or n to t. It follows therefrom that the wastage of husbands is expressed by the formula – tds/n, and the wastage of wives is – sdt/n. Therefore, the total numbers of the deceased husbands and wives are, respectively,

$$-\int tds/n \text{ and } -\int sdt/n.$$

However, since the general ratio between s and t is only given by mortality tables, the integrals above can only be partly [*per partes*] given. I have obtained the numbers just adduced by this method.

My essay, whatever it is, makes it clear that many variations occur with mankind and that there are many reciprocities that may be more thoroughly and better determined by calculation than by countless observations made until now.

Notes

1. {I can only say that the *Acta Helvetica* had been published in Basel.}
2. {The Russian text says *Only before the lapse*, etc.}

References

1. Bernoulli, D. (1766), Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir. *Werke*, Bd. 2. Hrsg. B.L. van der Waerden. Basel, 1982, pp. 235 – 267.
2. --- (1768), De usu algorithmi infinitesimalis in arte coniectandi specimen. Ibidem, pp. 276 – 287.
3. --- (1768), De duratione media matrimoniorum, pro quacunque conjugum aetate, aliisque quaestionibus affinibus. Ibidem, pp. 290 – 303.
4. Sheynin, O. (1977), Daniel Bernoulli's work on probability. In *Studies in History of Statistics and Probability*, vol. 2. Editors, Sir Maurice Kendall & R.L. Plackett. London, 1977, pp. 105 – 132.

3. Kh.O. Ondar. On the Works of Davidov in Probability Theory and on His Methodological Views

Istoria i Metodologia Estesvennykh Nauk, vol. 11, 1971, pp. 98 – 109…

Foreword

Avgust Yulievich Davidov (1823 – 1885) was Professor at Moscow University from 1853 and author of studies in mechanics and of various mathematical problems. His research in probability bordered on the work of Laplace and Poisson, and he also paid much attention to the application of statistics in medicine. In Russia, Davidov is not forgotten, but the only study of his findings is the one that I am presenting now in translation. Note that I had not checked Ondar's formulas and am not satisfied with his mathematical reasoning. I have left out some parts of the text (repetitions, hardly important details, discussion of Davidov's philosophical outlook); in Note 1 I omitted the titles of the mentioned books from Davidov's private library.

I myself, some 25 or 30 years ago, attempted to get hold of ref. [6] but only found an unreadable microfilm of some of its pages, and I notice that Ondar had passed it over in silence.

*   *   *

Davidov published ten contributions on the theory of probability [1 – 10]. Modern literature on the history of mathematics [11 – 15] cites only some of them, and, for that matter, only in passing. Usually only five of his works are mentioned whereas the other ones are not even included in the listing of his writings on probability. However, these are of certain interest as providing a more comprehensive characteristic of Davidov's efforts in this field.

The Department of rare books and manuscripts of the Gorky Library at Moscow University is keeping Davidov's private collection that includes many treatises on probability [1] as well as a large number of separate articles on this subject. He left many marginal notes in these sources and it is possible to assume that he knew and studied the contemporaneous materials on probability. At least as far as this branch of mathematics is concerned, we hesitate therefore to agree with Vygodsky [12, p. 153], that Davidov *did not like to bother himself with reading* [scientific] *literature.*

[1] We turn now to consider Davidov's works in probability and we begin with contribution [4] since it is the most valuable of all of them; as is witnessed by [1], Davidov expounded its contents to students of Moscow University. The study of [4] will allow us to judge, to a certain extent, his scientific standards and his style, and to see the findings contained there.

In his Introduction, Davidov indicates that conclusions, arrived at in statistics without taking into account stochastic laws, are not trustworthy. He regrets that the theory of probability did not become generally applied because of its involved mathematical tools, and he aims at deriving a formula for usage in many statistical problems.

We shall now describe his further reasoning almost verbatim, since, in our opinion, he is making use of the concepts of distribution function and density [2]. Suppose that some event dependent on n magnitudes

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

takes place when they satisfy a certain condition, for example when

$$F(\xi_1, \xi_2, \ldots, \xi_n) = 0. \quad (1)$$

Let us consider an event {stochastically}determined by a more general condition

$$a \le F(\xi_1, \xi_2, \ldots, \xi_n) \le b. \quad (2)$$

Assuming that $F(x)$ is the probability

$$P(0 < \xi < x) = F(x), \quad (3)$$

we shall determine the probability of the existence of (2). Davidov indicates that

$$F(x + dx) \approx F(x) + F'(x)dx$$

or

$$P(0 < \xi < x + dx) \approx F(x) + F'(x)dx.$$

If

$$F'(x)dx = f(x)dx, \quad (4)$$

then

$$P(\xi = x) \approx f(x)dx \quad (5)$$

and

$$P(m < x < n) = \int_m^n f(\xi)\, d\xi. \quad (6)$$

Let now $\alpha_i \le \xi_i \le \beta_i$; $i = 1, 2, \ldots, n$; then

$$\int_{\alpha_i}^{\beta_i} f_i(\xi_i)d\xi_i = 1. \quad (7)$$

Of course [4, p. 3], the ξ's are supposed to be continuous.

Had Davidov taken the lower boundary of ξ equal to $-\infty$ rather than to zero, expression (3) would have been the modern definition of distribution function, and (4) would have corresponded to the definition of density. […]

We now take up Davidov's further deliberations. Suppose that in inequalities (2) $a = c - d$, $b = c + d$, then

$$-1 < \{[F(\xi_1, \xi_2, \ldots, \xi_n) - c]/d\} < 1 \quad (8)$$

or, denoting the fraction by z,

$$-1 < z < 1. \quad (9)$$

Introducing a new function $\Phi(t)$ equal to 1 if $|z| < 1$ and to 0 if $|z| > 1$ {how about $z = 1$?}, we have

$$P(|z| < 1) = \int_{\alpha_1}^{\beta_1} f_1(\xi_1)d\xi_1 \int_{\alpha_2}^{\beta_2} f_2(\xi_2)d\xi_2 \ldots \int_{\alpha_n}^{\beta_n} f_n(\xi_n)\, \Phi(z)\, d\xi_n.$$

Davidov does not specify that the ξ's are independent; however [16, p. 130], "according to a contemporaneous tradition, no-one indicated" this condition [3].

Noting that

$$I = \int_0^\infty \sin x \cos xz\, (dx/x) =$$

$$(1/2)\int_0^\infty \sin[(1 + z)\, x](dx/x) + (1/2)\int_0^\infty \sin[(1 - z)\, x](dx/x),$$

$$\int_0^\infty k \sin u\, (du/u) = (\pi/2) \text{ if } k > 0 \text{ and } -(\pi/2) \text{ if } k < 0$$

so that

$$I = (\pi/2) \text{ if } |z| < 1 \text{ and } 0 \text{ if } |z| > 1$$

and we have

$$\Phi(z) = (2/\pi)\int_0^\infty \sin x \cos xz\, (dx/x),$$

$$P(|z| < 1) =$$

$$(2/\pi)\int_0^\infty \sin x\, (dx/x)\int_{\alpha_1}^{\beta_1} f_1(\xi_1)d\xi_1 \int_{\alpha_2}^{\beta_2} f_2(\xi_2)d\xi_2 \ldots \int_{\alpha_n}^{\beta_n} f_n(\xi_n) \cos xz\, d\xi_n.$$

Taking into account the Euler formula

$$e^{xzi} = \cos xz + i\sin xz, \qquad (10)$$

Davidov obtained

$$P(|z| < 1) =$$

$$\text{Re}\, [(2/\pi)\int_0^\infty \sin x\, (dx/x)\int_{\alpha_1}^{\beta_1} f_1(\xi_1)d\xi_1 \int_{\alpha_2}^{\beta_2} f_2(\xi_2)d\xi_2 \ldots \int_{\alpha_n}^{\beta_n} f_n(\xi_n)\, e^{xzi}\, d\xi_n], \qquad (11)$$

an expression which we shall call his main formula […]

Let now

$$z(\xi_1, \xi_2, \ldots, \xi_n) = c + c_1\xi_1 + c_2\xi_2 + \ldots + c_n\xi_n \qquad (12)$$

where the parameters are some real numbers and suppose that $\alpha_i = -\infty$ and $\beta_i = +\infty$ and $f_i(\xi_i) = \exp(-\xi^2)/\sqrt{\pi}$. Then, on the strength of (11),

$$P(|c + c_1\xi_1 + c_2\xi_2 + \ldots + c_n\xi_n| < 1) =$$

$$(1/\sqrt{\pi})\int_0^{y_1} \exp(-y^2)dy + (1/\sqrt{\pi})\int_0^{y_2} \exp(-y^2)dy \qquad (13)$$

where

$$y_{1,2} = (1 \pm c)/\sqrt{c_1^2 + c_2^2 + \ldots + c_n^2}\,.$$

[…] If, in formula (12), $c = o$ and the other parameters are unity, the problem is reduced to studying the behavior of a sum of independent normal random variables.

Davidov made use of a method based on the application of the Dirichlet discontinuity factor (above). According to Liapunov [17],

> *One of the most elegant among the old methods is the one based on the application of the discontinuity factor. Glaisher* [18] *used it to solve the problem under consideration.*

And in 1901, in a letter to Markov, he [19] also wrote:

You are asking whether I am sure that Glaisher was the first to introduce the discontinuity factor into the calculus of probability. I can only answer that Laurent published his Traité du calcul des probabilités in 1873, whereas Glaisher had published his papers in 1872. According to Sleshinsky, Cauchy thought of applying this method, but, as far as I know, he had not developed his idea. Leaving Cauchy out, I do not know whether anybody had used the method of discontinuity factor before 1872.

Davidov applied that method in 1854, considerably earlier than Glaisher did.

In his article [2] Davidov explicated the principles of probability theory and its applications for readers unfamiliar with higher mathematics. He began by considering various random events ({the results of} drawing balls from a box or {of} tossing coins, changes in air temperature, etc). While discussing the law of large numbers, he indicated that the innumerable diversity of random phenomena, which, taken separately, do not obey any definite law, "reveal a surprising and unexpected regularity as the number of the phenomena becomes ever more considerable". He illustrated this proposition by lucid and intelligible examples.

He then posed a problem: How much can the fraction *m/n* derived from observations differ from the actual probability of the studied event. He constructed an appropriate table giving the answer sought […] by issuing from the normal distribution.

[2] A few words about the article [3]. Here, Davidov mainly repeated the contents of his former work and extended the explication onto a more general domain,– statistics, whereas previously he had considered only medicine {medical statistics}. Here, he provided a more precise and simple table; his tables became popular. Zhukovsky et al [15, p. 37] testified that physicians and other readers from various places often asked him to send them a copy. In addition, Zenetz [20] compiled a contribution based on the abovementioned papers (and presented a copy to Davidov). With the latter's permission, he reprinted Davidov's table so as to criticize some investigations in medical statistics.

Another author included a chapter, *The knowledge of the statistical method in therapy*, in his book [21]. It contained an extract from [2] with Davidov's examples replaced by illustrations from therapeutics. Davidov's table was also reproduced there complete with detailed explanation of its use. The celebrated physician Erisman [22, pp. 1 – 46] most comprehensively explicated the contents of [2]. He set high store by it and by [5] and reprinted Davidov's table in full. Lutzenko [23] also included an extract from the same table and offered an example of its use. It follows that the explication of probability in medicine was then timely and that Davidov originated work in this direction in Russia.

[3] Let us now consider Davidov's contribution to population statistics. In 1857 he delivered a report [5] explaining intelligibly and in detail the concepts of the mean and the probable error and the law of large numbers. He correctly maintained that means represented nothing but a simplified and shortened expression of this law. Davidov then went on to describe the compilation of tables of mean mortality of the male Orthodox population of Russia for 1849 to 1853. His method consisted in comparing the number of deaths over several years distributed by age with the number of births during those years when the died were born.

In 1886 Davidov published a study of the laws of mortality in Russia [10]. Its first part was a short essay where he described the history of the compilation of mortality tables and expressed his high opinion of Buniakovsky [24]. At the same time, he justifiably criticized the latter's table since Buniakovsky had chosen 1862, the year of minimal mortality, and had thus somewhat lessened the merits of his table. In the second part of his work Davidov compiled a mortality table for Russia up to the age of 20 with yearly intervals. All this testifies that he was very interested in problems of population statistics in Russia.

[4] Davidov's works in mechanics and mathematical analysis show that he adhered to materialistic views concerning the theory of knowledge. He did not doubt that the laws of nature were objective, he believed in the power of science closely connected with life and technology, and he approached the study of the regularities of nature in a genuinely scientific way. […]

There exist completely opposing views on Davidov's outlook as expressed in his work on probability. Kolman [25, p. 74] maintains that "grains of materialism are to be only found in the work of Chebyshev and his school, of Markov and others". On the other hand, Efimochkina [14, p. 244] believes that Davidov was "a progressive scientist and materialist who clearly understood the role of probability theory and its place among other sciences". Druzhinin [13, p. 14], basing his conclusions on only one of Davidov's contributions [8], states that

In advancing the concept of subjective probability Davidov comes close to Laplace's proposition according to which probability is caused partly by ignorance, and partly by knowledge …

[…] Davidov [1, p. 2] thought that randomness had certain causes and that [1, p. 8] "The real difference between the so-called random and necessary phenomena consists in the difference between their causes". At the same time, he [1, p. 6] mistakenly treated the categories of necessity and randomness, thinking, as most contemporaneous natural scientists did, that all phenomena and processes were necessary:

Thus, when assuming, in the accepted sense, a difference between random and non-random phenomena, we should admit that this difference is essentially reduced to our knowledge or ignorance of all the conditions on which the phenomenon is dependent …

Laplace's inference is undoubtedly felt here, see the beginning of his *Essai philosophique sur les probabilités.*

In his course of lectures [1] as well as in his articles [2; 3] Davidov distinguished between objective, subjective, and philosophical probabilities [4]. […] The second one [1, p. 58] "has nothing in common with the phenomenon itself". Philosophical probabilities were those based on analogies. Thus, since the Earth had some features in common with the other planets of the Solar system, the conclusion that they were also inhabited was rather probable.

Davidov [2, p. 16] repeatedly stressed that the mathematical theory of probability exclusively considered objective probabilities since […] "only they can be precisely determined […]"

[…] In his later contributions [7; 8] Davidov does not mention philosophical probabilities anymore. In 1882 he {no such date in the References} stressed that it was impossible to apply subjective probabilities

to study the phenomenon itself, and all the mistakes made by celebrated scientists were mostly due to their having confused subjective and objective probabilities. A good example of such a mistake is provided by Poission's book where he applied probability to legal proceedings.

At the time, Davidov's views concerning probability were progressive. Indeed, almost all treatises published even at the end of the 19th, and the beginning of the 20th century defined probability as a measure of our ignorance. […]

Davidov's interest in problems in probability was incidental. He turned his main attention to mechanics and to other branches of mathematics which is perhaps why his achievements in probability were not as considerable. None the less, in our opinion, his methods of investigation, especially when regarded from a historical viewpoint, are of unquestionable interest. Appreciating probability theory at its true value, Davidov laid the basis for its teaching at Moscow University and widely popularized it by publishing his mimeographed courses, by delivering public lectures and putting out popular articles. […]

Notes

1. For example, those of S.F. Lacroix, 1833; V.A. Jahn, 1839; Poisson, 1841, this being the German translation of his celebrated book; G. Hagen, 1867.

2. {The first to introduce distribution functions was Poisson, see my paper in the *Arch. Hist. ex. Sci.*, vol. 18, 1978, pp. 252 – 253.}

3. {Laplace twice mentioned independence, see my paper in the same periodical, vol. 17, 1977, p. 11.}

4. {The first to introduce philosophical probabilities, unyielding to quantification, was Cournot (*Exposition de la théorie des chances,* etc, 1843, Chapter 17).}

References

A.Yu. Davidov (Items 1 and 6 – 9 are mimeographed editions)

1. *Лекции математической теории вероятностей* (Lectures on Mathematical Theory of Probability). First semester. M., 1854.

2. Application of probability theory to medicine. *Moskovsk. Vrachebn. Zhurnal*, book 1, 1854, pp. 54 – 91. (R)

3. The application of the conclusions of probability theory in statistics. *Zhurnal Ministerstva Narodn. Prosveshchenia*, No. 11, 1855, pp. 45 – 109. (R)

4. Application of probability theory in statistics. In *Учебно-литературные статьи профессоров и преподавателей Московского университета по случаю его столетнего юбилея* (Pedagogic and Literary Articles Written by Professors and Instructors of Moscow Univ. on the Occasion of Its Centenary). M., 1855. Separate paging.

5. *Теория средних величин с приложением ее к составлению таблиц смертности* (Theory of Mean Magnitudes with Its Application to the Compilation of Mortality Tables). In *Речи и отчет, произнесенные в торжественном собрании Московского Унив.* (Speeches and Report Made at the Grand Meeting at Moscow Univ.). M., 1857. Separate paging.

6. *Теория соединения результатов наблюдений, приводящая к наименьшим погрешностям или способ наименьших квадратов* (Theory of Combining the Results of Observations Leading to Least Errors, Or, the Method of Least Squares). M., 1869.

7. *Теория вероятностей* (Theory of Probability). M., 1879/1880.

8. *Математическая теория вероятностей* (Mathematical Theory of Probability). N.p., n.d.

9. *Теория вероятностей* (Theory of Probability). M., 1884/1885.

10. On mortality in Russia. *Izvestia Obshchestvo Liubitelei Estestvoznania, Antropologii i Etnografii*, vol. 49, pt. 1, 1886, pp. 46 – 66. (R)

Other Authors

11. Ya.I. Veinberg, Editor, Reminiscences about Davidov. *Izvestia Obshchestvo Liubitelei Estestvoznania, Antropologii i Etnografii*, vol. 51, Supplement, 1887. (R)
12. Vygodsky, M.Ya. Mathematics in Moscow University in the second half of the 19th century. IMI, vol. 1, 1948, pp. 141 – 183. (R)
13. Druzhinin, N.K. The notion of the law of large numbers in the Russian literature on theoretical statistics. *Nauchn. Trudy Moskovsk. Gosudarstven. Ekonomich. Inst*., vol. 5, 1957, pp. 104 – 122. (R)
14. Efimochkina, E.P. *К истории развития теории вероятностей в России в XIXв.* (On the History of the Theory of Probability in Russia in the 19th Century). M., 1953. A Candidate Dissertation. (R)
15. Zhukovsky, N.E., Nekrasov, P.A., Pokrovsky, P.M. The life and the works of Davidov. MS, vol. 15, No. 1, 1890, pp.1 – 57. (R)
16. Gnedenko, B.V. On Lobachevsky's work in the theory of probability. IMI, vol. 2, 1949, pp. 129 – 136. (R)
17. Liapunov, A.M. Sur une proposition de la théorie des probabilités. *Bull*. [*Izvestia*] AN Psb, sér. 5, t. 13, 1900, pp. 359 – 386.
18. Glaisher, J.W.L. On the law of facility of errors of observation and on the method of least squares. *Mem. Roy. Astron. Soc*., vol. 39, 1872, pp. 75 – 124.
19. Liapunov, A.M. Letter to A.A. Markov dated Oct. 1900. Archive, Russian Acad. Sci., Petersburg branch, Fond 173, Inventory 1, 1, No. 11. (R)
20. Zenets, M.K. The application of the theory of probability to the teaching of acute contagious diseases. *Zdorovie*, No. 80, 82 and 83, 1878. (R)
21. Manassein, V.A. *Лекции общей терапии* (Lectures on General Therapeutics), pt 1. Psb, 1879.
22. Erisman, F.F. *Курс гигиены* (Course in Hygiene)., vol. 2. M., 1887.
23. Lutsenko, I.M. On the importance of probability theory in medicine. *Vrach*, No. 50, 1888. (R)
24. Buniakovsky, V.Ya. Essay on the laws of mortality in Russia and on the distribution of the Orthodox population by ages. *Zapiski* AN Psb*,* vol. 8, Suppl. No. 6, 1865. Separate paging. (R)
25. Kolman, E. On the burning importance of the theory of probability. *Pod Znamenem Marksisma*, No. 2, 1934. (R)

4. B.V. Gnedenko. On Ostrogradsky's Work in the Theory of Probability

Istoriko-Matematich. Issledovania, vol. 4, 1951, pp. 99 – 123…

1. The name of our outstanding compatriot, Mikhail Vasilievich Ostrogradsky, whose 150th birthday occurred this year (he was born Sept. 12 (24) 1801 {and died 20 Dec. 1861 (1 Jan. 1862)}, is firmly included in the history of science, and, above all, in the history of mechanics, mathematical physics and mathematical analysis. Excellent surveys of his work concerning these directions of science had appeared [30; 20; 27; 17; 26; 28], but Ostrogradsky's work in probability theory has until now only been considered cursorily [1]. This might be explained by the fact that in this field he had only solved isolated particular problems that remained far from the main direction of the development of the theory as paved later on by Chebyshev. Whereas his work in mechanics, analysis and mathematical physics are inseparably included in modern science, his results in the field of probability are less important. Although from the viewpoint of the problems and aims of today the lack of attention to these results is rather natural, it can hardly be completely justified since we should also take into account whether they were interesting during his time.

The essays on Ostrogradsky's mathematical work that I have seen mention only three of his writings on probability [1; 3; 7]. One more unpublished manuscript kept at the Archive of the Soviet Academy of Sciences is sometimes cited. There, Ostrogradsky solved the following problem: Two gamblers decided to play a certain number of games and have already played some games. Determine the probability that one of them wins if the probabilities of their winning any single game are unequal.

However, two more of Ostrogradsky's popular notes [4; 5] directly connected with probability are completely forgotten. They are very interesting for achieving a more comprehensive picture of his work and methodological views. His memoir [2] might also be attributed to probability theory [2].

I describe the just listed six works as well as the drafts kept at the Manuscript Section at the State Public Library of the Ukrainian Academy of Sciences; I did not study Ostrogradsky's manuscripts kept in other archives. In 1858, Ostrogradsky had delivered 20 lectures on probability theory, and there are indications that three of them were published but I was unable to find them. Finally, A.N. Krylov [3], in an unpublished note [16] presented to the Presidium of the Soviet Academy of Sciences with a view of bringing out Ostrogradsky's complete works, mentioned his writings on insurance. Krylov himself, as he stated, made extensive use of them while serving on the retirement fund of the Naval Department. Here are his own words:

In 1856, in accord with the Paris treatise, Russia was deprived of the right to have a fleet on the Black Sea. A large number of office workers had to be fired, and, to improve their circumstances, it was decided to establish a retirement fund at the Naval Department, and to begin paying pensions in 1859. Life insurance was then a novelty, and calculations connected with the work of the retirement funds, or with determining the amount of pensions in accord with the pertinent deductions from wages were known still less. For this reason, both mathematicians who were members of the Petersburg Academy of Sciences, Ostrogradsky and Buniakovsky, were included in the committee charged with drawing up a charter of the fund. They have indeed made all the necessary calculations and provided their theoretical justification. The transactions of the committee were published without delay; they contain a remarkable note by Ostrogradsky and a number of joint notes written by him and Buniakovsky.

The last sentence is inaccurate: there are no such joint papers in the transactions of the Committee [25] which none the less contain several notes written by Buniakovsky and Ostrogradsky's note [6] (and official materials).

I shall dwell somewhat on this note. The Introduction to the transactions puts on record the exceptionally conscientious attitude of Ostrogradsky and his colleagues towards the work of the Committee, the importance and urgency of whose establishment I have indicated above. I quote:

Our noted scientists, academicians Ostrogradsky, Buniakovsky, and Veselovsky [4], were invited to participate in the work of the Committee. Not only had they fully consented; throughout all the work, they actively took part in the discussion and preparation of the entire project as its permanent members.

Ostrogradsky's work is best described in his own words:

We present the solution of the problem of the minimal pension and carry it through to such simplicity that it will not cause trouble even for those least experienced in arithmetic; it will be sufficient for them only to understand the addition of numbers. Three tables provide the solution. […]

The published tables contained the calculation of the sum accumulated after depositing 1, 2, …, 9 rubles for 1, 2, …, 52 years at the rate of interest of 4% (Table 1); the calculation of the sum accumulated by payments of 1, 2, …, 9 rubles yearly made in four-month instalments after 1, 2, …, 52 years (Table 2); and of minimal pensions paid out in accord with yearly deductions of 1, 2, …, 9 rubles for persons aged 41 – 80 years (Table 3). […]

Ostrogradsky's work was a valuable contribution to the theory of insurance, and retirement funds continued to use it for a long time in their practical work.

2. {Gnedenko briefly discusses Laplace's *Théorie* [19] and his philosophical outlook calling it *mechanistic-materialistic*.} Along with this confidence in the power of the human mind; in its unbounded cognitive faculty; in the absence of unknowable things or phenomena, Laplace now and then formulates idealistic propositions. In some problems he assumes probabilities of events on the basis of the idealistic principle of insufficient reason that consists in the following: If the probability of some event is unknown, assume as its reasonable value a certain number according to our inner considerations [5]. {Gnedenko also quotes Laplace's celebrated statement [18, 1995, p. 2] that an omnipotent intelligence would be able to comprehend everything.}

Obviously, we cannot say either at the first, the second, or at any other toss of a coin that the probability of its certain face is 1/2; it is simply unknown. Its determination, the estimation of its value, should not be done by such dubious means [6] that deprive the very concept of probability of its objective role as a quantitative characteristic of definite real phenomena.

I indicate this defect in Laplace's reasoning because Ostrogradsky followed suit. They both perceived that the essence of the theory of probability consisted not in those logical speculations that might be accomplished by playing upon its insufficiently definite concepts; not in determining knowledge of a phenomenon out of its complete ignorance by means of analytic calculations,– but in being an important tool for studying regularities of a special kind that take place in mass phenomena. They saw the justification of the theory in its ability to solve important practical and scientific problems. Indeeed, Laplace had applied probability for solving astronomical problems and studying population statistics, and Ostrogradsky indicated that one of his investigations directly bore on decreasing the work necessary when accepting delivery of large consignments at the War Department. Yes, but at the same time they both allowed themselves such expressions as might lead to an opinion that they found philosophical premises for using probability in *insufficient reason*.

Laplace's book [19] was the main source from where Ostrogradsky drew the topics for his work in probability. Poisson's (1781 – 1840) well-known writings [23; 24] did not considerably influence him, because, for one thing, he obtained some results before that French scholar did, and his scientific interests were sufficiently firmly established before these writings have appeared.

3. Ostrogradsky's work belonged to the period when physics had not yet formulated serious problems for the theory of probability. The writings of Clausius and Maxwell were published only during his last years, and Lomonosov's ideas about the molecular structure of matter and his kinetic theory of heat were already all but forgotten; not only did nobody think about their mathematical substantiation, no-one even recalled them. Applications to the theory of artillery firing also remained in a rudimentary state. As a result, great store was being set by applications to *moral sciences* [7]. Laplace [18, 1995, p. 65] justified the importance of this approach by stating that

Since most of our opinions are founded on the probability of testimony, it is very important to submit such evidence to the calculus.

One of the most important problems facing the moral sciences was, as Laplace (Ibidem, p. 78) put it, to determine

The probability that the decision of a court which may convict only by a given majority will be equitable, that is to say, will conform to the true solution of the question …

A large number of writings at the end of the 18^{th}, and in the beginning of the 19^{th} century were devoted to solving such problems and neither did Ostrogradsky avoid them. In his first work on probability he [1] had indeed investigated the probabilities of errors made by tribunals. One of the main principles underpinning the solution was the assumption that the inferences made by different judges were independent, and that they arrived at correct decisions with one and the same probability. If a large number of judges participate in solving an issue, then a tribunal, whose decision is reached by a majority verdict, occurs to be, practically speaking, infallible. But the mistake in this conclusion, as I have noted elsewhere, consists, above all, in ignoring the fact that, because of their class composition, courts of law do not judge cases with closed eyes. In addition, as remarked by Bernstein [9, p. 192],

This proposition is evidently wrong because it does not account for the fact that all the judges base their judgements on the same testimonies and material evidence. All of them will therefore gain an understanding of a simple case more or less in the same way; however, if intricate circumstances will mislead some of them, then a mistake will become more probable for the other ones. In other words, independence between the opinions of the judges is lacking, and this fact radically changes the situation.

4. In those times, no Russian literature on probability existed at all, whereas the need to originate it so as to attract attention to the solution of quite practical problems of the theory of insurance and of statistics was already felt. Not without reason had Professor N.D. Brashman, in 1841, speak about it definitely enough during a festive meeting of Moscow University. Discussing the establishment of insurance societies in Russia, he [10, pp. 30 – 31] dwelt on the need of a rigorous mathematical justification of their calculations:

Who can fail to see, with extreme regret, the absolute disregard, in the academic institutions, of one of the most important branches of mathematics? Only a few universities teach the rudiments of the theory of probability, and, up to now, there does not exist a single Russian work, or translation, on advanced or even elementary probability. True, the publications of the Academy of Sciences include important considerations with regard to the calculus of probability, but the rich treasures contained there since the time of Euler are the property of the educated world *outside Russia rather than of its Russian part. We hope that Russian scientists will soon try to make up this deficiency.*

The Russian treatise [11] written by the eminent mathematician and statistician Buniakovsky (1804 – 1889) only appeared in 1846. Until then, apparently only three pertinent contributions were published in Russia: the speech of Pavlovsky[21], Professor at Kharkov University, delivered in 1821; an article by Prince Kozlovsky [15]; and Chebyshev's Master Dissertation [13] [8]. In addition, the periodicals of the Academy of Sciences and public journals began to publish papers by Ostrogradsky and Buniakovsky. Chebyshev's remarkable paper [14], that might be considered as the beginning of a new period in the development of the theory of probability, appeared in print six years after Ostrogradsky's death.

5. In 1834, Ostrogradsky [1] solves a problem already considered in a particular case by Condorcet and Laplace: Assuming that the correctness of each judge is known, determine the probability of passing a wrong verdict for a tribunal consisting of a given number of judges. Contrary to Condorcet and Laplace, Ostrogradsky showed that, for judges of equal correctness, this probability only depended on the majority of the votes rather than on the number of judges. Later, in 1837, Poisson confirmed this result. Ostrogradsky began his memoir by posing the following problem:

Supposing that the boundaries of the correctness of each judge are known, the author offers analytical expressions concerning various possible cases for the probabilities of a mistake made by a tribunal consisting of a given number of judges. If the boundaries of the correctness are the same for each judge […], *these probabilities only depend on the majority of the votes.* […] *Laplace and Condorcet thought that such a conclusion contradicted common sense*.

He then criticized Laplace's opinion

concerning the extreme difference between the probability of a mistaken decision reached unanimously by a tribunal consisting of 12 judges and a majority of 12 in a tribunal of 212 judges.

Ostrogradsky went on to provide his solution based on the Bayes formula and assumed that the law of distribution of the correctness of each judge between the allowed boundaries was uniform. He made essential use of this restriction but had not at all mentioned it. I do not adduce his formulas since their derivation under the conditions indicated is not difficult and the problem itself is not interesting in our time. And I have already pointed out that it was formulated badly.

6. In his memoir [2] Ostrogradsky corrected inaccuracies made by Laplace [19]. Let y_x be a function of an integral non-negative x. Then, according to Laplace, its *génératrice* (generating function) u is determined from equality

$$u = y_o + y_1 t + y_1 t^2 + \ldots + y_n t^n + \text{etc.}$$

However, the generating function of y_{x+1} will not be, as he stated, u/t, but $(u - y_o)/t$; for y_{x+2} it will be $(u - y_o - y_1 t)/t^2$; and, in general, the generating function of y_{x+i} will be not u/t^i; as Laplace had written, but

$(u - y_o - y_1 t - y_{x-1} t^{i-1})/ t^{i-1}$.

It follows that the generating function of the finite difference Δy_x *will be not* $[(1/t) - 1]u$ *but* $[(1/t) - 1]u - y_o/t$; *similarly, for* $\Delta^2 y_x$ *it will be*

$$u[(1/t) - 1]^2 - (y_o/t^2) - [(y_1 - 2 y_o)/t]$$

rather than $u[(1/t) - 1]^2$, *etc.*

Ostrogradsky then indicated Laplace's mistake made in determining the generating function of the sum $\sum y_x$ and showed that the solution of finite equations became therefore absurd. Indeed, suppose that it is required to determine y_x from the equation

$$0 = ay_x + a_1 y_{x+1} + a_2 y_{x+2} + \ldots + a_n y_{x+n}$$

where $a, a_1,\ldots, a_n$ are constants. Then, as Ostrogradsky noted,

when making use of generating functions in Laplace's form we arrive at equation

$$0 = u\,[a + (a_1/t) + (a_2/t^2) + \ldots + (a_n/t^n)]$$

or at

$$0 = a + (a_1/t) + (a_2/t^2) + \ldots + (a_n/t^n)$$

from which we can determine t. However, in accord with its nature, this variable should remain absolutely indefinite.

Ostrogradsky then pointed out that his own formula led to equality

$$u = \frac{a_n y_o + (a_{n-1} y_o + a_n y_1)t + \ldots + (a_1 y_o + a_2 y_1 + \ldots + a_n y_{n-1})t^{n-1}}{a_n + a_{n-1}t + \ldots + at^n}.$$

In concluding his note, he made known that he had applied his formulas to problems in interpolation, extended them onto the case of two arguments, and solved many problems reducible to finite equations pertaining to probability theory [9].

7. In 1846 Ostrogradsky, in the largest article [3] from those under our discussion, was solving the following problem: An urn contains a given total number of white and black balls. Some of them are drawn from the urn and their color is recorded. It is required to determine the probability of the various possible compositions of the urn, and the probability that the number of white balls in the urn does not exceed given bounds.

Ostrogradsky had not here betrayed his scientific principles: he considered his study as a problem with possible important applications in everyday life rather than simply an analytical exercise. Here are his own words:

To understand the importance of this problem, note that it will occur when encountering difficulties connected with receiving a large number of articles possessing certain properties, and spending some time on each of *them so as to become convinced of these properties. Military suppliers are often obliged to carry out such work. For them, the balls in the urn are the supplied articles; white balls, for example, are those possessing the required properties whereas the black ones fail to have them. The drawing of a certain number of articles* {of balls} *in order to record their color is tantamount to examining a part of the supplied articles and evaluating their quality. Fix this part, let it amount to five, six, or seven articles per hundred, and select it randomly from all the totality. After drawing, studying, and recording them, those that might be accepted will determine the probability that the total number of acceptable* {unacceptable} *articles will not exceed the boundaries assigned beforehand. This determination is done in the same way as if the numbers of the white, and black balls in the urn had to be determined.* […] *Thus, the solution of the problem proposed by us might serve to diminish, approximately to* 1/20, *the mechanical and most frequently very tiresome work of examining a very large number of sacks with flour or rolls of cloth.*

Ostrogradsky's solution cannot satisfy us since he issued from a faulty notion that probability was a measure of our ignorance. Consequently, he thought that probabilities were uniquely determined even when we either would be unable to say anything about them, or would have assumed his inference as a new hypothesis. To illustrate, we adduce the first of Ostrogradsky's arguments that he put forward when solving the following preliminary problem: The general number of white and black balls in an urn is known, but its composition is not; l balls are drawn. Determine the probability that, among these, m are black and n white, $m + n = l$. Since, Ostrogradsky reasons, among the l balls there can be

$$0, 1, 2, \ldots, n, \ldots, (l-1), l$$

white, and, respectively,

$$l, (l-1), \ldots, m, \ldots, 1, 0$$

black balls, the probability of each of these possibilities is $1/(l + 1)$. He apparently had not noticed that these possibilities were not necessarily equally probable.

He calculates a certain probability by two different methods and arrives at an equality (in our usual notation)

$$\sum_{x=n}^{s-m} C_x^{\ n} \cdot C_{s-x}^{\ m} = C_{s+1}^{\ m+n+1}$$

calling this result a *small-scale theorem in the calculus of finite differences*. The final expression for the conditional probability that the urn containing s balls has x white balls and y black ones, given that among the l drawn balls n are white and m black, is

$$C_x^{\ n} \cdot C_y^{\ m} \div C_{s+1}^{\ l+1}.$$

For calculations, Ostrogradsky recommends to use the following formula {it is extremely involved; among other parameters, it includes the Bernoulli numbers. I am omitting it}. He then adduces a small table of the Bernoulli numbers, considers a numerical example and offers rather large tables for facilitating practical calculations.

8. The unfinished article [4] does not contain any analytical calculations; it is interesting mostly in a methodological sense. Here, we once again convince ourselves that, from the viewpoint of scientific methodology, Ostrogradsky was materialistically oriented although by far not consistently so. We find there isolated critical remarks on the then recently appeared book of Buniakovsky [11]. To acquaint ourselves with the nature of these objections, and to imagine as clear as possible Ostrogradsky the popularizer, we quote his own words with which he begins his article (pp. 29 and 30):

The theory of insurance cannot be explicated without making use of the analysis of probability theory on which it is founded – we shall therefore try to offer a clear idea about the essence of probability, then we shall turn to the theory of insurance and, while considering it, we shall also speak about the principles which it borrows from the science of probabilities.

Of course, it is possible not to speak about probabilities, and, after citing contributions devoted to this subject, to proceed directly to the theory of insurance. We act otherwise believing that such citations are hardly ever pleasant for the reader, whom I would have to refer to foreign sources since in the Russian writings known to us the theory of probability is expounded not quite clearly or correctly.

We could have confirmed our words by a critical analysis of such writings, but this will be out of place here. Nevertheless, begging pardon for the digression, let us say a few words about the introduction to one of these. The author wants to show that probability, which he calls likelihood, is a magnitude. In proving this, he attempts to convince us that some likelihoods can be higher or lower than other ones. Then, after offering all the arguments, sufficient, according to his opinion, for complete conviction, he infers that likelihood, just like any other mathematical magnitude, should be measured and admits of a measure.

Thus, the learned author believes that likelihoods are mathematical magnitudes only because some of them can exceed or be less than other ones. This opinion is not quite correct. Indeed, do not we really say, and, while saying it, do not we clearly understand, that such-and-such a scientist is more conscientious than another one; that a Frenchman is braver than a German; that a reader is more sensible than an author, etc. Conscience, bravery, wisdom can therefore be larger or smaller and are mathematical magnitudes, they can be measured, expressed in numbers, and various operations might be carried out on them. By such reasoning we could have considerably widened the scope of the mathematical sciences; fundamentals of mathematical theories of unscrupulousness, absurdity, etc can thus appear.

This sharp criticism was undoubtedly directed against Buniakovsky [11] [10], although it ought to be said, – not against its main defect, which was the author's subjective approach to probability: Ostrogradsky himself, as we shall see now, was guilty of the same. He continued:

The word probability has no meaning in itself, it always concerns some phenomenon or event. A phenomenon is usually called probable when believing that it will rather happen than fail to occur. On the contrary, a

phenomenon is considered improbable when the reasons for its failure are thought to be more convincing than those favoring it.
We note in passing that the analysis and estimation of the circumstances accompanying the phenomenon can be very different for different persons. A phenomenon considered probable, or even certain by one man will be thought improbable by another one. Almost all wagers, games of chance, etc take place because of such differences in opinion. We say almost because bets are sometimes meant to be lost to a needed high-ranking person.
Geometers attach another meaning, different from the usual one […] *to the word probability. For them, anything whose impossibility or certainty is not rigorously proved is probable. Thus, in the eye of geometers even a phenomenon, that all other people will consider unusual, might possess some probability; and, on the contrary, for a geometer, a phenomenon considered certain might be only probable. True, it will, generally speaking, be highly* probable.
Nature has no probabilities [11]. *All that occurs in nature is doubtless and certain. Probability is a corollary of man's weakness; it concerns us, exists for us and can serve only us. Its study is an important, and even a necessary supplement to those few truths that we know with a relative certainty.*
If a phenomenon depends on several other phenomena or cases, some of them producing it, and the others opposing it, and if, furthermore, they are such that for us,– we repeat: for us,– there is no reason for preferring some of them to the other ones, then the probability of the expected phenomenon is measured by a fraction […]. *It follows* […] *that we can only determine the probability of such phenomena that can be directly or obliquely separated into cases in whose occurrences we do not perceive any preferences of one over another.*
Five balls are in a bowl. […] *We do not know why one ball will be drawn rather than another one. A reason, about which we speak, undoubtedly exists, but it is absolutely unknown to us, and, since we are unable to prefer one ball over the other ones, all of them present for us equally possible cases. He, who would have known the arrangement of the balls in the bowl, and would have been able to calculate the movement of the hand drawing a ball,– he would have said beforehand which ball will be extracted; for him, there would have been no probability.*

For Ostrogradsky, therefore, probability is nothing but the cognizing subject's measure of certainty. All the inferences of probability theory are thus deprived of the objective essence independent of that subject. It is interesting to note that at the same time Ostrogradsky, when dealing with concrete practical problems, completely renounces his subjectively idealistic interpretation of probability, and regards it as a quite objective numerical measure of phenomena of a certain kind. It is sufficient to recall, for example, his belief in a real possibility of reducing to 1/20 the tiresome work of examining the quality of delivered articles to convince ourselves that, concerning problems in natural sciences, Ostrogradsky was a materialist – a spontaneous materialist, as we would say now.

In the first part of his paper Ostrogradsky argues with the followers of Saint-Simon who had declared, as is seen from a passage quoted by him, that if all the balls were perfectly similar and the bowl was such that all of them had equal chances of being drawn, then no ball will ever be drawn, and the entire theory of probability becomes therefore senseless. Here we

encounter, in another setting, the generally known story about the Buridan's ass. We shall not repeat Ostrogradsky's critical remarks since they occur, in essence, in the previous passages.
In the second part of his article Ostrogradsky offers a popular exposition of the concept of expectation. To illustrate, he asks:

> *Suppose that you obtained a lottery ticket, of course not for free; it costs something, and at the same time you bought a hope of winning. But did not you pay too much? It wouldn't be bad to know how much does it* {really} *cost.*

And he adds ironically:

> *We note in passing that the newspaper advertisements announcing lotteries always omit an important circumstance without which you have no means to estimate your hopes of winning. You usually read: The lottery has so much articles of excellent quality to be won, the tickets costing so much can be obtained in such-and-such a place. The total number of tickets is never mentioned, and without it you are unable to calculate the probability of winning, i.e., you cannot determine an element required for estimating the chances. Thus, you pay for a hope of winning not knowing its price, and you are always paying in excess.*

After offering such introductory explanations of the meaning of expectation illustrated by lotteries and the expected winnings, Ostrogradsky goes over to insurance. He had not achieved much: he posed a question about the methods of calculating the cost of insuring an article, considered a numerical example and, having determined this cost in one case, advised his readers:

> *Do not pay more; on the contrary, try to pay less so as to gain something. Do not worry about the insurance society: it will not incur losses.*

I was unable to find the promised sequel.

9. The note [5] is similar in essence to the paper on insurance [4]. It is also devoted to explaining to the general public that all kinds of chance undertakings,– lotteries, games,– are disadvantageous to them, but fetch a sure and easy profit for the businessman in charge. Somewhat earlier, in 1840, Buniakovsky [12] expressed the same idea: a lottery, he argued, was "a tax levied on ignorance" [12].
Ostrogradsky considered in detail the game of craps, where several dice are thrown on a table and the winner is determined by the number of points achieved. He calculated the probabilities of getting various scores in a throw of two, three, …, twelve dice, and on an adduced plate he offered an appropriate table. Ostrogradsky then asked his readers: "In Pavlovsk, you have undoubtedly seen this game played with eight dice. Would you like to find out whether it is profitable?" He did not know the going payouts and restricted his explanation with purely theoretical calculations and an introduction of a simplest idea of expectation. He concluded by formulating a rule for determining the most profitable (the most probable) number of points. […] Ostrogradsky's inference is the same as the one he vividly expressed earlier [4]:

> *You say that such-and-such had won the main prize and all at once gained a fortune – so why won't I also win the same prize? You are unwittingly*

comparing yourself with the winner, and do not want to admit that it is much more natural to place yourself among the losers since they are by far more numerous.

10. In 1858, the Scientific Council of the Mikhailovsky Artillery School, where Ostrogradsky taught for many years, resolved to establish optional courses to help cadets develop scientific initiative and widen their mental outlook. Ostrogradsky announced an elective course in the theory of probability. A brief report about it is to be found in [22, p. 265], and we quote:

Hundreds of cadets gathered for the first lecture of the celebrated professor. Ostrogradsky, well dressed in a tail-coat and wearing his order, described to his audience, with singular grace and remarkable simplicity, making hardly any use of the blackboard, the origin of probability theory and its principles. Everyone was delighted and all at once several listeners expressed their wish to write down and print the lectures. The second lecture was still attended by very many cadets, but this time Ostrogradsky had to use chalk and he was less enthused. The third one gathered a small number of listeners, only five attended the fourth one, and this number did not increase. Three persons were present on the last (*the twentieth*) *lecture and the printing of the notes was discontinued after the third one.*

That the number of listeners had sharply declined towards the end of the course is a rather usual phenomenon that also takes place in our time. For us, it is essential that three of the lectures were printed, apparently by a lithographic process. I have not seen this edition that deserves a detailed study because the first lectures are especially interesting in the methodological sense, and, besides, because they allow to judge to a certain extent the scope of the entire course. It is likely that the brief historical essay and some general reasoning on probability kept at the State Ukrainian Public Library [8, sheet 904] indeed constitute an outline of the introductory lecture of 1858. Since it is of possible interest, I reproduce its subject-matter in full. I invented some words which I was unable to decipher in accord with the context and placed them in brackets.

The theory of probability ought to be attributed to the sciences of the modern times since its real beginning does not go back further than the mid-17th century. True, some subjects belonging to this science have been known in very remote times: calculations based on mean longevity were made continually; marine insurance was practised, the number of chances in games of chance, although only in the simplest of them, were known; and the stakes in fair games were determined.
However, such conclusions were not subordinated to any rules. […]
Before the mid-17th century problems concerning probabilities were not subjected to mathematical analysis, and no rigorous general rules for their solution had existed.
Pascal, and, following him, Fermat, geometers of the 17th century, are justly considered the founders of the science of probability. The first problem belonging to this science was rather difficult. {Ostrogradsky describes the problem of points.} *It is remarkable that the name of Chevalier de Méré, a man of the world with no* [success in the field of – B.G.] *mathematical sciences, remains forever in their history.*
However, we ought to deal not with the history of the calculus of

probability, but with the science itself. The first question presents itself here: What should be meant by probability? Let us analyse several phrases containing this word and see what sense should be assigned to it.
Tomorrow, it is said, the weather will probably be fine. […]

Here we again encounter the same methodological approach to the concept of probability as we already saw in [4].

11. Ostrogradsky's last article on probability theory [7] was published in 1859, *i.e.*, at the very end of his creative life. It illustrates how a more general point of view simplifies exposition, and how a problem, whose elucidation in former times had taxed the wits and efforts of the best minds, is clarified to the utmost. Nowadays the Bayes formula belongs to the most elementary results of the theory; it is derived almost trivially when issuing from the definition of conditional probability and the formula of total probability [13]. In Ostrogradsky's times, however, it was necessary to consider numerous cases, and to formulate principles, only then passing onto the derivation of the formula itself in one or another form, under some conditions or other. Ostrogradsky's article was indeed devoted to deriving this formula.

He indicated that Laplace had considered without proof only the case in which the prior probabilities of the hypotheses coincided and put forth this formula as a principle. Later Gauss proved Laplace's principle under the same conditions {no reference provided}[14]. Ostrogradsky derived the Bayes formula without assuming that the prior probabilities of the hypotheses coincided. In concluding, he remarked that Poisson had also proved the same formula but that that proof seemed to him insufficiently straightforward.

12. Among Ostrogradsky's manuscripts at the State Ukrainian Public Library [8] there are pages devoted to probability theory, and we have fully reproduced the essence of one sheet. Sheets 443 – 466 contain drafts of [3]. In addition to calculations included in its final text, there are many pages with numerous uncompleted attempts at proving the identity of some expressions. Other pages contain initial formulations, problems, a study of the change of the probability depending o some parameters (this is also included in the final text), an outline of a derivation of the main formula of §7 {omitted there} as well as numerical examples. Sheets 520 – 525, united by a general heading *Probabilities of future events*, only contain fragments which apparently preceded the compilation of [7].

13. We have examined Ostrogradsky' works in the field of probability and became convinced in that all of them were devoted to subjects then strongly exiting science. A correction of Laplace's mistaken conclusions and a derivation of results later obtained by Poisson testify to a sufficient extent that also in probability, although this discipline was situated on the borders of his scientific interests, Ostrogradsky kept abreast with his time. […] His work on probability provide an additional feature for scientifically characterizing him and reveal the breadth of his interests […]

Notes

1. {General sources concerning Ostrogradsky also include Gnedenko (1951; 1984) and Maistrov (1974, pp. 181 – 187) who closely follows the paper now under translation and reprints a number of quotations from Ostrogradsky's writings. Seneta (2001) contributed a new paper on my subject. References to my Notes follow after their main list below.}

2. {Ostrogradsky also attempted to generalize the concept of moral expectation; the only relevant source is Fuss (1836, pp. 24 – 25), see my paper in the *Arch. Hist. Ex. Sci.*, vol. 16, 1976, pp. 170 – 171.}

3. {Aleksei Nikolaevich Krylov (1863 – 1945), a naval architect and applied mathematician.}

4. {Apparently Konstantin Stepanovich Veselovsky (1819 – 1901), an economist and statistician.}

5. Laplace (*Traité de Méc. Cél.*, t. 3, p. xi; my paper in the *Arch. Hist. Ex. Sci.*, vol. 17, 1977, p. 5) thought that hypotheses should be *rectifiant sans cesse par de nouvelles observations*, etc. It is hardly reasonable to criticize classics without a thorough preliminary research.}

6. {Gnedenko did not elaborate.}

7. {Many authors had been opposing the application of probability to jurisprudence; Mill, in 1843, called it an *opprobrium of mathematics* and Poincaré stated that, in law courts, men behave like the *moutons de Panurge*; see Gnedenko & Sheynin (1992, p. 280n) and my paper in the *Arch. Hist. Ex. Sci.*, vol. 9, 1973, p. 296, with a description of pertinent events in the mid-19th century. Note also that Laplace [19, p. 523] (and obviously Poisson) considered the case of independent jurors; they apparently thought of investigating the ideal conditions, – of estimating the utmost possibilities of law courts. Then, contrary to Gnedenko's context (also see end of his §5), such applications of probability are nowadays being resumed (see Heyde & Seneta 1977, p. 34), who provide several references, and Eggleston 1983).}

8. {Also Zernov (1843).}

9.{I specify a few points. First, the memoir was lost and only its Extrait was published. Second, Ostrogradsky provided an exact reference to Laplace [19, pp. 9ff]. Third, he indeed showed that Laplace's mistake could have led to absurdity, but his example was invented rather than borrowed from [19]. In 1959 – 1961 Ostrogradsky's *Полное собрание трудов* (Complete Works) were published in Kiev in three volumes and the Exrait was included in vol. 3 with a commentary by Gnedenko (pp. 347 – 348). Yes, Gnedenko mentioned that the memoir itself was lost, but he did not specify the other points, see above.}

10. {In effect, Buniakovsky went back on his own words and adopted an objective approach; see my paper in the *Arch. Hist. Ex. Sci.*, vol. 43, 1991, p. 202.}

11. {This statement is much too restrictive; nature has both necessity and randomness.}

12. Cf. Petty (1662, p. 64): *A lottery* […] *is properly a Tax upon unfortunate self-conceited fools* …

13. {Gnedenko had not dwelt on the importance of the Bayes formula and approach or on the pertinent protracted debates in the 20th century.}

14. {In 1809, Gauss justified the principle of inverse probability for the case of equal probabilities of the various hypotheses.}

References

M.V. Ostrogradsky

I mentioned his *Complete Works* in Note 9, see above.

1. Extrait d'un mèmoire sur la probabilité des erreurs des tribunaux. *Bull. Scient.*, No. 3, 1834, pp. xix – xxv. The *Bulletin* was appended to the *Mém. Acad. Imp. Sci. St.-Pétersb.*, 6me sér., Sci. math., phys. et naturelles, t. 3, this being Sci. Math. et phys, t. 1.

2. Sur le calcul des fonctions génératrices. *Bull. Scient.*AN Psb., t. 1, No. 10, 1836, pp. 73 – 75.

3. Sur une question des probabilités. *Bull. Phys.-Math.* AN Psb., t. 6, No. 21 – 22, 1846, pp. 321 – 346.

4. On insurance. *Finsky Vestnik*, vol. 13, No. 1 1847, pp. 29 – 34; No. 2, pp. 40 – 44. (R)

5. Dicing. Ibidem, No. 3, pp. 29 – 32. (R)

6. Note on the retirement fund [25, pp. 93 – 101 + Tables]

7. Sur la probabilité des hypothèses d'àprès les événements. *Bull. Phys.-Math.* AN Psb, t. 17, No. 33, 1859, pp. 516 – 522.
8. Manuscripts kept at the manuscript section, Ukrainian State Public Library, Ostrogradsky's Papers.

Other Authors

9. Bernstein, S.N. (1946), *Теория вероятностей* (Theory of Probability). Fourth ed. M. – L.
10. Brashman, N.D. (1841), *О влиянии математичеких наук на развитие умственных способностей* (On the Influence of Mathematcal Sciences on the Development of Intellectual Faculties). M.
11. Buniakovsky, V.Ya. (1846), *Основания математической теории вероятностей* (The Principles of the Mathematical Theory of Probability). Psb.
12. --- (1840), Thoughts about some unfounded notions in social life. *Maiak*, pt. 3, pp. 80 – 94. (R)
13. Chebyshev, P.L. (1845), *Опыт элементарного анализа теории вероятностей* (Essay on the Elementary Analysis of the Theory of Probability). *Полное собрание сочинений* (Complete Works), vol. 5. M.– L., 1951, pp. 26 – 87.
14. --- (1867), Les valeurs moyennes. *J. math. pures et appl.*, t. 12, pp. 177 – 184.
15. Kozlovsky, Prince (1836), On hope. *Sovremennik,* vol. 3, pp. 23 – 47. (R)
16. Krylov, A.N. (Manuscript), To the memory of Ostrogradsky. Rossiysk. Akad. Nauk, Archive, V, Inventory 1 – 0, No. 11. (R)
17. Lakhtin, L.K. (1901), Ostrogradsky's work in analysis.MS*,* vol. 22, No. 4, pp. 540 – 554. (R)
18. Laplace, P.S. (1814), *Essai philosophique sur les probabilités. Oeuvr. Compl.*, t. 7, No. 1. Paris, 1886. Separate paging. English translation : *Philosophical Essay on Probabilities*. New York, 1995.
19. --- (1812), *Théorie analytique des probabilités. Oeuvr. Compl.*, t. 7. Paris, 1886.
20. Liapunov, A.M. (1902), Ostrogradsky's merits in mechanics [29, pp. 115 – 118].
21. Pavlovsky, A.F. (1821), *О вероятности* (On Probability). Kharkov.
22. Platov, A., Kirpichev, L. (1870), *Исторический очерк образования и развития Михайловского артиллерийского училища* (Historical Essay on the Establishment and Development of the Mikhailovsky Artillery School). Psb.
23. Poisson, S.D. (1837), Sur la probabilité du tir à la cible. *Mémorial de l'artill.*, No. 4, pp. 59 – 94.
24. --- (1837), *Recherches sur la probabilité des jugements,* etc. Paris.
25. Anonymous (1858), *Предположение об учреждении ... пенсионной кассы* (Proposal on Establishing a Retirement Fund at the Naval Department). Psb.
26. Sabinin, E.F. (1901), Ostrogradsky. MS, vol. 22, No. 4, pp. 499 – 531. (R)
27. Steklov, V.A. (1902), On the work of Ostrogradsky in mathetical analysis [29, pp. 118 – 127].
28. Tikhomandritsky, M.A. (1902), An essay on Ostrogradsky's work in pure mathematics [29, pp. 92 – 115].
29. Tripolsky, P.I. (1902), *M.V. Ostrogradsky.* Poltava. (R)
30. Zhukovsky, N.E. (1901), On the scientific work of Ostrogradsky in mechanics. *Полное собрание сочинений* (Complete Works), vol. 9, 1937; vol. 7, 1950. M. – L.

References to Notes

Eggleston, R. (1983), *Evidence, Proof and Probability*, 2nd edition. London.

Fuss, P.N. (1836), Compte rendu de l'Académie pour l'année 1835. In *Recueil des actes de la séance … de l'Acad. Imp. Sci. St.-Pétersb. 29 Dec. 1835.* Psb. Separate paging.
Gnedenko, B.V. (1951), Ostrogradsky. *Uspekhi Matematich. Nauk,* vol. 6, pp. 1 – 25. (R)
--- (1984), *Ostrogradsky*. M. (R)
Gnedenko, B.V., Sheynin, O. (1992), Theory of probability. In *Mathematics of the 19th century*, vol. 1. Editors, A.N. Kolmogorov, A.P. Youshkevich. Basel, pp. 211 – 288. Originally published in Russian in 1978.
Heyde, C.C., Seneta, E. (1977), *I.J. Bienaymé*. New York.
Maistrov, L.E. (1974), *Probability Theory. Historical Sketch*. New York. Originally published in Russian in 1967.
Petty, W. (1662), Treatise of taxes and contributions. In *Economic Writings*, vol. 1. Cambridge, 1899, pp. 1 -97.
Seneta, E. (2001), Ostrogradsky as probabilist. *Ukrainsk. Matematich. Zhurnal, vol. 53, pp. 1038 – 1046.*
Zernov, N. [*E.*] (1843), *Теория вероятностей* (Theory of probability). M.

Part 2. The Petersburg School of the Theory of Probability

5. S.N. Bernstein. Chebyshev's Work in the Theory of Probability (1945).

In *Собрание сочинений* (Coll. Works), vol. 4. N.p., 1964, pp. 409 – 433

It is obvious that Bernstein had not studied the early history of probability. This apparently goes to show the general situation, at least in Russia in 1945, when he published his paper, and perhaps in 1964, when it was reprinted. Several of my Notes indicate his pertinent particular mistakes and shortcomings.

1. The number of Chebyshev's writing in probability is not large. There are only four of them: his Master Dissertation [1] followed by a paper [2]; then, after more than 20 years, an article [3] published in two periodicals at the same time; and, finally, after another 20 years, his last writing [4]. Papers [2] and [3] are reprinted in vol. 1 (1899) of his collected works published by the Petersburg Academy of Sciences, and article [4] is in vol. 2 (1907) of the same source. Neither his dissertation [1], nor his course of lectures in probability read at Petersburg University from 1860 to 1882 and published only recently from the extant notes taken in 1879 – 1880 by his illustrious student, Liapunov, were included in this collection.

That Chebyshev, barely graduating from the [Moscow] university, chose probability as the subject of his Master Dissertation, testifies that he became especially interested in this peculiar branch of mathematics from his earliest years.

2. At its origin, the theory of probability is known to have been far from the general movement of the natural sciences, and the only experiments by whose means its most important notions and main principles had been developed and specified, were provided by games of chance. It was on these grounds that Jakob Bernoulli, more than 200 years ago, discovered his celebrated theorem that offered a clue to understanding the origin of mass regularities out of independent individual fortuities and represented the first rigorously proved, although very particular, formulation of the law of large numbers.

Laplace's classical researches, and, above all, the De Moivre – Laplace limit theorem, constituted the next stage in the development of probability theory. This proposition established the limiting law of probability

$$G[(x-a)/\sigma] = (1/\sqrt{2\pi}) \int_{-\infty}^{(x-a)/\sigma} \exp(-z^2)dz$$

for the deviations of the number (x) of the occurrences of some random event E from the expectation $a = np$ of x in n Bernoulli trials with E having a constant probability p [and the variance σ^2 of x being {now} known to equal $np(1-p)$]. The law of large numbers and the Laplacean normal law $G(x)$, both of them being of fundamental methodological importance, were thus established in their simplest forms already in the 18th century [1]. Another of Laplace's considerable merit was that he had foreseen the extreme generality of these laws and led the theory of probability out from the narrow confines of games of chance onto the vast domain of scientific natural sciences [2].

Owing to Laplace's influence, the first half of the last {of the 19th} century was marked by a heightened interest in probability and emotion for its applications. Many of these were, however, insufficiently justified; and some, even those supported by Laplace himself and Poisson, were so obviously mistaken that Mill later on well-deservedly qualified them as a *mathematical opprobrium* [3].As a result of ensuing failures {?}, the enthusiasm gave place to disappointment and West-European mathematicians were becoming ever more convinced that the theory was only a mathematical entertainment of sorts not admitting of essential scientifically substantiated applications and hardly warranting the attention of serious scientists.

3. Chebyshev regarded the crisis point reached by probability theory in another way. Owing to the practicality natural to him, and without dwelling on justifying it as a method of scientific research by deep philosophical considerations (that only became possible in our time), he understood that, since there existed real random phenomena (games of chance) that confirmed with maximal precision its simplest mathematical inferences, the substance and the more general conclusions of the theory should find practical applications. To illustrate: the applicability of geometric formulas to the calculation of the area of a rectangle leads to their applicability to the cases of more complicated but precisely enough determined flat figures.

It was only necessary to construct, within the context of the same main notions, more flexible but precisely defined theoretical patterns and to adjust them to the properties of the random events and magnitudes observed in reality. As he said in his dissertation,

The science of probability, that goes by the name of the theory of probability, has as its subject the determination of the probability of an event given its {the event's} connection with events whose probabilities are known.

In particular, for Chebyshev the pivotal task was the solution of the two principal problems on which the subsequent destiny of the theory depended: To formulate precisely, and to prove rigorously for the widest possible class of random phenomena, first, the law of large numbers; and, second, the limit theorem for sums of independent random variables. Chebyshev formulated these problems, by no means equally difficult, while in his first youth. He brilliantly solved the first of them [3] whereas the second one remained the object of his deep thoughts to the end of his life, and, in essence, he completely solved it only in 1887 [4]. We shall now try to outline, as far as it is possible when studying his published works, the main stages of Chebyshev's development as a specialist in probability.

4. We shall not dwell on Chebyshev's first paper [1]; it is important only in its goal which was to formulate precisely the general theorems of the theory of probability and to prove them while advancing to the forefront inequalities and the estimation of the error of limiting formulas. For Chebyshev, the law of large numbers as well as the limiting normal law thus made sense not as mathematical characteristics of some infinite sets, but as approximations to quantitative relations observed in sufficiently numerous and really existing structures of some random elements. And it was essential that the word *sufficiently* be quite definitely interpreted in a mathematical way.

