

Evgeny Slutsky

Collected Statistical Papers

Selected and Translated by Oscar Sheynin
Assisted by Guido Rauscher and Claus Wittich

Berlin, 2010

ISBN 3-938417-82-X
© Oscar Sheynin, 2010

www.sheynin.de

Contents

Foreword

I. *Theory of Correlation and*

Elements of the Doctrine of the Curves of Distribution, 1912. Foreword

II. Statistics and mathematics, 1916

III. On the logical foundation of the calculus of probability, 1922

IV. On some patterns of correlation connection

and the systematic error of the correlation coefficient, 1923

V. On a new coefficient of mean density of population, 1923

VI. On calculating the state revenue

from the emission of paper money, 1923

VII. Mathematical notes on the theory of emission, 1923

VIII. On the law of large numbers, 1925

IX. Al. Tschuprow, 1926

X. On the distribution of errors [on the law of distribution] of the correlation coefficient in homogeneous connected series, 1932

XI. On the existence of connection between the solar constant and temperature, 1933

XII. On the solar constant, 1934

XIII. On the eleven year periodicity of sunspots, 1935

XIV. Statistical experiment as a method of investigation.

Critical notes on the problem Earth – Sun, 1935

XV. G. Rauscher, O. B. Sheynin, C. Wittich, The correspondence between E. E. Slutsky and V. I. Bortkevich, 2007

XVI. Autobiography, 1939

XVII. Autobiography, 1942

XVIII. O. Sheynin, Slutsky: Commemorating the 50th anniversary of his death, 1999

XIX. N. S. Chetverikov, The life and scientific work of E. E. Slutsky, 1959

XX. B. V. Gnedenko, N. V. Smirnov, Foreword to Slutsky's *Selected Works*, 1960

Foreword

1. General Information

1.1. For a first approximation to Evgeny Evgenievich Slutsky's (1880 – 1948) biography see [xix]. I have also included other materials about him [xviii]; [xx] and his own autobiographies [xvi; xvii], regrettably very short. Among obituaries I single out those written by Kolmogorov and Smirnov, both in 1948 and quoted by Chetverikov [xix]. Much information about Slutsky is contained in several Russian archives and still largely unstudied.

Slutsky was an outstanding scholar remembered for his achievements in economics, statistics and theory of probability. *As an economist, he enjoys worldwide renown as one of the forerunners of econometrics* (Zarkovitch 1956, p. 338/1977, p. 484). See [xv, Note 20]. Slutsky saw that his economic studies became impossible; mathematical methods had only entered Soviet economics in the 1960s, and, for that matter, with great difficulties; the Conjuncture Institute, where he had been a consultant, was shut down and statisticians in general became muzzled (Sheynin 1998; 2008, pp. 365 – 367); theological issues seriously interested him, but he could only discuss them with relatives and closest friends. In other words, he had been experiencing the usual fate (by far not its worst possible version) of the Soviet intelligentsia.

Theoretical statistics was Slutsky's stepping stone to probability; moreover, two of his papers here included [iii; viii] were devoted to the theory of probability, but at least chronologically they belong to the *statistical period* of Slutsky's life and directly bear on statistics. Two papers [vi; vii] treated the emission of paper money, and one [v] dwelt on the density of population, both subjects important but rarely discussed by statisticians. Also important were his studies of the correlation theory. In applications, he considered as *most fruitful* his geophysical contributions [xvi], but later he [xvii] stated that the appropriate period of his life was definitively lost owing to the impossibility of carrying out comprehensive studies.

I believe that the loss was only comparative, with respect to what was possible under more favourable conditions. Incidentally he many times expressed his (failed) intentions to further his work in the same direction. And I ought to stress that during the *statistical period* of his life, Slutsky remained one of the very few leading Soviet statisticians and that he time and time again referred to Chuprov, officially considered a scholar refusing to return to Russia. At the same time, Slutsky invariably calculated and provided his numerical results with superfluous (and therefore dangerous) digits. I [vii, Note 5] remarked on the most glaring example of this habit. Other unpleasant features are insufficient and sometimes careless explanation of his subject and the really bad, and again sometimes carelessly written English summaries to his geophysical papers.

In spite of the above, calculations were Slutsky's strong point which is clearly seen in his geophysical works. Here is Kolmogorov's pertinent opinion (1948/2002, p. 71): Slutsky was

Not embarrassed by corrupting the purity of his method [of solving problems when the analytic approach had failed]. If tables became necessary, [...] he was prepared to spend years compiling them.

Kolmogorov certainly meant Slutsky's noteworthy contribution, the table of the Γ -function.

From time to time, and especially at anniversaries of the October (old style) 1917 coup d'état, essays on the state of various sciences were being published. I (2005) have collected translations of such contributions on probability and statistics, and it is not difficult to find there many references to Slutsky. Kolmogorov, in 1935, 1938 and 1948 stressed the importance of his work on random functions and placed him alongside Wiener and Lévy (in 1935), and together with himself in 1938. In 1948, in a joint publication with Gnedenko, he repeated the latter statement and singled out Slutsky (1937). Then, in 1947, Kolmogorov named Khinchin, himself and Slutsky as the originators of the Moscow school of probability.

Smirnov, in 1948 (not in the obituary of the same year) stated that Slutsky, Khinchin and Kolmogorov largely created the theory of continuous stochastic processes and Gnedenko, in 1970, noted that Bernstein and Slutsky were the first Soviet authors on the theory of probability and mathematical statistics.

The tradition of publishing fundamental essays had a horrible ideological aspect. Thus, Khinchin (1937), of all men, wrote a servile contribution falsely describing the situation of science in pre-revolutionary Russia and comparing it with the alleged splendid position of the day, and that at the time when the Great Terror was in full swing!

Acknowledgement. It is my pleasant duty to mention Magister Guido Rauscher (Vienna) and Dr. Claus Wittich (Geneva). All three of us jointly published [xv] and it was G. R. who had discovered the Bortkiewicz papers (including his correspondence with Slutsky) in Uppsala. He had also found out that important and still largely unstudied material concerning Slutsky is kept in RGALI (Russian State Archive for Literature and Arts).

Claus Wittich partly edited my translation of [vi] and sent me the text of [xiii]; incidentally, that contribution had appeared both in Russian and English, and I have just reprinted the English version. I have also profited from two of his unpublished texts of 2005 and 2007 which he put at my disposal, Biographical notes on, and Bibliographical notes on selected sources concerning Slutsky.

I will now formulate some comments on most of the included papers. References to literature mentioned there are included in the Bibliographies to those papers, but I am providing the information about the sources mentioned above right now:

Khinchin A. Ya. (1937), The theory of probability in pre-revolutionary Russia and in the Soviet Union. *Front Nauki i Tekhniki*, No. 7, pp. 36 – 46. Translation: Sheynin (2005, pp. 40 – 55).

Sheynin, O. (1998), Statistics in the Soviet epoch. *Jahrbücher f. Nationalökonomie u. Statistik*, Bd. 217, pp. 529 – 549.

---, compiler and translator (2005), *Probability and Statistics. Soviet Essays*. Berlin. Also at www.sheynin.de

--- (2008), Romanovsky's correspondence with K. Pearson and R. A. Fisher. *Archives Intern. d'Histoire des Sciences*, t. 58, No. 160 – 161, pp. 365 – 384.

Slutsky E. (1937), *Qualche proposizione relative alla teoria delle funzioni aleatorie*. *Giorn. dell. Istituto Italiano degli Attuari*, t. 8, No. 2, pp. 3 – 19.

Zarkovitch S. S. (1956), Note on the history of sampling methods in Russia. *J. Roy. Stat. Soc.*, vol. A119, pp. 336 – 338. Reprinted in Kendall M., Plackett R. L. (1977), *Studies in the History of Statistics and Probability*, vol. 2. London, pp. 482 – 484.

1.2. Comments on Separate Papers

[iii] Kolmogorov (1948/2002, p. 69) stated that Slutsky “was the first to draw a correct picture of the purely mathematical essence of probability theory” and cited the paper here translated (“the present paper”, as I shall call it) and a later contribution (Slutsky 1925). Earlier, Kolmogorov (1933) referred to both these articles but did not mention the former in the text itself; curiously enough, that inconsistency persisted even in the second Russian translation of Kolmogorov’s classic published during his lifetime (Kolmogorov 1974, pp. 54 and 66).

Several years after 1922 Slutsky [viii, Note 2] remarked that back then he had not known Bernstein’s work (1917) which “deserves a most serious study”.

In his Commentary, B. V. Gnedenko (Slutsky 1960, p. 284) most approvingly cited a passage here italicized in § 5 and, on p. 285, concluded that Slutsky had

Correctly and deeply (and apparently for the first time) approached the construction of the theory of probability in a rigorous and purely mathematical way. His paper played an important part in forming contemporary ideas about the foundations of the theory of probability and occupies a noticeable place in its history.

This English translation of [iii] first appeared in Sheynin (2005).

[iv] In a letter of 1924 to Chetverikov, Chuprov (Sheynin 1990/1996, p. 49) commented:

I have recently received from Slutsky reprints of his papers. For me, the work [the present article] is very interesting; both in its approach and in the results obtained it accords with what I had arrived at for the correlation coefficient.

There seems to be no investigation of the systematic error of that coefficient in Chuprov’s published works; however, Slutsky himself several times referred to Chuprov and Chuprov (1923, Appendix) contains all the formulas from the beginning of § 3 to (7) inclusively. Both that contribution and the present paper had appeared at about the same time. He (1925) later mentioned Slutsky’s paper in the appended Review of Literature but (§ 5 of Chapter 6) only discussed the systematic error of the correlation coefficient in a few lines and noted that it became essential in cases of a small number of observations.

I have not found any comparatively recent references to the systematic error of the correlation coefficient, but I quote Prokhorov (1999):

For a large number of independent observations having one and the same near-normal distribution, the sample correlation coefficient is close to the real coefficient. In all other cases, the correlation ratio is recommended instead.

Nevertheless, it is still possible that Slutsky's contribution deserves to be recalled. Finally, I note that Slutsky wrote *normal* distribution in inverted commas although it was high time for dropping them.

[v] The author published this paper in a periodical intended for a broader circle of statisticians which apparently explains the somewhat excessively detailed calculations. True, he later published there much more mathematically oriented contributions. This time, I nevertheless think that in his context the remark (§ 3) about the incommensurability of certain areas was absolutely unwarranted.

In §§ 6 and 7 Slutsky calculated populations per square *versta* (an old Russian unit of length, 1.0668km). I replaced his figures passing on to densities per square kilometre. For the sake of brevity I usually omitted the "*sq. km*". Slutsky also applied an old Russian unit of area, see Note 2.

Valentei (1985, article *Slutsky* on p. 409) stated: "In demography, his name is connected with the so-called coefficient of mean social density of population". And on p. 329, in the article *Density of population*, that density is mentioned along with physical density.

Social density of population is also known in the English language literature (and possibly universally), but I am not sure that in a strict sense.

[vi – vii] Slutsky compiled this contribution "at the request of my [of his] friend Prof. L. N. Iasnopolsky" [xv, Letter No. 4]. The subject of his study was indeed important as witnessed, first, by his reference to a paper by Schmidt, the future (from 1935) academician and, much later, vice-president of the Academy of Sciences of the Soviet Union, and, second, by the appearance of his second paper [vii] published by the Conjecture Institute.

I suspect that [vi, formulas (34) to (38)] notation J^1 should have been J' .

[viii] Among the obvious features of this contribution are Slutsky's numerous and most respectful references to Chuprov, and a similar attitude to the law of large numbers (incidentally, he almost always writes these words in inverted commas whereas I abbreviate them as LLN) which leads him to excessive philosophizing.

Khinchin (1928) later published a short paper on the strong law of large numbers in the same periodical. He (pp. 124 – 125) approvingly mentioned Slutsky in connection with the stochastic limit [viii, § 2] although did not explain that notion, nor did he provide any exact reference but he (p. 125) stated that "The true basis of the statistical applications of the law of large numbers is the strong rather than its usual notion". Khinchin did not criticize Chuprov or Slutsky [viii]; still, the very absence of anything resembling their philosophical deliberations speaks for itself. He described the conditions for the strong law of large numbers to take place, and it was he who apparently introduced that term into Russian scientific literature. As a tiny diversion, I note that he (p. 124) wrongly believed that statisticians had "successfully" estimated probability by issuing from frequency and referring to the LLN. On the contrary, as witnessed at least by Chuprov and Slutsky, they remained here at a loss. Even in 1923, in a letter to Chetverikov, Chuprov (Sheynin 1990/1996, p. 97) acknowledged that he did not see any possibility of "throwing a formal logical bridge across the crack separating frequency from probability".

In § 2 (p. 5 of the Russian text) Slutsky states that probability is the stochastic limit of frequency. Now, this is the inverse law of large numbers (Sheynin 2010) which Bernoulli had not (but thought he) proved; exactly

that was his aim. Neither he (nor De Moivre after him), nor Slutsky were mistaken, but the precision of establishing probability through frequency is lower than the precision of frequency when determined through probability. It was only Bayes who understood this circumstance, and Slutsky should have elaborated on his statement. He did not, however, find himself on the path to the frequentist theory of probability, see below, so that the situation is somewhat indeterminate.

The same year Slutsky published a companion German paper (1925) translated into Russian in Slutsky (1960). There, in a commentary, Gnedenko (pp. 285 – 286) highly estimated Slutsky's introduction of stochastic asymptotes (also in § 6 of the Russian paper here translated); however, that notion seems to be forgotten, or, rather, not incorporated into one of the various kinds of convergence applied in probability. Second, Gnedenko noted that Slutsky had criticized Mises (without mentioning him) because the LLN deals with the stochastic rather than “usual” limit. In a weaker form, that criticism is also expressed here, in § 16.

Slutsky tediously discussed the then recent upheaval of geometry (§ 10). He could have mentioned that the entire development of mathematics, beginning with its emergence as the result of introducing natural numbers, consisted in such upheavals. Slutsky also referred to Hilbert's axiomatic approach to geometry and considered his work as a pattern to be followed in probability. It is difficult to understand, however, why did he pass over in silence Hilbert's explicit and famous demand (wish) to see the probability theory axiomatized. A related Slutsky's remark was contained in his letter to Markov back in 1912 [xviii, § 3]: *I consider it possible to develop all the Pearsonian theories by issuing from rigorous abstract assumptions.*

Slutsky consistently applied the terms *random variable* and *theory of probability*. The present Russian terms are *random magnitude* (regrettably), although Khinchin (1928) followed Slutsky, and *theory of probabilities*. Markov, the conservatively inclined great scientist, denied *random variable* (or *magnitude*) and used instead the decidedly worse expression *indefinite magnitude*; incidentally, the translators of Ondar (1977) had inadmissibly modernized him. Cantelli (1916a, p. 192) was likely the first to introduce the term *random variable* (in Italian), see Mises (1964, p. 52, Note 2), a posthumous contribution.

Concerning the random variable, Slutsky followed Chuprov (1922, at the very beginning); on the other hand, at least sometimes Chuprov (1909/1959, p. 13) wrote *theory of probabilities*. Then, Slutsky (beginning of § 5) also introduced *distribution of probabilities* (law of distribution) of a discrete random variable and, in his Note 8, properly mentioned Chuprov (1922, at the very beginning). I did not find anything similar in Czuber (1903/1908). Markov (1900/1924) also introduced it earlier than Slutsky, but only on p. 74, in a chapter on the LLN, and did not name it at all.

Slutsky did not apply the notation of the type \bar{x} for the arithmetic mean (I myself introduced it in the translation) although he himself did so previously, for example in a paper in the same periodical [iv, § 1], and called it *usual*.

The most important point is, however, that, issuing from the paper translated below, Slutsky “arrived at the notion of stochastic process”, see [xvii].

[ix] Slutsky's description of Chuprov's *Ocherki* (Essays) (1909) was quite consistent with its general appraisal. Markov's opinion (1911/1981, p. 151), was neglected: they lacked "that clarity and definiteness that the calculus of probability requires". Even more: the reader of our time will be lost in the ocean of Chuprov's general and hardly necessary considerations and simply will not understand his stress on philosophy and logic at the expense of mathematics.

Markov (*Materialy* 1991, p. 195), however, was outspoken; in a letter of 1910 to V. A. Steklov, the future vice-president of the Russian Academy of Sciences, he wrote:

From the mathematical point of view, [the Ocherki] contain much more nonsense than [the dissertation of Orzensky, a Russian statistician not mathematically oriented]; it is certainly necessary to reject it.

I (Sheynin 2009b, pp. 5 – 9) severely criticized that contribution; here, I briefly repeat some of my considerations leaving aside such points as Chuprov's timid (at best) disapproval of Bortkiewicz' alleged law of small numbers (Sheynin 2008); his strange failure to discuss randomness; his mistaken belief in Cournot's "canonical" proof of the law of large numbers and hesitant attitude about its meaning for statistics.

Chuprov (1905; 1906; 1909) discussed *nomological* and *ontological* relations; the former applied to certain phenomena universally or at least in general, the latter's action was restricted in space and/or time, but was it really necessary to introduce these concepts into statistics? He borrowed those concepts from German philosophers Rickert and Windelband whom historians of that science barely remember, but statisticians have definitely forgotten about them.

What Chuprov could have noted, but obviously did not know, was the appearance, in 1825, of the so-called *numerical method*, actually known in various branches of natural sciences. It was based on facts almost without any theories; an example can be a chart of the starry heaven containing thousands of stars, see Sheynin (1982, § 4). The existence of that method compels me to disagree, at least partly, with Nekrasov, who, in 1896, while approving Chuprov's (unpublished) student dissertation, inserted a marginal comment on its p. 4 (Sheynin 1990/1996, p. 85) containing this passage:

Concerning [force, space, time, probability] philosophers have written full volumes of no use for physicists or mathematicians. Mill, Kant and others [certainly including Windelband and Rickert] are not better, but worse than Aristotle, Descartes, Leibniz [...].

At the time, Nekrasov was a most serious scientist but later, as far as probability and statistics was concerned, he became almost a non-entity (Sheynin 2003).

Much later Chuprov himself (Bortkevich & Chuprov 2005, Letter No. 162 of 1921) remarked that "These last years" he was "turned aside" from philosophy to mathematics. For that matter, Chetverikov [xix, § 2] made the same remark about Slutsky. Still, in 1925, in a letter to Slutsky (Sheynin 1990/1996, p. 49), Chuprov stated that he considered the analysis in [viii], a paper with an excessive emphasize on philosophy, "as perfect".

There is much more to criticize in Chuprov's works. Thus, he stated his problems in a most general manner, and his formulas became therefore extremely complicated and hardly checked by anyone during the latest decades. Here is the relevant opinion of Romanovsky (1930, pp. 416 and 417) concerning formulas of the theory of correlation: Being "of considerable theoretical interest", they are "almost useless; extremely unwieldy [...] and hardly studied".

A special cause here was Chuprov's bad system of notation. In one contribution, he (1923) even introduced horrible expressions with two-storey subscripts and superscripts of the same structure, – in all, four storeys supplementing the main line at the same time!

On the other hand, Slutsky had not noticed either Chuprov's work as a public figure (he published more than 60 newspaper articles) or reviewer (more than 20 reviews of statistical literature published during the last decade of his life and about a dozen before that). And no one knew that Chuprov advocated an all-out intervention of the West to crush Bolshevism since the "fabula narratur [...] about the fate of the European culture".

This is a statement from his letter of 1919 to someone from the Russian Liberation Committee in London. The documents of that Committee (30 volumes) are kept in the British Library in London, but for about 90 years the archivists there have been unable to compile an inventory of those materials (which once more testifies to the scornful and extremely harmful Western attitude to Russian science and culture). The quotation above is from Add 54437, pp. 123 – 128, the only code known to me.

I also note that Slutsky's note is too short likely because Mises, the Editor of the *Zeitschrift*, stipulated that it should be no more than one page long [xv, Letter No. 9].

[x] Here, as in [xi], Slutsky applied the Fisher z -transformation. The problem he solved (experimentally and not really rigorously), see the title of the contribution, was indeed difficult. This is proved by the appearance of a paper (Hawkins 1989) whose author solved a similar problem analytically, certainly not knowing about his predecessor and only more than half a century later.

One of Slutsky's previous contributions (1927) to which he referred does not really belong to the set of his works here translated, but Ondar (1977/1981, p. 144, Note 4) had made an interesting remark about it. There, Slutsky (1927/1960, p. 101) anticipated Bernstein in considering a *returning chain* (a *random bridge*).

[xi] The subject of this study is certainly important. However, for those times his explanations were insufficient. As also in his other papers, he provided many numerical results with too many significant digits thus misleading his readers. Then, he applied both the probable, and the mean square error and, moreover, without mentioning that the former was much inferior.

In his § 6, Slutsky noted that the errors of the measured values of solar constant were too large but he did not study how that fact influenced his conclusions. Moreover, that constant, as he himself noted in the very beginning of this paper, should have been measured outside the earth's atmosphere, and it was therefore necessary to say something about the impossibility of ensuring it in those times. Finally, his study essentially depended on the measurement of temperatures made by previous authors,

but he said nothing about its precision so that his investigation was not altogether complete even as a discussion of their work.

Chetverikov [xix, end of § 8] reported that Slutsky had also studied annual rings of giant sequoias of Arisona covering about two thousand years aiming to reach some conclusions about the changes of solar constant in time, but that his materials were lost during his move to Tashkent in wartime, in 1941. Also in § 8 Chetverikov noted that Slutsky's statistical papers provided a bridge for his future stochastic contributions.

[xii] Some criticism can be repeated from the above lines, but Slutsky did not mention the probable error here and the few numerical calculations were acceptable.

[xiii] Slutsky's note is difficult to understand, partly due to the brevity expected in the pertinent source and to the carelessness of the communicator. His main actual result is that the periodicity of sunspots is 11.1 years, but that the problem "deserves further study". In essence, this was not new: in 1901, without taking into account northern lights, Newcomb (Sheynin 2002, p. 155) had determined the same figure so that Slutsky's main result was that that periodicity did not change. Nowadays, its strict constancy is denied and the period is held to be approximately 11 years.

In accordance with the regulations, Slutsky's note was additionally published in one of the three main European languages, namely in English.

[xiv] The very title of this paper brings to mind the later Monte Carlo method. It was Buffon who, in 1777, decisively introduced geometric probability and an appropriate experiment into the theory of probability. His celebrated example (the fall of a needle on a set of parallel lines) prompted the generally known Laplace's comment that that experiment can provide a value for the number π . However, his comment can serve as a good illustration of Slutsky's remark (end of § 1) that experiments become unnecessary once the pertinent problem is solved theoretically.

Slutsky's main problem (comparison of series of observations of two phenomena) could have been solved by applying, for example, the Spearman coefficient of rank correlation proposed in 1906. It is suited for comparing series consisting of the same number of terms, whereas Slutsky introduced a trick allowing him to drop that restriction by artificially lengthening the shorter series (which certainly should not be too short). He was thus able to estimate the probability that the phenomena were independent or not.

However, the comparison of harvests with solar activity over a large number of years was meaningless (see Note 14). In the 19th century several authors (Sheynin 1984, pp. 159 – 160) qualitatively studied the influence of solar activity on meteorological phenomena, and Slutsky himself [xi] investigated it as well.

[xv] We publish the extant letters of the correspondence between Evgeny Evgenievich Slutsky (1880 – 1948) and Vladislav Iosifovich Bortkevich, or Ladislaus von Bortkiewicz (1868 – 1931) that constitutes a part of the latter's posthumous archive kept at the Manuskript & Musik Abteilung of the Library of Uppsala University (Sweden), Kapsel 7, and recently discovered by Guido Rauscher. Slutsky partly, and Bortkiewicz completely adhered to the (Russian) system of spelling drastically changed in 1917 – 1918. It is perhaps noteworthy that there are no extant letters written by Slutsky from Moscow after mid-1926, – when the political situation in

Russia began to worsen drastically and his own circumstances became precarious although he continued to correspond with Western colleagues such as Ragnar Frisch (Chipman 2004).

On that point see Bortkevich & Chuprov (2005, Note 178.2 on p. 305) and [xviii, § 2]. The first source quotes Ptukha's letters to Bortkevich dated 19 Febr. and 22 May 1927:

Remnants of the previous leadership of the Central Statistical Directorate and of the zemstvo statistics are being rooted out.

For Slutsky and Chetverikov life is not sweet at all.

On Ptukha see Note 1.

Bortkiewicz' letters are obviously drafts. Their reading is difficult and we were unable to decipher some words; such cases are denoted by [?]. Then, he crossed out many lines, sometimes not clearly enough and in many cases did not write out words completely. Some words and expressions in Slutsky's letters are underlined (here italicized), and there are cases when this was done very crudely, most likely by Bortkiewicz.

Among other topics, Slutsky dwelt on logical and philosophical issues connected with statistics, and it is opportune to note (Chetverikov [xix, § 10]) that in the mid-1940s he

Even with some irritation refused to discuss purely logical concepts although he had been unable to disregard the then topical criticism levelled by Fisher against the problem of calculating the probabilities of hypotheses (of the Bayes theorem).

In his letters, first from Kiev, then from Moscow, Slutsky invariably indicated his address: Nesterovskaia St. 17, flat 8, and Mashkov St. 17/15 (by N. S. Chetverikov, Chuprov's closest student), respectively.

Bortkiewicz is known to have corresponded with Slutsky since the latter (Letter No. 3) had adopted a term suggested by his colleague. That they exchanged letters from time to time was not, however, ascertained, and only recently Bortkiewicz' correspondence with Ptukha and Chetverikov came to light (Bortkevich & Chuprov 2005, p. 10). Actually, although having lived in Germany for 30 years (and about seven years before 1897), Bortkiewicz had retained ties with Russia (Sheynin 2001, p. 228; Bortkiewicz & Chuprov 2005, pp. 9 – 12).

Here is an excerpt from a letter of Chuprov to Bortkiewicz of 13.2.1923 from Dresden (Ibidem, p. 250) which apparently led to the correspondence between the latter and Slutsky:

I have recently received a letter from Ptukha. [...] I also received a letter from Evg. Evg. Slutsky, again from Kiev. He had been in Moscow, attended the stat. conference, and obtained there my address from Chetv. He tells me, among other things, that a mathematician from Central Asia [Bortkiewicz remarked here: Romanovsky] read out a report in which he arrived in a similar way at some of my results which I had published in Biometrika. Amusing! It would be good if you will be able to send him some of your reprints, and especially Homogenität etc. He has again returned to math.

stat. Delivers a course and is working himself in that field. Laments the absence of recent literature. It would be possible to send them through his relative N. Wolodkewitsch, Parkstrasse 4 [?] Berlin-Südende¹.

[xix] The essay below complements other pertinent sources, notably Kolmogorov (1948). Regrettably, however, two negative circumstances should be mentioned. First, Chetverikov quoted/referred to unpublished sources without saying anything about their whereabouts. Second, Chetverikov's mastery of mathematics was not sufficient, – he himself said so before adducing a long passage from Smirnov (1948), – and I had to omit some of his descriptions. As compared with the initial version of this essay, the second one lacks a few sentences; I have inserted them in square brackets. Then, being able to see the texts of Slutsky's autobiographies, I note that Chetverikov quoted them somewhat freely (although without at all corrupting the meaning of the pertinent passages).