The main principles of the theory of probability as indicated by Chebyshev in the beginning of his dissertation, which did not differ in essence from those adopted by his predecessors, are reduced to the following:

If, under known circumstances, out of a definite number of different events one must necessarily occur, and there is no special reason for expecting some of these events rather than the other ones, then we distinguish such events by calling them equally possible occasions. Thus, 1 and 0 are the limits of the probability of events; it reaches the former by increasing in cases of necessary events; and it reaches the latter by decreasing in cases of impossible events. For all other events, having neither necessity nor impossibility, probability remains different from 1 and 0. We consider it approximately doubtless that events will, or will not take place if their probabilities differ but little from 1 or 0. Such are all the conclusions inferred from observations and testimonies.

Chebyshev undoubtedly saw well enough the inadequacy of the *elementary algebraic method* that he applied. In particular, he [1] proved the Poisson theorem (the law of large numbers) only for a restricted number of different probabilities. While seeking new approaches, he apparently only discovered a general elementary proof of this theorem complete with an appropriate estimation of error when [1] was being printed. The substantiation represented the first example of an extremal reasoning characteristic for his entire later work. It was contained in Chebyshev's article [2] [4] that is a valuable original supplement to his dissertation. I adduce the typical for Chebyshev beginning of this article:

The subject of this note will be the proof of the following proposition: It is always possible to assign such a large number of trials [5] *that the probability that the ratio of the number of the occurrences of some event E to the number of trials will not deviate from the arithmetic mean of the probabilities of E beyond given boundaries no matter how close they are one to another, and that that probability will be arbitrarily close to certainty.*
This main proposition of probability theory that includes as a particular case the Jakob Bernoulli law, was derived by Poisson [19, *Chapter* 5] *who issued from approximate calculations of some rather complicated definite integral. However, regardless of the cleverness of the method used by the celebrated Geometer, it does not provide the limit of the error allowed by his approximate analysis, and because of this uncertainty in the value of the error the demonstration does not possess appropriate rigor.*

Chebyshev's proof is based on the following remark. The probability P_m that the event E will occur not less than m times in μ trials is equal to some expression, symmetric with respect to all the p_k, $k = 1, 2, \ldots$, which are the probabilities of E at trials k, and linear in each p_k, for example, with respect to p_1 and p_2. Thus,

$$P_m = U + V(p_1 + p_2) + Wp_1p_2$$

where U, V, W do not depend on p_1 or p_2 anymore. Therefore, if $0 < a = p_1 + p_2 \le 1$ then $P_m = U + Va + Wp_1p_2$ takes its maximal value either at $p_1 = p_2 = a/2$, or at $p_1 = 0$. The case of $a > 1$ is reduced to the previous one by replacing the expression of P_m by a similar formula including $q_1 = 1 - p_1$ and

$q_2 = 1 - p_2$. A theorem of a specifically *Chebyshev type* is now easily derived: The maximal value of P_m corresponds to such values $p_1, p_2, \dots, p_\mu$ that

$$p_1 = p_2 = \dots = p_\rho = 0,\ p_{\rho+1} = \dots = p_{\rho+\sigma} = 1,$$
$$p_i = (S - \sigma)/[(\mu - (\rho + \sigma)],\ \rho + \sigma < i \le \mu,\ p_1 + p_2 + \dots + p_\mu = S,$$

and $\rho \ge 0$ and $\sigma \ge 0$ are some definite numbers. Thus,

$$P_m \le \frac{(\mu-\rho-\sigma)!}{(m-\sigma)!(\mu-m-\rho)!}\left(\frac{S-\sigma}{\mu-\rho-\sigma}\right)^{m-\sigma}\left(\frac{\mu-S-\rho}{\mu-\rho-\sigma}\right)^{\mu-m-\rho}\cdot$$
$$\left(1+\frac{\mu-m-\rho}{m-\sigma+1}\frac{S-\sigma}{\mu-S-\rho}+\dots\right).$$

The terms of the sum decrease more rapidly than a geometric progression with ratio

$$\frac{\mu-m-\rho}{m-\sigma+1}\cdot\frac{S-\sigma}{\mu-S-\rho};$$

replacing them by such a progression, we find that

$$P_m < \frac{(\mu-\rho-\sigma)!}{(m-\sigma)!(\mu-m-\rho)!}\left(\frac{S-\sigma}{\mu-\rho-\sigma}\right)^{m-\sigma}\left(\frac{\mu-S-\rho}{\mu-\rho-\sigma}\right)^{\mu-m-\rho+1}\cdot\frac{m-\sigma}{m-S}.$$

Chebyshev then notes that, if $m > S + 1$, the right side of this inequality increases with the decrease of the integers $\rho > 0$ and $\sigma > 0$, so that, assuming that $\rho = \sigma = 0$, we derive the simple inequality

$$P_m < \rho_m m/(m - S) \qquad (1)$$

where ρ_m is the probability that the number of the occurrences of the event E is m when $p_i = p = S/\mu$ (*i.e.*, when the Poisson pattern is replaced by appropriate Bernoulli trials). After a similar estimation of the probability that E will occur not more than $m < S - 1$ times, the demonstration is easily concluded. Thus, the very precise value of the lower boundary, derived in [1] for the number of trials μ sufficient for the probability of the deviations exceeding the given limits to be lower than some arbitrarily small magnitude, can be applied to the Poisson general case.

The idea about the existence of an inequality of type (1) might have possibly occurred to Chebyshev in connection with the fact that could have hardly escaped his notice,– that for μ independent trials corresponding to a given mean probability $p = S/\mu$ the variance of m is maximal if all the probabilities are equal one to another. Be that as it may, the peculiar clever demonstration of this inequality was only an elegant episode in Chebyshev's work, unconnected with the method of expectations or moments from which he was then far and which he created much later when solving concrete problems belonging to other branches of mathematics. Laplace's *transcendental* methods did not satisfy him whereas the usual algebraic methods occurred to be too weak for enabling to extend essentially the domain of reliable applications of the theory of probability.

5. During the next years after defending his Master Dissertation Chebyshev is known to have written a number of remarkable works on integrating algebraic functions by means of elementary functions and to have immortalized himself by solving classical problems on the distribution of primes that did not yield to the efforts of greatest mathematicians.

The solution of problems concerning algebraic integrals led Chebyshev to a deep study of the properties of algebraic continued fractions which soon became his favorite tool, exceptionally powerful and fruitful in his hands. However, it is difficult to assume that already then Chebyshev had foreseen how he will use it in probability. Such an idea might have first crossed his mind after he, in 1855, had applied continued fractions to interpolation by the method of least squares [6]. It would have been an essential stretch of imagination to attribute Chebyshev's works on the method to probability. Their direct aim was to change expediently the technique of calculation so as to specify when required the obtained approximate expressions in the simplest and shortest way; the formal algebraic transformations do not possess here any stochastic meaning. For Chebyshev, the deep theoretical essence of interpolation by least squares consisted not in its possible connection under more or less arbitrary assumptions with probability; for him, it was most important that the method provided a natural constructive approach to expanding an arbitrary empirical function in a series of polynomials converging (in some sense, in a best way) over a given interval of any length whereas the Taylor – Maclaurin series only sufficiently approximates functions at small values of the independent variable.

Owing to his instrument of continued fractions, Chebyshev went by algebraic means far beyond classical algebra and entered the immense field of the general theory of functions. The entire domain of the unreliable *transcendental* methods of analysis now became accessible to his precise trustworthy methods of augmented algebra. And, by their means, he must {now} appropriately formulate and solve in all algebraic rigor the abovementioned main problems of the theory of probability.

Indeed, a parabolic interpolation of a function $\varphi(x)$ given its values at m points x_i by least squares is known to consist in determining a polynomial $P_n(x)$ of degree $n < m + 1$ under the condition that the mean square error (with assigned weights $\theta^2(x)$)

$$I_n^2 = \sum_{i=1}^{m} [P_n(x_i) - \varphi(x_i)]^2 \theta^2(x_i), \quad \sum_{i=1}^{m} \theta^2(x_i) = 1 \qquad (2)$$

be as small as possible. For a finite m the minimal value of I_n^2, decreasing with the increase in n, vanishes when $n = m - 1$ and $P_n(x)$ then becomes the Lagrange interpolation polynomial.

Chebyshev's essential and practically very useful simplification of the solution of this algebraic problem, on whose details we shall not touch, consisted in that he represented the Lagrange interpolation polynomial $P_{m-1}(x)$ as

$$P_{m-1}(x) = \sum_{k=0}^{m-1} A_k \psi_k(x) \qquad (3)$$

where $\psi_k(x)$ were such polynomials of degree $k \le m - 1$ that for *any* $n < m - 1$ the polynomials

$$P_n(x) = \sum_{k=0}^{n} A_k \psi_k(x) \qquad (4)$$

ensured the minimal value of I_n^2. It occurred that this property was possessed by the denominators $\psi_i(x)$ of the consecutive convergents of the continued fractions emerging from

$$\sum_{i=1}^{m} \theta^2(x_i)/(x - x_i)$$

which were thus orthogonal polynomials with weights $\theta^2(x)$. In other words, they obeyed the conditions

$$\sum_{i=1}^{m} \theta^2(x_i)\, \psi_s(x_i)\, \psi_i(x_i) = 0,\ s \neq i. \qquad (5)$$

Chebyshev soon extended his conclusions onto the case of $m \to \infty$ so that the finite sum (2) was replaced by an integral. Solving the problem of bringing the integral

$$I_n^2 = \int_a^b [P_n(x) - \varphi(x)]^2 \theta^2(x)\, dx \qquad (6)$$

to its minimal value in a similar way and basing his work on the algorithm of continued fractions, he created the general theory of expanding an arbitrary function in a series of orthogonal polynomials. Thus, it is impossible not to admit that Chebyshev was the founder of this central direction of the modern theory of functions. True, his expansions in the most important orthogonal polynomials were largely formal and he did not appropriately examine the conditions for their convergence. By analogy with a finite m, when the minimal value of the mean square error I_n^2 vanishes at $n = m - 1$, Chebyshev argued without proof that in his examples the same happened as $n \to \infty$ also in the case of (6).

6. Chebyshev's general ideas described above came into contact with probability theory in his paper [7] read in 1859. There, for the first time, he provided the expansion of an arbitrary function $F(x)$ on the entire real axis into polynomials

$$\psi_l(x) = \exp(kx^2)\,[d^{\,l} \exp(-kx^2)/dx^l] \qquad (7)$$

which, as he showed, were the denominators of the convergent continued fractions

$$\sqrt{k/\pi} \int_{-\infty}^{\infty} \frac{\exp(-ku^2)}{x-u}\, du \qquad (8)$$

so that

$$F(x) = \sum_{l=0}^{\infty} A_l \psi_l(x), \qquad (9)$$

$$l!(2\mathrm{k})^l A_l = \sqrt{k/\pi} \int_{-\infty}^{\infty} \exp(-kx^2)\psi_l(x)F(x)\, dx.$$

If $f(x) = F(x) \exp(-x^2/2)$, the series (9) becomes the density of any normed law

$$\int_{-\infty}^{\infty} xf(x)\,dx = 0, \int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} x^2 f(x)\,dx = 1$$

with

$$f(x) = [\exp(-x^2/2)/\sqrt{2\pi}\,][1 + \sum_{l=3}^{\infty} A_l \psi_l(x)]. \tag{10}$$

After integrating term by term we obtain

$$\int_{-\infty}^{x} f(x)\,dx = (1/\sqrt{2\pi}\,) \int_{-\infty}^{x} \exp(-x^2/2)\,[1 + \sum_{l=3}^{\infty} A_l \psi_l(x)]\,dx. \tag{10bis}$$

Chebyshev applies this formula in his memoir [4] where the coefficients A_l are expressed in terms of the moments of

$$X_n = (1/\sqrt{B_n}\,) \sum u_i,–$$

of the normed sum of independent variables u_i. It is therefore natural to assume that, after discovering the formal expansion (10bis), Chebyshev already in 1859 more or less clearly saw the approach to proving the limiting theorem of the theory of probability. The demonstration should have established some boundary for

$$(1/\sqrt{2\pi}\,)| \int_{-\infty}^{x} \exp(-x^2/2) \sum_{l=3}^{\infty} A_l \psi_l(x)\,dx|$$

tending to zero when the number n of the terms u_i infinitely increases. Since the coefficients

$$A_l = (1/l!) \int_{-\infty}^{\infty} \psi_l(x) f(x)\,dx$$

are completely determined by the moments (*i.e.*, by the expectations of X_n^k, $k \le l$), it is also obvious that the latter acquire special importance. However, Chebyshev was certainly fully aware that the analysis that confronted him was of the same order of difficulty as the problem of imparting necessary rigor to Laplace's *transcendental* methods.

7. From 1860 Chebyshev took over from Buniakovsky the teaching of probability theory at Petersburg University which apparently imparted an additional stimulus for directing his thoughts to this field. It seems, however, that for a number of years the limit theorem did not yield to his efforts and remained an unattainable goal. Much later, in his course [5], Chebyshev explicated the Laplace method without mentioning the method of moments and turned the students' attention to its inadequacy:

In its present state, mathematical analysis cannot derive this boundary {of the ensuing error} *in any satisfactory fashion.*

On the contrary, in 1866 Chebyshev brilliantly solved the first problem which was to provide a general elementary proof of the law of large numbers easily understood by average students. Introducing absolute clarity into mathematical definitions and reasoning into probability, and, in particular, precisely establishing the general properties of expectations, he was able to arrive at his celebrated derivation of the so-called *Chebyshev inequality* from which directly followed the classical formulation of the law of large numbers:

If the expectations of the squares of [independent]*variables* u_1, u_2, u_3, … *do not exceed a given finite boundary, the probability that their arithmetic mean N differs from the arithmetic mean of their expectations less than by some given magnitude tends to unity as N increases to* ∞.

Chebyshev reported this finding to the Academy of Sciences in 1866 and published it next year [3] in two periodicals. One of these carried, in the same volume, a reprint of a memoir by the French mathematician Bienaymé that included an inequality similar to the Chebyshev inequality written out in a somewhat less general form. This fact to some extent deprives our great compatriot of priority. Markov [17d, 1924, p. 89] used the term *Bienaymé – Chebyshev inequality* and justified it in the following way (p. 92):

We connect the two names, Bienaymé and Chebyshev, with this remarkable simple inequality since Chebyshev first clearly formulated and proved it, but Bienaymé had much earlier indicated the main idea of the demonstration. The inequality itself, only accompanied by some particular assumptions, can be found in his memoir [12].

Bienaymé's memoir aimed at defending Laplace's stochastic justification of the MLSq which Cauchy had called in question. In getting to know Bienaymé's work, Chebyshev apparently found ideas akin to his own notions. Accordingly, we shall quote the entire beginning of his report [8] […]{see [23, p. 261]}.

Actually, Bienaymé's article had not contained other applications of the method that he recommended, but his reasoning was completely in keeping with Chebyshev's own train of thought. This was the reason why the latter, being prepared for the solution of the general problem posed by Bienaymé by his previous research and perceiving in its solution a sure path to proving the limit theorem, continued his deliberations in this direction. As a result, he discovered his celebrated {integral} inequalities [8]. Let

$$c_0 = \int_A^B f(x)dx,\ c_1 = \int_A^B xf(x)dx,\ \ldots,\ c_m = \int_A^B x^m f(x)dx$$

so that formally

$$\int_A^B \frac{f(x)dx}{z-x} = \frac{c_o}{z} + \frac{c_1}{z^2} + \ldots + \frac{c_m}{z^{m+1}} + \ldots \qquad (11)$$

If $\varphi(z)/\psi(z)$ is one of the convergents obtained by expanding the left side of (11) into a continued fraction, and if $z_1, z_2, \ldots, z_l, z_{l+1}, \ldots, z_n, \ldots, z_m$ are the roots of the equation (of degree m) $\psi(z) = 0$ arranged in increasing order, then, each time when $f(x)$ remains positive between boundaries A and B,

$$\sum_{i=l+1}^{n-1} \frac{\varphi(z_i)}{\psi'(z_i)} < \int_{z_l}^{z_n} f(x)dx < \sum_{i=l}^{n} \frac{\varphi(z_i)}{\psi'(z_i)} \qquad (12)$$

and the boundaries in these inequalities cannot be brought closer to each other.

In short, the inequalities (12) represent the fact that for all possible positive functions $f(x)$, for which the expansion (11) into a continued fraction has the same convergent $\varphi(z)/\psi(z)$ (*i.e.*, the same moments $c_0, \ldots, c_m$), the extreme values of the integral in (12) correspond to the case in which the continued fraction (11) is finite.

8. This fundamental theorem shows that, having established inequalities (12), Chebyshev found a general method for solving the main problem of the theory of probability: To what extent is the law of probability $f(x)$ of a random variable determined by the given appropriate expectations of consecutive degree (or the so-called *moments*), which, rather than being arbitrary, are known to satisfy some conditions?

In addition, these inequalities open the way for answering the question of whether $f(x)$ is determined uniquely when the expectations of all the degrees of the variable are given (the problem of uniqueness in the theory of moments).

Indeed, it follows from (12) that a necessary and sufficient condition for this uniqueness is that the difference between the consecutive roots z_l and z_{l+1} of the denominators $\psi(z)$ (of degree m) of the appropriate convergents tended to zero for any l. Indeed, in this, and only in this case any real number a is the limit of z_l and the difference between the extreme parts of (12) tends to vanish. Thus, Chebyshev now found the clue to proving the limit theorem: a deeper study of the continued fraction (8), whose most important properties he came to know already in 1859 (§6), will show that the totality of all the moments of the normal law uniquely determines it. Then, it still remained necessary to specify the demonstration of the fact that, when the number n of independent variables $u_1, u_2, \ldots, u_n$ increases, the magnitudes

$$\mathrm{E}(u_1 + u_2 + \ldots + u_n)^l / B_n^{l/2},$$

where

$$B_n = \mathrm{E}(u_1 + u_2 + \ldots + u_n)^2 = \mathrm{E}u_1^2 + \mathrm{E}u_2^2 + \ldots + \mathrm{E}u_n^2, \ \mathrm{E}u_i = 0,$$

have as their limits the corresponding moments of order l of the normed Laplace – Gauss normal law. In the sequel, we shall see that this fact was apparently obvious to Chebyshev already in his young years, when he had been studying classical literature, and that it served him as a guiding star.

However, he still had no rivals in the new fields of mathematics created by him, and he did not hurry either with publishing the proof of the inequalities discovered in 1873 [8] or with surmounting the last obstacles separating him from his goal contemplated long ago. Indeed, Chebyshev's efforts during the entire next decade had been directed to the other branches of his versatile creative work, mostly to examining and constructing mechanisms on the basis of his theory of functions least deviating from zero, which was another aim also delayed by him for a long time.

In the meantime, however, Markov published a paper [17a] with a general, very clever and simple proof of the abovementioned Chebyshev inequalities. After the appearance of this remarkable work, whose author deeply penetrated the essence of the ideas, and simplified the methods of his teacher, Chebyshev at once proceeded to conclude his prolonged investigation of the limit theorem for the sums of a large number of independent variables.

9. Chebyshev made a decisive step in this direction in a memoir [9] reported to the Academy of Sciences in 1886. After some modification of the extreme parts of inequalities (12) which followed from the identical dependences between the numerators and the denominators of the consecutive convergents of continued fractions, Chebyshev applied here his inequalities for estimating the boundaries of the integral

$$\int_{-\infty}^{v} f(x)\,dx,\ f(x) > 0$$

if

$$\int_{-\infty}^{\infty} f(x)\,dx = 1,\ \int_{-\infty}^{\infty} xf(x)\,dx = 0,\ \int_{-\infty}^{\infty} x^2 f(x)\,dx = (1/q^2),\ \ldots,$$

$$\int_{-\infty}^{\infty} x^{2m-2} f(x)\,dx = \frac{1\cdot 3 \ldots (2m-3)}{q^{2m-2}}\ \int_{-\infty}^{\infty} x^{2m-1} f(x)\,dx = 0. \qquad (13)$$

The transformation of inequalities (12) mentioned above leads to the conclusion that, if two functions, $f(x) \geq 0$ and $f_1(x) \geq 0$, satisfy conditions (13) (*i.e.*, if they are determined by the moments up to the $(2m-1)$-th order inclusively) corresponding to the normal law

$$(q/\sqrt{2\pi})\int_{-\infty}^{v} \exp(-q^2x^2/2)\,dx,$$

then, for any m,

$$\left|\int_{-\infty}^{v} f(x)\,dx - \int_{-\infty}^{v} f_1(x)\,dx\right| < 1/(T_0 + T_1 + \ldots + T_{m-1}) \qquad (14)$$

where $T_i = \psi_i^2(v)/(i!q^i)$ and

$$\psi_l(z) = \exp(q^2z^2/2)\,[d^l \exp(-q^2z^2/2)/dz^l]$$

are the abovementioned denominators of the convergents of the continued fractions emerging from

$$(q/\sqrt{2\pi})\int_{-\infty}^{\infty} \frac{\exp(-q^2x^2/2)}{z-x}\,dx.$$

Chebyshev derived the required lower boundary of the sum

$$S_{m-1} = T_0 + T_1 + \ldots + T_{m-1} \qquad (15)$$

which increases to infinity with an increasing m (so that the law having all the normal moments should be identical with the normal distribution) by a curious trick. The finite equation

$$\psi_l(v) + q^2v\psi_{l-1}(v) + (l-1)\,q^2\psi_{l-2}(v) = 0$$

($\psi_o(v) = 1$) which determines the polynomials $\psi_l(v)$ connecting the denominators of the consecutive convergents easily provides an equation for $T_l(v)$:

$$lT_l - (q^2v^2 - l + 1)T_{l-1} + (q^2v^2 - l + 1)T_{l-2} - (l-2)\,T_{l-3} = 0.$$

Chebyshev solved this latter by introducing a generating function

$$\theta(t) = T_o + T_1t + T_2t^2 \ldots + T_nt^n + \ldots \qquad (16)$$

After simple transformations the differential equation

$$(1 + t)\,(1 - t^2)\,\theta'(t) = [q^2v^2 - (q^2v^2 - 1)t + t^2]\,\theta(t)$$

is obtained from which

$$\theta(t) = C\frac{\exp[q^2v^2t/(1+t)]}{\sqrt{1-t^2}} \qquad (17)$$

where $C = 1$ since $\theta(0) = T_o = 1$. It immediately follows from (17) that the infinite series of non-negative numbers ($T_1 + T_2 + \ldots + T_n + \ldots$) diverges which means that the sum (15) increases unboundedly as $m \to \infty$. This, however, was not sufficient since, as we saw, for Chebyshev a limiting formula established for $m \to \infty$ was less essential than its error for a finite m, and this very fact reflects the main aim and the virtue of his inequalities (12).

It is interesting that Chebyshev connected his trick, by whose means he obtained the lower boundary for the sum (15) and which is applicable to any power series with non-negative coefficients, with a simple particular case of inequalities (12); and that it actually served as the basis of his classical proof of the law of large numbers. It should also be noted that Chebyshev introduced here, although in other notation, Stieltjes integrals:

So as to apply to the sums

$$\sum_0^\infty T_lt^l,\ T_o + T_1t + \ldots + T_{m-1}\,t^{m-1}$$

what was there [in inequalities (12) – S.B.] *provided for integrals, we represent them as integrals*

$$\int_0^\infty Yt^xdx,\ \int_0^{m-1} Yt^xdx$$

where Y is a function equal to zero at all values of x not adjacent to 0, 1, 2, …; *and for values of x infinitely close to* 0, 1, 2, … *taking such values that the integrals*

$$\int_0^\omega Ydx,\ \int_{1-\omega}^1 Ydx,\ \int_{2-\omega}^2 Ydx,\ \ldots$$

unboundedly approach T_o, T_1, T_2, …*as* ω *tends to zero. For this function, defined as just stated, we shall have*

$$\int_0^\infty Yt^xdx = \sum_{l=0}^\infty T_lt^l = \theta(t),\ \int_0^{m-1} Yt^xdx = T_o + T_1t + \ldots + T_{m-1}\,t^{m-1}.$$

It is sufficient to replace Ydx by dT where T is the appropriate step-function to obtain the Stieltjes integral. It is therefore apparent that Chebyshev regarded the introduction of these integrals into probability theory as a technical necessity; he actually made use of a similar tool in his investigations.

Assuming that

$$m - 1 = [t\,\theta''(t)/\,\theta'(t)] + 1 \qquad \text{(i)}$$

Chebyshev thus arrived [6] at

$$\sum_{i=0}^{m-1} T_i t^{\,i} \geq \theta(t) - \frac{t[\theta'(t)]^2}{t\theta''(t)+\theta'(t)} \quad (0 \leq t < 1) \qquad (18)$$

for any function $\theta(t)$ with non-negative coefficients. This inequality cannot obviously be improved since for

$$\theta(t) = T_o + T_{m-1}\, t^{\,m-1}$$

it reduces to $T_o + T_{m-1}t^{\,m-1} \geq T_o$. If (i) persists, inequality (18) all the more leads to

$$S_{m-1} = \sum_{i=0}^{m-1} T_i > \theta(t) - \frac{t[\theta'(t)]^2}{t\theta''(t)+\theta'(t)}\,. \qquad (19)$$

Issuing from (19) and (17), Chebyshev obtained, after some transformations and simplifications of the right side of (19),

$$S_{m-1} > [2/(3\sqrt{3})]\frac{(m-3)^3\sqrt{m-1}}{(m^2-2m+3)^{3/2}}\frac{1}{(q^2v^2+1)^3}\,. \qquad (20)$$

In virtue of (14) the following theorem takes place: *If function* $f_1(x)$ *remaining positive satisfies* (13), *then*

$$\left| \int_{-\infty}^{v} f_1(x)\,dx - (q/\sqrt{2\pi}\,) \int_{-\infty}^{v} \exp(-q^2x^2/2)\,dx \right| <$$

$$(3\sqrt{3}/2)\,\frac{(m^2-2m+3)^{3/2}(q^2v^2+1)^3}{(m-3)^3\sqrt{m-1}} \qquad \text{(14bis)}$$

(*i.e.*, *tends to zero when n increases not slower than a magnitude of order* $1/\sqrt{m}$).

As stated above, Chebyshev could have rightfully formulated this rigorously proved fundamental result in the now generally accepted notation for the Stieltjes integral.

10. Chebyshev made the last step in concluding the proof of the limit theorem on sums of independent random variables in 1887 [4]. The first of the two theorems mentioned in the title of [4] was the proposition of 1867 (the law of large numbers) that Chebyshev, as though summing up all his investigations in probability, reproduced here in the beginning of this paper before proceeding to the formulation and demonstration of the second theorem.

It only remained for him to prove that, as $n \to \infty$,

$$\lim \frac{E(u_1 + u_2 + ... + u_n)^{2k}}{B_n^{\;k}} = 1{\cdot}3\ldots(2k-1),\ \lim \frac{E(u_1 + u_2 + ... + u_n)^{2k+1}}{B_n^{\;k+1/2}} = 0 \qquad (21)$$

where $Eu_i = 0$, $Eu_i^2 = a_i^{(2)}$, $Eu_i^m = a_i^{(m)}$, $B_n = a_1^2 + a_2^2 + \ldots + a_n^2$. Indeed, he believed it obvious that if the sequence of the laws of probability $P_n(x)$ is such that

$$\lim \int_{-\infty}^{\infty} x^k dP_n(x) = c_k = \int_{-\infty}^{\infty} x^k dP(x),$$

where the moments c_k uniquely determined some law $P(x)$, then $P_n(x) \to P(x)$. Markov was the first to prove with complete rigor this property which is really easy to establish by means of the Chebyshev inequalities.

It should be admitted that Chebyshev's proof of the limiting equalities (21) did not entirely comply with the demands of rigor which he himself enunciated in his youth [1; 2]: here, he applied a method based on a Laplace transform differing from his usual methods of mature age. Had he not hurried with the publication of this last writing which appeared only a few months after the one considered just above [9], he would have possibly replaced the extremely interesting heuristic *transcendental* method by a more rigorous algebraic demonstration probably known to him for a long time.

Somewhat changing Chebyshev's notation, we shall describe the train of his thought s indicating the essence of its deficiency and explaining how, without any special difficulties, he could have removed it. He considered the function

$$\chi_n(s) = E \exp \frac{s(u_1 + u_2 + ... + u_n)}{\sqrt{B_n}} = \prod_{k=1}^{n} E \exp(su_k/\sqrt{B_n}\,) =$$

$$1 + (s^2/2) + \sum_{m=3}^{\infty} (s^m/m!)\, E\, [(u_1 + u_2 + \ldots + u_n)/\sqrt{B_n}\,]^m \qquad (22)$$

where s was an arbitrary constant. Does this expectation exist? Of course it does for purely imaginary values of s, but Chebyshev had not specified this, and, in addition, without providing any explanation, he expanded

$$E \exp(su_k/\sqrt{B_n}\,) = 1 + \frac{s^2 a_k^{(2)}}{2B_n} + \ldots + \frac{s^m a_k^{(m)}}{m! B_n^{\;m/2}} + \ldots \qquad (23)$$

into a power series. This, however, will not converge at any $|s| > 0$. For this expansion to become valid for $|s| < 1$ it would have been sufficient, say, to introduce the restriction

$$[|a_k^{(m)}|/B_n^{\;m/2-1}] \le m! a_k^{(2)} \lambda_n \ (m > 2) \qquad (24)$$

where $\lambda_n \to 0$ as $n \to \infty$. Then, if $|s| \le 1/2$, it will follow from (23) that

$$|E \exp(su_k/\sqrt{B_n}\,) - 1 - \frac{s^2 a_k^{(2)}}{2B_n}| < \frac{a_k^{(2)}\lambda_n}{B_n} \sum_{m=3}^{\infty} |s^m| < 2\lambda_n[a_k^{(2)}/B_n]|s|^3. \qquad (25)$$

Taking the logarithms of (23), Chebyshev obtained a formal expansion

$$\ln \mathrm{E} \exp(su_k/\sqrt{B_n}) = \ln\left(1 + \frac{s^2 a_k^{(2)}}{2B_n} + \ldots\right) = \frac{s^2 a_k^{(2)}}{2B_n} + \frac{s^3 a_k^{(3)}}{3!B_n^{3/2}} + \ldots$$

Without condition (24) this operation is invalid; but, having introduced this restriction, we arrive at

$$\ln\left(1 + \frac{s^2 a_k^{(2)}}{2B_n} + \ldots\right) = \ln\left[1 + \frac{s^2 a_k^{(2)}}{2B_n}(1 + \theta_k\lambda_n)\right],\ |s| \le 1/2,\ |\theta_k| < 2,$$

$$\ln \chi_n(s) = \sum_{k=1}^{n} \ln\left[1 + \frac{s^2 a_k^{(2)}}{2B_n}(1 + \theta_k\lambda_n)\right] = (s^2/2)(1 + \theta_k\lambda_n) + \ldots +$$

$$(-1)^m \frac{s^{2m}}{m(2B_n)^m} \sum_{k=1}^{n} [a_k^{(2)}(1 + \theta_k\lambda_n)]^m + \ldots,\ |\theta_k| < 2. \tag{26}$$

Thus, excepting the first one, all the coefficients of this power series that converges at $|s| \le 1/2$ will tend to zero if

$$\{\sum_{k=1}^{n} [a_k^{(2)}]^m/B_n^m\} = \{\sum_{k=1}^{n} [a_k^{(2)}]^m/[\sum_{k=1}^{n} (a_k^{(2)})]^m\} \to 0. \tag{27}$$

However, in virtue of (24),

$$\sum_{k=1}^{n} a_k^{(2m)} \le B_n^{m-1} \sum_{k=1}^{n} a_k^{(2)}\lambda_n(2m!) = (2m!)B_n^m\lambda_n,$$

or, for every given m,

$$\{[\sum_{k=1}^{n} a_k^{(2m)}]/B_n^m\} < (2m!)\lambda_n \to 0 \text{ as } n \to \infty,$$

and, since $[a_k^{(2)}]^m \le a_k^{(2m)}$, (27) follows. Therefore, at $|s| \le 1/2$

$$\lim \ln \chi_n(s) = (s^2/2) \text{ as } n \to \infty$$

and, in accord with a property of Taylor series, all the coefficients of the series for $\chi_n(s)$, convergent at $|s| \le 1/2$, tend to the coefficients of the series

$$\exp(s^2/2) = \sum_{i=0}^{\infty} [s^{2i}/(2i)!]1\cdot 3 \ldots (2i-1)$$

respectively [7].

We see now that condition (24) is sufficient for making Chebyshev's reasoning quite rigorous without adding anything essential to it. Therefore, it should be considered in justice that he had in essence completely proved the limit theorem on sums of independent variables under the {additional} condition (24). This restriction covers all the practically important cases. Indeed, it is fulfilled if [8]

$$|u_k|/\sqrt{B_n} < \lambda_n \ (\lambda_n \to 0 \text{ as } n \to \infty) \qquad (28)$$

since then, if $m > 2$,

$$|a_k^{(m)}| \le a_k^{(2)}(\lambda_n \sqrt{B_n})^{m-2}$$

and

$$\{|a_k^{(m)}|/B_n^{m/2-1}\} \le \lambda_n^{m-2}\, a_k^{(2)} < m!\, a_k^{(2)}\, \lambda_n. \qquad (24\text{bis})$$

Regrettably, Chebyshev had not introduced any restriction similar to (24) and the limit theorem in his excessively general form was wrong.

In 1898 Markov [17b] proved by direct algebraic calculation that the limiting equalities (21) were valid if only

$$\sum_{k=1}^{n} \{|a_k^{(m)}|/B_n^{m/2}\} \to 0 \qquad (29)$$

for all integer values of $m > 2$. This condition is somewhat weaker than restriction (24), and Markov had thus filled in the technical gap in Chebyshev's brilliant proof based on the method of moments.

11. Markov was undoubtedly the most outstanding exponent of Chebyshev's direction of, and ideas in probability theory. He was the closest to his teacher in the nature and the sharpness of his mathematical gift. Whereas Chebyshev, especially towards the end of his life, as well as in his lectures, sometimes deviated from his own demands of clear formulations and rigor of proof in probability, Markov's classical course in the calculus of probability and his original memoirs were specimens of preciseness and lucidity of exposition [9], which contributed to the maximal extent to the transformation of probability theory into one of the most perfect branches of mathematics and to the wide dissemination of Chebyshev's direction and methods. As we shall see below, the later deep analysis of dependences between observed random phenomena made by Markov in the Chebyshev spirit enabled him to extend essentially the domain of probability by introducing into consideration dependent random variables.

My aim does not include the reviewing of all the modern achievements in probability. These would have been impossible without the robust mathematical base of the theory constructed by Chebyshev and consolidated by Markov. I would only like to indicate the stages of the subsequent history of the problems connected with the LLN and the limit theorem, which, all of them, directly adjoin Chebyshev's investigations and develop his ideas.

Let us first dwell on the LLN. Markov's simple remark [17d] directly following from Chebyshev's inequality [3] that the restriction $(B_n/n^2) \to 0$ where $B_n = \mathrm{E}(u_1 + u_2 + \ldots + u_n)^2$ is sufficient for the applicability of the law to variables $u_1, u_2, \ldots, u_n$ ($\mathrm{E}u_i = 0$), led him, and, later, other authors, to a number of interesting sufficient conditions for the same fact both for independent and dependent variables u_i. And, by introducing supplementary variables u_i' such that $u_i' = u_i$ when $|u_i| \le L_n$, $i \le n$, and $u_i' = 0$ if $|u_i| > L_n$, Markov became able to prove the applicability of the same law in a number of other cases in which $\mathrm{E}u_i^2$ (and therefore B_n) do not exist. For example, he generalized the Chebyshev theorem [3] in the following way: If for some $p > 1$ $\mathrm{E}|u_i|^p < L$ where L is some constant, then the law is applicable to these independent variables u_i.

Among other developments obtained in a similar way, we indicate a remarkable in its simplicity and maximal (in some sense) generality theorem due to Khinchin [21b]: If all the independent variables u_i obey one and the same law of probability and Eu_i exist, then the LLN is applicable to u_i. The derivation of more or less general necessary conditions for the applicability of this law to independent variables is also based on the Chebyshev inequality and on a simple remark [11d] that the law cannot be applied to variables u_i in the particular case in which their laws of probability are symmetric and such that there exists a positive number c for which the probability that $|u_i| \geq cn$ at least for one i ($i \leq n$) does not tend to zero as n increases.

The Kolmogorov theorem [13] based on an appropriate development of Chebyshev's classical reasoning essentially supplements Chebyshev's inequality: The probability that all the inequalities

$$|u_1 + u_2 + \ldots + u_i| \leq t\sqrt{B_n},\ i = 1, 2, \ldots, n$$

for independent variables u_i are valid is higher than $1 - 1/t^2$. Bernstein [11g] generalized these inequalities onto dependent variables u_i having E$u_i = 0$ for any values of $u_1, u_2, \ldots, u_{i-1}$.

The determination of a more precise higher boundary than $(1 - 1/t^2)$ for the probability in the Chebyshev inequality

$$|u_1 + u_2 + \ldots + u_i| \leq t\sqrt{B_n},\ B_n = c_1^{(2)} + c_2^{(2)} + \ldots + c_n^{(2)}$$

irrespective of n necessarily demands that some more or less essential restrictions be introduced. We [11b] note, for example, the estimate

$$Q < \exp(-t^2)$$

of the probability Q that

$$u_1 + u_2 + \ldots + u_n \geq 2t\sqrt{B_n}.$$

It is valid for all values of n and $t > 0$ if $t \leq [\sqrt{B_n}/(2H)]$ under the condition that

$$|\mathrm{E}u_k^m| < (1/2)c_k^{(2)} H^{m-2} m!$$

where H is some constant. The proof of this proposition is also some development of Chebyshev's classical reasoning. In addition, the same estimate is obtained for dependent variables provided that E$u_i = 0$ and E$u_i^2 = c_i^{(2)}$ for any previous value of u_k ($k < i$); furthermore, this estimate persists [11g] when the inequality (30) is replaced by the appropriate inequalities as in the Kolmogorov theorem.

12. The work of Chebyshev was of no lesser fundamental importance for the further study of the conditions for the application of the limit theorem. It is necessary to dwell here first of all on the case in which the variables u_i are independent. We have already discussed Markov's specifications of Chebyshev's proof based on the method of moments that the former introduced and thus made the latter's reasoning irreproachably rigorous by stipulating that the variables u_i not only possess moments, *i.e.*, expectations of any order, but in addition satisfy the demand (29). Markov [17c] obtained this result in 1898.

Soon after that, in 1901, Liapunov [16a; 16b] remarkably weakened the restriction (29). He also was one of Chebyshev's closest students, and he experienced the powerful influence of his teacher. It is known, for example, that it was Chebyshev who suggested to him the problem of the figures of equilibrium of a rotating liquid that came to occupy a central place in his studies; incidentally, this fact testifies that Chebyshev's scientific interests went beyond those branches of mathematics where his personal original creative work had manifested itself. Nevertheless, Chebyshev's influence upon Liapunov, who, with respect to the power of his gift, was second to none either in Russia or in the West, was not really exceptional. Liapunov, better than the other representatives of the Petersburg school, understood, and was able to appreciate the achievements of the West-European mathematicians of the second half of the last {of the 19th} century **10** who had outlined precise boundaries for the methods of classical *transcendental* analysis, and made them no less reliable than Chebyshev's algebraic methods. This very fact explains why Liapunov approached Chebyshev's problems more independently than his other students. Whereas Markov perfected the methods of his great teacher and applied them to new problems, Liapunov, when thinking about the essence of the limit theorem, understood that the method of moments did not facilitate the problem and only shifted its main difficulty elsewhere. Indeed, the Laplace transform, applied by Chebyshev in his last memoir of 1887, consisted in considering the expectation

$$\mathrm{E}e^{sx} = \chi(s) = \int_{-\infty}^{\infty} e^{sx} f(x)\, dx$$

for an arbitrary value of the parameter s. Instead of the totality of moments $c_m = \mathrm{E}x^m$, the function

$$\chi(s) = \sum_{m=0}^{\infty} c_m s^m/m!$$

can itself characterize the density $f(x)$ or the integral law of probability

$$F(x) = \int_{-\infty}^{x} f(x)\, dx$$

if only the series (22bis) converges, which, as we saw, is also an essential condition for the correctness of the relevant portion of Chebyshev's reasoning.

However, it is not at all necessary to use the expansion (22bis) into a power series; there is even no need for supposing that the moments c_k exist provided we shall only assign purely imaginary values to s, $s = it$, since the integral

$$\theta(t) = \chi(it) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx = \int_{-\infty}^{\infty} e^{itx} dF(x) \qquad (31)$$

exists for all real values of t.

The function $\theta(t)$ which is now called the *characteristic function* of the law of distribution of variable x, uniquely determines the law $F(x)$ irrespective of the existence of c_k. In addition, when considering the case of

$$X_n = \sqrt{1/B_n}\,(u_1 + u_2 + \ldots + u_n) \qquad \text{(ii)}$$

Liapunov proved that, for the characteristic function of the law of probability of X_n to converge to

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

i.e., to the characteristic function of the normed normal law

$$(1/\sqrt{2\pi}\,)\int\limits_{-\infty}^{x} \exp(-x^2/2)\,dx,$$

it is sufficient that, at least for one $\delta > 0$,

$$\{\sum_{k=1}^{n} [a_k^{(2+\delta)}]/B_n^{1+\delta/2}\} = (1/M_{n,\delta}) \to 0,\ a_k^{(2+\delta)} = \mathrm{E}|\,u_k|^{2+\delta}. \qquad (32)$$

This condition is thus sufficient for the applicability of the limit theorem to suns of independent variables u_i. Following Chebyshev's precept, Liapunov also indicated the higher boundary of the error of the limit theorem. For example, if $\delta = 1$, the order of this error is $(\ln M_{n,1})/M_{n,1}$.

So, Liapunov established with perfect rigor and by an absolutely different method the limit theorem for sums of independent variables under a much more general condition than Markov. The latter, however, became able to exonerate the method of moments. Supplementing it by a short reasoning similar to that, which he had applied for generalizing the conditions for the applicability of the LLN, he fully demonstrated the limit theorem under the general Liapunov condition.

Later, in 1926, when applying in essence the same remark as Markov did, I [11c] provided a still more general condition for the application of the limit theorem [11] that also covered such cases in which $\mathrm{E}u_i^2$ did not exist. For example, if the variables u_i might take values $\pm\sqrt{m}$, where m is any positive integer, with probabilities $p_m = 3/(\pi^2 m^2)$, so that $\mathrm{E}u_i = 0$ and

$$\mathrm{E}u_i^2 = (6/\pi^2)\sum_{m=1}^{\infty} (1/m) = \infty,$$

then the probability

$$P[(t_o/\pi)\sqrt{6n\ln n} < \sum_{i=1}^{n} u_i < (t_1/\pi)\sqrt{6n\ln n}\,] \to (1/\sqrt{2\pi}\,)\int\limits_{t_o}^{t_1} \exp(-t^2/2)\,dt.$$

In connection with this illustration it should be noted that the meaning of the limit theorem, depending on whether some or other conditions are satisfied, is not quite the same. Whereas the very method of proof applied by Chebyshev and Markov testified that, under the latter's conditions, all the moments of the sum (ii) tended to the normed normal moments (for example, $\lim \mathrm{E}X_n^4 = 3$ as $n \to \infty$), the Liapunov demonstration does not show it at all; and, since under condition (32) $\mathrm{E}|u_i|^{2+\mu}$ possibly does not exist at $\mu > \delta$, $\mathrm{E}|X_n|^{2+\mu}$ will also be meaningless. It can be proved [11h] that the Liapunov condition with a given $\delta > 0$ is necessary and sufficient for $\mathrm{E}|X_n|^p$ with $p \le 2 + \delta$ to have the corresponding normal moment as its limit. In the example just above the Liapunov condition is not fulfilled for any $\delta > 0$ and we have here, as $n \to \infty$,

$$\lim \mathrm{E}\{\frac{\pi\,|\,u_1 + u_2 + \ldots + u_n\,|}{\sqrt{6n\ln n}}\}^p = (1/\sqrt{2\pi}\,)\int\limits_{-\infty}^{\infty} |x|^p \exp(-x^2/2)\,dx$$

only for $p < 2$. The Norwegian mathematician Lindeberg [15] derived a condition sufficient for $\mathrm{E}X_n^2 \to 1$; here, $p = 2$.

In general, during the last twenty years a large number of studies have thrown light, from various points of view, on some important properties of the normal law and on the conditions for its applicability to sums of independent variables. Nevertheless, their findings are far less fundamental than the result attained by Chebyshev and his students, Markov and Liapunov.

It should be noted, however, that during the same period a number of absolutely new problems have been formulated and solved. The most interesting in its essence was the discovery of the general form of the limiting law for the sums of independent variables in those exceptional cases in which it is not normal. The pertinent subtle investigations, the most important of which were due to Kolmogorov, Khinchin and Lévy, were based on the appropriate use of the Liapunov method of characteristic functions. They demanded a deep study of the properties of these functions and revealed an unexpected fact: The so-called Poisson law is in some sense even more universal than the Laplacean normal distribution [21c]. It is curious that the Petersburg mathematicians had completely overlooked the Poisson law which statisticians call the law of small numbers [12] and which also plays a considerable part in physical applications of probability.

13. Turning now to the generalization of the main laws of the theory of probability onto dependent variables, we recall that Chebyshev himself did not study this subject; however, since Markov was the founder of this central field of modern probability, and consistently developed Chebyshev's ideas in general and the method of moments in particular, we must briefly discuss this point. Following Chebyshev in that he believed that each new general domain of research should be approached by considering appropriate typical problems having a clearly seen and real meaning and admitting of a precise mathematical formulation, Markov displayed here his deep flair of an outstanding natural scientist and directed his attention to a class of dependent random variables exceptionally important for applications. He called his objects *chains*, and their present-day scientific designation is *Markov chains* [13].

The British biologist and statistician Galton was the first to consider such dependences under some particular assumptions. Drawing on rather extensive experimental materials, he attempted to explicate the Darwinian theory of heredity in a mathematical form. However, we do not find any satisfactory mathematical investigation of the properties of stochastic patterns similar to the Markov chains either in his works, or in the writings of physicists who came to study them almost at the same time as Markov did. The closeness of Markov's ideas to the notions that originated simultaneously in various fields of natural sciences testifies to their vitality and methodological importance. Indeed, a Markov chain is a stochastic transformation of a usual deterministic process; it is characterized by the fact that its dynamic state at a given moment fully determines its further course irrespective of all the previous states. More precisely, Markov defined a chain as a sequence of such random variables (or events, or states) $X_1, X_2, \ldots, X_n, \ldots$ that, after X_n takes some definite value, the probability that the next variable X_{n+1} takes one or another value becomes fully determined and does not depend on the values of the previous variables X_i, $i < n$. Excluding the case in which the chain unboundedly approaches a determinate process, Markov showed [14] that the dependence between two links of the chain, X_i and X_{i+h}, sufficiently far from one another, weakens so rapidly that , as n increases, the law of probability of X_n tended to a law independent from the initial state X_1. Under some very general assumptions, this important theorem might justify the hypothesis of a uniform distribution of probabilities at set conditions that forms the basis of various physical theories. By generalizing, as indicated above, the Chebyshev inequality and under the same restrictions Markov also derived an extension of the LLN as a corollary to the fact that the dependence between remote links of a chain was very weak.

The problem of the limiting applicability of the normal law to the sum of terms constituting a chain presented more considerable difficulties. The method of moments demanded the calculation of the expectations

$$\mathrm{E}\left(\sum_{i=1}^{n}(X_i/\sqrt{B_n}\right)^m$$

for all positive integer values of m and a proof that they tended to the corresponding normal moments. In the simplest cases the method of generating functions led to this goal, but the algebraic method similar to that which Markov had applied in 1898 to independent variables, proved here very unwieldy and Markov derived the demanded result under rather general conditions only owing to his mastery.

Later on I had generalized his findings basing my work on the Liapunov method of characteristic functions. This, however, was not here directly applicable: the advantage of these functions consisting in that, for independent terms, a characteristic function of a sum of two variables is equal to the product of such functions for each of these terms, falls away for the case of dependent variables. It was therefore necessary to supplement this method by the idea of separating chains into sections in such a way, that, after eliminating a number of intermediate links, the limiting law of probability for the entire sum would not differ from that of the remaining sum consisting of sufficiently weakly connected one with another, and in some precisely defined sense almost independent groups of terms. This idea is apparently applicable not only to Markov chains, but also to more general classes of dependent variables. The most difficult and interesting is the study of the case in which the chain is singular, *i.e.*, very close to a strictly deterministic process. The discussion of the findings derived in this direction is beyond the boundaries of this paper. I only note that this approach and many useful supplementary pertinent lemmas directly adjoin the problems examined by Chebyshev. Thus [11f], the following theorem is in his spirit: For any Markov chain the variance $B(S_n)$ of the sum of variables $X_1, X_2, \ldots, X_n$, $S_n = X_1 + X_2 + \ldots + X_n$, satisfies the inequality

$$B(S_n) \geq \beta_2 + \beta_4 + \ldots + \beta_{2[n/2]}$$

where β_h is the mean conditional variance of X_h after the values of X_{h-1} and X_{h+1} become known. The boundary is actually attained when $X_1, X_3, \ldots$ are determinate and $X_2, X_4, \ldots$ are independent. The problem about the limiting law of S_n for the case in which it differs from the Laplace – Gauss distribution is of a considerable theoretical and practical interest. The appropriate contributions are mostly closely connected with Markov chains, but the field in question also makes use of a number of new ideas and is too extensive for being discussed here.

I am concluding this brief review of the development of Chebyshev's ideas in which, his celebrated students, Markov and Liapunov, participated above all. We see that the theory of probability owes to the work of Chebyshev and his school its present-day maturity that ensures its reliable application to most diverse real phenomena. The genius of Chebyshev and his associates, who, in this field, left West-European mathematicians far behind, overcame the crisis of probability that arrested its development a hundred years ago.

Notes

1. {Bernstein mentioned De Moivre (see above) but had not discussed his finding; and Laplace's work on the CLN did not date back to the 18th century.}

2. {Bernstein overlooked two other fields of application of the early probability theory, viz., population statistics and the treatment of observations.}

3. {See Note 7 to Gnedenko's essay on Ostrogradsky in this book.}

4. Its subtitle is *Extrait d'un mémoire russe sur l'analyse élémentaire de la théorie des probabilités*, and, although the main part of [2] is lacking in the Dissertation [1], this is obviously a reference to [1]. It might be concluded that [2] had existed in some form when Chebyshev defended his dissertation in 1846 in Moscow. S.B. {A later commentary on [1] is Prokhorov (1986, §7).}

5. It should have been mentioned that the trials were independent. Actually, here and in all his works in general, Chebyshev only considered independent events. Poisson's essential mistake, which Chebyshev indeed corrected, consisted in that he had indiscriminately applied his theorem to any events.

6. Inequality (18) can be easily obtained directly by noting that

$$\theta(t) - \sum_{i=0}^{m-1} T_i t^i = \sum_{i=m}^{\infty} T_i t^i \le [1/(m-1)] \sum_{i=m}^{\infty} iT_i t^i \le \frac{[t\theta'(t)]}{m-1}$$

whence

$$\sum_{i=0}^{m-1} T_i t^i \ge \theta(t) - \frac{[t\theta'(t)]}{m-1} = \theta(t) - \frac{t[\theta'(t)]^2}{t\theta''(t)+\theta'(t)}.$$

7. {The equality just above is obviously wrong.}

8. It was subsequently shown that the most general form of the limit theorem can be easily derived when issuing from condition (28) [11c].

9. {Students hardly liked Markov's treatise [17d]; in 1910, he himself (Ondar 1977, p. 21) owned that he "often heard" that his presentation of the MLSq was "not sufficiently clear". Later editions were not better. And the opening lines of Markov's memoir [17e] present a specimen of unreadable phrases.}

10. {Nevertheless, Liapunov (1895, pp. 19 – 20) called Riemann's ideas "extremely abstract" and his investigations "pseudo-geometric", having nothing in common with Lobachevsky's "deep geometric studies". Then, Markov hid his *deep flair of an outstanding natural scientist*: he did not even hint at the possibility of applying his chains in natural sciences. A possible explanation of his attitude is found in his letter of 1910 to Chuprov (Ondar 1977, p. 52): "I shall not go a step out of that region where my competence is beyond any doubt". I also note that it was Bernstein himself [11c, §16] who introduced the term *Markov chain*. Finally, Bernstein had not elaborated his remark (below) about the prehistory of chains. I myself (*Arch. Hist. Ex. Sci.*, vol. 39, 1989, pp. 364 – 365) listed several pertinent developments (but did not mention Galton).}

11. Feller [20] proved that this condition was also necessary under the natural assumption that each term was negligible as compared with the entire sum.

12. {Bernstein overlooked Markov's studies of the stability of statistical series. They were partly connected with the law of small numbers; hence, with the Poisson law.}

13. {It was Bernstein himself [11c, §16} who introduced the term *Markov chains*. Markov hid his *deep flair of an outstanding natural scientist*: he did not even hint at the possibility of applying his chains in natural sciences. A partial explanation of his attitude is found in his letter of 1910 to Chuprov (Ondar 1977, p. 52): "I shall not go a step out of that region where my competence is beyond any doubt". Finally, concerning the prehistory of Markov chains (see below): Bernstein regrettably did not elaborate his remark about Galton. I myself (*Arch. Hist. Ex. Sci.*, vol. 39, 1989, pp. 364 – 365) listed several facts pertaining to that prehistory (but did not mention Galton).}

14. Almost at the same time Poincaré [18] independently arrived at the same result in some particular cases.

References

P.L. Chebyshev

1. (1845), *Опыт элементарного анализа теории вероятностей* (Essay on an Elementary Analysis of the Theory of Probability). Reprinted: *Полное собрание сочинений* (Complete Works), vol. 5. M. – L., 1951, pp. 26 – 87.

2. (1846), Démonstration élémentaire d'une proposition générale de la théorie des probabilités. *J. reine u. angew. Math.*, Bd. 33, pp. 259 – 267.

3. (1867), Des valeurs moyennes. *J. math. pures et appl.*, t. 12, pp. 177 – 184.

4. (1887, in Russian), Sur deux théorèmes relatifs aux probabilités. *Acta math.*, t. 14, 1890 – 1891, pp. 305 – 315.

5. (1936), *Теория вероятностей* (Theory of Probability). Lectures read 1879/1880 as taken down by A.M. Liapunov. M.– L. Translated into English: Berlin, 2004.

6. (1855, in Russian), Sur les fractions continues. *J. math. pures et appl.*, t. 2, 1858, pp. 289 – 323.

7. (1859), Sur le développement des fonctions à une seule variable. *Bull. Cl. phys.-math.* AN Psb, t. 1, pp. 193 – 200.

8. (1874), Sur les valeurs limites des intégrales. *J. math. pures et appl.*, t. 19, pp. 157 – 160.

9. (1887, in Russian), Sur les résidus intégraux qui donnent des valeurs approchées des intégrales. *Acta math.*, t. 12, 1888 – 1889, pp. 287 – 322.

(1899 – 1907), *Œuvres*, tt. 1 – 2. Psb., in Russian and French, contain contributions [2; 3; 6; 7; 9] in t. 1 and [4; 8] in t. 2. French edition reprinted (New York, 1962).

Other Authors

10. Bernoulli, J. (1713, in Latin), *Wahrscheinlichkeitsrechnung*. Frankfurt/Main, 1999.

11a. Bernstein, S.N. (1922), Sur le théorème limite du calcul des probabilités. *Math. Ann.*, Bd. 85, pp. 237 – 241.

11b. --- (1924) On a modification of the Chebyshev inequality and on the error of the Laplace formula. *Собрание сочинений*, vol. 4 (Coll. Works). N.p., 1964, pp. 71 – 79. (R)

11c. --- (1926), Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Math. Ann.*, Bd. 97, pp. 1 – 59.

11d. --- (1946), *Теория вероятностей* (Theory of Probability), 4th edition. M. – L. First edition, 1927.

11e. --- (1928), Addition à l'article Sur les sommes de quantités dépendantes. *Doklady* AN SU, pp. 55 – 60.

11f. --- (1936), Détermination d'une limite inférieure de la dispersion des sommes des grandeurs liées en chaine singuliére. MS, vol. 1, pp. 29 – 37.

11g. --- (1937), On some modifications of the Chebyshev inequality. *Doklady* AN SU, vol. 17, pp. 275 – 277. (R)

11h. --- (1939), Several remarks on the Liapunov limit theorem. Ibidem, vol. 24, pp. 3 – 7. (R)

12. Bienaymé, I.J. (1853), Considérations à l'appui de la découverte de Laplace, etc. *C.r. Acad. Sci. Paris*, t. 37, pp. 309 – 324. Reprinted 1867.

13. Kolmogorov, A.N. (1928), Über die Summen durch den Zufall bestimmten unabhängigen Grossen. *Math. Ann.*, Bd. 99, pp. 309 – 319.

14. Laplace, P.S. (1812), *Théorie analytique des probabilités. Oeuvr. Compl.*, t. 7. Paris, 1886.

15. Lindeberg, J.W. (1922), Eine neue Herleitung des Exponentialgesetzes, etc. *Math. Z.*, Bd. 15, pp. 211 – 225.
16a. Liapunov, A.M. (1900), Sur une proposition de la théorie des probabilités. *Izvestia* (*Bull.*) AN Psb, sér. 5, t. 13, No. 4, pp. 359 – 386.
16b. --- (1901), Nouvelle forme du théorème sur la limite des probabilités. *Zapiski* (*Mém.*) AN Psb, sér. 8, cl. phys.-math., t. 12, No. 5, pp. 24.
17a. Markov, A.A. (1884), The proof of some of Chebyshev's inequalities. *Избранные труды* (Sel. Works), M. – L., 1948, pp. 15 – 24. (R)
17b. --- (1899), The law of large numbers and the method of least squares. In Markov (1951, pp. 231 – 251). Translated in this book.
17c. --- (1898), Sur les racines de l'équation … *Izvestia* (*Bull.*) AN Psb, sér. 5, t. 9, pp. 435 – 446.
17d. --- (1900), *Исчисление вероятностей* (Calculus of Probability). Psb. Later editions: 1908, 1913, 1924. German translation: Leipzig – Berlin, 1912.
17e. --- (1906), Extension of the law of large numbers, etc. Translated in this book.
17f. --- (1908), Extension of the limit theorem, etc. Translated in this book.
17g. --- (1910), Investigation of the general case of trials, etc. Translated in this book.
17h. --- (1911), On a case of trials connected into a generalized chain. In Markov (1951, pp. 417 – 436).
17i. --- (1912), On trials connected into a chain by non-observed events. Ibidem, pp. 437 – 463.
17j. --- (1907, in Russian), Recherches sur un cas remarquable d'épreuves dépendantes. *Acta math.*, t. 33, 1910, pp. 87 – 104.
17k. --- (1915), Application of the method of expectations to connected series of magnitudes. *Izvestia* (*Bull.*) AN Psb, sér. 6, t. 9, pp. 1453 – 1484. (R)
Markov, A.A. (1951), *Избранные труды* (Sel. Works). N.p.
18. Poincaré, H. (1912), *Calcul des probabilités.* Paris.
19. Poisson, S.-D. (1837), *Recherches sur la probabilité des jugements*, etc. Paris.
20. Feller, W. (1935 – 1937), Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung. *Math. Z.*, Bd. 40, pp. 521 – 559; Bd. 42, pp. 301 – 312.
21a. Khinchin, A.Ya. (1928), Sur la loi forte des grands nombres. *C.r. Acad. Sci. Paris*, t. 186, pp. 285 – 287.
21b. --- (1929), Sur la loi des grands nombres. Ibidem, t. 188, pp. 477 – 479.
21c. --- (1938), *Предельные законы для сумм независимых случайных величин* (Limit Laws for Sums of Independent Random Variables). M.

References to Notes

Liapunov, A.M. (1895), Chebyshev. In Chebyshev, P.L., *Избранные математические труды* (Sel. Math. Works). M. – L., 1946, pp. 9 – 21. (R)
Ondar, Kh.O., Editor (1977, in Russian), *Correspondence between Markov and Chuprov*, etc. New York, 1981.
Prokhorov, Yu.V. (1986), The law of large numbers and the estimation of the probabilities of large deviations. In Bernoulli, J. *О законе больших чисел* (On the Law of Large Numbers). Editor, Yu.V. Prokhorov. M., pp. 116 – 150. (R)

6. N.A. Usov. A Remark concerning the Chebyshev Theorem

. *Matematich. Sbornik*, vol. 2, 1867, pp. 93 – 95 of second paging

Translator's Foreword

The author was Ostrogradsky's student (Dobrovolsky & Kiro 1967, p. 125). He published his note immediately after Chebyshev's pertinent contribution had appeared (and he referred to its Russian version which was issued at the same time as the French version mentioned in the References). I have insignificantly changed Usov's text by replacing, where possible, integrals, in accord with his notation, by a or a_1.

Usov's attempt was forgotten; the justification of the arithmetic mean and of the MLSq in general by a reference to Chebyshev (1867) is usually ascribed to Yarochenko (1893a; 1893b) although Maievsky had preceded the latter. Markov, however, reasonably rejected such efforts, see my paper in this book where I cite Markov's negative and reasonable opinion about such efforts.

That the treatment of observations had been highly topical is once again proved by Chebyshev's remark (1887) accompanying his proof of the CLT. That proposition, as he stated, "led" to the MLSq,– but only, as I add, in the Laplacean sense of treating a large number of observations. And the Gauss mature substantiation of the method did not depend on any law of error.

* * *

A rather remarkable corollary easily follows from Chebyshev's theorem (1867). The principle of the arithmetic mean is known to be admitted as a *postulatum* for the derivation of the type of the function $\varphi(\Delta)$ of the probability of error Δ. The theorem of Mr. Chebyshev provides a rigorous demonstration of that principle.

Suppose that N observations were made to determine some magnitude u and resulted in

$$A_1, A_2, \ldots, A_n.$$

Denoting in Mr. Chebyshev's theorem

$$x = u - A_1,\ y = u - A_2,\ z = u - A_3, \ldots$$

we see that all the possible values of $x, y, z,$ will be situated between some boundaries, call them $-L$ and M. According to the definition assumed by Mr. Chebyshev, the expectations of $x, y, z, \ldots$ will be, respectively,

$$a = b = c = \int_{-L}^{M} \Delta\varphi(\Delta)\, d\Delta$$

and those of $x^2, y^2, z^2, \ldots,$

$$a_1 = b_1 = c_1 = \int_{-L}^{M} \Delta^2\varphi(\Delta)\, d\Delta$$

where $\varphi(\Delta)\, d\Delta$ is the probability that the error of observation is situated between Δ and $(\Delta + d\Delta)$.