A special point concerns terminology. Slutsky's term "pseudo-periodic function" also applied by Smirnov, see above, and retained in the English translation of Slutsky's paper [17], is now understood in another sense, see *Enc. of Mathematics*, vols 1 – 10, 1988 – 1994.

Chetverikov, moreover, applied a similar term, quasi-periodic function, in the same context. It is now understood differently and, in addition, does not coincide with "pseudo-periodic function" (Ibidem). Note that Seneta (2001) applies the adjective *spurious* rather than *pseudo*. Unlike Chetverikov and Kolmogorov, he also mentions Slutsky's discovery [13] that, if a sequence of random variables $\{\xi_i\}$ tends in probability to a random variable ξ , then $f(\xi_i)$, where f is a continuous function, tends in probability to $f(\xi)$.

Nikolai Sergeevich Chetverikov (1885 – 1973) was Chuprov's closest student. In 1923 – 1929 he worked at the Conjecture Institute, later in various other institutions. Four years was in prison or labour camp as a saboteur (1931 – 1934?), repressed once more (1937 – 1946?) which at least meant prohibited to live in large cities. Published two collection of articles (1963; 1975), translated many Chuprov's papers from German as well as Cournot (1843). See Sheynin (1990/1996, § 7.7), Komlev & Manellia (1990), Manellia (1998) and [xix]. As discovered by G. Rauscher, many of his unpublished and unstudied manuscripts are kept at the Moscow branch of the Archive of the Russian Academy of Science (Fond 1650).

The English translation of this essay first appeared in Sheynin (2005).

Chetverikov was certainly unable to publish some important facts, and I additionally report Kluikin's archival findings (2009, pp. 78 – 82). In the end of 1929 Slutsky experienced a nervous breakdown caused by the general political climate and the situation in the statistical circles and did not publish anything in 1930 and 1931. In 1941, he was evacuated to Tashkent rather than Kazan and for a few years dismissed from the Mathematical Institute. The cause of that intrigue remains unknown.

In 1942, in a letter to Chetverikov, Slutsky commented on the book Boiarsky et al (1930):

The consequences of the evil, which persons known to you, namely Ia[strem]sky & Co., had inflicted on our statistics, were insurmountable.

I myself (1998, p. 533) quoted from the telling Preface to that source as translated by Chuprov's student Anderson. Kluikin also noted that Slutsky had been seriously considering theological issues (which he could have only discussed with his relatives and closest friends).

I

Théorie de la corrélation et Traité abrégé de courbes de fréquence Manuel pour servir à l'étude de quelques méthodes principales de la statistique moderne

**Annales de l'Institut Commercial de Kiew
vol. 16, 1912, 208pp.**

The entire book (whose title was also provided in Russian) is translated
(Berlin, 2009; also at www.sheynin.de)
Here, only the introductory sections are included

Annotation

This is a translation of Slutsky's contribution of 1912 which was intended for Russian readers. He described the Pearson's theory of correlation drawing on the pertinent work of that founder of biometry and on many other British authors. At the time, Markov failed to appraise it properly although Chuprov had at once realized its value (and a few years later compiled a very positive reference for Slutsky), and even in 1948 Kolmogorov called it "important and interesting".

Contents

Foreword by Translator

0. Introduction

Part 1. Elements of the Doctrine of Curves of Distribution

1. General notion of curves of distribution or frequency curves
2. The moments of distribution
3. The mean deviation and the coefficient of variation
4. Probable errors
5. The Gaussian law and its generalization by Pearson
6. Justification of the method of moments
7. Determining the empirical moments
8. Deriving parabolic curves fitting experimental data
9. The normal frequency curve (the Gaussian curve). Deviations from the normal type
10. Calculating the coefficients of the Pearson curves

Notes

Part 2. Theory of Correlation

Chapter 1. Correlation between Two Magnitudes

11. The notion of correlation dependence
12. The correlation table
13. The regression lines
14. Examples
15. The correlation coefficient
16. Formulas for the regression coefficients and the correlation coefficient
17. Other formulas for the correlation coefficient
18. The mean square error of the regression coefficient
19. The straight lines of regression
20. Calculating a correlation coefficient, an example
21. The general population and the random sample
22. Probable errors and coefficients of correlation between constants for the normal distribution
23. The probable error of the difference
24. Probable errors in case of a normal distribution
25. The difference method of determining the correlation coefficient

26. Curvilinear regression
27. Calculating the coefficients of the regression curve
28. Calculating the coefficients of the regression curve, continued
29. Correlation ratio
30. Dependence between the correlation ratio η and the correlation coefficient r
31. Correlation and causal dependence
32. Methods of instantaneous average and successive differences

Chapter 2. Correlation between Three Or More Magnitudes

33. The main theorem of the theory of linear regression
34. The case of three variables
35. Examples
36. The partial correlation coefficients
37. The general case: correlation between n variables
38. The case of four variables
39. Normal correlation. The equation of distribution
40. The main properties of the normal distribution function. The Edgeworth theorem
41. On the probability of a system of deviations correlatively connected with each other
42. A test for conformity of a theoretical with an experimental distribution
43. A test for conformity of theoretical with an empirical regression line

Notes

Additional Remarks

1. On terminology
2. On the method of moments

Tables

Bibliography

Foreword by Translator

1. Slutsky: life and work
2. The book on the theory of correlation
3. Foreword to Slutsky (1960)

1. Slutsky: Life and Work

1.1. General information. Evgeny Evgenievich Slutsky (1880 – 1948) was an economist, statistician and mathematician, in that chronological order. His life and work are described in Kolmogorov (1948), Smirnov (1948), Chetverikov (1959), Allen (1950), Sheynin (1999), Seneta (2001), with pertinent archival and newspaper sources quoted in Sheynin (1990). Slutsky himself (1938 and 1942, published 1999) compiled his biography. In two other unpublished pieces Wittich (2004; 2007) provides valuable data on Slutsky's life and a pertinent annotated bibliography. In another unpublished paper Rauscher & Wittich (2007) collected information about Slutsky the poet and connoisseur of literature, a side of his personality (as well as his being an artist) that remains unknown. Kolmogorov (1948/2002, p. 72) called Slutsky “a refined and witty conversationalist, a connoisseur of literature, a poet and an artist”.

Slutsky's works include his student diploma (1910), the book of 1912 translated below, a paper (1914) which directly bears on a subject discussed in that book, and a most important economic contribution (1915), see also Chipman & Lenfant (2002) and Chipman (2004). His *Selected Works* (1960) contains his biography written by B. V. Gnedenko and an almost complete list of his works. In my § 3 below, I translate its Foreword.

In 1899, Slutsky enrolled in the mathematical department of Kiev university, was drafted into army with others for participating in the students' protest movement; released after nationwide shock; expelled in 1902 for similar activities and banned from entering any other academic

institution. In 1902 – 1905 studied mechanical engineering at Munich Polytechnic School; obviously gained further knowledge in mathematics and physics, but remained disinclined to engineering. In 1905 was able to resume learning in Russia, graduated with a gold medal from the Law faculty of Kiev University (end of 1910). His book of 1912 ensured him a position at Kiev Commercial Institute. Became professor at a successor organisation of that institute but had to move to Moscow because of an official demand that teaching ought to be in the Ukrainian language.

Worked as consultant (a very high position) at the Conjuncture Institute and Central Statistical Directorate. Owing to the beginning of the Stalinist regime with horrible situation in statistics (Sheynin 1998), abandoned these occupations, turned to the applications of statistics in geophysics. Did not find suitable conditions for research, became engaged in mathematics. Worked at Moscow State University, received there the degree of Doctor of Physical and Mathematical Sciences *honoris causa* and (Slutsky 1942/2005, p. 145)

was entrusted with the chair of theory of probability and mathematical statistics. [...] However, soon afterwards I convinced myself that that stage of life came to me too late, that I shall not experience the good fortune of having pupils. My transfer to the Steklov Mathematical Institute also created external conditions for my total concentration on research [...]

Until the end of his life Slutsky had been working at that Institute of the Academy of Sciences, became eminent as cofounder of the theory of stationary processes, died of lung cancer. Was happily married, but had no children. From 1912 to Chuprov's death in 1926 maintained most cordial relations with him.

A special remark is due to Allen (1950, pp. 213 – 214):

For a very long time before his death Slutsky remained almost inaccessible to economists and statisticians outside Russia. [...] His assistance, or at least personal contacts with him would have been invaluable.

Slutsky compiled his book in a very short time; in a letter to Markov of 1912 he (Sheynin 1990/1996, p. 45) explained that he had “experienced a direct impetus from Leontovich's book [1909 – 1911] [...] as well as from information reaching me [...]”. So had he meant 1909 or 1911? He was more specific elsewhere (Slutsky 1942/2005, p. 142): “In 1911, I became interested in mathematical statistics, and, more precisely, in its then new direction headed by Karl Pearson”.

Slutsky possibly read some statistics at the Law faculty, but hardly much; he did not mention anything of the sort in his published works. So it seems that in about a year, all by himself, he mastered statistics and reached the level of a respected author!

1.2. A special publication: Slutsky's correspondence with Bortkiewicz, 1923 – 1926 (Wittich et al 2007). I describe some of Slutsky's letters.

Letter No. 3, 25.9.1923. Slutsky made 3000 statistical trials to study whether equally probable combinations occurred independently from the size and form of bean seeds, cf. § 42 of his translated book. He never heard

that automatic registering devices were applied in such experiments and even invented something of the sorts “out of boredom”.

Letter No. 7, 16.5.1926. Slutsky had to move to Moscow because of “some discord with the Ukrainian language”, cf. § 1.1 above, most warmly mentioned the deceased Chuprov. He works as a consultant at the Conjunction Institute “together with Chetverikov” (Chuprov’s closest student and follower) and “had to become” consultant also at Gosplan (State Planning Committee), an extremely important and influential Soviet institution. I venture to suppose that the situation there also became difficult and real scientific work was even considered subversive. Anyway, nothing is known about Slutsky’s work there so that he apparently soon quit it.

Letter No. 10, 14.6.1926. Slutsky discussed his paper of 1915 and stated

I would have now ended it in an essentially different manner. For uniqueness (to an additive constant) of the definition of the function of utility it is not necessary to demand that on each hypersurface of indifference there exists a pair of such benefits that

$$\frac{\partial^2 U(x_1, x_2, \dots, x_n)}{\partial x_i \partial x_j} = 0.$$

It is sufficient to be able to draw a line cutting a number of such hypersurfaces along which the marginal utility remains constant, and this is in principle always possible. This result can also be obtained by elementary considerations.

Then Slutsky refers to his not yet published paper (1927); see also Chipman (2004).

2. The Book on the Theory of Correlation

2.1. Opinions about it. The book was published, as stated on its title-page, in the *Izvestia (Annales)* of the Kiev Commercial Institute, and, as mentioned by several authors, appeared independently later the same year. Sections 25, 28 and 43 (these numbers conform to those adopted in the translation) contained “additions to the Pearson theories”, see Slutsky’s letter to Markov of 1912 (Sheynin 1990/1996, pp. 45 – 46). As mentioned out of place in a footnote to its Introduction, Slutsky reported on his work to the Kiev Society of Economists. Those “Pearson theories” are what the whole book is about, and it is hardly out of order to mention my paper (2010) on that scientist.

2.1.1. Chuprov. He (Sheynin 1990/1996, p. 44) published a review of Slutsky’s book stating that its author “gained a good understanding of the vast English literature” and described it “intelligently”. He “most energetically” recommended the book to those having at least “some knowledge of higher mathematics”. At the time, Chuprov was not yet critically inclined towards the Biometric school; he changed his attitude later, no doubt having been turned in the mathematical direction by his correspondence with Markov (Ondar 1977).

Apparently in 1916, Chuprov (Sheynin 1990/1996, p. 45) compiled Slutsky’s scientific character which contained a phrase: in Slutsky’s person

“Russian science possesses a serious force”, but he obviously did not imagine how correctly he assessed his new friend!

There also (p. 29) I published an archival letter written by N. S. Chetverikov to Chuprov at the end of 1926. He most favourably described the situation at the Conjecture Institute (where he himself held a high position) and informed his correspondent, already terminally ill, that Kondratiev was inviting him to join their staff. He added, however, that the general situation in the Soviet Union was unclear.

2.1.2. Pearson. He rejected both manuscripts submitted by Slutsky (Sheynin 1990/1996, pp. 46 – 47). In 1913, Slutsky wrote Chuprov about that fact and asked his advice stating that at least in one instance the reason for the rejection “astonished” him. Chuprov did fulfil Slutsky’s request and, accordingly, Slutsky successfully published one of his manuscripts (1914). I (Sheynin 2004, pp. 227 – 235, not contained in the original Russian paper) made public three of Slutsky’s letters to Pearson of 1912.

2.1.3. Markov. Continental mathematicians and statisticians, and especially Markov utterly disapproved of the Biometric school and I myself have described vivid pertinent episodes (Sheynin 1990/1996, pp. 120 – 122; 2007). In his letters to Chuprov Markov (Ondar 1977/1981, letters 45 and 47, pp. 53 and 58) remarked that Slutsky’s book (no doubt partly because of that general attitude) “interested” him, but did not “attract” him, and he did not “like it very much”.

More can be added. A few years later, Markov (1916/1951, p. 533, translation p. 212) critically mentioned the correlation theory: it “simply” [?] aims to discover linear [?] dependences, and, when estimating the appropriate probable errors, “enters the region of fantasy [...]”. This statement was based on an unfortunate application of that theory by a Russian author, but Linnik (Markov 1951, p. 670; translation, p. 215), who commented on Markov’s memoir, explained that the conclusions of the correlation theory depended on the knowledge of the appropriate general population. Slutsky, in 1912, did several times mention the general population (also see below) but certainly not on the level of mid-20th century. However, Markov could have well noted Slutsky’s conclusion (§ 22) to the effect that the correlation method should not be applied when observations are scarce (which was the case discussed by Markov).

Markov’s attitude shows him as a mathematician unwilling to recognize the new approaches to statistics and even to the theory of probability (and denying any optimal properties of the method of least squares), see Sheynin (2006). Markov had time to prepare the last edition of his treatise that appeared posthumously (1924). There, he somewhat softened his views towards the correlation theory and even included Slutsky’s book in a short list of references to one of its chapters.

Upon reading Slutsky’s book Markov asked Grave, a professor at Kiev university, about the new author. Dmitry Aleksandrovich Grave (1863 – 1939) was active in many branches of mathematics and he also published a treatise on insurance mathematics (in the same volume of the Kiev Commercial Institute *Izvestia* as Slutsky). In a letter to Markov of 1912 Grave (Sheynin 1999/2004, p. 225) informed his correspondent that neither he himself, nor the lawyers, professors at that Institute, had understood Slutsky’s report (see § 2.1 above), that they desired to acquaint themselves

with the Pearson theories and asked him to explicate it properly. Grave, however, finds it “repulsive” to read Pearson.

Grave also told Markov about his conversation with an unnamed university professor of political economy who had explained that Slutsky was “quite a talented and serious scientist” not chosen to study as postgraduate “because of his distinct sympathy with social-democratic theories”.

2.1.3. Slutsky explained himself in an apparently single extant letter to Markov of 1912 (Sheynin 1990/1996, p. 45 – 46). Improvements of his manuscript “were hindered by various personal circumstances” and he “decided to restrict myself [himself] to a simple concise description” the more so since it will help those Russian statisticians who are unable to read the original literature. He then prophetically stated that “the shortcomings of Pearson’s exposition are temporary” and that his theories will be later based on a “rigorous basis” as it happened with mathematics of the 18th and 19th centuries. He added a most interesting phrase: “I consider it possible to develop all the Pearsonian theories by issuing from rigorous abstract assumptions”.

Slutsky also mentioned Nekrasov: when his book (1912) had appeared, he began to think that

My [his] work was superfluous; however, after acquainting myself [himself] more closely with Nekrasov’s exposition, I [Slutsky] became convinced that he [Nekrasov] did not even study the relevant literature sufficiently.

In § 31 (Note 31.1) Slutsky praised the same book; perhaps he did not yet read it “more closely”: after ca. 1900, Nekrasov’s contributions on the theory of probability and statistics became almost worthless (and utterly disgusted Markov), see Sheynin (2003).

In a letter to Chuprov of the same year Slutsky (Sheynin 1990/1996, p. 44) noted that Grave “actively participates” in the dispute (between Markov and him) and added that Markov “gave me [him] a good dressing-down”. [...] It was easy for Markov “to discover a number of weak points”.

2.1.4. Kolmogorov (1948/2002) published Slutsky’s obituary which clearly shows his personal ties with the deceased. He (p. 68) stated that the book of 1912 “became a considerable independent contribution to [mathematical statistics and] remains important and interesting”. On the same page Kolmogorov listed “the main weakness[es] of the Biometric school:

Rigorous results on the proximity of empirical sample characteristics to the theoretical ones existed only for independent trials.

Notions of the logical structure of the theory of probability, which underlies all the methods of mathematical statistics, remained at the level of the 18th century results.

The third and last weakness concerned the incompleteness of the published statistical tables.

Kolmogorov indirect advice of applying Slutsky's book at least as a background was not, however, followed; even Slutsky's examples of statistically studying various problems had hardly ever been cited.

2.1.5. Some general remarks about the book. Information provided above, at the end of § 1.1, explains why Slutsky was unable to add a few pages about Pearson, his followers (and Galton!), or to be at least somewhat more critical. He certainly understood that the work of that great scientist was far from rigorous (see § 2.1.3 above), but on this point he only expressed himself about the method of moments (*Additional remarks*). Slutsky also felt that statistics ought to be based on the theory of probability; he said as much, although not quite generally, at the end of his § 32, and stated, in a letter to Markov (§ 2.1.3 above), that that approach was achievable.

On the other hand, the reader will not fail to note that Slutsky also became quite familiar with the practical side of statistics; his book abounds with pertinent remarks! And he also properly provided a lot of original examples of applying correlation theory.

Slutsky (the end of § 2.1.3 above) acknowledged that Markov had "discovered a number of weak points" in his book. For my part, I believe that he had succeeded by and large to provide a good general picture of his subject, but I ought to say the following.

1. He made a mistake in his reasoning on weighing observations, see my Note 28.1, in § 28 which contained his "additions to the Pearson theories", see § 2.1 above. I mentioned another mistake in Note 16.1.

2. His explanations were sometimes inadequate or even lacking, see Notes 3.1, 4.3, 16.2, 40.1 and 41.2.

3. An author ought to show readers not only the trees, but the wood as well, and I especially note that Slutsky had not stated expressly and simply that a zero correlation coefficient does not yet signify independence. His explanation (beginning of both §§ 19 and 29) is not quite sufficient, and in § 31 he only discusses correlation and causality.

4. He offered a faulty example (Note 31.3).

5. He introduced confusing notation (Note 18.5).

Slutsky's system of numbering the sections and formulas was not the best possible. Now, in the translation, sections are numbered consecutively (not separately for each part), and the numbering of the formulas allows to locate them quite easily; thus, formula (3.2) is the second numbered formula in § 3. The Notes (by Slutsky, signed E. S., and my own, signed O. S.) are numbered the same way.

I have omitted some pieces of the original text such as elementary explanations (even concerning the calculation of determinants), mathematical derivations and tables of data which after all can be looked up in the English literature described by Slutsky. Then, I have not included the numerous figures and, accordingly, had to modify their accompanying description.

Acknowledgements. Magister Guido Rauscher sent me his joint unpublished material (Rauscher & Wittich 2006) and photostat copies of Slutsky (1938; 1942), of the Contents of Slutsky (1910) and of the entire book translated below. From Dr Claus Wittich I received his unpublished contributions (2004; 2007).

Bibliography

- Allen, R. G. D.** (1950), The work of Eugen Slutsky. *Econometrica*, vol. 18, pp. 209 – 216.
- Chetverikov, N. S.** (1959, in Russian), The life and work of E. E. Slutsky. Translation: in the present collection [xix].
- Chipman, J. S.** (2004), Slutsky's praxeology and his critique of Böhm-Bawerk. *Structural Change and Economic Dynamics*, vol. 15, pp. 345 – 356.
- Chipman, J. S. & Lenfant, J.-S.** (2002), Slutsky's 1915 article: How it came to be found and interpreted. *Hist. Polit. Economy*, vol. 34, No. 3, pp. 553 – 597.
- Eliseeva, I. I., Volkov, A. G.** (1999), Life and work of E. E. Slutsky. *Izvestia Sankt-Peterburgsk. Universitet Ekonomiki i Finansov*, No. 1, pp. 113 – 121. In Russian.
- Kolmogorov, A. N.** (1948, in Russian), Obituary: E. E. Slutsky. *Math. Scientist*, vol. 27, 2002, pp. 67 – 74.
- Markov, A. A.** (1900), *Ischislenie Veroiatnostei* (Calculus of Probabilities). Fourth edition: Moscow, 1924.
- (1916), On the coefficient of dispersion. In Markov (1951). Translated in Sheynin, O. (2004), *Probability and Statistics. Russian Papers*. Berlin, pp. 206 – 215. Also at www.sheynin.de
- (1951), *Izbrannye Trudy* (Sel. Works). Moscow.
- Ondar, Kh. O.**, Editor (1977, in Russian), *Correspondence between Markov and Chuprov on the Theory of Probability and Mathematical Statistics*. New York, 1981. Russian original contained 90 significant mistakes, most of them necessarily retained in the translation (Sheynin 1990/1996, pp. 79 – 83).
- Rauscher, G. & Wittich, C.**, Editors (2006), *E. E. Slutsky's Papers, Fond 21333, Russian State Archive of Literature and Art* [Moscow]. Unpublished.
- Seneta, E.** (1992), On the history of the strong law of large numbers and Boole's inequality. *Hist. Mathematica*, vol. 19, pp. 24 – 39.
- (2001), E. E. Slutsky. In *Statisticians of the Centuries*. New York, pp. 343 – 346.
- Sheynin, O.** (1990, in Russian), *Aleksandr A. Chuprov. Life, Work, Correspondence*. Göttingen, 1996.
- (1998), Statistics in the Soviet epoch. *Jahrbücher f. Nationalökonomie u. Statistik*, Bd. 217, pp. 529 – 549.
- (1999, in Russian), Slutsky: commemorating the 50th anniversary of his death. Translation in present collection [xviii].
- (2003), Nekrasov's work on the central limit theorem. *Arch. Hist. Ex. Sci.*, vol. 57, pp. 337 – 353.
- (2006), Markov's work on the treatment of observations. *Hist. Scientiarum*, vol. 16, pp. 80 – 95.
- (2007), Integrity is just as important as scientific merits. *Intern. Z. f. Geschichte u. Ethik d. Naturwissenschaften, Technik u. Medizin*, Bd. 15, pp. 289 – 294.
- (2010), Karl Pearson a century and a half after his birth. *Math. Scientist*, vol. 35, pp. 1 – 9.
- Slutsky, E. E.** (1910), *Teoria Predelnoi Poleznosti* (Theory of Marginal Utility). Kiev. Diploma thesis. Vernadsky Ukrainian Nat. Library. Ukrainian transl.: Kiev, 2006.
- (1914), On the criterion of goodness of fit of the regression lines and on the best method of fitting them to the data. *J. Roy. Stat. Soc.*, vol. 77, pp. 78 – 84.
- (1915, in Italian), On the theory of the budget of the consumer. In *Readings in Price Theory*. G. J. Stigler, K. E. Boulding, editors. Homewood. Ill., 1952, pp. 27 – 56.
- (1916, in Russian), Statistics and mathematics. Review of Kaufman, A. A. (1916), *Teoria i Metody Statistiki* (Theory and Methods of Statistics). Moscow. Third edition. *Statistichesky Vestnik*, No. 3 – 4, pp. 104 – 120. Translation in present collection [ii].
- (1927, in German), A critique of Böhm-Bawerk's concept of value and his theory of the measurability of value. *Structural Change and Economic Dynamics*, vol. 15, 2004, pp. 357 – 369.
- (1938, published 1999, in Russian), Autobiography. Translation in present collection [xvi].
- (1942, published 1999, in Russian), Autobiography. Translation in present collection [xvii].

--- (1960), *Izbrannye Trudy* (Sel. Works). Moscow. Economic publications not included. Omission without indication of (mainly foreign) references in footnotes, distortions in translations (comment by Wittich 2007).

Smirnov, N. V. (1948), E. E. Slutsky. *Izvestia Akad. Nauk SSSR*, ser. math., vol. 12, pp. 417 – 420.

Wittich, C. (2004), *Biographical Notes on E. E. Slutsky*. Unpublished. Put at my disposal by author.

--- (2007), *Bibliographical Notes on Selected Sources concerning E. E. Slutsky*. Unpublished. Put at my disposal by author.

Wittich, C., Rauscher, G., Sheynin, O. B., Editors (2007, in Russian), Correspondence between E. E. Slutsky and V. I. Bortkevich. *Finansy i Biznes*, No. 4, pp. 139 – 154. Authors ordered in accord with Russian alphabet. We had not seen the proofs and the paper contains a number of misprints; in particular, four of my papers are attributed to Chipman. Translation in present collection: [xv].

II

Statistics and mathematics. Review of Kaufman (1916)

Statistika i matematika.

Statistichesky Vestnik, No. 3 – 4, 1915 – 1916, pp. 104 – 120

[1] Kaufman's treatise, now in its third edition, is certainly an outstanding phenomenon in our educational statistical literature, and not only in our as testified by the reviews of its German edition (1913) written by the most notable representatives of the European statistical thought¹. This third edition will also obviously find many friendly readers the more so since in its main parts and especially in its first theoretical part it is entirely recast as compared to 1912.

However, those who attentively followed the evolution of Kaufman's work will not fail to note that at least in one respect this third edition is not a simple development of the previous one but as though some new stage in the author's statistical Weltanschauung. Indeed, the author intended both the second and the third editions as a manual for those wishing to prepare themselves for working in statistics but lacking that mathematical background necessary for *entirely* mastering statistical theory and methods.

The author (1912, p. 235) believed (and believes) that

It is hardly possible to master consciously the principles of the statistical theory [...] without [its] connection with the main principles of the theory of probability.

He therefore devoted sufficient efforts and place to provide his readers with a possibly more distinct idea about both the theory of probability and its application for solving fundamental issues of the theory of statistics. As to the practical application of the formulas and tricks of the higher statistical analysis, the author (p. 236) properly and tactfully warned those insufficiently prepared:

Thoroughly perceive the boundaries of your competence. [...] In particular, certainly abstain from mechanically applying final formulas provided by mathematical statistics without being quite clearly aware of their intrinsic meaning and sense, otherwise misunderstanding can often result.

For consoling his readers he (p. 235) stated that he was sure that “In its current state, statistics still leaves for them an infinitely broad area of activity”. True, he (p. 234) apparently did not entirely get rid of his serious doubts about the issue of the interrelations between statistics and mathematics and while acknowledging that “It is *hardly possible to resolve* the difference of opinion” among the representatives of the statistical theory (my italics), he even avoided any attempt to clear up this matter in his manual. Given these circumstances, his practical way out, as mentioned above, to which he became inclined, could have only been welcomed. This is all the more so since the general outline of his introduction to the theory of statistics and a number of other instances (see, for example, his very indicative remarks on pp. 131 and 132) allow to think that it were considerations about the difficulties of mathematical methods rather than doubts about the principles themselves that compelled the author to hesitate.

[2] That practical dualism is not at all specifically peculiar to statistics and is observed in other sciences and reflects the distinction between the individual features of the researchers and the subjects of their work (theoretical and practical astronomy, theoretical and experimental physics etc). However, in this third edition it became transformed into a dualism different in principle, the dualism between statistical theory and practice (p. 148):

As a rule, because of the very properties of this [statistical – E. S.] material statistical analysis does not allow, and because the structure based on that data is coarse and at the same time complicated, does not demand the application of formulas of the calculus of probability. However, this does not at all contradict the fact that each such structure is entirely based on the principle of probabilities.

But the author (p. 153) also keeps to his previous divide between statisticians who “follow and will follow the routes demanding application of more or less complicated forms of mathematical analysis” and others who “while treating [...] statistical material and interpreting its results, may restrict their efforts to elementary methods of calculation”.

This motif now seems rather inconsistent with the previous. Indeed, how is it possible to reconcile the right of a purely practical distinction only founded in essence on the division of labour between the researchers and the abovementioned standpoint negative in principle, or the author’s statements (p. 152) that such procedures as the construction of frequencies of distribution, adjustment of series etc “not only do not help to elucidate the real features of the studied phenomena, but, on the contrary, can provide ideas corrupting reality” and that “the method of correlation does not add anything essential to the results of elementary analysis”?

Choose one or the other: either these procedures and methods are useless and therefore harmful and ought to be altogether abandoned; or, they are useful, but demand an understanding of their essence, meaning and boundaries of application which is at least partly possible even in a treatise intended for readers lacking sufficient mathematical background.

The dualism of the author’s point of view which is not objectively resolved in those texts becomes nevertheless somewhat explained after reading that (p. 147)

The issue of our right to apply [in the area of general statistics – E. S.] the methods of the calculus of probability is in any case left open, or, as he adds, open for me.

Objectively speaking, this pronouncement certainly only confuses the matter since the reader remains ignorant of the basis on which, as the author believes, his own arguments against applying the calculus of probability are weakened and he is led to adopt it in the practical sense on the one hand and to candidly ignore it on the other.

And still, if I am allowed to express my general feeling, I ought to say that the main and the specific for the author is apparently at present not this previous hesitation and doubts which reflected the former stage of his scientific evolution, but the formed and almost firmly established conviction in that the statistical analysis does not either allow or demand probability theory.

I have thus returned to the quote from p. 148 with which I started to describe the present viewpoint of Professor Kaufman and I think that after all stated above I am compelled at least conjecturally to adopt it as the expression of the real opinion of the author, to assume it as the starting point and main object of my critical remarks below.

However, I have to begin elsewhere. Indeed, I am sure that the indicated dualism between statistical theory and practice is rooted much deeper, i. e., not in the author's understanding of the role of probability theory, but in his ideas about the essence of statistics, and that issue is not yet clarified in contemporary literature in any sufficient measure.

[3] Kaufman adheres here to the now apparently dominant point of view that statistics is a method or methodological doctrine and not at all a science with its own special subject of research. And I personally would have been prepared to adjoin somewhat the critical aspect of his considerations, provided he had sharpened his reasoning to allow for Chuprov's view whose idea of statistics as an ideographic science² he does not regrettably even mention in spite of its certainly being the most powerful argument possessed by the camp which Kaufman criticizes. True, I think that even that argument cannot be upheld, but Kaufman did not prove that. I will not dwell on this difficult point because of lack of space and the more so since here I am not really at any variance with Kaufman.

Distinctions between us start further, and exactly where Kaufman believes to have concluded the issue, where he recognizes the methodological essence of statistics. Let us ask ourselves is it in essence really indifferent, as he (p. 17) thinks, whether "to discuss statistics as a *supplementary science*, or simply as a *methodological doctrine*". When allowing for the author's considerations, we, as it seems to me, ought to conclude, first of all, that he does not sufficiently clearly distinguish between the various versions of the term *statistics* (see his pp. 15 – 18) and does not follow up to conclusion the reasoning on the place of the statistical method in the system of logical knowledge.

I begin with the issue of the method itself. As a method, statistics is certainly not a science, but a technique, that is, a system not of reasoning, but of tricks, rules and patterns of practical cognizing work, whether applied systematically or not, conscientiously or unconscientiously, for scientific or

practical goals. Just the same as addition and subtraction remain arithmetical operations independently from who is applying them and what for. This will become quite clear after analysing the contents of statistics as a methodological *doctrine*. We will find there, in particular, a number of propositions concerning even the most simple procedure of the statistical technique, enumeration of the elements of a totality and its necessary conditions and forms.

The methodology of enumeration based on the analysis of its very nature allows us to see how practice is conditioned by the general properties of totalities on the one hand, and the properties of known logical operations on the other. To oppose, as Kaufman (pp. 13 – 14) does, statistical method and statistical art by issuing from indications external in regard to operations themselves, i. e., from the aims of the work, is in essence wrong even without allowing for the difficulty of drawing the necessary boundary which he mentions.

Whether the enumeration of social masses, say, is applied practically (e. g., for the aims of administration as *statistical art* according to Kaufman) or for knowledge (*statistical method*), is of no consequence. Not technically, as the author believes, but according to its essence the nature of the operation will be the same as will be the conditions for it to be properly done; consequently, the corresponding reasoning belonging to statistics as a methodological doctrine will also be the same in both cases.

And the last inference: since this reasoning does not change with the aims of the operation, it follows that the location of the boundary depends not on Kaufman's decision or otherwise, but on the essence of the matter. And, incidentally, this means that a discussion of the issues concerning the contents of a science is not idle, is not to be decided by opportunistic considerations of expediency; no, it is important and, if properly formulated, fosters the deepening and the solution of the most general problems of science.

[4] I think that it is just as impossible to agree with Kaufman's arguments about the nature of statistics as a methodological science. As a system of considerations, statistics, understood in that sense, is necessarily either a science or its part. Kaufman (p. 16) compares it with the doctrine of measuring devices which allegedly cannot be isolated as a special science. However, if that doctrine is not a *special* science, it is a part of another one, – of which, it ought to be asked, of the science which it provides with the means of research, or of that on which it is logically based?

Both alternatives fall away almost at once; the former, because measuring devices such as clocks and microscopes serve all or many sciences and purely practical needs as well, and the latter, since a complete theory of one and the same device as of an ideographic item demands the application of many sciences, such as mathematics, mechanics, physics, chemistry, psychology (recall the personal equation in astronomy) rather than one, etc.

After thinking it over, the entire issue of attributing the theory of measuring devices to a certain *theoretical* science becomes absolutely mistaken because the peculiar logical structure of such a doctrine is overlooked here. The considerations constituting that doctrine are united into a system in a manner absolutely different from that, applied in any theoretical science. Here, the systematic connection is conditioned not by the objective relations between things and their various aspects, but by their

teleological function with these things being seen as the means for attaining the aims of the researcher.

Hence the natural grouping of separate technical disciplines according to the pursued goals intersected with their partition for the sake of achieving maximal possible homogeneity of the contents according to the essence of the underlying theoretical doctrines.

And it is now also understandable why, in the process of teaching and elucidating, some technical disciplines are more closely adjoined to those sciences from which they derive their theoretical elements (for example, the doctrine of physical measurements) whereas others are in the neighbourhood of those sciences which make use of their results (e. g., the doctrine of devices for psychological measurements). Finally, still other disciplines in addition possess external independence (metallurgy or the doctrine of fibrous substances). All this, however, is an issue of teaching and elucidation and has no direct bearing on the logical essence of the relevant doctrines.

[5] These considerations justify the independence of statistics as a technical or practical science which according to some tests admits in addition of separation into statistical methodology and statistical technique and at the same time leads us really earnestly to the problem of statistics as an independent theoretical science. Actually, any practical doctrine, as Husserl (1900 – 1901) had discovered in an inimitable masterly way, certainly assumes some underlying theoretical doctrine justifying its propositions. Indeed, for proving the possibility of some goal by definite means we ought to perceive the connection between means and goal as between cause and effect. And the study of such connections leads us to a totality of considerations constituting a system whose main point is the essence and properties of the subject rather than of the goals.

We thus arrive at an analysis of the theoretical considerations on which statistical methodology is built. Isolating that which relates to the properties of, first, judgements and concepts, i. e., to logic, and then of the properties of quantitative images upon which it [logic]³ is operating, i. e., of mathematics, we nevertheless obtain some remainder for which no acknowledged sanctuary is in existence, which remains uncoordinated and homeless until we perceive its special theoretical essence and provide it with the missing unity in the system of judgements fully deserving the name of theoretical statistics.

All the existing various propositions of the doctrine of totalities and their general properties only provisionally adjoining methodological problems will belong here. We will thus have, first of all, the doctrine of the main formal properties of totalities; then, of their quantitative and structural forms (which now constitutes an essential part of the so-called *Kollektivmasslehre*, that is, of the doctrine of frequencies and surfaces of distribution, of means etc); then, also included will be a generalized formal doctrine of population, or, more correctly, of totalities of a changing composition, whose elements emerge, change their state and disappear, be they [individuals of a] population, trees in a forest or atoms⁴.

Finally, here also belongs the doctrine of the machinery of causes determining the frequency of phenomena rather than of separate events. All this is not a logical doctrine of the world of judgement and concepts, but

statistical doctrines of the world of phenomena in [the entirety of] their forms and mutual conditionality.

Whether to separate them as a special subject for elucidation and teaching, certainly depends on our arbitrary opinion, but a special science emerges not by arbitrariness but [because of the existence of] intrinsic ties [of the appropriate components], cognized as something objectively compelling, as establishing a systematic likeness and unity of the corresponding relations as well as of the considerations expressing our knowledge of their properties and ties [between them].

[6] And now we approach the issue of the relation of the calculus of probability to statistics. It only suffices to compare the contents of some purely mathematical treatise on the former with statistical reasoning on probability; for example, the contribution of Markov with the writings of von Kries [1886] or Chuprov, and the deep intrinsic heterogeneity of the problems, methods and of the very spirit of these writings becomes striking. And further considerations will show that these distinctions are based on the difference between the subjects.

Calculus of probability is a purely mathematical science⁵. How something is occurring is of no consequence to it; it deals not with *factual* but *possible* frequencies, not with their real causes but their *possible* probabilities. And the concept of probability itself is there quite different, is generalized and abstract. As soon as some number is arbitrarily assigned as the weight of each possible event and a number of definitions is made use of, the basis is prepared for building in a purely abstract way infinitely many purely abstract castles of combinations in the air, and of going over from those weights to the weights of various derivative possibilities (for example, of some groups of repeated occurrences of events).

For the *calculus* of probability, any enrichment of the concept of probability as compared with the above is useless, it would have nothing to do with it. Throwing a bridge from that ethereal atmosphere of mathematical speculations to the region of real events is only possible by abandoning the ground of the *calculus* of probability and entering the route of studying the real world with its machinery of cause and effect. Only thus we obtain knowledge about the ties between frequency and probability, justify [the assumptions of] the law of large numbers and find the basis for applying the calculus of probability to studies of reality.

Chuprov investigates free causal connections⁶; von Kries discovers the causal underpinning of games of chance and the actual justification for the tendency of frequencies to coincide with probabilities; Venn and Edgeworth attempt to build the very notion of probability on the concept of frequency⁷, – but nothing mentioned has any relation to the mathematical science of the calculus of probability. Here [in statistics], the mind operates not with ideal forms and quantities but with real things and phenomena although considered from an extremely general viewpoint⁸.

Above, I did not add anything to the essence of the doctrines of theoretical statistics, I had not even demanded the creation of such a science (always a somewhat dangerous enterprise) and only mentioned a number of existing doctrines and their intrinsic ties [with each other?]. If, however, it occurred that these doctrines constitute the main theoretical contents of statistical methodology, then I will hardly be mistaken when stating that

statistics as a theoretical science does exist, that collective items, totalities considered as such, to whichever area they belong, are its subject.

Incidentally, it also follows all by itself that since statistics studies quantitative properties which we cannot ignore because of their part in the relations and ties peculiar to the subject of statistics, statistics should be indeed considered a mathematical science, i. e., one of those sciences in which mathematical methods are essential and unavoidable⁹.

Then, it is natural that also in practical applications of theoretical statistics and statistical methodology, that is, in the practice of concrete statistical work on empirical data, mathematical methods are also unavoidable, and that there exists no essential boundary between various chapters of statistical practice in regard to the *subject* of study. It is only possible to distinguish in each chapter more elementary and more complicated problems, and, in general, problems of one and another theoretical type.

[7] The study of Kaufman's viewpoint only corroborates, as I believe, these considerations. Indeed, after formulating his essential objections to the application of the theory of probability to statistics he finally arrives at a conclusion whose considerable significance I ought to deny decisively. He (p. 147) assigns as the area of such application the set of *simplest* phenomena of population statistics and a certain part of phenomena in natural sciences, then (true, somewhat hesitatingly, see above) refuses to agree that the calculus of probability is applicable to *general statistics*.

It is impossible to be satisfied by such a decision. Indeed, *general statistics* (an expression that the author himself writes in inverted commas) is obviously a heterogeneous group of problems lacking any intrinsic ties. And, if only the theory of probability can at all be applicable to analysing reality, the necessary boundaries and conditions can depend not on the concrete properties of the totalities, but on their formal properties on the one hand and on the properties of the problems to be solved on the other. It is exactly in this direction that a manual of statistics ought to guide the beginners.

Kaufman (p. 147) expresses himself in the sense that for the areas mentioned (population statistics etc)

The existence of the prerequisites for the [application of] the theory of probability can be considered justified a posteriori, and the application of its methods here does not in principle excite objections anymore.

That the author is hardly in the right here can be already seen by the quotations from Markov and von Kries that he provides there. Indeed, even in the area where "the validity of applying the elements of the calculus of probability is *least* doubtful" (Kaufman, p. 145, his italics), the former denies the right of statisticians to justify *in principle* their practice (tables of mortality whose usefulness he does not deny) "by referring to the formulas of the calculus of probability".

As to the latter, since he is against the application of those formulas, his viewpoint concerns not one or another area of statistics (population or general statistics), i. e., not real objects but formally traced problems.

[8] Turning to the essence of the matter, inasmuch as it is possible in the boundaries of this paper, I am issuing from Markov's demand, that is, from the need to ascertain in each separate case whether the trials were

independent, the probability was invariable and [the appropriate cases] equally possible. Under such restrictions, all the applications of the calculus of probability to statistics are partitioned in two main groups: in one of them, the applicability is justified a posteriori by proving that those conditions are fulfilled at least approximately; and in the other one, applications are substantiated a priori.

In the first instance we are dealing with predicting probable frequencies of some phenomenon by known frequencies of other facts (urn experiments, heredity, insurance etc). In the second case we have to do with comparing reality with a theoretical pattern for which Markov's demands are postulated a priori. Simplest examples here again are experiments with urns, coins etc only considered in their different logical aspect.

Here, we compare the actual frequency with its value expected with one or another probability under the conditions of constancy, independence (or a definite dependence) and equal possibilities. And we do not act differently when studying the fluctuations of the sex ratio at birth or death etc. by the Lexian or any other similar method.

The same standpoint underlies the method that Kaufman discusses under a somewhat unfortunate name *differential*. Thus, when comparing for example the percentage of peasants lacking a horse of their own in two different localities so as to find out how significant is the difference, the real basis for the comparison is some imagined totality of individual farms, some imagined nation where the conditions determining the number of horseless peasants are assumed to be everywhere the same and the distribution of the farms over the territory is purely accidental¹⁰. And, issuing from that image, we calculate the probability that the difference mentioned could have been not less than in reality.

We are thus able to imagine at least the order of the probability of a correct judgement about whether the observed difference may be explained as being purely accidental, or whether we should assume as its basis either some detectable in principle causes or the insufficient accuracy of the data.

Kaufman correctly states that for such a conclusion it is not necessary to determine invariably the value of the appropriate probabilities, but he fails to notice that practitioners are infinitely many times guilty, also in our zemstvo statistics, of absolutely unfounded decisive inferences made from insufficient data. To oppose such arbitrary conclusions and to train systematically the feeling for the digits (so valued by the author), the calculation of probabilities or estimation of their order by determining mean square or probable errors and other measures of probable deviations ought to be practised incomparably oftener than it is done now.

Incidentally, it should be noted that Kaufman, when referring to von Kries for corroborating his views, hardly noted that he (p. 244) discussed that very method of applying the calculus of probability to statistics calling it *Untersuchende Methode*. He allowed its application for studying mass social phenomena even in case of large numbers (*in erhebliche Umfänge*). Kries very highly appreciates the investigations of Lexis and argues that they simply constitute a variety of that same method.

I believe that the contemporary statistical literature (above, I myself did not say anything essentially new)¹¹ sufficiently justified the application of the method under discussion to statistics and that, according to the train of thought leading to that substantiation, no partition whatsoever of statistics

into areas, as Kaufman attempts to accomplish, can hardly be supported by any perceptible logical foundation¹².

[9] Not less shaky is the author's understanding of the application of the patterns and formulas of the theory of probability to sampling. As von Knies rightfully remarked, its embryos in the conjectural statistics¹³ of political arithmeticians developed by mathematicians (by Laplace!) failed to be sufficiently justified by an exhausting criticism of the empirically applied methods of isolating the sample. The work of Kiaer [at the turn of the 19th century] which in a sense marked a new stage suffered from the same shortcoming.

But Kaufman himself (p. 98) demands sampling with mechanical selection, that is, as I understand him, with a purely accidental choice, and he admits that such a procedure "provides a full guarantee of typicality, of *representativeness* of the results of sampling". This, however, is indeed what is needed for a justified application of the calculus of probability to sampling. Yule and Bowley deal only with this method [of sampling] and I am unable to understand how Kaufman (p. 97) could have concluded that, according to the Bowley method, it was indifferent whether to snatch at random 100,000 individual farms of a province or to select as the sample the entire population of its two uyezds [districts], or of an entire longitudinal strip.

This statement, may the author excuse me, is a misunderstanding pure and simple. And when he comes to deny the importance of the sample size (contrary to what is already gaining the upper hand also in our practice) as opposed to its relative size I cannot but perceive here the results of the same misunderstanding¹⁴.

I will dwell, for example, on his (Ibidem) reproach of sampling for extinguishing those qualitative nuances, those varieties of phenomena which exist in real life when considered in large masses, and are exhibited ever more distinctly with their increase. This is of course true, but we must not overlook that, on the other hand, the more considerable is the mass, the simpler and more curtailed usually ought to be the programme [of its investigation] so that in most cases, on the contrary, only sampling can allow us to approach reality from such different sides and therefore to perceive it more or less fully and distinctly in all its variety, see for example Westergaard (1890, pp. 205, 207 and some other places).

Then, for recognizing any *nuance* a corresponding *absolute* size of the sample is needed so that, having formulated beforehand definite cognitive theoretical or practical goals, we will be able to determine the corresponding sample size. No flair will help here since it did not guarantee even such an experienced investigator as Kaufman¹⁵ against an entirely mistaken recognition of the decisive importance of the relative size of the sample. Only a systematic application of tests provided by the calculus of probability, if, certainly, the researcher possesses all the other qualities peculiar for a good worker, can ensure the success of sampling. A critical discussion of the experience already at hand could have indicated all this with indisputable clarity.

I ought to add that this issue is by no means academic only. Exactly for the practitioner the problem of establishing the number of elements to be described, invariably connected with financial considerations, often determines whether the investigation will take place or not.

[10] My paper has already become too lengthy and I cannot consider in detail separate propositions made by the author concerning particular issues. Although unable to agree with many of them, I must abandon his views on separate methods of mathematical statistics and will only point out some of his shortcomings. Thus, I am inclined to believe that the uninitiated will be put to difficulties by the three definitions of probability (by Laplace, Bortkiewicz and Vlasov [1909]) with which the author (p. 49) begins his exposition, and that such readers will understand them the less the more they will ponder over them. As it seems, this is especially true in regard to Laplace's definition provided out of the author's context.

Just the same (pp. 50 – 51), Czuber's definition is not understandable without a long explanation. I imagine that it would have been more advantageous to expound the principles of the theory of probability by examples with balls etc making use of the most elementary concept of probability as the ratio of the favourable cases to all of them and only to deepen this idea afterwards by indicating other possibilities. Here, however, the "logical foundation of the notion of probability" cannot be avoided since otherwise the reader will be confused by those various definitions rather than assimilate them. Moreover, the discussion of these issues (the viewpoints, say, of Venn, Cournot, von Kries, Chuprov) are much more important for understanding the beginnings of the theory of statistics than many other parts of the author's exposition, and, in addition, they are more readily understood.

Thus, I think that the derivation of the probability integral (pp. 76 – 78) could have been omitted since the reader will not be able to conclude it; it would have been better to explain instead the general train of thought leading to it and its significance and meaning. The author (p. 81) provided the appropriate approximate calculations, but it would have been better to choose an example allowing in addition to calculate the same probability in an elementary way by adding up probabilities of separate cases. For the beginner, this would mean much and it will also clearly indicate that the integral only provides approximations. Then, I think that the generalization of the law of large numbers (pp. 82 – 83) based on the [Bienaymé –] Chebyshev inequality can also be omitted, but, on the contrary, that it would have been apparently better to prove the Bayes theorem and to explain it in more detail. Indeed, it was the source of so many logical sins!

[11] And in general, it seems to me that for the goals attempted by Kaufman the volume of mathematics could have been lessened, but that the selected minimal information should have been worked out in rather more detail. Then, it would have been easier for the reader to learn how to calculate and to use the formulas of the calculus of probability at least at the minimal possible level as well as to apply the table of the probability integral which would have been useful to adduce at least in an abridged form as was the case with Chuprov (1909) or even in a more abridged way as Westergaard did.

Among the minor shortcomings [...]. These, however, are trifles which will hardly dumbfound a shrewd reader. More essential, as it seems to me, is the statement (pp. 121 – 129) that the naturalism of the coincidence of the empirical so-called check of the formulas of the calculus of probability with the theoretically predicted for games of chance is explained because "the law [of random deviations – E. S.] itself was, after all, derived from the

results of such experiments and games”. This, however, seems to be an accidental lapse.

Then, the author unjustly attributes to Bortkiewicz (to the law of small numbers) the ascertaining of the “theoretical distribution of the fluctuations” of small numbers without mentioning that the appropriate main formula is due to Poisson. Finally, I would argue against the use of the expression *method of moments* not in regard to the method competing with least squares for drawing a curve, but to the calculation of means by issuing from data grouped into intervals of equal length (p. 531). This will result in the use of an absolutely definite term in an extraneous manner which is hardly sufficiently justified.

Some probabilities are calculated wrongly [...], 0.995 instead of 0.95, 0.999979 instead of 0.997. The figure on p. 566 is scarcely vivid since only a quarter of the correlation diagram is shown, but the exposition of the calculation of the correlation coefficient itself seems to be sufficiently clear even for a beginner which of course was not easy to attain. I only think that the author with his knowledge of explaining could have included in his lengthy treatise rather more practical advice on, and patterns of calculation and not to refer readers so often to other sources either only helpful to a few because of linguistic difficulties (Yule) or insufficiently suited for the beginners (my own contribution of 1912) or, finally, to those entirely unsuited for his aims because of mistakes made (M. B. Gurevich).

In particular, I bear in mind the calculation of means, index numbers and more elementary methods of smoothing series which the author also admits for certain purposes. The inclusion of the formulas and tables due to Pareto and provided by Benini¹⁶ transforms the application of the method of least squares to calculating smoothing curves of the first four degrees into a childish occupation possible perhaps even for a school student of the third form and it would have compelled many practitioners to thank the author heartily. And this will be desirable to see in the likely deservedly soon to appear next edition of his generally speaking excellent treatise.

Notes

1. I mention Lexis (1913): Kaufmann’s contribution “fills an important gap” and occupies “a special place” in the German statistical literature; he manages [makes do] with elementary mathematics which is a favourable circumstance. Then, Lexis believes that the theory of probability assumes equally possible cases and that the law of large numbers ought to be justified by empirical data. In short, I do not discern here a pioneer in the field of statistics.

Slutsky did not say anything about previous studies of the same subject but later he refers to several authors. However, it is opportune to add a few lines (Sheynin 1999). Slutsky mentions, and italicizes the term theoretical statistics but avoids mathematical statistics, a term that appeared at least in 1869 (Zeuner), and he did not define *statistics*. That statistics is a method (see his text a bit below) was stated in 1860 (Fox); and Pearson’s maxim (1892, p. 15) certainly comes to mind: “Unity of all science consists alone in its method, not in its material”. And it was Alphonse DeCandolle who first stated, in 1833, that statistics was a branch of mathematics. I note finally that later scholars, Pearson and Fisher, held that statistics was (“essentially” – Fisher) a branch of applied mathematics. O. S.

2. According to Chuprov (1909), who followed the German philosophers Windelband and Rickert, various sciences are either ideographic or nomographic (rather than nomothetic, as those philosophers called it). The former described reality (history), the latter studied regularity.

Late in life, in his reviews of several books, Chuprov again stated that statistics was an ideographic science although mostly having to do with quantitative data (which is not the case with history, allegedly an ideographic science). However, the literature concerning

philosophy of probability does not anymore mention those philosophers and anyway even history is not a science without discussing regularities. O. S.

3. Judgements and concepts rather belong to philosophy. The *it* in the next sentence is not altogether clear and is only one example of Slutsky's careless style. And in § 10 Slutsky mentioned definitions of probability offered by four authors whereas they really were either general considerations or comments. O. S.

4. Atoms do not *disappear*. O. S.

5. Yes, purely mathematical, but, at that time, not yet belonging to pure mathematics. O. S.

6. Chuprov (1909/1959, p. 133) set great store by free causal connections but I am not at all satisfied by his considerations. Their existence, as he reasoned, led to an unavoidable recognition of the need for probabilities, but he did not mention either correlation or randomness. O. S.

7. At the time, Mises had not yet formulated his frequentist theory of probability. O. S.

8. Venn (1866/1888, p. 88) expresses this idea very distinctly:

There is, it seems to me, a broad and important distinction between a material science which employs mathematics and a formal one which consists of nothing but mathematics.

And on p. 40:

During these [...] chapters we have been entirely occupied with laying what may be called the physical foundations of Probability.

See also pp. 41 and 265 – 266. I quote Venn because both von Kries and Chuprov, as it seems to me, were not altogether just in respect to him. His empiricism is not at all as coarse as can be judged by their opinion and in any case he is not guilty of simply identifying probabilities with empirical frequencies. E. S.

9. It seems to me that these considerations answer Kaufman's objection (p. 151) to my statement that statistics "is a mathematical science": "This is certainly not the case. Statistics is not mathematics". I agree with the latter words, but hope that he will also agree that neither is physics the same as mathematics. E. S.

10. Slutsky several times uses this not quite acceptable expression obviously having in mind a uniformly distributed random variable. O. S.

11. I believe that it is superfluous to corroborate this statement by quotations and references or name some names since any such attempt may be objected to by saying that all this is a *mathematical school*. In regard to at least this issue the essence consists not at all in opposing a *school*. When the debate is about a substance of something, it would be strange to group authors into schools according to their attitude towards propositions sufficiently clearly established by most authoritative scholars from, let us say, Laplace to leading contemporary figures of statistical thought.

I should hardly qualify this statement by adding that, when referring to experts, I do not wish to doubt that hesitations and debates are justifiable. E. S.