Then, in accord with the theorem of Mr. Chebyshev, the boundaries for the magnitude

$$u - \Sigma A_i/N \qquad (1)$$

are

$$\int_{-L}^{M} \Delta\varphi(\Delta)\, d\Delta \pm (1/t)\, \sqrt{a_1 - a^2}$$

and the probability P that this magnitude will not exceed those limits are, for any t,

$$P > 1 - t^2/N \qquad (2)$$

and, consequently, for any t, it will tend to 1 as N increases.

If positive and negative errors are equally possible, then $|L| = M$, $\varphi(-\Delta) = \varphi(\Delta)$ and obviously

$$\int_{-L}^{L} \Delta\varphi(\Delta)\, d\Delta = 0$$

and we may conclude: The probability(2) that the magnitude (1) differs from zero not more than by

$$(1/t)\sqrt{\int_{-L}^{L} \Delta^2 \varphi(\Delta) d\Delta}$$

tends to unity as N increases.

It is easy to see that a_1 is the mean of the squares of the errors. Assuming that

$$\varphi(\Delta) = (h/\sqrt{\pi}) \exp(-h^2\Delta^2)$$

we can obtain

$$a_1 = 1/(2h^2).$$

I think that this remark will be useful for teaching.

References

Chebyshev, P.L. (1867), Des valeurs moyennes. *J. math. pures et appl.*, t. 12, pp. 177 – 184.

--- (1887, in Russian), Sur deux théorèmes relatifs aux probabilités. *Acta math.*, t. 14, 1890 – 1891, pp. 305 – 315.

Dobrovolsky, V.A., Kiro, S.N. (1967), Teaching of mathematics in special technological and military schools. In *История отечественной математики* (History of National Mathematics), vol. 2. Editor, I.Z. Stokalo. Kiev, 1967, pp. 119 – 126. (R)

Yarochenko, S.P. (1893a), On the theory of the method of least squares. *Zapiski Novoross. Univ.*, vol. 58, pp. 193 – 208 of second paging. (R)

--- (1893b), Sur la méthode des moindres carrés. *Bull. sciences math.*, sér. 2, t. 17, pp. 113 – 125.

7. S.N. Bernstein. The Petersburg School of the Theory of Probability

Uchenye Zapiski Leningradsk. Gosudarstven. Univ., No. 55 (ser. math. sci., No. 10),

Report read at the plenary meeting of the scientific conference devoted to the 120-th anniversary of Leningrad University

Translator's Foreword

This contribution, also valuable, obviously served Bernstein as a point of departure for his later essay on Chebyshev (included in this book). Most of my critical historical notes inserted there are relevant here also. Even here the author passed over in silence the alienation of the Petersburg school from the then powerful and fruitful European direction towards abstract structures, cf. my Note 10 to his essay.

* * *

[1] Some of the most illustrious pages of the past that we may proudly recall at the University celebrations are the scientific achievements of the Petersburg school of the theory of probability. The aim of my address is to throw light on the leading part, and the importance of the works of the three main representatives of this school, Academicians Pafnuty Lvovich Chebyshev, Aleksandr Mikhailovich Liapunov and Andrei Andreevich Markov, who laid a firm foundation of the modern theory of probability and its applications to scientific natural sciences.

Indeed, at the time of its origin probability theory had little in common with the general progress of these sciences, and games of chance provided the only experimental basis for the development and specification of its most important notions and principles. It was on these grounds that Jakob Bernoulli, more than two hundred years ago, had discovered his celebrated theorem that offers a clue to the understanding of the origin of regularities inherent in mass independent individual chances, and represents the first rigorously proved, even if a very particular formulation of the LLN.

Laplace's classical research, and, in the first place, the widely known De Moivre – Laplace limit theorem, constituted the next most important stage in the development of the theory of probability. This proposition establishes the limiting law of distribution for the deviations of the number of occurrences of some random event from its expectation when the trials correspond to one and the the same elementary Bernoullian pattern and are repeated many times over. The LLN and the normal law of distribution, or the so-called Gaussian law of random errors, that generalizes the Laplace theorem {?}, are fundamentally important for our Weltanschauung. They were thus discovered in their most simple form already in the 18th century.

Another of Laplace's prominent merits consisted in that he foresaw the universality of these laws and was the first to lead the theory of probability out from the narrow range of games of chance onto the wide arena of scientific natural sciences. Because of his influence, the first half of the last {the19th} century was marked by a heightened interest to, and enthusiasm for probability theory, but many of its generalizations and applications were insufficiently substantiated, and some of them, even those supported by Laplace himself and Poisson, were so manifestly wrong, that later on John Stuart Mill quite deservedly qualified them as a *real opprobrium of mathematics*.

Because of these failures experienced by the theory of probability in its new field, enthusiasm gave way to disappointment and scepticism. Rare exceptions excluding, West European mathematicians of the second half of the last {of the 19th} century believed that the theory was nothing but a peculiar mathematical amusement that did not permit important scientifically substantiated applications and barely deserved any serious study.

[2] However, at that time the Petersburg school of the theory of probability appeared on the stage of the world science. Instead of the scepticism, understandable owing to the previous mistakes, it was able to elicit some indications for a radical revision of the theory's classical ideas and methods and for their adaptation to a profound analysis of random phenomena in all their actual variety. Specifying the formulation of the problems of probability theory, Petersburg mathematicians brilliantly surmounted the obstacles that had checked its development. In particular, they solved, in a measure sufficient for all its practical applications, both of the main contemporaneous problems, viz., the discovery of the *conditions for applying the LLN, and of the normal law.*

It might be assumed that the Petersburg school had originated in 1837: from then onward, the theory of probability became a discipline continually taught at Petersburg University. Until 1850, its first lecturer had been Professor Vikenty Aleksandrovich Ankudovich among whose listeners was the future head of the Petersburg school, Chebyshev. From 1850 to 1860, the course was read by Academician Victor Yakovlevich Buniakovsky, who, like his predecessor, had been developing the ideas of Laplace and Poisson, and, attaching great importance to the applications of the theory, largely promoted interest in probability among the Russian public. The flourishing of the theory of probability at Petersburg University began in 1860, when the reading of the course passed to Chebyshev. In 1882, Markov, one of the two of his celebrated students, replaced him, whereas the second one, Liapunov, never delivered any courses in probability and devoted to this discipline not more than two or three of his contributions [1] having, however, utmost importance. Markov continued to teach until the end of his days, and his treatise, matchless both in precision and clearness of exposition [2], was no less influential than his original works in transforming the theory of probability into a rigorous mathematical discipline.

In 1922, with the death of Markov, the Petersburg school had lost its last eminent representative and actually ceased to exist, but its ideas spread the world over and left an ineffaceable trace in science. Petersburg mathemati-cians generalized the principles of the theory of probability so as to bring the widest possible classes of random events and variables within reach of mathematical analysis. Unlike their predecessors, they formulated the assumptions of their propositions with extraordinary preciseness and only attempted to apply the concept of probability to such phenomena whose occurrence or non-occurrence could be experimentally checked. Thus, for them, the LLN as well as the limit theorems made sense not as abstract characteristics of fictitious infinite collectives, but as approximations to quantitative relations observed in sufficiently numerous and actually existing aggregates of some random elements.

[3] This point of view, which puts in the foreground the establishment of the appropriate inequalities for the finite case, and the estimation of the errors of limit theorems, is already outlined in Chebyshev's first work of 1846. There, while offering a new proof of the generally known Poisson theorem, the author justified his research by stating that

Regardless of the cleverness of the method used by the illustrious geometer, it does not provide the boundary of the possible error.

Filling in this gap, Chebyshev offered a very precise estimate of the studied error and became able to ascertain the approximation furnished by the Poisson formula when applied to a given and sufficiently long but finite series of trials. The same tendency characterizes all the later investigations accomplished by Chebyshev; in particular, it marks his discovery, made twenty years later, of the form of the LLN remarkable for its generality and embracing all the practically met with totalities of independent random variables. The strikingly simple idea of his proof {of this form of the law} is based on an inequality that states that if the variance of a random variable is sufficiently small, its considerable fluctuations are very unlikely.

The {Bienaymé – }Chebyshev inequality is extremely general so that the estimate of the probability that it provides is often insufficient for practical applications. However, the authors of all of its subsequent specifications, obtained at the expense of appropriate restrictions, have essentially applied Chebyshev's idea. Just the same, the later generalizations of the applicability of the LLN to totalities of either independent or dependent variables, which provide an exhaustive solution of the problem concerning this law, and which were accomplished mostly by Markov, and, in part, by later authors, were based on one or another development of Chebyshev's reasoning.

The same should be said about the proofs of the so-called strong LLN which was beyond the scope of the Petersburg school, and which, like its ordinary counterpart, is applicable to all practically met with totalities of independent random variables. As Academician Kolmogorov has recently shown, it is indeed applicable to all independent variables satisfying the Chebyshev condition of restricted variations. It should be noted in this connection that the generally accepted shortened formulation of the strong LLN is a proposition concerning an infinite repetition of independent trials. These are actually unrealizable, do not admit of experimental checks, and are therefore inadmissible from the point of view of the Petersburg school. Nevertheless, the really concrete substance of the strong law is expressed with perfect preciseness by means of the inequalities of the Petersburg school, and its proof is based on exactly this fact.

For example, if we, for the sake of simplicity, restrict our attention to the repetition of independent Bernoulli trials with p being the probability of the occurrence of event A, and $p_n = m/n$, the frequency of its occurrence in n trials, the strong LLN (in this case, the strong Bernoulli theorem) is reduced to the following proposition: For any given magnitudes $\varepsilon > 0$ and $\eta > 0$ it is possible to indicate such a sufficiently large number n_o, that the probability of the simultaneous realization of the inequality $(p_n - p) < \varepsilon$ for all values of n, $n_o \le n < n + k$, will be greater than $1 - \eta$ however large are the values of $k > 0$. In this form, the only one suitable for applications, the strong LLN does not therefore differ in essence from the classical LLN; it is only some supplement of the latter. As already noted above, all the recently made relevant theoretical investigations, at least insofar as they can be applied to natural sciences, were essentially based on Chebyshev's ideas.

Among the two problems mentioned above that gave rise, in the last {the 19th} century, to the crisis of probability theory, the one connected with the LLN was solved in the most general way, and by essentially simpler mathematical means than those {earlier} applied by the classics for obtaining its more or less particular solutions. This only became possible because Chebyshev and his followers attentively and precisely analyzed the general properties of the most diverse random variables. On the contrary, the solution of the second problem concerning the normal law presented profound mathematical difficulties and demanded the creation and use of new analytical methods. The complete investigation and solution of the problem about the conditions for the applicability of the normal law in the limiting case to the sums of very large numbers of independent random terms should therefore be considered as the most outstanding achievement of the Petersburg school; it was jointly accomplished by all three of its great representatives.

Chebyshev made the first essential step towards solving this problem. His way consisted in approximately determining a monotone function representing the integral law of distribution of a random variable, given some finite number of its successive moments, *i.e.*, expectations of its consecutive integral powers. Owing to his profound investigation of the extreme values of definite integrals, very important in other problems of mathematical analysis as well, Chebyshev inferred, in particular, that the normal law of distribution was completely determined by the sequence of all of its moments. Using arguments that were not, however, completely rigorous, he then ascertained that the successive moments of the sums of a very large number of independent variables tended to normal moments, and concluded that the limit of the distribution for the sum of such variables was the normal law. However, the two last links of his reasoning demanded essential specification, and this was done by Markov.

[4] Here, I have to enter into some mathematical details without which it is impossible to conceive the successive stages in the formulation and solution of the problem concerning the normal law. Following Chebyshev, Markov only considered sums

$$S_n = x_1 + x_2 + \ldots + x_n$$

for such independent random variables x_i that possessed expectations or central moments $C_i^{(p)}$ of all integer powers p. Calculating the central moments of their sum S_n, he proved, with an indisputable mathematical rigor peculiar to all his investigations, that for {$B_n = \text{var}\, S_n$} all integral values of p, if

$$M_n^{(p)} = (1/B_n^{p/2})\sum_{i=1}^{n} C_i^{(p)} \to 0,$$

$$\lim \frac{ES_n^{\,p}}{B_n^{\,p/2}} = (1/\sqrt{2\pi}\,)\int_{-\infty}^{\infty} \exp(-t^2/2)t^p dt = A_p = \sum_{i=1}^{n} C_i^{(p)},\ n \to \infty.$$

He then ascertained that, if the moments of the sum $S_n/\sqrt{B_n}$ tended to the normal moments A_p of the corresponding orders, the distribution of this sum tended to the normal law. Thus, basing his reasoning on Chebyshev's ideas, Markov, in 1898, for the first time discovered a precise and sufficiently general formulation of the conditions for the applicability of the normal law of distribution to the sums of independent random variables. These conditions are, therefore, that, as n increases, $M_n^{(p)}$ tends to vanish for all integral values of $p > 2$. In particular, this is fulfilled if all the magnitudes x_i remain restricted, and $B_n \to \infty$ as n increases indefinitely, *i.e.*, if $x_i/\sqrt{B_n} \to 0$. Because of this, Markov's result embraced all the former rigorously ascertained forms of the limit theorem, and, in addition, covered a considerable part of its practical applications since the terms x_i were indeed usually restricted. Nevertheless, when schematically representing natural phenomena by a theory, the last condition is sometimes too inconvenient. If, in such cases, one discards it, but is reconciled with the existence of the moments of all orders, it can be difficult to verify whether these moments obey all the Markov conditions.

The appearance, in {1900 and} 1901, of Liapunov's work was therefore an extremely important event, both theoretically and practically. Applying quite another method, that of characteristic functions, he formulated a remarkably simple and general condition for the applicability of the normal law. The idea of his method goes back to Laplace, Fourier and Dirichlet, but its rigorous realization presented, as it seemed, insurmountable mathematical difficulties. It is Liapunov's considerable historic merit that he had overcome this obstacle after which the method of characteristic functions secured for itself the central place in mathematical investigations in modern probability theory. Liapunov's sufficient condition for the applicability of the normal law to the sum S_n of independent random variables x_i in the limiting case consisted in that only one of the Markov restrictions had to be fulfilled: $M_n^{(p)}$ should tend to zero at least for one value of $p > 2$ where p was not necessarily an integer.

Soon after Liapunov had discovered this general, and very convenient for applications condition, Markov obtained the same result as a corollary of his own theorem by means of a simple, but essentially important remark, which contemporary authors are using very often in similar cases, apparently without knowing about Markov's priority. It consists in the following: Liapunov's condition $M_n^{(p)} \to 0$ for some $p > 2$ means that, although the terms x_i can be arbitrarily large, the probability of their taking values of the order of $\sqrt{B_n}$ is so low, that we shall not change the limiting law of distribution for their sum by assuming that such large values are impossible; and, as stated above, the Markov theorem will be valid for the thus changed sum.

By referring to the same remark it was not difficult for me, in 1926, to replace the Liapunov sufficient condition by a somewhat more general one, according to which the individual terms x_i can even posses no finite variances. In 1936, Feller, a young German mathematician, proved that the new condition was not only sufficient, but, in a sense, necessary for the applicability of the normal law to the sums of independent variables in the limiting case.

One circumstance, not yet noted by anyone, should, however, be paid attention to. In a sense rather interesting for applications, the Liapunov condition $M_n^{(p)} \to 0$ for a given p is at the same time sufficient and necessary. Indeed, the infinite totality of the Markov conditions expresses not only that the distribution of $S_n/\sqrt{B_n}$ has the normal law as its limit, but, in addition, that all of its moments tend to normality since both these facts are inseparable when the Chebyshev – Markov method of moments is applied. On the contrary, the Liapunov condition for some single $p > 2$, whose inalienable corollary is incidentally the fulfilment of the pertinent conditions for all smaller values of p, does not assume anything about the moments of the higher orders, which, generally speaking, can even be non-existent.

It can be shown, however, that the Liapunov condition of the order of p is equivalent to $E(S_n/\sqrt{B_n})^p$ having as its limit the corresponding normal expectation; for integral and even values of p this directly follows from Markov's reasoning. This condition for power p, or the equivalent single demand that the moment of this order for magnitudes $(S_n/\sqrt{B_n})^p$ has as its limit the corresponding normal moment, is therefore necessary and sufficient for the law of distribution of the sum $(S_n/\sqrt{B_n})$ to tend to the normal law, and, at the same time, for its moments not higher than p to tend to the normal moments. For $p = 4$ this remark is also of some practical importance, since, when fitting empirical curves of distribution, statisticians usually make use of the moments up to and including the fourth order. It follows that the so-called excess being close to zero is, under appropriate conditions, necessary and sufficient for the closeness of the pertinent distribution to normality. Owing to Liapunov's and Markov's investigations, which supplemented each other, the historic merit of the complete elucidation and final solution of the general problem about the conditions for the applicability of the normal law to sums of independent random variables thus exclusively belongs to the Petersburg school of the theory of probability.

[5] The last stage in the development of this school that took place during the first two decades of our {of the 20th} century was only connected with Markov. After the problems posed by the LLN and the normal law were solved for the case of any independent variables, Markov, for the first time ever, precisely posed and began to extend these laws onto dependent variables. He thus sketched the main line of the further development of probability theory in the 20th century, *i.e.*, the investigation of all regularities occurring when a large number of random variables somehow connected with each other are aggregated. As stated above, Markov, without experiencing any particular difficulties, extended Chebyshev's reasoning onto dependent variables and offered an essentially exhaustive solution of the problem of the LLN for this case as well. When applying this law to mass aggregates of elements depending on each other, we need not anymore be afraid of repeating bygone mistakes which were denounced at the time as a mathematical opprobrium.

When studying the dependences between random events or variables, Markov, who possessed a remarkable flair for natural sciences, paid special attention to, and profoundly investigated an important class of sequences of dependent events which he called *chains* and which are now known in science as *Markov chains*. The English scientist Galton was the first to consider some particular cases of such kinds of dependences. Already in the last {the 19^{th}} century, having a rather vast and diverse experimental material at his disposal, he attempted to convert the Darwinian theories of heredity and evolution into mathematical form. Neither he, nor the physicists, who also came to construct stochastic patterns similar to Markov chains almost at the same time as the Russian mathematician did, have offered their precise definition, or, still less, any satisfactory mathematical analysis of their properties.

The closeness of Markov's ideas to those, having emerged at the same time in diverse fields of natural sciences, testify to their vitality and to the wide scope of their possible applications. Indeed, a Markov chain is a stochastic transformation of a usual determinate process whose dynamic state at a given moment fully defines its further course independently of all its previous states. More specifically, according to Markov, a chain is such a sequence of random variables $x_1, x_2, \ldots, x_n, x_{n+1}$ that, after x_n takes some definite value, the probabilities that the subsequent variable x_{n+1} takes some values are fully determined independently of the values that the previous variables $x_1, x_2, \ldots, x_{n-1}$ had.

Excluding the limiting case in which the value of the variable x_n fully defines the next link, and the chain is thus transformed into a determinate process, Markov proved that the dependence between the links weakens with the increase in the distance between them, and went on to infer that the law of distribution for the variable x_n tends to some *limiting stationary law* independent from the initial value of x_1. In particular, this theorem, that finds many applications, can, under some very general suppositions, justify the underpinning principle of the kinetic theory of gases, *i.e.*, that of the uniform distribution of energy among the phases for established conditions. The extension of the LLN onto Markov chains is a simple corollary of the fact that the dependence between remote links of the chain is very weak.

The problem about the limiting applicability of the normal law to the sum of terms constituting a chain presented more essential difficulties. Applying the method of moments, Markov also solved it under rather general suppositions. The further generalization of these results was achieved later, and I shall not dwell on this topic.

Concluding my review of the investigations of Markov himself into dependent variables, I shall only cite

1) His curious philological phonetic application of chains to studying the alternation of vowels and consonants in literary works; and

2) One of his last contributions of a somewhat different nature. There, he indicated and analyzed a stochastic pattern leading to limiting laws of distribution represented by the Pearsonian curves. He thus revealed the essence of the conditions under which the use of these curves in statistics was justified.

[6] Only one point is left now, viz., the listing of the main directions of the modern probability theory that have been developed under the direct influence of the Petersburg school. First of all, I mention the systematization and detailed study of all the limiting laws obeyed by the sums of an infinitely increasing number of independent terms, including the case in which the Liapunov condition does not hold. These investigations threw additional light on the special importance of the normal law in probability theory, where its role is similar to that played by the principle of inertia in mechanics. On the other hand, the Liapunov method of characteristic functions was extended onto the case of several sums so that the theory of normal correlation became thus justified. Finally, the investigation of dependent variables that followed along the way paved by Markov proved most fruitful and important. Here, we now have essentially new methods and results relating to the Markov chains themselves and to their various generalizations; above all I should cite the so-called theory of stochastic differential equations which is undoubtedly destined for a great future. It is not my aim to review modern contributions belonging to this direction; there, the theory of probability, now remote from its engendering elementary problems on random events in games of chance, is a powerful mathematical tool for cognizing nature.

For its transformation from a mathematical amusement into a method of natural sciences the theory of probability is mainly obliged to the Petersburg school which accomplished this fundamental progress leaving West European mathematicians far behind.

Note

1.{Liapunov left four contributions on the theory of probability, and he actually read a course in probability for three years, see B.V. Gnedenko, On the work of Liapunov in the theory of probability. IMI, vol. 12, 1959, pp. 135 – 160, §1; translated in P.A. Nekrasov, *The Theory of Probability*. Berlin, 2004.}

2.{See Note 9 to Bernstein's essay on Chebyshev translated in this book.}

8. Oscar Sheynin. On the Relations between Chebyshev and Markov

. *Istoriko-Matematich. Issledovania*, to appear

Markov regarded Chebyshev with highest respect and became one of the two editors of the first (Russian and French) sufficiently complete collection of his works [1]. And he sharply opposed Nekrasov when the latter had begun to belittle the importance of Chebyshev's proof of the CLT. At the same time, however, the relations between Chebyshev and Markov had not been serene. I describe the appropriate facts (§1), and, in §2, make public the extant part of Markov's report devoted to the memory of his teacher. Finally, I have to say a few more words about Nekrasov (§3). My paper only deals with the theory of probability and the MLSq.

1. Chebyshev and Markov

1.1. Chebyshev's memoir (1887). The appearance of Nekrasov's paper (1898a) devoted to the memory of Chebyshev led to a discord between him and Markov. The latter (1898b) had accused him of underestimating the Chebyshev memoir and then (1899b) remarked that Nekrasov, in his paper, had not mentioned it. Here are Nekrasov's pertinent statements (1898b; 1900, p. 384).

The theorems considered in Chebyshev's memoir are due to Laplace; Chebyshev only devised a better proof for them.

I set infinitely high store by his immortal memoir [Chebyshev 1867] *which is a greatest contribution to science* [to probability theory]. *And I consider his memoir* [Chebyshev 1887] *as of minor importance since it contains that, which was sufficiently rigorously proved much earlier and included in generally known treatises* [2]. *It is only interesting as being one of the successful applications of Chebyshev's great inventions to* [ingenuity in considering] *problems exhausted earlier.*

Markov did not restrict his activities here to criticizing Nekrasov. I quote the opening lines of his memoir (1899a):

Chebyshev's memoir (1887) *is important. Regrettably, this fact is obscured by two circumstances, viz., by involved derivations and insufficient rigor of reasoning*

And Markov went on to eliminate Chebyshev's shortcomings in the proof of the CLT. In a subsequent work Markov (1913, p. 321) called Chebyshev's memoir "remarkable".

1.2. Bienaymé. Two issues connect Bienaymé and Chebyshev (Heyde & Seneta 1977; Gnedenko & Sheynin 1978, pp. 258 and 261 – 262); Sheynin (1989, pp. 346 – 348). First, in 1853 Bienaymé proved that, in modern notation,

$$P(\bar{x} - E\bar{x} \geq \alpha) \leq \text{var}\,\bar{x}\,/\alpha^2 \qquad (1)$$

where $\bar{x}$ was the arithmetic mean of a series of observations and $\alpha > 0$. Bienaymé had not, however, isolated his formula from the general context of his memoir.

Markov (1924, p. 92; also in the edition of 1913) called (1) the Bienaymé – Chebyshev inequality and mistakenly stated that Bienaymé had only indicated the main idea of its proof, restricted, moreover, by some conditions. Recall, however, that Chebyshev had not only independently proved that inequality but applied it for deriving most important corollaries to his LLN.

Second, again in 1853, Bienaymé investigated the order of the even moments of sums of independent observational errors whereas Chebyshev applied the moments for proving the CLT.

Markov (1912, p. 218) declared that he had "developed" and "extended" "the Bienaymé method". Soon afterwards, in the Introduction to the German edition of his treatise [see Ref. Markov (1924)], he mentioned the "remarkable Bienaymé – Chebyshev method" [3]. Nekrasov (1911, p. 436), however, thought that Chebyshev had exhausted the "idea of Bienaymé" and referred to Chebyshev (1874). It was in this connection that Markov returned to that issue in 1912. Moreover, he justly added that Chebyshev, who had highly respected Bienaymé (Gnedenko & Sheynin 1978, pp. 261 – 262), had never said anything of the sort.

The method of moments is less effective than the method of characteristic functions and had not been generally applied in the 20th century (Zolotarev 1999). Already Liapunov (1900, p. 125 of the Russian translation) had called Markov's proof of the CLT "too involved and unwieldy" and Bernstein (1945, p. 427) stated that the method of moments did not "facilitate" the proof of the CLT but rather "transferred all the difficulties elsewhere". Nevertheless, that method is still applied in mathematical statistics for solving concrete problems (Soloviev 1997, p. 12).

1.3. The Method of Least Squares. Markov (1899a, pp. 249 – 250) criticized Maievsky (1881) (and other unnamed Russian authors) for an unsuccessful attempt to justify the optimal property of the arithmetic mean (the Gauss postulate). Maievsky had cited the Chebyshev form of the LLN whereas Markov remarked that his substantiation was hollow. It is remarkable that Chebyshev "communicated" Maievsky's manuscript to the physical and mathematical section of the Petersburg Academy of Sciences (Mandryka 1954, p. 81). It is unclear, however, whether he actually noted the pertinent attempt.

Anyway, Chebyshev had hardly been interested in the treatment of observations: he (1936, pp. 249 – 250) wrongly described the estimation of their precision (Sheynin 1994, p. 336) [4] and he did not describe definitely enough the application of the CLT to the theory of errors. Before formulating that theorem, he remarked, that, given a large number of observations, it led to the MLSq; actually, to the normal distribution whose existence is not, however, necessary for justifying the method. In addition, Gauss had on principle only considered the treatment of a finite number of observations (and Markov, see above, could have well indicated this fact also).

1.4. The Offence. And so, while fully understanding the importance of Chebyshev's work, Markov nevertheless felt it necessary to criticize him directly (§1.1) and perhaps indirectly (§1.3). And he also had grounds for feeling himself offended by his teacher. Bernstein (1945, pp. 416 – 417) indicated that the path from the Chebyshev inequalities (Chebyshev 1874) to the CLT had already then, after 1874, been obvious and noted, on p. 419, that Chebyshev, who had no rivals, was in no hurry either to prove those, only formulated by him inequalities, or to surmount the last obstacles.

Markov (1884), however, without waiting for Chebyshev, proved these inequalities. Krein (1951) described their history and I shall only repeat that Stieltjes had independently substantiated them a bit later, and then (1885) expressed his regret that he did not, at the time, know about Markov's proof.

After that, Chebyshev, on the contrary, speeded up his stochastic investigations (Bernstein 1945, pp. 419 – 420 and 423) now collected, together with his other contributions, in vol. 3 of his *Полное СобраниеСочинений* (Complete Works). But he had not even once mentioned either Markov or Stieltjes. No wonder that Markov voiced his dissatisfaction, although only in a private letter to the eminent geologist A.P. Karpinsky (1847 – 1936) (Grodsensky 1987, pp. 62 – 63). In 1885 [5], as he wrote, Chebyshev "published without proof and without mentioning me his solution, and he is continuing in the same vein".

All this resembles the well-known attitude of Laplace and Gauss to other scientists. Here, however, is a reasonable opinion about the latter (Biermann 1966, p. 18):

What is forbidden for usual authors, should be really allowed for Gausses, and in any case we ought to respect the considerations that guided him.

This means that suchlike attitudes are forgiven by future generations only with respect to Gausses,– and therefore to Chebyshev, as I would add.

2. Markov's Report

2.1. Some Explanation. The centenary of Chebyshev's birth occurred in 1921. Liusternik (1965, p. 23) reported that in the spring of that year Moscow University and the Moscow Mathematical Society had received an invitation from the Academy of Sciences, then located in Petrograd (former, and present-day Petersburg) to participate in a scientific conference devoted to that occasion. Liusternik, who was present there, indicated that it was convoked on Markov's initiative.

This statement might be complemented. Here is Markov's own earlier pronouncement (Grodzensky 1987, p. 63):

> *This day should have been commemorated by a special grand meeting; however, since our life is going on in an abnormal form, any celebrations whatsoever are hardly appropriate.*

But he added that he did not exclude such a possibility if only the physical and mathematical section of the Academy decided that a meeting is necessary.

More than seven reports were read at the conference, Markov's included (Liusternik (1965, p. 25):

Markov seemed physically weak and old, but his eyes shone provocatively. He spout caustic hints. He read a report imparting his recollections of Chebyshev. It was regrettably neither written down, nor taken down short-hand, and all that, what a man, who had for several decades rubbed shoulders with Chebyshev, was able to say, was irrevocably lost.

A certain text is however partly extant. It was written by Markov and devoted to Chebyshev, but has neither date nor signature. Markov's authorship is nevertheless doubtless because the text includes such phrases as "… in 1883 I was able …", and no-one except for himself could have written them. The text has no "reminiscences" [6]; it is possible, however, that the organizers of the conference had asked Markov to change somewhat the subject of his future speech (or that he himself included them later on). In any case, the text was likely written on the occasion of Chebyshev's centenary.

The extant text is kept at the Archive of the Russian Academy of Sciences, *Fond* 173, Inventory 1, sheet 60, and written on a form with the following letterhead:

Deputy Chairman, Vasileostrov district, Petrograd Commercial & Industrial Union. Vas[ilievsky] Ostr[ov], Nikolaevskaia Embank. 11, Tel. 77627

Markov's references were incomplete, some of them lacking altogether. He abbreviated many words; his first phrase, for example, ran as follows: "Cheb works th prob make up consid part of his scient activ and are very remark".

2.2. The Text

Chebyshev's works on the theory of probability make up a considerable part of his scientific activity and are very remarkable. Nevertheless, they had not attracted either foreign or even Russian scientists. Such an attitude might be explained by the usual understanding of probability theory as being an applied science for which mathematical rigor is unnecessary. An additional circumstance is that Chebyshev was concerned not with new propositions of the theory, but with new methods for establishing those main theorems which are generally considered as having been quite ascertained by Laplace and Poisson. Although Chebyshev had been well known in France, and especially in Paris, neither Bertrand (1888), nor Poincaré (1896) mentioned him at all. Czuber, in his essay (1899) on the development of the theory of probability, twice cited Chebyshev's works, but his indications did not provide any clear idea about them.

Apart from two short notes, which appeared in the Liouville and Crelle periodicals {in *J. math. pures et appl.* and *J. reine und angew.Math.*}, Chebyshev's first published work is his [Master's] dissertation *Essay on an Elementary Analysis of the Theory of Probability* of 1845, which, as stated in its very title, was devoted to probability theory. Note that it appeared a year earlier than the generally known fundamental work of another of our late academicians, Buniakovsky (1846). Note also that the teaching of the theory of probability at Moscow University, where Chebyshev had obtained his higher education, only began in 1850, even later than in Petersburg. In Petersburg {where Chebyshev had been working} the subject of the dissertation was proposed not by a university professor [7] (who is not {none of whom are} mentioned either in its Introduction or main text, but by the curator of the Moscow educational region, Count {S.G.} Stroganov.

{An obvious gap in the text appears here.}

This note had been unnoticed until the beginning of the 1880s when it caused a lively exchange of opinions among Petersburg scientists, and mostly between me and my respected teacher, K.A. Posse, whom I am greeting as the eldest student of Chebyshev present at our meeting [8]. Professor Korkin {Chebyshev's student, 1837 – 1908} even doubted that the Chebyshev inequality was valid. Finally, in 1883 I was able not only to prove it, but to solve his problem {Markov 1884a}. And I was just in time because a little later the eminent Dutch scientist Stieltjes {Krein 1951} arrived at the same inequality and at the problem of moments not knowing that Chebyshev had published these inequalities in 1873[9]. And Stieltjes' proof fully coincided with mine [10].

My dissertation that appeared in 1884{1884b} prompted Chebyshev to return to his problem [11] with which a large part of his work accomplished during the last ten years of his life had been connected. From among nine contributions of that period only two ("On the simple components" and "Sur les expressions approchées de la racine carrée) have nothing in common with the method of moments [12].

the results of Chebyshev's inferences found application in ballistics as witnessed by Maievsky's book (1881) which I have already mentioned and by his reference to the writing by the French artillery Captain Jouffret. Finally, the German three-volume treatise

Regarding the other ones, it might be assumed that they only constitute a part of a contemplated series whose implementation was thwarted by his death. Perhaps he intended to supplement the previous conclusions by estimating the pertinent deviations caused by finiteness.

Two years separate this period from the previous one: from 10 May 1883 until 8 October 1885 Chebyshev had not communicated a single work to the Academy [13]. He concluded by the memoir "Sur les sommes qui dépendent des valeurs positives d'une fonction quelconque" communicated in 1894 and published posthumously in 1895. I consider the three first contributions of that period, closely connected with each other, as being the main ones. These are (1885), communicated 8 October 1885, the next one, "Sur les résidus intégraux", communicated 18 November 1886, and, finally, (1887), communicated 10 March 1887.

3. Nekrasov

On Nekrasov's work in the theory of probability, mostly on the CLT, see Seneta (1984), Soloviev (1997) and Sheynin (2003). Nekrasov apparently proved the CLT for the case of large deviations, but he attempted to justify his innovation by an arbitrary interpretation of Chebyshev's formula of the theorem. Chebyshev (1887) had indicated that the integral of the exponential function of the negative square had "any" limits whereas Nekrasov (1911, p. 449) declared that Chebyshev had considered variable limits. Liapunov's statement (1901, pp. 62 – 63) has been left unnoticed. He explained that he was not interested in the case of variable limits, but nevertheless adduced some pertinent remarks.

Commentators believe that Nekrasov's unthinkable mistakes and absurd pronouncements were caused by the burden of his administrative work; by his inadmissible desire to unite mathematics, religion and philosophy; and even (Mikhailov & Stepanov 1985, p. 225) by a mental illness. It is not amiss to add, however, that Nekrasov was indeed able to absorb the ideas of the religious philosopher V.S. Soloviev, who (Radlov 1900, p. 786) had stated, that "philosophy offers its hand to religion"; that "the basis of veritable knowledge is mystic or religious perception"; and that (p. 787) the "system of veritable knowledge is a thorough synthesis of theology, rational philosophy and positive science". So had not Nekrasov subordinated mathematics to religion and (poorly understood) philosophy?

It was the same Nekrasov, who had previously been quite another person, and, in 1896, wrote a marginal note on the text of Chuprov's candidate dissertation (Sheynin 1990, p. 85):

We [mathematicians] *are content with notions concerning force, space, time, probability, etc. Concerning these same subjects philosophers have written full volumes of no use for physicists or mathematicians. From our point of view, Mill, Kant and others are not better but worse than Aristotle, Plato, Descartes, Leibniz …*

Notes

1. Markov Jr (Markov 1951, pp. 605 – 606) described a remarkable episode connected with the preparation of Chebyshev's collected works (vols 1 – 2, 1899 and 1907, in Russian and French; French edition reprinted: New York, 1962). In 1902, after a hardly justified revocation of the election of Maxim Gorky to honorary membership of the Academy of Sciences, Markov sent in his resignation from that body. His application was refused and he resumed the editorial work "apparently having been afraid of handing over that responsible work to someone else". And it was Chebyshev on whose proposal Markov was elected, in 1886, adjunct of the Academy of Sciences.

2. Here, Nekrasov referred to Laurent (1873, pp. 144 – 165) who nevertheless (on pp. 144 – 145) had not estimated the error of his approximations. Can it be that Nekrasov deliberately indicated wrong page numbers?

3. Seneta (1998, p. 297) somehow decided that Markov had defended Bienaymé's priority in spite of his "volatile" nature. Markov, however, always acted in accord with what he considered (although sometimes mistakenly) just; he admired Chebyshev's talent, but would have never attributed to his teacher more than the deserved. True, Markov (1914, p. 162) once remarked that he "sometimes" called the method of moments only after Chebyshev, but this did not change his attitude towards Bienaymé.

4. Chebyshev stated that the Gauss formula for estimating the variance was based either on the existence of a large number of observations, or on the "law of hypotheses" that "does not deserve great confidence". By that law he apparently meant the existence of a uniform prior distribution of observational errors only assumed by Gauss in his first justification of the MLSq; that assumption was not even required because of the introduced at the same time postulate of the arithmetic mean. Then, Gauss never considered the case of a large number of observations, a fact that also bears on the last sentence of my §1.3 and could have been indicated by Markov, see above. Finally, when comparing the Gauss formula with its previous version, Chebyshev, unlike Gauss (whom he never mentioned), had not indicated that it provided an unbiased estimate of the variance.

5. Markov had in mind the memoir (1885) read 8 October 1885. Indeed, he was not mentioned there.

6. There were no "caustic hints" either. These could have been levelled, first and foremost, against Nekrasov.

7. Markov hardly meant Buniakovsky who had become professor in 1846.

8. Had Posse been absent, which was unlikely, Markov would have omitted the appropriate phrase. Alternatively, he could have written his text after reading the report.

9. The plural form (inequalities) is correct, the singular form (above) is wrong.

10. The next lines up to the "three-volume treatise" were crossed out. The phrase "Regarding the other ones" apparently meant "Regarding the other seven memoirs". I mentioned Maievsky (see below) in §1.3, but he is not cited in the previous lines of the extant text.

11. Cf. Bernstein's relevant pronouncement in §1.4.

12. A chronological list of Chebyshev's works is included in vol. 5 of his *Complete Works* and it only contains the second of those memoirs (published in 1889). On the other hand, it has another memoir of the same year, "Sur le système articulé le plus simple …" obviously not connected at all with the method of moments. All nine of Chebyshev's last memoirs were included in the first edition of Chebyshev's collected works, see Note 1.

13. Cf. Note 5.

References

Bernstein, S.N. (1945), On Chebyshev's work on the theory of probability.Translated in this book.

Bertrand, J. (1888), *Calcul des probabilités.* Paris. Second edition, 1907; reprint, 1970.

Biermann, K.-R. (1966), Über die Beziehungen zwischen Gauss und Bessel. *Gauss-Ges. Göttingen*, Bd. 3, pp. 7 – 20.

Buniakovsky, V.Ya. (1846), *Основания математической теории вероятностей* (Principles of Mathematical Theory of Probability). Psb.

Chebyshev, P.L. (1867), Des valeurs moyennes. *J. math. pures et appl.*, t. 12, pp. 177 – 184.

--- (1874), Sur les valeurs limites des intégrales. Ibidem, t. 19, pp. 157 – 160.

--- (1885, in Russian), Sur la répresentation des valeurs limites des intégrales par les résidus intégraux. *Acta Math.*, t. 9, 1886, pp. 35 – 56.

--- (1887, in Russian), Sur deux théorèmes relatifs aux probabilités. Ibidem, t. 14, 1890 – 1891, pp. 305 – 315.
--- (1936), *Теория вероятностей. Лекции 1879 – 1880* (Theory of Probability. Lectures of 1879 – 1880). M. Translation: Berlin, 2004.
--- (1946 – 1951), *Полное Собрание Сочинений* (Complete Works), vols 1 – 5. M. – L.
Czuber, E. (1899), Die Entwicklung der Wahrscheinlichkeitstheorie und ihrer Anwendungen. *Jahresbericht Deutscher Mathematiker-Vereinigung*, Bd. 7, No. 2. The whole issue.
Gnedenko, B.V., Sheynin, O.B. (1978, in Russian), Theory of probability. In Kolmogorov, A.N., Yushkevich, A.P., Editors (1992), *Mathematics of the 19th Century*, vol. 1. Basel, pp. 211 – 288. Reprinted in 2001.
Grodzensky, S.Ya. (1987), *Markov.* M. (R)
Heyde, C.C., Seneta, E. (1977), *Bienaymé.* New York.
Krein, M.G. (1951), Chebyshev's and Markov's ideas in the theory of the limiting values of integrals and their further development. *Uspekhi Matematich. Nauk*, vol. 6, No. 4, pp. 3 – 120. (R)
Laurent, H. (1873), *Traité du calcul des probabilités.* Paris.
Liapunov, A.M. (1900), Sur une proposition de la théorie des probabilités. *Izvestia Imp. Akad.Nauk St. Pétersb.*, sér. 5, t. 13, pp. 359 – 386.
--- (1901), Answer to Nekrasov. *Zapiski Kharkov Univ.*, vol. 3, pp. 51 – 63.
Liusternik, L.A. (1965), Report at the jubilee sitting of the Moscow Mathematical Society. *Uspekhi Matematich. Nauk*, vol. 20, No. 3 (123), pp. 21 – 30. (R) That periodical is being translated as *Russian Math. Surveys.*
Maievsky, N.V. (1881), *Изложение способа наименьших квадратов и применения его преимущественно к исследованию результатов стрельбы* (Explication of the Method of Least Squares and of Its Applications Mostly for Investigating the Results of Artillery Firing). Psb.
Mandryka, A.P. (1954), *Maievsky.* Moscow. (R)
Markov, A.A. [Sr] (1884a), Proof of some of Chebyshev's inequalities. In author's book *Избранные труды по теории непрерывных дробей и теории функций наименее уклоняющихся от нуля* (Sel. Works on the Theory of Continued Fractions and the Theory of Functions Least Deviating from Zero). M.– L., 1948, pp. 15 – 24.
--- (1884b), *О некоторых приложениях алгебраических непрерывных дробей.* (On Some Applications of Algebraic Continued Fractions). Doctor's Dissertation. Psb.
--- (1899a), The law of large numbers and the method of least squares. In Markov (1951, pp. 233 – 251). Translated in this book.
(1899b), An answer [to P.A. Nekrasov]. *Izvestia Fiziko-Matematich. Obshchestva Imp. Kazan Univ.,* ser. 2, vol. 9, No. 3, pp. 41 – 43.
--- (1912), A rebuke to P.A. Nekrasov. MS, vol. 28, pp. 215 – 227.
--- (1913), The theorem on the limit of probability for the Liapunov case. In Markov (1951, pp. 321 – 338).
--- (1914), Bicentennial of the law of large numbers. In Ondar, Kh. O., Editor (1977, in Russian). *Correspondence between Markov and Chuprov*, etc. New York, 1981, pp. 158 – 163.
--- (1924), *Исчисление вероятностей* (Calculus of Probability). M. Previous editions: 1900, 1908, 1913. German translation of second edition: Leipzig – Berlin, 1912.
(1951), *Избранные труды* (Sel. Works). No place.
Markov, A.A. [Jr] (1951), Biography of Markov Sr. In Markov (1951, pp. 599 – 613). Translated in this book.
Mikhailov, G.K., Stepanov, S.Ya. (1985), On the history of the problem of the rotation of a solid about a fixed point. IMI, vol. 28, pp. 223 – 246. (R)

Nekrasov, P.A. (1898a), General properties of mass independent phenomena, etc. MS, vol. 20, pp. 431 – 442.
--- (1898b), Letter to Markov of 7 Oct. 1898. Archive of Russian Academy of Sciences, *Fond* 173, Inventory 1, *Delo* 53, No. 1.
--- (1900), Concerning Markov (1899b). MS, vol. 21, pp. 379 – 386.
(1911), On the principles of the law of large numbers, the method of least squares and statistics. MS, vol. 27, pp. 433 – 451.
Poincaré, H. (1896), *Calcul des probabilités.* Paris. Second edition, 1912, reprinted in 1923 and 1987.
Radlov, E.L. (1900), V.S. Soloviev. His religious and philosophical views. *Энциклопедич. словарь Брокгауза и Ефрона* (Brockhaus & Efron Enc. Dict.), halfvol. 60, pp. 785 – 792. (R)
Seneta, E. (1984), The central limit theorem and linear least squares in pre-revolutionary Russia. *Math. Scientist*, vol. 9, pp. 37 – 77.
--- (1998), Bienaymé (1798 – 1878): Criticality, inequality, and internationalization. *Intern. Stat. Rev.*, vol. 66, pp. 291 – 301.
Sheynin, O. (1990, in Russian), *Chuprov.* Göttingen, 1996.
--- (1989), Markov's work on probability. *Arch. Hist. Ex. Sci.*, vol. 39, pp. 337 – 377; vol. 40, p. 387.
--- (1994), Chebyshev's lectures on the theory of probability. Ibidem, vol. 46, pp. 321 – 340.
--- (2003), Nekrasov's work on probability: the background. *Arch. Hist. Ex. Sci.*, vol. 57, pp. 337 – 353.
Soloviev, A.D. (1997), Nekrasov and the central limit theorem of the theory of probability.IMI, ser. 2, vol. 2 (37), pp. 9 – 22. Transl.: DHS 2799, 2004, pp. 199 – 208.
Stieltjes, T.J. (1885), Note à l'occasion de la réclamation de M. Markoff. *Oeuvr. Compl.*, t. 2. Berlin, 1993, pp. 430 – 431 (518 – 519). Double paging.
Zolotarev, V.M. (1999), The method of moments. In Prokhorov, Yu.V., Editor (1999), *Вероятность и математическая статистика. Энциклопедия* (Probability and Math. Statistics. An Encyclopedia). M., p. 360.

Translations of the following Russian sources are available in P.A. Nekrasov, *Theory of Probability*. Berlin, 2004: Liapunov (1901); Markov (1898; 1899b; 1912; 1913; Nekrasov (all those referred to).

9. A.M. Liapunov. On the Gauss Formula for Estimating the Measure of Precision of Observations

Manuscript, first published by the present translator in *Istoriko-Matematich. Issledovania*, vol. 20, 1975, pp. 319 – 328.

Suppose that n observations are made to determine m magnitudes $x, y, z, \ldots$ and they directly provide the values of the following n linear functions

$$a_ix + b_iy + c_iz + \ldots \ (i = 1, 2, \ldots, n).$$

Denote these values given by observations by $g_1, g_2, \ldots, g_n$. We will have

$$a_ix + b_iy + c_iz + \ldots = g_i + \varepsilon_i \ (i = 1, 2, \ldots, n)$$

where $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are the errors made.

Assuming that among the n linear functions under consideration there are at least m mutually independent ones ($n > m$), and that the approximate equalities

$$a_i x + b_i y + c_i z + \ldots = g_i \ (i = 1, 2, \ldots, n)$$

are free from *constant error*, *i.e.*, that the expectations of the *possible error* in each observation are equal to zero, we apply the MLSq to determine the unknowns $x, y, z, \ldots$

To this end, we denote the expectation of the square of the possible error of the i-th observation {cf. formula (4) below} by k/p_i, and, assuming that the *weights* of the observations $p_1, p_2, \ldots, p_n$ are known, we look for $x, y, z, \ldots$ under the condition that the expression

$$\sum_{i=1}^{n} p_i(a_i x + b_i y + c_i z + \ldots - g_i)^2$$

is minimum.

Let $\xi, \eta, \zeta, \ldots$ be the values of the unknowns thus determined. Then magnitudes

$$a_i\xi + b_i\eta + c_i\zeta + \ldots - g_i = \omega_i$$

will represent the *corrections of the observations* obtained by the MLSq.

Suppose that it is required to determine the magnitude k on which the *measure of precision* [1] $\sqrt{p_i/2k}$ of each observation depends. For this purpose Gauss [2] proposes to use the formula

$$k = \sum \frac{p_i\omega_i^2}{n-m} \qquad (1)$$

which, however, is not yet rigorously established. It is not therefore amiss to indicate the following proposition on which it might be founded: *Suppose that* λ_i *is the possible error of the i-th observation and* ρ_i *is the corresponding correction obtained by the MLSq. If, no matter how long the observations be continued* (supposing that it is possible to continue them indefinitely), *the ratio*

$$\mathrm{E}\lambda_i^4/(\mathrm{E}\lambda_i^2)^2$$

remains less than some limit, then the probability of the inequality

$$\left|\frac{\sum p_i\rho_i^2}{n} - k\right| < \alpha,$$

no matter how small the positive number α *would be, can be made arbitrarily close to* 1 *by making n sufficiently large.*

Denote

$$\sum p_i a_i\lambda_i = E_a, \sum p_i b_i\lambda_i = E_b, \sum p_i c_i\lambda_i = E_c, \text{ etc.}$$

Then, expressing the magnitudes ρ_i by means of $\lambda_1, \lambda_2, \ldots, \lambda_n$, we shall have, as it is known,

$$\sum p_i\,\rho_i^2 = \sum p\lambda_i^2 - \varphi(E_a; E_b; E_c; \ldots), \qquad (2)$$

where $\varphi(E_a; E_b; E_c; \ldots)$ stands for some *positive definite* quadratic form of variables E_a, E_b, E_c; … [3]

Hence we find that

$$\left|\frac{\sum p_i\rho_i^2}{n} - k\right| < \left|\frac{\sum p_i\lambda_i^2}{n} - k\right| + \varphi/n.$$

Consequently,

$$\mathrm{E}\left|\frac{\sum p_i\rho_i^2}{n} - k\right| < \mathrm{E}\left|\frac{\sum p_i\lambda_i^2}{n} - k\right| + \mathrm{E}\varphi/n.$$

However, it is known that

$$\mathrm{E}\varphi = mk.\text{[4]} \tag{3}$$

On the other hand, noting that

$$\mathrm{E}\lambda_i^2 = k/p_i \tag{4}$$

and supposing that $\mathrm{E}\lambda_i^4 = k^2 q_i/p_i^2$, we have [5]

$$\mathrm{E}\left|\frac{\sum p_i\lambda_i^2}{n} - k\right| < \left[\mathrm{E}\left(\frac{\sum p_i\lambda_i^2}{n} - k\right)^2\right]^{1/2}$$

or

$$\mathrm{E}\left|\frac{\sum p_i\lambda_i^2}{n} - k\right| < (k/\sqrt{n})\sqrt{(\sum q_i/n) - 1}\,.$$

We thus obtain the inequality

$$\mathrm{E}\left|\frac{\sum p_i\rho_i^2}{n} - k\right| < (k/\sqrt{n})\sqrt{(\sum q_i/n) - 1} + mk/n$$

from which, making use of the Chebyshev reasoning [6], our proposition is indeed deduced.

Suppose that

$$\mathrm{E}\lambda_i^4/(\mathrm{E}\lambda_i^2)^2 = q_i < L. \tag{5}$$

The obtained inequality provides

$$\mathrm{E}\left|\frac{\sum p_i\rho_i^2}{n} - k\right| < (k/\sqrt{n})\sqrt{L - 1} + mk/n.$$

Evidently, however,

$$\mathrm{E}\left|\frac{\sum p_i\rho_i^2}{n} - k\right| > (1 - P)\alpha \qquad (6)$$

where P is the probability of the inequality

$$\mathrm{E}\left|\frac{\sum p_i\rho_i^2}{n} - k\right| \le \alpha.$$

Therefore, we have

$$(1 - P)\alpha < (k/\sqrt{n})\sqrt{L-1} + mk/n$$

whence

$$P > 1 - [k/(\alpha\sqrt{n})]\,[\sqrt{L-1} + (m/\sqrt{n})] \qquad (7)$$

which indeed proves the theorem.

Basing our reasoning on this theorem, we may consider each expression of the type

$$\frac{\sum p_i\omega_i^2}{f(n)}$$

as an approximate expression for k. Here, $f(n)$ is any function of n for which

$$\lim\,[f(n)/n] = 1 \text{ as } n = \infty.$$

However, from among all such expressions there is only one that will be free from a constant error; namely that, for which

$$\mathrm{E}\frac{\sum p_i\rho_i^2}{f(n)} = k.$$

This expression is $f(n) = n - m$ which gives us

$$k = \frac{\sum p_i\omega_i^2}{n-m}. \qquad (8)$$

Notes and Commentary by Translator

1. According to Gauss [2, §§178 – 179], the *measure of precision*, h, is the parameter of the normal distribution

$$\varphi_1(x) = (h/\sqrt{\pi}) \exp(-h^2x^2).$$

In 1823 he did not retain this term and the normal law itself lost its uniqueness, and, furthermore, appeared, in §§ 7 and 9, in another form:

$$\varphi_2(x) = [1/(h\sqrt{\pi})] \exp(-x^2/h^2).$$

On the other hand, Gauss formally introduced weight only in [3, §7].

2. See [3, §§38 and 39].

3. In notation of the classical error theory formula (2) becomes

$$\sum p_i \rho_i^2 = \sum p\lambda_i^2 - \{[paa](x-\xi)^2 + [pbb](y-\eta)^2 + [pac](z-\zeta)^2 + 2[pab](x-\xi)(y-\eta) + 2[pac](x-\xi)(z-\zeta) + 2[pbc](y-\eta)(z-\zeta)\} \quad \text{(i)}$$

where, for example, $[pab]$ is the sum of terms $p_ia_ib_i$ and, without loss of generality, the number of the unknowns is three. Magnitudes $[paa]$, $[pab]$, …, $[pcc]$ are the coefficients of the appropriate system of normal equations.

4. My formula (i) leads to [6, pp. 81 – 89]

$$E(\sum p_i\lambda_i^2 - \sum p_i\rho_i^2) = k\{Q_{11}[paa] + Q_{12}[pab] + Q_{13}[pac] + Q_{21}[pab] + Q_{22}[pbb] + Q_{23}[pbc] + Q_{31}[pac] + Q_{32}[pbb] + Q_{33}[pcc]\}.$$

By definition of the magnitudes Q_{ij} the right side is equal to $3k$ (or, generally, to mk), hence Liapunov's formula (3).

5. The next line does not directly follow from the context, but it represents a particular case of Liapunov's inequalities [7, p. 112].

6. Liapunov apparently thought that *Chebyshev's reasoning* was his (or, rather, the Bienaymé – Chebyshev) inequality. Formula (6) below can be deduced at once by means of elementary transformations made while proving it [5, §32]. Let $P(\xi \le \alpha) = p$. Then

$$1 - p = P(\xi > \alpha) \le (1/\alpha) \int_{|x|>\alpha} |x|\psi(x)\,dx \le (1/\alpha)E|\xi| \quad \text{(ii)}$$

where $\psi(x)$ is the density law of the random variable ξ.

Seneta [8, p. 54] in essence noted that, more to the point, (ii) was the Markov inequality for positive random variables.

* * *

Liapunov possibly wrote his note at about the time when he was working on his main papers on probability theory; that is, somewhat before 1900, see Gnedenko [4]. Suppose that n observations $g_1, g_2, \ldots, g_n$ are made to determine m unknowns, $x, y, \ldots$ In the ideal case,

$$a_ix + b_iy + \ldots + g_i = 0, \; i = 1, 2, \ldots, n.$$

Residual free terms ω_i appearing after solving these equations by least squares can be used to estimate the precision of observations. In Liapunov's notation, see formula (1), the estimator of the squared mean square error of unit weight is [3, §38]

$$k = \frac{\sum p_i\omega_i^2}{n-m}.$$

Liapunov mistakenly thought that this formula was *not yet established rigorously* and set out to prove it anew, and to determine whether it was good enough as an estimator. He cited Gauss but had not mentioned any definite contribution so that possibly he did not then read the *Theoria combinationis* [3]. It is likely, however, that he acquainted himself with this memoir later (for example, on Markov's advice) and that, accordingly, he abstained from publishing his manuscript.

Liapunov undoubtedly came to know the Gauss formula from his teacher, Chebyshev, whose lectures on probability [1] were eventually published from notes taken by him, Liapunov. I [9] studied these lectures and indicated, in §5.3, that Chebyshev had not treated the Gauss formula adequately. Now I add that, consequently, Liapunov could have indeed inferred that it was not established rigorously. Cf. also Note 1 above.

What had Liapunov achieved? Yes, he proved the Gauss formula anew. And he also derived inequality (7) from which it followed that, for any law of error, the estimator (8) of the variance was consistent. This, however, was already evident from Gauss' findings [3, §40]. Seneta [8, p. 54] indicated this fact, but he cited Gauss' additional formula that only pertained to the normal distribution.

References

1. Chebyshev, P.L. (1936), *Теория вероятностей* (Theory of Probability). Lectures read 1879/1880 as written down by A.M. Liapunov. M. – L. Translation: Berlin, 2004.

2. Gauss, C.F. (1809), *Theoria motus*, etc. German translation of extract in Gauss, C.F. (1887), *Abhandlungen zur Methode der kleinsten Quadrate*. Hrsg. A. Boersch & P. Simon. Vaduz, 1998, pp. 92 – 117.

3. --- (1823), Theoria combinationis, etc. German translation: Ibidem, pp. 1 – 53.

4. Gnedenko, B.V. (1959), On the work of Liapunov in the theory of probability. IMI, vol. 12, 1959, pp. 135 – 160. Translated in Nekrasov, P.A. (2004), *Theory of Probability*. Berlin.

5. --- (1968), *Lehrbuch der Wahrscheinlichkeitsrechnung*. Berlin. Several previous Russian editions.

6. Helmert, F.R. (1872), *Die Ausgleichungsrechnung,* etc. Leipzig. Later editions: 1907, 1924.

7. Heyde, C.C., Seneta, E. (1977), *I.J. Bienaymé.* New York.

8. Seneta, E. (1984), The central limit problem and linear least squares in pre-revolutionary Russia: the background. *Math. Scientist*, vol. 9, pp. 37 – 77.

9. Sheynin, O. (1994), Chebyshev's lectures on the theory of probability. *Arch. Hist. Ex. Sci.*, vol. 46, pp. 321 – 340.

10. A.A. Markov. The Law of Large Numbers and the Method of Least Squares

Избранные труды (Sel. Works). N.p., 1951, pp. 233 – 251.

{This memoir consists of extracts from letters written to Professor A.V. Vasiliev}

23 September 1898

Chebyshev's memoir [2] is important. Regrettably, its importance is obscured by two circumstances, viz., by involved derivations and insufficient rigor of reasoning. The theorem, that Chebyshev proves there, has for a long time been considered true, but it was established by means of extremely non-rigorous methods. I do not say *proved*, since I do not admit non-rigorous demonstrations when perceiving no possibility of making them rigorous.

So, Chebyshev's aim was not to establish the theorem, even if non-rigorously, but to prove it. Its well-known derivation is not rigorous but easy. On the contrary, Chebyshev's demonstration is very involved since it is based on preliminary investigations. Therefore, its advantage over the previous proof can only consist in rigor of reasoning. However, Chebyshev's derivation is expounded in such a way that its rigor may be doubted. A question therefore arises, whether Chebyshev's proof is distinguished from the previous one not in essence, but only by needless complexity, or can it be made rigorous. Your essay on the works of Chebyshev [10] has strengthened my long-standing desire to simplify, and, at the same time, to make his analysis quite rigorous.

I begin with the **Theorem** on expectations that makes up the main subject of his memoir. It can be formulated thus: *If the expectations of the independent magnitudes*

$$x_1, x_2, \ldots, x_n \qquad (1)$$

are zero, and those of the power x_n^k *for any integral positive number k remain finite as* $n \to \infty$*, then, under the last-mentioned condition, each of the differences*

$$\mathrm{E}\left(\frac{x_1+x_2+\ldots+x_n}{\sqrt{n}}\right)^m - A_m\left[E\left(\frac{x_1+x_2+\ldots+x_n}{\sqrt{n}}\right)^2\right]^{m/2} \to 0 \qquad (2)$$

where m is an integral positive number and

$$A_m = (2^{m/2}/\sqrt{\pi})\int_{-\infty}^{\infty} t^m \exp(-t^2)dt.$$

I provide an elementary proof of this proposition [1].According to the well-known generalization of the Newton formula, we have

$$\left(\frac{x_1+x_2+\ldots+x_n}{\sqrt{n}}\right)^m = \sum \frac{m!}{\alpha_1!\alpha_2!\ldots\alpha_i!}\frac{S^{\alpha_1,\alpha_2,\ldots,\alpha_i}}{(\sqrt{n})^m}$$

where

$$\alpha_1, \alpha_2, \ldots, \alpha_i \qquad (3)$$

are integral positive numbers (not zeros) satisfying the condition

$$\alpha_1 + \alpha_2 + \ldots + \alpha_i = m \qquad (4)$$

and S represents a symmetric function of numbers (1) that can be determined by one of its terms,

$$x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_i^{\alpha_i}.$$

The number of the terms of this kind is not larger than

$$n(n-1)(n-2)\ldots(n-i+1).$$

It is equal to this product when all the numbers (3) are different, and less than it otherwise. Turning now to expectations, we find that the expectation of

$$\left(\frac{x_1+x_2+\ldots+x_n}{\sqrt{n}}\right)^m$$

is equal to the sum

$$\sum \frac{m!}{\alpha_1!\alpha_2!\ldots\alpha_i!}\mathrm{E}\frac{S^{\alpha_1,\alpha_2,\ldots,\alpha_i}}{(\sqrt{n})^m}.$$

Here, in accord with one of the conditions of the theorem, all such terms, in which any of the numbers (3) is equal to unity, vanish. Therefore, only such terms remain for which each of these numbers is not less than 2. Supposing therefore that

$\alpha_1 \ge 2, \alpha_2 \ge 2, \ldots, \alpha_i \ge 2,$

we conclude from equality (4) that the maximal value of i in the remaining terms will be $(m - 1)/2$ for odd values of m, and $m/2$ for its even values. It is also important to note that for an even m the number i only attains its maximal value $m/2$ if

$\alpha_1 = \alpha_2 = \ldots = \alpha_i = 2.$

Otherwise, $i < m/2$.