12. Also here I indicate that the description of the *differential method* (pp. 139 – 141) is hardly understandable to a beginner, and in essence hardly correct. The interpretation of the formula ... [without consulting Kaufman's treatise the following lines will not be clear. In essence, the matter is rather elementary]. My remark (1912) concerning the probable error of the difference of dependent variables Kaufman (pp. 140, 143, 146) interprets to his advantage, but wrongly, without allowing for my statement elsewhere (1912, p. 100/2009, § 23). [...] E. S.

13. Bortkiewicz (1904, p. 825) used the same expression in the sense of *sampling*. O. S.

14. I take the opportunity to remark that it seems wrong to attribute to Bowley, as became usual apparently because of Chuprov [1912], the principle of composing the sample from purely accidentally snatched elements. The point is that this is the *only* method of sampling prompted by the calculus of probability and it was known long ago; in any case, Laplace had used it. As to the statistical aspect of the problem, it consists not in the principle of randomness as such, but in the technical tricks needed to achieve a purely accidental selection of observations, and here the last word is far from being pronounced. In 1903, at the Berlin session of the International Statistical Institute, March, as the author of a resolution on Kiaer's report adopted by its demographic section, quite rightly, as I believe, objected to him by connecting the only correct version of sampling with Laplace's

investigations rather than with Bowley. E. S. On the history of sampling see You Poh Seng (1951). O. S.

15. Kaufman had indeed published many concrete statistical investigations, but I doubt that they were ever seriously reviewed. O. S.

16. Chuprov (1925) later also referred to Benini (1906) and noted that he was unable to get hold of the relevant Pareto memoir (which he did not name either). O. S.

Bibliography

- Benini R.** (1906), *Principii di statistica metodologica*. Torino.
- Bortkiewicz L. von** (1904), Anwendungen der Wahrscheinlichkeitsrechnung auf Statistik. *Enc. math. Wiss.*, Bd. 1/2. Leipzig, pp. 822 – 851.
- Chuprov A. A.** (1905), Die Aufgabe der Theorie der Statistik. *Schmollers Jahrb. f. Gesetzgebung, Verwaltung u. Volkswirtschaft im Dtsch. Reich*, Bd. 29, No. 2, pp. 421 – 480.
- (1909), *Ocherki po Teorii Statistiki* (Essays on the Theory of Statistics). Moscow, 1910, 1959.
- (1912, in Russian), Sampling. Translation incorporated in Sheynin (1997).
- (1925, in Russian), Review of Khotimsky V. (1925), *Vyravnivanie Statisticheskikh Riadov po Metodu Naimenshikh Kvadratov* (Adjustment of Statistical Series by the Method of Least Squares). Moscow – Leningrad. *Russk. Ekonomich. Sbornik* (Prague), No. 2, pp. 166 – 168.
- Cournot A. A.** (1843), *Exposition de la théorie des chances et des probabilités*. Paris, 1984.
- Husserl E.** (1900 – 1901, in German), *Logical Investigations* (1913). London, 1973.
- Kaufman (Kaufmann) A. A.** (1909), *Teoria Statistiki* (Theory of Statistics). Moscow.
- (1912), *Teoria i Metody Statistiki* (Theory and Methods of Statistics). Later editions: Moscow, 1916, 1922 (posthumously reprinted edition), 1928 (posthumous, with many material added by actual coauthors stressing the importance of mathematical methods).
- (1913), *Theorie und Methoden der Statistik*. Tübingen.
- Kries J. von** (1886), *Die Principien der Wahrscheinlichkeitsrechnung*. Tübingen, 1927.
- Lexis W.** (1913), Review of Kaufmann (1913), *Schmollers Jahrbuch f. Gesetzgebung, Verwaltung u. Volkswirtschaft in Deutsch. Reich*, Bd. 37, pp. 2089 – 2092.
- Pearson, K.** (1892), *Grammar of Science*. London. Many later editions.
- Sheynin O.** (1997), Chuprov's early paper on sampling. *Jahrbücher f. Nationalökonomie u. Statistik*, Bd. 216, pp. 658 – 671.
- (1999), Statistics, definitions of. *Enc. of Statistical Sciences*, 2nd edition, vol. 12, 2006, pp. 8128 – 8135.
- Slutsky E. E.** (1912), *Teoria Korreliatsii* (Theory of Correlation). Kiev. Regarding the existing translation see [i].
- Venn J.** (1866), *Logic of Chance*. London, 1888. [New York, 1962.]
- Vlasov A. K.** (1909), *Teoria veroiatmostei* (Theory of Probability). Moscow.
- Westergaard H.** (1890), *Grundzüge der Theorie der Statistik*. [Jena, 1928, coauthor H. C. Nybolle.]
- You Poh Seng** (1951), Historical survey of the development of sampling theories and practice. *J. Roy. Stat. Soc.*, vol. A114, pp. 214 – 231. Reprint: Kendall M. G., Plackett R. L. (1977), *Studies in the History of Statistics and Probability*, vol. 2. London, pp. 440 – 457.

III

On the Logical Foundation of the Calculus of Probability¹

K voprosu o logicheskikh osnovakh ischislenia veroiatnostei.

Report at the Third All-Union Statistical Congress, Nov. 1922.

Vestnik Statistiki, No. 9 – 12, 1922, pp. 13 – 21.

In a somewhat modified form in *Sbornik Statei Pamiati* (Festschrift) *N. A. Kablukova*, vol. 1. Moscow, 1925, pp. 254 – 262.

Finally, in Slutsky's posthumous *Izbrannye Trudy* (Sel. Works).

Moscow, 1960, pp. 18 – 24

[1] The calculus of probability is usually explicated as a purely mathematical discipline, and it is really such with respect to its main substance when considered irrespective of applications. However, the pure mathematical nature of one element that enters the calculus from its very beginning is very questionable: any detailed interpretation of that element involves our thoughts in a domain of ideas and problems foreign to pure mathematics. Of course, I bear in mind none other than the notion of probability itself. As an illustration, let us consider, for example, the classical course of Academician Markov.

In the Introduction to its second edition (1908), he declares that he will treat the calculus of probability as a branch of mathematics, and each attentive reader knows how strictly he regards his promises. Markov shows this strictness at once, in the extremely typical of him comment on the second line of the very first page. There, he elaborates on the word *we*:

The word we is generally used in mathematics and does not impart any special subjectivity to the calculus of probability.

Let us however compare this pronouncement with Markov's definition of equipossibility that he offers on p. 2:

We call two events equally possible if there are no grounds for expecting one of them rather than the other one.

He adduces a note saying that, according to his point of view,

Various concepts [...] are defined not so much by words, each of which in turn demands a definition, as by our attitude to them, which is ascertained gradually.

It is doubtless, however, that, in the given context, this remark should only be considered as a logically hardly rightful way out for the author's feeling of some dissatisfaction with his own definition. That subjective element, whose shadow he diligently attempted to drive out by his remark on the first page, appears here once more so as to occupy the central position in the structure of the main notion, which must serve as the foundation for all the subsequent deliberations.

In my opinion, there is a means for countering this difficulty; there exists an absolutely drastic measure that many will, however, be bound to consider

as cutting rather than untangling the Gordian knot. The legend tells us, that, nevertheless, such an attitude proved sufficient for conquering almost the whole world. I shall sketch the idea of my solution.

First of all, it is necessary to introduce the main notions defined in the spirit of strict formalism by following the classical example of Hilbert's *Grundlagen* (1899). Such notions as *event*, *trial*, *the solely possible events*, *(in)compatible events*, etc, ought to be established in this way, *i.e.* with the removal of all the concepts concerning natural sciences (time, cause, etc). Let us call the complex of the solely possible and incompatible events A, B, \dots, H an *alternative*, and the relation between them, *disjunction*. Then, instead of introducing the notion of equipossibility, we shall proceed as follows.

[2] We will consider such relations, which take place if some number is associated with each of these solely possible and incompatible *events*, under the condition that, if any of them (for example, A) is in turn decomposed into an alternative (either α , or β , or γ , ..., or η), then the sum of those numbers, that occur to be associated with $\alpha, \beta, \gamma, \dots, \eta$, will be equal to the number associated with A .

The association just described should be understood as the existence of some one-valued but not one-to-one relation R between the *events* included in the alternative and the numbers. In addition, R is the same for all the *events* and possesses the abovementioned formal property, but in essence it remains absolutely arbitrary in the entire domain of the calculus considered as a purely mathematical discipline. It can even happen that, in the context of one issue, each term of an alternative is connected with some number by relation R , whereas each term of another alternative is in turn connected with some number by relation R' not identical with R ; the relation R'' will take place for a third alternative, and so on. If, in addition, a formal connection between the relations R, R', R'', \dots is given, purely mathematical complications, which the classical calculus of probability had never studied in a general setting, will arise. Leaving them aside, I return however to the simplest type.

Suppose that an alternative can be decomposed into the solely possible and incompatible *events* with which some fundamental relation R connects numbers equal one to another. I shall call such elementary *events* isovalent, and I shall introduce the notion of valency of an event as a proper fraction whose numerator is equal to the number of the elementary *events* corresponding to the given *event*, and whose denominator is the number of all the solely possible elementary and incompatible *events* included in the given alternative.

It is absolutely obvious that this foundation formally completely coincides with the classical foundation; hence, all the former's purely mathematical corollaries will formally be the same. The word *probability* will everywhere be substituted by *valency*; the formulation of all the theorems will, *mutatis mutandis*, persist [with necessary alterations]; all the proofs will remain valid. The only change consists in that the very substance of the calculus will not now have any direct bearing on probability.

For example, the addition theorem will be formulated thus: If A and B are events incompatible one with another, the valency of the event "either A or B " is equal to the sum of their valencies. The multiplication theorem will be: For compatible events A and B , the valency of the event "both A and B " is

equal to the valency of one of them multiplied by the conditional valency of the other one; etc.

The purport of any theorem obviously remains purely formal until we, when somehow applying it, associate some material sense with the fundamental relation R ; that is, until we fix the meaning of those numbers, that in the given case are attached to the terms of the alternative. Knowing the sense in which such and such *events* are isovalent, we will be logically justified, on the grounds of our calculus, to state that some other definite *events* will also be isovalent or have such and such valency, again in the same sense. It will now be naturally inconsistent to call our science calculus of probability; the term *disjunctive calculus* will apparently do².

[3] This science will be as formal and as free of all the non-mathematical difficulties as the theory of groups. There, we are known to be dealing with some things, but it remains indefinite with which exactly. Then, we have to do there with some relation that can conjugate any two things one with another so that the result of this operation will be some third thing from the same totality. Under these conditions, the theory of groups develops an involved set of theorems, mathematically very elegant and really important for various applications. Within the bounds of the theory itself, the material substance of that set remains indefinite which leads to formal purity and variety of applications, and is indeed one of the theory's most powerful points. If the group consists of natural numbers, and the main operation providing a third thing out of the two given ones is addition, we obtain one possible interpretation; if the main operation is multiplication, we arrive at another interpretation; then, when compiling a group out of all possible permutations of several numbers, we get a still differing interpretation; and if, instead, we consider all possible rotations of some regular polyhedron, we have a yet new interpretation, etc.

In our case, we also have something similar. The formal notion of valency can have more than one single sense, and the meaning of the theorems known to us long since in their classical form is in essence also many-valued. Their nature remains however hidden and is only dwelt with during disputes, to a considerable extent fruitless, on the notion of probability. I shall attempt to sketch several possible interpretations of the calculus of alternatives.

First of all, we certainly have its classical form. We come to it by replacing isovalency by equipossibility and substituting probability for valency. This change may be considered from a purely formal, and from a material point of view. When keeping to the former, which is the only interesting one for a mathematician, we introduce the concepts purely conventionally. Suppose that the *possibility* of an event can be higher than, or equal to the *possibility* of another event. Presume also that two events, each decomposable into the same number of other solely and equally possible incompatible events, are themselves necessarily equally possible. Then, irrespective of either a more definite meaning of, or of the conditions for equipossibility, we are able to introduce, in the usual way, the notion of *probability* in its purely mathematical aspect. Or, otherwise, when keeping closer to the reasoning above, we may say: Suppose that *possibility* can be expressed numerically and that the *possibility* of an event is equal to the sum of the *possibilities* of those solely possible and incompatible events into which it is decomposable. Then, etc, etc.

This deliberation is tantamount to the following. We have a finished formal mathematical calculus complete with its notions and axioms. When applying it, we suppose that those axioms, that underpin the formal disjunctive calculus, are valid for some chosen concept, – for example, as in our case, for *possibility*. Thus, we presume that *possibilities* can be expressed by numbers; that all the terms of a given alternative are connected with these numbers by a one-to-one correspondence; that these latter obey those formal relations which we introduced for the numbers connected with the former by valency. From a purely mathematical viewpoint, this is apparently quite sufficient for passing on from the calculus of alternatives to the calculus of probability.

It is obvious, however, that all this only covers one aspect of the matter, and that here also exists another, material, so to say, side lying entirely beyond the bounds of purely mathematical ideas and interests. Indeed, for settling the issue of whether all the abovementioned notions and axioms categorically rather than conditionally suit the concepts *possibility* and *probability*, we ought to know what exactly do we mean by these notions. It is clear that this problem is of an absolutely special type requiring not a mathematical, but an essentially different phenomenological and philosophical approach. I think that for my formulation of the issues, the line of demarcation appears with sufficient clearness as though all by itself.

[4] Let us now go somewhat farther in another direction. I have remarked that the calculus of alternatives admits not a single interpretation, but rather a number of them, and this formal generality is indeed one of its most important logical advantages over the classical calculus of probability. So as to justify this idea, I ought to indicate at least one more of its differing interpretations. Let us have a series of trials where each of the events A, B, \dots, H is repeated several times. The numbers of these repetitions, *i.e.*, the actual absolute frequencies of the events, are uniquely connected with the events because each of them has one certain frequency. This relation is not biunique, because, inversely, two or more events can have one and the same frequency. Then, if some event is decomposable into several solely possible and incompatible kinds, the sum of their frequencies is equal to the frequency of the given event. Frequency thus satisfies those conditions under which I introduced the concept of valency into the calculus of alternatives.

We may therefore replace valency by relative frequency and thus obtain a number of theorems with respect to the latter without repeating all the deliberations or calculations, but jokingly, so to say, by a single substitution of the new term into the previous formal statements. Thus, we will have the addition and the multiplication theorems for frequencies, absolutely analogous to the known propositions of the calculus of probability. How far-reaching are such similarities? Obviously, they go as far as the general foundation of definitions and axioms do. Had there been no other independent entities except these notions and axioms in the disjunctive calculus, or, respectively, the calculus of probability, then the calculus of frequencies would have formally covered the entire contents of both the two former calculuses. This, however, is not so. The issue of repeated trials and the concept of frequency enter the calculus of probability at one of its early stages; in addition, we should naturally find out whether, when assuming

our second interpretation of the calculus of alternatives, the formal conditions that correspond to that stage can be, and are actually satisfied.

[5] More interesting is a third interpretation. It goes much farther, covers a large and perhaps even the entire domain of our calculus provided only that we can agree with a purely empirical understanding of *probability*. Suppose that it makes sense to consider any number of trials under some constant conditions. Presume also that there exists a law on whose strength the relative number of the occurrences of any of the alternatively possible events must tend to some limit as the number of trials increases. This limiting relative frequency [These ... frequencies] apparently satisfies [satisfy] those conditions, under which I introduced the notions of isovalency and valency of events. Hence, as far as the general foundation of the axioms reaches, all the theorems of the calculus of valency will possess this, as well as the classical interpretation. That the analogy goes very far is unquestionable. To say nothing about the almost trivial addition and multiplication theorems, it also covers the doctrine of repetition of events including such of its propositions as the Jakob Bernoulli, and the [De Moivre –] Laplace theorems. Small wonder that sometimes all civic rights are granted to this interpretation. Thus, we find it as a special favourite in the British school. What has it to do with the classical interpretation? Does it entirely cover the latter? And, if not, where do they diverge? Only in the understanding of the sense of the theorems, or perhaps in the extent of mathematical similarity? Until now, there are no definitive answers to any of these questions.

A rigorous revision of all the fundamentals of the calculus of probability, a creation of a rigorous axiomatics and a reduction of the entire structure of this discipline to a more or less visible mathematical form, are necessary. This however is only possible on the basis of a complete formalization of the calculus with the exclusion from it of all not purely mathematical issues. Neither probability, nor the potential limiting frequency possess such a formal nature. The calculus of probability should be converted into a disjunctive calculus as indicated above, and only then will it enter the system of mathematical sciences as its branch and become definitive, a quality which it is still lacking, and enjoy equal logical rights with the other branches.

My solution is however something more than a simple methodical device for disentangling the issues of the logic of the calculus of probability. So as to convince ourselves in this fact, suffice it to imagine that nature of logical purity which our calculus will obtain as a result of the indicated conversion, which is something objective, as are all the borders separating the sciences one from another. We reveal them, but we do not create them. Indeed, it needs only to compare with each other even those few theorems whose statements in terms of *probabilities*, *frequencies* and potential *limiting frequencies* are unquestionable.

Let us only imagine three such absolutely parallel series of definitions, axioms and theorems explicated independently, and, consequently, roughly speaking, separately from each other in three different treatises devoted to three supposedly separate calculuses respectively. In each case we will have independent series of ideas, definitions and proofs. We ask ourselves, whether the similarity between them is objective or subjective. The answer is self-evident. The general pattern and the course of reasoning are the same.

Once we perceive this, we also observe that the likeness exists irrespective of our subjective arbitrariness.

[6] It may be objected, that the formalization of the calculus of probability postulated here avoids exactly the most essential and the most interesting for theoretical statistics issues. This, however, is no objection. The essence of *probabilities*, the relations between *probability* and *limiting frequency*, and between the calculus of probability and the real course of things, – all these problems are important and interesting, but they are of another logical system, and, moreover, such, whose proper statement is impossible without solving simpler and logically more primitive problems. Their definitive and complete solution, as dictated by the entire development of mathematical thought, lies exactly in the direction whose defence is the subject of my study. We only have to dart a look on the issues, concerning the essence of the notion of probability and of its relation to reality, for understanding with full clearness their utter distinction from the formal mathematical problems comprising the subject of the disjunctive calculus and its axiomatics and logic. Thus, only a logical and phenomenological analysis absolutely not of a formal mathematical nature can indicate that probability is a category unto itself, completely independent of the notion of *limiting frequency*.

Now I allow myself a remark as a hint of a solely preliminary nature. Suppose that we have a number of frequencies which must surely approach some limit as the number of repetitions [of trials] increases unboundedly. It does not however follow at all that in some initial part of the trials the [studied] event could not have been repeated with a frequency essentially different from its limiting value. Suppose for example that a sharper deals the cards unfairly; that he cheats relatively less as the game goes on; and that in the limit, as the number of rounds increases unboundedly, each card will appear with frequency $1/52$ as it should have happened under fair circumstances³. Even without knowing anything about the law governing the composition of the series of trials, we would nevertheless be sure to discover, after observing the actual behaviour of the frequencies, that, with probability extremely close to certainty, the probability of the event during the first series of the trials diverges from its limiting value not less than by such-and-such amount. True, the notion of limiting frequency can also be applied to the proportion of right and wrong judgements, but neither here is the issue definitively decided: just as in the case above, we may ask the [same] questions about the probability of judgement, about the frequency and the probability of that proportion⁴.

[7] The same is true with respect to the possibility of applying the calculus of probability to empirical experience. Not the latter guides us when we establish the calculus' theorems, but, on the contrary, they, and only they, provide us with a prior compulsory clue for regulating it. From the calculus of probability we borrow the type of that law, which, following N. A. Umov⁵, we might have called the law of chaos, of complete disorder. There exist domains of phenomena where the chain of causes and effects on the one hand, and the arrangement of ideographic information⁶ on the other hand, ensure, in conformity with natural laws, the regularity of such a sequence: if the occurrence (non-occurrence) of some event is denoted by A (by B), then, in the limit, as the number of trials increases unboundedly, A ought to appear with the same relative frequency both in the entire series

and after any combination of the events; equally often after A , and after B ; after AA , AB , or BB ; after AAB , AAA , ABA , etc, etc.

That such domains actually exist is shown by experience, but only when the idea of probability guides it and provides the very pattern of the law of chaos and the tests for establishing its action in one or another field, and for appraising the judgement which establishes it. Hence, in this respect the notion of probability also becomes indispensably necessary and logically primary. Is it even possible to justify the natural philosophical premises of the law of chaos without applying the notion of probability? I think that this is questionable.

Now, I have however went out of the boundaries of my main subject although this was apparently not quite useless for its elucidation. My concluding remarks will perhaps amplify the purely logical arguments by a vivid feeling, caused not by a logically formal consideration, but by direct vision and comprehension of the essence of things and issues.

Notes

1. After my text had appeared in *Vestnik Statistiki*, I [see 1925a] improved some formulations making them more intelligible and introduced a few editorial corrections, but I did not change anything in essence. E. S.

2. After my report was published, Professor Bortkiewicz, in a letter [to me], kindly suggested this term [xv, Letter No. 3]. E. S.

Khinchin (1928, p. 126) mentioned the disjunctive calculus “according to Slutsky’s known terminology”. O. S.

3. I have omitted some details in this passage because Slutsky had not explained the essence of the game. Also note that his example is actually directed against the frequentist (Mises) theory of probability. O. S.

4. Some explanation is lacking. O. S.

5. Russian physicist (1846 – 1915). Slutsky provided no reference. O. S.

6. Ideography, the science of single facts, of history. This notion goes back to the philosophers Windelband and Rickert. Also see Sheynin (1990/1996, p. 98) and Foreword, comment on [viii]. O. S.

Bibliography

Bernstein, S.N. (1917, in Russian), An essay on an axiomatic justification of the theory of probability. *Sobranie Sochinenii* (Coll. Works), vol. 4. No place, 1964, pp. 10 – 60.

Translation Sheynin (2005, pp. 49 – 111). Also at www.sheynin.de

Khinchin, A. Ya. (1928, in Russian), The strong law of large numbers and its significance for mathematical statistics. *Vestnik Statistiki*, No. 1, pp. 123 – 128.

Kolmogorov, A.N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin.

--- (1948, in Russian), Obituary: Evgeny Evgenievich Slutsky. Translation (2002): *Math. Scientist* vol. 27, pp. 67 – 74.

--- (1974). Second Russian translation of Kolmogorov (1933). Moscow.

Markov, A. A. (1900), *Ischislenie Veroiatnostei* (Calculus of Probability). Subsequent editions: 1908, 1913 and 1924. German translation of the second edition: *Wahrscheinlichkeitsrechnung*. Leipzig – Berlin, 1912.

Sheynin, O. (1990, in Russian), A. A. *Chuprov: Life, Work, Correspondence*. Göttingen, 1996.

--- (2005), *Probability and Statistics. Russian Papers of the Soviet Period*. Coll. translations of various authors. Berlin. Also www.sheynin.de

Slutsky, E.E. (1925). Über stochastische Asymptoten und Grenzwerte. *Metron*, Bd. 5, No. 3, pp. 3 – 89.

--- (1960), *Izbrannye Trudy* (Sel. Works). Moscow.

IV

On Some Patterns of Correlation Connection and the Systematic Error of the Correlation Coefficient

О некоторых схемках корреляционной связи и о систематической ошибке коэффициента корреляции. *Vestnik Statistiki*, No. 1 – 3, 1923, pp. 31 – 50

1. Chetverikov (1921) has recently considered a few patterns of connection between causes and effects leading to correlation between random variables. All those patterns can be generalized to appear as particular cases of a still more [?] general pattern (Yule 1912, Problem 6, p. 227 and Answer on p. 365).

The correlation coefficient for the last mentioned generalization can be easily and rigorously determined by the method of moments. Consider the correlation between two variables, both being sums of two terms supposing that their connection consists in that one term is common for both. And so, let

$$x = u + w, y = v + w \quad (1)$$

where u , v and w are random variables, independent from each other in respect of probabilities. Denoting expectation by E , we therefore have

$$Eu^r w^s = Eu^r Ew^s, Ev^r w^s = Ev^r Ew^s, Eu^r v^s = Eu^r Ev^s. \quad (2)$$

Note that for deriving the formula of the correlation coefficient it is not necessary to presume absolute independence in Chuprov's sense (1922, p. 241/2004, § 1.2)¹; it is sufficient for formula (2) to take place at $r = s = 1$. In usual notation, for some random variable z ,

$$\bar{z} = Ez, \mu_2(z) = \sigma_z^2 = E(z - \bar{z})^2,$$

we will say that the correlation coefficient is

$$r_{xy} = \frac{E(x - \bar{x})(y - \bar{y})}{\sigma_x \sigma_y}, \quad (3)$$

where, as is known and easy to prove,

$$E(\bar{x} - x)(\bar{y} - y) = Exy - \bar{x}\bar{y}.$$

Let us now derive the formula for the correlation coefficient between two random variables obeying the above conditions. We have

$$\begin{aligned} E(x - \bar{x})(y - \bar{y}) &= Exy - \bar{x}\bar{y}, \\ Exy &= E(u + w)(v + w) = Euv + Euw + Evw + Ew^2 = \\ \bar{u}\bar{v} + \bar{u}\bar{w} + \bar{v}\bar{w} + Ew^2 &= (\bar{u} + \bar{w})(\bar{v} + \bar{w}) - \bar{w}^2 + Ew^2 = \\ \bar{x}\bar{y} + Ew^2 - \bar{w}^2 &= \bar{x}\bar{y} + E(w - \bar{w})^2. \end{aligned}$$

Therefore,

$$E(x - \bar{x})(y - \bar{y}) = \mu_2(w),$$

$$r_{xy} = \frac{\mu_2(w)}{\sqrt{\mu_2(x)\mu_2(y)}} = \frac{\sigma_w^2}{\sigma_x \sigma_y} = \frac{\sigma_w^2}{\sigma_{u+w} \sigma_{v+w}}. \quad (4a, b)$$

It is not difficult to prove that formulas (4) will also be valid for

$$x = A(u + w), y = B(v + w)$$

where A and B are some constant coefficients².

This formula allows us to introduce the following pattern of connection between two random variables with u , v and w being the causes, and x and y , the effects. *If two phenomena have a common cause in addition to special causes independent both from each other and from that common cause, then, for any law of distribution of probabilities, in case the causes are additive and the effects proportional to them, the correlation coefficient will be equal to the ratio of the mean square of the common cause and the geometric mean of of their mean squares of the causes.*

2. Consider now some particular cases.

a) Suppose that the causal connection admits of being described in the following way. The special causes are proportional to the numbers u and v of white balls having been extracted with replacement from two urns after n_1 and n_2 trials with probabilities p_1 and p_2 ; the common cause is proportional to the number w of white balls extracted from a third urn in n_3 trials with probability p_3 . Then

$$x = Au + Bw = B[(A/B)u + w], y = Cv + Dw = D[(C/D)v + w].$$

Now, in accordance with the above, and directly applying formula (4b), we will have

$$r_{xy} = \frac{\sigma_w^2}{\sigma_{(A/B)u+w} \sigma_{(C/D)v+w}}.$$

It is known that

$$\sigma_u = \sqrt{n_1 p_1 q_1}, \sigma_v = \sqrt{n_2 p_2 q_2}, \sigma_w = \sqrt{n_3 p_3 q_3}$$

where $q_i = 1 - p_i$, $i = 1, 2, 3$.

In addition, since the second moment of a sum of independent magnitudes is equal to the sum of their second moments,

$$\sigma_{(A/B)u+w}^2 = \sigma_{(A/B)u}^2 + \sigma_w^2 = \left(\frac{A}{B}\right)^2 \sigma_u^2 + \sigma_w^2 = \frac{A^2 n_1 p_1 q_1 + B^2 n_3 p_3 q_3}{B^2},$$

$$\sigma_{(C/D)u+w}^2 = \sigma_{(C/D)v}^2 + \sigma_w^2 = \left(\frac{C}{D}\right)^2 \sigma_v^2 + \sigma_w^2 = \frac{C^2 n_2 p_2 q_2 + D^2 n_3 p_3 q_3}{D^2}.$$

$$r_{xy} = \frac{BDn_3 p_3 q_3}{\sqrt{A^2 n_1 p_1 q_1 + B^2 n_3 p_3 q_3} \sqrt{C^2 n_2 p_2 q_2 + D^2 n_3 p_3 q_3}}.$$

If $A = B$ and $C = D$,

$$r_{xy} = \frac{n_3}{\sqrt{n_1 p_1 q_1 + n_3 p_3 q_3} \sqrt{n_2 p_2 q_2 + n_3 p_3 q_3}}. \quad (5)$$

Finally if $p_1 = p_2 = p_3 = p$,

$$r_{xy} = \frac{n_3}{\sqrt{(n_1 + n_3)(n_2 + n_3)}}$$

which is Chetverikov's formula. If $n_1 = n_2$, as in the trials made by Darbishire [1907], and we assume for the sake of convenience that $n_1 = n_2 = n$ and $n_3 = m$, then

$$r_{xy} = \frac{m}{n + m}.$$

Only under the conditions stated concerning the coefficients A , B , C , and D and equality of the probabilities, it is possible to represent the correlation coefficient as a *ratio of the number of common elementary causes of two phenomena to the geometric mean of all the elementary causes of both*.

b) We may obtain the same results by considering another pattern. Let there be n_1 special elementary causes of phenomenon x , each of them contributing either a positive or negative elementary component A to x with probabilities p_1 and q_1 and suppose that the number of positive A 's is u . Also, let similar conditions hold for the other causes as well. Then

$$x = [Au - A(n_1 - u)] + [Bw - B(n_3 - w)] = A(2u - n_1) + B(2w - n_3),$$

$$y = [Cv - C(n_2 - v)] + [Dw - D(n_3 - w)] = C(2v - n_2) + B(2w - n_3)$$

All the mean square magnitudes σ will obviously be here twice as large as in the previous case a) so that the formulas for the correlation coefficient provided above will also be valid.

c) As our last example, we consider three groups of mutually independent random variables

$$x_1, x_2, \dots, x_{n_1}; y_1, y_2, \dots, y_{n_2}; z_1, z_2, \dots, z_{n_3},$$

with differing values having different probabilities according to their own arbitrary laws possible for each group. Let

$$u = \sum_{i=1}^{n_1} x_i, \quad v = \sum_{j=1}^{n_2} y_j, \quad w = \sum_{k=1}^{n_3} z_k,$$

required is the correlation coefficient between x and y as defined by (1).

The conclusions of § 1 are of course valid here also. Namely, since

$$\sigma_u^2 = \sum_{i=1}^{n_1} \sigma_{x_i}^2, \quad \sigma_v^2 = \sum_{j=1}^{n_2} \sigma_{y_j}^2, \quad \sigma_w^2 = \sum_{k=1}^{n_3} \sigma_{z_k}^2, \quad \sigma_x^2 = \sigma_u^2 + \sigma_w^2, \quad \sigma_y^2 = \sigma_v^2 + \sigma_w^2,$$

$$r_{xy} = \frac{\sum_{k=1}^{n_3} \sigma_{z_k}^2}{\sqrt{\left[\sum_{i=1}^{n_1} \sigma_{x_i}^2 + \sum_{k=1}^{n_3} \sigma_{z_k}^2 \right] \left[\sum_{j=1}^{n_2} \sigma_{y_j}^2 + \sum_{k=1}^{n_3} \sigma_{z_k}^2 \right]}}.$$

If

$$\sigma_{x_1} = \sigma_{x_2} = \dots = \sigma_{y_1} = \sigma_{y_2} = \dots = \sigma_{z_1} = \sigma_{z_2} = \dots,$$

which can take place, for example, when the laws of distribution are the same for all random variables of the discussed problem, then (5) takes place.

We will obtain such a correlation coefficient when compiling sums of numbers according to the following rule. Let the terms of a number sequence follow one another absolutely irregularly and take the values 0, 1, 2, ..., 8, 9 with one and the same probability 1/10; such, as we may believe, is a sequence of the last digits in a seven-digit table of logarithms. From that sequence, isolate subsequences of n_1 , n_3 , and n_2 numbers, add up the first n_1 numbers with the first n_3 numbers, then the latter with the next n_1 numbers, etc³.

3. When wishing to compare the theoretical formulas derived above with the results of some appropriate experiment, we will have to allow for the fact that the formula for calculating the correlation coefficient by issuing from the empirical frequencies is corrupted by some systematic error. Let the empirical values of the variables be x' and y' , their arithmetic mean, \bar{x}' and \bar{y}' , then the empirical value of the correlation coefficient⁴

$$\rho = \frac{\sum (x' - \bar{x}')(y' - \bar{y}')}{\sqrt{\sum (x' - \bar{x}')^2 \sum (y' - \bar{y}')^2}}$$

will be only corrupted by a random error if

$$E\rho = r_{11} = \frac{E(x - \bar{x})(y - \bar{y})}{\sqrt{E(x - \bar{x})^2 E(y - \bar{y})^2}}.$$

According to the newest investigations (Chuprov 1922, p. 267/2004, § 4.2A), this, however, does not take place, and the systematic error is (to the terms of order $1/N$)

$$\varepsilon_p = E\rho - r_{11} = \frac{1}{N} \left[\frac{1}{4} r_{22} r_{11} + \frac{3}{8} r_{11} (r_{40} + r_{04}) - \frac{1}{2} (r_{31} + r_{13}) \right] + \dots \quad (6)$$

Here, N is the number of all cases in the given empirical totality, and

$$r_{ij} = \frac{\mu_{ij}}{\mu_{20}^{j/2} \mu_{02}^{j/2}},$$

$$\mu_{ij} = E(x - \bar{x})^i (y - \bar{y})^j, \quad \mu_{20} = E(x - \bar{x})^2, \quad \mu_{02} = E(y - \bar{y})^2.$$

For the mean square random error of the correlation coefficient, to terms of the order of $(1/N)$, we have (Chuprov 1922, p. 269/2004, § 4.2B)

$$\sigma_p^2 = E(\rho - \bar{\rho})^2 = \frac{1}{N} \left[r_{22} \left(1 + \frac{1}{2} r_{11}^2 \right) - r_{11} (r_{31} + r_{13}) + \frac{1}{4} r_{11}^2 (r_{40} + r_{04}) \right] + \dots \quad (7)$$

and it is somewhat interesting to apply the few theoretical patterns of correlation connection for ascertaining, at least for those cases, how large can the systematic error of the usual formula for the correlation coefficient be as compared with its random error.

To achieve that goal, we have to find the expressions for all the needed moments. By applying the method of § 1, I obtain after easy algebraic work, again for x and y obeying relations (1),

$$\begin{aligned} \mu_4(x) &= \mu_4(u) + \mu_4(w) + 6\mu_2(u)\mu_2(w), \\ \mu_4(y) &= \mu_4(v) + \mu_4(w) + 6\mu_2(v)\mu_2(w), \\ \mu_{22}(xy) &= \mu_2(x) + \mu_2(y) + \mu_4(w) - \mu_2^2(w), \\ \mu_{31}(xy) &= \mu_4(w) + 3\mu_2(u)\mu_2(w), \quad \mu_{13}(xy) = \mu_4(w) + 3\mu_2(v)\mu_2(w), \end{aligned}$$

so that

$$\begin{aligned} r_{11} &= \frac{\mu_2(w)}{\sqrt{\mu_2(x)\mu_2(y)}}, \quad r_{22} = 1 + \frac{\mu_4(w) - \mu_2^2(w)}{\mu_2(x)\mu_2(y)}, \\ r_{40} &= \frac{\mu_4(x)}{\mu_2^2(x)} = \frac{\mu_4(u) + \mu_4(w) + 6\mu_2(u)\mu_2(w)}{\mu_2^2(x)}, \quad (8) \\ r_{04} &= \frac{\mu_4(y)}{\mu_2^2(y)} = \frac{\mu_4(v) + \mu_4(w) + 6\mu_2(v)\mu_2(w)}{\mu_2^2(y)}, \\ r_{31} &= \frac{\mu_4(w) + 3\mu_2(u)\mu_2(w)}{\sqrt{\mu_2^3(x)\mu_2(y)}}, \quad r_{13} = \frac{\mu_4(w) + 3\mu_2(v)\mu_2(w)}{\sqrt{\mu_2^3(y)\mu_2(x)}}. \end{aligned}$$

4. We will now apply the formulas of § 3 for the particular patterns of § 2; to simplify our problem, we will assume those patterns in a more specific way.

Pattern a). Suppose we have three urns; n extractions are made from each of the first two, and m extractions from the third one. In each case, the probability of the occurrence of a white ball is p , and the random numbers of the extracted white balls are u , v , and w respectively. We will consider the correlation between x and y obeying conditions (1) and denote $n + m = s$.

The moments of the so-called binomial distribution are known. For the first urn, say, they are

$$\begin{aligned}\bar{u} &= E(u) = np, \quad \mu_2(u) = E(u - \bar{u})^2 = npq, \\ \mu_3(u) &= E(u - \bar{u})^3 = npq(q - p), \\ \mu_4(u) &= E(u - \bar{u})^4 = 3n^2 p^2 q^2 + npq(1 - 6pq).\end{aligned}$$

Similar formulas are valid for the other urns, and, because the trials are independent, for the extractions from the first two urns taken together, i. e., for x and y , as well. Noting that

$$r_{xy} = r_{11} = \frac{m}{s}, \quad (9)$$

introducing

$$c = \frac{1 - 6pq}{pq} \quad (10)$$

and issuing from the results of § 3, we arrive after insignificant transformations at

$$r_{22} = 1 + 2r_{11}^2 + \frac{c}{s} r_{11}, \quad r_{40} = r_{04} = 3 + \frac{c}{s}, \quad r_{31} = r_{13} = \left(3 + \frac{c}{s}\right) r_{11}.$$

With an increasing s , these expressions approach the corresponding moments of the so-called “normal” distribution

$$r_{22} = 1 + 2r_{11}^2, \quad r_{40} = r_{04} = 3, \quad r_{31} = r_{13} = 3r_{11}.$$

Calculating as previously to the terms of the order $1/N$, we have for the systematic and random errors of the correlation coefficient

$$\varepsilon_p = E\rho - r_{11} = -\frac{r_{11}(1 - r_{11})[1 + r_{11} + (c/2s)]}{2N} + \dots, \quad (11)$$

$$\sigma_p = \sqrt{E(\rho - \bar{\rho})^2} = \frac{1}{\sqrt{N}} \sqrt{(1 - r_{11}^2)^2 + \frac{c}{2s} r_{11}(1 - r_{11})(2 - r_{11}) + \dots} \quad (12)$$

and, as the first approximation,

$$\frac{\varepsilon_p}{\sigma_p} = -\frac{1}{2\sqrt{N}} \frac{r_{11}[1+r_{11}+(c/2s)]}{\sqrt{(1+r)^2 + \frac{c}{2s} \frac{r_{11}(2-r_{11})}{1-r_{11}}}} + \dots \quad (13)$$

When considering this expression, we note that in two cases, when $c < 0$ and $r_{11} \rightarrow 1$, and when r_{11} is arbitrary and $c \rightarrow \infty$, it will apparently infinitely increase. In the first case, the left side of (13) cannot actually become as large as desired. Let for example $p = q = 1/2$, and, see (10), $c = -2$. We know, see (9) and the conditions of the problem, that m and s are whole numbers. For having $r_{11} = 0.99$, say, we ought to extract at least 99 balls from the third urn and one ball from each of the two first ones. In general, n must at least be equal to 1. Then

$$\frac{n}{s} = \frac{1-r_{11}}{r_{11}} = \frac{1}{s}, \quad s = \frac{r_{11}}{1-r_{11}}$$

and formula (13) will provide

$$\frac{\varepsilon_p}{\sigma_p} = -\frac{1}{2\sqrt{N}} \frac{r_{11}^2 + 2r_{11} - 1}{\sqrt{r_{11}^2 + 3r_{11} - 1}}$$

This magnitude takes its maximum value when $r_{11} = 1$, when further $\varepsilon_p/\sigma_p = -1/\sqrt{3N}$. For $N = 12$ and 27 it is $-1/6$ and $-1/9$ etc. and in any case the systematic error of the correlation coefficient is several times less than its random error. The difference as compared with the “normal” distribution is also small. At $c = 0$ or finite c and $s = \infty$ formulas (11) – (13) are transformed into formulas of the “normal” distribution.

Then, for $r_{11} = 1$, $\varepsilon_p/\sigma_p = -1/2\sqrt{N}$ which only very little differs from the above.

Pattern b). If, for example, q approaches zero, c , see formula (10), can become arbitrary large. For $r_{11} = 1/2$ formula (13) provides

$$\frac{\varepsilon_p}{\sigma_p} = -\frac{1}{4\sqrt{N}} \sqrt{1+(c/3s)}$$

Let $p = 99/100$, or $9999/10000$, $q = 1/100$ or $1/10000$; with $s = 10$, $\varepsilon_p/\sigma_p = -0.51/\sqrt{N}$; and $-4.6/\sqrt{N}$. If $N = 25$, then $\varepsilon_p/\sigma_p = -0.1$ and -0.9 etc.

A closer examination, however, reveals the following circumstance. When separately calculating ε_p and σ_p the second example will provide $\varepsilon_p = -2.5$ and $\sigma_p = 2.75$. First of all, the absolute value of ρ cannot exceed 1 so that the result obtained must be attributed to an inadmissible use of approximate values of these two magnitudes, to neglecting the terms of the order $1/N^2$. Second, it is also obvious that, when extracting 5 balls from any of the three urns containing 9999 white balls and 1 black ball, we should expect in an overwhelming number of cases all the extracted balls to be white.

In each series of 5 extractions, one after another, we ought to have $x = u + w = 5 + 5 = 10$ and $y = v + w = 5 + 5 = 10$. Assuming $N = 25$ as we did in our example, we will have to make $25 \cdot 15 = 375$ extractions for compiling 25 values of both x and y , and the probability that not a single black ball will occur is $(9999/10000)^{375} = 0.97$. It follows that in an overwhelming number of cases the series of 25 values of x and y will consist of the *same* numbers with zero deviations from their means. The correlation coefficient will then take the empirical expression $\rho = 0/0$.

In other words, *no* conclusion can be formulated under the stated conditions. If we increase N so that the chance of at least one black ball occurring during all $15N$ trials will be at least $1/2$, we will have to choose $N = 502$ approximately, then $\varepsilon_\rho = 0.125$, $\sigma_\rho = 0.61$ and $\varepsilon_\rho/\sigma_\rho = 0.2$. For any sufficiently large N and the correlation coefficient to be reasonably formulated, that ratio will even become essentially smaller. The general conclusion is that the studied theoretical patterns, in all practically interesting cases, always provide a systematic error of the correlation coefficient several times less than the random mean square error.

5. Turning now to *Pattern c*) of § 2, I only consider the case in which all the variables

$$x_1, x_2, \dots, y_1, y_2, \dots, z_1, z_2, \dots$$

obey the same law of distribution. Suppose also that, as it was in § 4,

$$n_1 = n_2 = n, n_3 = m, n + m = s.$$

Then u, v, w, x and y will all be sums of n, n, m, s and s random variables obeying the same law. For the sake of brevity let also

$$\mu_2(x_1) = \mu_2(x_2) = \dots = \mu_2(y_1) = \mu_2(y_2) = \dots = \mu_2(z_1) = \mu_2(z_2) = \dots = \mu_2$$

and similarly denote the fourth moment of all these variables simply by μ_4 . Then

$$\begin{aligned} \mu_2(v) = \mu_2(u) &= E \left[\sum_{i=1}^n x_i - E \sum_{i=1}^n \bar{x}_i \right]^2 = E \left[\sum_{i=1}^n (x_i - \bar{x}_i) \right]^2 = n\mu_2, \\ \mu_4(v) = \mu_4(u) &= E \left[\sum_{i=1}^n x_i - E \sum_{i=1}^n \bar{x}_i \right]^4 = E \left[\sum_{i=1}^n (x_i - \bar{x}_i) \right]^4 = \\ &= E \left[\sum_{i=1}^n (x_i - \bar{x}_i)^4 \right] + E \left[6 \sum_{i=1}^n \sum_{j=1, j \neq i}^n (x_i - \bar{x}_i)^2 (x_j - \bar{x}_j)^2 \right] = \\ &= \sum_{i=1}^n [E (x_i - \bar{x}_i)^4] + 6 \sum_{i=1}^n \sum_{j=1, j \neq i}^n [E (x_i - \bar{x}_i)^2 E (x_j - \bar{x}_j)^2] = \\ &= n\mu_4 + 3n(n-1)\mu_2^2. \end{aligned}$$

Similarly, when replacing n by m and s , we will have

$$\mu_2(w) = m\mu_2, \mu_2(x) = \mu_2(y) = s\mu_2,$$

$$\mu_4(w) = m\mu_4 + 3m(m-1)\mu_2^2, \mu_4(x) = \mu_4(y) = 3\mu_4 + 3s(s-1)\mu_2^2.$$

Substituting now the obtained expressions in formula (8) and denoting after Pearson $\beta_2 = \mu_4 / \mu_2^2$ we will have after some easy work

$$r_{11} = \frac{m}{s}, r_{22} = 1 + 2r_{11}^2 + \frac{r_{11}}{s}(\beta_2 - 3),$$

$$r_{40} = r_{04} = 3 + \frac{\beta_2 - 3}{s}, r_{31} = r_{13} = 3r_{11} + \frac{r_{11}}{s}(\beta_2 - 3).$$
(14)

For $\beta_2 = 3$ or finite β and $s = \infty$ these will again become the moments of the “normal” distribution. And, when substituting (14) into (6) and (7), we find that

$$\varepsilon_p = E\rho - r_{11} = -\frac{r_{11}(1-r_{11})[1+r_{11}+(\beta_2-3)/2s] + \dots}{2N},$$
(15)

$$\sigma_p = E(\rho - \bar{\rho})^2 = \frac{1}{\sqrt{N}} \sqrt{(1-r_{11}^2)^2 + \frac{\beta_2-3}{2s} r_{11}(1-r_{11})(2-r_{11}) + \dots},$$
(16)

i. e., formulas similar to (11) and (12). Their analysis on the lines of § 4 will lead to the same conclusions, but we leave it for the readers.

We only provide a numerical example. Let us take 330 cards, writing 0 on 81 of them; 1, on 49; 2, on 25; 3, on 9; 4, on 1; 5, on 1, and again, 6, 7, 8 and 9 on 9, 25, 49 and 81 cards respectively. When extracting these cards with replacement, the occurring randomly variable numbers will have values 0, 1, 2, ..., 8, 9 with probabilities proportional to

81, 49, 25, 9, 1, 1, 9, 25, 49, 81

(equal to 81/330, ...). They are also proportional to the squares of the deviations of the values of the variable from its arithmetic mean for an unlimited number of trials, 4.5. Calculation shows that that distribution is characterized by the following moments:

$$Ex = \bar{x} = 4.5, \mu_2(x) = 14.65, \mu_4(x) = 252.0625,$$

$$\beta_2 = \frac{\mu_4(x)}{\mu_2^2(x)} = 1.18, \beta_2 - 3 = -1.82.$$

When composing groups of 11 numbers from a sequence of the random values of such a variable, then composing N pairs of numbers by adding the first 10, and the last 10 of them, the terms of each pair will be correlated and, as proved above, the correlation coefficient will be 9/10. Determining now ε_p and σ_p by formulas (15) and (16) we will have

$$\varepsilon_p = -0.0814/N, \sigma_p = 0.164/\sqrt{N}.$$

Although the initial distribution, as we saw, was extremely peculiar and very remote from the “normal” law, they only differ from the same magnitudes for that distribution by 5% and 14% respectively. Even for a modest value of $N = 25$, the ratio of the systematic error to the random will only be about 1/10.

In general, the above apparently allows us to conclude that, for any theoretical value of the systematic error, the practical asinificance of that error of the empirical correlation coefficient, as revealed by the latest investigations, is happily either completely or almost of no consequence.

Appendix

Here is an illustration. I chose 1500 one-digit numbers separated in their order in groups of three⁵. Three vertical series, 500 numbers in each, were thus formed; suppose them to be the values of random variables u , w and v having possible values 0, 1, 2, ..., 8, 9. Assume that for each of the three series the probability of each value is 1/10 and does not depend either on the previous or subsequent values of the same variable or of the other ones. These three assumptions 1) Of a definite law of probability. 2) Of the constancy of that law, and 3) Of the independence of the variables, are apparently the most likely if not absolutely necessary.

Adding up in pairs the numbers of the first two rows, then the second and the third, we obtain the values of variables

$$x = u + w, y = v + w.$$

We also compile a column of the absolute values of $(x - y)$, directly calculate how many times each value of the [three new] variables does occur and obtain the following table (Table 1).

Now we have to calculate the correlation coefficient between x and y . As shown above, it should be equal to 1/2 if only the adopted hypotheses are obeyed. Denoting $\sum n_x x$ by $\sum n$ etc, and empirical magnitudes with an additional stroke above, we have

$$\sum x = 4453, \sum y = 4651, \sum x^2 = 47783, \\ \sum y^2 = 51729, \sum (x - y)^2 = 7354,$$

$$\sum xy = \frac{1}{2}[\sum x^2 + \sum y^2 - \sum (x - y)^2] = 46079,$$

$$\bar{x}' = \frac{1}{N} \sum x = 8.906, \bar{y}' = \frac{1}{N} \sum y = 9.302,$$

$$\mu_2'(x) = (\sigma_x')^2 = \frac{1}{N-1} [\sum x^2 - N(\bar{x}')^2] = 16.281727, \sigma_x' = 4.035062,$$

$$\mu_2'(y) = (\sigma_y')^2 = \frac{1}{N-1} [\sum y^2 - N(\bar{y}')^2] = 16.964726, \sigma_y' = 4.118826,$$

$$\mu_{11}(xy) = \frac{1}{N-1} [\sum xy - N\bar{x}'\bar{y}'] = 9.333054,$$

$$r'_{xy} = \frac{\mu'_{11}(xy)}{\sigma'_x \sigma'_y} = \frac{9.33054}{16.619720} = 0.56157.$$

For comprehending the significance of the obtained deviation from the theory, let us derive the corresponding theoretical constants. Consider a random variable with values 0, 1, 2, ..., 8, 9 with probabilities 1/10. Its constants [parameters] will coincide with the corresponding values of u , w , and v :

$$\begin{aligned}\bar{z} &= E(z) = \frac{1}{10}(0+1+ \dots + 8+9) = 4.5 = \bar{u} = \bar{v} = \bar{w}, \\ \mu_2(z) &= \sigma_z^2 = E(z - \bar{z})^2 = \frac{1}{10}[(0-4.5)^2 + (1-4.5)^2 + \dots + (9-4.5)^2] = \\ 8.25 &= \mu_2(u) = \mu_2(v) = \mu_2(w), \\ \sigma_z &= 2.872281 = \sigma_u = \sigma_v = \sigma_w, \\ \mu_4(z) &= E(z - \bar{z})^4 = \frac{1}{10}[(0-4.5)^4 + (1-4.5)^4 + \dots + (9-4.5)^4] = \\ 120.8625 &= \mu_4(u) = \mu_4(v) = \mu_4(w), \\ \beta_2 - 3 &= \frac{\mu_4(z)}{\mu_2^2(z)} - 3 = -1.22425.\end{aligned}$$

And, as shown above in a general way,

$$\begin{aligned}\bar{x} = \bar{y} &= 2E(z) = 9.0, \\ \mu_2(x) = \mu_2(y) &= 2\mu_2(z) = 16.50, \quad \sigma_x = \sigma_y = 4.062019, \\ \mu_4(x) = \mu_4(y) &= 2\mu_4(z) + 3.2\mu_2^2(z) = 650.1.\end{aligned}$$

Now, by applying known formulas we calculate the mean square errors

$$\begin{aligned}\sigma_{\bar{u}} = \sigma_{\bar{v}} = \sigma_{\bar{w}} &= \frac{\sigma_z}{\sqrt{N}} = 0.12845, \quad \sigma_{\bar{x}} = \sigma_{\bar{y}} = \frac{\sigma_x}{\sqrt{N}} = 0.18166, \\ \sigma_{\sigma_u} = \sigma_{\sigma_v} = \sigma_{\sigma_w} &= \sqrt{\frac{\mu_4(z) - \mu_2^2(z)}{\mu_2(z)}} : 2\sqrt{N} = 0.05657, \\ \sigma_{\sigma_x} = \sigma_{\sigma_y} &= \sqrt{\frac{\mu_4(x) - \mu_2^2(x)}{\mu_2(x)}} : 2\sqrt{N} = 0.10700.\end{aligned}$$

Finally, the mean square error of the correlation coefficient r_{xy} which is equal to 1/2 will be

$$\sigma_{r_{xy}} = \sqrt{\left[1 - \left(\frac{1}{2}\right)^2\right]^2 - \frac{\beta_2 - 3}{2 \cdot 2} \cdot \frac{1}{2} \left(1 - \frac{1}{2}\right) \cdot \left(2 - \frac{1}{2}\right)} : \sqrt{500} = 0.02992$$

with systematic error

$$\varepsilon_{r'_{xy}} = \text{Er}'_{xy} - r_{xy} = -\frac{1}{2500} \left[\frac{1}{2} (1-1/2)^2 + \frac{\beta_2 - 3}{2 \cdot 2} \cdot \frac{1}{2} (1-1/2) \right] = -0.00030.$$

This is absolutely insignificant. For an infinitely repeated series of 500 pairs of numbers x and y irreproachably complying with theoretical assumptions the arithmetic mean of the empirical values of the correlation coefficient should tend not to $1/2$, but to $(1/2 - 0.00030) = 0.49970$. The difference only constitutes 1% of the mean square error. Denoting the deviation of the empirical correlation coefficient from its theoretical value (otherwise, its *error*) by Δr_{xy} , we have

$$\Delta r_{xy} = 0.56157 - 0.49970 = 0.06187,$$

$$\frac{\Delta r_{xy}}{\sigma_{r_{xy}}} = \frac{0.06187}{0.02992} = 2.068.$$

The “error” of the empirical correlation coefficient so essentially exceeding its mean square error is actually not extraordinary. If our hypotheses were indeed valid, such an excess would have been considered a rather, but not excessively rare random occasion. For the “normal” distribution of errors, which for $N = 500$ can be admitted as a first approximation, the probability of an error equal or larger in absolute value is $3.8/100$. However, had there been no prior considerations for our hypotheses, that result should have led us to conclude that the probability of their being valid is rather low.