On the other hand, on the strength of one of the conditions of the theorem, the ratio

$ES/[n(n - 1)(n - 2) \ldots (n - i + 1)]$

should remain finite as $n \to \infty$. Therefore, for $i < m/2$ each of the ratios $ES/(\sqrt{n})^m$ should tend to zero as $n \to \infty$. We infer at once that, as $n \to \infty$, the expectation of the expression

$$\left(\frac{x_1 + x_2 + \ldots + x_n}{\sqrt{n}}\right)^m$$

should tend to zero if m is not even. Otherwise, once more, of course, as $n \to \infty$,

$$\mathrm{E}\left(\frac{x_1 + x_2 + \ldots + x_n}{\sqrt{n}}\right)^m - (m!/2^{m/2})\frac{ES^{2,2,\ldots,2}}{n^{m/2}} \to 0.$$

Let us compare now

$$\frac{ES^{2,2,\ldots,2}}{n^{m/2}} \text{ and } \left[E\left(\frac{x_1 + x_2 + \ldots + x_n}{\sqrt{n}}\right)^2\right]^{m/2}.$$

Denote for the sake of brevity $\mathrm{E}x_i^2 = c_i$, then

$$\mathrm{E}\left(\frac{x_1 + x_2 + \ldots + x_n}{\sqrt{n}}\right)^2 = \frac{c_1 + c_2 + \ldots + c_n}{n}.$$

As to the expectation of $S^{2,2,\ldots,2}$, it is equal to a symmetric function of the numbers

$$c_1, c_2, \ldots, c_n \qquad (5)$$

determined by one of its terms, $c_1 \cdot c_2 \ldots c_n$. Applying now the generalized Newton formula to the power

$$\left(\frac{c_1 + c_2 + \ldots + c_n}{n}\right)^{m/2},$$

we have

$$\left(\frac{c_1 + c_2 + \ldots + c_n}{n}\right)^{m/2} = \sum \frac{(m/2)!}{\mu_1!\mu_2!\ldots\mu_j!}\ \frac{C^{\mu_1,\mu_2,\ldots,\mu_j}}{n^{m/2}} \qquad (6)$$

where $\mu_1, \mu_2, \ldots, \mu_j$ are integral positive numbers (not zeros) obeying the condition

$$\mu_1 + \mu_2 + \ldots + \mu_j = m/2 \qquad (7)$$

and C denotes a symmetric function of numbers (5) that can be determined by one of its terms $c_1^{\mu_1} c_2^{\mu_2} \ldots c_j^{\mu_j}$.

Condition (7) shows that $j < m/2$ for all the expressions (8) under our consideration excepting one, $C^{1, 1, \ldots, 1}$, which coincides with $ES^{2, 2, \ldots, 2}$. Therefore, all the terms of the sum in (6) tend to zero as $n \to\infty$ except for one of them,

$$(m/2)![C^{1, \ldots, 1}/n^{m/2}] = (m/2)![ES^{2, \ldots, 2}/n^{m/2}].$$

It follows that the difference

$$\left(\frac{c_1 + c_2 + \ldots + c_n}{n}\right)^{m/2} - (m/2)!\ [ES^{2, \ldots, 2}/n^{m/2}]$$

tends to zero as $n \to\infty$. Comparing this result with the one obtained above, we infer that, as $n \to\infty$, the difference (2), where

$$A_m = \frac{m!}{2^{m/2}(m/2)!} = 1\cdot3\cdot5\ \ldots\ (m-1) = (2^{m/2}/\sqrt{\pi}) \int_{-\infty}^{\infty} t^m \exp(-t^2)dt$$

should tend to zero. Here, m is any given even positive number.

The theorem is thus proved for even values of m. For its odd values it was proved above because in this case the last integral vanishes and

$$\lim E\left(\frac{x_1 + x_2 + \ldots + x_n}{\sqrt{n}}\right)^m = 0 \text{ as } n \to\infty$$ [2].

5 October

To arrive, after all, to the well-known limiting probability expressed by the definite integral

$$(1/\sqrt{\pi}) \int_t^{t'} \exp(-x^2)\, dx \qquad (9)$$

we must transform the theorem of the first letter into the following

Theorem. *As $n \to\infty$, each difference*

$$E\left(\frac{x_1 + x_2 + \ldots + x_n}{\sqrt{2E(x_1 + x_2 + \ldots + x_n)^2}}\right)^m - (1/\sqrt{\pi}) \int_{-\infty}^{\infty} t^m \exp(-t^2)dt$$

tends to zero. This transformation does not present any difficulties but demands a new condition; namely, we must assume that

$$E(x_1 + x_2 + \ldots + x_n)^2$$

increases unboundedly with n and in such a manner that its ratio to n does not tend to zero; or, better, cannot be arbitrarily small. This condition also plays an important part in the long ago known derivation of an approximate expression of the probability [3].

I shall show, making use of a particular example, that, if this condition is violated, the integral (9) does not represent either the approximate value, or even the limit of probability. Let the values of x_k be $-1/N_k$ and 1 with probabilities $N_k/(N_k + 1)$ and $1/(N_k + 1)$ respectively. Let also

$$(1/N_1), (1/N_2), \ldots \tag{11}$$

be a convergent series of positive numbers whose sum is less than 1. Under these conditions it is easy to express the probability that

$$x_1 + x_2 + \ldots + x_n < 0. \tag{12}$$

Indeed, this sum will be negative if all its terms (1) are negative, and positive if any of its terms is positive, *i.e.*, equal to 1. Therefore, according to the multiplication theorem, the probability that (12) is fulfilled is expressed by the product

$$[N_1/(N_1 + 1)]\cdot[N_2/(N_2 + 1)] \ldots [N_n/(N_n + 1)] \tag{13}$$

equal to

$$\{1 - [1/((N_1 + 1)]\}\cdot\{1 - [1/((N_2 + 1)]\} \ldots \{1 - [1/((N_n + 1)]\} =$$
$$1/\{[1 + (1/N_1)]\,[1 + (1/N_2)] \ldots [1 + (1/N_n)]\}.$$

It follows that the probability sought is higher than

$$1 - [1/((N_1 + 1)]\cdot - [1/((N_2 + 1)] - \ldots - [1/((N_n + 1)]$$

but lower than

$$1/[1 + (1/N_1) + (1/N_2) + \ldots + (1/N_n)].$$

Until now, the numbers

$$N_1, N_2, \ldots, N_n \tag{14}$$

were restricted by one condition,

$$1 \geq (1/N_1) + (1/N_2) + \ldots$$

to which we may add the following one:

$$1/2 \geq [1/((N_1 + 1)]\cdot + [1/((N_2 + 1)] + \ldots$$

Thus, as $n \to \infty$, the probability of condition (12) being fulfilled approaches a limit larger than 1/2. On the contrary, this limit will be less than 1/2 once the numbers (14) are chosen in such a manner that

$$1 = (1/N_1) + (1/N_2) + \ldots$$

For example, if

$$N_k = (k + 1)(k + 2) - 1,$$

then

$$[1/(N_1 + 1)] + [1/(N_2 + 1)] + \ldots = [1/(2\cdot3)] + [1/(3\cdot4)] + [1/(4\cdot5)] + \ldots = 1/2$$

and the probability considered has as its limit a number larger than 1/2; if, however, $N_k = k(k + 1)$, the limit of the probability is less than 1/2.

This result shows that the Chebyshev theorem is not applicable to our example: according to the theorem, the limit of the probability for the condition (12) to be fulfilled must be equal to unity.

Turning now to the expectation (10) we see that it is equal to

$$(1/N_1) + (1/N_2) + \ldots + (1/N_n) \tag{15}$$

and remains finite as $n \to \infty$. This indeed is the reason why the Chebyshev theorem is not applicable here.

9 October

To form a better judgement on the situation, it seems useful to me to dwell in addition on a modification of my example. I have previously supposed that the positive and the negative values of each number x_k are neither equal in absolute value nor equally probable. However, it is usually assumed that observational errors only differing in their signs are equally probable, and I modify my example accordingly.

Suppose that the values of x_k are

$$1, (-1/N_k), (1/N_k), 1$$

with probabilities

$$1/[2(N_k+ 1)], N_k/[2(N_k+ 1)], N_k/[2(N_k+ 1)], 1/[2(N_k+ 1)]$$

respectively. Then the expectation of x_n^2 will still be

$$[1/(N_k+ 1)][1 + (1/(N_k)] = 1/N_k.$$

Denote the sum of the series (11), which we suppose is convergent, by $2t^2$ and consider the probability that the sum in (12) is contained within the boundaries

$$\pm t\{2[(1/N_1) + (1/N_2) + \ldots + (1/N_n)]\}^{1/2}.$$

It is easy to show that, once we appropriately choose the series of numbers $N_1, N_2, \ldots$, the integral·

$$(2/\sqrt{\pi})\int_0^t \exp(-x^2)\, dx \tag{9a}$$

will not be the limit of this probability.

Indeed, this probability is higher than that for the sum in (12) to be contained within the boundaries

$\pm [(1/N_1) + (1/N_2) + \ldots + (1/N_n)]$.

This latter probability is, however, higher than that for all numbers (1) to be either 1 or – 1; *i.e.*, it is larger than the product (13). After establishing this fact, choose

$N_1 = 2\cdot3$, $N_2 = 3\cdot4$, $N_3 = 4\cdot5$, etc.

The sum (15) will then be equal to 1/2; therefore, $t = 1/2$ and

$$(2/\sqrt{\pi})\int_0^t \exp(-x^2)\,dx = (2/\sqrt{\pi})\int_0^{1/2} \exp(-x^2)\,dx < 0.53.$$

The product (13), that represents a magnitude smaller than the probability considered, is larger than $[1 - (1/N_2) - (1/N_3) - \ldots]$, and therefore larger than

$6[1 - (1/3)]/7 = 0.57 \ldots$

It is thus obvious that the integral (9a) does not represent an approximate value of the probability and cannot be its limit as $n\to\infty$ [4, 5].

Now, in accord with the wish formulated in your letter, I turn to the MLSq. Without assuming any definite law of probability for the errors of the separate observations, it is possible to arrive at the method by issuing from the following propositions:

1) We consider only such approximate equalities, which, according to our assumptions, do not contain any constant error.

2) To each approximate equality we assign a certain weight and we suppose that the weights of the different equalities are inversely proportional to the expectations of the squares of the errors.

3) We estimate the merit of each approximate equality by its weight, and, accordingly, for each of the unknowns we determine such an equality whose weight is maximal.

To my mind, only this justification of the MLSq is rational; it was indicated by Gauss. I consider it rational mainly because it does not obscure the conjectural nature of the method. Keeping to this substantiation, we do not ascribe the ability of providing the most probable, or the most plausible results to the MLSq and only consider it as a general procedure furnishing approximate values of the unknowns along with a hypothetical estimation of the results obtained [6].

Astronomers like another substantiation, according to which the MLSq is presented as providing the most probable values of the unknowns. Since this {earlier} justification is also due to Gauss, I feel that it is useful to quote his own opinion formulated in his letter to Bessel as provided by Czuber [3, p. 289] [7].

It takes one to be very obstinate to persist in considering the most probable hypotheses, Gauss' words notwithstanding. I return to that justification which I believe to be the only rational one. According to my opinion, it provides everything needed for practice, but it does not furnish the probable error [8]. I believe that this doubtful magnitude is not required; if, however, it should be determined without fail, the well-known expression for the probability, $\int\exp(-x^2)dx$, should be assumed for each error separately, independently of whether it was derived from one or from several observations.

Here, we do not anymore need to determine the limit of the probability, or prove some inequalities concerning probabilities. However, it is possible to ask whether there exists a reasonable substantiation for admitting this formula for each observation separately. One of the relevant grounds known to me is that this admission allegedly tallies with practice, which is, however, difficult to ascertain. Another one consists in that the error of each observation might be supposedly considered as a totality of many errors independent of one another. Then, only a reference to the theorem on the limit of probability is needed to arrive at the desired formula; such a reasoning is to be found in Poincaré [8, p. 181]. However, the theorem on the limit of probability can only be ascertained with many restrictions; and, in addition, the idea that an error is a sum of many independent errors should be attributed, in my opinion, to the realms of fancy.

Let us now pass on to the derivation of the MLSq based on the theorem according to which the limit of probability is expressed by the well-known integral. Considering the well-known formula as a limiting expression, and not admitting it for a single observation, we have no right of admitting it either for 2, or 10, or 100, or, in general, for any given number of observations. Actually, however, we always treat a given number of observations, and we may only speak about the limit after artificially inventing their continuation. We are engaged not in a derivation for the limiting case, to which we come after continuing our observations until infinity, but by issuing from a finite number of observations. The theorem, according to which the limit of probability is expressed by a well-known integral, gives no indication of how far does the probability, in each given case, deviate from this integral. The question concerning the magnitude of these deviations remains open; they can be quite considerable, and, what is the main point, different in different cases. Therefore, when considering the well-known integral as the approximate value of the probability, and making it maximal, we are nevertheless unable to maintain that the probability itself will at the same time attain its maximal value. If, however, the value of the integral remains invariable, we cannot assert that the probability itself will also be such. The theorem on the limit of probability cannot therefore serve as a proof that the MLSq provides the most plausible results, *i.e.*, such, that maximal probability is attained for given limits of error, or narrowest limits of error are attained for a given probability.

Attempts at substantiating the MLSq by issuing from the theorem proved by Chebyshev [1] are also found in the Russian literature on probability theory. One such endeavour, very vague in its essence, might be perceived in Maievsky's well-known writing [4] [9]. There, in §§31 and 32, some kind of a *principle of arithmetic mean* is considered. In §31 the author busies himself with deriving it for the case of a very large number of observations without establishing its meaning. And only in §32 is it possible to conjecture that it is actually the well-known Gauss assumption that the arithmetic mean of the observed magnitude represents the most probable value of the measured unknown [10]. However, while considering the reasoning in §31, we ascertain that nothing of the sort follows from it since the author invariably discusses the arithmetic mean without comparing it with other means.

The conclusion in §31 is that, as the number of observations increases to ∞, the probability that the arithmetic mean of the observed magnitudes differs from the true value of the unknown less than by some given quantity, tends to unity. However, nothing follows from this proposition, since, as the number of observations increases to ∞, the same property also holds for other (linear) means. Finally, I know that some Russian mathematicians strove to derive the MLSq by means of a wrong application of the Chebyshev theorems which state that the probability is larger than a certain magnitude. Bearing in mind this proposition, they demanded that the probability should exceed a number assigned beforehand, and sought the narrowest boundaries of the error corresponding to a given inequality for the probability. They forgot that an uncountable set of different numbers obeys the same inequality and that the probability can exceed some given number even if the Chebyshev inequality does not reveal that fact.

Right now I am unable to recall any other justification of the MLSq. It seems that I had a look at everything, and I can only repeat, that, according to my opinion, the only rational substantiation is based on a direct study of the expectation of the square of the error. It is to this formulation that I adhere in my lectures.

Notes, partly by Yu.V. Linnik [6, pp. 654 – 655]

1. I discovered it long ago, but have not yet published it. M. [It advantageously differs from the not altogether correct one due to Chebyshev. L.]

2. A similar reasoning is in [8, pp. 168 – 186]. M.

3. Markov offered such a derivation in [5]. L.

4. An interesting but more complicated example can be found in the well-known book [9]. It is also possible to compile such examples in which the expectations of the squares represent divergent series. M.

5. The examples provided by Markov represent cases in which a condition, now called *maximal disregard*, is violated. At the end of his paper, Markov puts forward illustrations where it is not violated, but the expectations of the squares constitute divergent series. These examples were needed for the second part of the paper (for the letter dated 5 October 1898), where the author criticizes the substantiation of the normal law of observational errors by means of the *hypothesis of elementary errors*. L.

6. In simplest cases, the Markov principle is applied for the derivation of the well-known formulas of the MLSq in the following way. Suppose that linear equations

$$a_{i1}\alpha_1 + a_{i2}\alpha_2 + \ldots + a_{r1}\alpha_r,\ i = 1, 2, \ldots, n$$

are measured so as to determine the unknown magnitudes $\alpha_1, \alpha_2, \ldots, \alpha_r$.

Denoting the errors of observation by Δ_i, we obtain random variables y_i,

$$y_i = \Delta_i + a_{i1}\alpha_1 + a_{i2}\alpha_2 + \ldots + a_{r1}\alpha_r.$$

Let now $\mathrm{E}\Delta_i = 0$ and $\mathrm{var}\Delta_i = \sigma^2/p_i$ where p_i are the *weights* of the observations. Then we shall choose the approximate value of α_k as

$$\lambda_{1k}y_1 + \lambda_{2k}y_2 + \ldots + \lambda_{nk}y_n = T_k$$

under the conditions

1) $\mathrm{E}T_k = \alpha_k$ (absence of constant errors) and

2) $\mathrm{var}T_k = \min$.

7. {See English translation in [7, 1977, p. 287]. In this letter dated 1839, Gauss maintained that the maximal probability of an estimate of an unknown was less important than its least variance.}

8. The calculation of the *probable error* demands that the variance of the observations, rather than only their relative weights, be known. L.

9. I only mention it for the sake of definiteness. A similar reasoning can be found in other books as well. M.

10. A different sense is attached to this proposition in modern mathematical statistics (the method of maximum likelihood). It is assumed here that the observational errors are normally distributed and independent. Under these conditions it is possible to establish a certain link between the method indicated and the Markov principle. L.

References

1. Chebyshev, P.L. (1867), Des valeurs moyennes. *J. math. pures et appl.*, t. 12, pp. 177 – 184. Russian version appeared at the same time.

2. --- (1887, in Russian), Sur deux théorèmes relatifs aux probabilités. *Acta Math*., t. 14, 1891, pp. 305 – 315.
3. Czuber, E. (1891), *Theorie der Beobachtungsfehler.* Leipzig.
4. Maievsky, N. (1881), *Изложение способа наименьших квадратов* (Exposition of the Method of Least Squares, etc.). Psb.
5. Markov, A.A. (1898), Sur les racines de l'équation … *Izvestia* AN Psb, sér. 5, t. 9, No. 5, pp. 435 – 446.
6. --- (1951), *Избранные труды* (Sel. Works). N.p.
7. Plackett, R.L. (1972), Discovery of the method of least squares. *Biometrika*, vol. 59, pp. 239 – 251. Reprinted in *Studies Hist. Stat. Prob.*, vol. 2. Editors, Sir Maurice Kendall & R.L. Plackett. London, 1977, pp. 279 – 291.
8. Poincaré, H. (1896), *Calcul des probabilités.* Paris.
9. Poisson, S.-D. (1837), *Rechereches sur la probabilité des jugements,* etc. Paris.
10. Wassilief, A. (1900), P.L. Tschebyschef und seine wissenschaftlichen Leistungen. In collection of two articles *P.L. Tschebyschef* (1900). Leipzig, pp. 3 – 55. French version in *Boll. Bibl. e Storia Sci. Mat*., t. 1, 1898.

11. A.A. Markov. The Extension of the Law of Large Numbers onto Quantities Depending on Each Other

On the strength of the LLN it is possible to maintain, with probability however close to certainty, that the arithmetic mean of sufficiently numerous magnitudes will differ arbitrarily little from the arithmetic mean of their expectations. This law was derived by Chebyshev [1] who considered the expectation of the square of the difference between the sums of these magnitudes $\{x_i\}$ and of their expectations. Namely, his deliberations made it clear that the indicated law must persist in all cases in which the square of that difference {call it Ez^2}, as the number of the magnitudes $\{n\}$ increases unboundedly, increases slower than the square of that number, so that the ratio of these squares has zero as its limit $\{[\mathrm{E}z^2/n^2] \to 0\}$.

Chebyshev restricted his derivations with the simplest and therefore the most important case of independent magnitudes in which, as he showed, the expectation of the indicated square {Ez^2}, as the number of magnitudes $\{n\}$ increased unboundedly, can only increase as rapidly as this number if the expectations of the squares of these magnitudes themselves {Ex_i^2} remain finite rather than increase unboundedly.

To be sure, even if we restrict our attention to independent magnitudes, Chebyshev's conditions do not at all exhaust all the cases to which the abovementioned law may be applied. However, we do not aim at discovering the conditions necessary and sufficient for the applicability of the LLN; we only indicate that Chebyshev's derivations might also be extended onto some rather general cases when the magnitudes depend one on another.

1. Above all, we may consider the case in which the connection between the magnitudes is such that the increase in any one of them leads to a decrease in the expectations of the other ones. Let us dwell on this case. Suppose that we consider magnitudes

$$x_1, x_2, \dots, x_n, \dots \qquad (1.1)$$

with expectations

$$a_1, a_2, \dots, a_n, \dots$$

Denote for the sake of brevity $x_i - a_i = z_i$. When considering the expectation of the square

$$(z_1 + z_2 + + \ldots + z_n)^2 = \\ z_1^2 + z_2^2 + \ldots + z_n^2 + 2z_1z_2 + 2z_1z_3 + \ldots + 2z_{n-1}z_n \tag{1.2}$$

we shall apply the well-known proposition that the expectation of a sum is equal to the sum of the expectations of its terms.

For independent magnitudes the expectations of products such as $z_l z_k$ are all equal to zero and the expectation of the square (1.2) is reduced to the sum of the expectations of the squares of $z_1, z_2, \ldots, z_n$. However, in our case the expectation of each product $z_l z_k$ is a negative number so that

$$E(z_1 + z_2 + + \ldots + z_n)^2 < Ez_1^2 + Ez_2^2 + \ldots + Ez_n^2. \tag{1.3}$$

To convince ourselves in this fact we assume that the totality of numbers

$$z_l', z_l'', \ldots, z_l^{(\omega)}$$

arranged in an increasing order represent all the possible values of z_l with probabilities

$$p_l', p_l'', \ldots, p_l^{(\omega)}$$

respectively**.** Finally, suppose that, if

$$z_l = z_l', z_l'', \ldots, z_l^{(\omega)}, Ez_k = a_k', a_k,'' \ldots, a_k^{(\omega)}.$$

In this notation

$$Ez_l z_k = p_l' z_l' a_k' + p_l'' z_l'' a_k'' + \ldots + p_l^{(\omega)} z_l^{(\omega)} a_k^{(\omega)}$$

and the expectations of the magnitudes z_l and z_k themselves, equal to zero, can be represented as sums

$$p_l' z_l' + p_l'' z_l'' + \ldots + p_l^{(\omega)} z_l^{(\omega)} \text{ and } p_l' a_k' + p_l'' a_k'' + \ldots + p_l^{(\omega)} a_k^{(\omega)}.$$

And since in our case, in accord with the assumption made,

$$a'_k > a''_k > \ldots > a^{(\omega)}{}_k,$$

we have, on the strength of the well-known Chebyshev inequality [3; 5],

$$\sum p_l^{(i)} z_l^{(i)} a_k^{(i)} < \sum p_l^{(i)} z_l^{(i)} \sum p_l^{(i)} a_k^{(i)} = 0.$$

It follows that the LLN is applicable to our case if only the expectation of z_n^2 remains finite rather than increases unboundedly with n.

The same inequality (1.3), and, consequently, the same inference about the applicability of the LLN, can also be easily made in the case in which the expectation of x_k for any k decreases with the increase in the sum

$$x_1 + x_2 + \ldots + x_{k-1}.$$

To show this, it is sufficient to apply the identity

$(z_1 + z_2 + + \dots + z_n)^2 = z_1^2 + z_2^2 + \dots + z_n^2 +$
$2z_1z_2 + 2(z_1 + z_2)z_3 + 2(z_1 + z_2 + z_3)z_4 + \dots$

The indicated conditions are fulfilled, for example, in the case in which the sum

$$x_1 + x_2 + \dots + x_n \quad (1.4)$$

is equal to the number of white balls among n balls extracted from a container under the following conditions:

1) The container initially had $2a$ white balls and $2b$ balls of other colors;
2) The drawn balls are not returned;
3) When $(a + b)$ balls are left in the container, a white balls and b balls of other colors are added to them.

In this case, just as in the well-known instance in which the ratio of the number of white balls to all the balls in the container, $a/(a + b)$, remains invariable, we may state, on the strength of the LLN, with probability however close to certainty, that the ratio of the number of the appeared white balls to the number of all the drawn balls will differ from $a/(a + b)$ less than by any given magnitude.

2. Repeating that we only provide sufficient but not necessary conditions, we shall dwell now on one such case onto which Chebyshev's conclusions might be extended,– on the case in which the influence of the magnitudes (1.1) on each other rapidly decreases with the increase in the distance between them. In our case the sum (1.4) will represent the number of the occurrences of some event A in n consecutive trials connected in such a manner that at each trial the probability of A has a quite definite value p' if A had occurred at the immediately previous trial, and another definite value p'' otherwise, whichever were the results of the rest previous trials; if, however, the results of all the trials remain indefinite, the probability of A is the same and equal to p, for each one of them.

Proceeding to consider this case, we denote for the sake of brevity

$$1 - p = q, \; 1 - p' = q', \; 1 - p'' = q''$$

and we remark that the numbers p, p' and p'' are connected one with another by a simple equality

$$p = pp' + qp''$$

so that any two of these numbers being given, it is not difficult to calculate the third one; namely,

$$p = p''/(1 - p' + p''), \; p' = 1 + p'' - p''/p, \; p'' = p\,(1 - p')/(1 - p).$$

Since x_l and x_k equal to 0 or 1 are the numbers of the occurrences of the event A at trials l and k, it is easy to establish the equalities

$$a_l = a_k = p.$$

It is not difficult either to convince ourselves that the expectation of

$$z_l z_k = x_l x_k - a_l x_k - a_k x_l + a_k a_l$$

is reduced to $Ex_l x_k - p^2$.
As to $Ex_l x_k$, it is equal to the probability of the occurrence of the event A at both trials l and k, which, on the strength of the multiplication theorem, can in turn be represented by the product of p,– of the probability of the occurrence of the event at trial l,– by some number R_k^l, – by the probability of the occurrence of A at trial k when it is known that it occurred at trial l.

Having thus arrived at the equality

$$Ez_l z_k = p\,(R_k^l - p),$$

we ought to derive R_k^l, which only depends, as it is not difficult to convince ourselves, on the difference $(k - l)$ and can therefore be more simply denoted as R_{k-l}.

If $k - l = 1$, then, owing to our conditions, we have

$$R_1 = p'. \qquad (2.1)$$

Then, it is not difficult to calculate consecutively

$$R_2 = R_1 p' + (1 - R_1)p'' = p'p' + q'p'',\ R_3 = R_2 p' + (1 - p_2)p'', \ldots,$$
$$R_{m+1} = R_m p' + (1 - R_m)p'' = p'' + (p' - p'')R_m.$$

The equation

$$R_{m+1} = p'' + (p' - p'')R_m$$

belongs to those, for whose solution it is not difficult to indicate a general formula; namely, it leads to

$$R_m = p + C(p' - p'')^m$$

where the constant C is determined by the condition (2.1) and occurs to be equal to $1 - p = q$. And so,

$$Ez_l z_k = pq\,(p' - p'')^{k-l}$$

and, consequently,

$$E[(z_1 + z_2 + + \ldots + z_{k-1})\, z_k] = pq\,[(p' - p'') + (p' - p'')^2 + \ldots + (p' - p'')^{k-1}].$$

It follows that, if $p' < p''$, the left side of this equality is negative, so that

$$E(z_1 + z_2 + + \ldots + z_n)^2 < npq.$$

If, however, $p' > p''$, then

$$E[(z_1 + z_2 + + \ldots + z_{k-1})\, z_k] < \frac{pq\,(p' - p'')}{1 - p' + p''}$$

and therefore

$$\mathrm{E}(z_1 + z_2 + + \dots + z_n)^2 < npq\left(1+\frac{2(p'-p'')}{1-p'+p''}\right) < npq\frac{1+p'-p''}{1-p'+p''}.$$

We thus obtained inequalities which clearly show that the LLN is applicable to this case. We may therefore maintain, with probability however close to certainty, that, given a sufficiently large number of our trials, the ratio of the number of the occurrences of the event A to the number of the trials will differ from p less than by any given number.

3. The expression for the expectation of the product $z_l z_k$ found in §2 can be derived when issuing from general formulas with which we shall deal now and which can serve as a foundation for further investigations. Keeping to the assumptions of §2, denote the probability that the event A occurs exactly m times in the first k trials by $P_{m,k}$. We may suppose that

$$P_{m,k} = V_{m,k} + U_{m,k}$$

where both magnitudes on the right side are the same as $P_{m,k}$ subject, however, to the following conditions:

1) $U_{m,k}$ is the same as $P_{m,k}$ if A did not occur at the last trial;

2) Otherwise, $V_{m,k}$ is the same as $P_{m,k}$.

We shall now introduce three more functions of an arbitrary number ξ connected with magnitudes $P_{m,k}$, $V_{m,k}$, and $U_{m,k}$:

$$\Phi_k = U_{o,k} + U_{1,k}\,\xi + U_{2,k}\,\xi^2 + \dots + U_{k-1,k}\,\xi^{k-1}, \quad (3.1a)$$
$$\Psi_k = V_{1,k}\,\xi + V_{2,k}\,\xi^2 + \dots + V_{k,k}\,\xi^k, \quad (3.1b)$$
$$\Omega_k = P_{o,k} + P_{1,k}\,\xi + P_{2,k}\,\xi^2 + \dots + P_{k,k}\,\xi^k = \Psi_k + \Phi_k. \quad (3.1c)$$

And we shall show now that the function Ψ_k can be defined as the coefficient of t^k in the development of a rather simple expression in increasing powers of an arbitrary number t. To this aim, applying the addition and multiplication theorems, we establish two equalities

$$U_{m,k} = q'V_{m,k-1} + q''U_{m,k-1},\ V_{m,k} = p'V_{m-1,k-1} + p''U_{m-1,k-1}.$$

Making use of these equalities and returning to (3.1a) and (3.1b), we obtain two equalities

$$\Phi_k = q'\,\Psi_{k-1} + q''\Phi_{k-1},\ \Psi_k = p'\,\xi\Psi_{k-1} + p''\,\xi\Phi_{k-1}$$

and, finally, after eliminating one of the functions, Φ or Ψ, we arrive at

$$\Phi_{k+1} - (p'\,\xi + q'')\Phi_k + (p'-p'')\xi\Phi_{k-1} = 0,$$
$$\Psi_{k+1} - (p'\,\xi + q'')\Psi_k + (p'-p'')\xi\Psi_{k-1} = 0.$$

But, owing to (3.1c), the following equation must also hold:

$$\Omega_{k+1} - (p'\,\xi + q'')\Omega_k + (p'-p'')\xi\Omega_{k-1} = 0.$$

Consequently, Ω_k can be determined as the coefficient of t^k in the development of the fraction

$$\frac{A+Bt}{1-(p'\xi+q'')t+(p'-p'')\xi t^2},$$

where A and B do not depend on t, in increasing powers of that magnitude. For determining A and B it remains to consider Ω_k at two values of k; let $k = 1$ and 2, then we easily derive

$$\Omega_1 = q + p\xi,\ \Omega_2 = qq'' + (qp'' + pq') + pp'\xi^2$$

so that $\Omega_o = 1$.

And, after developing our fraction in increasing powers of t and restricting the development to the two first terms, we find that

$$A + (B + p'\xi + q'')t = \Omega_o + \Omega_1 t = 1 + (q + p\xi)t$$

so that $A = 1$, $B = (p - p')\xi + q - q''$.

It is not difficult to express the second of these equalities as

$$B = (p'' - p')\cdot(q\xi + p).$$

And so, finally,

$$\frac{1+(p''-p')(q\xi+p)t}{1-(p'\xi+q'')t+(p'-p'')\xi t^2} = \Omega_o + \Omega_1 t + \Omega_2 t^{\,2} + \ldots$$

This formula can serve as a starting point for further investigations, see my paper [6]. In particular, it is not difficult to apply it for deriving the results of the previous section, but I shall not dwell on this point.

4. To throw light on our issue, I adduce an example in which the LLN is not applicable. Suppose that, as in §1, the sum

$$x_1 + x_2 + \ldots + x_n$$

is equal to the number of white balls from among n balls drawn from a container one after another. Let, however, the conditions for the drawings be different; we shall fix them as follows:

1) The initial number of white balls in the container was a; and of those of other colors, b.

2) Each drawn ball is returned back together with another one of the same color.

In this case, the increase in the sum

$$x_1 + x_2 + \ldots + x_{k-1}$$

leads without fail to the increase in the expectation of x_k, so that the expectation of

$$(z_1 + z_2 + \ldots + z_{k-1})z_k > 0.$$

Having in mind the calculation of the expectation of

$$(z_1 + z_2 + \ldots + z_n)^2,$$

we note that it can be expressed by the sum

$$\sum_{m=0}^{n} \{m - [na/(a+b)]\}^2 P_{m,n,a,b} \qquad (4.1)$$

where P is the probability that m white balls will occur among n of the drawn ones. It is not difficult to convince ourselves that

$$P_{m,n,a,b} = C_n^{\,m} \frac{C_{a+m-1}^{\;m}\, C_{b+n-m-1}^{\;n-m}}{C_{a+b+n-1}^{\;n}}.$$

The following simple equalities result:

$$mP_{m,n,a,b} = [na/(a+b)]P_{m-1,n-1,a+1,b},$$

$$m(m-1)P_{m,n,a,b} = \frac{n(n-1)a(a+1)}{(a+b)(a+b+1)}P_{m-2,n-2,a+2,b}.$$

They lead to

$$\sum_{m=0}^{n} mP_{m,n,a,b} = [na/(a+b)], \sum_{m=0}^{n} m(m-1)P_{m,n,a,b} = \frac{n(n-1)a(a+1)}{(a+b)(a+b+1)}.$$

Together with an obvious equality

$$\sum_{m=0}^{n} P_{m,n,a,b} = 1$$

they make it possible to determine the sought sum (4.1). To prove this we decompose that sum into three parts,

$$\sum_{m=0}^{n} m(m-1)P_{m,n,a,b} - \{[2na/(a+b)] - 1\} \sum_{m=0}^{n} mP_{m,n,a,b} +$$

$$n^2[a/(a+b)]^2 \sum_{m=0}^{n} P_{m,n,a,b}.$$

It remains to make simple calculations after which we obtain

$$\sum_{m=0}^{n} \{m - [na/(a+b)]\}^2 P_{m,n,a,b} = \frac{nab(n+a+b)}{(a+b)^2(a+b+1)} \qquad (4.2)$$ [1].

Issuing from this result, we can show that the LLN is not here applicable, so that, given a sufficiently small ε, the probability of the inequalities

$$\varepsilon \le (m/n) - [a/(a+b)] \le \varepsilon$$

cannot be arbitrarily close to 1 no matter how large is n.

Indeed, let us denote the probability that these inequalities do not hold by β and take into account that

$$\{(m/n) - [a/(a+b)]\}^2$$

cannot be larger than ξ, the larger of the numbers $[a/(a+b)]^2$ and $[b/(a+b)]^2$. Issuing from the derived formula it is not difficult to ascertain the inequality

$$\xi\beta > \frac{ab}{(a+b)^2(a+b+1)} - \xi^2,$$

which shows that, if only $\xi < \sqrt{ab/[(a+b)^2(a+b+1)]}$, β cannot be arbitrarily small.

It is interesting to note that in our example the most probable value of *m*, which we shall denote by μ, is determined by simple inequalities

$$(a-1)n - b + 1 \le (a + b - 2)\mu,\ (a-1)n + a - 1 \le (a + b - 2)\mu.$$

In addition, the ratio μ/n, when *n* unboundedly increases, has as its limit

not $a/(a + b)$, but $(a - 1)/(a + b - 2)$.

Nevertheless, we naturally cannot expect that, for large values of *n*, the ratio *m/n* will be arbitrarily close to $(a - 1)/(a + b - 2)$ since the expectation of $[(m/n) - g]^2$ takes its minimal value at $g = a/(a + b)$, which, as is seen from the formula derived by us, tends, as $n \to \infty$, to

$$\frac{ab}{(a+b)^2(a+b+1)} \ne 0.$$

In the simplest case, in which $a = b = 1$, the LLN obviously cannot be applied; here, all the suppositions $m = 0, 1, 2, \dots, n$ have the same probability $1/(n + 1)$.

16 Jan. 1907

5. It is possible to generalize considerably the conclusions of §2. Namely, instead of the number of the occurrences of some event we may consider the sum of magnitudes connected into a chain in such a manner that, when one of them takes a definite value, the next ones become independent of those preceding it. Let

$$x_1, x_2, \dots, x_k, x_{k+1}, \dots \qquad (5.3a), (5.3b), \text{etc}$$

be an infinite series of magnitudes connected in such a way that, for any *k*, once x_k is known, x_{k+1} does not depend on $x_1, x_2, \dots, x_{k-1}$.

Let also the totality of numbers

$$\alpha, \beta, \gamma, \dots \qquad (5.2)$$

represent all the possible values of any of our magnitudes and the system

$$p_{\alpha\alpha}, p_{\alpha\beta}, p_{\alpha\gamma}, \dots, p_{\beta\alpha}, p_{\beta\beta}, p_{\beta\gamma}, \dots, \text{etc} \qquad (5.3)$$

be the probabilities for x_{k+1} to have a definite value when x_k is given. The first subscript indicates the value of x_k, and the second one, the assumed value of x_{k+1}; for example, if $x_k = \beta$, the probability of $x_{k+1} = \gamma$ will be $p_{\beta\gamma}$.

To prevent too complicated notation and reasoning, we assume that these probabilities do not depend on *k*. Let finally

$$p_\alpha^{(k)}, p_\beta^{(k)}, p_\gamma^{(k)}, \dots \qquad (5.4)$$

be the probabilities that x_k takes the values (5.2) as long as all the magnitudes $x_1, x_2, x_3, \dots$ remain indefinite.

The numbers (5.3) should certainly be positive; in addition, they should satisfy equalities

$$p_{\alpha\,\alpha} + p_{\alpha\,\beta} + p_{\alpha\,\gamma} + \ldots = 1,\ p_{\beta\,\alpha} + p_{\beta\,\beta} + p_{\beta\,\gamma} + \ldots = 1, \text{ etc},$$

but they are not restricted otherwise. As to the numbers (5.4), they should of course satisfy the condition

$$p_\alpha^{(k)} + p_\beta^{(k)} + p_\gamma^{(k)} + \ldots = 1$$

and they cannot be assigned independently of all the values of k. On the contrary, issuing from their values for $k = 1$, it is possible to calculate all of them consecutively by means of simple formulas

$$p_\alpha^{(k+1)} = p_{\alpha\,\alpha}\, p_\alpha^{(k)} + p_{\beta\,\alpha}\, p_\beta^{(k)} + \ldots,\ p_\beta^{(k+1)} = p_{\alpha\,\beta}\, p_\alpha^{(k)} + p_{\beta\,\beta}\, p_\beta^{(k)} + \ldots, \text{ etc.}$$

Turning now to expectations, we denote, as previously, $\mathrm{E}x_k = a_k$ as long as all the magnitudes (5.1) remain indefinite, and let

$$A_\alpha^{(i)},\ A_\beta^{(i)},\ A_\gamma^{(i)},\ \ldots \qquad (5.5)$$

be the expectations of x_{k+i} when $x_k = \alpha, \beta, \gamma, \ldots$ It is now not difficult to establish the following equalities:

$$a_k = p_\alpha^{(k)}\,\alpha + p_\beta^{(k)}\,\beta + p_\gamma^{(k)}\,\gamma + \ldots,\ \ a_{k+i} = p_\alpha^{(k)}\, A_\alpha^{(i)} + p_\beta^{(k)}\, A_\beta^{(i)} + p_\gamma^{(k)}\, A_\gamma^{(i)} + \ldots,$$

$$A_\alpha^{(i)} = p_{\alpha\,\alpha} A_\alpha^{(i-1)} + p_{\alpha\,\beta}\, A_\beta^{(i-1)} + p_{\alpha\,\gamma}\, A_\gamma^{(i-1)} + \ldots,$$
$$A_\beta^{(i)} = p_{\beta\,\alpha}\, A_\alpha^{(i-1)} + p_{\beta\,\beta}\, A_\beta^{(i-1)} + p_{\beta\,\gamma}\, A_\gamma^{(i-1)} + \ldots, \text{ etc.}$$

Issuing from these equalities, we shall prove that all the magnitudes

$$a_{k+i},\ A_\alpha^{(i)},\ A_\beta^{(i)},\ A_\gamma^{(i)},\ \ldots \qquad (5.6)$$

tend to one and the same limit as i increases unboundedly. To attain this aim, we note, first of all, that, on the strength of the equalities above, the number a_{k+i} is situated between the largest and the smallest numbers from among (5.5) all of which, in turn, are situated between the largest and the smallest numbers from among

$$A_\alpha^{(i-1)},\ A_\beta^{(i-1)},\ A_\gamma^{(i-1)},\ \ldots \qquad (5.7)$$

Again issuing from the same equalities, we derive the difference

$$A_\alpha^{(i)} - A_\beta^{(i)} = (p_{\alpha\,\alpha} - p_{\beta\,\alpha})\, A_\alpha^{(i-1)} + (p_{\alpha\,\beta} - p_{\beta\,\beta})\, A_\beta^{(i-1)} + \ldots$$

The sum

$$(p_{\alpha\,\alpha} - p_{\beta\,\alpha}) + (p_{\alpha\,\beta} - p_{\beta\,\beta}) + \ldots = 0.$$

Therefore, separating the positive and the negative numbers of the system

$$(p_{\alpha\,\alpha} - p_{\beta\,\alpha}),\ (p_{\alpha\,\beta} - p_{\beta\,\beta}),\ \ldots \qquad (5.8)$$

from each other, and replacing the latter numbers by their absolute values, we will get two totalities of positive numbers constituting equal sums.

It is also important to note that these sums are less than unity because their terms are less than the numbers in (5.3a) and (5.3b), respectively. Therefore replacing some of the numbers (5.7) by their maximal number, and the others, by their minimal number, and denoting by $\Delta^{(i-1)}$ the difference between the two last-mentioned numbers, we can establish the inequality

$$|A_\alpha^{(i)} - A_\beta^{(i)}| < h\Delta^{(i-1)}$$

where h is the sum of all the positive numbers in system (5.8) and is {therefore} less than unity.

Absolutely the same conclusion can be reached concerning the difference of any of the two numbers taken from the system (5.5). Therefore, considering $\Delta^{(i)}$, the difference between the maximal and the minimal from among the numbers of that system, instead of $A_\alpha^{(i)} - A_\beta^{(i)}$, we can obtain the inequality

$$\Delta^{(i)} < H\Delta^{(i-1)}$$

where H is some constant number situated between 0 and 1. This inequality shows that $\Delta^{(i)}$ tends to zero as i increases unboundedly.

It follows that, as $i \to \infty$, all the magnitudes (5.6) tend to one and the same limit differing from it less than by $\Delta^{(i)}$. And we also have

$$\Delta^{(i)} < CH^{i}$$

where C and H are constant numbers and $0 < H < 1$.

Turning now to the expectation of

$$[(x_1 - a_1) + (x_2 - a_2) + \ldots + (x_n - a_n)]^2, \tag{5.9}$$

we shall decompose this square into the same terms as in §2 and introduce $z_k = x_k - a_k$. And, on the strength of the proven, reasoning absolutely in the same way as we did in §2, we easily arrive at inequalities

$$E[z_k(z_1 + z_2 + \ldots + z_{k-1})] < D(H + H^2 + \ldots + H^{k-1}),$$
$$E[(x_1 - a_1) + (x_2 - a_2) + \ldots + (x_n - a_n)]^2 < Gn$$

where D and G are constant numbers.

On the other hand, when comparing the expectations of (5.9) and of

$$(x_1 + x_2 + \ldots + x_n - na)^2,$$

where $a = \lim a_{k+i}$, $i = \infty$, we find that their difference is equal to

$$[(a_1 - a) + (a_2 - a) + \ldots + (a_n - a)]^2$$

and persists when n increases unboundedly. Consequently, under this condition, the expectation of

$$\{[(x_1 + x_2 + \ldots + x_n)/n] - a\}^2$$

must tend to zero.

It immediately follows that the LLN is applicable to this case: no matter how small are the positive numbers ε and η, the probability that the inequalities

$$-\varepsilon < [(x_1 + x_2 + \ldots + x_n)/n] - a < \varepsilon$$

take place is higher than $(1 - \eta)$ for all sufficiently large values of n.

And so, independence of magnitudes is not a necessary condition for the existence of the LLN.

25 March 1907 [2]

Notes and Commentary by N.A. Sapogov [6, pp. 660 – 662]

1.{As first indicated by Chuprov, this equality can be derived from one of Bohlmann's formulas, see Ondar (1977, pp. 13 – 15) who actually provided this derivation. Moreover, the translators of this source noted that Markov's example was a special case of Pólya's urn scheme (Eggenberger & Pólya 1923). Also see Feller (1957, §5.2) with a reference to Friedman (1949).}
2. The date of publication of the memoir is given as 1906. S.

Markov considers the conditions for the applicability of the LLN to sums

$$S_n = X_1 + X_2 + \ldots + X_n$$

of dependent random variables X_i, and he only discusses sufficient conditions. His results are based on a simple but important remark that the law is applicable to S_n if

$$(\mathrm{var}S_n/n^2) \to 0 \text{ as } n \to \infty.$$

In the beginning of his work Markov applies a *Chebyshev inequality*, which, in the needed form, is formulated thus:

$$z_l^{(1)} \le z_l^{(2)} \le \ldots \le z_l^{(\omega)};\ a_k^{(1)} \ge a_k^{(2)} \ge \ldots \ge a_k^{(\omega)};$$
$$p_l^{(i)} \ge 0,\ i = 1, 2, \ldots, \omega,\ p_l^{(1)} + p_l^{(2)} + \ldots + p_l^{(\omega)} = 1;$$

$$\sum_i p_l^{(i)} z_l^{(i)} a_k^{(i)} \le \sum_i p_l^{(i)} z_l^{(i)} \sum_i p_l^{(i)} a_k^{(i)}. \quad (1)$$

Supposing that $\omega > 1$, we convince ourselves that this inequalityis valid in the following way. Obviously,

$$a_k^{(1)} \ge \sum_{i=1}^{\omega} p_l^{(i)} a_k^{(i)} \ge a_k^{(\omega)}$$

so that there exists such a subscript i_o, $1 \le i_o \le \omega$, for which

$$a_k^{(i)} - \sum_{i=1}^{\omega} p_l^{(i)} a_k^{(i)} \ge 0,\ i \le i_o,\ a_k^{(i)} - \sum_{i=1}^{\omega} p_l^{(i)} a_k^{(i)} < 0,\ i > i_o.$$

Then

$$\sum_{i=1}^{\omega} \{p_l^{(i)} z_l^{(i)}[a_k^{(i)} - \sum_{i=1}^{\omega} p_l^{(i)} a_k^{(i)}]\} = \left(\sum_{i=1}^{i_0} + \sum_{i=i_0+1}^{\omega}\right)\{\text{the same}\} \le$$
$$z_l^{(i_0)} \sum_{i=1}^{\omega} \{p_l^{(i)}[a_k^{(i)} - \sum_{i=1}^{\omega} p_l^{(i)} a_k^{(i)}]\} = 0,$$

hence (1).
One of Markov's examples of dependent magnitudes to which the LLN was applicable was that of a simple homogeneous chain; a chain was indeed considered for the first time in this work. In §5 he studied such a chain under the condition that all transition probabilities $p_{\alpha\beta}$ were positive. Note that this strong restriction was not necessary. Markov's reasoning and all of his conclusions would have persisted if only he had demanded that there existed at least one such subscript, ν, that, for any μ, $p_{\mu\nu} > 0$.
From among later works devoted to the LLN for dependent magnitudes I mention [1; 11].

References

1. Bernstein, S.N. (1918), On the law of large numbers. *Собрание сочинений* (Coll. Works), vol. 4. N.p., 1964, pp. 61 – 65. (R)

2. Chebyshev, P.L. (1867), Des valeurs moyennes. *J. math. pures et appl.*, t. 12, pp. 177 – 184.
3. --- (1882), On approximate expression of integrals through other integrals having the same limits. *Полное собрание сочинений* (Complete Works), vol. 3. M. – L., 1948, pp. 128 – 131.
4. Eggenberger, G., Pólya, G. (1923), Über die Statistik verketteter Vorgänge. *Z. angew. Math. Mech.*, Bd. 3, pp. 279 – 289.
5. Korkine, A.N. (1883), Sur un théorème de Tschébycheff. *C.r. Acad. Sci. Paris*, t. 96, pp. 326 – 327.
6. Markov, A.A. (1907, in Russian), Recherches sur un cas remarquable d'épreuves dépendantes. *Acta math.*, t. 33, 1910, pp. 87 – 104.
7. --- (1951), *Избранные труды* (Sel. Works). N.p.
8. Ondar, Kh.O., Editor (1977, in Russian), *Correspondence between Markov and Chuprov*. Translated by C. & M. Stein. New York, 1981.
9. Sapogov, N.A. (1950), On the law of large numbers for dependent magnitudes. *Izvestia Akad. Nauk SSSR*, ser. Math., vol. 14, No. 2, pp. 145 – 154.

12. A.A. Markov. The Extension of the Limit Theorems of the Calculus of Probability onto a Sum of Magnitudes Connected into a Chain

I have shown [3] that the well-known LLN, established by Chebyshev for independent quantities, extends onto many cases of dependent quantities. I think that special attention among these instances deserve magnitudes connected into a chain in such a manner, that, when the value of one of it gets known, subsequent magnitudes become independent from those preceding it.

When considering one such case, a simplest case as it might be said, I [4] proved for it a theorem on the limit of expectation from which followed the limit expression of the probability in the form of the well-known Laplace integral. This result suggests that the theorem must also take place for other cases of quantities connected into a chain.

My proof was based on a peculiarity of the given instance, viz., on the symmetry of the expressions sought with respect to p and q; small changes are, however, sufficient for making the demonstration independent from the indicated feature so that the possibility of extending such deliberations and the appropriate conclusions onto other cases becomes clear. Without returning to the studied instance, we shall consider another, more general case and thus ascertain the proof of a general nature.

1. Let

$$x_1, x_2, \ldots, x_k, x_{k+1}, \ldots \qquad (1.1)$$

be a series of quantities connected into a chain. Suppose also that only three values,

$$-1, 0, 1 \qquad (1.2)$$

are possible for each of them, and that the first, the second, and the third triplet of the system of numbers

$$p, q, r; p_1, q_1, r_1; p_2, q_2, r_2$$

represent the probabilities of $x_{k+1} = -1$, 0, and 1 when $x_k = -1$, 0, and 1 respectively. Of course, these probabilities cannot be negative and must satisfy the conditions

$$p + q + r = 1, p_1 + q_1 + r_1 = 1, p_2 + q_2 + r_2 = 1. \qquad (1.3)$$

We shall suppose, in addition, that none of them is equal to unity; it is especially important to assume this with respect to p, q_1, and r_2 so that our series (1.1) will not be a simple repetition of the same number. Finally, we must introduce three more numbers,

$$P,\ Q,\ R, \tag{1.4}$$

the probabilities of x_1 taking values (1.2) respectively.

Under these conditions we shall consider the probabilities of various assumptions about the magnitude of the sum

$$x_1 + x_2 + \ldots + x_n \tag{1.5}$$

for any given n, beginning with small values of this parameter, then going over to its large values and to $n \to \infty$. If $n = 1$, our sum is reduced to one number, x_1, and, accordingly, it can only take three values (1.2) with probabilities (1.4).

For $n = 2$ we have a sum of two numbers, $x_1 + x_2$, for which, as is not difficult to see, five values

$$-2, -1, 0, 1, 2$$

with probabilities

$$Pp,\ Pq + Qp_1,\ Pr + Qq_1 + Rp_2,\ Qr_1 + Rr_2,\ Rr_2,$$

respectively are possible. For the sum $(x_1 + x_2 + x_3)$ there are already seven values […] with probabilities […] respectively.

Increasing now the number n consecutively by unity, and acting as previously in the article mentioned [2], we represent the probability of the equality

$$x_1 + x_2 + \ldots + x_n = m \tag{1.6}$$

for any given values of n and m as a sum $(\overline{P}_{m,n} + P_{m,n} + \hat{P}_{m,n})$ of the probabilities

$$\vec{P}_{m,n},\ P_{m,n},\ \hat{P}_{m,n} \tag{1.7}$$

of the same equality taken under an additional condition of $x_n = -1, 0, 1$. Applying this notation, it is not difficult to establish the following equations

$$\overline{P}_{m,n+1} = p\overline{P}_{m+1,n} + p_1 P_{m+1,n} + p_2\hat{P}_{m+1,n};\ P_{m,n+1} = q\overline{P}_{m,n} + q_1 P_{m,n} + q_2\hat{P}_{m,n};$$
$$\hat{P}_{m,n+1} = r\overline{P}_{m-1,n} + r_1 P_{m-1,n} + r_2\hat{P}_{m-1,n} \tag{1.8}$$

where $n = 1, 2, 3, \ldots$

We can make use of these equations for consecutively calculating the probabilities (1.7) at $n = 2, 3, 4, \ldots$; it is only necessary to take into account the main equalities

$$\overline{P}_{-1,1} = P,\ P_{0,1} = Q,\ \hat{P}_{1,1} = R \tag{1.9}$$

and to note that all the other expressions $\overline{P}_{m,1}, P_{m,1}, \hat{P}_{m,1}$, which are the probabilities of impossible events, must be zero. We thus find that

$$\overline{P}_{-2,2} = pP, \overline{P}_{-1,2} = p_1Q, \overline{P}_{0,2} = p_2R; P_{-1,2} = qP, P_{0,2} = q_1Q, P_{1,2} = q_2R;$$
$$\hat{P}_{0,2} = rP, \hat{P}_{1,2} = r_1Q, \hat{P}_{2,2} = r_2R$$

whereas the other expressions, $\overline{P}_{m,2}, P_{m,2}, \hat{P}_{m,2}$, are zero. We obtain then

$$\overline{P}_{-3,3} = ppP, \overline{P}_{-2,3} = pp_1Q + p_1qP, P_{-2,3} = pqP, \hat{P}_{-2,3} = 0, \ldots$$
$$\overline{P}_{-2,3} + P_{-2,3} + \hat{P}_{-2,3} = P(pq + qp_1) + Qp_1p, \text{ etc.}$$

Without considering the case in which

$$\begin{vmatrix} p & p_1 & p_2 \\ q & q_1 & q_2 \\ r & r_1 & r_2 \end{vmatrix} = 0,$$

we may introduce in addition three numbers, $\overline{P}_{0,0}, P_{0,0}, \hat{P}_{0,0}$, determining them by the equations

$$P = p\overline{P}_{0,0} + p_1P_{0,0} + p_2\hat{P}_{0,0}; Q = q\overline{P}_{0,0} + q_1P_{0,0} + q_2\hat{P}_{0,0}; \;\; R = r\overline{P}_{0,0} + r_1P_{0,0} + r_2\hat{P}_{0,0}$$

on whose strength we have $\overline{P}_{0,0} + P_{0,0} + \hat{P}_{0,0} = 1$.

At the same time, assigning a zero value to the symbols $\overline{P}_{m,0}, P_{m,0}, \hat{P}_{m,0}$, $m \neq 0$, we may extend the deduced equations onto the case $n = 0$ as well.

2. Introducing a supplementary variable t and its functions

$$\overline{\varphi}_n = \sum \overline{P}_{m,n}t^m, \varphi_n = \sum P_{m,n}t^m, \hat{\varphi}_n = \sum \hat{P}_{m,n}t^m, \Phi_n = \overline{\varphi}_n + \varphi_n + \hat{\varphi}_n \qquad (2.1)$$

we can, first, determine the probability of the equality (1.6) as the coefficient of t^m in the development of Φ_n in powers of t; and, second, derive

$$t\overline{\varphi}_{n+1} = p\overline{\varphi}_n + p_1\varphi_n + p_2\hat{\varphi}_n; \varphi_{n+1} = q\overline{\varphi}_n + q_1\varphi_n + q_2\hat{\varphi}_n;$$
$$(1/t)\hat{\varphi}_{n+1} = r\overline{\varphi}_n + r_1\varphi_n + r_2\hat{\varphi}_n \qquad (2.2)$$

from the equations (1.8).

From (2.2) follows one and the same linear finite equation with an unknown function Φ_n for all the functions $\overline{\varphi}_n, \varphi_n, \hat{\varphi}_n, \Phi_n$. We can rather simply represent it in a symbolic manner

$$\begin{vmatrix} p - t\Phi & p_1 & p_2 \\ q & q_1 - \Phi & q_2 \\ r & r_1 & r_2 - (1/t)\Phi \end{vmatrix} \Phi^n = 0 \qquad (2.3)$$

where, after performing the operations indicated, Φ^{n+3}, Φ^{n+2}, Φ^{n+1}, Φ^n, should be replaced by Φ_{n+3}, Φ_{n+2}, Φ_{n+1}, Φ_n and $n = 0, 1, 2, \ldots$

We note the symbolic nature of the equation (2.3) because it clearly reveals how our conclusions are extended onto other, more general cases in which x_k can take more than three different values. In an ordinary manner this equation is represented as

$$\Phi_{n+3} - A\Phi_{n+2} + B\Phi_{n+1} - D\Phi_n = 0$$

where

$$A = (p/t) + q_1 + r_2 t,\ B = [(pq_1 - p_1 q)/t] + pr_2 - p_2 r + (q_1 r_2 - q_2 r_1)t,$$
$$D = pq_1 r_2 - pq_2 r_1 + p_1 q_2 r - p_1 q r_2 + p_2 q r_1 - p_2 q_1 r. \qquad (2.4)$$

On the other hand, in accord with our notation we have $\Phi_o = 1$ and a direct study of the cases $n = 1$ and $n = 2$ provides

$$\Phi_1 = (P/t) + Q + Rt,\ \Phi_2 = (Pp/t^2) + [(Pq + Qp_1)/t] + Pr + Qq_1 + Rp_2 + (Qr_1 + Rq_2)t + Rr_2 t^2.$$

Issuing from these, we can determine, on the grounds of equation (2.3), all the other functions Φ_n. And it is not difficult to see that all of them can be determined as coefficients of z_n in the expansion of some rational function
$f(t; z)/F(t; z)$ in increasing powers of this new supplementary variable z. Here,

$$F(t; z) = -\begin{vmatrix} pz - t & p_1 z & p_2 z \\ qz & q_1 z - 1 & q_2 z \\ rz & r_1 z & r_2 z - (1/t) \end{vmatrix} = 1 - Az + Bz^2 - Dz^3 \qquad (2.5)$$

and $f(t; z)$ is an integral function of the second degree with respect to z; it is determined by the first three terms of the formula

$$f(t; z)/F(t; z) = \Phi_o + \Phi_1 z + \Phi_2 z^2 + \ldots + \Phi_n z^n + \ldots \qquad (2.6)$$

A simple multiplication of the series just above by $F(t; z)$ provides

$$f(t; z) = \Phi_o + (\Phi_1 - A\Phi_0)z + (\Phi_2 - A\Phi_1 + B\Phi_o)z^2.$$

For our purpose it is important to note that at $t = 1$ all functions Φ_n must reduce to unity and that consequently

$$f(1; z)/F(1; z) = [1/(1 - z)]. \qquad (2.7)$$

3. We shall apply the deduced formulas to calculate the expectations of various powers of the sum

$$x_1 + x_2 + \ldots + x_n - na \qquad (3.1)$$

and we shall select a in such a way that the expectation of the first power of this sum remains finite as $n \to \infty$.

In accord with the definition of the function Φ_n, its coefficient of any given power of t is the probability that the sum (1.5) is equal to the pertinent exponent; therefore, the probability that the sum (3.1) is equal to a given number can be determined by expanding the product

$$t^{-na}\Phi_n \qquad (3.2)$$

in powers of t. In this expansion, the probability sought is the coefficient of the power of t whose pertinent exponent is equal to the given number.

It is not difficult to conclude now that the expectation of

$$(x_1 + x_2 + \ldots + x_n - na), \qquad (3.3)$$

where i is any integral positive number, can be obtained by making use of the product (3.2) in the following way. Assume that $t = e^u$, compose the derivative $d^{\,i}[t^{-na}\Phi_n]/du^{\,i}$ and calculate it at $u = 0$. The result thus obtained will provide the expectation sought; the transfer from t to e^u was necessary for the exponents to persist under differentiation.

At the same time, it is not difficult to see that the series

$$\Phi_o + t^{-a}\Phi_1 z + t^{-2a}\Phi_2 z^2 + \ldots + t^{-na}\Phi_n z^n + \ldots$$

is obtained from the series in (2.6) after replacing z by the product $t^{-a}z$. It follows, on the grounds of formula (2.6), that

$$f(t; zt^{-a})/F(t; zt^{-a}) = \Phi_o + t^{-a}\Phi_1 z + t^{-2a}\Phi_2 z^2 + \ldots \qquad (3.4)$$

and the expectation of (3.3) might be determined as the coefficient of z^n in the expansion of

$$d^{\,i}\{[f(e^u; ze^{-au})/F(e^u; ze^{-au})]/du^{\,i}\}_{u=0} \qquad (3.5)$$

in increasing powers of z.

4. Let us dwell now on the expectation of the first power of (3.1) so as to determine a. According to our conclusion, this expectation is expressed by the coefficient of z^n in the development of

$$\frac{f'_{u=0}(e^u; ze^{-au})}{F(1;z)} - \frac{f(1;z)}{F(1;z)} \cdot \frac{F'_{u=0}(e^u; ze^{-au})}{F(1;z)} \qquad (4.1)$$

in increasing powers of z. Symbols $f'_{u=0}$ and $F'_{u=0}$ denote the values of derivatives at $u = 0$.

In order to study the expansion of the indicated function in increasing powers of z we shall first of all consider its development into partial fractions; for our purpose it is important to isolate only one fraction with denominator $(1 - z)^l$. Obviously, only simple factors of the integral function $F(1; z)$ can be the denominators of the partial fractions sought.

One of its factors is $(1 - z)$ since its value at $z = 1$ is represented by the determinant

$$-\begin{vmatrix} p-1 & p_1 & p_2 \\ q & q_1-1 & q_2 \\ r & r_1 & r_2-1 \end{vmatrix},$$

which, on the strength of the main conditions (1.3), is zero. Turning now to the other simple factors of the function $F(1; z)$, we suppose that

$$F(1; z) = (1 - z)\cdot(1 - y_1 z)\cdot(1 - y_2 z). \tag{4.2}$$

The numbers 1, y_1 and y_2 are the three roots of the equation

$$\begin{vmatrix} p-y & p_1 & p_2 \\ q & q_1-y & q_2 \\ r & r_1 & r_2-y \end{vmatrix} = 0. \tag{4.3}$$

We shall prove that its root $y = 1$ is simple, and that the moduli of the other roots are less than unity. For this purpose, we compose the derivative with respect to y of the left side of the equation (4.3) and assume that $y = 1$. The result obtained can be represented as a sum of three differences

$$[q_2r_1 - (1 - q_1)\cdot(1 - r_2)] + [p_2r - (1 - p)\cdot(1 - r_2)] + [p_1q - (1 - p)\cdot(1 - q_1)]$$

none of which can be positive and which can only vanish all at the same time in the cases excluded by us, when at least one of the numbers p, q_1, r_2 is unity.