Let us attempt to reveal the possible cause for the deviations of the empirical correlation coefficient from the indications of the theory. We express r'_{xy} in the following way

$$\begin{aligned} \sigma'_x \sigma'_y r'_{xy} &= \frac{1}{N-1} \sum [(u-w) - (\bar{u}' + \bar{w}')] [(v+w) - (\bar{v}' + \bar{w}')] = \\ &= \frac{1}{N-1} \sum [(u - \bar{u}') + (w - \bar{w}')] [(v - \bar{v}') + (w - \bar{w}')] = \\ &= \frac{1}{N-1} \sum [(w - \bar{w}')^2 + \frac{1}{N-1} \sum (u - \bar{u}') (v - \bar{v}') + \\ &= \frac{1}{N-1} \sum (u - \bar{u}') (w - \bar{w}') + \frac{1}{N-1} \sum (v - \bar{v}') (w - \bar{w}'), \\ \sigma'_x &= \sqrt{\frac{1}{N-1} [\sum (u - \bar{u}')^2 + \sum (w - \bar{w}')^2 + 2 \sum (u - \bar{u}') (w - \bar{w}')]}, \\ \sigma'_y &= \sqrt{\frac{1}{N-1} [\sum (v - \bar{v}')^2 + \sum (w - \bar{w}')^2 + 2 \sum (v - \bar{v}') (w - \bar{w}')]} \end{aligned}$$

and therefore

$$r'_{xy} = \frac{\sigma_w'^2 + \sigma_u' \sigma_v' r'_{uv} + \sigma_u' \sigma_w' r'_{uw} + \sigma_v' \sigma_w' r'_{vw}}{\sqrt{\sigma_u'^2 + \sigma_w'^2 + 2\sigma_u' \sigma_w' r'_{uw}} \sqrt{\sigma_v'^2 + \sigma_w'^2 + 2\sigma_v' \sigma_w' r'_{vw}}}$$

For independent u , v and w the expectations of r'_{uv} , r'_{uw} , and r'_{vw} vanish (Chuprov 1922, p. 248/2004, end of § 2.2A), and for a more or less considerable N , r'_{xy} will be sufficiently precisely, to the order of its systematic error, $\varepsilon_{r'_{xy}}$ equal to

$$\frac{\sigma_w'^2}{\sqrt{\sigma_u'^2 + \sigma_w'^2} \sqrt{\sigma_v'^2 + \sigma_w'^2}}$$

If now u , v and w have a single common law of distribution remaining constant during the entire empirical series, then $E\sigma_u'^2 = E\sigma_v'^2 = E\sigma_w'^2$ and, again for a sufficiently large N , we will have sufficiently precisely

$$Er'_{xy} = \frac{\sigma_w'^2}{2\sigma_w'^2} = \frac{1}{2}$$

Neglecting errors of the order of the systematic error of the correlation coefficient, we thus conclude that any deviation of the empirical correlation coefficient from theoretical indications can result from three and only three causes. *First*, random errors of r'_{uv} , r'_{uw} , r'_{vw} on the one hand, and of σ_u' , σ_v' , σ_w' on the other hand. *Second*, non-random “errors” of r'_{uv} , r'_{uw} , r'_{vw} , i. e., their non-random deviations from zero owing to connections between u , v and w . *Third*, non-random “errors” of σ_u' , σ_v' , and σ_w' caused by the difference between the laws of distribution of those variables. When analysing empirical data, these causes can be distinguished one from another and the sources of “error” in the r'_{xy} revealed, sometimes practically certainly, but sometimes only more or less probably, depending on the magnitude of the corresponding deviations. If these deviations are not clearly expressed, the problem can certainly remain entirely open.

Turning now to our numerical series, we can compile anew a table of the distribution of the values of u , v and w (Table 2). We are then able to calculate [...].

Calculating now the other general characteristics of the distribution of our variables, we compare them with each other and with the earlier derived corresponding magnitudes for x and y , see Table 3. We are now able, first, to establish the source of the error of r'_{xy} . Noting that the product $\sigma_x' \sigma_y'$ exceeds its theoretical value, we see that $\sigma_w'^2 / \sigma_x' \sigma_y'$ exceeds 0.5 only because of the rather essential positive error of σ_w' . It led to r'_{xy} deviating by 0.035201, or by 7.04% of its theoretical value. The other part of the error (5.27% of the same magnitude) occurred because of a positive correlation between u and v on the one hand and, on the other hand, between v and w .

How essential are all these deviations? Or, how high is the probability of a random origin of their entire totality? Restricting our attention to nine

magnitudes, \bar{u}' , \bar{v}' , \bar{w}' , σ'_u , σ'_v , σ'_w , r'_{uv} , r'_{uw} and r'_{vw} whose errors should be mutually independent for independent u , v and w , we can apply the Pearson criterion (Slutsky 1912, pp. 192 – 193/2009, § 43). Namely, squaring the ratio of the “errors” to the corresponding mean square errors (Table 3, last column), and adding up those squares, we have $\chi^2 = 17.35$. For nine independent variables, i. e., for $n' = 10$, the Elderton table [1902] leads to $P = 0.044$.

A probability of such an order does not empower us to any decisive conclusions. There is nothing special in that we have accidentally encountered a combination of random numbers naturally occurring once in roughly 23 cases. On the other hand, had there been no prior considerations for our hypotheses about the chosen numbers, it would still have been more probable that they, the hypotheses, did not quite comply with reality.

It is not easy to see anywhere something absolutely random, and there is nothing inconceivable in that the chosen one-digit numbers were slightly connected with each other owing to the connection between the four- and five-digit numbers of births and deaths in those consecutive years which they describe. That curious problem demands, however, a special investigation beyond the framework of the present paper.

Explanation of Tables

I am only explaining the three tables, all of them in the Supplement. O. S.

Table 1. Magnitudes x and y , as explained in text, are sums of uniformly distributed random variables and have possible values $0(1)18$. The Table provides the number of the occurrences of each of those values, then the same for $|x - y|$.

Table 2. Magnitudes u , v and w are the one-digit empirical numbers chosen from the *Jahrbuch* (1913). The Table provides the number of occurrences of each of their values, separately for all three of those variables.

Table 3. It provides the theoretical and empirical values of the arithmetic means of u , v , w and of their standard deviations; the same for r_{uv} , r_{uw} and r_{vw} , and for x , y and r_{xy} . The main column shows the difference between the empirical and theoretical values of each of the mentioned magnitudes divided by their appropriate theoretical value.

Notes

1. Chuprov stated that for random variables to be mutually independent, the law of distribution of any one of them ought to persist whichever possible values are taken by the other variables. According to the modern definition, the densities $f(x_i)$ of random variables ξ_i and the density $f(x_1, x_2, \dots, x_n)$ of $\{\xi_i\}$ should then obey the condition

$$f(x_1, x_2, \dots, x_n) = f(x_1) f(x_2) \dots f(x_n)$$

with a similar condition imposed on their distribution functions. O. S.

2. Indeed, if $u + w = \xi$, $v + w = \eta$, then $x = A\xi$, $y = B\eta$. But

$$E(x - \bar{x})(y - \bar{y}) = ABE(\xi - \bar{\xi})(\eta - \bar{\eta}), \quad \sigma_x = A\sigma_\xi, \quad \sigma_y = B\sigma_\eta,$$

therefore

$$r_{xy} = r_{\xi\eta} = \frac{\sigma_w^2}{\sigma_\xi \sigma_\eta},$$

QED. E. S.

3. This is unclear. O. S.

4. I have written x', y', \dots instead of the obviously mistaken x^1, y^1, \dots . The same mistake occurred in the beginning of the Appendix in spite of Slutsky's direct indication that he is applying strokes. O. S.

5. Slutsky mentions the source, the tables of births and deaths in the *Jahrbuch* (1913), from which he had taken his numbers, and explains how he had chosen them, 1499 in all, from it. O. S.

We repeated one digit, a 9, and obtained 1500. In one case, 4 was mistakenly chosen instead of 5. Having in mind the goal of the experiment, the ensuing errors can be thought as of no consequence. E. S.

Bibliography

Chetverikov, N. S. (1921, in Russian), The theory of the Darbishire experiments. *Vestnik Statistiki*, No. 5 – 8, pp. 250 – 254.

Chuprov, A. A. (1922, in Russian), On the expectation of the ratio of two mutually dependent random variables. *Trudy Russkikh Uchenykh za Granitsei*, vol. 1. Berlin, pp. 240 – 271. Translation in Chuprov, A. A. (2004), *Statistical Papers and Memorial Publications*. Berlin. Also at www.sheynin.de

--- (1923), Aufgaben und Voraussetzungen der Korrelationsmessung. *Nordisk Statistisk Tidsskrift*, Bd. 2; No. 1, pp. 24 – 53.

--- (1925), *Grundbegriffe und Grundprobleme der Korrelationstheorie*. Leipzig – Berlin.

Darbishire, A. D. (1907), *Some Tables for Illustrating Statistical Correlation*. *Mem. and Proc. Manchester Lit. & Philos. Soc.*, vol. 51, No. 16.

Elderton, W. P. (1902), Tables for testing the goodness of fit of theory to observation. *Biometrika*, vol. 1, pp. 155 – 163.

Jahrbuch (1913), *Statistisches Jahrbuch der Stadt Berlin*, 32. Jg. Berlin.

Prokhorov, A. V. (1999), Correlation. In Prokhorov, Yu. V., Editor (1999), *Veroiatnost i Matematicheskaia Statistika. Enziklopedia* (Probability and Math. Statistics. An Enc.). Moscow, pp. 264 – 265.

Sheynin O. (1990, in Russian), *Chuprov. Life, Work, Correspondence*. Göttingen, 1996.

Slutsky E. E. (1912, in Russian), *Teoria Korreliatsii*. English translation: *Theory of Correlation*. Berlin, 2009. Also at www.sheynin.de Cf. [i].

Yule, G. U. (1911), *Introduction to the Theory of Statistics*. London, 1912.

V

On a New Coefficient of Mean Density of Population

O novom koeffiziente srednei plotnosti naselenia.
Vestnik Statistiki, No. 4 – 6, 1923, pp. 5 – 19

1. The ratio of population to the area of the territory it occupies, if distributed uniformly, is the density of population; otherwise, it is the mean density. However, for any considerably irregular distribution that latter number absolutely wrongly characterizes the conditions of life of the population.

Let us consider a fictitious example. Given a country consisting of two regions with populations 10 *mln* and only 99 thousand and areas 100 thousand and 9,900 thousand *sq km* respectively. The densities will be 100 and 0.01 whereas the mean density is 1.0099. In any comparison with the more uniformly populated countries, that last-mentioned figure will only mislead because almost 99% of the population, or, in other words, all the population is certainly living, generally speaking, under the same conditions of density as the inhabitants of other countries having density 100.

It might be argued that the inadequacy of the mean density is a property common to all other mean magnitudes. This, however, is wrong, as I will now show. More precisely, without denying that a mean value as such possesses some shortcomings, we will see that the mean density is a number formed logically correctly, but illogically applied.

Let us have some territory divided into small districts with differing densities. Marking off the density on the *x*-axis, and dividing the entire interval into parts, we can construct rectangles with small bases with their areas proportional, first, to the sum of territories inhabited with the corresponding density; or, second, to the *population*. We thus obtain diagrams showing the *distribution of the territory*, or of the *population according to the density*. The mean values of the indications will be absolutely different since the weights will be areas in the first case, and populations, in the second.

Let the areas be s_1, s_2, \dots, s_m with populations n_1, n_2, \dots, n_m and densities c_1, c_2, \dots, c_m . Then the mean densities will be respectively¹

$$\bar{c} = \frac{[cs]}{\sum_{i=1}^m s_i} \equiv \frac{N}{S}, \quad \bar{\gamma} = \frac{[cn]}{\sum_{i=1}^m n_i} \equiv \frac{[cn]}{N}. \quad (1; 2)$$

Here, N is the population of the territory, and S , its area. The second mean can also be calculated as

$$\bar{\gamma} = \frac{1}{N} \sum_{i=1}^m \frac{n_i^2}{s_i} = \frac{1}{N} [cs]. \quad (3)$$

The logical mistake usually made is that the mean density of the first kind, i. e. the mean density or the arithmetic mean of the densities is

calculated as an *indicator of the territory*, but applied for characterizing population and the conditions of life for which the *arithmetic mean of densities as an indication of the population itself* would have been required. To explain this on a simplest example: densities 10 and 20 lead to mean density 15, but the first case is valid for equal areas uniformly populated with the above densities; and the second instance, when we have equal populations in both regions.

We may distinguish between the two means calling them *mean physical* and *mean social* densities, respectively. Let us calculate the latter for our fictitious example above:

$$\bar{\gamma} = \frac{100 \cdot 10 \cdot 10^6 + 0.01 \cdot 99 \cdot 10^3}{10.099 \cdot 10^6} = 99.02,$$

a number very close to the density for the prevailing number of people. Under the conditions stated, that should seem to be quite natural.

In concluding, it would perhaps not be amiss to note the connection between the two means. Formula (3) provides

$$\bar{\gamma} = \frac{\sum_{i=1}^m c_i^2 s_i}{N} = \frac{1}{\bar{c}} \frac{1}{S} \sum_{i=1}^m c_i^2 s_i = \frac{m_{2(c)}}{\bar{c}}. \quad (4)$$

In other words, the social density of the population is equal to the mean square density with weights being the corresponding areas divided by the mean physical density. Or, denoting as usual the mean square deviation of density by σ_c and the coefficient of variation, i. e. σ_c / \bar{c} , by v_c , and taking into account the known identity

$$m_{2(c)} = \sigma_c^2 + \bar{c}^2,$$

we have

$$\bar{\gamma} = \frac{\sigma_c^2 + \bar{c}^2}{\bar{c}} = \bar{c} + \sigma_c v_c \quad (5)$$

from which, in particular, it follows that always $\bar{\gamma} \geq \bar{c}$.

It is also evident that for these magnitudes to be equal to each other it is necessary that $\sigma_c = 0$, i. e., that $c_1 = c_2 = \dots = c_m$. In all cases, in which the social density is calculated for territories with differing densities, $\bar{\gamma} > \bar{c}$.

2. The above method of calculating the mean social density is only applicable without restrictions when the partial regions s_1, s_2, \dots, s_m are populated absolutely uniformly. Otherwise that mean only provides an approximate magnitude whose logical meaning should yet be ascertained and whose precision is not yet known. On the face of it, it is only clear that the more uniformly is the population distributed over each district, the closer is the calculated magnitude to the real mean (social) density, and the other way round. But what is that real mean?

Suppose that the population is a continuous function of the coordinates of the appropriate surface. This means that, for each element of area Δs taken about an arbitrary point and having population Δn , there should exist the limit of the ratio $\Delta n/\Delta s$ as Δs tends to zero. Then that limit equal to γ will be the density at a given point, γds will be the population of an infinitely small element of area ds and the mean social density will be represented by the integral

$$\bar{\gamma} = \frac{1}{N} \int \gamma^2 ds \quad (6)$$

taken over all the appropriate area.

Now, another case. Let

$$\Delta_1 s > \Delta_2 s > \dots \Delta_i s > \dots$$

be a number of elementary areas situated about some point and unboundedly approaching zero. It can occur that, although the population is not a continuous function of the area, the ratios

$$\Delta_1 n/\Delta_1 s, \Delta_2 n/\Delta_2 s, \dots, \Delta_i n/\Delta_i s, \dots, \Delta_n n/\Delta_n s, \dots$$

constitute a so-called semi-convergent series. That means that its terms, beginning with some i , will ever nearer approach some constant magnitude C so that the [absolute] difference between them becomes less than some somehow reasonably assigned error, but, beginning with some farther subscript l , these terms move away from that quasi-limit C . We may agree to consider that magnitude as the density of the population at the appropriate point and formula (6) will again provide the social mean density for the entire territory. From the logical side, the situation here is about the same as in the similar concepts of statistical physics.

We will attempt, however, to regard our issue from another viewpoint, and to formulate the following problem:

Is it possible to attach a meaning to the concept of mean density logically consistent with any distribution of the population and admitting quite an adequate approach to such an essentially discrete function as the inhabitation of a territory?

I believe that that will be possible after the concept of density of population be subjected to a certain logical transformation. When the density is calculated in the usual way, individuals are treated as some absolutely abstract units without at all considering their relations with the territory, such as work, mobility, ownership etc. Only one aspect is taken into account, viz., whether the given individual is within the boundaries of a certain territory. And, if the territory is separated by several systems of successive partitions, such as provinces, districts, etc., only the smallest of those is always considered as though the inhabitant of the district, say, has no relations with the larger parts. When calculating the mean (physical) density, obvious absurdities are not met with although some difficulties sufficiently described in geographical literature do occur. For example, if a man lives in one district, but works elsewhere, to which population should he be attached? And, how should we deal with uninhabited, or very sparsely

inhabited areas (forests, highland pastures, various wastelands, etc.) interspersed among populous tracts?

Due to the insufficiency of its logical basis, the concept of mean social density suffers considerably greater. Indeed, when calculating the mean physical density, a move of an individual from one sub-region to another one as well as the change of the boundaries between them are only reflected in the mean densities of these districts whereas the mean density of the entire territory does not change. At the same time, however, after any change of those boundaries, the mean social density generally takes a new value. Since the interior boundaries are arbitrary, the mean social density lacks objective meaning.

This can be explained by a *reductio ad absurdum* if we understand the usual concept as the restriction of the individual's relation with the territory only to his "being within". Separate all the territory into such tiny plots that not more than one individual inhabits each of them. Then, applying the general principle of calculation, we will obtain a zero density for each uninhabited plot and the same [positive] density for the other ones. For plots of 2, 1, 0.5sq. *m* we will find elementary densities and, therefore, mean social densities equal to 1/2, 1, 2 people/sq. *m* depending on our arbitrary choice rather than on some objective conditions.

3. For establishing a logically well-grounded concept of density of population we ought to attach to it the meaning of some relationship between individual and territory and derive just as many versions of that concept as reasonable relationships will be taken into account. Within the context of my paper it is not important which of these will be useful for the practitioner or be theoretically interesting.

Let us consider some possible relation *R* between an individual and a point within the pertinent territory. Suppose that all the points connected with some individual by *R* constitute one or several entire plots and call it, or their totality, *the region inhabited by him with respect to R*, or, shorter, his region. Consider at first the case in which the relation *R* does not admit of any quantitative gradations, that for the individual there is no difference between the points of his region. Then we may say that such a region is uniformly inhabited by him, and its density will obviously be 1/*S* where *S* is its area. As to the population, if other individuals are lacking, it is equal to 1 for the entire region; for its half it is 1/2, and for any part of the region with area *s* it is *s/S*.

Let us take into consideration all the individuals inhabiting some territory. Determine the boundaries of the regions where each of them is living and consider one of the regions formed by the intersection of such boundaries but not being intersected by any of them. All the points of each region will be homogeneous in the sense that each point is conjugated by the relation *R* with one and the same number of the same individuals. The area of a region will possess one and the same density in all its parts and we will call such regions *simple parts* of the territory.

Suppose that some territory is divisible in ω simple parts with areas $s_1, s_2, \dots, s_\omega$ and let part s_i be "inhabited" by m_i individuals whose regions are S_{i1}, S_{i2}, \dots . Then, according to the above, the total population of such a simple part will be

$$v_i = s_i/S_{i1} + s_i/S_{i2} + \dots \text{ (} m \text{ terms)} \quad (7)$$

and its density

$$\gamma_i = 1/S_{i1} + 1/S_{i2} + \dots \text{ (} m \text{ terms).} \quad (8)$$

Obviously, $\sum s_i = S$ and $\sum v_i = N$ and we may apply formulas (2) and (3) for calculating the mean social density. Then

$$\bar{\gamma} = \frac{[\gamma v]}{N} = \frac{\sum_{i=1}^{\omega} \gamma_i^2 s_i}{N} = \frac{\sum_{i=1}^{\omega} \frac{v_i^2}{s_i}}{N}. \quad (9)$$

Suppose also that we may subdivide all the simple parts into elementary plots having one and the same area Δs . Then, since the densities in all the separate simple parts are the same, we will have for each such part

$$\sum_s \gamma^2 \Delta s = \gamma_i s_i \quad (10)$$

and therefore

$$\bar{\gamma} = \frac{1}{N} \sum_s \gamma^2 \Delta s. \quad (11)$$

Since, in the general case, the areas of the simple parts are incommensurable, formula (11) can only be approximate. However, in the limit we will always have the precise expression

$$\bar{\gamma} = \frac{1}{N} \sum_s \gamma^2 ds. \quad (12)$$

All the previous considerations can be easily generalized to the case in which the relation of the individual to the territory is of a quantitative nature. Then, for each individual there will exist some function of the density describing his inhabiting an appropriate territory with respect to R , and the density of population at every point will be equal to the sum of these partial densities. Formula (12), or, approximately, (11), will then represent the mean social density of all the territory.

Example. Ten people, 6 prisoners and 4 warders, are walking over a plot of, say, 1 *desiatina*². The latter walk all over the plot, whereas the former are free to walk over 1/4 of it. Then the “population” of the smaller plot is $6 + 4(1/4) = 7$ people and the density is $7:(1/4) = 28$ people. The rest part of the plot, only available for the warders, has population $4 \cdot 3/4 = 3$ people and density $3:3/4 = 4$. The mean social density is $28 \cdot 7 + 4 \cdot 3 = 20.8$ as against the mean physical density of 10 people.

4. Both the population and the density can change not only in space but over time as well and the same two concepts have to be distinguished now also. When calculating the mean density and assigning the appropriate time intervals as the weights, we obtain the physical mean; if, however, we admit

instead the *time during which the population has lived* [in the territory] given the corresponding densities, we get the mean social density.

Suppose that the constant population of a coastal zone with area 10sq. km is 10 people, but that during 1/50 of a year this zone is being overflowed with an industrial army 10 thousand strong. Then the zone's density will be 1,001 during 1/50 of a year and 1 man during the rest of the year. Calculating the mean with weights 1/50 and 49/50 we have

$$\bar{c} = 21.0 \text{ people.} \quad (13)$$

If, however, we take into account that during the first period 10,010 people will live $10,010(1/50) = 200.2$ man-years, and during the second period the 10 people will live $10(49/50) = 9.8$ man-years, the mean social density will be

$$\bar{\gamma} = \frac{1001 \cdot 200.2 + 1 \cdot 9.8}{200.2 + 9.8} = 954.3 \text{ people.} \quad (14)$$

The first number (13) is the arithmetic mean as an indicator of the territory, almost uninhabited for the greater part of the year, whereas the second mean (14) is an indicator of the conditions of life of certain social masses, of 10,010 people during 1/50 of a year, and of 10 people during 49/50 of a year. Each of these means has its own special meaning and they cannot replace one another.

It is of certain interest to consider in general the mean social density for an entire period. Suppose that t_1, t_2, \dots, t_k are elementary periods of time, $s_1, s_2, \dots, s_\omega$, elementary plots of the territory, $v_{i1}, v_{i2}, \dots, v_{ik}$, and $\gamma_{i1}, \gamma_{i2}, \dots, \gamma_{ik}$, the populations and densities for the i -th plot during the appropriate time intervals. Let, in addition, l_{ij} be the total time of life passed by the population of the i -th plot during the j -th interval of time. Then, denote by \bar{v}_i and $\bar{\gamma}_i$ the mean population and density for the whole period of time

$$T = t_1 + t_2 + \dots + t_k$$

for the i -th plot, and by l_i the total time passed by its population during period T . We also ought to denote the total time passed by all the population during all the period by L , the mean population of the territory over the entire period by \bar{N} and the mean social density for all the territory and the whole period by $\bar{\gamma}_T$. Similar magnitudes for the whole territory over the j -th elementary interval of time will be L_j, N_j and $\bar{\gamma}_{ij}$.

Define the mean social density over time period T as

$$\bar{\gamma}_T = \frac{1}{L} \sum_{i=1}^{\omega} \sum_{j=1}^k l_{ij} \gamma_{ij}. \quad (15)$$

It is also possible to represent this expression in one of the following ways. *First*, when summing over time, we obtain

$$\bar{\gamma}_T = \frac{1}{L} \sum_{i=1}^{\omega} l_i \bar{\gamma}_i. \quad (16)$$

In other words, the mean social density for all the territory over a certain period of time is obtained by calculating those densities over the whole period for each elementary plot and then taking the mean with weights equal to the time passed by the inhabitants of each plot over all the period.

Second, when summing over all the area for each of the separate elementary time periods we get

$$\bar{\gamma}_T = \frac{1}{L} \sum_{j=1}^k t_j [N_j \frac{1}{N_j} \sum_{i=1}^{\omega} v_{ij} \gamma_{ij}] = \frac{1}{L} \sum_{i=1}^k L_j \bar{\gamma}_{ij}. \quad (17)$$

Thus, the same mean social density can be determined when deriving it for all the territory over each elementary period of time and then taking their mean with weights equal to the time passed by all the population during each of those periods.

Third, let us denote by $m_{2(\gamma_i)}$ the mean square of the density for the i -th plot over all the period with weights simply equal to the duration of the corresponding elementary periods:

$$m_{2(\gamma_i)} = \frac{1}{T} \sum_{j=1}^k \gamma_{ij}^2 t_j. \quad (18)$$

Now, noting that

$$l_{ij} = t_j v_{ij} = t_j \gamma_{ij} s_i$$

we will represent expression (15) in the following way:

$$\bar{\gamma}_T = \frac{1}{L} \sum_{i=1}^{\omega} \sum_{j=1}^k \gamma_{ij}^2 t_j s_i = \frac{T}{L} \sum_{i=1}^{\omega} [\frac{1}{T} \sum_{j=1}^k \gamma_{ij}^2 t_j].$$

Since

$$\frac{L}{T} = \frac{1}{T} \sum_{j=1}^k N_j t_j = \bar{N},$$

we finally obtain

$$\bar{\gamma}_T = \frac{1}{\bar{N}} \sum_{i=1}^{\omega} m_{2(\gamma_i)} s_i. \quad (19)$$

This expression (19) is curiously similar to formula (3). The latter is transformed into it if the square of the density for an elementary plot is replaced by the physical mean square of the densities for each plot over all the time period and the population is understood as the *mean population* over the same period.

5. My aim does not include the study of the methodological issues that will inevitably arise when any attempt is made to apply the general theoretical patterns to concrete empirical material. This constitutes a problem for geography rather than for theoretical statistics and I restrict my further exposition to a few remarks.

It is clear that only rarely and, for that matter, in absolutely exceptional cases, the researcher will be able to determine the density by issuing from the distribution of each individual among the appropriate territory. In general, even for the most detailed investigation, it will be necessary to confine the study to establishing social masses more or less homogeneous with respect to their relation with the territory, as well as of rather homogeneous regions. As a first approximation, it is possible to indicate the following.

The researcher should single out the areas of a) settlements of the urban type; b) agricultural settlements and those belonging to them complete with tracts of land under usual agricultural use; he should isolate areas c) with extremely rare permanent populations (forests, steppes, highland pastures); and, finally, those d) having absolutely no populations (swamps, sands, etc.).

The density for areas a) is calculated under the assumption that all the inhabitants uniformly use the entire area of the appropriate settlement, and nothing else. For areas b), again assuming a uniform use, but also allowing for the time when the population is working outside its territory (in logging areas, on summer pastures, and, also seasonally, in cities). For areas c), taking into account both the sparse permanent and the more dense temporary populations. In areas d) the density should be assumed nonexistent.

Only such investigations of typical regions spread over a number of small tracts can give us an idea about the real distribution of the population and the social mean densities of the separate tracts. After all the territory of some country is covered by a sufficient number of such test tracts, we can also hope to obtain, according to the principles of sampling, both the overall picture of the distribution of the social densities and the mean social density for the entire country. In those countries, where, as for example in Germany, many researchers have already been working with great persistence and considerable input in that direction, [even] on the scale of the smallest administrative units, the preparatory stage for calculating the mean social density is partly concluded. We, however, were hardly engaged in solving such problems and should begin almost from the beginning.

6. The following example based on slightly generalized data on the Moscow district for ca. 1910 illustrates how different can the mean physical and social densities be. The area of the territory is $2,670 \text{ sq. km}$, and population 2 mln of which 1.5 mln live in the main city of the district. For the sake of simplicity we will suppose that all the rest (0.5 mln) constitute the rural population uniformly distributed over the whole area.

The density of the population in Moscow is 158 thousand and in rural areas $500 \cdot 10^3 / 2670 = 187$ people. The mean social density is

$$\frac{158 \cdot 10^3 \cdot 1.5 + 187 \cdot 0.5}{1.5 + 0.5} = 115.2 \text{ thousand people}$$

whereas the mean physical density is $2 \cdot 10^6 / 2670 = 749$ people.

Given such a great difference between the densities for the urban and rural populations, a single mean density, be it physical or social, cannot provide a complete picture. In particular, the means for these groups of inhabitants should be additionally calculated. However, our example suggests other considerations as well.

It is impossible to derive the mean social density of the population of a country by issuing from mean physical densities for any large sub-territories, such as, for instance, our provinces or even districts and disregarding either absolutely uninhabited tracts (the tundra in the far North, the sands in Central Asia [now being beyond Russia] or the compact gatherings of the population over insignificant urban areas. Therefore, the mean density, calculated, say, for the Russian Empire in 1915 as the mean of the district data with weights, equal for example to the appropriate populations, cannot at all be considered as any, even rough approximation to the real value of the mean social density.

Such an average possesses nevertheless its own, special logical meaning distinct both from that of the mean physical and mean social density. Indeed, after finding out³ that there are, per square kilometre, in Belgium, 240.0; in England, 138.7; in Germany, 112.2; and in France, 73.0 people, we, for all that, learn something, although imperfectly, about the conditions of life in those countries.

In other words, the mean physical density can (with certain reservations and cautiously) be considered as a numerical measure of the conditions of life of the appropriate social masses. It is therefore absolutely logically justifiable to introduce such a notion as the mean of physical densities but calculated with weights representing the populations of the appropriate countries or their parts. It is easy to indicate the deficiencies of that notion, but all that may be stated against it is based on the shortcomings of its foundation, i. e., of the usual mean (physical) density. Unless and until it is considered possible to apply that mean for characterizing the conditions of life, the mean (physical) density, calculated, however not for the mean square kilometre, but for the mean individual, possesses, as I think it does, its *raison d'être*.

7. As an example, let us consider the density of the population in Russia in 1915. It is strange to discover⁴ that, contrary to the data for the countries of Western Europe (§ 6), the density for Russia as a whole was represented by the number 8.3. After considering everything that was said above, it seems absolutely doubtless that that number will not do at all for comparison. If we desire to have a *single* number for all the population of Russia it can only be the social-physical mean of the provinces or districts⁵.

It is quite natural that the mean density with weights assigned in accordance with the appropriate population is 55.7 instead of 7.76 as the same data provide by the usual method. And it is not difficult to establish that 89.0% of all the population of Russia was living with a higher density. Then, when calculating the mean density weighted in accordance with the population, we obtain⁶

5% of the population,	mean density 293.2
10%,	198.1
25%	125.3 [...]

While adducing these numbers only as an illustration, I do not provide any comment.

Notes

1. Unlike Slutsky, I introduce here and below the Gauss notation such as

$$[ab] = a_1b_1 + a_2b_2 + \dots + a_nb_n. \quad \text{O. S.}$$

2. An old Russian unit of area, or, actually, several essentially differing from each other units. For Slutsky's fictitious example, this fact is not really important. O. S.

3. *Statisticheskii Ezegodnik Rossii* (Russian Statistical Yearbook) for 1915, p. 59. E. S.

4. *Ibidem* (Note 3), pp. 58 – 59. E. S.

5. I only adduce Slutsky's final result. Issuing from the same source (Note 3), he arrived at the following conclusion:

5% of population occupied 0.16% of territory, density 112.0
10 and 25%, occupied 0.55 and 2.1% of the territory respectively, density 93.4 and 66.2,
etc. O. S.

6. Here, Slutsky adduced a note explaining the details of his calculations. O. S.

Bibliography

Valentei, D. I., Editor (1985), *Demografichesky Enziklopedichesky Slovar* (Demographic Enc. Dict.). Moscow.

VI

On Calculating the State Revenue from the Emission of Paper Money

K voprosu o vychislenii dokhoda gosudarstva ot emissii.
Mestnoe Khoziastvo (Kiev), No. 2, 1923, pp. 39 – 62. Appended to Iasnopolsky (1923)

The following lines have the task to study the methods of calculating the state revenue from issuing paper money and the real value of this operation. In another logical context it would have been a part of this construction to relate it to a formal theory of the issue of money, but a justification of this point of view does not enter into our present task. The aim of this sketch is mostly methodological. We check some methods of treating the data pertaining to money emission adopted in our literature and outline some new approaches. Although the lack of space does not enable us to exhaust the subject even to some extent, we nonetheless hope that our attempt will not be useless for those engaged in a scientific study of money emission.

1. Proceeding to our problem, we begin by establishing which magnitudes ought to be investigated. We denote the quantity of all paper money issued up to a certain moment t by u_t , or, more briefly, by u . This is what is called the *grand issue* during all the time until the given moment. The quantity of paper money being *en route* [in transit] and in bank tills we shall designate by w_t or w . The quantity of paper money actually put into *circulation* to moment t , i.e. actually paid out from the bank tills, shall be m_t or simply m .

If the banks' working cash might be neglected – that is, if we could assume that the banknotes transferred from the reserve fund to the circulating fund began generally circulating in the national economy at the moment of the transfer – then w_t would coincide with the so-called *small issue*. The following relation obviously exists between the three magnitudes introduced above:

$$u_t = m_t + w_t \quad (1)$$

In our notation, the issue during some time interval, for example for the period from $t = 0$ to $t = 1$, will be: the grand issue, $(u_1 - u_0)$; and the small issue, $(m_1 - m_0)$. As to the real value of the total mass of all the paper money issued up to a given moment, it is determined either by multiplying the pertinent magnitude by the value of a monetary unit at the moment, p_t , or by dividing it by the so-called index of the given moment i_t , since, indeed,

$$i_t = (1/p_t), p_t = (1/i_t). \quad (2)$$

Denoting now the real value of all the grand, and the small issue by U_t and M_t respectively, we thus obtain

$$U_t = p_t u_t = u_t/i_t, M_t = p_t m_t = m_t/i_t. \quad (3; 4)$$

The quantity of paper money *put into circulation* (m_t) does not coincide with that *actually circulating* because a part of it is lost for whatever reason, and another part, even though a small one, is hoarded. However, we will assume that the quantity of circulating money changes proportionally to m_t

which, for short periods of time, is certainly close to reality and might be admitted as a first approximation. The reader will see that this assumption is of a very restricted import for the sequel. Indeed, we only need the simplest case, in which the level of prices changes proportionally to the quantity of circulating money (and, consequently, proportionally, on the strength of this assumption, to the quantity of the issued paper money m_t), as some tentative norm and a convenient transitional link to a more complicated and more close to reality theoretical construction.

2. We will understand the state revenue from the emission during a certain period of time as the real value of the paper money at the moments when the pertinent issues had been put out into circulation. Denote the revenue by J and suppose that small issues $\Delta_1 m, \Delta_2 m, \dots, \Delta_n m$, were made many times and that each time the sums $\Delta_i m$ were put out into circulation as a single whole. If the value of money at the moments of the issues was p_1, p_2, \dots, p_n , then

$$J = p_1 \Delta_1 m + p_2 \Delta_2 m + \dots + p_n \Delta_n m, \quad (5)$$

$$\Delta_1 m + \Delta_2 m + \dots + \Delta_n m = m_t - m_0. \quad (6)$$

The problem of calculating the revenue is encountered because both the separate sums $\Delta_i m$ and the corresponding values of the monetary unit are unknown.

We will first determine the revenue J under the simplest hypothesis that the value of money is inversely proportional to the total amount of money put out into circulation m so that the value of that amount remains constant:

$$p_0 m_0 = p_1 m_1 = \text{Const.} \quad (7)$$

Various assumptions may be adopted concerning the separate issues of paper money. It might be supposed that the money was issued in equal portions; that the issues constituted a geometric progression; or, finally, that the issue was going on continuously. The last-mentioned supposition is the most convenient for calculations and its deviation from the real situation is negligible. Indeed, the state revenue is realized at the moments when money is paid out to civil servants, contractors, etc. Thousands of banks are paying out moneys, negligible as compared with the total mass of a monthly issue, during different hours of a business day and the general picture is that of an almost continuous current.

Suppose that an elementary issue is dm , and its value, pdm . Then, owing to our assumption of an inverse proportionality between p and m ,

$$p = M_0/m. \quad (8)$$

The elementary revenue is thus

$$dJ = M_0(dm/m)$$

and the total revenue is

$$J = M_0 \int_{m_0}^{m_t} (dm/m) = M_0 \ln(m_t/m_0) = M_0 \ln n, \quad (9)$$

if n denotes the multiplier by which the quantity of money put into circulation had increased during the given period.

To convince ourselves how close to reality is this formula under the same hypothesis of inverse proportionality of p and m , we shall additionally calculate the revenue under another assumption. Suppose that the money is put into circulation in s equal portions of size Δm . Then the quantity of the circulating money will be

$$m_0 + \Delta m, m_0 + 2\Delta m, \dots, m_0 + s\Delta m$$

and the respective values of the monetary unit,

$$p_0[m_0/(m_0 + \Delta m)], p_0[m_0/(m_0 + 2\Delta m)], \dots, p_0[m_0/(m_0 + s\Delta m)].$$

The revenue will then be

$$J = p_0 \Delta m \{ [m_0/(m_0 + \Delta m)] + [m_0/(m_0 + 2\Delta m)] + \dots + [m_0/(m_0 + s\Delta m)] \}.$$

Denoting $(\Delta m/m_0) = h$, we may write that as

$$J = p_0 \Delta m \{ [1/(1 + h)] + [1/(1 + 2h)] + \dots + [1/(1 + sh)] \}. \quad (10)$$

Calculating the sum according to the Euler formula, we will obtain¹

$$J = M_0 \left\{ \ln \frac{1 + (s+1)h}{1+h} + \frac{h}{2} \left[\frac{1}{1+h} - \frac{1}{1+(s+1)h} \right] + \frac{h^2}{12} \left[\frac{1}{(1+h)^2} - \frac{1}{[1+(s+1)h]^2} \right] - \frac{6h^4}{720} \left[\frac{1}{(1+h)^4} - \frac{1}{[1+(s+1)h]^4} \right] + \dots \right\}. \quad (11)$$

If, as above, we denote the rate of the increase in the issue, m_t/m_0 , by n , then

$$h = \Delta m/m_0 = [\Delta m/(m_t - m_0)] [(m_t - m_0)/m_0] = (n - 1)/s.$$

It follows that the expression under the sign of logarithm in (11) will be

$$\frac{1 + [(s+1)/s](n-1)}{1 + [(n-1)/s]} = \frac{n + [(n-1)/s]}{1 + [(n-1)/s]}.$$

It tends to n as $s \rightarrow \infty$, and, since all the other terms then tend to zero, we have from (11)

$$\lim J = M_0 \ln n, \quad s \rightarrow \infty, \quad \Delta m \rightarrow 0, \quad (12)$$

which coincides with what we obtained above when assuming a continuous current.

Suppose that $s = 30$, $\Delta m = (1/30)m_0$, then $h = \Delta m/m_0$, $n = 2$ and formula (11) leads to

$$J = M_0 \{ \ln(61/31) + (1/2)[(1/31) - (1/61)] + (1/12)[(1/31^2) - (1/61^2)] - (6/720)[(1/31^4) - (1/61^4)] + \dots \} = 0.6949M_0.$$

Formula (9) provides $J = M_0 \ln 2 = 0.6931M_0$, only differing from the former magnitude by 1.2%. However, our assumption that, during the period when the amount of money in circulation increased twofold [$n = 2$], there were only 30 issues, is of course too rough. Supposing that $s = 1000$ and $h = 1/1000$, we will indeed find out, as it is easy to show, that formula (11) provides a result only differing from $J = M_0 \ln 2$ by 0.04%. We may therefore consider the formula (9) a quite proper expression for the state revenue from the issue of paper money under our main abovementioned assumption.

3. In our literature, this revenue is usually determined by dividing the amount of money put into circulation by the index for the middle of the relevant period of time. Suppose for the sake of convenience that the beginning of that period is the zero point of time and that its length is unit (for example, 1 month), then the index for the middle of the period will be denoted as $i_{1/2}$ and the approximate formula under our discussion will then be

$$J' = [(m_1 - m_0)/i_{1/2}]. \quad (13)$$

Its first check ought to consist in comparing it with formula (9), which, under the hypothesis that the level of prices is proportional to the amount of money, should be regarded as strict. But we must know the law guiding the movement of the issue in time.

If the amounts of the issues as existed at the first day of each month from 1917 to 1923 are plotted on a logarithmic scale, then, as shown by Schmidt (1923), for the most part of that period the thus obtained line might be replaced to a high precision by a number of linear segments, most of them rather long. Assuming that this law is valid for each of the comparatively short (for example, monthly) periods taken separately, for which the approximate formula (13) can indeed only be reasonably applied, we will obtain the law of the movement of the issue as²

$$\ln m = a + gt. \quad (14)$$

For $t = 0$ and 1 we have

$$a = \ln m_0, \quad g = \ln m_1 - \ln m_0 = \ln(m_1/m_0) = \ln n.$$

Thus we arrive at

$$m_t = m_0 e^{gt}, \quad g = \ln n \quad (15)$$

or, which is sometimes more convenient,

$$m_t = m_0 n^t. \quad (16)$$

Note that n is here understood as the *norm of the increase in the issue per unit time* and it would be most expedient to call it the *rate of the issue*.

Suppose now that at some moment T the level of prices becomes such that formula (13) provides an exact expression of the revenue when $i_{1/2}$ is replaced in it by i_T . According to our hypothesis

$$(i_t/i_0) = (m_t/m_0)$$

so that, because of (16),

$$i_t = i_0 n^t \quad (17)$$

Therefore, $i_T = i_0 n^T$ and T can be obtained from the equation

$$\frac{m_1 - m_0}{i_0 n^T} = M_0 \ln n \quad (18)$$

Since

$$\frac{m_1 - m_0}{i_0} = \frac{m_1 - m_0}{m_0} \frac{m_0}{i_0} = (n - 1)M_0,$$

we obtain from (18)

$$n^T = [(n - 1)/\ln n].$$

Therefore, on the one hand, we will have

$$T = \frac{\ln[(n - 1)/\ln n]}{\ln n} \quad (19)$$

and, on the other hand, on the strength of (17) and (2),

$$(p_T/p_0) = (i_0/i_T) = [\ln n/(n - 1)]. \quad (20)$$

Table 1 provides the values of (p_T/p_0) and T for some values of n . Column 4 indicates how many days apart from the middle of a 30-days month is the moment for which the index furnishes the exact magnitude of the revenue when the monthly issue is divided by that index. We see that for $n < 2$ this difference ($30T - 15$) is less than a day, which of course is small enough. Finally, the last column shows the error of the approximate value of the revenue in accordance with formula (13). It is obtained in the following way. Inserting in (13), instead of $i_{1/2}$ its value as provided by formula (17), i.e., $i_0 \sqrt[n]{n}$, we have

$$J = [(m_1 - m_0)/i_0 \sqrt[n]{n}] = [(n - 1)/\sqrt[n]{n}]M_0.$$

Dividing this by the expression for J as given by formula (9), we obtain

$$(J'/J) = \frac{1}{\sqrt{n}[\ln n/(n-1)]}. \quad (21)$$

Using the magnitudes calculated in accordance with this formula, we have derived for the sake of distinctness the relative percentage errors of the empirical formula (13) and entered them in the last column of Table 1. We see that for $n < 2$ the error of formula (13) does not exceed 2% and is [even] less than 1% until n does not exceed 1.6 plus. Since until 1923 the monthly rate of the issue never exceeded 2, and in most cases it was not more than 1.6, the application of formula (13) obviously involves more or less admissible errors especially when we take into consideration the influence of the other sources of error. Nevertheless, formula (9) is a more proper expression of the revenue from the issue.

4. The proportionality of the level of prices and the amount of money put out into circulation can only be considered as a tentative assumption. The general level of prices depends not only on that factor, but also on a number of other causes including such a forcible factor as the volume of the turnover [?]. During the time period, with which our study is concerned, the volume of commodities marketed in our country had experienced very considerable changes and the rapidity of the money circulation had certainly not remained constant either. And we indeed see that the actual movement of the prices had been to a large extent independent of the movement of the issues.

To gain a foothold for theoretical deliberations let us consider the curve depicting the movement of the real value of all the issues, see Diagram 1 in Iasnopolsky's article (1923). We see that M does not remain constant, as it would have been had the level of prices been proportional to the amount of money put out into circulation, but experiences considerable changes. Generally, they are, however, comparatively rather smooth, do not have very sharp jumps. Therefore, when considering a more or less short period of time, it is seen that each small arc of the curve M , which is a function of time, can be replaced with a rather proper approximation either by a parabola of the second degree or even by a [segment of a] straight line. For very short periods, such as a month or a fortnight, the assumption of linearity for the curve $M = f(t)$ is probably rather close to reality.

Let us dwell on this assumption. It means that

$$M = a + bt. \quad (22)$$

Hence, since $M_t = m_t p_t$,

$$p = (a + bt)/m. \quad (23)$$

Suppose that the amount of money put out into circulations increased by an infinitesimal magnitude dm ; the infinitesimal revenue will then be

$$dJ = [(a + bt)/m]dm$$

and the revenue accrued during time interval $(0; t)$ will be represented by the integral

$$J = \int_0^t [(a + bt)/m] dm. \quad (24)$$

Differentiating (15) with respect to t , we will have

$$dm = gm_0 e^{gt} dt = gmdt$$

and, when substituting this expression in (24), we obtain

$$J = g \int_0^t (a + bt) dt = g[at + (1/2)bt^2]. \quad (25)$$

However, assuming that $t = 0$ we obtain on the strength of (22) $a = M_0$ and it follows that $bt = M_t - M_0$. Owing to (15),

$$gt = \ln (m_t/m_0)$$

and we find that

$$J = gt[a + (1/2)bt] = \ln (m_t/m_0)[M_0 + (1/2)(M_t - M_0)].$$

And, finally, again denoting $(m_t/m_0) = n$, we obtain

$$J = (1/2)(M_0 + M_t)\ln n. \quad (26)$$

The reader will see that this formula only differs from (9) in that the value of all the paper money put out into circulation up to the beginning of the period (M_0) is replaced by the arithmetic mean of the same value for the beginning and the end of the period. Now, (26) can be written as

$$J = (1/2) M_0[1 + (M_t/M_0)]\ln n$$

and since $M_0 = m_0/i_0$, $M_t = m_t/i_t$, $(m_t/m_0) = n$ and (i_t/i_0) can be denoted by k , the revenue from the issue will be expressed as

$$J = (1/2)[1 + (n/k)]M_0 \ln n. \quad (27)$$

5. Following Falkner (1923), we may call the ratio J/M_0 the relative efficiency of the issue. Only it is hardly correct to consider this magnitude, as he does (p. 54), the *portion of the circulated consumer values extracted by means of the issue*, since, even in accordance with the simplest theoretical pattern, the value of the mass of commodities is equal not to M , but to M multiplied by the rapidity of the circulation of the banknotes.

Denoting the relative efficiency by η , that is, replacing J/M_0 by η , we obtain from (27)

$$\eta = (1/2)[1 + (n/k)]\ln n. \quad (28)$$

It is seen now that, if the general level of prices increases in the same number of times as the amount of money put out into circulation, the relative efficiency of the issue will be

$$\eta = \ln n \quad (29)$$

and is thus expressed by the natural logarithm of the rate. We will indeed call this magnitude, tentatively of course, the *normal relative efficiency*. If the rate of the issue is higher than the rate of prices ($n > k$), then $\eta > \ln n$; if it is lower ($n < k$), then $\eta < \ln n$.

If n does not exceed 2, $\ln n$ can be expanded into a series

$$\ln n = \ln[1 + (n - 1)] = (n - 1) - (1/2)(n - 1)^2 + (1/3)(n - 1)^3 - \dots \quad (30)$$

The magnitude $(n - 1)$ is the *norm of the increment* of the issue. If the rate = 1.25, this norm is 0.25 or 25%, etc. Formula (30) shows that, given a *more or less weak rate*, it is possible to neglect the second and the following terms of this expansion and then the *normal relative efficiency will be expressed by the norm of the increment of the issue*. For example, if $n = 1.1$,

$$\ln n = 0.1 - 0.005 + 0.00033 - 0.000025 + \dots$$

and, assuming that, approximately, $\ln n = n - 1 = 0.1$, or 10%, we thus make an absolute error less than (1/2)% although its relative (with respect to the 10%) magnitude is nevertheless 4.7%. For $n = 1.2$, the approximate value of the normal relative efficiency would have been 0.2, or 20% instead of $\ln 1.2 = 0.182$, or 18.2% and its relative error, already 10%. This error rapidly increases with the further increase in n as it can be seen in Table 2.

It was the closeness of the relative efficiency to the norm of increment, given small values of n , that caused Falkner's mistake (1923, p. 54). He attempted to explain the considerable difference between these two indicators in 1921 and 1922 by the divergence of the rate of prices and the rate of the issue. We see that this is wrong because, even if the level of prices is strictly proportional to the volume of the issue, when having the norm of the increment equal to 100% ($n = 2$), the relative efficiency is only 69.3%; and efficiency of only 109.9% corresponds to an increment of 200%, etc.

The formula for the relative efficiency derived above can be checked against the data for 1922 (Table 3). The actual efficiency (the method of whose calculation is indicated below) is shown in column 6. Entered alongside are the relative theoretical (column 5) and normal (column 4) efficiencies. Columns 2 and 3 show the rates of the issue (n) and prices (k).

When transferring the figures of columns 2 – 6 of Table 3 to a diagram (Fig. 1), we obtain a clear picture of the fluctuations of all the pertinent magnitudes that confirms the deliberations above. The empirical efficiency, as we see, rather closely follows all the windings of the theoretical curve. As long as the rate of issue overcomes that of the prices (from May to September), both curves are situated above the curve of the normal efficiency; and below it in the opposite case (January – April and October – November). When both rates coincide, then the curves of relative efficiency (both the theoretical and empirical curves) coincide, or almost coincide,

with the norm (April and December). Since the selected period [1922] belongs to those with most sharp fluctuations of both rates, we may apparently conclude that the theory explicated above corresponds to reality in general [for any periods]. This means that we are entitled to base our future calculations as well on the hypothesis that, at least for periods not longer than a month, the curve of the value of the total mass of paper money is [actually] rectilinear.

6. Let us check further the empirical formula for calculating the revenue from the issue (formula (13)) which we discussed above when assuming the hypothesis that the level of prices was proportional to the amount of money put out into circulation. Assuming now that the index is moving independently and a hypothesis of a linear M , and recalling formulas (23) and (2), we will have for the index the equation

$$i_t = [m_t / (a + bt)].$$

At $t = 0$ we have, according to formula (22), $a = M_0$ and, at $t = 1$, $b = M_1 - M_0$, so that the previous expression becomes

$$i_t = \frac{m_t}{M_0 + (M_1 - M_0)t}.$$

Substituting M_0 by m_0/i_0 , and M_1 , by m_1/i_1 , and denoting, as before, $m_1/m_0 = n$ and $i_1/i_0 = k$, we obtain without difficulties

$$i_t = \frac{i_0}{1 + [(n/k) - 1]t} \frac{m_t}{m_0}.$$

Substituting now, instead of (m_t/m_0) , its value n^t given by equation (16), we finally obtain

$$i_t = \frac{i_0 n^t}{1 + [(n/k) - 1]t}. \quad (32)$$

For the middle of a unit period, substituting $t = 1/2$ in (32), we will find that

$$i_{1/2} = \frac{i_0 \sqrt{n}}{(1/2)[(n/k) + 1]}$$

whence, substituting this expression in formula (13), we get

$$J' = (1/2) \frac{[(n/k) + 1](m_1 - m_0)}{i_0 \sqrt{n}} = (1/2) \frac{[(n/k) + 1](n - 1)}{\sqrt{n}} M_0. \quad (34)$$

However, under the assumed hypothesis, the real value of the revenue from the issue is determined by formula (27). Therefore, the ratio of the approximate expression of the revenue J' to its real value J becomes equal to

$$(J'/J) = \frac{1}{\sqrt{n}[\ln n/(n-1)]}. \quad (35)$$

This exactly coincides with the same expression derived above (formula (21)) for the case in which the level of prices changes proportionally to the amount of the issued money. In practice, the error of this formula in most cases does not exceed 1%; only once in 1922, with $n = 1.8$, was it larger than this figure; for such value of n Table 1 shows that the error is 1.4%. This error does not at all depend on the ratio of the rate of prices to the rate of the issue which refutes the opinion recently expressed by Bazarov (1923, p. 23)³ that

During periods of a slow sinking of the course [this method allegedly] systematically exaggerates the real revenue from the issue and that, to the contrary, during periods of especially rapid sinking of the course, this method systematically underestimates the magnitude sought.

This question is important, and we will therefore subject formula (13) to one more check. Indeed, we will assume that, just as the issue, the index is changing by a geometric progression, only having another common ratio. Then $m_t = m_0 e^{gt}$, $i_t = i_0 e^{ht}$. Omitting the derivation so as not to overstep the assigned volume of this paper, we only provide the end result:

$$J = \int_0^1 \frac{dm}{i} = \frac{(n/k) - 1}{\ln(n/k)} M_0 \ln n. \quad (36)$$

For the middle of the period the index is

$$i_{1/2} = i_0 \sqrt{k} \quad (37)$$

and, substituting this last expression into the approximate formula (13) and dividing the result by (36), we will find that⁴

$$\frac{J'}{J} = \frac{\frac{1}{\sqrt{n}[\ln n/(n-1)]}}{\frac{1}{\sqrt{n/k} \{ \ln(n/k) / [(n/k) - 1] \}}}. \quad (38)$$

For $n < 1.6$, the numerator of this expression, as we saw in Table 1, differs from 1 not more than by 0.7% whereas the denominator repeats the numerator with n being replaced by n/k . During the entire year 1922, this ratio never exceeded 1.51 [see Table 3, July 1922]. Table 1 shows that for $n/k = 1.5$ the denominator of (38) ought to differ from 1 less than by 0.7%. Under that hypothesis, this is the maximal extent of the influence of the divergence of the rate of prices from that of the issue, even when so sharp fluctuations as occurred in 1922 are present, on the error of the formula (13). And, contrary to Bazarov, a slower increase in the index for a given n does not exaggerate the discussed approximate value of the revenue but tends to

underestimate it. Be that as it may, suchlike influences are absolutely insignificant as compared with other errors.