Ascertaining thus that unity is a simple rather than a multiple root of the equation (4.3), we go over to the other roots, y_1 and y_2, of the same equation which differ from unity. Let y be one of these other roots, it does not matter which. It is possible to select a system of numbers α, β, γ, not all of them being zeros, satisfying the equations

$$\alpha y = p\alpha + q\beta + r\gamma,\ \beta y = p_1\alpha + q_1\beta + r_1\gamma,\ \gamma y = p_2\alpha + q_2\beta + r_2\gamma. \tag{4.4}$$

All the differences $(\alpha - \beta)$, $(\alpha - \gamma)$, $(\beta - \gamma)$ cannot vanish either, otherwise the equations (4.4) lead to $y = 1$.

Consider the difference having the maximal modulus; let it be $(\alpha - \beta)$. Accordingly, subtracting βy from αy provides us with the following equality

$$(\alpha - \beta)y = (p - p_1)\alpha + (q - q_1)\beta + (r - r_1)\gamma.$$

The absolute values of the coefficients $(p - p_1)$, $(q - q_1)$, and $(r - r_1)$ of α, β, γ are less than unity; and, since their sum is zero, one of them has a sign opposite to the sign of the two others, and is equal to their sum in absolute value. For the sake of definiteness let us assume that $(p - p_1)$ and $(q - q_1)$ are of one sign, and $(r - r_1)$ is of the other sign. Replacing now in the equality above the difference $(r - r_1)$ by an equal magnitude, the sum of the differences $(p_1 - p)$ and $(q_1 - q)$, we have

$$y = [(p - p_1)\cdot(\alpha - \gamma) + (q - q_1)\cdot(\beta - \gamma)]/(\alpha - \beta)$$

so that

$\mod y \le |p - p_1| + |q - q_1| = |r - r_1| < 1.$

Thus, the moduli of the coefficients y_1 and y_2, included in the development (4.2) of the function $F(1; z)$ in simple factors, are less than unity. Therefore, the coefficients of z^n in the well-known expansion of the fractions of the type

$$[1/(1 - y_1 z)^l],\ [1/(1 - y_2 z)^l]$$

in increasing powers of z must tend to zero together with $(1/n)$.

As to the development of fractions of the type $[1/(1 - z)^l]$ in increasing powers of z, we have

$$[1/(1 - z)^l] = 1 + E_1 z + E_2 z^2 + \ldots + E_n z^n + \ldots,$$

where, on the strength of the formula

$$E_n = (n + 1)(n + 2) \ldots (n + l - 1)/(l - 1)!,$$

$$\lim[E_n/n^{l-1}] = [1/(l - 1)!],\ \lim[E_n/n^{l-1+\varepsilon}] = 0,\ n \to \infty,$$

where ε is any given positive number.

Applying our conclusions to function (4.1), we infer that in its expansion in increasing powers of z the coefficient of z^n, equal to the expectation of (3.1), increases unboundedly with n, if, and only if, $(1 - z)$ cannot be cancelled out of the second term of the expression (4.1).

Desiring that this expectation will not increase unboundedly with n, and taking into account that, on the strength of the equality (2.7), the integral function $f(1; z)$ cannot include the factor $(1 - z)$, we must assign to a such a value that this factor will be included in the function $F'_{u=0}(e^u; ze^{-au})$.

We thus arrive at the equation

$$F'_{u=0}(e^u; ze^{-au}) = 0$$

that can easily be reduced to the following form

$$a[dF(1; z)/dz]_{z=1} + [dF(t; 1)/dt]_{t=1} = 0. \qquad (4.5)$$

The coefficients of this equation, which admits of one and only one solution, can be calculated by means of the formula

$$[dF(1; z)/dz]_{z=1} =$$
$$[q_2 r_1 - (1 - q_1)\cdot(1 - r_2)] + [p_2 r - (1 - p)\cdot(1 - r_2)] + [p_1 q - (1 - p)\cdot(1 - q_1)], \qquad (4.6)$$
$$[dF(t; 1)/dt]_{t=1} =$$
$$(1 - q_1)\cdot(1 - r_2) - q_2 r_1 - (1 - p)\cdot(1 - q_1) + p_1 q.$$

5. We go on to consider the higher powers of the sum (3.1) with the determined value of a. In accord with the above, the expectation of (3.3) can be derived as the coefficient of z^n in the expansion of the function (3.5) in increasing powers of an arbitrary number z.

Proceeding to study this function, we set, for the sake of brevity,

$$f(e^u; ze^{-au}) = U,\ F(e^u; ze^{-au}) = V,\ d^iU/du^i = U^{(i)},\ d^iV/du^i = V^{(i)}. \qquad (5.1)$$

In this notation we have, according to the formula for differentiating a product,

$$d^i(U/V)/du^i = [Ud^i(1/V)\, du^i] + (i/1)\, U'd^{i-1}(1/V)/du^{i-1} + \ldots \qquad (5.2)$$

and, making use of the formula for differentiating a function of a function, we obtain

$$d^i(1/V)/du^i = \sum \frac{i!\,j!}{V^{j+1}} \frac{(-V')^{\lambda}}{\lambda!} \frac{[(-1/2)V'']^{\mu}}{\mu!} \frac{[(-1/6)V''']^{\nu}}{\nu!} \ldots \qquad (5.3)$$

where the summation should extend over all the possible totalities of integral numbers

$$j, \lambda, \mu, \nu, \ldots \qquad (5.4)$$

satisfying two equations

$$\lambda + \mu + \nu + \ldots = j, \; \lambda + 2\mu + 3\nu + \ldots = i \qquad (5.5)$$

and inequalities

$$0 < j \le i, \lambda \ge 0, \mu \ge 0, \nu \ge 0, \ldots$$

We want to prove that, as $n \to \infty$, the ratio of expectation of (3.3) to $\sqrt{n}$ tends to

$$(1/\sqrt{\pi}) C^{i/2} \int_{-\infty}^{\infty} t^i \exp(-t^2)\, dt \qquad (5.6)$$

where C is some constant. For this purpose, bearing in mind the conclusions above, it is necessary to ascertain that the denominator of the rational function of z, represented by the value of the derivative $d^i(U/V)/du^i$ at $u = 0$ can include factor $(1 - z)$,– of course, after appropriate cancellations,– only in a power not higher than $(i + 1)/2$ for odd values of i, and not higher than $[(i/2) + 1]$ for its even values.

Considering one of the products included in formula (5.2)

$$[U^{(l)} d^{i-l}(1/V)/du^{i-l}]_{u=0},$$

we note that the factor $(1 - z)$ cannot be present in the denominator of this product in a power higher than in the denominator of its factor

$$[d^{i-l}(1/V)/du^{i-l}]_{u=0}.$$

Concerning this second factor and issuing from the formula (5.3) with number i replaced by $(i - 1)$, we can, however, derive important conclusions. Indeed, it is not difficult to see that, with the selected value of a, the general term of formula (5.3) taken at $u = 0$ is such an irreducible fraction whose denominator includes the multiplier $(1 - z)$ in a power not higher than $(j + 1 - \lambda)$ because the function V' at $u = 0$ includes this multiplier whereas the function V is not divisible by $(1 - z)^2$. On the other hand, issuing from the conditions imposed on the magnitudes (5.4), it is not difficult to derive the inequality

$$\lambda \ge 2j - i \qquad (5.7)$$

that restricts the values of λ.

Until $2j < i$, the difference

$$j + 1 - \lambda \qquad (5.8)$$

where $\lambda \geq 0$ obviously remains less than $[(i/2) + 1]$. On the strength of inequalities (5.7) this restriction persists also when $2j > i$. And only if $j = i/2$, if such value is possible, and $\lambda = 0$, it can attain the value $[(i/2) + 1]$. Dwelling on the supposition

$$i = 2j, \lambda = 0,$$

which is possible only for even values of i, we find that only one term of the formula (5.3),

$$[i!/2^{i/2}](- V'')^{i/2}/V^{i/2+1} \qquad (5.9)$$

corresponds to this assumption. Indeed, equations (5.5) and their associated inequalities then lead to $\mu = 1/2$.

It follows that for odd values of i not a single fraction, to which the derivatives

$$[d^i(1/V)/du^i], [d^{i-1}(1/V)/du^{i-1}], [d^{i-2}(1/V)/du^{i-2}], \ldots$$

are reduced if $u = 0$, can include in its denominator, of course after appropriate cancellations, the factor $(1 - z)$ in a power higher than $(i + 1)/2$; for even values of i only the first of these fractions obtained from the derivative $[d^i(1/V)/du^i]$ can include in its denominator the factor $(1 - z)$ in the power $[(i/2) + 1]$. If, however, the expression (5.9) at $u = 0$ is subtracted from this first fraction, the difference, after appropriate cancellations, is reduced to a fraction whose denominator includes the factor $(1 - z)$ in a power lower than
$[(i/2) + 1]$.

Comparing this result with formula (5.2) and taking into account notation (5.1), we conclude that the expression studied (3.5) is reduced, for odd values of i, to such a rational fractional function of z whose denominator includes the factor $(1 - z)$ in a power lower then $[(i/2) + 1]$. It immediately follows, on the strength of the investigations in §§3 and 4, that for odd values of i the ratio

$$[\mathrm{E}(x_1 + x_2 + \ldots + x_n - na)^i]/n^{i/2} \qquad (5.10)$$

must tend to zero as $n \to \infty$. Neither is it difficult to indicate the partial fraction with denominator $(1 - z)^{i/2+1}$ which should be isolated from the expression (3.5) so that the denominator of the residual will include only lower powers of $(1 - z)$, and not $(1 - z)^{i/2+1}$ itself.

Indeed, this fraction must coincide with the one whose isolation from the expression

$$[i!/2^{i/2}]\,(U/V)_{u=0}\,(- V''/V)^{i/2}{}_{u=0}$$

leaves a fraction that does not include the multiplier $(1 - z)$ in a power higher than $[(i/2) - 1]$ in its denominator. Therefore, taking into account the equality (2.7), we find that the fraction sought can be represented as

$$[i!/2^{i/2}]\,[C^{i/2}/(1 - z)^{i/2+1}].$$

Here, C is determined by the formula

$$C/2 = F''_{u=0}(e^u;\ e^{-au})/F'_{z=1}(1;\ z). \qquad (5.11)$$

After arriving at such a conclusion and comparing it with the investigations of §§3 and 4, we satisfy ourselves without any serious additional work that for even values of i the ratio (5.10) tends, as $n \to \infty$, to

$$\frac{i\,!}{2^i (i/2)!} C^{i/2}.$$

It remains to take into consideration that the integral (5.6) vanishes if i is odd and is equal to

$$\frac{i\,!}{2^i (i/2)!}$$

otherwise, and we arrive at the final conclusion. If a is determined by the equation (4.5) and C is given by the formula (5.11), then, both for odd and even values of i, as $n \to \infty$,

$$\lim[\mathrm{E}(x_1 + x_2 + \ldots + x_n - na)^i]/n^{i/2} = (1/\sqrt{\pi})C^{i/2} \int_{-\infty}^{\infty} t^i \exp(-t^2)\, dt.$$

Consequently, as $n \to \infty$, the probability of the inequalities

$$t_1\sqrt{Cn} < x_1 + x_2 + \ldots + x_n - na < t_2\sqrt{Cn} \qquad (5.12)$$

where t_1 and t_2 are any given numbers with $t_2 > t_1$, must tend to

$$(1/\sqrt{\pi}) \int_{t_1}^{t_2} \exp(-t^2)\, dt.$$

6. After considering a case of quantities connected into a chain and noting that it does not present any peculiarities but differs from the other ones only in the simplicity of the initial data, we may in a few words extend our deductions onto the general instance of such quantities studied in the abovementioned article [3].

Keeping to the notation in this paper, suppose that

$$\alpha, \beta, \gamma, \ldots \qquad (6.1)$$

are the various possible values of the numbers in the chain

$$x_1, x_2, \ldots, x_k, x_{k+1}, \ldots, \qquad (6.2)$$

and that the system

$$p_{\alpha\alpha}, p_{\alpha\beta}, p_{\alpha\gamma}, \ldots;\ p_{\beta\alpha}, p_{\beta\beta}, p_{\beta\gamma}, \ldots;\ p_{\gamma\alpha}, p_{\gamma\beta}, p_{\gamma\gamma}, \ldots, \text{etc} \qquad (6.3)$$

represents the probabilities, given the values of x_k, for x_{k+1} to take definite values; the subscripts of the p's indicate the given x_k and the supposed x_{k+1}, respectively.

The numbers (6.1) and (6.3) determine our final conclusions, but, to pose a quite definite problem, we ought to introduce additionally numbers

$$p'_\alpha,\ p'_\beta,\ p'_\gamma,\ \ldots$$

that, just as the numbers (1.4) of the particular case did, fail to appear in the final conclusions, and are the probabilities of the equalities

$$x_1 = \alpha,\ x_2 = \beta,\ x_3 = \gamma,\ \ldots$$

respectively as long as all the x's of our chain remain indefinite.

Turning to considering various assumptions about the value of the sum (1.5), we denote the probability of the equality (1.6) by symbol P_{mn} and introduce a function Φ_n of an arbitrary number t

$$\Phi_n = \sum P_{mn}\, t^m. \tag{6.4}$$

Acting now in the same way as we did in the particular case, we partition P_{mn}:

$$P_{mn} = P_{mn,\alpha} + P_{mn,\beta} + P_{mn,\gamma} + \ldots \tag{6.5}$$

and introduce a number of functions

$$\Phi_{n,\alpha} = \sum P_{mn,\alpha}\, t^m,\ \Phi_{n,\beta} = \sum P_{mn,\beta}\, t^m,\ \ldots \tag{6.6}$$

where the new symbols are the probabilities of the same equality (1.6) and of an additional restriction expressed by one of the equalities $x_n = \alpha;\ \beta;\ \gamma;\ \ldots$

Having this notation at our disposal, it is not difficult, in accord with the conditions of the problem, to establish the equations

$$P_{mn,\alpha} = P_{m-\alpha,n-1,\alpha}\, p_{\alpha\alpha} + P_{m-\alpha,n-1,\beta}\, p_{\beta\alpha} + \ldots,$$
$$P_{mn,\beta} = P_{m-\beta,n-1,\alpha}\, p_{\alpha\beta} + P_{m-\beta,n-1,\beta}\, p_{\beta\beta} + \ldots,$$

then passing on from them to equations

$$\begin{aligned} \Phi_{n,\alpha} t^{-\alpha} &= p_{\alpha\alpha}\Phi_{n-1,\alpha} + p_{\beta\alpha}\Phi_{n-1,\beta} + p_{\gamma\alpha}\Phi_{n-1,\gamma} + \ldots \\ \Phi_{n,\alpha} t^{-\beta} &= p_{\alpha\beta}\Phi_{n-1,\alpha} + p_{\beta\beta}\Phi_{n-1,\beta} + p_{\gamma\beta}\Phi_{n-1,\gamma} + \ldots \end{aligned} \tag{6.7}$$

One and the same homogeneous difference equation follows now for all the functions $\Phi_{n,\alpha}$, $\Phi_{n,\beta}$, $\Phi_{n,\gamma}$, …as well as for $\Phi_n = \Phi_{n,\alpha} + \Phi_{n,\beta} + \Phi_{n,\gamma} + \ldots$ It can be symbolically represented in a rather simple way as

$$\begin{vmatrix} p_{\alpha\alpha} - t^{-\alpha}\Phi & p_{\beta\alpha} & p_{\gamma\alpha} & \ldots \\ p_{\alpha\beta} & p_{\beta\beta} - t^{-\beta}\Phi & p_{\gamma\beta} & \ldots \\ p_{\alpha\gamma} & p_{\beta\gamma} & p_{\gamma\gamma} - t^{-\gamma}\Phi & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{vmatrix} \Phi_n = 0 \tag{6.8}$$

where, after performing the indicated operations, Φ^n, Φ^{n+1}, Φ^{n+2}, … should be replaced by Φ_n, Φ_{n+1}, Φ_{n+2}, … respectively.

On the strength of this equation all the functions Φ_0, Φ_1, Φ_2, …, Φ_n, … can be determined as coefficients in the expansion of some rational function $f(t;\,z)/F(t;\,z)$ of a second supplementary number z in its increasing powers. The denominator of this function is determined as

$$F(t;\,z) = \begin{vmatrix} p_{\alpha\alpha}z - t^{-\alpha} & p_{\beta\alpha}z & p_{\gamma\alpha}z & \dots \\ p_{\alpha\beta}z & p_{\beta\beta}z - t^{-\beta} & p_{\gamma\beta}z & \dots \\ p_{\alpha\gamma}z & p_{\beta\gamma}z & p_{\gamma\gamma}z - t^{-\gamma} & \dots \\ \dots & \dots & \dots & \dots \end{vmatrix}. \qquad (6.9)$$

Before going on to further conclusions, it is necessary to note that we are only considering such chains (6.2) for which the appearance of some numbers (6.1) does not definitively exclude the possibility of the appearance of other numbers. This important condition can be expressed in the following way: the determinant

$$\begin{vmatrix} u & p_{\beta\alpha} & p_{\gamma\alpha} & \dots \\ p_{\alpha\beta} & v & p_{\gamma\beta} & \dots \\ p_{\alpha\gamma} & p_{\beta\gamma} & w & \dots \\ \dots & \dots & \dots & \dots \end{vmatrix} \qquad (6.10)$$

with arbitrary elements u, v, w, … is not reducible to a product of several determinants of the same type.

This, however, is not sufficient for our purpose, and we ought to suppose [1] that the determinant above cannot be explicitly reduced to a product of several determinants when $u = p_{\alpha\alpha}$, $v = p_{\beta\beta}$, $w = p_{\gamma\gamma}$, either.

Both in the particular and in the general case our conclusions concern the expectation of the power (3.3) where a is determined by demanding that at $i = 1$ this expectation will not increase unboundedly with n. Having established, as before, that it can be expressed by the coefficient of z^n in the expansion of the value of the derivative (3.5) in increasing powers of z, we note that, for extending the inferences made in the particular case onto the general instance, we ought to consider the development of the function $F(1;\,z)$ in linear factors

$$F(1;\,z) = \pm\,(1 - z)\cdot(1 - y_1z)\cdot(1 - y_2z)\,\dots \qquad (6.11)$$

and to prove that the multiplier $(1 - z)$ occurs here only once and that the moduli of y_1, y_2, … are less than unity.

In other words, we ought to satisfy ourselves that unity is a simple rather than a multiple root of the equation

$$\begin{vmatrix} p_{\alpha\alpha} - y & p_{\beta\alpha} & p_{\gamma\alpha} & \dots \\ p_{\alpha\beta} & p_{\beta\beta} - y & p_{\gamma\beta} & \dots \\ p_{\alpha\gamma} & p_{\beta\gamma} & p_{\gamma\gamma} - y & \dots \\ \dots & \dots & \dots & \dots \end{vmatrix} = 0 \qquad (6.12)$$

and that the moduli of the other roots of the same equation, y_1, y_2, …, are less than unity.

The following proposition regarding determinants can serve to prove the first point. If all the elements of the determinant

$$\begin{vmatrix} u & -b_1 & -c_1 & \dots \\ -a_1 & v & -c_2 & \dots \\ -a_2 & -b_2 & w & \dots \\ \dots & \dots & \dots & \dots \end{vmatrix}$$

satisfy inequalities

$$a_k \ge 0,\ b_k \ge 0,\ c_k \ge 0,\ \dots \qquad (6.13)$$

and inequalities

$$u \ge a_1 + a_2 + \dots,\ v \ge b_1 + b_2 + \dots,\ w \ge c_1 + c_2 + \dots,\ \dots \qquad (6.14)$$

it cannot be negative; and it can vanish only in the extreme case in which all the inequalities (6.14) are reduced to equalities, or in which it becomes, owing to the reduction of some of the inequalities (6.13) to equalities, a product of several determinants of the same type with, moreover, one of the latter being such that all the inequalities similar to (6.14) are reduced to equalities [2].

We become convinced in this important proposition when considering *u, v, w*, … as variables and noting that the derivatives of our determinant with respect to them are expressed by similar determinants of a lower order; such a reasoning enables us to extend gradually this theorem from a determinant of the second order, for which it is evident, onto a determinant of the third; then to one of the fourth order, etc.

It immediately follows from the proposition thus proved that under our conditions the derivative of the left side of equation (6.12) with respect to y does not vanish at $y = 1$. Indeed, after being multiplied by ± 1, it can be represented as a sum of the determinants

$$\begin{vmatrix} 1-p_{\beta\beta} & -p_{\gamma\beta} & \dots \\ -p_{\beta\gamma} & 1-p_{\gamma\gamma} & \dots \\ \dots & \dots & \dots \end{vmatrix} + \begin{vmatrix} 1-p_{\alpha\alpha} & -p_{\gamma\alpha} & \dots \\ -p_{\alpha\gamma} & 1-p_{\gamma\gamma} & \dots \\ \dots & \dots & \dots \end{vmatrix} + \dots$$

obeying the conditions of the theorem and not presenting the extreme case.

Thus, unity, which on the strength of the conditions

$$p_{\alpha\alpha} + p_{\alpha\beta} + p_{\alpha\gamma} + \dots = 1,\ p_{\beta\alpha} + p_{\beta\beta} + p_{\beta\gamma} + \dots = 1,\ \dots$$

must satisfy equation (6.12), cannot be its multiple root. Turning to the other roots of this equation, we assume that y is any one of them. A system of numbers

$$\alpha',\ \beta',\ \gamma',\ \dots \qquad (6.15)$$

not all of them zeros and satisfying the equations

$$\alpha' y = p_{\alpha\alpha}\alpha' + p_{\alpha\beta}\beta' + p_{\alpha\gamma}\gamma' + \dots,\ \beta' y = p_{\beta\alpha}\alpha' + p_{\beta\beta}\beta' + p_{\beta\gamma}\gamma' + \dots,$$
$$\gamma' y = p_{\gamma\alpha}\alpha' + p_{\gamma\beta}\beta' + p_{\gamma\gamma}\gamma' + \dots,\ \dots$$

can be chosen accordingly. These numbers cannot coincide because the equations above will then lead to $y = 1$. Taking this into account, we suppose at first that the numbers (6.15) do not have a common modulus either. Then, owing to our conditions, among the equations to which the numbers (6.15) are subordinated, there must be at least one, the coefficient of y in whose left side will be one of these numbers with the greatest modulus and whose right side will also include, with a non-zero coefficient, another one of these numbers (6.15) with a lesser modulus. It follows immediately, since the sum of the coefficients of the numbers (6.15) in the right sides of each of these equations is unity, that

$$\text{mod } y < 1. \tag{6.16}$$

Let us suppose now that all the numbers (6.15) have the same modulus. We know, however, that not all of them are equal one to another. Their totality can therefore be partitioned in two groups with respect to the first number, α': the numbers in the first group being equal, and in the second group, unequal to α'. The numbers in the second group differ from α' in their arguments. On the other hand, because of one of our main conditions, it is impossible to partition the totality of the sums

$$p_{\alpha\alpha}\alpha' + p_{\alpha\beta}\beta' + p_{\alpha\gamma}\gamma' + \ldots;\ p_{\beta\alpha}\alpha' + p_{\beta\beta}\beta' + p_{\beta\gamma}\gamma' + \ldots;\ \ldots$$

in two totalities in such a way that the first sums will include, with non-zero coefficients [3], only the numbers equal to α' and the second sums will include only the other numbers.

Therefore, one of these sums will certainly include, with non-zero coefficients, numbers both equal and unequal to α', and its modulus, equal to the product (mod y) (mod$\cdot\alpha'$), must be less than mod α' since for numbers with differing arguments the modulus of their sum is less than, and not equal to the sum of their moduli. The inequality (6.16), which we should have proved, follows now immediately.

Having thus proved that in developing the function $F(1; z)$ in simple multipliers (6.11) the factor $(1 - z)$ only enters in the first power, and that the moduli of the coefficients $y_1, y_2, \ldots$ are less than unity, we may transfer without change all the further deliberations, that we have applied to the particular instance, onto the general case. Since our reasoning remains entirely invariable, we shall not repeat it and only adduce the final conclusion.

If, under the stipulated conditions and notation, we define numbers a and C by equalities {cf. (5.11)}

$$a = F'_{t=1}(t;\ 1)\ /\ F'_{z=1}(1;\ z),\quad C/2 = F''_{u=0}\,(e^u;\ e^{-au})/F'_{z=1}(1;\ z), \tag{6.17}$$

which does not lead either to impossible or indeterminate results, then, for any integral positive i {see end of §5}, as $n \to \infty$,

$$\lim[\mathrm{E}(x_1 + x_2 + \ldots + x_n - na)^i]/n^{i/2} = (1/\sqrt{\pi})C^{i/2} \int_{-\infty}^{\infty} t^i \exp(-t^2)\,dt$$

and, consequently, as $n \to \infty$, the probability of the inequalities

$$t_1\sqrt{Cn} < x_1 + x_2 + \ldots + x_n - na < t_2\sqrt{Cn} \tag{5.12}$$

where t_1 and t_2 are any given numbers with $t_2 > t_1$, must tend to

$$(1/\sqrt{\pi})\int_{t_1}^{t_2} \exp(-t^2)\,dt.$$

It can also be noted about the numbers a and C that, as $n \to \infty$, they are equal to the limits of

$$[E(x_1 + x_2 + \ldots + x_n)/n\,],\ 2[E(x_1 + x_2 + \ldots + x_n - na)^2/n]$$

respectively. This remark allows to extend our conclusions also onto such cases in which the totality (6.1) is not exhausted by a finite number of terms. It is also possible to extend our conclusions onto generalized chains whose each term is directly connected not with one, but with several preceding terms.

Notes and Commentary by N.A. Sapogov [5, pp. 662 – 665]

1. Our conclusions may be extended onto many of the excluded cases.
2. A similar proposition is in Minkovsky's note [6].
3. Previously, I [3] have only considered only the simplest assumption that among these coefficients none is equal to zero.

{In several works} Markov proves the applicability of the CLT to sums

$$S_n = X_1 + X_2 + \ldots + X_n$$

of random variables X_i […]. {Here,} he studies variables X_i forming a simple homogeneous chain and especially dwells on the assumption that the possible values of the variables are –1, 0, and 1. To calculate the moments of S_n, he introduces generating functions

$$\Phi_n(t) = \sum_m p_{mn} t^m$$

with p_{mn} being the probabilities of the equalities $S_n = m$, and studies the generating function of the functions $\Phi_n(t)$

$$\sum_n \Phi_n(t)\, z^n = f(t;\, z)/F(t;\, z). \qquad (1)$$

The main role in the ensuing calculations was played by the denominator of this fraction, $F(t; z)$, which, as it appeared, was a polynomial of the third degree with respect to z. Under some restrictions of a general nature concerning transition probabilities, it turned out that $F(1; z)$ had a simple root $z = 1$ and two other ones whose moduli were less than 1. On the contrary, the function $f(t; z)$ did not play an important part nor did it influence the results of calculation as $n \to \infty$.

Concerning the moments $E[(S_n - an)/\sqrt{n}]^i$ where a was an adequately chosen number, Markov showed by investigating the right side of (1), that, for any integral $i = 0, 1, 2, \ldots$ and $n \to \infty$,

$$E[(S_n - an)/\sqrt{n}]^i \to (1/\sqrt{\pi})C^{i/2}\int_{-\infty}^{\infty} t^i \exp(-t^2)\,dt \qquad (2)$$

where

$$C = [2\partial^2 F(e^u;\, e^{-au})/\partial u^2]_{u=0} / \partial F(1;\, z)/\partial z_{z=1}.$$

The proof was largely based on the fact that $z = 1$ was a simple, and, furthermore, the only root of the function $F(1; z)$ having maximal modulus. A similar reasoning leading under these conditions to (2) can be reproduced for many other cases concerning the variables S_n […]

In concluding his work, in §6, Markov studied variables X_i with values

$$\alpha, \beta, \gamma, \ldots \qquad (3)$$

He proved that the CLT was applicable to S_n under conditions, which, in our opinion, were not formulated distinctly enough. He said that he was considering *only such chains for which the appearance of* {certain} *numbers* (3) [did] *not exclude definitively the possibility of the appearance of other numbers*.

In other words, using modern terminology, he studied homogeneous chains whose matrixes of transition probabilities (stochastic matrixes) were simple [1, Addendum 5]; or, all the states of the chain formed one *class of essential states* [2]. Then Markov put forward this condition: *The determinant* (6.10) *with arbitrary elements u, v, w, … is not reducible to a product of several determinants of the same type*, and, also, *the determinant above cannot be explicitly reduced to a product of several determinants when* $u = p_{\alpha\alpha}$, $v = p_{\beta\beta}$, $w = p_{\gamma\gamma}$, … *either*. He then stated that under these conditions the characteristic equation (6.12) has only one root of mod $y = 1$, namely $y = 1$ whereas all the other ones have moduli less than 1. It is known that in the case of a simple stochastic matrix a necessary and sufficient condition for the equation (6.12) to have only one root, maximal in modulus, namely, $y = 1$, is, in Bernstein's terminology, the regularity of this matrix; in other words, its period must be equal to 1. […]

In concluding, we note that, for being irreproachable, Markov's inferences about the applicability of the CLT to variables S_n should be supplemented by demanding that the non-negative constant C in equality (6.17) […] be positive.

References

1. Bernstein, S.N. (1946), *Теория вероятностей* (Theory of Probability). M. – L. Fourth edition.

2. Kolmogorov, A.N. (1937), Markov chains with a countable set of possible states. *Bull. Moskovsk. Gos. Univ*., vol. 1, No. 3. (R)

3. Markov, A.A. (1906), Extension of the law of large numbers, etc. Translated in this book.

4. --- (1907, in Russian), Recherches sur un cas remarquable d'épreuves dépendantes. *Acta math.*, t. 33, 1910, pp. 87 – 104.

5. --- (1951), *Избранные труды* (Sel. Works). N.p.

6. Minkowski, H. (1900), Zur Theorie der Einheiten in den algebraischen Zahlkörpern. *Nachr. Kgl. Ges. Wiss. Göttingen,* math.-phys. Kl., 1900, pp. 90 – 93.

13. A.A. Markov. An Investigation of the General Case of Trials Connected into a Chain

When investigating the simplest case of trials connected into a chain, we [5] have ascertained the possibility of extending the LLN onto it. Then, in addition, we [6] established for it the theorem on the limit of expectation, from which, on the strength of the works of Chebyshev and myself, follows the theorem on the limit of probability. Our conclusions were based, in their most difficult stage, on the use of generating functions, which, for that case, were expressed rather simply. Their consideration also allowed us to extend the same inferences onto any homogeneous [1] chain of magnitudes [7].

Owing to the complexity of the generating functions themselves, this method will hardly simplify the investigation of non-homogeneous chains. Therefore, desiring to extend the theorems on the limits of expectation and probability onto magnitudes connected in such chains, we were compelled to return to those methods, which we had successfully applied to various cases of independent magnitudes and which consisted in decomposing, on the strength of the Newton formula, the expectation of the power of the sum into separate sums, then isolating from them those sums, that increase more rapidly than any other ones, of course with the increase in the number of magnitudes under consideration.

Here, we shall dwell on the general case of trials connected into a non-homogeneous chain that does not contain such links where the probability of the event can happen to be either zero or unity, or be arbitrarily close to these numbers. Note that the word *probability*, as was already established [6], means here, for each trial, not one but three numbers corresponding to different assumptions.

1. Beginning our investigations, we ought first of all to establish the relevant conditions and introduce notation. We shall consider an unbounded series of consecutive trials distinguishing them, in their order, by numbers 1, 2, 3, … At each of these trials, either some event E or its contrary event F can occur. Our trials are connected into a chain in such a way that for any integral positive number k we have

1) The probability of event E at the $(k + 1)$-th trial, until the results are not at all known, has a definite value p_{k+1}; the probability of F under the same condition is $q_{k+1} = 1 - p_{k+1}$.

2) The probability of event E at the $(k + 1)$-th trial takes another definite value, p'_{k+1} if the results of the consequent trials are indefinite as before, but the directly preceding trial, i.e., the k-th trial, furnished event E, whatever were the results of the other ones, *i.e.*, of the first, the second, …, and the $(k - 1)$-th trial. The probability of event F under these new conditions is $q'_{k+1} = 1 - p'_{k+1}$.

3) Finally, the probability of E at the $(k + 1)$-th trial takes a third definitive value, p''_{k+1} if the results of the consequent trials remain indefinite as before, but the directly preceding trial, *i.e.*, the k-th trial, had not provided event E whatever were the results of the other ones. The probability of event F in this last case is $q''_{k+1} = 1 - p''_{k+1}$.

On the strength of these conditions, trial $(k + 1)$ is only connected with the previous ones through the k-th trial and becomes independent from them once the last-mentioned one is determined. We must complement the data indicated above by the probabilities p_1 and $q_1 = 1 - p_1$ of the events E and F at the first trial supposing that the results of all of the trials remain indefinite.

The numbers

$$p_1, p'_2, p''_2, p'_3, p''_3, \dots$$

may be assigned arbitrarily; and, issuing from them, it is not difficult to calculate consecutively $p_2, p_3, \dots, p_k, p_{k+1}, \dots$ by means of a simple general formula

$$p_{k+1} = p_k\, p'_{k+1}, + q_k\, p''_{k+1}. \tag{1.1}$$

This formula can also serve for another purpose: by its means, we can express p'_{k+1}, p''_{k+1}, q'_{k+1}, q''_{k+1} through

$$p_{k+1}, q_{k+1}, p_k, q_k \text{ and } \delta_{k+1} = p'_{k+1} - p''_{k+1}.$$

Indeed, denoting δ_{k+1} as above, we derive from formula (1.1)

$$p_{k+1} = p'_{k+1}\,(p_k + q_k) - \delta_{k+1}\,q_k = p'_{k+1} - \delta_{k+1}\,q_k$$

and can therefore immediately establish a number of simple equalities

$$p'_{k+1} = p_{k+1} + \delta_{k+1}\,q_k,\; q'_{k+1} = q_{k+1} - \delta_{k+1}\,q_k,$$
$$p''_{k+1} = p_{k+1} - \delta_{k+1}\,p_k,\; q''_{k+1} = q_{k+1} - \delta_{k+1}\,p_k. \qquad (1.2)$$

We may thus introduce into our calculations p_{k+1}, q_{k+1}, δ_{k+1} instead of p'_{k+1}, p''_{k+1}, q'_{k+1} and q''_{k+1} and we shall indeed do it so as to simplify our deductions as much as possible. It should be noted that p'_{k+1} and p''_{k+1} are only restricted by inequalities $0 \le p'_{k+1} \le 1$ and $0 \le p''_{k+1} \le 1$ so that $-1 \le \delta_{k+1} \le 1$. If, however, we introduce p_{k+1} and δ_{k+1} instead of p'_{k+1} and p''_{k+1}, then, as can easily be ascertained, the numbers δ_{k+1} ought to be restricted by the new inequalities

$$-(p_{k+1}/q_k) \le \delta_{k+1} \le (q_{k+1}/q_k),\; -(q_{k+1}/p_k) \le \delta_{k+1} \le (p_{k+1}/p_k),$$

whereas the numbers p_{k+1} must only obey inequalities $0 \le p_{k+1} \le 1$.

2. Let us dwell now on a problem, important for our aim, of calculating the probability of event E at the k-th trial, at first under the condition that it had occurred at the i-th trial ($i < k$), and then under the contrary condition. Denoting for this purpose the former probability by $P_k^{(i)}$, and the latter by $Q_k^{(i)}$, and considering different values of k, we deduce simple equations

$$P_{k+1}^{(i)} = P_k^{(i)}\,p'_{k+1} + (1 - P_k^{(i)})\,p''_{k+1} = p_{k+1} - \delta_{k+1}\,p_k + \delta_{k+1}\,P_k^{(i)},$$
$$Q_{k+1}^{(i)} = Q_k^{(i)}\,p'_{k+1} + (1 - Q_k^{(i)})\,p''_{k+1} = p_{k+1} - \delta_{k+1}\,p_k + \delta_{k+1}\,Q_k^{(i)}.$$

At the same time we note, that, because of our notation,

$$P_{i+1}^{(i)} = p'_{i+1} = p_{i+1} + \delta_{i+1}\,q_i,\; Q_{i+1}^{(i)} = p''_{i+1} = p_{i+1} - \delta_{i+1}\,p_i.$$

Issuing from these equalities, it is not difficult to deduce consecutively

$$P_{i+2}^{(i)} = p_{i+2} + \delta_{i+1}\,\delta_{i+2}\,q_i,\; Q_{i+2}^{(i)} = p_{i+2} - \delta_{i+1}\,\delta_{i+2}\,p_i,$$
$$P_{i+3}^{(i)} = p_{i+3} + \delta_{i+1}\,\delta_{i+2}\,\delta_{i+3}\,q_i,\; Q_{i+3}^{(i)} = p_{i+3} - \delta_{i+1}\,\delta_{i+2}\,\delta_{i+3}\,p_i,$$

and, in general, to arrive at

$$P_k^{(i)} = p_k + \delta_{i+1}\,\delta_{i+2}\,\ldots\,\delta_k\,q_i,\; Q_k^{(i)} = p_k - \delta_{i+1}\,\delta_{i+2}\,\ldots\,\delta_k\,p_i. \qquad (2.1)$$

Thus, under the stated condition, the probability of the event E at the k-th trial is determined when the result of the i-th trial ($i < k$) is known.

Neither is it difficult to establish the formulas for calculating the probabilities $P_k^{(i)}$ and $Q_k^{(i)}$ of event E at the i-th trial under the condition that E took place, or alternatively, that it did not occur at the k-th trial (as before, $i < k$). These last-mentioned formulas directly follow from formula (2.1) and from the simple equalities

$$p_i P_k^{(i)} = p_k P_i^{(k)},\; p_i(1 - P_k^{(i)}) = q_k Q_i^{(k)}.$$

Indeed, we have

$$P_i^{(k)} = p_i + \delta_{i+1}\,\delta_{i+2}\,\ldots\,\delta_k\,p_iq_i/p_k,\ Q_i^{(k)} = p_i + \delta_{i+1}\,\delta_{i+2}\,\ldots\,\delta_k\,p_iq_i/q_k. \qquad (2.2)$$

Concerning formulas (2.1) and (2.2), we note that, in accord with what we said in the introduction, neither of the numbers p_k or q_k can be zero or arbitrarily close to it; therefore, the inequalities

$$1 > p^\circ > p_k > p > 0 \qquad (2.3)$$

hold for some constant magnitudes p° and p and any subscript k. Introducing a similar assumption concerning all the numbers p'_{k+1} and p''_{k+1}, we ought also to establish the general inequalities

$$1 < -\delta < \delta_{i+1} < \delta < 1$$

where δ is the same for all values of k. These inequalities will play an important part in our conclusions about the limiting values of the studied expressions.

It is of course possible to pose the question on the extension of the final deductions also on such cases, for which some of the numbers p_k, $1 - p_k$, and δ_{k+1} are arbitrarily close to unity. However, here we shall not consider this difficult problem leaving it for other researchers. On the contrary, not wishing to ascertain how rapidly is one rather involved expression increasing, we believe it reasonable to restrict, in this paper, the numbers δ_{k+1} by some additional inequalities; being complied with, they ensure that the investigation of this problem presents no difficulty. We shall introduce these inequalities at an appropriate time; for the present, those established above are sufficient.

They are sufficient, for example, for deriving from formulas (2.1) and (2.2) the inequalities

$$|P_k^{(i)} - p_k| < \delta^{k-i},\ |Q_k^{(i)} - p_k| < \delta^{k-i},$$
$$|P_i^{(k)} - p_i| < \delta^{k-i}/4p^\circ,\ |Q_i^{(k)} - p_i| < \delta^{k-i}/[4(1 - p^\circ)]$$

on whose strength we can infer that all these differences tend to zero when $(k - i) \to \infty$. However, this last conclusion can of course be also justified in many other cases on which we do not dwell.

3. Turning now to consider the totality of the first n trials, denote the number of the occurrences of event E in them by m, and, in accord with a well-known rule, let

$$m = x_1 + x_2 + \ldots + x_k + \ldots + x_n. \qquad (3.1)$$

Here, x_k is unity if event E occurs at trial k and zero otherwise, when at this trial happens event F. We shall study the possible values of m assuming that the results of all the trials remain indefinite.

Under this assumption we shall consider the expectations of various powers of the difference

$$m - (p_1 + p_2 + \ldots + p_n) \qquad (3.2)$$

equal to the sum

$$(x_1 - p_1) + (x_2 - p_2) + \ldots + (x_n - p_n) \qquad (3.3)$$

and compare these expectations with powers of n. We attempt to isolate the main part of each expectation which determines the law of its increase as $n \to \infty$.

The expectation of the first power of the difference (3.2) is obviously zero since

$$Ex_1 = p_1,\ Ex_2 = p_2,\ \ldots,\ Ex_n = p_n.$$

Neither is it difficult to compile the expectation of its square. To achieve this we make use of the well-known equality

$$[(x_1 - p_1) + (x_2 - p_2) + \ldots + (x_n - p_n)]^2 = \sum(x_k - p_k)^2 + 2\sum(x_i - p_i)\,(x_k - p_k)$$

with $k = 1, 2, \ldots, n$ in the first sum, and $i = 1, 2, \ldots, n - 1$ and $k = i + 1, \ldots, n$ in the second one. We then calculate the expectations of the squares $(x_k - p_k)^2$ and the products $(x_i - p_i)\,(x_k - p_k)$. Applying formulas (2.1), we deduce

$$E(x_k - p_k)^2 = p_k q_k^2 + q_k p_k^2 = p_k q_k,$$

$$E(x_i - p_i)\,(x_k - p_k) = p_i q_i\,(P_k^{(i)} - p_k) - q_i p_i\,(Q_k^{(i)} - p_k) =$$
$$p_i q_i\,(P_k^{(i)} - Q_k^{(i)}) = \delta_{i+1}\,\delta_{i+2} \ldots \delta_k\, p_i q_i.$$

Therefore, for $i = 1, 2, \ldots, n$,

$$E[m - (p_1 + p_2 + \ldots + p_n)]^2 = \sum p_i q_i\,(1 + 2\delta_{i+1} + 2\delta_{i+1}\delta_{i+2} + \ldots 2\delta_{i+1} \ldots \delta_n) =$$

$$p_1 q_1\,(1 + 2\delta_2 + 2\delta_2\delta_3 + \ldots 2\delta_2\,\delta_3 \ldots \delta_n) +$$
$$p_2 q_2\,(1 + 2\delta_3 + 2\delta_3\delta_4 + \ldots 2\delta_3\,\delta_4 \ldots \delta_n) + \ldots +$$
$$p_{n-1}\,q_{n-1}\,(1 + 2\delta_n) + p_n\,q_n. \qquad (3.4)$$

Our expression for the expectation of the square of (3.2) enables us to extend the LLN in a very simple way onto this case (onto the Bernoulli theorem) if only all the $|\delta_{i+1}|$ remain less than one and the same number δ, which, in accord with our abovementioned condition, is itself less than unity. Indeed, if

$$-\delta < \delta_2 < \delta,\ -\delta < \delta_3 < \delta,\ \ldots,\ -\delta < \delta_n < \delta < 1,$$

not a single sum from among

$$1 + 2\delta_n,\ 1 + 2\delta_{n-1} + 2\delta_{n-1}\delta_n,\ \ldots,\ 1 + 2\delta_2 + 2\delta_2\delta_3 + \ldots + 2\delta_2\delta_3 \ldots \delta_n \qquad (3.5)$$

attains the value

$$[(1 + \delta)/(1 - \delta)] = 1 + 2\delta + 2\delta^2 + \ldots + 2\delta^n + \ldots$$

so that the formula (3.4) provides the inequality

$$E[m - (p_1 + p_2 + \ldots + p_n)]^2 < n\,(1 + \delta)/[4(1 - \delta)].$$

The LLN is derived from this inequality by means of a well-known reasoning; therefore, with probability arbitrarily close to certainty, we may, for a sufficiently large n, state that the difference

$$[m/n] - [(p_1 + p_2 + \ldots + p_n)/n]$$

is less in absolute value than any given number.

For our further deliberations concerning the expectations of the higher powers of the difference (3.2), it is important that the ratio

$$E[(x_1 - p_1) + (x_2 - p_2) + \ldots + (x_n - p_n)]^2/n \tag{3.6}$$

will not become arbitrarily small. We do not dwell on the question of whether the inequalities indicated above are sufficient for the last requirement to be fulfilled; we shall only indicate several cases in which it is not difficult to ascertain that this condition is indeed obeyed. First, if all the magnitudes

$$\delta_2, \delta_3, \ldots \tag{3.7}$$

are positive, then each sum (3.5) is not less than unity, so that, with the general inequalities (2.3) persisting, formula (3.4) leads to a simple inequality

$$\{E[(x_1 - p_1) + (x_2 - p_2) + \ldots + (x_n - p_n)]^2/n\} > p\,(1 - p°). \tag{3.8}$$

If, however, some of the numbers (3.7) are negative, but δ, introduced earlier, is less than 1/3, then each sum (3.5) is larger than the positive number

$$[(1 - 3\delta)/(1 - \delta)] = 1 - 2\delta - 2\delta^2 - 2\delta^3 - \ldots,$$

and, owing to formula (3.4), we can also establish a simple inequality: the left side of (3.8) is larger than

$$(1 - 3\delta)\,p\,(1 - p°)/(1 - \delta).$$

Having thus ascertained the existence of cases in which the ratio (3.6) cannot be arbitrarily small, we shall assume further on that this condition is fulfilled, leaving room for doubting whether it represents a new restriction, or holds as a corollary of those established earlier [2].

4. Before considering the expectation of

$$[(x_1 - p_1) + (x_2 - p_2) + \ldots + (x_n - p_n)]^m$$

for any given integral positive number m, we shall dwell on the particular case of $m = 3$ whose investigation can serve for throwing light on some peculiarities of our forthcoming reasoning. We must prove that the expression

$$E\left(\frac{x_1 - p_1 + x_2 - p_2 + \ldots + x_n - p_n}{\sqrt{E(x_1 - p_1 + x_2 - p_2 + \ldots + x_n - p_n)^2}}\right)^3 \tag{4.1}$$

tends to zero as $n \to \infty$. For this purpose, we note that it must tend to zero if the same is true for the ratio

$$E[(x_1 - p_1) + (x_2 - p_2) + \ldots + (x_n - p_n)]^3/n^{3/2} \tag{4.2}$$

which, in accord with the condition formulated at the end of §3, only differs from (4.1) by a finite factor.

Then we pay attention to the well-known simple equality

$$[(x_1 - p_1) + (x_2 - p_2) + \ldots + (x_n - p_n)]^3 = \sum(x_i - p_i)^3 + 3\sum(x_i - p_i)^2(x_k - p_k) + 3\sum(x_i - p_i)(x_k - p_k)^2 + 6\sum(x_i - p_i)(x_k - p_k)(x_r - p_r)$$

where

$$1 \le i \le n;\ 1 \le i < k \le n;\ 1 \le i < k \le n; \text{ and } 1 \le i < k < r \le n,$$

respectively. The equality follows as a particular case from the generalized Newton formula with respect to a power of a sum. Consequently, we consider the expectations of the four sums.

The expectation of the first of these is equal to the sum $\sum p_i\, q_i(q_i^2 - p_i^2)$ consisting of n terms not one of them attaining 1/4 in absolute value; therefore, the ratio $E\sum(x_i - p_i)^3/n$ remains finite, and the ratio $E\sum(x_i - p_i)^3/n^{3/2}$ tends to zero as $n \to \infty$. Turning to the second sum, we find that

$$E[(x_i - p_i)^2(x_k - p_k)] = p_i q_i^2\,(P_k^{(i)} - p_k) + q_i p_i^2(Q_k^{(i)} - p_k) = p_i q_i\,(q_i - p_i)\,\delta_{i+1}\delta_{i+2}\ldots\delta_k$$

so that

$$E[(x_i - p_i)^2(x_k - p_k)] = p_1 q_1(q_1 - p_1)[\delta_2 + \delta_2\,\delta_3 + \ldots + \delta_2\,\delta_3\ldots\delta_n] + p_2 q_2(q_2 - p_2)[\delta_3 + \delta_3\,\delta_4 + \ldots + \delta_3\,\delta_4\ldots\delta_n] + \ldots + p_{n-1}q_{n-1}\,(q_{n-1} - p_{n-1})\delta_n.$$

This equality reveals that the ratio

$$E\sum[(x_i - p_i)^2(x_k - p_k)]/n$$

also remains finite. Therefore

$$\lim\{E\sum[(x_i - p_i)^2(x_k - p_k)]/n^{3/2}\} = 0 \text{ as } n \to\infty \tag{4.3}$$

since all the products $p_1q_1(q_1 - p_1)$, $p_2q_2(q_2 - p_2)$, …, $p_{n-1}q_{n-1}\,(q_{n-1} - p_{n-1})$ are less than 1/4 in absolute value and any sum

$$\delta_{i+1} + \delta_{i+1}\,\delta_{i+2} + \delta_{i+1}\,\delta_{i+2}\,\delta_{i+3} + \ldots + \delta_{i+1}\,\delta_{i+2}\ldots\delta_n$$

remains under our conditions less than $\delta/(1 - \delta)$.

In a similar way we obtain for the third sum

$$E[(x_i - p_i)(x_k - p_k)^2] = p_i q_i[P_k^{(i)}q_k^2 + (1 - P_k^{(i)})p_k^2 - Q_k^{(i)}q_k^2 - (1 - Q_k^{(i)})p_k^2] = p_i q_i(q_k - p_k)\,\delta_{i+1}\,\delta_{i+2}\ldots\delta_k$$

and easily conclude that (4.3) is indeed valid.

It remains to consider the expectation of the last sum which consists of the products $(x_i - p_i)(x_k - p_k)(x_r - p_r)$ with $i < k < r$. The expression for the expectation of such a product can be obtained by various means. We shall base our calculations on the fact that x_i and x_r become independent one from another once the value of x_k is determined. And two values, 0 and 1, with probabilities p_k and q_k respectively, are possible for x_k. At the same time, it is not difficult to see that, when $x_k = 1$, the expectations of x_i and x_r are equal to $P_i^{(k)}$ and $P_r^{(k)}$; or, $Q_i^{(k)}$ and $Q_r^{(k)}$ respectively when $x_k = 0$.

Therefore, taking into account formulas (2.1) and (2.2), we determine that, for $i < k < r$,

$$\begin{aligned}
&E[(x_i - p_i)(x_k - p_k)(x_r - p_r)] = \\
&p_k q_k[(P_i^{(k)} - p_i)(P_r^{(k)} - p_r) - (Q_i^{(k)} - p_i)(Q_r^{(k)} - p_r)] = \\
&p_i q_i (q_k - p_k)\,\delta_{i+1} \dots \delta_k \delta_{k+1} \dots \delta_r
\end{aligned}$$

so that

$$E\sum[(x_i - p_i)(x_k - p_k)(x_r - p_r)] = \sum p_i q_i (q_k - p_k)\,\delta_{i+1} \dots \delta_k \delta_{k+1} \dots \delta_r = \sum p_i q_i C_i$$

where, respectively,

$1 \le i < k < r \le n$ for the first two, and $1 \le i \le n$ for the third sum;

$$\begin{aligned}
&C_i = \sum(q_k - p_k)\,\delta_{i+1} \dots \delta_k \delta_{k+1} \dots \delta_r = \sum(q_k - p_k)\delta_{i+1} \dots \delta_k D_k \ (i < k < r \le n), \\
&D_k = \sum \delta_{k+1} \dots \delta_r = \delta_{k+1} + \delta_{k+1}\delta_{k+2} + \dots \delta_{k+1}\delta_{k+2} \dots \delta_n.
\end{aligned}$$

And, consecutively considering D_k and C_i, we obtain simple inequalities

$$|D_k| < \delta/(1 - \delta),\ |C_i| < [\delta/(1 - \delta)]^2$$

from which we immediately arrive at

$$[|E\sum(x_i - p_i)(x_k - p_k)(x_r - p_r)|/n] < (1/4)[\delta/(1 - \delta)]^2$$

and finally

$$\lim\{E\sum[(x_i - p_i)(x_k - p_k)(x_r - p_r)]/n^{3/2}\} = 0 \text{ as } n \to \infty.$$

Thus, as $n \to \infty$, all the four ratios of $E\sum(x_i - p_i)^3$, $E\sum(x_i - p_i)^2(x_k - p_k)$, $E\sum(x_i - p_i)(x_k - p_k)^2$, and $E\sum(x_i - p_i)(x_k - p_k)(x_r - p_r)$ to $n^{3/2}$ tend to zero, and, together with them, the ratio (4.2) must also tend to zero on the strength of the abovementioned simple equality.

5. Finally, let us now pass over to general conclusions without dwelling separately on the case of $m = 4, 5, \dots$ Issuing from the generalized Newton formula, we obtain the equality for any given positive integer m

$$\begin{aligned}
&[(x_1 - p_1) + (x_2 - p_2) + \dots + (x_n - p_n)]^m = \\
&\sum G \sum (x_i - p_i)^\alpha (x_j - p_j)^\beta \dots (x_r - p_r)^\lambda,\ G = m!/(\alpha!\beta! \dots \lambda!) \qquad (5.1)
\end{aligned}$$

and the summations are over $\alpha, \beta, \dots, \lambda$ and $i, j, k, \dots, r$ respectively extending onto all the totalities of integral positive numbers

$$\alpha, \beta, \dots, \lambda \qquad (5.2)$$

satisfying the condition

$$\alpha + \beta + \ldots + \lambda = m \qquad (5.3)$$

and onto the totalities of integral positive numbers

$$i, j, k, \ldots, r \qquad (5.4)$$

obeying the inequalities

$$i < j < k < \ldots < r \leq n. \qquad (5.5)$$

It is important to remark that the number of the sums extended over $i, j, k, \ldots, r$ in (5.1) does not increase unboundedly with n but remains finite. Dwelling on one such sum, we shall consider its expectation, *i.e.*, the expectations of its terms

$$(x_i - p_i)^\alpha (x_j - p_j)^\beta (x_k - p_k)^\gamma \ldots (x_r - p_r)^\lambda. \qquad (5.6)$$

The expectation of such a product is represented by a sum of a finite number of terms of the kind

$$(\xi_i - p_i)^\alpha (\xi_j - p_j)^\beta \ldots (\xi_r - p_r)^\lambda A(B + G\delta_{i+1} \ldots \delta_j)\,(C + H\delta_{j+1} \ldots \delta_k) \ldots \qquad (5.7)$$

Here, each of the numbers $\xi_i, \xi_j, \xi_k, \ldots, \xi_r$ is equal to unity or zero, A is the probability of the equality $x_i = \xi_i$; the sum $(B + G\delta_{i+1} \ldots \delta_j)$ is the probability of the equality $x_j = \xi_j$ when the former equality is already established, and the sum $(C + H\delta_{j+1} \ldots \delta_k)$ is the probability of the equality $x_k = \xi_k$ when the equality $x_j = \xi_j$ is already established, etc. It is important to note that all the numbers $A, B, G, C, H, \ldots$ always remain finite since they are determined by the equalities

$$A = p_i \text{ or } q_i,\ B = p_j \text{ or } q_j,\ G = \pm p_i \text{ or } \pm q_i,\ C = p_k \text{ or } q_k,\ H = \pm p_j \text{ or } \pm q_j, \ldots$$

Neither is it difficult to represent the compiled expression for the expectation of the product (5.6) as a sum

$$L + M_1\delta_{i+1}\,\delta_{i+2} \ldots \delta_j + M_2\delta_{j+1}\,\delta_{j+2} \ldots \delta_k + \ldots + N_1\delta_{i+1}\,\delta_{i+2} \ldots \delta_j\,\delta_{j+1} \ldots \delta_k + \ldots + \ldots \qquad (5.7)$$

arranged according to the products

$$\delta_{i+1}\,\delta_{i+2} \ldots \delta_j,\ \delta_{j+1} \ldots \delta_k, \ldots \qquad (5.8)$$

and to their various combinations into involved products. On the strength of the explanations provided, we may state that, for any given system of exponents (5.2), all the magnitudes

$$L, M_1, M_2, \ldots, N_1, \ldots \qquad (5.9)$$

remain finite no matter how large n will be, even if $n \to \infty$ and whichever values the subscripts (5.4) only obeying the inequalities (5.5) will take.

Pursuing a definite purpose, we shall not study the composition of the expressions (5.9); we shall only indicate that the elimination from the sum (5.7) of all the terms containing a definite product from our totality of products (5.8) is equivalent to severing the chain of magnitudes

$$x_i, x_j, x_k, \ldots, x_r \tag{5.10}$$

at the respective place. Instead of one chain we shall thus obtain two chains, and the expectation under consideration will be replaced by a product of two expectations.

If, however, only those terms that do not contain some of the products (5.8) are left in the sum (5.7), the chain of the magnitudes (5.10) will be separated into several isolated chains, and instead of the expectation under consideration we shall obtain a product of several expectations. For example, when eliminating all the terms containing the factor $\delta_{j+1} \ldots \delta_k$ from (5.7), the remaining sum will be equal to

$$\mathrm{E}\,(x_i - p_i)^\alpha (x_j - p_j)^\beta\, \mathrm{E}(x_k - p_k)^\gamma \ldots (x_r - p_r)^\lambda.$$

If, however, we only leave there those terms that do not contain either $\delta_{i+1} \ldots \delta_j$ or $\delta_{j+1} \ldots \delta_k$, we shall obtain, instead of the expectation considered,

$$\mathrm{E}\,(x_i - p_i)^\alpha\, \mathrm{E}(x_j - p_j)^\beta\, \mathrm{E}(x_k - p_k)^\gamma \ldots (x_r - p_r)^\lambda.$$

Finally, the first term *L,* to which our sum is reduced after rejecting all those containing the factors (5.8), is of course equal to the product of the expectations of the separate exponents,

$$\mathrm{E}\,(x_i - p_i)^\alpha\, \mathrm{E}(x_j - p_j)^\beta \ldots \mathrm{E}(x_r - p_r)^\lambda.$$

We shall make use of the deliberations about partitioning the chain into several separate chains for determining the main term of the expression

$$\mathrm{E}\sum(x_i - p_i)^\alpha\, (x_j - p_j)^\beta\, (x_k - p_k)^\gamma \ldots (x_r - p_r)^\lambda, \tag{5.11}$$

i.e., of that term that plays the decisive role in our investigation.

However, as a preliminary, we shall establish the formula for expectation (5.11): it is equal to

$$\begin{aligned}&\sum L + \sum M_1\delta_{i+1}\, \delta_{i+2} \ldots \delta_j + \ldots +\\ &\sum N_1\delta_{i+1}\, \delta_{i+2} \ldots \delta_j\, \delta_{j+1} \ldots \delta_k + \ldots + \ldots\end{aligned} \tag{5.12}$$

This follows from the expression (5.7) for the expectation of the product (5.6). The sums extend over all possible values of the subscripts satisfying the inequalities (5.5).

We can accomplish the summing in several stages, by transforming it into multiple sums, only extending it each time over one subscript. Arranging these separate operations in the order of decreasing subscripts $r > \ldots > k > j > i$ so as to compile the appropriate sum, we must assign all possible values exceeding the next subscript but not exceeding n, to each subscript occupying a definite place. The number of sums is equal to the number of exponents (5.2).

Turning now to the problem of the possible rapidity of the increase in our multiple sums comprising the formula (5.12) as $n \to \infty$, we may, with respect to each of them separately, partition the just mentioned consecutive summings over the subscripts $r, \ldots, k, j, i$ into two groups in such a manner, that, as far as our purpose is concerned, all the operations in the first one may be compared with repeating the same number n times, and the summings in the second group,– with a geometric progression with common ratio δ. And it is not difficult to conclude from here that the ratio of the considered multiple sum to the power of $n,$ equal to the number of the appropriate sums in the first group, must remain finite as $n \to \infty$. In particular, for $\sum L$ we should attribute all the summations over the separate subscripts to the first group, and, consequently, in those cases in which we cannot establish that $L = 0$, we may only maintain that the ratio $\sum L/n^\sigma$, where σ is the number of the exponents (5.2), must remain finite.

For any other sum in (5.12) the number of operations in the first group is less than σ so that its ratio to n^σ must tend to zero as $n \to \infty$; for example, the sum

$$\sum M_1\delta_{i+1}\ \delta_{i+2}\ \ldots\ \delta_j$$

taken over the subscript j should be attributed to the second group and there will be left only $(\sigma - 1)$ summings in the first one, and

$$\sum N_1\delta_{i+1}\ \delta_{i+2}\ \ldots\ \delta_j\ \delta_{j+1}\ \ldots\ \delta_k$$

taken either over j or over k should be attributed to the second group, so that only $(\sigma - 2)$ summings will be left in the first one.

We may therefore state that the difference

$$[\mathrm{E}\sum(x_i - p_i)^\alpha(x_j - p_j)^\beta\ (x_k - p_k)^\gamma\ \ldots\ (x_r - p_r)^\lambda/n^\sigma] - \sum L/n^\sigma$$

must tend to zero as $n \to \infty$. At the same time,

$$\sum L = \sum \mathrm{E}(x_i - p_i)^\alpha\ \mathrm{E}(x_j - p_j)^\beta\ \mathrm{E}(x_k - p_k)^\gamma\ \ldots\ \mathrm{E}(x_r - p_r)^\lambda$$

will always be the main term of the expression (5.12) excepting such cases in which some of the exponents (5.2) are unity. In these instances, which we shall now consider, the sum $\sum L$ vanishes since all of its terms are zero.

6. Dwelling now on the case in which some of the exponents (5.2) are unity, we introduce for the sake of brevity the following expressions:

1) For subscripts $i, j, k, \ldots, r$ satisfying inequalities (5.5) the products (5.8) will be the coefficients of connection between i and j, j and k, … respectively.

2) We shall call the terms of (5.2), appearing in the product (5.6), the exponents of the subscripts (5.4).

These expressions allow us to formulate, concisely and clearly, the following proposition: *If the exponent of any of the subscripts* (5.4), *satisfying inequalities* (5.5), *in the product* (5.6) *is unity, then each term of the expression* (5.7) *for the expectation of* (5.6) *contains at least one coefficient of connection of this subscript with its adjacent subscripts*. In particular, if $\alpha = 1$, *expression* (5.7) *must include as a common multiplier the coefficient of connection between i and the next subscript, j, equal to the first product in* (5.8); *for* $\lambda = 1$, *the common multiplier will be the coefficient of connection between r and the directly preceding subscript*.

This proposition can be proved briefly and simply by means of our previous reasoning concerning the partition of the chain (5.10) into several isolated chains. Indeed, after isolating one subscript from among (5.4), and excluding all terms containing the coefficients of connection of this subscript with its adjacent ones from expression (5.7), the remaining sum will represent, as we have already remarked, not the expectation of the product (5.6), but rather the product of three, or two expectations. And one of these will be a term of the series

$$\mathrm{E}(x_i - p_i)^\alpha, \mathrm{E}(x_j - p_j)^\beta, \mathrm{E}(x_k - p_k)^\gamma, \ldots, \mathrm{E}(x_r - p_r)^\lambda \qquad (6.1)$$

corresponding to the isolated subscript.

Our remark concerns all cases. If, however, the exponent of the selected subscript is unity, the corresponding term of the series (6.1) is zero because

$$E(x_i - p_i) = E(x_j - p_j) = E(x_k - p_k) = \ldots = E(x_r - p_r) = 0.$$

And, of course, our product of the three expectations, equal to the sum of all the terms of expression (5.7) excepting those excluded by us, also vanishes. It follows that in this case expression (5.7) only consists of these eliminated terms which include as a multiplier at least one of the coefficients of connection between the selected subscript and its adjacent subscripts. Thus, our proposition is proved.

It will serve as a basis for our further conclusions about the sums composing (5.12) for the special case in which some of the exponents (5.2) are unity. Concerning these sums we have already indicated that they are reduced to consecutive summings over separate subscripts arranged in a decreasing order. As to these summings, we have separated them in two groups comparing the operations in the first one with a repetition of the same number, and the summations in the second group with a decreasing geometric progression.

We have accordingly concluded that the ratio of each of our sums of (5.12) to the power of n, equal to the number of the corresponding summations in the first group, must remain finite as $n \to \infty$. Of course, the number of exponents (5.2) can serve in all cases as the highest boundary for the number of these summations. For our purpose, this boundary is sufficient if only among (5.2) there are no exponents equal to unity. To specify: if under condition (5.3) $\alpha = \beta = \gamma = \ldots = \lambda = 2$, then the number of exponents (5.2) is $m/2$ and we can establish that, as $n \to \infty$,

$$\lim\{[E\sum(x_i - p_i)^2(x_j - p_j)^2(x_k - p_k)^2 \ldots (x_r - p_r)^2/n^{m/2}] - [Ec_{ii}c_{jj}\ldots c_{rr}/n^{m/2}]\} = 0 \qquad (6.2)$$

where $i < j < \ldots < r < n$,

$$c_{ii} = E(x_i - p_i)^2,\ c_{jj} = E(x_j - p_j)^2,\ \ldots,\ c_{rr} = E(x_r - p_r)^2.$$

If among the exponents (5.2) none are equal to unity, and, moreover, if some of them are larger than 2, then their number is less than $m/2$ and, on the same grounds, as $n \to \infty$,

$$\lim[E\sum(x_i - p_i)^{\alpha}(x_j - p_j)^{\beta}(x_k - p_k)^{\gamma} \ldots (x_r - p_r)^{\lambda}/n^{m/2}] = 0. \qquad (6.3)$$

If, however, some of the exponents (5.2) are unity, we must establish another highest boundary smaller than their number for the number of the summings in the first group.

In other words, we must now establish some lowest boundary differing from zero for the number of summings in the second group. And, since for each term this number is equal to the number of the included coefficients of connection, our problem is reduced to determining the lowest boundary of this latter. In order to establish such a boundary, we shall consider the totality of the coefficients of connection included in any of the non-vanishing terms of (5.12), or, which is the same, in its corresponding terms in expression (5.7).

On the strength of the proven proposition, for each subscript whose exponent is unity, there must be in this totality at least one coefficient of connection with its adjacent subscripts. And one common coefficient can only correspond to two such subscripts if the series of subscripts (5.4) arranged in an increasing order contains no intermediate subscripts, *i.e.*, if these two are next to each other. It is not difficult to conclude now, that for each term of (5.12), not vanishing but really included into it, the number of summations of the second group is not less than one half the number of all the exponents (5.2) equal to unity and can only be equal to this magnitude when all the subscripts whose exponent is unity are partitioned into separate pairs of adjacent subscripts, and the term selected by us contains all the coefficients of connection of these pairs considered separately, but does not contain any other coefficients.

If, therefore, the totality of the exponents (5.2) consists of α' unities, β' twos, γ', δ', … numbers equal to 3, 4, … respectively, then

$$m = \alpha + \beta + \gamma + \ldots + \lambda = \alpha' + 2\beta' + 3\gamma' + 4\delta' + \ldots$$

and we may choose

$$(\alpha'/2) + \beta' + \gamma' + \delta' + \ldots \qquad (6.4)$$

as the highest boundary of the number of summings in the first group for all the summings considered; it is only attained under the indicated conditions.

On the other hand, it is not difficult to see that, if only not all the numbers γ', δ', … are zeros, the sum (6.4) is less than half the sum $\alpha' + 2\beta' + 3\gamma' + 4\delta' + \ldots$ Therefore, we may extend equality (6.3) onto all the totalities of the exponents (5.2) satisfying the condition (5.3) and not solely consisting of twos and pairs of adjacent unities.

7. Having thus extended the equality (6.3) and issuing from it and formulas (5.1) and (5.12), we immediately conclude that for odd values of m the ratio

$$\mathrm{E}[(x_1 - p_1) + (x_2 - p_2) + \ldots + (x_n - p_n)]^m/n^{m/2}$$

must tend to zero as $n \to \infty$. Indeed, if the sum m of exponents (5.2) is odd, their totality cannot only consist of pairs of adjacent unities and of twos. And, since the ratio (3.6), remaining finite, cannot be arbitrarily small (§3), the conclusion just mentioned might be represented, as $n \to \infty$, by a simple formula, valid for any odd positive m

$$\lim \mathrm{E}\left(\frac{x_1 - p_1 + x_2 - p_2 + \ldots + x_n - p_n}{\sqrt{E2(x_1 - p_1 + x_2 - p_2 + \ldots + x_n - p_n)^2}}\right)^m = (1/\sqrt{\pi}) \int_{-\infty}^{\infty} t^m \exp(-t^2)dt \qquad (7.1)$$

whose right side vanishes when m is odd.

To arrive at the same formula (7.1) for even powers of m as well, we ought to determine definitively the main part of (5.11) for totalities of exponents (5.2) which only consist of twos and of separate pairs of adjacent unities. According to §6, the main part of (5.12) for such totalities of exponents (5.2) will be that, which includes all the coefficients of connection of the separate pairs and does not include any other coefficients; the number of summations in the first group in this term is indeed equal to $m/2$.

And, on the strength of the earlier reasoning on the severance of the chain into several chains, it is not difficult to conclude that, in the cases considered by us, this term, which is the main part of (5.11), is equal to the sum of the products adequately composed of the expectations of the squares of some of the differences $(x_i - p_i)$, $(x_j - p_j)$, …, $(x_r - p_r)$ and of the expectations of the pairwise products of some other differences.

This conclusion might be represented, as $n \to \infty$, by the following formula:

$$\lim\{\mathrm{E}[\sum(x_e - p_e)(x_f - p_f)(x_g - p_g)\,(x_h - p_h)\ldots/n^{m/2}] -$$

$$[\sum c_{ef}\, c_{gh} \ldots /n^{m/2}]\} = 0 \qquad (7.2)$$

where

$$c_{ef} = \mathrm{E}(x_e - p_e)(x_f - p_f),\ c_{gh} = \mathrm{E}(x_g - p_g)(x_h - p_h),\ \ldots$$

and the subscripts

$$e, f, g, h, \ldots \qquad (7.3)$$

whose number in each product

$$(x_e - p_e)(x_f - p_f)(x_g - p_g)\,(x_h - p_h) \ldots \qquad (7.4)$$

is m, are only restricted by the inequalities

$$e \le f < g \le h < \ldots \le n. \qquad (7.5)$$

The duality of the signs $\le$ is eliminated when all the square multipliers of the products (7.4) are ascertained.

Formula (7.2) incorporates the previous equality (6.2) as a particular case when $e = f$, $g = h$, … For the sake of convenience we shall additionally introduce new notation:

$$c_{ii} = \mathrm{E}(x_i - p_i)^2 = D_{ii},\ 2c_{ij} = \mathrm{E}(x_i - p_i)(x_j - p_j) = D_{ij}$$

where $i < j$; the second subscript j in the expression for D_{ij} will never be less than the first one, i. Making use of this new notation, we can first of all represent the expectation of the square

$$[(x_1 - p_1) + (x_2 - p_2) + \ldots + (x_n - p_n)]^2$$

as the sum $\sum D_{ij}$ extended over all subscripts i and j obeying the inequalities $i \le j \le n$. Unlike it was in formula (7.2), the sign $\le$ between i and j is not fixed but remains dual.

Then, on the grounds of formulas (5.1), (6.3) and (7.2) it is not difficult to establish the following formula {as $n \to \infty$} for even values of m

$$\lim\{\mathrm{E}[(x_1 - p_1) + (x_2 - p_2) + \ldots + (x_n - p_n)]^m/n^{m/2}] - [m!/(2n)^{m/2}]\sum D_{ef}\, D_{gh} \ldots\} = 0 \qquad (7.6)$$

where all the values satisfying the inequalities (7.5) should be assigned to the subscripts (7.3) whose number in each product $D_{ef}\, D_{gh} \ldots$ is m. Once more, the signs $\le$ between e and f, g and h, … are not fixed but remain dual as in the just mentioned sum $\sum D_{ij}$.

It remains to compare the sum

$$\sum D_{ef}\, D_{gh} \ldots = S \qquad (7.7)$$

that appears in formula (7.6) with the power $(\sum D_{ij})^{m/2}$. This power contains our sum S with a coefficient $(m/2)!$ as well as other products of the same kind,

$$D_{ef}\, D_{gh} \ldots \qquad (7.8)$$

which do not, however, obey our inequalities (7.5).

We ought to prove that the ratio of the sum of these products, taken of course with the coefficients adequate for composing the difference

$$(\sum D_{ij})^{m/2} - (m/2)!S, \qquad (7.9)$$

to $n^{m/2}$ tends to zero as $n \to \infty$. Considering the totality of the products (7.8) making up the difference (7.9), we shall establish a certain order among the factors so as to avoid their multiple repetition and to ascertain the composition of the studied totality. We shall always arrange the first subscripts $e, g, \ldots$ in an increasing order; and, if they coincide, we shall arrange the second subscripts in the same way.

Having thus established the order of the multipliers of the product (7.8), we can easily solve the question of whether this product is included in the sum S, or, on the contrary, whether it belongs to that totality which makes up the difference (7.9) and which we ought to study. If the product (7.8) belongs to the totality under study, then one of the second subscripts $f, h, \ldots$ must be not less than the subsequent first subscript; for example, $f \ge g$.

We shall make use of this remark for composing the sum which includes all the terms of the difference (7.9) and which therefore provides,– after replacing the magnitudes D_{ij} by their absolute values, $|D_{ij}|$,– a number certainly larger than this difference. Consider the totality of those products (7.8), which include one and the same multiplier D_{rs}, such that the second subscript of the next factor is not larger than s. It is not difficult to note that this totality constitutes a part of the one that follows after fulfilling the operations indicated in the expression

$$(m/2)[(m/2) - 1]D_{rs}(\sum D_{ij})^{m/2-2}(D_{rr} + D_{r\,r+1} + D_{r\,r+2} + \ldots + D_{r+1\,r+1} +$$
$$D_{r+1\,r+2} + \ldots + D_{ss} + D_{s\,s+1} + D_{s\,s+2} + \ldots).$$

Its first factor, $(m/2)[(m/2) - 1]$, determines the coefficient with which our products might enter in the first term of (7.9). Replacing in this expression all the D_{ij} by their absolute values, and summing the obtained expressions over all possible subscripts r and s, we arrive at the sum which we shall denote by Ω and which certainly exceeds the absolute value of the difference (7.9). On the other hand (§3), we have inequalities

$$|D_{rs}| < \delta^{s-r}/2,\ \sum|D_{ij}| < n\,(1+ \delta/[4(1-\delta)],$$
$$|D_{rr}| + |D_{r\,r+1}| + |D_{r\,r+2}| + \ldots < 1/[2(1-\delta)], \ldots$$
$$|D_{ss}| + |D_{s\,s+1}| + |D_{s\,s+2}| + \ldots < 1/[2(1-\delta)]$$

so that

$$\Omega < (m/2)[(m/2) - 1]\frac{n}{4(1-\delta)}\left(\frac{n(1+\delta)}{4(1-\delta)}\right)^{m/2-2}(1 + 2\,\delta + 3\,\delta^2 + 4\,\delta^3 + \ldots) <$$
$$(m/2)[(m/2) - 1]\frac{n}{4(1-\delta)^3}\left(\frac{n(1+\delta)}{4(1-\delta)}\right)^{m/2-2}.$$

Consequently, the absolute value of the difference (7.9) must also be less than the last expression just above and we can therefore establish the formula

$$\lim\{[(\sum D_{ij})^{m/2}/n^{m/2}] - [(m/2)!S/n^{m/2}]\} = 0 \ \{\text{as } n \to \infty\} \qquad (7.10)$$

where S is defined by (7.7).

It remains to compare formulas (7.10) and (7.6) and to take into account the well-known equality

$$(1/\sqrt{\pi})\int_{-\infty}^{\infty} t^{m}\exp(-t^2)\,dt = [1\cdot3\cdot5\ \dots\ (m-1)/2^{m/2}]$$

to obtain immediately, for any even m, the formula (7.1) which we have previously established for odd exponents m. Having ascertained formula (7.1) for all integral exponents m, we may issue from it and conclude, as it is known, that, for any given numbers t_1 and $t_2 > t_1$ and $n \to \infty$, the probability of the persistence of the inequalities

$$t_1 < \frac{x_1 - p_1 + x_2 - p_2 + \dots + x_n - p_n}{\sqrt{E2(x_1 - p_1 + x_2 - p_2 + \dots + x_n - p_n)^2}} < t_2,$$

or, which is the same, of

$$t_1 < \frac{m - p_1 - p_2 - \dots - p_n}{\sqrt{E2(m - p_1 - p_2 - \dots - p_n)^2}} < t_2,$$

as $n \to \infty$ tends to

$$(1/\sqrt{\pi})\int_{t_1}^{t_2} \exp(-t^2)\,dt.$$

As to the expectation of the square

$$[m - (p_1 + p_2 + \dots + p_n)]^2, \qquad (7.11)$$

it is determined by the previously established formula (3.4).

We are thus extending the theorem about representing the limit of probability by the Laplace integral onto the general case of trials connected into a chain under some restrictions that are indicated above and occasioned by the course of our reasoning.

8. In concluding our article, we return to the expectation of (7.11), which, in accord with formula (3.4), is equal to

$$p_1 q_1 (1 + 2\delta_2 + 2\delta_2\delta_3 + \dots 2\delta_2\,\delta_3 \dots \delta_n) +$$
$$p_2 q_2 (1 + 2\delta_3 + 2\delta_3\delta_4 + \dots 2\delta_3\,\delta_4 \dots \delta_n) + \dots + p_{n-1}\,q_{n-1}\,(1 + 2\delta_n) + p_n\,q_n.$$

We shall try to ascertain that the condition put forth in §3 does not introduce any new restrictions except those indicated in the introduction. For this purpose we deduce from formula (1.2) simple equalities

$$p_{k+1}\,q_{k+1} = (p''_{k+1} + \delta_{k+1}\,p_k)\,(q'_{k+1} + \delta_{k+1}\,q_k) =$$
$$\delta^2_{k+1}\,p_k\,q_k + \delta_{k+1}(p_k\,q'_{k+1} + q_k\,p''_{k+1}) + p''_{k+1}\,q'_{k+1}$$

and

$$p_{k+1}\,q_{k+1} = (p'_{k+1} - \delta_{k+1}\,q_k)\,(q''_{k+1} - \delta_{k+1}\,p_k) =$$
$$\delta^2_{k+1}\,p_k\,q_k - \delta_{k+1}\,(p_k\,p'_{k+1} + q_k\,q''_{k+1}) + p'_{k+1}\,q''_{k+1}.$$

Now we infer that the difference $[p_{k+1}\, q_{k+1} - \delta^2_{k+1}\, p_k\, q_k]$ is not less than one of the products $p''_{k+1}\, q'_{k+1}$ and $p'_{k+1}\, q''_{k+1}$ and cannot therefore be arbitrarily small, if only, as stated in the introduction, among the numbers

$$p'_{k+1},\ p''_{k+1},\ q'_{k+1},\ q''_{k+1} \qquad (8.1)$$

not a single one is arbitrarily small. Taking this into account, we denote

$$p_{k+1}\, q_{k+1} - \delta^2_{k+1}\, p_k\, q_k = \Delta_k$$

and replace consecutively, in the left side of equality (3.4),

$p_n\, q_n$, by an equal magnitude $\delta^2_n\, p_{n-1} + \Delta_{n-1}$,
$p_{n-1}\, q_{n-1}$, by an equal magnitude $\delta^2_{n-1}\, p_{n-2}\, q_{n-2} + \Delta_{n-2}, \ldots$,
$p_2\, q_2$, by an equal magnitude $\delta^2_2\, p_1\, q_1 + \Delta_1$.

We thus transform formula (3.4) into

$$\mathrm{E}[m - (p_1 + p_2 + \ldots + p_n)]^2 = T_n^2\Delta_{n-1} + T_{n-1}^2\Delta_{n-2} + \ldots + T_2^2\Delta_1 + T_1^2\Delta_0$$

where

$$\Delta_0 = p_1 q_1,$$
$$T_n = 1,\ T_{n-1} = 1 + \delta_n = 1 + \delta_n T_n,\ T_{n-2} = 1 + \delta_{n-1} + \delta_{n-1}\,\delta_n = 1 + \delta_{n-1}\, T_{n-1}$$

and in general

$$T_{k-1} = 1 + \delta_k T_k = 1 + \delta_k + \delta_k\,\delta_{k+1} + \ldots + \delta_k\,\delta_{k+1} + \delta_n.$$

We also have

$$T^2_{k+1} + T^2_k = (1 + \delta_k^2)T^2_k + 2\delta_k T_k + 1 =$$
$$(1 + \delta_k^2)\{T_k + [\delta_k/(1 + \delta_k^2)]\}^2 + [1/(1 + \delta_k^2)],$$

from which we conclude that for any subscript k the sum in the left side cannot be arbitrarily small, but must always remain not less than $[1/(1 + \delta^2)]$. Therefore, if only the conditions indicated in the introduction are fulfilled, the ratio

$$\mathrm{E}[m - (p_1 + p_2 + \ldots + p_n)]^2/n$$

cannot be arbitrarily small either.

Thus, the conclusions of this article do not demand other restrictions except those indicated in the introduction. They consist in that the numbers (8.1), at all values of k, remain larger than some constant positive number. The question about the possible relaxation of the restrictions imposed by us remains open.

Notes and Comment by N. A. Sapogov [8, pp. 666 – 668]

1. We think that this term can describe the chains that we have considered in contrast to chains of other dependent magnitudes.
2. We shall remove this doubt at the end of our paper.

Markov proves by the method of moments that the CLT is applicable to the sums ($X_1 + X_2 + + \dots + X_n$) of random variables X_i connected into a simple *non-homogeneous* chain under the assumption that the possible values of X_i are 0 and 1 and

$$0 < p_0 < p'_i < 1 - p_0,\ 0 < p_0 < p''_i < 1 - p_0$$

where p'_i and p''_i are the transition probabilities

$$p'_i = P(X_i = 1 / X_{i-1} = 1),\ p''_i = P(X_i = 1 / X_{i-1} = 0)$$

and p_o is a constant number not depending on n.

Markov left open the question of whether the CLT was applicable to the sums of X_i if p_o can approach zero infinitely. Bernstein [1] arrived at the same result by another method and proved that the CLT persisted if $p_o = 1/n^\alpha$ with the constant $\alpha < 1/5$, but that it was perhaps non-applicable for $\alpha = 1/3$. In two other writings he [2; 3] proved the same proposition for the case of $\alpha < 1/3$.

I [9] have shown the applicability of the CLT under the condition that $p_o = \varphi(n)/n^{1/3}$ where $\varphi(n)$ was an arbitrary function infinitely increasing as $n \to \infty$. Linnik [4] considered a more general case of a non-homogeneous chain of restricted variables X_i with an arbitrary finite number of possible values $x_i^{(k)}$,
$1 \le k \le k_i$ for each X_i and proved that the CLT is applicable if all the transition probabilities differ from 0 and 1 not less than by $1/n^\alpha$ with the constant $\alpha < 1/3$ and

$$(1/k_i)\sum_{k=1}^{k_i} [x_i^{(k)} - (1/k_i)\sum_{k=1}^{k_i} x_i^{(k)}]^2 \ge c > 0.$$

I [10] obtained a similar result also for the case of chains of h-dimensional vectors.

References

1. Bernstein, S.N. (1926), Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Math. Annalen*, Bd. 97, pp. 1 – 59.
2. Bernstein, S.N. (1926), Sur les sommes de quantités dépendantes. *Izvestia Akad. Nauk SSSR*, ser. 6, vol. 20, pp. 1459 – 1478.
3. Bernstein, S.N. (1928), Addition à l'article … [to the previous paper]. *Doklady Akad. Nauk SSSR*, ser. A, pp. 55 – 60.
4. Linnik, Yu.V. (1949), *Izvestia Akad. Nauk SSSR*, ser. math., vol. 13, pp. 65 – 94.
5. Markov, A.A. (1906), The extension of the law of large numbers, etc. Translated in this book.
6. Markov, A.A. (1907, in Russian), Recherches sur un cas remarquable d'épreuves dépendantes. *Acta math.*, t. 33, 1910, pp. 87 – 104.
7. Markov, A.A. (1908), The extension of the limit theorem, etc. Translated in this book.
8. Markov, A.A. (1951), *Избранные труды* (Sel. Works). N.p.
9. Sapogov, N.A. (1947), *Doklady Akad. Nauk SSSR*, No. 2, vol. 58, pp. 193 – 196.
10. Sapogov, N.A. (1949), Ibidem, No. 2, vol. 69, pp. 133 – 135.

14. A.A. Markov On the Coefficient of Dispersion

Translator's Foreword

As Markov indicated in the beginning of his paper, the first to consider the expectation of the coefficient of dispersion was Chuprov. Indeed, in 1916 Chuprov informed Markov in a private letter that he had proved that EQ = 1. Markov, however, was the first to publish this result and in a few months he communicated Chuprov's relevant contribution to a periodical of the Petersburg Academy of Sciences, see pp. 112 – 113 of my book *Chuprov. Life, Work, Correspondence.* Göttingen, 1996. Chuprov's paper appeared in my translation in Chuprov, *Statistical Papers and Memorial Publications*. Berlin, 2004, pp. 39 – 47.

* * *

[1] Without dwelling on the importance of the coefficient of dispersion for statistics, I aim here, first, to prove the proposition stated by Professor Chuprov, that its expectation for independent trials with a constant probability is exactly equal to unity provided that we do not extract the square root {of the coefficient} but adhere to the definition adopted in my book [3]; and, second, to establish, for the case of an equal number of trials {in each series}, a rather simple expression for the expectation of the square of the deviation of this coefficient from unity. The expectation thus obtained is somewhat in excess of its exact value [1].

I connect the proposition about the expectation of the coefficient of dispersion with the name of Professor Chuprov because, as far as I know, he was the first to consider not the numerator and the denominator of this fractional expression taken separately one from another, but the fraction itself, and arrived at the abovementioned conclusion, at least for the case of equal series. As to the second problem, Professor L. Bortkiewicz had long ago discovered its solution of sorts, but it was connected with such assumptions, which we, taking care of rigor and clarity of definitions, and of absolute rigor of deductions, cannot admit.

[2] Let us consider the general case of several series of independent trials with a constant probability. Denote the series, σ in number, by 1, 2, …, σ; the number of trials in series i by s_i; the corresponding number of appearances of event E by x_i with

$$n = s_1 + s_2 + \dots + s_\sigma \text{ and } m = x_1 + x_2 + \dots + x_\sigma$$

and, finally, the constant probability of event E by p, and let the coefficient of dispersion for the considered totality of series be

$$Q = \frac{n(n-1)\sum s_i[(x_i/s_i) - (m/n)]^2}{(\sigma-1)m(n-m)} = \frac{(n-1)[\sum (nx_i^2/s_i) - m^2]}{(\sigma-1)m(n-m)},$$

$$Q^2 = \frac{(n-1)^2}{(\sigma-1)^2}\cdot$$
$$\frac{\sum[n^2x_i^4/s_i^2] + 2\sum[n^2x_i^2x_j^2/s_is_j] - 2m^2\sum[nx_i^2/s_i] + m^4}{m^2(n-m)^2}.$$

These expressions lose their meaning if $m = 0$, or if $m = n$, when both their numerators and denominators vanish. In these exceptional cases we shall assume that $Q = Q^2 = 1$. To calculate the expectations of Q and Q^2 we must multiply their expressions by the probability P of the totality of numbers

$$x_1, x_2, \dots, x_\sigma \qquad (1)$$

and compile the sums $\sum PQ$ and $\sum PQ^2$ for all possible totalities (1). The probability P is equal to

$$P = \frac{s_1!}{x_1!(s_1-x_1)!}\cdot\frac{s_2!}{x_2!(s_2-x_2)!}\ \dots\ \frac{s_\sigma!}{x_\sigma!(s_\sigma-x_\sigma)!}p^m q^{n-m}$$

where $q = 1 - p$.

We shall break up the summation of PQ and PQ^2, which should not be confused with the operations $\sum_i$ or $\sum_{i,j}$ (as in the formulas above), in two stages. For the first summation we shall suppose that m is invariable, and we shall only vary this number alone during the second one. This sequence of operations can be written down as

$\sum\sum PQ$ and $\sum\sum PQ^2$ with $m = 0, 1, 2, \dots, n$ for the external sums and
$x_1 + x_2 + \dots + x_\sigma = m$ for the internal ones.

During the first summation, the denominators $m(n-m)$ and $m^2(n-m)^2$ of the expressions Q and Q^2 respectively remain constant, so that the problem of determining the inner sums reduces to calculating

$$\sum P,\ \sum Px_i^2,\ \sum Px_i^4,\ \sum Px_i^2x_j^2 \qquad (2)$$

where i and j denote some two indices taken from our system $1, 2, \dots, \sigma$ and remaining invariable during the summation.

To determine all these sums we introduce $(\sigma + 1)$ arbitrary magnitudes $\xi_1, \xi_2, \dots, \xi_\sigma, t$ and construct their function

$$W = [pt\exp(\xi_1) + q]^{s_1} \dots [pt\exp(\xi_i) + q]^{s_i} \dots [pt\exp(\xi_\sigma) + q]^{s_\sigma}$$

which can be decomposed into terms

$$Pt^{x_1+x_2+\dots+x_\sigma}e^{x_1\xi_1+x_2\xi_2+\dots+x_\sigma\xi_\sigma}.$$

To determine the first of the sums in (2) by means of W, it is only necessary to equate all the ξ's to zero; develop the thus obtained expression $W = (pt + q)^n$ in powers of t; and take the coefficient of t^m, equal to $C_n^m p^m q^{n-m}$.

In a similar way, the coefficients of t^m in the expressions

$$\partial^2 W/\partial\xi_i^2,\ \partial^4 W/\partial\xi_i^4,\ \partial^4 W/\partial\xi_i^2\partial\xi_j^2 \text{ at } \xi_1 = \xi_2 = \dots = \xi_\sigma = 0$$

will be equal to the other sums in (2) respectively. We must only consider these sums for

$$0 < m < n \qquad (3)$$

since, in accord with the definition adopted,

$$\sum PQ = \sum PQ^2 = q^n \text{ if } m = 0 \text{ and } p^n \text{ if } m = n.$$

To calculate the expectation of Q, we begin with the first of these sums. We have, for $\xi_1 = \xi_2 = \ldots = \xi_\sigma = 0$,

$$\partial^2 W / \partial \xi_i^2 = s_i pt(pt + q)^{n-1} + s_i(s_i - 1)\, p^2 t^2 (pt + q)^{n-2}$$

so that, for $x_1 + x_2 + \ldots + x_\sigma = m$,

$$\sum P(nx_i^2/s_i) = [C_{n-1}^{m-1} + (s_i - 1)\, C_{n-2}^{m-2}]np^m q^{n-m}.$$

Consequently,

$$\sum P[(nx_1^2/s_1) + (nx_2^2/s_2) + \ldots + (nx_\sigma^2/s_\sigma) - m^2] =$$
$$\frac{(n-2)!n}{(m-1)!(n-m)!}[\sigma(n - 1) + (n - \sigma)(m - 1) - m(n - 1)]p^m q^{n-m} =$$
$$n(\sigma - 1)\, C_{n-2}^{m-1} p^m q^{n-m}$$

and $\sum PQ = C_n^m p^m q^{n-m}$.

We have deduced this formula under condition (3) and it was previously established for $m = 0$ and $m = n$. Carrying out the second summing, we obtain Professor Chuprov's proposition

$$\sum \sum PQ = (p + q)^n = 1.$$

Turning now to the expectation of $(Q - 1)^2$, that, on the strength of the proven is equal to the difference

$$\mathrm{E}Q^2 - \sum_{m=0}^{n} C_n^m p^m q^{n-m},$$

we restrict our attention to the case in which all the numbers $s_1, s_2, \ldots, s_\sigma$ take one and the same value s. In this particular case, as shown by direct calculation [2], Q is reduced to unity when $m = 1$ and $m = n - 1$. Therefore, not only the highest powers of p and q, but also the products $p^{n-1}q$ and $p\, q^{n-1}$ vanish in the expression $\mathrm{E}Q^2 - (p + q)^n$. And, because of the formulas (where $\xi_1 = \xi_2 = \ldots = \xi_\sigma = 0$)

$$(1/s)\, \partial^4 W / \partial \xi_i^4 = pt(pt + q)^{n-1} + 7(s - 1)p^2t^2(pt + q)^{n-2} +$$
$$6(s - 1)(s - 2)\, p^3t^3(pt + q)^{n-3} + (s - 1)(s - 2)(s - 3)p^4t^4(pt + q)^{n-4}$$

and

$$(1/s^2)\, \partial^4 W / \partial \xi_i^2 \partial \xi_j^2 =$$
$$p^2t^2(pt + q)^{n-2} + 2p^3t^3(s - 1)(pt + q)^{n-3} + (s - 1)^2 p^4t^4(pt + q)^{n-4},$$

we find without much difficulties that, for $x_1 + x_2 + \ldots + x_\sigma = m$,

$$\sum P(Q^2-1) = \frac{(n-4)!(n-1)np^m q^{n-m} R^{(m)}}{(\sigma-1)^2 (m-1)! m^2 (n-m)!(n-m)^2},$$

$$R^{(m)} = \sigma^2(n-3)(n-2)(n-1)^2 + 7\sigma^2(s-1)(n-3)(n-2)(n-1)(m-1) +$$
$$6\sigma^2(s-1)(s-2)(n-3)(n-1)(m-1)(m-2) +$$
$$\sigma^2(s-1)(s-2)(s-3)(n-1)(m-1)(m-2)(m-3) +$$
$$n\,\sigma(\sigma-1)(n-3)(n-2)(n-1)(m-1) +$$
$$2n\,\sigma(\sigma-1)(s-1)(n-3)(n-1)(m-1)(m-2) +$$
$$n\,\sigma(\sigma-1)(s-1)^2(n-1)(m-1)(m-2)(m-3) -$$
$$2m^2\sigma(n-3)(n-2)(n-1)^2 - 2m^2(n-\sigma)(n-3)(n-2)(n-1)(m-1) +$$
$$m^3(n-3)(n-2)(n-1)^2 - (\sigma-1)^2 m(n-m)^2(n-3)(n-2).$$

To facilitate the study of the rather involved expression for $R^{(m)}$ it is possible to take advantage of the circumstance that Q does not change its value when all the x_i's are replaced by the differences$(s - x_i)$ so that m is replaced by $(n - m)$. Accordingly, we have

$$p^{n-m}q^m \sum P(Q^2-1) = p^m q^{n-m} \sum P(Q^2-1)$$

where $x_1 + x_2 + \ldots + x_\sigma$ is equal to m and $(n - m)$ respectively, and therefore

$$mR^{(m)} = (n-m)R^{(n-m)}.$$

It follows, first of all, that $R^{(m)}$ includes the factor $(n - m)$. Then, establishing by direct calculation that $R^{(1)} = 0$, we reveal the presence of not only the factor $(m - 1)$ but also of the factor $(n - m - 1)$ in the expression $R^{(m)}$ which is an integral function of the third degree with respect to m. Therefore, $R^{(m)}$ is divisible by the product $(n - m)(n - m - 1)(m - 1)$, and for its complete determination it only remains to consider the coefficient of m^3 which is equal to the expression

$$A = \sigma^2(s-1)(s-2)(s-3)(n-1) + n\sigma(\sigma-1)(s-1)^2(n-1) +$$
$$(n-3)(n-2)(n-1)^2 - (\sigma-1)^2(n-3)(n-2) - 2(n-\sigma)(n-3)(n-2)(n-1)$$

where $n = s\sigma$. Combining the three last terms of A, we find that their sum is equal to

$$-\sigma^2(s-1)^2(n-3)(n-2)$$

which enables us to isolate the factor $\sigma^2(s - 1)$ from A.

We carry out the subsequent calculation in the following way:

$$A = \sigma^2(s-1)[(s-2)(s-3)(n-1) +$$
$$(n-s)(s-1)(n-1) - (s-1)(n-2)(n-3)] =$$

$$\sigma^2(s-1)[(s-2)(s-3)(n-1) - (s-2)(s-1)(n-1) +$$
$$(n-2)(s-1)(n-1) - (s-1)(n-2)(n-3)] =$$

$$2\sigma^2(s-1)[-(n-1)(s-2) + (n-2)(s-1)] =$$
$$2\sigma^2(s-1)(n-s) = 2\sigma^2 s(s-1)(\sigma-1).$$

Thus, for $x_1 + x_2 + \ldots + x_\sigma = m$,

$$\sum P(Q^2 - 1) = \frac{2\sigma^2 s(s-1)}{(\sigma s-2)(\sigma s-3)(\sigma-1)} \frac{(m-1)(n-m-1)}{m(n-m)} C_n^{\ m} p^m q^{n-m}$$

and

$$E(Q-1)^2 = \frac{2\sigma^2 s(s-1)}{(\sigma s-2)(\sigma s-3)(\sigma-1)} \sum_{m=1}^{n-1} \frac{(m-1)(n-m-1)}{m(n-m)} C_n^{\ m} p^m q^{n-m} \qquad (4)$$

from which, since the expression $(m-1)(n-m-1)/[m(n-m)]$ invariably remains less than unity, immediately follows the inequality

$$E(Q-1)^2 < \frac{2\sigma^2 s(s-1)}{(\sigma s-2)(\sigma s-3)(\sigma-1)}.$$

For $\sigma \geq 5$ it is not difficult to derive now a very simple inequality [3]

$E(Q-1)^2 > 2/(\sigma - 1)$.

For large values of n this highest boundary for the expectation of $(Q-1)^2$ also represents its approximate value since the main terms in the sum included in (4) are those for whom m is close to np and the difference $(n-m)$ is close to nq, and for whom therefore the expression $(m-1)(n-m-1)/[m(n-m)]$ differs but little from unity if $(n-s_\sigma)$ is a large number [4].

[3] Being prompted by an article of Tikhomirov [3], I take this opportunity to say a few words about the fashionable theory of *correlation*. Its positive side is not significant enough and consists in a simple usage of the MLSq to discover linear dependences. However, not being satisfied with approximately determining various coefficients, the theory also indicates their probable errors, and enters here in the realm of imagination, hypnotism and belief in mathematical formulas that actually have no sound scientific foundation. Of such nature is, for example, the Pearson formula which plays an important part in Tikhomirov's paper and which is adduced after a vain reference to my book [5] [5].

The hypnotism of the theory of correlation manifests itself in the following words of the same paper:

When r is equal to zero, it is said that there is no correlation between the elements, and in this case it is absolutely impossible to estimate the variability of the other element when issuing from the deviations of the first one.

Actually, however, it is not difficult to construct any number of connections which will not at all be detected by the coefficient of correlation, and such, that, nevertheless, the changes of one element will determine those of the other one. Some indication of such cases is also found in Tikhomirov's article (p. 34). The passage even begins there by declaring that "It is important to point out …", but, contradicting these words, remains futile and ends by stating that

It is said in this case that the variables are not correlated, but, at the same time, are not independent one from another.

The following opinion (p. 26) should also be attributed to the realm of belief:

The equation of regression only means that, knowing the value of x_i, it is possible to say that the most probable value of y_i will be $r\sigma_r/\sigma_1$.

The example given in Table 4 (p. 43) can serve for describing propositions based on the theory of correlation. For 23 pairs of magnitudes Δx and Δy entered there, the coefficient of correlation occurred to be small (0.09), whereas its probable error, comparatively large (0.14). The author inferred that the existence of correlation between these magnitudes cannot be considered proven; I shall not by any means try to refute such an indefinite conclusion.

However, if the last ten pairs instead of all the 23 of them are taken, the coefficient will exceed 0.7, whereas the Pearsonian probable error will fall down to 0.1 which is less than 1/6 of this coefficient. Then, in accord with the rule indicated on p. 24, we shall have to make an absolutely different conclusion:

In practice, it is usually considered that the existence of correlation between the elements studied is entirely certain when the coefficient of correlation is not less than six times its probable error. This is equivalent to having many thousand chances to one in favor of the connection actually existing.

To avoid misunderstanding, or possible debates about 10 being a small number and 23 being a sufficiently large number, I remark that even in the theory of observational errors I do not attach any great importance to the so-called probable errors and only consider them as a means for conjecturally comparing the merit of different observations. As to the coefficient of correlation, it has a quite definite value and can be calculated without any error when studying a given totality of numbers. If, however, this totality is considered as a part of an absolutely unknown totality, then the coefficient of correlation for the latter cannot be determined either when the former consists of 10 pairs, or of 23, or of a much larger number of pairs. If desired, it is of course possible to consider the number calculated for the given totality as an approximate value of this coefficient whose very existence is doubtful; but one cannot help to call the probable error of such a determination, and the connected calculation of chances, pure fiction [6].

Comments by Yu.V. Linnik [3, pp. 668 – 670]

1. The coefficient of dispersion in the form derived by Markov enables to establish some test of homogeneity of the σ series of independent trials; namely, the test for examining the hypothesis on the constancy of the probability of the event observed in these series. If this hypothesis is true, the values of the expectation and the variance of Q will be those indicated by Markov and allowing, for example, the use of the Chebyshev inequality.

2. The calculation can be carried out thus. When $m = 1$, and $s_i = s$, $i = 1, 2, \dots, \sigma$ we have $n = \sigma s$, $x_1 + \dots + x_\sigma = 1$ so that $x_i = 1$ for one i, and all the other x's are zero. Therefore, Q is necessarily

$$Q = \frac{n(n-1)s}{(\sigma-1)(n-1)}\{[(1/s) - (1/n)]^2 + [(\sigma - 1)/n^2]\} =$$

$$\frac{\sigma s^2}{\sigma-1}\{[(1/s) - (1/\sigma s)]^2 + [(\sigma - 1)/\sigma^2 s^2]\} = 1.$$

When $m = n - 1$, $x_i = s - 1$ for one i, and all the other x's are s. In this case

$$Q = \frac{n(n-1)s}{(\sigma-1)(n-1)}\left((\sigma-1)[1-\frac{\sigma s-1}{\sigma s}]^2+[\frac{s-1}{s}-\frac{\sigma s-1}{\sigma s}]^2\right)=$$

$$\frac{\sigma s^2}{\sigma-1}\left(\frac{\sigma-1}{(\sigma s)^2}+\frac{(\sigma-1)^2}{(\sigma s)^2}\right)=1.$$

3. We have

$$\frac{2\sigma^2 s(s-1)}{(\sigma s-2)(\sigma s-3)(\sigma-1)} = \frac{2[1-(1/s)]}{(\sigma-1)[1-(2/\sigma s)][1-(3/\sigma s)]}.$$

It remains to prove that for $\sigma \geq 5$

$$1 - (1/s) < [1 - (2/\sigma s)]\cdot[1 - (3/\sigma s)].$$

But, under this condition,

$$[1 - (2/\sigma s)]\cdot[1 - (3/\sigma s)] \geq [1 - (2/5s)]\cdot[1 - (3/5s)] =$$
$$1 - (1/s) + (6/25)\cdot(1/s^2) > 1 - (1/s).$$

4. This proposition follows, for example, from the Chebyshev inequality written down in the form

$$P[(m - np) > n\xi] < pq/n\xi^2$$

with an appropriate choice of ξ.

5. Tikhomirov (p. 23) writes down the Pearson formula as "the probable error is equal to $0.674(1 - r^2)/\sqrt{n}$". Strictly speaking, this formula is meaningless already because the sample coefficient of correlation r is a random variable. The rigorous sense, and the condition for applying the Pearson formulas are as follows: For samples of size $n \to \infty$ from a normal bivariate general population $(X; Y)$ with ρ ($|\rho| \neq 1$) being the coefficient of correlation between X and Y, the sample coefficient of correlation r tends to normality, its expectation approaches ρ, and its variance is asymptotically equal to $(1 - \rho^2)^2/n$, see [1]. In the absence of grounds for assuming that the general population is normal, the Pearson formula of course has no scientific foundation.

6. Markov indicates here that it is impossible to speak about the distribution of the sample coefficient of correlation without knowing the type of the initial "general" distribution $(X; Y)$. In our time, there exist calculations of such a distribution for a normal $(X; Y)$. It occurs that the distribution of the sample coefficient of correlation only depends on the "general" coefficient of correlation ρ and the size of the sample n. In this case, given a large number of observations, it is possible to estimate ρ.

References

1. Kuzmin, P.O. (1939), Sur la loi de distribution du coefficient de corrélation dans les tirages d'un ensemble normal. *Doklady* AN SU, vol. 22, No. 6, pp. 298 – 301.
2. Markov, A.A. (1900), *Исчисление вероятностей* (Calculus of Probability). Psb. Later editions: 1908, 1913, 1924; German translation, 1912.
3. Markov, A.A. (1951), *Избранные труды* (Sel. Works). N.p.

4. Tikhomirov, G. (1915), The method of correlation and its application in meteorology. *Geofizich. Sbornik*, vol. 2, No. 3, pp. 21 – 48. (R)

15. A.A. Markov. On the Coefficient of Dispersion for Small Numbers

Strakhovoe Obozrenie, No. 2, 1916, pp. 55 – 59

For small numbers, the coefficient of dispersion cannot be large. Since some statisticians [1; 2, pp. 131 – 134] are apparently inclined to exaggerate the importance of this fact perceiving there a peculiar law of small numbers, and a proof of a special stability of these numbers, it seems that it would be helpful to show its independence of the theory of probability. This is indeed the aim of my note.

We begin with a simplest assumption that, in M series of N observations in each, some event occurred a times in m series and had not occurred at all in $(M – m)$ series. The square [1] of the coefficient of dispersion will here be

$$Q^2 = \frac{(M-m)\,(ma)^2 + m[(M-m)a]^2}{(M-1)\,ma\,[M-(ma/N)]}.$$

Disregarding the fraction ma/N because of the supposed very large value of N, we have

$$Q^2 = [(M-m)/(M-1)]a.$$

Consequently, in this simplest example the coefficient of dispersion is, generally, somewhat less than $\sqrt{a}$ and only reaches this extreme value at $m = 1$. Then, it is not difficult to see that the same upper boundary $\sqrt{a}$ persists for Q in case of differing numbers {of successes} not exceeding a. The coefficient Q can only reach its indicated maximal value when the system {of successes} consists of zeros and a single a ($m = 1$).

If, however, the system several times includes a as well as other numbers less than a but differing from zero, the coefficient of dispersion must be much less than $\sqrt{a}$. Thus, for a system consisting of zeros and unities, the coefficient will be without fail less than unity. The same holds if a small number of twos is added to a large number of unities. For example, the coefficient of dispersion for a system consisting of 15 zeros (non-occurrences), 9 unities and two successes in one single case [1, pp. 18 – 19] is naturally less than unity.

When changing this example but leaving the previous number of zeros and the total number of unities and twos (10 unities; 9 unities and two successes in 1 case; …; 10 twos; with 15 zeros in each system), we will have 11 systems with the maximal value of the coefficient of dispersion $\sqrt{2\cdot 15/24} = \sqrt{1.25} = 1.11\ldots$ for the system consisting of 15 zeros and 10 twos.

Some light can also be thrown by an examination of all possible combinations from four numbers not exceeding 3. If the order of the numbers is taken into account, but the totality of four zeros be eliminated, there will be $4^4 – 1 = 225$ such combinations.

For the 15 combinations in which neither number exceeds 1, the coefficient of dispersion naturally does not exceed unity and only reaches it in four cases,

0001, 0010, 0100, 1000.

For the 81 – 16 = 65 combinations in which a two is included but a three is not, the coefficient remains not larger than $\sqrt{2}$. It reaches this limit in four cases, – in combinations of three zeros and a two; and it takes the value $\sqrt{4/3}$ in six cases (two twos and two zeros), and the value $\sqrt{11/9}$ in 12 cases (2 zeros, a unity and a two). The other 65 – 22 = 43 combinations in which a two occurs but not a three, the coefficient of dispersion is less than 1. For 256 – 81 = 175 combinations with a three, the coefficient remains not larger than $\sqrt{3}$ and only reaches this limit in four combinations (three zeros and a three). Then, the coefficient reaches

$\sqrt{2}$, in 6 cases, – 0033, 0303, ...; and in 12 cases, – 0013, 0103, …;
$\sqrt{9/5}$, in 12 cases, – 0023, 0203, …; $\sqrt{9/7}$, in 12 cases, – 0133, 1033, ...;
$\sqrt{19/15}$, in 12 cases, – 0113, 1013, …; $\sqrt{10/9}$, in 24 cases, – 0123, 1023, ...;

The other 175 – 82 = 93 combinations of the considered type provide a coefficient of dispersion not exceeding unity. A similar study of a large number of terms and their considerable upper boundary is of course very difficult. Therefore, instead of accomplishing it on a large scale, I shall dwell on the combinations resulting from another example also adduced in [1] as a confirmation of the law of small numbers:

144 times, 0; 91 times, 1; 32 times, 2; 11 times, 3; 2 times, 4.

I shall consider all the 124 pertinent combinations without changing the numbers of zeros, of threes and fours as well as leaving intact the previous total number of unities and twos. In general, these combinations are

144 times, 0; x times 1; (123 – x) times, 2; 11 times, 3, 2 times, 4.

For whichever x the coefficient of dispersion will be

$$(280/279)\left(\frac{623-3x}{287-x}-\frac{287-x}{280}\right).$$

For investigating the behavior of this function at 124 considered values of x we calculate its derivation with respect to x; it is equal to

$$(280/279)\left(\frac{1}{280}-\frac{238}{(287-x)^2}\right).$$

For $x \le 28$ the derivative is positive and it becomes negative for $x \ge 29$. Therefore, the coefficient of dispersion reaches its maximal values at $x = 28$ and 29 and its minimal values at $x = 0$ and 123. Calculations provide the following values for the square of the coefficient:

at $x = 0, 28, 29, 123$, $Q^2 = 1.15, 1.16, 1.16, 0.97$, respectively.

We see that in all 124 cases the coefficient remains close to unity and less than 1.09.

It is thus confirmed that the coefficient of dispersion can be close to unity for most various combinations which have no relation to the theory of probability and that the smallness of the coefficient for small numbers does not represent anything special [2].

As to the application of the remarkable Poisson formula to small numbers, it might be considered well-founded provided that a large totality of homogeneous series of observations is investigated; otherwise, it is only a skilfull play with numbers.

Notes

1. I keep to the generally adopted term although I consider the taking of the root unnecessary.

2.{Markov repeated this statement in several letters to Chuprov and in a letter to Kaufman, see Sheynin (1990, p. 111). Quine & Seneta (1987) argued, contrary to Markov and without citing him, that the smallness of the coefficient of dispersion is understandable from a stochastic point of view, and I am inclined to agree with them.}

References

1. Bortkiewicz, L. von (1898), *Das Gesetz der kleinen Zahlen*. Leipzig.
2. Kaufman, A.A. (1916), *Теория и методы статистики* (Theory and Methods of Statistics). M. Third edition.
3. Quine, M.P., Seneta, E. (1987), Bortkiewicz' data and the law of small numbers. *Intern. Stat. Rev*., vol. 55, pp. 173 – 181.
4. Sheynin, O. (1990, in Russian), *Chuprov.* Göttingen, 1996.

16. The Correspondence between A.A. Markov and B.M. Koialovich

Translator's Foreword

Professor Boris Mikhailovich Koialovsky (1867 – 1941) was Markov's *gifted student* [1, p. 220] and *an eminent scientist* [2, p. 415]. He taught at several academic institutions, notably at Petersburg Technological Institute, and achieved certain success in studying differential equations [3]. His correspondence with Markov characterizes himself, to some extent describes Markov's exposition [4] of the MLSq and once more confirms that Markov was diligently studying the relevant literature. The letters are kept at the Archive, Russian Academy of Sciences, *Fond* 173, Inventory 1; the additional numbers provided below show the place of the letters in that *Fond*.

****

Letter 1, Koialovich – Markov, 25.9.1893 (10, No. 1)

Dear Andrei Andreevich! I received your letter and am thankful for the attention to my lectures. I personally considered them as an urgent work only written for giving the students some manual for recalling my lectures. […]

I regret very much that my lectures made such an impression on you; and I regret even more that I felt in your letter an extremely unexpected opinion about me with which I cannot at all agree because I consider it undeserved and unjust. […] Do you, Andrei Andreevich, really think that I am unacquainted with the true essence of the {Stirling} formula? […]

It is of course not for me to judge to what extent I was able to explain the notion of probability. I shall only dwell on that example to which you object. {Taken by itself, the author's explanation is not sufficient; apparently, however, the matter was not important.} There remained for me a choice between two methods: either to base my reasoning on the notion of probability of error, or on considering, as Chebyshev did in his lectures, whose manuscript notes I had in my possession, a sum of randomly selected magnitudes. The second possibility, which I naturally prefer, was impossible for me owing to lack of time. Then, I add that in any explication of the MLSq we have to rely on more or less arbitrary assumptions. This, as I think, cannot and should not be avoided since the matter concerns a practical method for treating observations rather than a construction of an abstract mathematical theory. I am far from desiring to prove somehow the MLSq and I have therefore used the term *justify* (p. 55) rather than *prove*. I think that justifying and proving are absolutely different notions.

I also note that from among the hypotheses which we undoubtedly have to introduce when explicating the MLSq, the one concerning the properties of the probability of error is in any case the most natural although I know very well the difficulties connected with it [1]. […]

I called the results provided by the MLSq quite trustworthy, because, as far as I know, such are all the main numerical results of our natural sciences, physics and astronomy.

You say that, once the data are sound, the results will always be good even without the MLSq. I agree absolutely, but from among these good results some might nevertheless be better than the others depending on how we combine the observations. […]

I did not at all attempt to connect the works of Sleshinsky [5] and Yarochenko [6; 7] with those of Chebyshev [2]. For me, they have only one common point: they explicate the MLSq.unlike Gauss, and it was in this sense that I have cited them.

I extremely regret that your works remained unknown to me and I apologize for omitting to cite them. They were unknown to me because I have not seen any references to them, and, as far as I can remember, you never mentioned them in our talks [3].

One of Chebyshev's works, I do not remember which one, had been in my hands, but for such a short time that I was unable to acquaint myself in detail with it. For me, it was sufficient to see that he explicated the MLSq not as Gauss did.

Concerning the function which I denoted by $\varphi(\Delta)$, I allow myself to turn your attention to the following. Let us imagine a series of magnitudes

$$-n\varepsilon, \ldots, -2\varepsilon, -\varepsilon, 0, \varepsilon, 2\varepsilon, \ldots, \Delta, \Delta+\varepsilon, \ldots, \Delta+n\varepsilon \tag{1}$$

and assume at first that ε is a finite although a very small magnitude. If we suppose that the error of observation can only take values indicated in the series (1), then the notion about its probability is quite clear for any ε. If we will now decrease ε indefinitely, then our hypothesis will come arbitrarily close to an assumption of a continuous error and ε will become the differential of the magnitude Δ. The function that I denoted by $\varphi(\Delta)$ will indeed express the probability {as studied} in the theory of probability. The aim of the theory is not to create the notion of probability anew, but to explain it and to make it accessible to measurement. By stating that a well-known ratio is probability, we do not explain the difficulty, but only keep ourselves away from such an explanation and I did not consider it possible to do so. […]

Concerning the notion of dependent and independent trials, I quite agree with you that I have left a gap here, although the explication below made it clear which trials I had discussed. […]

I am turning now to the main point of our disagreement, to the MLSq. You reproach me as your student for avoiding your adopted description of the method. Believe me, Andrei Andreevich, from among your students there are hardly many who respect you more than I do; however, even for all this respect, I reserve for myself the right to a free choice in the matters of science and I can never renounce this right. You yourself would hardly be pleased if your students were to restrict their teaching by blindly imitating your lectures. And the reasons why I have chosen a different explication of the MLSq are these.

1) I believe that your exposition will be too difficult for my listeners [4].

2) There is one point in your account which I was never able to understand either myself or during our conversations. Here it is. On p. 159 of your lectures (1891) [8] you say:

Let us introduce, in accord with the above, magnitudes x', x'', …, $x^{(n)}$ representing the possible result of the first, the second, …, the n-th observation.

It is this place that I was unable to understand. What do you mean by the possible results of observation? Under what conditions are they possible and how do they differ one from another? I could not understand this without introducing once more the concept of the probability of error and I therefore preferred to decide in favor of an error equal to Δ and to assume that $\varepsilon = d\Delta$ is an infinitesimal.

The function φ(Δ) is thus not only not senseless, but it is even very beneficial in that it eliminated the need for introducing, while explicating the MLSq, the concept of density [5]. […]

I hope that my explanations will serve to eliminate the misunderstandings and disagreements which appeared between us.

Letter 2, Koialovich – Markov, 2.10.1893 (10, No. 8)

[…] I know very well that my course is corrupted by many shortcomings which should be eliminated later on. […] {Concerning the example discussed in Letter 1}I say clearly enough that the point only concerns the choice of the most probable event; and an event remains most probable no matter how low is its probability if only it is higher than the probabilities of the other events. […]

You write that I attempt to obscure both this notion (of equally possible cases) and the concept of probability. I believe that I have the right to protest against such a charge because it rings oddly with respect to a man who nevertheless tries his best to reveal the truth. I can obscure the truth unintentionally but I cannot *attempt* to do so; that would imply such motives which you are unable to assume that I have. […]

I am asking you to turn your attention to the following difficulties, which, as it seems to me, naturally arise when keeping to your viewpoint. You write: Why may we not say in the theory that the well-known relation is probability? I answer: We certainly may, and we may say that anything is probability, but for whom will such a theory be compulsory and interesting? Will not everyone have the right to say: Am I concerned with theories operating on notions concocted by you yourself which perhaps have no object {no representation} in reality? Will not all this theory become *Übungen für den Verfasser* (as apparently Weierstrass expressed it) [6] […]

Therefore, you will probably agree that I have grounds for remarking that we do not explain the difficulties but only keep ourselves away from them. […]

I shall now dwell on the MLSq. I am naturally acquainted with Laplace's account and completely agree with you that I should have indicated it. I had certain motives for citing Sleshinsky and Yarochenko; I myself have a low opinion about them, and especially with regard to the former, but these motives are subjective and not obligatory for anyone else.

You write that you perceive that I am not at all acquainted with the works of Chebyshev and others. This surprises me very much. Indeed, I held them in my hands and read them. Or, do not you trust me?

Concerning your account of the MLSq I allow myself to indicate the following. I know of course that your p. 159 carries a reference to previous explanations (probably to p. 157); I have not mentioned this because, to my utmost understanding, nothing is explained either on p. 157 or elsewhere. Even your latest explanation, for which I am of course sincerely thankful, did not explain to me my most serious perplexity, namely:

As far as I understand you, you consider each separate observation as a value of a possible result. Thus, a series of results

$$a_1, a_2, \ldots \qquad \text{(A)}$$

is possible for each measurement, and one of them is realized. I am prepared to understand all this concerning one observation. However, if there are, for example, two observations, then I cannot understand the difference between the series of all the possible results of the first observation (A) and the similar series for the second measurement (B) [7]. The problem will be certainly solved at once if you say that the probabilities of the same error in these two series are different, but you will hardly want to introduce the notion of probability of error in your exposition. […]

I never said that *all* the numerical results of physics and astronomy are trustworthy, but only that all the *principal* results are such. This is indeed very different. […]

Letter 3, Markov – Koialovich, 1897 (149)

[…] Try to eliminate the outlined difficulties not only by unfounded declarations and you will have to admit that all your calculations were nothing but an obscurity of the problem that distracted us from the essence of the matter [9].

Letter 4, Koialovich – Markov, 5.2.1909 (10, No. 14)

[…] That you did not read not only my book, but even its Contents, is obvious since you discovered an important gap (*not a single word is there about the* […]). This imaginary gap is filled in in §47, pp. 73 – 74.

Before advising me to consult generally known textbooks and unnecessarily to reassure Posse [8], you should have had a look at his course in integral calculus of 1891, pp. 128 – 189. […] My definition of the concept of function will hardly mislead those readers who will […] also read note 2 on pp. 39 – 40. […] [9]

Notes

1.{See belowNote 5.}
2.{Koialovsky had apparently already read the publications of those two authors.}
3.{The only relevant work is apparently [8]. However, Koialovsky refers to this source below, and the only possible explanation seems to be that he somehow thought about a few of Markov's writings on probability in general.}
4.{Cf. Markov's own words which concerned the *end* of his manual [4, 1908], *i.e.*, the MLSq [9, Letter 15 of 1910, p. 21]: *To my regret, however, I have often heard that my presentation is not clear enough.* On this point, and in connection with Notes 5 and 6 below, also see [10].}
5.{Koialovich thus distinguishes between φ(Δ) and the density. Anyway, the main point concerning Markov's manual [4, 1908] is that, when discussing the treatment of observations, he did not say clearly enough that their errors were random, and that, therefore, they, practically speaking, possessed some density. At the time, the term *random variable* (or, in Russian, *random magnitude*) was only coming into existence, but at least, beginning with Gauss onwards, astronomers and geodesists have been discussing random errors.}
6.{In actual fact Markov himself, in the very beginning of his treatise [4, 1908], expressed his own dissatisfaction with the classical definition of probability.}
7.{Koialovsky did not write out any series (B).}
8.{Konstantin Aleksandrovich Posse (1847 – 1928), Honorary Member of the Petersburg Academy of Sciences. His textbook in differential and integral calculus had been widely used for about 25 years.}
9.{Letters 3 and 4 did not discuss probability.}

References

1. *Научное наследство* (1991) (Scient. Heritage), vol. 17. L.
2. Stokalo, I.Z., Editor (1967), *История отечественной математики* (History of National Mathematics), vol. 2. Kiev.

3. *Нов. энциклопедич.словарь* (ca. 1915) (New Enc. Dict.), vol. 23, p. 46.
4. Markov, A.A. (1900), *Исчисление вероятностей* (Calculus of Probability). Psb. Later editions, 1908, 1913, 1924; German translation, Leipzig – Berlin, 1912.
5. Sleshinsky, I.V. (1893), On the Chebyshev theorem. *Zapiski Novorossiysk. Univ.*, vol. 59, pp. 503 – 506 of sect. 1. (R)
6. Yaroshenko, S.P. (1893), On the theory of the method of least squares. *Zapiski Novorossiysk. Univ.*, vol. 58, pp. 193 – 208 of sect. 2. (R)
7. --- (1893), Sur la méthode des moindres carrés. *Bull. sci. math.*, sér. 2, t. 17, pp. 113 – 125.
8. Markov, A.A. (1884, 1885, 1888, 1891), *Теория вероятностей* (Theory of Probability). Mimeogr. editions. Psb.
9. Ondar, Kh.O., Editor (1977, in Russian), *Correspondence between Markov and Chuprov*. New York.
10. Sheynin, O.B. (1989), Markov's work on probability. *Arch. Hist. Ex. Sci.*, vol 39, pp. 337 – 377.

17. Yu.V. Linnik, N.A. Sapogov, V.N. Tikhomirov
A Sketch of the Works of Markov in the Theory of Probability

In Markov, A.A. *Избранные труды* (Sel. Works). N.p., 1951, pp. 614 – 642.

I left aside the first part of this sketch devoted to Markov's work in the theory of numbers. The authors never mentioned Nekrasov; see P.A. Nekrasov, *Theory of Probability*. Berlin, 2004.

1. Markov's works in this field are mainly devoted to the CLT, as it is now called, for sums of independent variables; to limit theorems for dependent variables, including variables connected into a chain pattern introduced by him himself; to urn problems and to issues in mathematical statistics, and in particular to justifying the MLSq.

2. For ten years, from 1898 to 1908, Markov examined problems connected with the CLT for independent summands. Overcoming considerable difficulties one after another, and disregarding the complexity of calculations, he applied Chebyshev's method of moments [1] for studying the distributions of sums of independent terms and showed that this method was capable of producing results of an almost exhaustive generality.

At the beginning of the indicated period, A.A. only considered the sums of such independent terms for which moments of any order existed, but at its conclusion he discovered a way to rid himself of that burdening assumption and thus to impart the necessary generality to his findings.

In a letter to Vasiliev [2] of 23.9.1898 Markov [1] examined a sum of such independent variables $X_1, X_2, \ldots, X_n$ that $\mathrm{E}X_i = 0$ and that for any integer k the k-th moment existed and was restricted in absolute value by a number C_k, generally depending on k, but not on n.

Supposing that $Y_n = (X_1 + X_2 + \ldots + X_n)/\sqrt{n}$, Markov proved that, as $n \to \infty$,

$$\mathrm{E}Y_n^{\,m} - A_m\,(\mathrm{E}Y_n^{\,2})^{m/2} \to 0, \qquad (1)$$

$$A_m = (2^{m/2}/\sqrt{\pi}) \int_{-\infty}^{\infty} t^{\,m} \exp(-t^2)\, dt.$$

Already Chebyshev (1887), studied a similar problem, but his formulations and proofs contained a gap lacking in the work of Markov.

The proof of (1) was based on the simple properties of the general formula of the Newton binomial, and of the expectations. It is still important as a component of all his works on the CLT for independent variables; and he applied its appropriate modifications in his later contributions devoted to dependent variables connected into a chain or in another way.

3. In his subsequent work Markov achieved success in two directions.

1) He constructed a tool enabling him to determine the limiting distribution of probabilities for a normed sum by issuing from relations of the type of (1) concerning the limiting behavior of the expectations of the various integral powers of this sum. This tool was based on the analytic theory of continued fractions and the related "Chebyshev inequalities".

2) He got rid of the restrictions connected with the existence and uniform boundedness of the moments of high orders by "curtailing" random variables.

Already in 1883 Markov studied the Chebyshev inequalities in the theory of continued fractions, and later he explained in detail how to use them in the problem of moments and its stochastic applications. These studies do not directly belong to probability, but they are conveniently formulated in stochastic terms.

4. In essence, two different cases were considered, viz., variables having a discrete distribution or a continuous density. The study of the general case would have led the proof out of the boundaries of the classical analysis of the 19th century to which he was keeping.

For the discrete case a sequence of random variables X having all the integral moments $\sum_x px^k$ is examined. Then the sum

$$\sum_x p/(z-x) \qquad (2)$$

for each random variable of the sequence is expanded into a formal continued fraction and the properties of the $(m-X)$ convergents $\psi_m(z)/\omega_m(z)$ are established. The denominator has simple real roots ξ so that

$$\psi_m(z)/\omega_m(z) = \sum_\xi [\rho/(z-\xi)]$$

with an equality valid for any polynomial $\Omega(z)$ of degree $\le (2m-1)$:

$$\sum_\xi \rho\Omega(\xi) = \sum_x p\Omega(x).$$

Issuing from this equality and applying very subtle considerations, the following Chebyshev inequalities are derived. If α is situated between two adjacent roots ξ' and ξ'' of the polynomial $\omega_m(z)$, then

$$\sum_{\xi<\xi'} \rho \le \sum_{x\le a} p \le \sum_{\xi\le\xi''} \rho. \qquad (3)$$

If a continuous density $f(x)$ is considered instead of the discrete distribution, then the formally written integral

$$\int_{-\infty}^{\infty} \frac{f(x)}{z-x}dx$$

should be considered instead of (2) and we derive

$$\sum_{\xi<\xi'} \rho \le \int_{-\infty}^{\alpha} f(x)\,dx \le \sum_{\xi\le\xi''} \rho. \qquad (4)$$

In particular, the formal expansion of

$$(1/\sqrt{\pi})\int_{-\infty}^{\infty} [\exp(-x^2)/(z-x)]\,dx$$

can be thus examined. We shall have

$$\sum_{\bar\xi<\bar\xi'} \bar\rho \le (1/\sqrt{\pi})\int_{-\infty}^{\alpha} \exp(-x^2)\,dx \le \sum_{\bar\xi\le\bar\xi''} \bar\rho \qquad (5)$$

where $\bar\rho$ and $\bar\xi$ denote the corresponding coefficients and roots. For an unboundedly increasing m the coefficients ρ and $\bar\rho$ become arbitrarily small.

If now, for each integer k in the sequence X_i of random variables, the relations

$$\sum_x px^k \to (1/\sqrt{\pi})\int_{-\infty}^{\infty} x^k \exp(-x^2)\,dx$$

are obeyed, then, for a fixed m, as it is easily seen,

$$\sum_{\xi<\xi'} \rho \to \sum_{\bar\xi<\bar\xi'} \bar\rho;\ \sum_{\xi\le\xi''} \rho \to \sum_{\bar\xi\le\bar\xi''} \bar\rho.$$

The difference between the sums in (4) and (5) only consists in the coefficients ρ and $\bar\rho$ for the roots ξ', ξ'' or $\bar\xi'$ and $\bar\xi''$, and, since ρ and $\bar\rho$ tend to zero as $m \to \infty$, the main relation

$$\sum_{x\le\alpha} \rho \to (1/\sqrt{\pi})\int_{-\infty}^{\infty} \exp(-t^2)\,dt$$

for X running through the sequence is thus proved.

5. This provided a full proof of the limit theorem for the case considered by Markov in 1898 when he assumed that all the moments existed and were uniformly bounded and that the ratio of the variance of the sum to the number of terms was bounded below. Such a theorem was less general than Liapunov's central proposition, but, in return, the method of proof enabled to derive relations of the dependence of the approximation of distributions on the approximation of the moments which are not furnished by the CLT.

Liapunov's fundamental works appeared in 1900 and 1901. There, he applied the method of characteristic functions rather than the Chebyshev – Markov method of moments. For independent random variables $Z_1, Z_2, \ldots, Z_n$, $EZ_k = a_k$, var $Z_k = b_k$, the existence of $b_k^{(2+\delta)} = E|Z_k - b_k|^{2+\delta}$ was also assumed with $\delta > 0$ being some constant. For the limit theorem to be valid only one condition was needed, viz.,

$$\sum_{k=1}^{n} b_k^{(2+\delta)}/(\sum_{k=1}^{n} b_k)^{1+\delta/2} \to 0 \text{ as } n \to \infty.$$

Subsequent investigations showed that Liapunov had attained almost exhaustive generality. As Markov [9] indicated, Liapunov had achieved a

generality of conclusions […] *far exceeding that secured by the method of expectations. It seemed even impossible to attain such generality by that method since it was based on considering such expectations, limitless in number, whose existence in the Liapunov case was not assumed.*

For about eight years, from 1900 to 1908, Markov studied such an application of the method of moments that would enable him to prove the limit theorem under the Liapunov conditions. In 1908 he succeeded. The main idea of the new application of the method of moments was the introduction, along with the main magnitudes $Z_k - a_k$, of supplementary "curtailed" random variables

$X_k = Z_k - a_k$ if $|Z_k - a_k| < N$, $X_k = 0$ otherwise.

Such variables already have any moments, and the distribution of their sum when the number N is chosen appropriately differs but little from the distribution of the initial sum. This Markov's idea is being successfully applied in stochastic studies until this very day.

6. Markov attached great importance to ascertaining the conditions for the applicability of the limit theorem for sums of random variables. Even in 1898, in his letters to Vasiliev [1], he constructed examples of sums of independent random variables which did not tend to the normal law owing to the comparatively slow increase in the variance of the sum. There also he criticized Poincaré's viewpoint on the applicability of the normal law in the theory of observational errors [3].

Then, Markov [9] offered examples of sums not approaching the normal distribution because of the violation of the conditions which in present terminology are called after Lindeberg and Feller.

We also note that Markov studied, although had not solved definitively, the extension of the limit theorem onto sums of independent two-dimensional vectors. In 1915 he [10] examined the sums

$$S_n = (s_{n1}; s_{n2}) = \sum_{i=1}^{n} (x_i; y_i)$$

of identically distributed and independent random vectors $(x_i; y_i)$, $i = 1, 2, \ldots$ and proved that after an appropriate norming the mixed moments of all the orders of s_{n1} and s_{n2} tend to the corresponding mixed moments of the bivariate normal law. The first rigorous proof of the two-dimensional limit theorem is due to Bernstein (1944).

7. Markov should be considered the founder of a very important and large section of the modern theory of probability devoted to studying dependent random variables. The main issues in which he was interested here related to the LLN and the CLT for sums of random variables.

8. In 1907 [4] Markov [2] indicated several sufficient conditions for the LLN to be applicable to the sums S_n of dependent random variables

$$S_n = \sum_{i=1}^{n} X_i. \qquad \text{(ii)}$$

They were obtained as a corollary of the applicability of the law to any X's possessing second moments if only $[\mathrm{var}\,(S_n)/n^2] \to 0$ as $n \to \infty$.

It seems that before the appearance of this work the very possibility of the existence of sufficiently general conditions for the applicability of the LLN to such sums was not obvious and Markov [2, p. 361] felt himself obliged to stress that the *independence of the variables is not a necessary condition for the existence of the law of large numbers*.

In §1 of his contribution, Markov proved that the law was applicable to S_n if the variances $\mathrm{var}\,(X_i)$, $i = 1, 2, \ldots$, were bounded and the connection between the X_i's was such that the increase in a particular value of one of them led to the decrease of the expectations of any other one. At the end of that section he proved that the LLN persisted even when the expectation $E(X_i)$ of each X_i decreased with the increase in $(X_1 + X_2 + \ldots + X_{i-1})$.

As a next example of dependent variables obeying the law Markov considered the sequence $X_1, X_2, ..., X_n, \ldots$ connected, according to modern terminology, into a simple homogeneous *Markov chain* with each of the X_i's only taking two values, 0 and 1. It was here that A.A. first introduced such sequences of dependent variables (or dependent *trials*) which he himself (although only later) named *chains*. In essence, the conclusion about the applicability of the LLN to S_n was obtained here as a corollary of the ergodic property of the simple homogeneous chain with two possible states that invariably takes place if the transition probabilities differ from 0 and 1.

In §4 Markov offered an example in which the law was not applicable to dependent and bounded variables. Its essence consisted in that the law did not apply if $|X_i| < L$ for all values of i and var $S_n > cn^2$ where the constant c was positive. Finally, at the end of his memoir, Markov considered variables connected into a simple homogeneous chain with the number of the possible values of the X_i's not being anymore bounded by 2. Once more he obtained the LLN for S_n as a corollary of the ergodic property of the chain.

9. Again in 1907 Markov [3] proved the applicability of the CLT to the sums S_n of random variables X_i connected into a simple homogeneous chain and only taking two possible values, 0 and 1, with the transition probabilities differing from 0 and 1.

A.A. indicated that his case offered an example (and, to our knowledge, the first example) of the applicability of the limit theorem to sums of dependent variables. He obtained the proof of his proposition by the method of moments basing his reasoning on the symmetry of some expressions with respect to p and q, the probabilities of the equalities $X_i = 1$ and 0 respectively. This symmetry followed from the special conditions of the case under consideration and cannot be made use of when the variables take more than two values.

10. The work of 1908 [4] is devoted to the study of these more general cases. Here, Markov began by examining with special thoroughness the assumption that the variables X_i, constituting the simple homogeneous chain, only take three values, – 1, 0 and 1, and modified the proof applied in his previous work in such a way that it became applicable to a large number of new problems concerning dependent variables X_i which he examined, one after another, in later contributions.

It is noteworthy that here, while calculating the moments of the sum S_n, Markov in essence introduced characteristic functions $\psi_n(\xi) = \exp[i\xi(S_n - an)]$ of variables $(S_n - an)$ where the constant a was determined under the condition that the expectation $E(S_n - an)$ remained bounded in absolute value as $n \to \infty$. Indeed, he began by constructing the function

$$\frac{f(t;z)}{F(t;z)} = \sum_n \Phi_n(t)z^n \qquad (5)$$

which was the generating function of the functions

$$\Phi_n(t) = \sum_n P(S_n = m)t^m,$$

which, in turn, were the generating functions of the probabilities $P(S_n = m)$. He then changed the variables, $t = e^u$, $z = z_1e^{-au}$, so that the function (5) became *the generating function of the characteristic functions* $\psi_n(\xi)$ for imaginary values of the argument ξ:

$$\frac{f(e^u;z_1e^{-au})}{F(e^u;z_1e^{-au})} = \sum_n \Phi_n(e^u)[z_1e^{-au}]^n = \sum_n \psi_n(-iu)z_1^n$$

since

$$\psi_n(-iu) = \sum_m e^{u(m-an)} P(S_n = m) = e^{-aun}\,\Phi_n(e^u).$$

Romanovsky (1949) later developed in detail the method of introducing characteristic functions $\psi_n(\xi)$ and applied it to Markov chains properly speaking and their various generalizations.

At the end of his work Markov extended his result on the applicability of the limit theorem to sums onto a more general case when variables connected into a chain had any possible values.

11. As stated above, A.A. also applied the same method in some of his other works. Thus, he [7] considered the following stochastic pattern: Let some *system S* randomly take one of its possible *states* $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k$ which change one into another at discrete moments of time $t_1 < t_2 < \ldots < t_n < \ldots$ Suppose that the probability of S being in state ε_i at moment t_n depends not only on its state at the immediately preceding moment t_{n-1} as it happens for simple chains, but on all m of its previous states $t_{n-m}, t_{n-m+1}, \ldots, t_{n-1}$, although not on n or on the states of S at moments t_l, $l > (n-m)$.

If random variable X_n, taking the value x_i when the system is in state ε_i corresponds to each moment t_n, then the sequence

$$X_1, X_2, \ldots, X_n, \ldots \qquad \text{(iii)}$$

is connected, by definition, into a generakized (homogeneous) chain. The problem that attracted Markov here consisted in studying the sums S_n of such variables under a particular assumption that $m = 2$ and that the X's take only two possible values, 0 and 1. He proved that under some restrictions imposed on the transition probabilities the CLT is applicable to the sums S_n. He also indicated an important fact distinguishing generalized and simple chains: for the former, the variance of the limiting distribution can happen to be identical with that for a corresponding series of independent variables.

Incidentally, this shows that, given the limiting normal distribution of S_n, it is impossible to decide whether the X_i's were quite independent or in a definite way connected into a generalized chain.

12. In 1911 Markov [6] solved by means of his method a problem previously considered by Bruns (1906) from another standpoint. Suppose that variables

$$X_n = W_n \cdot W_{n+1},\ n = 1, 2, \ldots$$

are the products of two adjacent terms of a sequence

$$W_1, W_2, \ldots, W_n, \ldots \qquad (6)$$

of independent random variables W_n only taking values 1 and 0 with probabilities α and β respectively, $\alpha + \beta = 1$. The sequence of dependent variables (iii) represents the simplest case of a Markov – Bruns chain, as it is now called.

The series (iii) is characterized by the property that only the adjacent terms X_i and X_{i+1} are dependent. For $|i-j| > 1$ the variables are independent. This problem prompted Markov to consider a more general case of arbitrarily distributed bounded variables X_n, $n = 1, 2, \ldots$ each two of which, X_i and X_j, can only be arbitrarily connected one with another if $|i-j| \le c$ where c

was some constant; otherwise he assumed that X_i and X_j were independent. By means of the method of moments A.A. proved now that the CLT was applicable to the sum (ii) provided that B_n = var $S_n > an$ where the constant a was positive.

Bernstein (1944) essentially generalized this problem. Applying the method of characteristic functions he proved that the CLT was applicable to sums of a wide class of such variables X_i that any of them could be connected one with another although only if the influence of some X_i on another X_j was, in a sense, sufficiently weak when the distance $|i - j|$ was large.

Returning to Markov [6], we note that, in concluding his memoir, he offered some extension of the simplest case. More precisely, he admitted that the W_n were not independent, but constituted a simple homogeneous chain.

13. Markov devoted [8] to a problem belonging, in modern terminology, to vector chains. Consider such a system S with possible states $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k$ that the change from one of them to another one occurs at discrete moments of time $t_1 < t_2 < \ldots < t_n < \ldots$ and represents a simple homogeneous Markov chain with transition probabilities having matrix

$$\left(P_{ij}\right),\ i, j = 1, 2, \ldots, k. \qquad \text{(iv)}$$

Let random vectors from an h-dimensional Euclidean space E_h with common possible values $\theta^{(i)} = (\xi_1^{(i)}; \xi_2^{(i)}; \ldots; \xi_h^{(i)})$, $i = 1, 2, \ldots, k$, correspond to moments t_n, $n = 1, 2, \ldots$ We assume that vector X_n has value $\theta^{(i)}$ if the system S is in state ε_i at moment t_n. The sequence of thus determined connected vectors $X_1, X_2, \ldots, X_n$ constitutes a *vector chain.*

The stochastic pattern studied by Markov might be considered a particular case of a vector chain in the Euclidean space E_2. Events *A, B, C* combined with events *E* and *F* constitute six possible states, *AE, BE, CE, AF, BF,* and *CF*. Let vectors $X_n = (x_{n1}; x_{n2})$ belonging to E_2 have values

$$(\xi^{(1)}; 1), (\xi^{(2)}; 1), (\xi^{(3)}; 1), (\xi^{(1)}; 0), (\xi^{(2)}; 0) \text{ and } (\xi^{(3)}; 0) \text{ respectively}$$

with $\xi^{(1)}$, $\xi^{(2)}$ and $\xi^{(3)}$ being arbitrary numbers. Markov examined the limiting distribution of the number of the occurrences of the event E in n trials as $n \to \infty$. This problem thus became a study of the limiting distribution of a sum of random variables

$$s_{n2} = (x_{12} + x_{22} + \ldots + x_{n2}).$$

The formulation of Markov's several problems differ in the assumptions regarding the matrix of transition probabilities (iv) and he solved them by his method explicated in detail elsewhere [4].

14. In all the contributions described above, Markov invariably had to do with homogeneous chains. In 1910 he [5] studied the limiting distribution of the sum S_n of vector variables X_i connected into a non-homogeneous chain. He assumed that the variables X_i can only take values 1 and 0 with transition probabilities

$$p'_i = P(X_i = 1 / X_{i-1} = 1),\ p''_i = P(X_i = 1 / X_{i-1} = 0)$$

and proved that the CLT was applicable to S_n if

$$p_o < p'_i < 1 - p_o,\ p_o < p''_i < 1 - p_o$$

where p_o was some positive number. The proof was once more accomplished by the method of moments, but the moments themselves were calculated quite differently as compared with the case of homogeneous chains. The manner of calculation can rather be considered similar to that which he applied for independent variables. It consisted in expanding the calculated expectation in accord with the Newton formula for a power of a polynomial into separate sums and isolating their main terms. In the process, A.A. made use of the inequality

$$\text{var}\, S_n > cn,\ c > 0,\ n = 1, 2, \ldots \qquad (7)$$

By means of a very clever consideration provided at the end of the memoir, he ascertained that (7) was always valid if the number p_o, introduced above, was strictly positive. In 1936 Bernstein, in order to extend the limit theorem onto more general non-homogeneous chains with the number of possible values of their terms not restricted by 2, had to generalize, first and foremost, the inequality (7).

We also note that Bernstein provided yet another method of proving limit theorems for non-homogeneous chains which enabled to generalize essentially Markov's findings. This method is also applicable to more general kinds of dependent variables and is explicated in detail in his abovementioned contribution (1944).

15. As a rule, Markov illustrated his general results by random variables occurring when considering urn patterns, *i.e.*, stochastic schemes realized by some number of urns containing balls of different colors. The balls served for making *trials*,– for being extracted one by one or otherwise in accord with precise rules specified in advance. The problem consisted in studying the limiting distribution of the number of balls of a certain color extracted during n trials as $n \to \infty$. Markov paid attention to such urn patterns not only for the sake of illustration, he also treated them as a subject in its own right.

16. Many scientists kept to, and developed Markov's ideas concerning chains. Naturally, homogeneous chains were studied in special detail. The Moscow stochastic school introduced new important notions here, and Bernstein with his Leningrad students continued to develop further the theory of non-homogeneous chains. We note that, apparently, especially important is that development of Markov's ideas about variables connected into a chain which is contained in the Kolmogorov theory of stochastic processes of the Markov type and the Bernstein theory of stochastic differential equations.

17. Ideas concerning mathematical statistics occupied an important place in A.A's work. We bear in mind the issues of justifying the MLSq examined in the abovementioned letters to Vasiliev in 1898 [1] and his treatise [13] as well as his investigation of the coefficient of dispersion.

When substantiating least squares, Markov in essence introduced new important notions identical with the now current concept of unbiased and effective statistics for estimating the constant parameters of laws of distribution when issuing from samples [5]. Markov defined them in common professional terms of observations or measurements corrupted by error. For substantiating the MLSq he assumed three propositions [1, p. 163] {The authors quote Markov, see p. of this book}.

He deduced the *rule of the arithmetic mean* in the simplest case of measuring one magnitude a when the results of independent observations of equal precision were $X_1, X_2, \ldots, X_n$ and without introducing any assumptions about the law of distribution of the errors, $\Delta = X - a$, except for the existence of expectation and variance and absence of a constant error ($E\Delta = 0$).

Indeed, when considering linear functions $\lambda_1 X_1 + \lambda_2 X_2 + \ldots + \lambda_n X_n$ *without a constant error* (unbiased) so that

$$E(\lambda_1 X_1 + \lambda_2 X_2 + \ldots + \lambda_n X_n) = a,$$

it occurs that among these the arithmetic mean, for which $\lambda_i = 1/n$, has the least variance.

Later investigations proved that under normality and in the sense of the Markov principles the arithmetic mean becomes most advantageous not only among all the linear functions of observation, but also among the wide class of all *regular* and unbiased statistics.

18. In a more general case m numbers (having no known connections with each other), $a_1, a_2, \ldots, a_m$, are determined, given the values of

$$b_l = y_l + \Delta_l,\ l = 1, 2, \ldots, n,\ n > m$$

where

$$y_l = A_1^{(l)} a_1 + A_2^{(l)} a_2 + \ldots + A_m^{(l)} a_m,\ i = 1, 2, \ldots, m;\ l = 1, 2, \ldots, n,$$

$A_i^{(l)}$ are known numbers and Δ_l are the random errors of observation possessing $E\Delta_l = 0$ and var $\Delta_l = h/p_l$ and p_l are the weights of the observations. Here, the same principles are applied. The numbers a_l are estimated by linear statistics under unbiasedness and minimal variance. These statistics coincide with those determined by least squares, hence the justification of the method from the viewpoint of the Markov principles.

19. Markov [11] studied the coefficient of dispersion by considering σ series of S_i independent trials in series i ($i = 1, 2, \ldots, \sigma$) where some event E with constant probability P of occurrence was observed. If X_i is the number of its occurrences in series i, $n = S_1 + S_2 + \ldots + S_\sigma$, $m = X_1 + X_2 + \ldots + X_\sigma$, then the coefficient of dispersion is

$$Q = n(n-1)\frac{\sum_i S_i[(X_i/S_i) - (m/n)]^2}{(\sigma - 1)\, m\, (n - m)}.$$

Markov calculated the expectation and the variance of Q and thus made it applicable for testing hypotheses on the constancy of P (the hypothesis of homogeneity of the series of trials) [6]. In a later work he [12] provided the asymptotic distribution of Q as $n \to \infty$ thus specifying the test for large values of n.

In these works Markov also criticized the use of the sample correlation coefficient for estimating the general coefficient when the type of distribution of the general population was unknown.

Notes

1. {More precisely, the Bienaymé – Chebyshev method. The same should be borne in mind in §5.}

2. {Markov's contribution [1] consisted of extracts from letters to Professor Vasiliev (Kazan).}

3. {Markov had indeed mentioned *for example* Poincaré in connection with the so-called hypothesis of elementary errors. In actual fact, it goes back to the first half of the 19$^{\text{th}}$ century (Hagen) and even to Daniel Bernoulli.}

4. {Bibliographies state the date of publication of [2] as 1906, but Sapogov (see his Note 2 to [2] in this book) noted that Markov had dated the last part of [2] as 1907.}

5. {The authors could have just as well attributed these concepts to Gauss. In general, their discussion of Markov's work on the MLSq is faulty; Markov hardly contributed anything to what was published by Gauss.}

6. {The authors should have mentioned Bortkiewicz and Chuprov if not Lexis, the originator of the *theory of stability of statistical series.*}

References

Except for Ref. [3], the French titles of Markov's contributions are those that he himself additionally provided for his Russian works. His memoirs [1; 2; 4; 5; 11] are translated in this book.

1. Markov, A.A. (1899), The law of large numbers and the method of least squares.

2. --- (1907), Extension de la loi des grands nombres, etc.

3. --- (1907, in Russian), Recherches sur un cas remarquable d'épreuves dépendantes. *Acta math.*, t. 33, 1910, pp. 87 – 104.

4. --- (1908), Extension des théorèmes limites du calcul des probabilités, etc.

5. --- (1910), An investigation of the general case of trials, etc.

6. --- (1911), Sur les valeurs liées qui ne forment pas une chaîne véritable [14, pp. 399 – 416].

7. --- Sur un cas d'épreuves liées an chaîne multiple [14, pp. 417 – 436].

8. --- (1912), Sur les épreuves liées en chaîne par les événements laissés sans observation [14, pp. 437 – 463].

9. --- (1913), The theorem on the limit of probability for the Liapunov cases [14, pp. 319 – 338]. Translated in P.A. Nekrasov (2004), *Theory of Probability.* Berlin.

10. --- (1915), Sur l'application de la méthode des ésperances mathématiques aux sommes liées. *Izvestia* AN Psb*,* ser. 6, t. 9, No. 14, pp. 1453 – 1484.

11. --- (1916), On the coefficient of dispersion.

12. --- (1920), Sur le coefficient de la dispersion. Note seconde [14, pp. 537 – 547].

13. --- (1900), *Исчисление вероятностей* (Calculus of Probability). Psb. Later editions: 1908, 1913, 1924; German translation: Leipzig – Berlin, 1912.

14. --- (1951), *Избранные труды* (Sel. Works). N.p.

15. Bernstein, S.N. (1944), Extension of the limit theorem of the theory of probability onto sums of dependent variables. *Uspekhi Matematich. Nauk,* vol. 10, pp. 65 – 114. (R)

16. Bruns, H. (1906), Das Gruppenschema für zufällige Ereignisse. *Abh. Sächs. Akad. Wiss. Leipzig*, math. – phys. Kl., Bd. 29, No. 8, pp. 577 – 628.

17. Chebyshev, P.L. (1887, in Russian), Sur deux théorèmes relatifs aux probabilités. *Acta math.*, t. 14, pp. 305 – 315.

18. Romanovsky, V.I. (1949), *Дискретные цепи Маркова* (Discrete Markov Chains). M.

18. A.A. Markov, Jr. The Biography of A.A. Markov

Translator's Foreword

Markov Jr was an eminent mathematician in his own right. He, in his own words, attempted to describe his father as a man and a citizen but the reader will perceive some inevitable political undertones; in one case (Note 5) I allowed myself to disagree with the author.

I list more recent work on the same subject in the References to this Foreword, and I mostly compiled/specified the references to the essay itself.

* * *

[1] My father, Andrei Andreevich Markov, was born on 2 June (old style) 1856 in Ryasan. He was a son of Andrei Grigorievich Markov, an official working at the Forestry Department and holding the rank of collegiate councillor, then resigning and becoming an attorney, or solicitor. Andrei Grigorievich's father, Grigori Markovich Markov, was a deacon in the countryside somewhere near Ryasan [1].

A.G's retirement was connected with his revealing the abuses practised by the higher-ranked officials at the Forestry Department. To cover the tracks of their fraud, the influential swindlers compelled him to retire. A.G. was married twice. By his first wife, Nadezhda Petrovna, a daughter of an official, he had six children: Petr, Pavel (who died in childhood), Maria, Evgenia, *Andrei* and Mikhail; by his second wife, Anna Iosifovna, he had three: Vladimir, Lidia and Ekaterina.

From among the father's brothers, Vladimir Andreevich (1871 – 1897) became well-known as a first-class mathematician, but tuberculosis brought him to his grave at the age of 26. A sister of my father, Evgenia Andreevna, was one of the first Russian women doctors. A psychiatrist by speciality, she worked for a long time in various hospitals for mental patients. She died in 1920. Also a doctor is Ekaterina Andreevna (born 1875), another sister of my father now living in Leningrad. {Yet another} sister, Lidia Andreevna, was a teacher in several high schools. She died in Leningrad in the beginning of 1942. Mikhail Andreevich, a brother of my father, who died at a great age shortly before the Great Patriotic War {1941 – 1945}, was for many years a forester in the Ukraine.

In the beginning of the 1860s, Andrei Grigorievich with his family moved to Petersburg. Upon his retirement he became the steward of the estate of Ekaterina Aleksandrovna Valvatieva who had two daughters, Maria (1860 – 1942) and Elizaveta.

[2] Andriusha Markov was a sickly child. He suffered from tuberculosis of a knee-joint and walked on crutches. He was nevertheless able to manage without them, but then he hopped on one leg with the other one never unbending at the knee. Still, he achieved perfection in both these modes of movement and was even able to play successfully race-and-catch.

When Andriusha was ten, the famous surgeon Kade {Cade? Cadet?} operated on him and unbended his leg thus enabling him to walk normally. True, he slightly limped all his life, but this did not prevent him from becoming a good walker and a lover of wanderings. He liked to repeat the words of a doctor addressed to a postman: *You will remain alive as long as you keep moving.*

In 1866 Andriusha entered the 5th Petersburg gymnasium. This *classical* educational institution with its red tape was not to the boy's liking. All through his life he retained a gloomy memory of this school whose teachers had tried not so much to teach, but rather to drill and to dull him. Especially conducive to this aim was the teaching of the classical languages (Latin and Greek) founded on cramming countless rules and exceptions of all kinds, {rhymed grammatical sayings}, etc. into the students' heads. In most of the subjects, he got on rather poorly and often brought home unsatisfactory marks. Mathematics appeared to be the only exception: in this discipline Andriusha invariably got the highest grade. Andrei Grigorievich once said:

I am not worried about my daughters Masha and Enia {Maria and Evgenia}*, they are learning successfully. But Andrei is giving me great trouble. The Director had again summoned me. Andrei does not want to study anything except mathematics.*

Even as a school student, Andrei was indeed very fond of mathematics, and he studied it on his own. For some time he thought that he had invented a new method of integrating ordinary differential equations with constant coefficients. He reported his finding to the well-known contemporaneous Russian mathema-ticians V.Ya. Buniakovsky, E.I. Zolotarev and A.N. Korkin. The first did not answer the letter of a school student whereas the other scientists thoroughly explained him that his method was not really new. Thus my father struck up an acquaintance with two professors at Petersburg University.

Andrei Grigorievich erred, however, when he thought that his Andrei was not interested in anything except mathematics. Actually, he was engrossed in the pamphlets of the great public figures of the 1860s, Chernyshevsky, Dobroliubov, Pisarev, under whose influence the best part of the student body had been then [2].

In one of his school compositions about {Pushkin's} *Evgeni Onegin* Markov treated it in Pisarev's spirit for which his teacher vouchsafed him the following remark: *You have read a lot of hack writers who had denied the sense of the beautiful*.

An incident nearly resulting in Andrei's expulsion from the gymnasium happened during his study in the graduation class. Once, during the prayer after classes, he was packing up his books. As ill luck would have it, the last lesson was conducted by the *German* who zealously sought for occasions to nag his students. He not so much prayed but rather watched after the students' behavior at the prayer. *You are disturbing the reverential feelings of the class*, he declared after the prayer. *I shall inform the Director*. When Kapustin, Andrei's school fellow who later became a physicist and Professor at Petersburg University, had attempted to plead for him and told the teacher that the reverential feelings of the class were not disturbed, the teacher got into a rage and shouted: *You are a none advocate* [3]! *Off with you to the punishment cell*!

The teacher complained to the Director who summoned Andrei Grigorievich Markov and declared to him that he will not tolerate *atheists or nihilists* in his gymnasium. Nothing is known about the means to which A.G. had to resort to pacify the Director, a three-fold turn-coat who changed his religion a few times and finally decided on Orthodoxy.

[3] In 1874 my father graduated from the gymnasium and entered Petersburg University. There, he attended the lectures of Chebyshev, who influenced his entire scientific work, as well as those of Professors Korkin and Zolotarev. In addition to reading lectures, the two last-mentioned mathematicians conducted study groups consisting of the best students. My father participated in them and was quick to solve the difficult problems formulated there. As he himself owned, his conversations with Korkin were the starting-point of many of his independent works.

On 31 May 1878 he graduated from the Mathematical Department of the Physical and Mathematical Faculty of Petersburg University as a candidate {of Sciences}. The same year he was awarded a gold medal for a composition on a topic proposed by the faculty, *On the integration of differential equations by means of continued fractions*, and was left at the University *to prepare himself for professorship*. In 1880 he defended his celebrated Master Dissertation, *On the binary quadratic forms of a positive determinant* which at once advanced him to the first ranks of Russian mathematicians. In 1881 he defended his Doctor Dissertation, *On some applications of algebraic continued fractions*.

[4] In 1880 he began teaching at Petersburg University as a Privat-Docent. In 1880/1881 and 1881/1882 he read a refresher course in differential and integral calculus, and, in 1883, he took over the *Introduction to Analysis* previously read by Yu.V. Sukhotsky and K.A. Posse. Also in 1883, Chebyshev had left the University and my father read, for the first time, a course in the theory of probability. He invariably read it from 1885/1886.

In 1883 he married Maria Ivanovna Valvatieva. My parents became acquainted already during their childhood, in the country-house of Ekaterina Aleksandrovna Valvatieva. Later, when my father became a student, and my mother, a school student with mathematics not coming easily to her, E.A. invited *the steward's son* as a mathematics tutor for her daughter.

He soon sought Maria's hand in marriage but was not given consent at once. For a long time my grandmother hesitated being careful not to give away her daughter to a man insufficiently, as she thought, prosperous and having no definite *position*. She only assented in 1883, when my father already was Privat-Docent, about to defend his Doctor Dissertation with the prospect of a professorship taking shape, and the wedding was celebrated.

On 13 December 1886, in accord with Chebyshev's proposal, my father was elected *adjunct* of the {Imperial Petersburg} Academy of Sciences becoming Extraordinary, and then Ordinary Academician on 3 March 1890 and 2 March 1896 respectively. In 1886 he was appointed Extraordinary Professor at Petersburg University, and Ordinary Professor in 1893.

He resigned in 1905 remaining Meritorious Professor but continued to read his course in probability theory. Attempting to find a useful practical application for his main scientific speciality, he actively participated in the calculations performed at the Pension fund at the Ministry of Justice both during its establishment and when reviewing its activities.

[5] The scientific significance of my father's work is discussed in detail in another article of this collection [5] and I shall not dwell on this subject. I would {rather} like to throw light on the other aspects of his life and to say what kind of a man and a citizen he was. He was an open-hearted, frank and bold man. He never betrayed his convictions; all his life he furiously battled with everything considered by him as stupid or harmful. His civil courage was staunch; he did not reckon either with the figures against whom he came out or with the possible consequences for himself. When objections were once raised to one of his suggestions because of its being in conflict with an *Imperial instruction*, he answered publicly: *I am talking business but I hear from you nothing except Imperial instruction.*

Here are a few facts and documents characterizing my father. It is known that on 25 February 1902 the united sitting of the Branch of Russian Language and Literature and the Section of Belles-Lettres of the Imperial Academy of Sciences elected A.M. Gorky [4] Honorary Academician and an official report about this fact appeared in the *Pravitelstvenny Vestnik*. It infuriated the Czar Nikolai II who decided to annul immediately the election. To this aim, he, through the Minister of People's Education P.S. Vannovsky, demanded from the *Most August* President of the Academy, the Grand Duke Konstantin Konstantinovich Romanov, that the same *Vestnik* publish, on behalf of the Academy, an announcement that the election was annulled. Such an announcement had indeed appeared, although at first (on 10 March) not from the Academy, but then (on 12 March), at a new request from Vannovsky, on its behalf. As stated in the announcement, the reason for the annulment was that Gorky was allegedly under open police surveillance. This was an example of rude arbitrary rule by the *autocrat* and by the President of the Academy of Sciences who trembled before him [5] and acted on the behalf of the Academy without the knowledge or consent of its General Assembly. The following statement was my father's response:

To the General Assembly of the Academy of Sciences
I have the honor of proposing to the Assembly that it should insist that the announcement {itself} *about the annulment of the election of Mr. Peshkov to honorary membership be declared null and void or corrected because, first, it was made on behalf of the Academy which actually did not annul his election; and, second, the stated reason for the annulment is insignificant.*
6 April 1902. A. Markov

He wanted to read out his statement at the sitting of the General Assembly, but the President did not permit it and the then Perpetual Secretary, N. Dubrovin, appended an instruction on the statement: *To be left outside the official records.* This meant that the statement was not to appear in the minutes of the Academy of Sciences. Thus the Czarist myrmidons attempted to deaden any public actions of the scientists directed against the Imperial arbitrary rule.

My father then made the next step: on 8 April 1902 he submitted a request for resignation to the President. I did not see its text and its very existence is only known from other documents.

The President naturally attempted to conceal it from the public. My father also stated that, accordingly, he will not further participate in the publication of Chebyshev's {collected} contributions by the Academy. His resignation was not, however, accepted, and he resumed his work of publishing Chebyshev's writings being apparently afraid of handing over this responsible duty to someone else.

In the beginning of 1905 he again protested against the annulment of Gorky's election. He submitted the following statement to the Academy of Sciences:

To the General Assembly of the Academy of Sciences

Wishing to believe, and to hope that the Imperial Decree of 12 December 1904, that acknowledged as urgent the <u>taking of real measures for protecting the full force of the law … so that its inviolable and equal for everyone execution would be considered the very first duty of all the authorities and regions subordinated to us,</u> – opens for Russia a new era of legality. I consider it my duty to recall to the General Assembly the unprecedented case of violating the law which concerns the Honorary Academician Peshkov who is still not entered in the academic list and is deprived of the possibility of enjoying the rights of honorary membership. The annulment of Mr. Peshkov's election to honorary academician was naturally announced in the newspapers as having been done as though by the Academy of Sciences; but we know that this announcement was false. Suchlike announcements can be valid only when limitless arbitrary rule is reigning and fall all by themselves, once the latter is removed.

I am therefore proposing to enter the name of Mr. Peshkov in the list of honorary academicians and to invite him, in accord with the law {with the by-laws}, *to participate in the life of the Academy.*

8 January 1905. Academician A. Markov [6]

In accordance with §§97 and 98 of the Academy's charter this statement was not permitted to be read out either. Gorky was only invited to a sitting of the Branch of Russian language and Literature and the Section of Belles-Lettres after the February revolution of 1917 [7].

In March 1903 my father submitted the following application to the Board of the Academy of Sciences:

Since, in accord with the Board's decision, a notification was sent to me about the procedures for deducting the payment for orders {from holders of decorations}, *I have the honor to ask most humbly to consider that I do not request any orders, nor do I wish to be decorated.*

23 March 1903. A. Markov

The essence of this application was naturally not that my father did not want to pay the deductions for the decorations which were then levied on their holders (the deductions were insignificant). The essence consisted in that he refused to be further decorated by the Czarist government and that he disdainfully regarded those orders that he already had. And he took the first occasion to declare this.

As is well known, on 3 June 1907 the Czarist government dissolved the disagreeable 2nd State Duma {the Parliament} and issued a new law on electing the 3rd Duma. It thus violated its own Manifesto of 17 October 1905 according to which it could only have promulgated new laws with the Duma's consent. And so, my father applied to the Board of the Academy of Sciences with the following request:

Since the convocation of the 3rd State Duma is connected with a violation of the law, so that it will not be an assembly of the people's representatives but some illegal medley, I have the honor of humbly begging the Board not to enter me in the list of voters. […]

S. Oldenburg, the then perpetual Secretary of the Academy of Sciences, appended the instruction: *No action to be taken*. In 1908, as a result of the students' unrest, the Ministry of People's Education issued a circular letter
where it attempted to put {some} police duties in the hands of university professors. My father submitted the following appropriate request to the Ministry:

To His Excellency, the Minister of People's Education
Concerning the well-known circular letter that was based on an interpretation formulated by the Senate and shown to me on 25 September in the office of Petersburg University, I consider it my duty to inform Your Excellency that I decidedly refuse to be a government agent at the University, although, in accord with the desire of the Physical and Mathematical Faculty, I reserve for myself the reading of lectures on the theory of probability.
2 October 1908. Academician A. Markov

An appended instruction of the Ministry reads: *To be returned to Academician Markov as improperly submitted.*
The *Most Holy* Synod is known to have excommunicated the great Russian writer, Lev Tolstoy, from the Russian Orthodox Church. To reveal all the absurdity of this act smacking of the Middle Ages, my father submitted a request to the Synod for being excommunicated from the Church as well. Here is its complete text [8].

To the Most Holy Governing Synod
I have the honor of humbly asking The Most Holy Synod to excommunicate me from the Church. I hope that a reference to my book [6] *can serve as a sufficient cause for my excommunication. There, my negative attitude to the legends underlying the Judaic and the Christian religions is clearly expressed. Here is a passage from this book* (*pp*. 213 – 214) {p. 320 in the edition of 1924}: <u>*Independently of the mathematical formulas on which we shall not dwell because we do not attach to them any great importance*</u> [9], <u>*it is clear that we should regard stories about incredible events allegedly having occurred in the bygone times with extreme doubt. And we cannot at all agree with Acad. Buniakovsky*</u> [1, p. 326] <u>*that it is necessary to isolate a certain class of stories which he considers it reprehensible to doubt.*</u>
<u>*In order not to deal with still stricter judges and to avoid accusations of shaking the foundations, we shall not dwell on this subject that does not directly bear on mathematics.*</u>
So that there will remain no doubt about what I am speaking, I adduce the appropriate passage from Buniakovsky's book: <u>*Some philosophers, pursuing reprehensible aims, attempted to apply the formulas concerning the lowering of the probability of testimonies and legends to religious beliefs and thus to shake them.*</u>
If the provided passage is not sufficient, I humbly beg to take into consideration that I do not see any essential difference between ikons and idols which are of course not gods but their pictures, and that I do not sympathize with religions which, like Orthodoxy, are supported by, and in turn lend their support to fire and sword.
12 February 1912. Academician A. Markov

It was not possible anymore to return this request as *submitted improperly*. The request caused a terrible commotion in the government camp. The newspapers of the Black-Hundred orientation raised a wild howl. The Petersburg Metropolitan sent a *spiritual pastor,* the archpriest Ornatsky, to *admonish and exhort* my father who declared, however, to be prepared only to speak with Ornatsky about mathematical issues, a decision that did not at all suit the clergyman. The Synod had to comply with my father's request; in this connection it was amusingly investigated whether or not he was a sectarian; whether he was baptised; who were his parents; etc. [11]

In 1913, at the government's bidding, the tercentennial of the House of Romanovs' was celebrated. To counterbalance this false Black-Hundred {?} jubilee, my father organized a scientific anniversary, the bicentennial of the law of large numbers. In one of his letters to Oldenburg, Aleksei Nikolaevich Krylov [12] wrote about this struggle of my father *for justice* with the previous government in the following way [3, p. 320]:

You certainly remember his sharp protest addressed to the Academy about Gorky's exclusion from the Academy on the instruction of the Minister Sipiagin [13]. *You apparently also remember that at the very first sitting of the Academy of Sciences in March 1917* [14] *Andrei Andreevich* {once more}*put forward an unanimously carried proposal to reenter Gorky in the list of honorary academicians. And how many times did Markov come out at the University with protests against the measures taken by the previous government and the police with respect to the University and the student body. For some time he was even discharged from professorship for these activities. Recall his protest against the Synod concerning Tolstoy's excommunication from the Church; as you know, such protests, always openly and explicitly stated, were countless, and Markov's very name and the scientific glory certainly gave them strength and dissemination that were not conducive to the consolidation of the previous government.*

[6] My father paid great attention to the teaching of mathematics in the high school. He vigorously protested against various harmful experiments in this field. In particular, P.A. Nekrasov, a {former} Professor at Moscow University, attempted to carry out such experiments. A member of the Black Hundred and a mystic, he aimed at transforming mathematics into a support for Orthodoxy and autocracy. In 1915, this former warden of an educational {of the Moscow educational} region, connected with the governing body of the Ministry of People's Education, came out together with P.S. Florov with a plan for introducing the theory of probability into the school curriculum. Actually, their plan came to the inculcation of their own confused and pseudo-scientific views on the theory of probability, mathematical statistics and mathematics in general in the students' minds. On my father's initiative, the Academy of Sciences set up an ad hoc commission which destructively criticized that plan [8]. {Accordingly,} it was not implemented, although Florov, who headed a non-classical high school {*Realschule*} in Uriupinsk, performed some {appropriate} experiments.

In the 1917/1918 academic year my father himself taught in a high school. […] He solicited the Academy of Sciences for a yearly trip *for continuing scientific work within Russia* and we spent the winter in Zaraisk. In this case, *we* were my parents, my mother's aunt Serafima Aleksandrovna Moskvina, and I. I had to transfer from the 5th form of the 8th Petrograd [15] gymnasium to the 6th form of the Zaraisk non-classical school. Mathematics in our class was taught by the Director, Gilweg, a diligent and pedantic person. He invariably demanded that all the figures during the lessons in geometry be drawn by ruler and compasses. He suddenly resigned and the higher forms were left without a mathematics teacher.

To help the school out of trouble and also perhaps satisfying his permanent need to teach (until then, he had indeed taught for 37 years in succession at the University), my father offered his free services as a mathematics teacher. His proposal was gratefully accepted and I thus became his official pupil. His first lesson somewhat frightened and puzzled us. Nothing resembled the outward carefulness of his predecessor. My father did not even write out the formulas exactly one under another one. Then {however} his young pupils began to get used to the *Professor* (as they called him) but for some time the habits acquired from Gilweg kept preventing a mutual understanding. Thus, my father once called one of the best school students to the blackboard and became furious when the pupil, instead of explaining the solution of the problem, began thoroughly drawing the appropriate figure with ruler and compasses. My father cried out angrily: *Our lesson is in geometry, not mechanical drawing*.

He put emphasize on the solution of problems. And he organized additional classes during vacations and on Sundays for those desiring to advance themselves in this direction. Many pupils gladly attended and benefited accordingly. The teachers' assembly presented him a letter of thanks for his short but intensive work, and I am still keeping it.

[7] In the fall of 1918 our family returned to Petrograd. My father, who invariably attempted to do everything in his power for the welfare of his mother country, overwhelmingly longed for his pedagogic work at the University which then began to return to normal.

In Petrograd he had to undergo an eye operation. For some years, he was suffering from glaucoma, a grave eye disease, that manifested itself in many representatives of our family at their old age. Dr. Vygodsky successfully operated my father, his eyesight improved and he resumed the reading of lectures at the University.

However, the move from Zaraisk to Petrograd told on the general state of his health. In 1920/1921 I walked him arm-in-arm to the lectures which had never been previously needed. He lectured on the theory of probability conforming to his well-known book. As a listener, I may testify that, although barely able to keep himself on his legs, he read them faultlessly. At the same time my father intensively worked on the fourth edition (published posthumously) of his book [6] which is known to differ considerably from its previous edition.

In the fall of 1921 my father became bedridden. He suffered from a grave kind of radiculitis accompanied by racking pain. In the spring of 1922 a new disease, aneurism, that developed in a leg, joined in. Bleeding began. By that time, however, my father became very tired of lying in bed. He began craving for open air, for being close to nature. The doctor who treated him permitted a move to the health center for scientists in Detskoe Selo (present name of the city, Pushkin). This was probably a mistake: the ride in a car unfavorably affected my father and the bleedings intensified. The head physician of the center decided that an immediate return to Petrograd for an operation, for the ablation of the aneurism, was necessary.

After the operation my father began feeling himself better and the temperature lowered, but in a few days there appeared threatening symptoms and {in particular} the temperature rose sharply. A conference of specialist doctors ascertained a general blood-poisoning and pronounced his condition hopeless. On 20 July 1922, at 10 p.m., my father passed away. He was buried in the Mitrofanovsky cemetery in Leningrad.

In 1923 the Academy of Sciences observed the first anniversary of his death. His friend, the then vice-president of the Academy, Vladimir Andreevich Steklov, vividly addressed the meeting describing my father as a scientist, a man and a citizen. Thirty years have now passed since he had died, but Markov the scientist did not, and will not die. His ideas and his findings,– the celebrated *Markov chains*, the proof of the {generalized} law of large numbers, the theorems on the minimal values of quadratic forms and his other splendid achievements were included in the main stock of science and will remain alive for centuries.

Notes

1. My father's biography{written by A.S. Besikovich} included in the posthumous edition of his textbook [6] mistakenly states that A.G. himself was a deacon. This mistake is repeated in some of his other biographies.

2.{Nikolai Gavrilovich Chernyshevsky (1828 – 1889), Nikolai Aleksandrovich Dobroliubov (1836 – 1861) and Dmitry Ivanovich Pisarev (1840 – 1868) were militant writers opposed to the Establishment and more or less advocating a peasant revolution. Pisarev (cf. below) was also a staunch antagonist of Pushkin's poetry.}

3.{I render the broken phrase that apparently included an intentionally preserved mistaken loan-translation made by the teacher of German.}

4.{The pen name of the eminent writer Peshkov, as properly named below.}

5.{The *trembling* is doubtful. The President (born in 1858 and in office since 1889 [4, p. 474]), was about ten years older than Nikolai II (born 1868) and enjoyed more than enough influence. Note that he had not complied at once.}

6.{I adduce a related letter written by Markov on 23 February 1905 (Archive, Russian Academy of Sciences, Fond 173, Inventory 1, 60, No. 27) to the President of the Academy of Sciences.

Your Imperial Highness

On 18 Febr. I received a note signed by Your Highness on 4 February. I have no desire to criticize it, but the signature of the Most August President of the Academy of Sciences compels me to make several comments.

And first of all I consider it necessary to state that I cannot change my convictions by order of the Authorities. Then, as Professor at a university teaching <u>Differential and Integral Calculus</u> and the <u>Theory of Probability</u>, I ought to turn Your benevolent attention to the fact that in general the academic institutions are not within the province of the President of the Academy of Sciences and that only persons fully mastering a certain subject may properly judge the methods of its teaching.

Finally, perceiving in the note a suggestion to retire, I have the honor to report to Your Imperial Highness that I shall immediately leave the Academy as soon as its General Assembly considers my membership superfluous.}

7.{Which preceded the *Great October* (old style) *Socialist Revolution* of 1917, as it was called by the Soviet authorities.}

8. Archive, Acad. Sci. Soviet Union, Fond 173, Inventory 1, No. 65.

9.{In one instance, Markov [7, p. 32] declared that the study of testimonies was the *weakest section of the theory of probability*.}

10.{Bearing in mind this date, several authors mistakenly assumed that Tolstoy (who died in 1910) was excommunicated in 1912. Actually, this happened in 1901, but, during Tolstoy's last days, the Synod discussed whether he should be admitted *to the bosom of the church*. Nevertheless, he remained excommunicated [9, p. 62n]. This goes to show that in 1912 the whole story was likely still well remembered but does not explain why Markov had not submitted his petition two (if not twelve) years earlier.}

11.{Actually, Markov was not excommunicated; the Synod resolved that he *had seceded from God's Church* and expunged him *from the lists of Orthodox believers* [2, p. 408].}

12. {A naval architect and an applied mathematician.}

13.{Above, in the same connection, the author named another minister.}

14. {Which means: the very first sitting after the February revolution; cf. Note 7.}

15.{The name of Petersburg from 1914 to 1924.}

References

1. Buniakovsky, V.Ya. (1846), *Основания математической теории вероятностей* (Principles of the Mathematical Theory of Probability). Psb.

2. Emeliakh, L.I. (1954), The case of Markov's excommunication from the church. *Voprosy Istorii Religii i Ateisma*, vol. 2, pp. 397 – 411. (R)

3. Krylov, A.N. (1936), *Собрание трудов* (Coll. Works), vol. 1, pt. 2. M.– L.

4. Levshin, B.V. et al, Compilers (1974), *Академия Наук СССР, персональный состав* (Academy of Sciences of the Soviet Union, Personnel), vol. 1. M.

5. Linnik, Yu. V. et al (1951), A sketch of the work of Markov, etc. Translated in this book.

6. Markov, A.A. (1900), *Исчисление вероятностей* (Calculus of Probability). Psb. Later editions: 1908, 1913, 1924; German translation: Leipzig – Berlin, 1912.

7. --- (1915), On Florov’s and Nekrasov’s draft. *Zhurnal Ministerstva Narodn. Prosveshchenia*, vol. 57, No. 5, pp. 26 – 34 of the section on current chronicle. (R)
8. Markov, A.A. et al (1916), Report of the Commission to discuss some issues concerning the teaching of mathematics, etc. *Izvestia* AN Psb, ser. 6, vol. 10, No. 2, pp. 66 – 80. Translated in P.A. Nekrasov (2004), *Theory of Probability.* Berlin.
9. Sheynin, O. (1990, in Russian), *Chuprov.* Göttingen, 1996.

Additional References

Grodzensky, S.Ya. (1987), *Markov*. M. (R)
Sheynin, O. (1989), Markov’s work on probability. *Arch. Hist. Ex. Sci.*, vol. 39, pp. 337 – 377.

Supplement

19. Oscar Sheynin.On the Probability – Theoretic Heritage of Cournot

. *Istoriko-Matematich. Issledovania*, vol. 7 (42), 2002, pp. 301 – 316.

1. General Information

Antoine Augustin Cournot (1801 – 1877), whose collected works have been recently appearing, [1] is known as a mathematician, philosopher, economist, educationist and author of memoirs. Cournot’s life and work are described by Feller (1961), Weinstein et al (1970) and in Cournot (1978). In mathematical analysis, he is meritorious for his treatise (1841) reprinted in 1984 with a preface by Dugac who states that, with regard to rigor of explication, Cournot followed Lacroix and agrees with Weierstrass in that he contributed to the development of the notion of function. Dugac also quoted Poisson (without, however, providing an exact reference) who had called Cournot rather an amateur of the highest rank than a geometer in the strict sense. Somewhat earlier Dugac devoted to Cournot a short note (1978). I remark that Dugac did not regrettably appraise the methodological value of Cournot’s treatise (1841).

I mostly discuss Cournot’s *Exposition* (1843) [2] in which he treated the theory of probability and its applications. After Laplace, in 1812, published his *Théorie analytique des probabilités*, the need to popularize that discipline began to be felt. Several authors attempted to accomplish that aim, – Lacroix, Cournot, De Morgan, and, in Russia, Buniakovsky. Following Lacroix, Cournot, as he stated in his Préface, destined his book for readers *qui n’ont cultivé les hautes parties des mathématiques* and he promised to correct the mistakes and to “dissipate the obscurities” left by *des plus habiles géomètres*.

Cournot had indeed only applied mathematical analysis in some of his footnotes but he thus did his readers a disservice [3]. At the same time, however, in Chapter 12 he applied terms of spherical astronomy and formulas of spherical trigonometry. His style was ponderous. Sentences of seven or eight lines and even longer (e.g., in §§19, 162, 201, 207) are often met with, and §117 begins by a phrase of 12 lines. Many elementary calculations (for example, in §§13, 70, 165 – 167, 170, 182, 203, 204) were wrong which, however, did not corrupt Cournot’s final conclusions, and, finally, he certainly referred to his predecessors too seldomly [4].

I especially note that Cournot in essence passed over in silence such scholars as De Moivre and Gauss. Up to the 20th century, mathematicians effectively forgot the latter’s definitive approach to the treatment of observations and recalled De Moivre’s part in proving the De Moivre – Laplace limit theorem even later. Then, although Cournot, in his Préface, cordially mentioned his late contemporary, Poisson (who died in 1840), and somewhat earlier devoted his treatise (1841) to Poisson’s memory, and, moreover, applied the latter’s term, *law of large* *numbers*, in a previous memoir (1838), he did not say a word about that law in his book [5]. Commentators are unanimous in that the reason for that attitude was the extremely negative (and nowadays largely forgotten) criticism of the LLN by his *excellent ami* (Préface, p. 6), Bienaymé (Heyde et al 1977, §3.3).

2. The Main Concepts of the Theory of Probability

2.1. The Theory of Probability. Its aim, as Cournot (1875, p. 181) formulated it, was to create methods for assigning numerical values to probabilities. He thus moved away from Laplace's natural-scientific viewpoint (Sheynin 1976, p. 176) who had thought that the theory was a means for discovering the laws of nature. Recall that Chebyshev (1845, p. 29) formulated the goal of probability theory as the calculation of the unknown probabilities [of events] given the probabilities of other [events] and Boole(1851, p. 251) was of the same opinion, he only replaced Chebyshev's events by propositions.

Cournot (1875, p. 180) had also stated that the tardy origin of the theory of probability was caused by pure chance, but many later authors including myself (1977, p. 204) have not agreed with him at all.

2.2. Probability. Cournot (1843, Préface) published Poisson's letter of 1836 where that scholar had explained that he distinguished between the terms *chance* and *probability* in the same way as Cournot did, and that he *"beaucoup insisté* [insisted] *sur cette difference".* A bit later Poisson (1837, pp. 30 and 31) specified that probability depended on our knowledge and was therefore subjective, whereas chance was objective. This understanding is now usually ignored [6]. Cournot explained his point of view in §48 rather than in the opening pages of his treatise and not clearly enough. He insisted, however (see my §5), that probabilities could be both subjective and objective as well as theoretical and posterior and provided a simple example (§86) in which the probability of a random event could have been only determined experimentally [7].Finally, he (§113) introduced probabilities unyielding to measurement and (§§233 and 240.8) called them philosophical. Such probabilities, as he thought, often served for making judgements. In Russia, Davidov (Ondar 1971), in the mid-century, recognized them and nowadays philosophical probabilities can be interpreted as expert estimates from whose treatment modern statistics cannot isolate itself.

Cournot (§18) also heuristically defined probability as the ratio of the extent (*étendue*) of the favorable chances to that of all of them; a modern definition replaced *extent* by *measure.* I stress that Cournot's definition included geometric probability that until him had been lacking any formula, and thus combined it with the classical discrete case. Newton considered geometric probability in a manuscript of 1664 – 1666, Jakob Bernoulli mentioned it, and Buffon and Laplace applied it; on its history see Sheynin (2003a). The formal birth of the earlier concept was, however, hardly noticed.

2.3. Density of Distribution. Mathematicians made use of continuous distributions since Niklaus Bernoulli [8]. In 1757 Simpson applied the continuous uniform and triangular laws to prove that (for them) the arithmetic mean was (stochastically) preferable to an isolated observation. Cournot (§§64 – 65) formally introduced a curve representing the *loi des probabilités des diverses valeurs de la variable* [9]. Also see his §§124 and 125; in §125 he maintained that the computation of (empirical) *tables* of probability was *comme le chef-d'oeuvre de la statistique.*

In addition, Cournot (§73) described the derivation of the density law of a function η of a random variable ξ. Let, in my notation, $\eta = f(\xi)$ and $\varphi_1(x)$ be the density of ξ. It is required to derive $\varphi_2(x)$, the density of η. Then, in §74, he generalized that problem by considering a function of two variables and examined two examples but did not provide any general formulas. For that matter, Cournot was mostly interested in such characteristics of the studied functions as their means, medians (a notion and a term that he himself introduced in §34), as *modules de convergence* (§126, – a statistic inversely proportional to variance). Already Bessel (1838, §§1 – 2) derived densities of random functions which possibly only remained known to astronomers. And of course many authors beginning with Simpson studied the stochastic behavior of the arithmetic mean in the theory of errors.

Following Poisson (1829, §6), Cournot (§81) introduced a mixture of densities. Let n_1, n_2, … observations have densities $f_1(x)$, $f_2(x)$, … Then

$$f(x) = \Sigma\, n_i f_i(x)/\Sigma\, n_i$$

will be the density of the totality of the observations. Cournot had not specified the type of the component functions but in §132 he referred to his formula and indicated that the numbers n_i could stand for observations made by means of different instruments or under differing conditions. It is possible that Cournot thus had not excluded an unification of various types of functions.

Here also he was mostly interested not in the function $f(x)$ in itself, but in its characteristics (cf. above) and his innovation was in agreement with his attempts to treat heterogeneous observations (cf. my §6.2.2).

2.4. Expectation. Cournot paid little attention to expectation [10] partly because he did not yet know the Bienaymé – Chebyshev inequality. He introduced expectation in §50 explaining it in commercial terms and only provided a numerical example in §61. Further on, when considering continuous random variables and not daring to make use of integrals, he only heuristically introduced them (§67).

2.5. The Law of Large Numbers. I noted, in my §1, that in 1843 Cournot had not used that term anymore. He (§86) indicated, however, that the "Jakob Bernoulli principle" allowed to consider posterior probability instead of the prior and in §115 he called it the *seule base solide* for all the applications of the theory of probability (see my §6.2). Many later statisticians missed his pronouncements and declared that the lack of equally possible cases in practice meant that it was impossible to apply the theory (Sheynin 2003b, §3.5.3).

Cournot (§38) vaguely stated that the application of the theory of probability to the real world ought to be justified mostly by *critique philosophique* rather than by *raisonnements mathématiques.* In this connection I note that Chuprov (1905, p. 60 of Russian translation) attributed to Cournot (1843, Chapter 4) a "real substantiation" of the LLN. That chapter (which indeed begins with §38) only describes in detail the notions of randomness (see my §3) and physical impossibility. Later on Chuprov (1909, pp. 166 – 168), without providing an exact reference, stated that Cournot's proof of that law was canonical and that the LLN did not represent "either a mathematical theorem or an independent logical principle of an equal worth with the law of causality". Chuprov first formulated his thoughts about the logical justification of that law in 1896 when Bortkiewicz regarded them "sceptically" (Sheynin 1990, pp. 95 – 97), and then in 1914. Suchlike reasoning [11] proved unsuccessful: it was the strong LLN that better described the connection between theory and reality [12].

3. Randomness

Cournot (§40) defined randomness [13] as an intersection of independent chains of events. He thus repeated the interpretation introduced by ancient scholars (Cournot 1843, commentaries by B. Bru on p. 306). In §45 Cournot mentioned a *théorie mathématique du hazard*, calling it the application *plus vaste de la science des nombres* [of the widest application of quantitative knowledge?] whose aim was to study *les actes des êtres vivants.* Bru (Ibidem) noted that Cournot had not repeated the Laplacean idea (actually due to ancient scholars) of randomness being caused by ignorance of pertinent causes and remarked that Cournot had attempted to construct, in the logical and philosophical sense, the calculus of probability on the basis of the abovementioned theory of randomness. In any case, Cournot (§43) connected randomness with the *impossibilité physique*, and one of his examples (the impossibility of standing a right circular cone on its vertex) actually meant that randomness appeared with unstable equilibrium. Several pronouncements made by later scholars including Maxwell obliquely contained the same connection between the two notions and Poincaré directly defined randomness as a phenomenon occasioned by small causes leading to considerable consequences.

Cournot's further deliberations on physical impossibility had not essentially added anything to the ideas put forward by Descartes, Leibniz, Jakob Bernoulli and Huygens on moral certainty (Sheynin 1977, pp. 204 – 205 and 250 – 252) whereas Dalembert (Todhunter 1865, §473) directly introduced physical impossibility. I resolutely deny the opinion of some commentators (Martin 1994) who attribute something new here to Cournot.

Cournot returned to the interpretation of randomness in three later contributions and considered there mathematical objects. He (1851, §33, Note 38) recalled Lambert's forgotten attempt to formalize the "randomness" of the decimal expansions of irrationalities which was based on his intuitive notion of normal numbers, as they are now called. In accord with his previous ideas about intersection of chains of events (see above), Cournot also concluded that there existed no essential difference between randomness and independence of causes.

He (1861, §61, pp. 65 – 66) then repeated his and Lambert's considerations regarding randomness (but not his own conclusion), and, finally, he (1879, pp. 177 – 179) applied Bienaymé's sign test of 1874 – 1875 (Heyde et al 1977, pp. 124 – 125; Sheynin 1995, pp. 84 – 85) to checking the randomness of the initial 36 digits of the decimal expansion of π (without pronouncing any definite opinion about that number). In his time, it was impossible to achieve anything else.

4. Addendum to §3: Natural Selection

Weinstein et al (1970) attributed to Cournot (1851, p. 287/169) the principle of natural selection. The passage they referred to had consisted of four sentences (one of them of 10, and another one, of 12 lines) in which Cournot had recognized hereditary variations. He (Ibidem, p. 119/74) also assumed that changes in external conditions led to the extinction of less adapted individuals, but he (1851, pp. 126 – 127/78) thought that random causes have played a secondary part in that process with predestination having been the main agent. Later on Cournot (1861, §232, p. 362/217; 1872, pp. 155 – 159/385 – 390) expressed similar ideas, and, in the second source, he apparently came to recognize Darwinism in some restricted form. Finally, Cournot (1875, p. 92ff and p. 103) repeated much of his previous considerations as well as his partial recognition of the Darwin hypothesis.

The application of stochastic ideas and methods in biology had begun in England at the close of the 19th century by combined efforts of biologists and mathematicians. Nothing of the sort happened on the Continent until much later; to some extent this may be explained by Cournot's passive viewpoint, but much more by Quetelet's religious attitude. The Belgian scholar, who died in 1874, had not mentioned Darwin even once; moreover, he (1846, p. 259) declared that the plants and the animals had remained as created.

5. The Bayes Approach

Cournot (§§86 – 95) paid much attention to the rule *attribuée à Bayes* (§88) and to related material. This rule, or theorem, as Cournot called it in the sequel, that was nevertheless lacking in the Bayes' memoir (1764 – 1765), stated that (in modern notation)

$$P(A/B\ ) = P(B/A_i)\sum_{j=1}^{n}\ P(B/A_J)P(A_j). \qquad (1)$$

The reprint of the first part of the Bayes memoir lacks any indication about its second part. Nevertheless, the latter indirectly contained a limit theorem. Bayes, as it follows from his note published in the same source and at the same time as the first part of his memoir, had not trusted the contemporaneous practice of making use of the sums of the first terms of divergent series, and apparently for this reason he did not complete his investigation. It was Timerding, the Editor of the German translation of his memoir, who proved that (Sheynin 1971)

$$P\{-z \le (\bar{p} - a)/[\sqrt{pq}/n\sqrt{n}] \le z\} \sim \sqrt{2/\pi} \int_0^z \exp(-x^2/2)dx. \qquad (2)$$

Here p and q are the numbers of the occurrence and non-occurrence of the studied event in n Bernoulli trials ($p + q = n$), $\bar{p}$ is the unknown probability of success in a single trial, $a = p/n = \mathrm{E}\,\bar{p}$ and $pq/n^3 = \mathrm{var}\,\bar{p}$. It is remarkable that Bayes was not satisfied with the finding of De Moivre

$$P[-z \le (\mu/n - p) \le z] \sim \sqrt{2/\pi} \int_0^z \exp(-x^2/2)dx$$

where μ was the number of the occurrences of the event in n Bernoulli trials, p and q, the known probabilities of success and failure in a single trial ($p + q = 1$) and $pqn = \mathrm{var}\,\mu/n$. A modern description of the Bayes memoir is in Hald (1998, Chapter 8).

It was the "Bayes approach", first and foremost, that Cournot desired to prove and correct (cf. my §1). He (§89) justly indicated that unsubstantiated (subjective) assumptions about prior probabilities of the events A_i, $i = 1, 2, \ldots, n$ in formula (1) make objective estimates of the posterior probabilities impossible.

Cournot (§95) remarked, however, that, given a large number of observations, probability $\bar{p}$ in formula (2) [14] may be assumed equal to the posterior probability p/n. He attempted to demonstrate this proposition, but hardly successfully. Nevertheless, it is implied by the Bernoulli LLN and Laplace proved it and indicated a more effective estimate of the rapidity of the appropriate convergence (Hald 1998, pp. 169 – 170).

6. Applications of the Theory of Probability

6.1. Astronomy, Mathematical Treatment of Observations, Insurance. Cournot (§145) thought that astronomy was the branch of natural sciences in which the successful application of probability theory might be expected first of all, and he even proposed the term *statistique des astres.* Indeed, the measurement of the proper motions of stars began in 1837(though at first only perpendicular to the line of vision), and, in the second half of the century, the regularities inherent in starry clusters came under examination [15].

Cournot himself, however, only studied statistically the parameters of planetary and cometary orbits. In the former case he (§148) assumed a natural supposition about the probability of a distance between the poles of two randomly situated great circles. Laplace, and later Newcomb issued, however, from other hypotheses; Bertrand (1888, pp. 6 – 7) repeated Laplace's and Cournot's reasoning without mentioning anyone and justly noted that the distinction between the obtained solutions as well as his celebrated paradox concerning the length of a "random" chord of a given circle showed that the expression "randomly" was not sufficiently definite (Sheynin 2003a).

Somewhat similar happened with the investigation of the cometary orbits. Poisson had stated that Cournot should have distinguished between two assumptions: either any point of the celestial sphere was, with one and the same probability, the pole of a cometary orbit; or, all the inclinations of the orbit were equally probable. Cournot (§149) did not agree, also see Bru's commentary on that section.

Cournot described the treatment of observations in more detail than necessary, but not on a sufficiently professional level, cf. my §1. In particular, he (§144) adduced a summary of the determinations of the length of a seconds pendulum [16], but his inference about the actual precision of that magnitude, although perhaps correct, was strangely enough based on the results of its first derivation (by Picard in 1671) [17]. It also remains unclear why Cournot had not discussed meridian arc measurements.

When describing insurance, Cournot hardly applied any new stochastic ideas or methods, but he had not failed to note that accidents were not independent from each other, cf. my §6.2.3.

6.2. Statistics. Applications to statistics are the most interesting in his book. Statistics, he (§103) declared, had blossomed exuberantly and [the society] should be on guard against its applications *prématurées et abusives* which can discredit it for some while and delay the time when it will underpin all the theories concerning the *organisation sociale* [18]. Indeed, already Lueder (1812, p. 9) complained about the appearance of "legions" of statistical data.

Statistics, as Cournot (§105) continued, ought to have its theory, rules and principles, it should be applied to physical, natural-scientific, social and political phenomena. He then formulated a few apparently obvious, but highly questionable at the time statements. The main aim of statistics was (§106) to *pénétrer autant que possible dans la connaissance de la chose en soi*, and (§120) to study the causes of phenomena. The theory of probability was applicable to statistics although the *élément variable* escaped our attention [?] and could only be described by philosophical probabilities (§113; see my §2.2). The *principe de Bernoulli* is the only foundation for applying probability theory and it satisfies the statisticians (§115).

The *Staatswissenschaft* (Statecraft), a flabby discipline studying political and economic features of separate states, was not interested in the causes of social phenomena and considered numerical data insufficient for its purposes. The opposing (sometimes tacit) views of Daniel Bernoulli, Laplace, Poisson (and Cournot, as seen above), coinciding with the opinion of the founders of political arithmetic, Graunt and Petty, only became prevailing by the end of the 19th century. Among the statisticians, the originator of the stochastic direction was Quetelet, who introduced mean (for a given society) inclinations to crime and to marriage, *i.e.*, the pertinent probabilities. He had not referred to the Poisson LLN (which would have made his conclusions more general) and did not stress that the calculated inclinations must not be applied to individuals. No wonder that after his death many statisticians abandoned Queteletism,– and probability theory in general. The LLN did not become an exception, it was denied on absurd grounds. Knapp (1872, pp. 116 – 117), for example, even stated that statisticians always had to do with a single observation, as when counting the population of a city [19]. I dwell now on Cournot's concrete ideas and recommendations.

1. The significance of empirical discrepancies. Laplace repeatedly studied the significance of small discrepancies between observational series (and between theory and observation). Poisson (Sheynin 1978, §5.2) formally explicated this issue and, like Laplace, restricted his attention to the case of a large number of observations. Cournot (§§107 – 117 and 128 – 129) followed Poisson; just as his predecessor (1824, §14; 1829, §§1 and 9), he indicated, for instance in §109, see my §6.2.2, that the probability of an event could be variable. Cournot (§§111 and 113) also remarked that apparently considerable discrepancies could contradict common sense (and be fictitious). However, his refusal to use mathematical formulas (see my §1) made his exposition cumbersome.

2. Series of Bernoulli trials (§§76 – 77). Suppose that the series consist of n_1, n_2, … trials with respective probabilities of success p_1, p_2, … Then, for the totality of the trials, the probability of success will be

$$p = \Sigma\, n_i\, p_i/\Sigma\, n_i.$$

Denote $n_i/n = k_i$, $1 - p_i = q_i$, $1 - p = q$, then

$pq = \Sigma k_i p_i(1 - \Sigma k_i p_i)$.

In quite an elementary way [20] Cournot proved that

$$pq > \Sigma k_i p_i q_i > (\Sigma k_i \sqrt{p_i q_i})^2. \quad (3)$$

Commentators (Stigler 1986, p. 197) indirectly indicated that already then formula (3) could have been applied for justifying stratified sampling.

3. Solidary (systematic) influences and judicial statistics. Cournot (§§104 and 117) isolated solidary causes acting in a similar way on a number of trials [21]. Random causes, as he indicated, only corrupted or even determined the results of separate events. This new characteristic of randomness did not contradict his main definition of that notion (see my §3).

Most interesting was Cournot's attempt (Chapter 15), also see the appropriate part of his paper (1838), to reveal the general causes influencing all the judges (jurors) during a hearing of a given case [22]. Suppose (§193) that for two judges the probabilities of a correct judgement are s_1 and s_2. Then the probability of the coincidence of their decisions, whether just or otherwise, will be

$$p = s_1 \cdot s_2 + (1 - s_1) \cdot (1 - s_2) \quad (4)$$

If $s_1 = s_2 = s > 1/2$, then (§206) $s = 1/2 + (1/2)\sqrt{2p - 1}$

where p can be determined statistically. If (§195), in addition, it were possible to derive, in a similar way, all the magnitudes included in equations of the type of (4), then a considerable discrepancy between theory and statistics will testify to an interdependence between the judges.

Then, again in §206, Cournot recommended to distribute the legal cases among several heads and form the same equations for each of those. As the number of the categories increased, then, as he believed, the sources of the miscarriage of justice for each of them will become independent [23].

I doubt that Cournot's recommendations were practicable but in any case he desired to treat statistical data corrupted by systematic causes, and Bertrand's statement (1888, pp. 325 – 326) that his predecessor had not noticed the dependence existing between the individual judgements was absolutely wrong.

Three more points. After Quetelet (and other scholars mentioned in this subsection) it was Cournot who introduced probability into statistics. Then, he formulated reasonable ideas about the administration of justice and judicial statistics, and, finally (in §225), he refused to apply probability for estimating the trustworthiness of a chain of reports. It is not amiss to recall that Markov (Sheynin 1989, p. 340n) declared that the study of testimonies was the "weakest section" of the theory of probability.

4. In my §6.2.1 I noted Cournot's qualified remark about possible contradictions between a formal treatment of statistical data and common sense. Issuing from such reasonable considerations and in obvious contradiction with what Cournot (§113, see my §6.2) had concluded, Stigler (1986, pp. 197 – 198) alleged that he had denied the possibility of applying probability to statistics.

I shall nevertheless adduce two of Cournot's critical remarks so as to stress his proper attitude to statistics. In §102 he indicated that statistical data could be separated in groups in many ways and that an unfortunate choice of the groups led to mistaken inferences. Then, he (1851, p. 245) remarked that most often statistical numbers *ne mesurer que des effets très complexes et très éloignés de ceux qu'il faudrait saisir pour avoir la théorie rationnelle des* *phénomènes* but he explained the situation by the *développement prodigieux, parfois maladroit ou prématuré, de ce que l'on nomme la statistique*, cf. his earlier similar statement in my §6.2.

7. Conclusions

In his innovations and in the suggested widening of the possibilities of statistics, Cournot (see my §§2.2, 2.3 and 6.2) mostly followed Poisson. This did not depreciate the methodological importance of his work because the recommendations made by his predecessor had remained little known. I especially note Cournot's efforts to treat dependent statistical observations and trials with variable probabilities of the studied event; his pertinent contribution could have considerably fostered the originating theoretical statistics.

Bortkievicz (1894 – 1896) several times approvingly mentioned Cournot. Still oftener Chuprov referred to him in his copy of that writing; it is that copy which the translator of Bortkiewicz had at his disposal. Edgeworth, the predecessor of Karl Pearson, was another admirer of Cournot. He (1885, p. 571) characterized the latter's discussion of posterior probability as *masterly*. Elsewhere Edgeworth (1926, p. 446) concluded that Cournot, "in his masterly work on chances, … [had] pointed out the bearing of the calculus of probabilities on statistics".

Nevertheless, here is the opinion of Anderson (1934, p. 539), a student of Chuprov, about the contemporaneous situation in statistics:

Germanic countries, despite the work of Lexis and his eminent pupil von Bortkiewicz, still continue under the influence of von Mayr's empirical school. … Only in very recent years has a faint start been made toward the acceptance of the English theories, and, in economic statistics, of American methods.

Recall also (my §1) that Cournot chose an unfortunate mathematical level for his book and that his style was flabby. Nevertheless, and the more so since that contribution was translated into German, it can only be regretted that neither Kries (1886), nor Czuber (1899; 1923), nor Kaufman (1928) mentioned him at any length. Even Lexis (1879), who studied the stability of statistical series by issuing from his distribution-free criterion, had not in essence noticed Cournot. Lexis attempted to determine whether the probability of a studied event varied during several series of observations, or whether they were dependent one on another. Already Cournot formulated these aims and at least Lexis could have approached his problems much earlier than he actually did [24]. On the other hand, Chuprov highly (perhaps too highly) praised Cournot. First, bearing in mind mathematics as well as philosophy and economics, he (1905, p. 60) called Cournot a man of genius. Later on, he (1909, p. 30) stated that Cournot was

One of the most original and profound thinkers of the 19th century, whom his contemporaries had failed to appreciate and who rates ever higher in the eyes of posterity.

Lastly, Chuprov (1926, p. 77) characterized the French scientist as "the real founder of the modern philosophy of statistics", cf. my §2.5, and this seems to be correct.

Notes

1. From 1973 onwards, no less than 13 volumes of his *Oeuvres complètes* have appeared. Except for Cournot (1843), which I specify by sections, I provide the page numbers of his contributions both in accord with their initial editions, and with the *OC*.

2. Its Russian translation is corrupted by mistakes, misprints and wrong terminology. Nevertheless, Russian readers might be grateful to the translator, N.S. Chetverikov, a student of Chuprov, who acquainted himself with Cournot's book on his teacher's recommendation

(Sheynin 1990, p. 56). Chetverikov was also coauthor of Weinstein et al (1970) where the French scholar was wrongly credited with the principle of natural selection (see my §4).

3. Cournot repeated Laplace's methodological mistake that his predecessor made in compiling his *Essai philosophique.* In §30 he formulated the LLN in words only; in §69, for instance, instead of providing the appropriate integral, he was compelled to say that a certain probability was tabulated (at the end of his book).

4. Bru, the Editor of its reprint, meticulously inserted them.

5. Bortkiewicz (Sheynin 1990, p. 42) thought that Poisson was not understood properly; his statement is not yet examined.

6. Bayes (1764, p. 137), for instance, applied, as he himself stated, both terms, chance and probability, on a par.

7. Khinchin (1961, No. 1, p. 95) enthusiastically mentioned that example (throws of an irregular die) but attributed it to Mises. Apparently exaggerating, the latter (1919, p. 57; 1931, Introduction) called Cournot his predecessor. Actually, however, Cournot (§240.5) only hinted that the limit of the statistical probability might be taken as the theoretical probability, and in a few instances (see above) stressed, as Jakob Bernoulli did, the importance of posterior probability.

8. He applied the uniform distribution for calculating the mean life of the last survivor from among a group of men (Todhunter 1865, §340), *i.e.*, the mean value of the appropriate order statistic. De Moivre derived the normal distribution in 1733, but this fact was forgotten, cf. my §1.

9. Note that already Gauss (1809, §175) introduced the density of distribution in the theory of errors. Actually following Poisson (1829, §1), modern mathematicians define it as the derivative of the integral distribution function. The term *random variable* had not appeared until the end of the 19th century and possibly even somewhat later (Sheynin 1990, pp. 122 – 123).

10. He did not at all recognize moral expectation (see his §§51 and 61). During its hey-day, that term became so widespread that Laplace (1812, p. 189) specified the classical term *expectation* by calling it *mathematical.*

11. In attempting to justify the LLN, Ellis (1854, p. 49) remarked that "On a long run of similar trials, every possible event tends ultimately to recur in a definite ratio of frequency".

12. In his Préface Cournot reasoned on the *état pénultième* characterizing the behavior of random phenomena (hasards). Stretching somewhat this idea, it might be interpreted as a subordination of the variable to the weak (rather than strong) LLN, but it seems better to disregard that passage altogether.

13. On the history of the concept of randomness see Sheynin (1991).

14. Cournot had not inserted that formula, I am only making use of the appropriate notation.

15. In this connection, as also concerning meteorology, medical statistics and public hygiene (neither of which Cournot had in essence mentioned), see my papers Sheynin (1982, 1984a; 1984b). Already in 1817, Humboldt had introduced isotherms and thus founded climatology, a discipline studying the mean values of meteorological elements, after which the study of the deviations from those means has begun. In medical statistics, a main issue was variolation of smallpox, a dangerous procedure preceding vaccination.

16. Cournot reduced all these measurements to the latitude of Paris.

17. Cournot (§130) reasonably stated that the precision of the final result did not increase unboundedly with the number of observations, but he (§§117, 138, 139, 162) only indirectly and not quite definitely explained the situation. Then, anticipating a resolution of the International Statistical Congress passed in 1869 (Sheynin 1986, p. 291, Note 18), he (§§107 and 127) recommended to include in the statistical reports not only the results obtained, but also their weights.

18. Cf. Cauchy (1845, p. 242): statistics provides a means for judging doctrines and institutions.

19. The standard argument was that equally possible cases presumed by the definition of probability did not exist, and that probabilities were not constant, see my §2.5. The gap between the theory of probability and the other branches of mathematics, soon beginning to develop rapidly, was not essential for statistics.

20. In a note attached to §82 Cournot estimated, in a similar manner, the precision of the function $f(x)$, see my §2.3.

21. Strangely enough, he had not mentioned systematic errors already isolated (although in a narrow sense) by Daniel Bernoulli in 1780.

22. Keeping to the established tradition (Condorcet, Laplace, Poisson), Cournot both in that memoir and in the previous chapter of his book first considered the case of independent judgements. Note that Laplace (1812, p. 523) only once, and in passing, indicated that he assumed such independence. Other scholars undoubtedly understood that their deliberations were restrictive and apparently wished to describe the administration of justice at least under ideal conditions.

23. Bru noted that the last-mentioned statement was arbitrary.

24. Chuprov (1918 – 1919) proved that the Lexis criterion did not meet the issue; nevertheless, his works belonged to a new stage in the development of statistics (Anderson 1934; Bauer 1955).

References

Abbreviation: AHES = *Arch. Hist. Ex. Sci.*

Anderson, O. (1934), Statistics: Statistical method. *Ausgew. Schriften*, Bd. 2. Tübingen, 1963, pp. 539 – 544. Editor, H. Strecker.

Bauer, R.K. (1955), Die Lexische Dispersionstheorie in ihrer Beziehungen zur modernen statistischen Methodenlehre. *Mitt.-Blatt math. Stat., Anwendungsgeb.*, Bd. 7, pp. 25 – 45.

Bayes, T. (1764 – 1765), Essay towards solving a problem in the doctrine of chances. Reprint of pt 1: *Biometrika*, vol. 45, 1958, pp. 293 – 315 and Pearson et al (1970, pp. 131 – 153). References in text to last-mentioned source.

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

Bessel, F.W. (1838), Untersuchungen über die Wahrscheinlichkeit der Beobachtungsfehler. *Abh.*, Bd. 3. Leipzig, 1876, pp. 372 – 391.

Boole, G. (1851), On the theory of probabilities. *Studies in Logic and Probability*, vol. 1. London, 1952, pp. 247 – 259.

Bortkiewicz, L. (1894 – 1896), Kritische Betrachtungen zur theoretischen Statistik. *Jahrbücher f. Nat.-Ökon. u. Statistik*, 3. Folge, Bd. 8, pp. 641 – 680; Bd. 10, pp. 321 – 360; Bd. 11, pp. 701 – 705. Russian translation in Chetverikov (1968, pp. 55 – 137).

Cauchy, A.L. (1845), Sur les secours que les sciences du calcul pouvent fournir aux sciences physiques ou même aux sciences morales. *Oeuvr. Compl.*, sér. 1, t. 9. Paris, 1896, pp. 240 – 252.

Chebyshev, P.L. (1845), Essay on an elementary analysis of the theory of probability. *Полное собрание сочинений* (Complete Works), vol. 5. M – L., 1951, pp. 26 – 87. (R)

Chetverikov, N.S., Editor (1968) *О теории дисперсии* (On the Theory of Dispersion). M. Collection of articles.

Chuprov, A.A. (1905), Die Aufgabe der Theorie der Statistik. *Schmollers Jahrb. f. Gesetzgebung, Verwaltung u. Volkswirtschaft im Deutsch. Reich*, Bd. 29, No. 2, pp. 421 – 480. References in text to Russian translation (Chuprov 1960, pp. 43 – 90).

--- (1909). *Очерки по теории статистики* (Essays on the Theory of Statistics). M., 1959.

--- (1918 – 1919), Zum Theorie der Stabilität statistischer Reihen. *Skand. Aktuarietidskr.*, t. 1, pp. 199 – 256; t. 2, pp. 80 – 133.
--- (1926), Theory of stability of statistical series. In Swedish. Reference in text to English translation in author's *Statistical Papers and Memorial Publications*. Berlin, 2004, pp. 74 – 90.
--- (1960), *Вопросы статистики* (Issues in Statistics). Coll. Articles. M.
Cournot, A.A. (1838), Sur les applications du calcul des chances à la statistique judiciaire. *J. math. pures et appl.*, t. 3, pp. 257 – 334.
--- (1841), *Traité élémentaire de la théorie des fonctions et du calcul infinitésimal.* OC, *t.* 6.1. Paris, 1984.
--- (1843). *Exposition de la théorie des chances et des probabilités.* OC, t. 1. Paris, 1984. Editor, B. Bru. German translation: Braunschweig, 1849. Russian translation: M., 1970.
--- (1851), *Essai sur les fondements de nos connaissances et sur les caratères de la critique philosophique.* OC, t. 2. Paris, 1975.
--- (1861), *Traité de l'enchaînement des idées fondamentales dans les sciences et dans l'histoire.* OC, t. 3. Paris, 1982.
--- (1872), *Considérations sur la marche des idées et des événements dans les temps modernes.* OC, t. 4. Paris, 1973.
--- (1875), *Matérialisme, vitalisme, rationalisme.* OC, t. 5. Paris, 1979.
--- (1978), *Etudes pour le centenaire de sa mort.* Paris.
Czuber, E. (1899), Die Entwicklung der Wahrscheinlichkeitstheorie und ihrer Anwendungen. *Jahresber. Deutsche Mathematiker-Vereinigung*, Bd. 7, No. 2. Second paging.
--- (1923), *Philosophische Grundlagen der Wahrscheinlichkeitsrechnung.* Leipzig – Berlin.
Dugac, P. (1978), Cournot et le calcul infinitésimal. In Cournot (1978, pp. 65 – 73).
Edgeworth, F.Y. (1885), Calculus of probability applied to physical research, pt. 1. *Writings in Probability, Statistics and Economics*, vol. 3. Cheltenham (UK) – Brookfield (US), 1996, pp. 569 – 578.
--- (1926), Cournot. In *Palgrave's Dictionary of Political Economy*, vol. 1. London, pp. 445 – 446. Possibly included in earlier editions of the *Dictionary.*
Ellis, R.L. (1854), Remarks on the fundamental principle of the theory of probability. *Math. And Other Writings*. Cambridge, 1863, pp. 49 – 52.
Feller, J. (1961), Cournot. *Dictionnaire de biographie Francaise*, t. 9. Paris, p. 983.
Gauss, C.F. (1809), Aus der Theorie der Bewegung der Himmelkörper. In Latin. German translation in author's *Abh. zur Methode der kleinsten Quadrate* (1887). Vaduz, 1998, pp. 92 – 117.
Hald, A. (1998), *History of Mathematical Statistics from 1750 to 1930.* New York.
Heyde, C.C., Seneta, E. (1977), *Bienaymé.* New York.
Kaufman, A.A. (1928), *Теория и методы статистики* (Theory and Methods of Statistics). M. – L. German translation of a previous edition: Tübingen, 1913.
Kendall, M.G., Plackett, R.L., Editors (1977), *Studies in History of Statistics and Probability*, vol. 2. London.
Khinchin, A.Ya. (1961), The Mises frequentist theory and modern ideas of the theory of probability. *Voprosy Filosofii*, No. 1, pp. 92 – 102; No. 2, pp. 77 – 89. English translation to appear in 2004 in *Science in Context.*
Knapp, G.F. (1872), Quetelet als Theoretiker. *Jahrbücher f. Nat.-Ökon. u. Statistik*, Bd. 18, pp. 89 – 124.
Kries, J. (1886), *Principien der Wahrscheinlichkeitsrechnung.* Tübingen, 1927.
Laplace, P.S. (1812), *Théorie analytique des probabilités.* OC, t. 7. Paris, 1886.
Lexis, W. (1879), Über die Theorie der Stabilität statistischer Reihen. *Abh. zur Theorie der Bevölkerungs- und Moralstatistik.* Jena, 1903, pp. 170 – 212.
Lueder, A.F. (1812), *Kritik der Statistik und Politik.* Göttingen.

Martin, T. (1994), La valeur objective du calcul des probabilités selon Cournot. *Math. Inf. Sci. Humain.*, t. 32, No. 127, pp. 5 – 17.
Mises, R. (1919), Grundlagen der Wahrscheinlichkeitsrechnung. *Sel. Papers*, vol. 2. Providence, Rhode Island, 1964, pp. 57 – 105.
--- (1931), *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik.* Leipzig – Wien.
Ondar, Kh.O. (1971), On Davidov's work on the theory of probability. Translated in this book.
Pearson, E.S., Kendall, M.G., Editors (1970), *Studies in History of Statistics and Probability*, [vol. 1]. London.
Poisson, S.D. (1824 for 1827, 1829 for 1832), Sur la probabilité des résultats moyens des observations. *Conn. des Temps*, second paging, pp. 273 – 302 and 3 – 22.
--- (1837), *Recherches sur la probabilité des jugements.* Paris.
Quetelet, A. (1846), *Lettres sur la théorie des probabilités.* Bruxelles.
Sheynin, O. (1971), On the history of some statistical laws of distribution. *Biometrika*, vol. 58, pp. 234 – 236. Reprint: Kendall & Plackett (1977, pp. 328 – 330).
--- (1974), Prehistory of the theory of probability. AHES, vol. 12, pp. 97 – 141.
--- (1976), Laplace's work on probability. AHES, vol. 16, pp. 137 – 187.
--- (1977), Early history of the theory of probability. AHES, vol. 17, pp. 201 – 259.
--- (1978), Poisson's work on probability. AHES, vol. 18, pp. 245 – 300.
--- (1982), On the history of medical statistics. AHES, vol. 26, pp. 241 - 286.
--- (1984a), On the history of the statistical method in astronomy. AHES, vol. 29, pp. 151 – 199.
--- (1984b), On the history of the statistical method in meteorology. AHES, vol. 31, pp. 53 – 93.
--- (1986), Quetelet as statistician. AHES, vol. 36, pp. 281 – 325.
--- (1989), Markov's work on probability. AHES, vol. 39, pp. 337 – 377.
--- (1990, in Russian), *Chuprov.* Göttingen, 1996.
--- (1991), On the notion of randomness from Aristotle to Poincaré. *Math., Inf. Sci. Humain.*, t. 29, No. 114, pp. 41 – 55.
--- (1995), Helmert's work in the theory of errors. AHES, vol. 49, pp. 73 – 104.
--- (2003a), Geometric probability and the Bertrand paradox. *Hist. scientiarum*, vol. 13, pp. 42 – 53.
--- (2003b), Social statistics: its history and some modern issues. *Jahrbücher f. Nat.-Ökon. u. Statistik*, Bd. 223, pp. 91 – 112.
Stigler, S.M. (1986), *History of Statistics.* Cambridge (Mass.) – London.
Todhunter, I. (1865), *History of Mathematical Theory of Probability.* New York, 1965.
Weinstein, A.L., Chetverikov. N.S. (1970), Cournot and his contribution to the theory of probability. In Russian translaton of Cournot (1843), pp. 5 – 24.

20. B.V. Gnedenko. A Review of the Current State of the Theory of Limiting Laws for Sums of Independent Terms

Uchenye Zapiski Tomsk. Gosudarstven. Pedagogich. Inst.,
No. 1, 1939, pp. 5 – 28.

Foreword by the Translator

This is an essay on an important issue initially published in a rare source. Regrettably, its first sections that describe the events in the 19th century are corrupted by mistakes, see my Note 2, and the type-setter accomplished his work extremely bad. In many cases, I corrected the formulas but I am not sure that everything is now correct. And, for example, the first term on the right side of formula (10) is apparently wrong.

* * *

1. I review the current state of one of the main chapters of the theory of probability, – of the theory of the limiting laws for sums of independent terms. During the last years a number of fundamental investigations enriched this chapter. Not only had they given an answer to the problems raised in the 19th century; they also widely generalized and extended their very formulation. The research done enables us to say that the main, the principal pertinent issues are already solved and that a general method, which makes it possible to solve isolated problems previously treated by most various tricks in one general way and, in addition, to solve a number of yet unyielding problems, was created in this theory.

Soviet scientists occupy leading positions in developing the theory of limiting laws as also in furthering other chapters of probability theory. Among those scholars who attained the most considerable success I ought to name, above all, Kolmogorov, Khinchin and Bernstein. The students of Khinchin and Kolmogorov, G.M. Bavli, A.A. Bobrov, Gnedenko and D.A. Raikov, made first-rate discoveries and P. Lévy,and his student Doeblin in France, and Feller in Sweden, obtained results of no lesser importance, but the basic role belongs to Soviet mathematicians and this fact will be revealed all by itself in the exposition below.

To make this review intelligible not only for specialists in probability, I provide, in a special section, brief definitions of such main notions as law of distribution and characteristic function. And, in pursuing the same goal, I offer a concise description of the process of integration in the Stieltjes sense. […]

4. […] If $\alpha(x)$ is a non-decreasing on $(a; b)$ function of bounded variation, then the *Stieltjes integral* of a continuous function $\beta(x)$, defined on the same interval, with respect to $\alpha(x)$ is the limit of the sum, as $n \to \infty$,

$$\sum_{k=0}^{n} \beta(x_k')[\alpha(x_{k+1}) - \alpha(x_k)]$$

where $x_k \le x_k' \le x_{k+1}$, $x_0 = a$, $x_{n+1} = b$ and, as $n \to \infty$,

$$\max (x_{k+1} - x_k) \to 0,\ 0 \le k \le n.$$

This limit is denoted as $\int_a^b \beta(x)\, d\alpha(x)$. […]

5. Chebyshev […] essentially furthered the study of limiting relations for sums. He had proved a theorem widely generalizing the LLN […] by a quite elementary but at the same time very powerful method of moments which later became one of the main methods of the theory of probability […]

A *characteristic function* of the law $F(x)$ is the function

$$f(t) = \int_{-\infty}^{\infty} e^{ixt} dF(x) = \mathrm{E}e^{ixt}.$$

Obviously, $|f(t)| \le 1$. We will denote such functions by

$$f(t) = \rho(t)\, e^{i\omega(t)}$$

and assume that $|\omega(t)| \le \pi$. […]

6. At the very end of the 19th century, Liapunov demonstrated a remarkable theorem on the convergence of the laws of distribution of sums of independent terms to the Gauss law.

The Liapunov Theorem. If independent random terms of the sequence

$$x_1, x_2, \ldots, x_n, \ldots \qquad (1)$$

have finite expectations $Ex_k = a_k$, variances $b_k = E(x_k - Ex_k)^2$ and moments $d_k = E|x_k - a_k|^{2+\delta}$ where δ is a positive constant, and if

$$\left(\sum_{k=1}^{n} d_k\right)^2 = \mathrm{o}\left(\sum_{k=1}^{n} (b_k)^{2+\delta}\right),$$

then, as $n \to \infty$,

$$P\left(\frac{\sum_{k=1}^{n}(x_k - a_k)}{[\sum_{k=1}^{n} b_k]^{1/2}} < x\right) \to (1/\sqrt{2\pi}) \int_{-\infty}^{x} (-z^2)dz. \qquad (2)$$

To prove his theorem, Liapunov had to apply a very involved analytical tool, and, in essence, to create the method of characteristic functions which became, at the hands of modern workers, one of the main and most powerful methods of the theory of probability. Soon, however, in order to rehabilitate the Chebyshev method of moments, Markov found a new substantiation of the Liapunov theorem which made use of this method. His approach proved to be quite sufficient for that purpose, but the method of moments is considerably weaker that the method of characteristic functions which is also applicable when the random variables do not have finite moments.

Chebyshev, Liapunov and Markov left their contemporaries behind; for more than two decades their work found no followers either in Russia or abroad. Only in 1922 Lindeberg provided a new sufficient condition for the sums of independent terms normed by the squared deviations and centered by the expectations to tend to the Gauss law.

The Lindeberg Theorem. If, for any positive constant τ,

$$\sum_{k=1}^{n} (1/B_n^2) \int_{|x-a_k|>\tau B_n} (x - a_k)^2 \, dF_k(x) \to 0 \text{ as } n \to \infty$$

where $F_k(x)$ is the law of distribution of the random variable x_k and

$$B_n^2 = (b_1^2 + b_2^2 + \ldots + b_n^2)^{1/2},$$

then (2) holds and

$$(1/B_n^2) \max b_k \to 0,\ 1 \leq k \leq n,\ n \to \infty.$$

This theorem was an essential step forward because Lindeberg had not assumed that the moments higher than the second one existed. In 1935 Feller proved that the Lindeberg condition was not only sufficient but also necessary for the convergence to the Gauss law. In 1933 Kolmogorov obtained the Lindeberg result anew by the method of differential equations proposed by I.G. Petrovsky. For a number of problems in the theory of probability this method proved especially convenient.

7. In 1925, Lévy, in his *Calcul des probabilités*, further developed the method of characteristic functions. He expounded it and proved the main theorems concerning the connection between the behaviour of the characteristic functions and of the corresponding laws of distribution and applied his theory both for proving the Lindeberg theorem and for constructing the theory of the so-called stable laws. Lévy formulated and solved the following problem: Which distributions $\Phi(x)$ can serve as limiting laws for the laws of distribution of the sums

$$S_n = (x_1 + x_2 + \ldots + x_n)/B_n$$

of independent terms x_k, $1 \le k \le n$, identically distributed in accord with the law $F(x)$ and normed by appropriately chosen positive constants B_n. It occurred that the class of these limiting laws coincided with that of stable laws (in Lévy's sense). He found a general formula for them: the logarithm of a characteristic function $\varphi(t)$ of a limiting law in the just indicated sense was expressed as

$$\lg \varphi(t) = i\gamma t - C_o|t|^\alpha[1 + i\beta t\omega(\alpha)/|t|] \quad (3)$$

where $C_o > 0$, $|\beta| \le 1$, $0 < \alpha \le 2$, and γ were real constants and $\omega(\alpha) = \mathrm{tg}(\pi\alpha/2)$ for $\alpha \ne 1$ and $= 0$ otherwise.

In 1936 Khinchin supplemented Lévy's research. He showed that for $\Phi(x)$ to be a limiting law for the laws of distribution of the sums

$$S_n = [(x_1 + x_2 + \ldots + x_n)/B_n] - A_n, \quad (4)$$

where A_n and $B_n > 0$ were some constants and (1) was a sequence of identically distributed independent random variables, it was necessary and sufficient that the logarithm of its characteristics function was the same as (3) but with $\omega(\alpha; t)$ instead of $\omega(\alpha)$: $\omega(\alpha; t) = \mathrm{tg}(\pi\alpha/2)$ for $\alpha \ne 1$ and $= (2/\pi)\lg |t|$ otherwise.

8. At about the same time (1922 – 1923), a systematic reconstruction of the theory of probability on the basis of the theory of functions of a real variable began in Moscow (Khinchin, E.E. Slutsky). A number of Khinchin's contributions devoted to the theory of limiting theorems for sums have then appeared. In particular, he proved the following proposition.

The Khinchin Theorem. Let mutually independent and identically distributed random variables (1) possess expectations $a = \mathrm{E}x_1 = \mathrm{E}x_2 = \ldots = \mathrm{E}x_n = \ldots$ Then, for any constant $\varepsilon > 0$,

$$P\{|\sum_{k=1}^{n} (x_k/n) - a| < \varepsilon\} \to 1 \text{ as } n \to \infty.$$

This theorem completes the investigation of the necessary and sufficient conditions for the LLN in the Chebyshev form when the terms are identically distributed. Elsewhere Khinchin proved that in the general case it was impossible to provide necessary and sufficient conditions only depending on the moments of all integral orders.

Khinchin's monograph [4] greatly influenced the taste for investigation in the field of probability. At the same time, there appeared works that gave an impulse to many later studies devoted to the so-called strong LLN (Khinchin and Cantelli) [1] and the law of iterated logarithm (Khinchin). I do not touch these important directions in the development of the classical theorems of probability theory.

9. In 1927, Kolmogorov's paper devoted to the study of the boundaries of applicability of the LLN had appeared. There, he offered necessary and sufficient conditions for applying this law not only in Chebyshev's formulation, but also in a more general setting. Kolmogorov asked himself, under what conditions was it possible to choose such real constants A_n, that, for any constant $\varepsilon > 0$ and independent random variables (1),

$$P\{|\sum_{k=1}^{n} (x_k/n) - A_n| < \varepsilon\} \to 1 \text{ as } n \to \infty.$$

In the Chebyshev theorem these constants are equal to the means of the expectations of the variables:

$$A_n = (1/n)\,(\mathrm{E}x_1 + \mathrm{E}x_2 + \ldots + \mathrm{E}x_n).$$

Later, in 1937, Feller somewhat generalized this Kolmogorov formulation of the LLN by studying the conditions to be imposed on the sequence of independent variables (1) so that, for a given sequence $\{B_n\}$ of positive constants, it would be possible to choose such real constants A_n, that, for any constant $\varepsilon > 0$,

$$P\{|\sum_{k=1}^{n} (x_k/B_n) - A_n| < \varepsilon\} \to 1 \text{ as } n \to \infty. \quad (5)$$

In addition to generalizing the formulation of the LLN (above), Kolmogorov, in the same contribution, essentially changed the very problem for sums of independent terms. Instead of the classical pattern dealing with a *sequence of random variables*, he proposed to consider a *sequence of series* of random variables independent within each such series, see §10. And his following theorem that only constituted a particular finding included in his paper, completed the investigation cycle on the LLN in Chebyshev's formulation.

The Kolmogorov Theorem. For a sequence (1) of independent random variables with expectations $\mathrm{E}x_1 = a_1$, $\mathrm{E}x_2 = a_2$, …, $\mathrm{E}x_n = a_n$, … to obey the relation

$$P\{[|\sum_{k=1}^{n} (x_k - a_k)|/n] < \varepsilon\} \to 1 \text{ as } n \to \infty$$

for any positive ε it is necessary and sufficient that, as $n \to \infty$, the conditions

$$\sum_{k=1}^{n} \int_{|x-a_k|>n} dF_k(x) \to 0, \sum_{k=1}^{n} (1/n) \int_{|x-a_k|<n} (x - a_k)\, dF_k(x) \to 0,$$

$$\sum_{k=1}^{n} \int_{|x-a_k|<n} (1/n^2)\,(x - a_k)^2 dF_k(x) \to 0$$

be satisfied.

The following proposition provides necessary and sufficient conditions for a more general formulation of the LLN.

The Kolmogorov – Feller Theorem. For the existence of a sequence of real constants A_n such that (5) is valid for any positive ε where $B_n > 0$ is a given sequence of constants and (1) is a sequence of independent variables, it is necessary and sufficient, that, for an arbitrary constant $\tau > 0$,

$$\sum_{k=1}^{n} \int_{|x|>\tau B_n} dF_k(x) \to 0; \sum_{k=1}^{n} (1/B_n^2) \int_{|x|<\tau B_n} x^2 dF_k(x) \to 0.$$

It is assumed here that the variables x_k are centered by medians:

$$F_k(-0) \le 1/2 < F_k(+0).$$

10. The classical application of the limiting laws in natural sciences consists in the following. It is supposed that a course of a given phenomenon is influenced by a very large number of independent random causes each of them little affecting the course as compared with their combined action. Being added together, the causes provide a resulting (naturally, random) cause that indeed determines the studied process. Consider now the sequence (1) of independent variables. Their action is approximately described by the sum

$$S_n = x_1 + x_2 + \ldots + x_n \quad (6)$$

normed and centered by the corresponding constants $B_n > 0$ and A_n; that is, by

$$\Delta_n = [(x_1 + x_2 + \ldots + x_n)/B_n] - A_n.$$

As the number of the terms increases, the sum Δ_n reproduces the actual phenomenon ever more precisely. We saw that such a pattern is quite satisfactory for a number of most important practical issues. […][2] However, as Kolmogorov remarked, many natural-scientific problems did not yield to it, and it was only applied by tradition. In his abovementioned contribution on the LLN, he indicated a more general scheme that enabled to cover the studied phenomenon in a much subtler way. He considered a sequence of series

$$x_{11}, x_{12}, \ldots, x_{1k_1}; x_{21}, x_{22}, \ldots, x_{2k_2}; \ldots; x_{n1}, x_{n2}, \ldots, x_{nk_n} \quad (7)$$

with the random variables being independent within each of these. The phenomenon under examination is described ever more precisely by the laws of distribution of the sums

$$S_n = x_{n1} + x_{n2} + \ldots + x_{nk_n}. \quad (8)$$

Here, we are not anymore interested in the dependence between the random variables of different series and it is now not necessary to norm and center the sums S_n.

The following general example can illustrate Kolmogorov's pattern. Let us study some phenomenon depending on the action of random causes. It can be approximately described by the variables of the first sequence, $x_{11}, x_{12}, \ldots, x_{1k_1}$. We then apply more precise methods of observation, allow for the previously rejected causes, separate the considered causes into groups, discard those which do not actually act, – and obtain the second series, $x_{21}, x_{22}, \ldots, x_{2k_2}$, for describing the phenomenon. Continuing this process of approxi-mating reality, we arrive at the next series in (7) and the influence of the separate terms decreases as n increases. This last-mentioned circumstance, justified by natural-scientific considerations, is thus formulated in the theory of probability: In the theory of limiting laws for sums, it is thought that the variables x_{nk} $(1 \le k \le k_n)$ are *asymptotically negligible*; that is, uniformly with respect to k $(1 \le k \le k_n)$ the relation

$$P(|x_{nk}| \ge \varepsilon) \to 0 \text{ as } n \to \infty$$

holds for any positive constant ε.

11. For understanding the sequel, we ought to become acquainted with an important class of the laws of distribution, with the so-called infinitely divisible laws. A law is infinitely divisible if a random variable subordinated to it can be represented, for an arbitrary natural number n, by a sum of n independent and identically distributed terms (6).

The development of the theory of time-homogeneous stochastic processes stimulated the interest to infinitely divisible laws. Bruno de Finetti initiated the study of the theories of both homogeneous processes and infinitely divisible laws. Only Kolmogorov (1932), however, attained essential progress here by deriving the formula for the logarithm of a characteristic function for an infinitely divisible law. He restricted his study to such laws having finite second moments. Soon Lévy (1934) extended the Kolmogorov formula onto the general case although without providing a rigorous proof. Khinchin (1937) furnished a full proof of the Kolmogorov – Lévy formula.

As a result of all these investigations, it occurred that for a law $\Phi(x)$ to be infinitely divisible it was necessary and sufficient that the logarithm of its characteristic function be expressed as

$$\lg \varphi(t) = i\gamma t + \int_{-\infty}^{\infty} [e^{iut} - 1 - \frac{iut}{1+u^2}]\frac{1+u^2}{u^2}dG(u).$$

Here, $G(u)$ is a non-decreasing function with bounded variation on $(-\infty; \infty)$ and γ is a real constant.

It is easy to understand that the Gauss law

$$\Phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^{x} \exp(x - z^2)\, dz$$

is infinitely divisible; indeed, in this case we ought to set, in the Kolmogorov – Lévy formula, $\gamma = 0$ and $G(u) = 0$ for $u < 0$ and $G(u) = 1$ for $u > 0$.

12. In 1934 – 1935, Bavli, a student of Kolmogorov, proved the following superb proposition{s}.

The First Bavli Theorem. If random variables $x_{n1}, x_{n2}, \ldots, x_{nk_n}$ are independent within each series,

$$\mathrm{E}\, x_{nk_n} = 0,\ \mathrm{E}x^2_{nk} < C,\ \max \mathrm{E}x^2_{nk} \to 0 \text{ for } 1 \le k \le k_n,\ n \to \infty$$

where C is a constant, then the laws of distribution of the sums (6) converge in measure to some sequence of infinitely divisible laws of distribution. It was Kolmogorov who formulated the idea about such a convergence of the laws of distribution of sums to a sequence of specially selected infinitely divisible laws; later he repeatedly expressed the hypothesis that this convergence takes place always if only the terms x_{nk} are asymptotically negligible.

As a corollary to his general theorem, Bavli obtained a proposition on the convergence of the laws of distribution of sums to a limiting law:

The Second Bavli Theorem. As n unboundedly increases, the necessary and sufficient condition for the functions of distribution of the sums (8) to converge to some limiting distribution function is this: it is possible, under the conditions of the first theorem, to select such a sequence of constants $\alpha_1, \alpha_2, \ldots, \alpha_n, \ldots$ that, as $n \to \infty$, the sequence of functions

$$\sigma_n(x) = \sum_{k=1}^{n} \int_{-\infty}^{x} \xi^2 dF_{nk}(\xi) + \alpha_n$$

tends to some non-decreasing function $G(x)$ at each point of its continuity.

13. In 1936, Khinchin became able to prove a remarkable theorem that determined the class of limiting laws for sums. It enabled better to establish the part of infinitely divisible laws and to simplify the proof of some previously established propositions.

The Khinchin Theorem. Each limiting law for a sequence of the laws of distribution of the sums (8) of terms asymptotically negligible and independent within each series is infinitely divisible. Inversely, each such law is limiting for some sums of such terms.

Also in 1936 Khinchin formulated the problem about the class of limiting laws for the classical pattern, *i.e.*, of the laws that can be limiting for the sequence of the laws of distribution of the sums (4) when the constants A_n and B_n are appropriately chosen. This class is naturally included in the class of infinitely divisible laws and Lévy soon exhaustively described it.

14. We ought to indicate now that essential progress was made in studying the convergence of the laws of distribution of sums to the Gauss law. In 1935, almost at the same time, Khinchin, Feller, and Lévy published contributions devoted to this issue. Khinchin provided a necessary and sufficient condition for the convergence of the laws of distribution of the sums (4) with appropriately chosen coefficients $B_n > 0$ and A_n of independent identically distributed terms to the Gauss law. Feller discovered necessary and sufficient conditions for the same convergence when the terms were not identically distributed.

The Feller Theorem. A sequence (1) of independent random variables is given, and it is required to choose such constants $B_n > 0$ and A_n that the laws of distribution of the sums (4) converge to the Gauss law and that the terms x_k/B_n $(1 \le k \le n)$ are asymptotically negligible. For this to be possible, the existence of such a sequence of positive constants C_n that, as $n \to \infty$,

$$\sum_{k=1}^{n} \int_{|x|>C_n} dF_k(x) \to 0, \sum_{k=1}^{n} (1/C_n) \int_{|x|<C_n} x^2\, dF_k(x) \to 0,$$

$$\sum_{k=1}^{n} (1/C^2{}_n)\{ \int_{|x|<C_n} x^2\, dF_k(x) - [\int_{|x|<C_n} x\, dF_k(x)]^2\} \to \infty$$

is necessary and sufficient.

In the same work Feller proved that the Lindeberg condition was not only sufficient, but also necessary for the convergence to the Gauss law of the laws of distribution of the sums normed by the variances

$$B_n = \sqrt{\sum_{k=1}^{n} E(x_k - Ex_k)^2}$$

and centered by the expectations $A_n = \mathrm{E}x_1 + \mathrm{E}x_2 + \ldots + \mathrm{E}x_n$.

Lévy also provided a necessary and sufficient condition for the convergence to the Gauss law; moreover, his formulation was extremely interesting and offered new ideas. His necessary and sufficient condition for the indicated convergence consisted in that the maximal term should be asymptotically negligible. At first his formulation was insufficiently clear and the proof unconvincing, but later he furnished a fundamentally new substantiation with an easily understandable formulation. Issuing from his result, Khinchin proved the following theorem ascertaining the basic importance of the Gauss law in the theory of limiting laws.

The Khinchin Theorem (cf. the theorem in §13). If the laws of distribution of sums (8) of terms independent within each series and asymptotically negligible converge to a limiting law, then the relation

$$P(\text{at least one } |x_{nk}| \ge \varepsilon,\ 1 \le k \le k_n) \to 0 \text{ as } n \to \infty$$

holding for any $\varepsilon > 0$ is a necessary and sufficient condition for the limiting law to be Gaussian.

15. In 1935 Khinchin introduced the notion of relative stability of sums for non-negative random variables. Suppose that the sequence (1) of independent random variables is given. The sums (6) are called relatively stable if such a sequence of positive constants B_n can be chosen that for any positive constant

ε, as $n \to \infty$,

$$P\{|\sum_{k=1}^{n} (x_k/B_n) - 1| < \varepsilon\} \to 1.$$

Evidently, relative stability is a natural formulation of the LLN for positive random variables.

At the same time Khinchin discovered a necessary and sufficient condition for relative stability of sums of identically distributed variables, and indicated that his condition was analogous to the convergence to the Gauss law in a similar case. Just after Feller's contribution had appeared, Khinchin formulated the conditions for relative stability in the general case and his student Bobrov soon proved this hypothetical theorem.

The Bobrov Theorem. A sequence (1) of non-negative random variables is given. For the sum (4) to be relatively stable, and the terms (x_k/B_n), $1 \le k \le n$, asymptotically negligible, the existence of such a sequence of positive constants C_n, that, as $n \to \infty$,

$$\sum_{k=1}^{n} \int_{C_n}^{\infty} dF_k(x) \to 0, \sum_{k=1}^{n} (1/C_n) \int_{0}^{C_n} xdF_k(x) \to \infty$$

is necessary and sufficient.

In a conversation, Khinchin also formulated a proposition similar to the Lindeberg theorem. I have soon proved it; and, together with Bobrov, we then substantiated it by another method. In 1938 Khinchin's student Raikov formulated this analogy noticed by Khinchin as a first-rate theorem.

The Raikov Theorem. Let (1) be independent random variables centered by their medians,

$$F_k(-0) \le 1/2 < F_k(+0),$$

and $B_n > 0$ and A_n be real constants. For the sequence of the laws of distribution of the sums (4) to converge to the Gauss law it is necessary and sufficient that the sequence of the sums of the squares of these variables,

$$x_1^2 + x_2^2 + \dots + x_n^2,$$

is relatively stable.

16. We turn now to considering the sequence (1) of identically distributed random variables with distribution function $F(x)$. Khinchin proved that for any infinitely divisible law of distribution $\Phi(x)$ it is possible to choose such a sequence (1) and such constants B_n and A_n that for some sequence of natural numbers k_n the laws of distribution of sums

$$S_{k_n} = \frac{x_1 + x_2 + \dots + x_{k_n}}{B_n} - A_n \qquad (9)$$

converge to it. It is said that the law $F(x)$ is attracted to $\Phi(x)$. This theorem shows the importance of a thorough study of the behavior of the laws of distribution for the sums (9). Lévy, Gnedenko and Doeblin attained a number of pertinent findings.

The totality of the laws of distribution $F(x)$ attracted to a given law $\Phi(x)$ is called the *domain of* its *partial attraction*, and if the attraction is restricted by the condition $k_n = n$, we are dealing with the *domain of complete attraction*. Understandably, only stable laws possess such a domain.

It occurred that if $F(x)$ belonged to a domain of partial attraction of a non-stable law, it was also attracted to a non-countable set of other laws differing one from another. It was also found that there exist absolutely divergent laws, – such that were not attracted to any limiting law (Lévy, Khinchin, Gnedenko, Doeblin), – as well as such (universal) laws which are attracted to all the infinitely divisible laws (Doeblin). Doeblin and Gnedenko determined the domains of complete attraction for stable laws.

17. In 1937, issuing from a lemma due to Khinchin, I was able to prove the following proposition{s}.

Theorem 1. Given sums (8) of terms asymptotically negligible and independent within each series. For these sums to converge to a limiting law it is necessary and sufficient that the laws of distribution $\Phi_n(x)$ of the sums

$$S'_n = x'_{n1} + x'_{n2} + \ldots + x'_{nk_n}$$

of infinitely divisible terms converge to a limiting law. The functions $\Phi_n(x)$ are determined by the logarithms of their characteristic functions

$$\lg \varphi(t) = \sum_{k=1}^{k_n} \{i\omega_{nk}(1)t + \int_{-\infty}^{\infty} [e^{iut} - 1 - \frac{iut}{1+u^2}]\frac{1+u^2}{u^2} dF_{nk}(u) + \omega_{nk}(1)\}$$

where $F_{nk}(x)$ is the law of distribution of the variable x_{nk}. The limiting laws for both sequences coincide.

To apply this proposition for deducing limiting laws, we apparently ought to indicate the conditions for the convergence of infinitely divisible laws.

Theorem 2. For the sequence $\Phi_1(x), \Phi_2(x), \ldots, \Phi_n(x), \ldots$ of infinitely divisible laws to converge to a limiting law $\Phi(x)$ it is necessary and sufficient that the following conditions be fulfilled.

a) $G_n(u) \to G(u)$ as $n \to \infty$ at the points of continuity of $G(u)$.

b) The total variations of the functions $G_n(u)$ converge to the total variation of $G(u)$.

c) $\gamma_n \to \gamma$ as $n \to \infty$.

The functions $G_n(u)$ and $G(u)$ as well as the constants γ_n and γ are determined by the Lévy – Kolmogorov formula for the laws $\Phi_n(x)$ and $\Phi(x)$.

Since the limiting law for infinitely divisible laws is itself infinitely divisible, Theorem 1 enables to derive as a corollary the Khinchin theorem from §13 on the class of limiting laws. As a second corollary of Theorems 1 and 2 I indicate the following proposition.

[**Theorem 3**]. Given, a sequence of sums (6) of independent and asymptotically negligible terms. For the convergence of their laws of distribution to some limiting law it is necessary and sufficient that there exist such non-decreasing functions $M(u)$ and $N(u)$ and such constants $\alpha > 0$ and γ that

a) $M(-\infty) = 0$, $N(+\infty) = 0$, $\int_{-\varepsilon}^{-0} u^2 dM(u) + \int_{\varepsilon}^{+0} u^2 dM(u) < +\infty$.

b) For $x < 0$, at the points of continuity of $M(x)$,

$$\sum_{k=1}^{k_n} F_{nk}(x) \to M(x) \text{ as } n \to \infty.$$

c) For $x > 0$, at the points of continuity of $M(x)$,

$$\sum_{k=1}^{k_n} [F_{nk}(x) - 1] \to N(x) \text{ as } n \to \infty.$$

d) $\sum_{k=1}^{k_n} \int_{-\infty}^{0} \frac{x}{1+x^2} dF_{nk}(x) \to \gamma$ as $n \to \infty$.

e) $\sum_{k=1}^{k_n} \{ \int_{|x|<\varepsilon} x^2 dF_{nk}(x) - [\int_{|x|<\varepsilon} x dF_{nk}(x)]^2 \} \to \alpha^2$ as $n \to \infty$ and $\varepsilon \to 0$.

The limiting law is determined by the Lévy – Kolmogorov formula where γ is given by condition d) and

$$G(u) = \int_{-\infty}^{u} \frac{u^2}{1+u^2} dM(u),\; u < 0;$$

$$G(u) = \int_{-\infty}^{-0} \frac{u^2}{1+u^2} dM(u) + \alpha + \int_{u}^{+0} \frac{u^2}{1+u^2} dN(u),\; u > 0. \qquad (10)$$

Theorem 3 enables to obtain, with hardly any work, all the propositions dealt with in this paper. Theorems 1 – 3 can thus be considered as the basis of a general method for arriving at limiting laws for sums of asymptotically negligible terms. The idea of this method is that we replace the sums of arbitrary independent terms by sums of closely connected with them infinitely divisible variables. We thus avoid the difficulties encountered when directly dealing with laws of distribution or when applying characteristic functions, but at the same time make use of the advantages of both these tools.

The laws of distribution are of a simple functional nature, but for them the addition of random variables is very complicated. In terms of characteristic functions this operation is simple, but their analytical nature is very involved. In the method that I propose, the determination of characteristic functions is reduced, in accord with the Lévy – Kolmogorov formula, to the derivation of a monotonically non-decreasing function with a bounded variation and the addition of random variables is replaced by the addition of these functions. And, for the practically most important functions (the Gauss, the Poisson, the stable laws), the functions $G(u)$ in the abovementioned formula are incomparably simpler than their corresponding laws of distribution. Owing to these circumstances, my method not only enables to unify the proofs of isolated (with respect to their derivation) facts, but to decrease essentially the calculations. Thus, the Feller theorem required lengthy explication (50 printed pages) from its author whereas its justification by my method is completed literally in a few lines.

18. As an example, I adduce a derivation of the conditions for the convergence of the laws of distribution of sums to the Gauss law for the Kolmogorov pattern. For the Gauss law the function $G(u)$ has a single point of growth $u = 0$ where it experiences a saltus equal to 1; the functions $M(u)$ and $N(u)$ in Theorem 3 of §17 are therefore identically equal to zero, $\alpha = 1$ and $\gamma = 0$. In accord with this proposition, for the convergence of the laws of distribution to the Gauss law it is necessary and sufficient, as $n \to \infty$, that

a) $\sum_{k=1}^{k_n} F_{nk}(x) \to 0$ for $x < 0$; $\sum_{k=1}^{k_n} [F_{nk}(x) - 1] \to 0$ for $x > 0$;

b) $\sum_{k=1}^{k_n} \int_{-\infty}^{\infty} \frac{x}{1+x^2} dF_{nk}(x) \to 0$;

c) $\sum_{k=1}^{k_n} \{ \int_{|x|<\varepsilon} x^2 dF_{nk}(x) - [\int_{|x|<\varepsilon} x dF_{nk}(x)]^2 \} \to 1$ as $\varepsilon \to 0$.

It is easy to understand that these conditions can be formulated otherwise: For the indicated convergence it is necessary and sufficient that, for any ε > 0 and $n \to \infty$,

$$\sum_{k=1}^{k_n} \int\limits_{|x|<\varepsilon} dF_{nk}(x) \to 0, \sum_{k=1}^{k_n} \int\limits_{|x|<\varepsilon} xdF_{nk}(x) \to 0,$$

$$\sum_{k=1}^{k_n} \{ \int\limits_{|x|<\varepsilon} x^2 dF_{nk}(x) - [\int\limits_{|x|<\varepsilon} xdF_{nk}(x)]^2\} \to 1.$$

We see that the obtained conditions are very close to those put forward by Feller.

19. I complete my paper by a short list of references and indicate some problems which can be solved by my method (§17).

For an initial acquaintance with the theory of probability:

1. Glivenko, V.I. *Теория вероятностей* (Theory of Probability), 1938. {New edition: 1939.}

For a more detailed study of the modern problems of the theory of limiting laws discussed above I recommend, in the first place,

2. Khinchin, A.Ya. *Предельные законы для сумм независимых случайных величин* (Limiting Laws for Sums of Independent Random Variables). M. – L., 1938.

It is impossible to fail recommending for a most thorough examination two of his other monographs:

3. *Асимптотические законы теории вероятностей* (Asymptotic Laws of the Theory of Probability). M. – L., 1936. {German version, 1933.}

4. *Основные законы теории вероятностей* (Main Laws of the Theory of Probability). M. – L., 1926 and 1932.

Among foreign sources I ought to indicate above all

5. Lévy, P. *Calcul des probabilités.* Paris, 1925.

6. Lévy, P. *Théorie de l'addition des variables aléatoires.* Paris, 1937.

The method expounded in §§17 and 18 along with its applications to a number of classical problems is described in my writing

7. К теории предельных теорем для сумм независимых случайных величин (On the Theory of Limiting Theorems for Sums of Independent Random Variables).{*Izvestia* AN SU, ser. math., No. 2, 1939, pp. 181 – 232.}

I do not include{other} papers published in periodicals; a pertinent detailed bibliography is contained in [6] and [7]; brief bibliographic indications are to be found in each monograph listed above.

Here, now, are the problems for independent investigation.

1. Theorem 3 of §17 enables to find the necessary and sufficient conditions for the convergence of the laws of distribution of sums of independent terms to any infinitely divisible law for the Kolmogorov pattern. For the classical scheme, however, this problem is not yet solved.

2. The conditions that make it possible to determine whether a given law belongs to the domain of partial attraction of a given infinitely divisible law are still unknown.

3. Two laws, $F_1(x)$ and $F_2(x)$, belong to the domain of partial attraction of the law $\Phi(x)$; does the law $F(x) = F_1(x) * F_2(x)$, that is, the law of distribution of the sum of independent terms having laws $F_1(x)$ and $F_2(x)$, belong to the domain of partial attraction of the same law $\Phi(x)$?

4. The properties of universal laws are yet unstudied; neither are the conditions for a given law to belong to the class of universal laws established.

5. Determine all the laws attracting a given law (Kolmogorov).

Notes

1.{The author should have mentioned Borel.}

2.{The author cited the theory of errors which he had indeed mentioned in §§4 and 5: I omitted both these cases; in one of them he (§4) wrongly stated that Gauss had justified his theory of errors in the same way (by the action of a large number of causes, etc). There he also mistakenly stated that De Moivre had only proved a particular case of the De Moivre – Laplace limit theorem.}