Had we assumed that during a month the level of prices sometimes rose and sometimes fell rather than moved in one direction, the error of the approximate formula (13) could have been considerably larger. This, however, has nothing to do with Bazarov's thesis, because, independently of the total result of all the movements of the index during a month, the error can be either positive or negative. For proving this proposition it is sufficient to imagine that first the highest, and then the lowest level of prices occurred at the middle of the pertinent period. For the time being, until the question is not studied more thoroughly, we may consider such fluctuations of the index as random deviations from a smoothly changing level and believe that the ensuing errors in calculating the revenue from the issue compensate each other in the general total for a number of months. Our calculations (below) apparently confirm this point of view.

7. We are now going over to the problem of interpolation because this operation can be necessary when calculating the revenue from the issue when the data are incomplete.

Figures of the *grand issue* (our u) for the first day of each month as well as the index of the *Statistika Truda* (Statistics of Labour), also for these days, are known to be officially published. Beginning with 1922, this index is also provided for the fifteenth day of each month and we have to make use of it since the other indices do not cover all the period. All our calculations are therefore of a very tentative importance; however, we must regrettably leave here completely aside the problem of indices in its relation to the issue. The reader will see that this circumstance can hardly shake our conclusions since they are only of a methodological nature.

For 1922, in addition to the figures of the grand issue, figures of the so-called small issue (our m) are also available, and the difference between them constitutes the total amount of paper money en route and in banks. These figures, especially when comparing them with those provided by Bazarov (1923, p. 24) for the first months of 1923, provoke doubts and for us their analysis is premature. Note only that we cannot at all agree with the respected author that the mean percent of money in the banks and en route, as compared with the total amount of money put out into circulation during a month, and calculated by making use of the data for those few months which he studies, can be taken as a norm for all the previous period. Without even mentioning that the conditions in the past could and should have differed *not* insignificantly as compared with the modern time, we indicate that the mean itself was derived from so few, and so differing one from another numbers, that its significance is extremely doubtful. We believe that here we ought to wait for further official publications and to abstain for the time being from judgement.

In the sequel, we will keep to the hypothesis adopted by a number of researchers that it takes two weeks for the money en route and deposited in banks until being circulated. Accordingly, the official figures of the issues for the first day of a month correspond to the amount of money put out into circulation (m) until the 15th of the same month. For the sake of brevity we will now simply call the time during which money is en route and in banks the bank period and denote it by τ . Assuming, of course absolutely tentatively, the hypothesis that $\tau = 1/2$ of a month, we shall study the

ensuing methods of calculation of the revenue from, and the relative efficiency of the issue. In concluding, we will attempt to estimate, even if approximately, how great is the error of these calculations depending on a mistaken determination of τ .

The first problem which we should examine is that of interpolation. In accordance with the adopted hypothesis, the amount of the issued paper money is only known for the middle of a month whereas the indices of prices, for its first day. True, we can calculate the monthly revenue by dividing the increment of the monthly issue by the index for the end of the month because the end of the month, or, more precisely, the first day of the next month, will be the middle of the corresponding period. Nevertheless, without knowing the index, we will be unable to determine the relative efficiency of the issue since we cannot calculate the value of the total amount of paper money at the beginning of the period (as of the middle of the month). And, second, we will only know the revenue from the 15th to the next 15th rather than from the first day to the next first day as is necessary for comparing and adding it to other state revenues. Third, we will be unable to determine the yearly revenue either because the total calculated revenue for 12 months will correspond to the period from January 15 to January 15 of the next year rather than to the calendar year.

For determining the relative efficiency when calculating the revenue from the middle of a month to the middle of the next month it is necessary to interpolate the prices, and, for solving the second problem, to interpolate the emission. We begin with the latter.

If the rate of issue during the studied period remains constant, we have (formula (15))

$$\ln m_1 = \ln m_0 + gt_1; \ln m_2 = \ln m_0 + gt_2; \ln m_3 = \ln m_0 + gt_3,$$

therefore

$$\frac{\ln m_2 - \ln m_1}{\ln m_3 - \ln m_1} = \frac{t_2 - t_1}{t_3 - t_1}, \ln m_2 = \ln m_1 + \frac{t_2 - t_1}{t_3 - t_1} (\ln m_3 - \ln m_1).$$

Supposing for the sake of convenience that $t_1 = 0$, $t_2 = t$ and $t_3 = 1$, we obtain a simple interpolation formula

$$\ln m_t = \ln m_0 + t(\ln m_1 - \ln m_0). \quad (39)$$

A practically convenient arrangement of calculations is shown in Table 4 which deals with figures describing the *grand issue* (u) and t being consecutively equal to 1/4, 1/2, 3/4. The second column shows the logarithms of the issues for 1 Dec. 1921, 1 Jan. 1922, etc. Consecutive differences between these logarithms are entered in the third column and, in the next column, 1/4 of the differences. Note that when applying five-place logarithmic tables, as it was done here, rounding off is better avoided because we ensure a constant checking when exactly calculating $(1/4)\Delta \ln u$ by means of an arithmometer. One revolution of its handle provides the figures of the fifth column; two, three, and four revolutions furnish those of the next two columns and the value of the logarithm for the next month.

The thus interpolated figures, along with the initial figures, provide the data of the grand issue for the first day of the month, for its first quarter, its middle, for the third quarter and for the first day of the next month. Had the money not been delayed in the banks at all ($\tau = 0$), the issue for January 1, for example, would have coincided with the amount of money issued up to that date. If the money is delayed for one week, the amount of the issued money will be equal to the figure of the grand issue for the third quarter of December (its logarithm is the first figure in column 7 of Table 5). Mid-December and its first quarter correspond to delays of two and three weeks respectively.

In this way we obtain Table 5 where the thus derived amounts of money put out into circulation on the first day of each month of 1922 are shown under the hypotheses that $\tau = 0, 1, 2, 3$ and 4 weeks.

Drawing on these figures, we shall formulate some essential conclusions in the sequel, but now we pass on to the next problem: Is it correct to apply the described method of interpolation if, as we know for sure, the rate of the issue changes from month to month, sometimes considerably, and there are no grounds for believing that these changes occur by jumps at the boundaries between the monthly intervals? Is it not more natural to suppose that, if during the previous month the rate of issue was lower, the pertinent causes were still probably acting during the first half of the current month; and that the second half of the current month can be influenced by those causes whose action will be quite felt in the next month by a further lowering, or, let us say, heightening of the rate? It would therefore be perhaps more correct to take into consideration not only the given month, but to allow also for the two, and perhaps four neighbouring months although the more distant influences hardly ought to be studied. If this idea is adopted, the appropriate interpolation can be described as follows. The logarithms of the figures of the monthly issues are shown on a diagram; the points corresponding to the beginning and the end of the months are connected not by rectilinear segments, as it would have been proper for the previous method of interpolation, but by curves. For each month a separate curve determined by its starting and end points as well as by points corresponding to one or two previous and the same number of subsequent months is drawn.

Without dwelling on the theory of interpolation for this case, we only provide an arrangement of the calculations in accordance with the Stirling [interpolation] formula which seems to be indeed suitable because it interpolates not one-sidedly, as, for example, the Newton formula does, – that is, not by means of the differences between the given month and a number of previous *or* subsequent months, – but by considering *all* of these differences.

The arrangement of the calculations is shown in Table 6. Here, in column 2, we again have the logarithms of the *grand issue* for the first day of each month beginning with October 1921. We begin with that date because we have to interpolate between 1 Dec. 1921 and 1 Jan. 1922 and wish to allow for the two subsequent, and two previous months. We will see below that this involves the necessity to restrict the calculations with third differences. For calculating the fourth, etc differences we should have considered not two, but three or more previous, and the same number of subsequent months but we consider the more distant influences illusive.

Column 3 shows the differences between the figures of the preceding column (Roman type), and the second and third differences are entered in columns 4 and 5 respectively. Each time we subtract the upper figure from the lower one and allow for their signs, and we write down the differences not on the same line with the subtrahend, as we did in Table 4, but *between* it and the minuend. The figures in italics, which are the arithmetic means of the corresponding differences, are entered afterwards.

When keeping to such an arrangement of the table, it is easy to see that the two [apparently] superfluous figures provided both in the beginning and at the end of the table enable us to allow for the third differences.

Column 6 provides the calculated logarithms of the issues for the middle of the appropriate months. Thus, the first figure of that column, 4.11127, is entered between the lines XII and I and corresponds to the issue for 15 December ($u_{XII+1/2}$). In accordance with the hypothesis that the bank period lasts two weeks, it should be equal to the logarithm of the amount of money issued on 1 January (m_1). Calculations are made by means of the Stirling [interpolation] formula, see for example Krylov (1911, p. 250)⁵,

$$f(a + th) = f(a) + (t/1)f^{(1)}(a) + (t^2/1 \cdot 2)f^{(2)}(a) + [(t + 1)t(t - 1)/1 \cdot 2 \cdot 3]f^{(3)}(a) + \dots \quad (40)$$

only allowing for the first four terms (including $f^{(3)}(a)$). Here, a is the value of the argument. In our case, this is the moment corresponding to the first of those two numbers between which interpolation is done; for example, if 1 Jan. is assumed as the starting moment of time and a month is considered as a unit, then $a = 0$ for 1 Jan., $a = 1$ for 1 Febr., etc. The difference between consecutive values of the argument is denoted by h ; in our case, $h = 1$ (one month) and t is a number situated between 0 and 1 and showing what part of h is the distance between the beginning of a period and the value of the argument for which we are calculating the intermediate value of the function. For example, if we wish to know the issue for 8 January (more precisely, for the boundary between the first and the second quarter of January), $t = 1/4$; then, $t = 1/2$ for the middle of the month, etc. Finally, $f^{(1)}(a)$, $f^{(2)}(a)$, etc are the numbers entered in the columns of differences in the same line as $f(a)$ for the beginning of the corresponding period; namely, $f^{(1)}(a)$ is the arithmetic mean of the first differences, $f^{(2)}(a)$ is the second difference, $f^{(3)}(a)$ is the mean of the third differences, etc.

To determine the issue for mid-January (say), we ought to take the logarithm of the issue for 1 Jan., $f(a) = 4.24400$, and three numbers from the same line,

$$f^{(1)}(a) = 0.24300, f^{(2)}(a) = -0.01556, f^{(3)}(a) = -0.049515.$$

We then calculate the coefficients:

$$t = 1/2; (t^2/1 \cdot 2) = 1/8; [(t + 1)t(t - 1)/1 \cdot 2 \cdot 3] = -1/16$$

and proceed further. [...] The result, $\ln u_{I+1/2} = 4.36665$, is shown in column 6 of the same Table 6 between the dates 1922, I and II. [...] Column 7 indicates the figures of the issue (of the grand issue) for the first day of each month; the next column shows the issue for mid-months calculated as

explained above (drawing on the logarithms in column 6). The same magnitudes calculated as geometric means by a simple linear interpolation of the logarithms and entered in Table 5 are shown once more in column 9 of Table 6 so as to enable a distinct comparison. Column 10 provides the discrepancies between columns 8 and 9, and the same discrepancies relative to the data in column 8 and expressed in per cents are in column 11. We see that these discrepancies are quite small; their absolute mean is only equal to 0.8%. For usual calculations, when having to do with such approximate data as, for example, the index of prices, corrupted by systematic errors amounting perhaps to 10 – 15%, and the issue, whose moment we do not know, – it could have occurred during the first, the second, or the third week of the month, – these discrepancies might certainly be neglected. However, we delay our judgement until making the next stage of the comparison; indeed, we ought to examine whether these discrepancies do not influence to a greater extent the calculations of the revenue from the issue.

But let us first study the prices. Without providing the calculations made by interpolating the logarithms of the indices in accordance with the two abovementioned methods, we only furnish their results, see Table 7. There, i denotes real indices as given in *Statistics of Labour*, NNo. 1 and 2 for 1923; i' and i'' are the indices for the middle of each month considered as unknown magnitudes and derived by interpolation from the first days in accordance with the geometric mean and the Stirling formula respectively.

When studying the differences between the figures obtained by both methods, we see that in the mean they amount to 2.2% and thus almost threefold exceed the same differences for the issues. This is indeed what should have been expected after some acquaintance with both curves. However, in spite of the differences, being although not large but still already comparatively noticeable, the complicated method of interpolation, if judged by the data for 1922 treated by us, is apparently not more advantageous than the simple method: the errors of these methods, 4.2 and 4.5%, as calculated by comparing their results with the real figures, almost coincide. True, the former distributes the error in both directions more uniformly than the latter: the sums of the percentage errors of i (columns 10 and 8) are $22.4 - 23.7 = -1.3$ and $21.3 - 28.1 = -6.8$ respectively. However, the difference between these sums is not really large and, with the data at hand, it is difficult to say to what extent it is not accidental. In general, both methods are approximately equivalent and even the less accurate of them provides comparatively suitable approximations.

8. We can subject the different methods of interpolation and different formulas for calculating the revenue to some critical comparison by calculating it for the separate months of 1922 by different methods.

It is obvious that, from among all possible approximations to this revenue, the best result is rendered by that which was obtained by making use of all the available data. The conclusion thus arrived at will serve for estimating the other methods, – those, among which we ought to choose when investigating the issues for the previous years and drawing on a lesser amount of materials. What is available for 1922 but remains unknown for the previous time is the index for the middle of the months which we must determine by interpolation.

And so, we have five magnitudes, i_0 , $i_{1/2}$, i_1 , m_0 and m_1 for each month of 1922. We can make use of them by calculating the revenue either for

monthly, or half-monthly periods. For the sake of distinctness, let us compare all the available formulas (enumerating them anew):

$$J^I = [m_1 - m_0]/i_{1/2} \quad (\text{I}),$$

$$J^{II} = M_0 \ln n, \quad (\text{II})$$

$$J^{III} = (1/2)(M_0 + M_1)\ln n = (1/2)[1 + (n/k)]M_0 \ln n, \quad (\text{III})$$

$$M_0 = m_0/i_0, M_1 = m_1/i_1, n = m_1/m_0, k = i_1/i_0.$$

First, consider the application of these formulas to the calculation of J for whole months at once. Formula (I), that, as we saw, should at $n < 2$ provide a result rather close to reality, is nevertheless only empirical and corrupted by some systematic error. There are hardly any grounds for preferring it rather than (II) or (III). It is unlikely to be the best one when the rate of the increase in prices is changing. Indeed, when applying it, we do not at all make use of two out of the five magnitudes listed above, – of the index for the beginning and the end of the month. And so, this formula is not suitable.

In our case, formula (II) is not the best one either; the assumptions on which it is founded (uniformity of the rate of issue and coinciding rates of prices and issues) contradict reality. [In addition,] it does not make use of all the facts or all data. Formula (III) is better. It allows for the discrepancy between the [rates of] the prices and the issues and makes use of four magnitudes, m_0 , m_1 , i_0 and i_1 , but still does not use $i_{1/2}$ and assumes a certain increase in the level of prices, – namely, such that corresponds to a linear increase (or decrease) in the total mass of paper money.

Considering [all] this, it seems most expedient to calculate the revenue from the issue for half-months in accordance with formula (III); it allows for all the known facts and its possible error is limited to narrowest boundaries.

And so, having chosen this method and keeping to the hypothesis of a two-week bank period, we calculate the half-monthly revenues from the issue from 1 Jan. 1922 to 1 Jan. 1923 drawing on the given indices (column 2 of Table 7) and on partly given, and partly derived by interpolation figures of the issues. All the initial data and the results of calculation are provided in Table 8.

After all that was stated above, this table hardly demands much explanation, its structure is clear. It begins with the index for 1 Jan. 1922 and the amount of the issue interpolated for 15 Dec. (see Table 6). In accordance with the assumed hypothesis, this amount, 12.920, is assumed as the amount of money *put out into circulation* (m) as of 1 Jan. 1923. Figures in column 4 are derived by dividing those of column 3 by figures in column 2 and represent the value of the money put out into circulation. Column 5 shows the arithmetic means of the pertinent figures in the previous column. In column 6 we have the coefficient of the increase in the issue during two weeks, column 7 indicates its natural logarithm and column 8 is the revenue from the issue for separate half-months calculated by multiplying the pertinent figures in columns 5 and 7. Columns 9 and 10 provide the revenue from the first day to the next first day, and from the 15th day to the next 15th day, respectively, obtained by summing the pertinent pairs of figures from column 8.

For the sake of comparison column 11 provides the revenue from the 15th day to the next 15th day, beginning with 15 Jan. 1922 until 15 Dec. of the

same year inclusive, calculated by means of formula (I), – that is, in accordance with the approximate and usually applied method. Neither logarithms, nor interpolation are here present: the consecutive differences of the figures of the issues, assumed as the amount of money put out into circulation from the 15th day to the next 15th day, are divided by the index of the first day of the next month. In our opinion, the coincidence is not bad at all. The discrepancies do not exceed 2.8% and their absolute mean is only 1.2%. And, the discrepancy of the appropriate total sum [266.1:265.41] (we obtain it by adding the figures in brackets in column 11 corresponding to two half-months both in the beginning and at the end of the year) is absolutely insignificant, only 1/4%. Those readers, who with certain distrust, understandable for a non-mathematician, scanned our long calculations and deliberations, will [now] perhaps gain some confidence in mathematics. And, in general, we hope for an agreement concerning the following statements:

1) *The empirical method of calculating the revenue from the issue is trustworthy and may be applied without being afraid of large errors.*

2) *The hypotheses adopted as a foundation for the calculations in accordance with formula (III), i.e., above all, the uniformity of the rate of issue, and, second, the uniformity (linearity) of the change of the value of the total amount of money put out into circulation, when applied to half-monthly periods, are close to reality.*

9. The relative reliability of the empirical method (formula (I)) of calculating the revenue does not save us from the trouble of making a number of comparisons between several methods. Indeed, when quite justifiably wishing to calculate the revenue from the issue for the previous years, for which no mid-monthly indices are available, not from the 15th to the 15th of the next month, but from the first day to the next first day, we would have either to interpolate the indices or apply formula (III). And in any case we would also be obliged to interpolate the issues. Even if the indices for the middle of the months were published, the pertinent problems would not have lost meaning: it is sufficient to become convinced, that the hypothesis of a two-week bank period ought to be replaced by another one, for all the problems connected with interpolation to reappear.

For the sake of comparison we have calculated the revenue from the issue from the first day to the next first day by several methods. Let us list them and introduce notation.

J : the revenue calculated by a method indicated above, see column 9 in Table 8. We will consider it as providing the maximal yet available approximation to reality and compare it with the other approximations.

J' : the same, calculated by formula (I).

J''' : the same, calculated by formula (III).

J_1' and J_1''' : the same, calculated by formulas (I) and (III) respectively by drawing on the amount of money put out into circulation as of the first day of the month (m) and interpolated by the first method (of geometric mean).

J_2' : the same as J_1' but with interpolation by the second method (by the Stirling formula). Note that J_1' , J_1''' and J_2' are only calculated by drawing on officially published data (on *actual* indices).

J_{1A}' : the revenue calculated in the same way as J_1' but by dividing by the index equal to the geometric mean of the indices for the beginning and the

end of the month rather than by the *actual* index for the middle of the month.

Magnitudes J , J_1' , J_2' , J_1''' and J_{1A}' are entered in columns 1 – 5 of Table 9 respectively which also shows the percentage differences; notation such as $\varepsilon_{3/1}$ means the ratio, expressed in per cents, of the difference between the figure in column 3 and that in column 1 divided by the latter.

This table provides some valuable indications.

1) First, it again confirms what we saw above: formula (1) provides a result closely situated to the most reliable approximation J both in Table 8, where we applied this formula to the initial data contained in an official publication and considering the interval from the 15th day to the next 15th day; and here, in Table 9, if applying it to the best from among the interpolated data. There, the deviations in the mean amounted to 1.2%, here, to 2.2% ($\varepsilon_{3/1}$). The use of the simple geometric mean somewhat worsens the result, – the deviation becomes equal to 3.5% ($\varepsilon_{2/1}$), – but even this cannot be considered bad.

2) The result is worse when applying formula (III) to figures derived by simple interpolation (J_1'''), or, when calculating the revenue by formula (I) by drawing on the same figures and again using interpolation. In both cases, the deviation from the most reliable approximation then amounts to 4.5%.

3) But these errors are not systematic, their positive and negative signs are distributed randomly. This is already evident since the yearly revenue is almost the same for all the methods of its calculation; deviations only amount to fractions of one per cent.

4) The very small difference between J_1''' and J_{1A}' only amounting to 1/2% in the mean, whereas the methods of their determination are absolutely different, is interesting indeed.

5) We may therefore conclude that, *for the time before 1922, the results of determining the revenue from the first day to the next first day by both available methods, – that is, a) when considering the figures of the small issue for each first day as geometric means of official data and calculating the revenue by formula (III); and b) when considering the indices for the middle of the months as geometric means of the indices for the beginning and the end of the appropriate months and drawing on the same figures but applying formula (I), – are of approximately the same precision. In short, when applying simple interpolation (of the logarithms), formulas (I) and (III) are apparently of the same precision.*

10. The mean error of calculating the revenue by interpolation of the data, as we may guess on the strength of the above, does not exceed 5% in the mean in either direction and is more or less compensated in the yearly totals. However, the situation concerning another error is quite different. I bear in mind the error caused by a wrong determination of the bank period τ . In Table 5 we had entered the values of m at $\tau = 0, 1, 3$ and 4 (weeks). Let us see how the change of τ influences the value of the total issue and the revenue, see Table 10 based on formula (I). The pertinent differences Δm between the first day of a month and the first day of the next month are divided by the index for the middle of the month, see column 2 of Table 8. Thus the values of J in Table 10 are derived whereas the values of M are the quotients of m divided by the respective index for the beginning of the month. We restricted our attention to three hypotheses, $\tau = 1, 2$ and 3.

In the fourth part of Table 10 the reader can see the deviations ($\varepsilon_{1/2}$ and $\varepsilon_{3/2}$) of the revenue calculated for $\tau = 1$ and 3 from that when $\tau = 2$ if the last-mentioned revenue is taken to be 100. The mean errors are here 9.85 and 8.87% and they are of a systematic rather than random nature (do not change their direction).

The yearly totals therefore also differ considerably. We shall determine these differences in per cents in a somewhat another way, not as we did when calculating the monthly revenues in the table. Namely, we shall suppose that the yearly total is 100 first for $\tau = 1$, and then for $\tau = 3$. This means that if the actual bank period was only 1 week, and assuming that $\tau = 2$, our estimate of the yearly revenue for 1922 would have been lower by 7.2% than the real figure. For an actual bank period of 3 weeks we would have overestimated the yearly revenue by 9.2%. These errors are great as compared with discrepancies amounting to fractions of one per cent, between the figures of the revenue as calculated by different methods of interpolation and calculation. The efforts of the researchers evidently ought to be directed to throwing light on this obscure point.

Explanation of Tables and Figures

Table 3, column 1: Months of 1922

Fig. 1, the legend: Empirical; theoretical; normal relative efficiency of the issue. The norm of the monthly increment of the issue; of the index of prices

Table 4, column 1: Date

Table 5, column 1: Date. The other columns: the amount of money put out into circulation as of the first day of the appropriate month (in 10^{12} of nominal roubles) assuming various bank periods (in weeks)

Table 6, column 1: Date. Last line: Mean absolute value.

Footnote: Issues entered in this [the seventh] column are taken from *Vestnik Finansov* No. 16 – 17, pp. 10 – 11. A misprint in the figure for Dec. 1922 is corrected. The figure for Febr. 1923 is calculated by drawing on later data that we had not possessed but was provided by Bazarov (1923, p. 24).

Table 7, columns 1 – 4, 6, 8, 10: Date; In thousands; Geometric mean; According to the Stirling interpolation formula; The same, %% with respect to i'' ; The same, with respect to i ; column 10 = column 8

The last line: The mean absolute values

Table 8, columns 1, 3, 4, 9 – 12: Date; in 10^{12} ; in 10^6 ; The same, for separate months from first day to first day; The same, from 15th day to 15th day; From 15th day to 15th day; Deviation of J' from J ; in per cents.

Line after 1.1: Total; Last line: Mean error

Footnote: We had no index for 15 Dec. 1922 and replaced it by the geometric mean of the indices for 1 Dec. 1922 and 1 Jan. 1923

Table 9, column 1: Date. The last 8 columns: Per cents.

Line after XII: Total; Last line: Mean values

Footnote to December, column 2: *Actual* magnitude unavailable and therefore calculated by applying the index interpolated for mid-December as a geometric mean; it is equal to the corresponding J_{1A}' .

Table 10, columns for m , M and J : τ given in weeks.

Last line: Total

Notes

1. Slutsky applied the Euler – MacLaurin summation formula. He made a mistake in its second term on the right side: the sign should have been *plus* rather than *minus*, see Korn & Korn (1968, § 4.8.5). After correction, the error studied by him changed its sign but retained its absolute value. O. S.

2. Slutsky's symbol for the natural logarithm was \lg_e ; sometimes, however, it was simply \lg , which is the standard Russian notation for the common logarithm. Believing that Slutsky always applied natural logarithms, I kept to the Russian (extremely convenient, in my opinion) notation \ln . O. S.

3. In 1930, Bazarov (real name, Rudnev) was called an *exposed saboteur* (Sheynin 1998, p. 537) and his fate could have only been tragic. O. S.

4. I was unable to understand the derivation of formula (38). O. S.

5. This formula is not readily found in the general mathematical literature whereas Korn & Korn (1968, § 20.5.3) provided it in a complicated form. O. S.

Bibliography

Bazarov, V. (1923, in Russian), A method of calculating the revenue accrued by the state from issuing paper money. *Ekonomich. Obozrenie*, No. 6, pp. 21 – 25.

Falkner, S. A. (1923, in Russian), The past and the future of the Russian system of issuing paper money. *Sozialistich. Khoziastvo*, No. 2 – 3, pp. 40 – 68.

Iasnopolsky, L. N. (1923, in Russian), Our money circulation at the epoch of the revolution. *Mestnoe Khoziastvo* (Kiev), No. 2, pp. 3 – 37.

Korn, G. O., Korn, Theresa M. (1961), *Mathematical Handbook* (1961). New York, 1968.

Krylov, A. N. (1911), *Lektsii o Priblizennykh Vychisleniakh* (Lectures on Approximate Calculations). Petersburg. [At least five more editions. Edition 6, Moscow, 1954.]

Schmidt, O. Yu. (1923, in Russian), Mathematical laws of the issue of paper money. *Vestnik Sozialistich. Akad.*, vol. 3.

Sheynin, O. (1998), Statistics in the Soviet epoch. *Jahrbücher f. Nat.-Ökon. u. Statistik*, Bd. 217, pp. 529 – 549.

VII

Mathematical Notes on the Theory of Emission¹

Matematicheskie zametki k teorii emissii.

Ekonomich. Bull. Koniunkturn. Inst., No. 11 – 12 (26 – 27), 1923, pp. 53 – 60

1. Pattern and Notation

The emission of government paper money can be thus described in a general way. An authorized organ of the government issues an appropriate act transferring a certain amount of money to some administrative body. Banknotes thus become legally existing as money; until then, they had only been not yet prepared manufactured technical means. Now, they are money issued by the government. To us, none of the subsequent technical details are of any consequence: where is the money kept; how much time does it take to make it available to local authorities when it concerns provinces; how much longer does it still stay in the special storehouse or any other such place; is it loaded into railway cars and transported, or is it already at the designed place, etc.

We will consider that, as soon as the act was issued, i. e., from the moment of the emission, it is entered into a government bank. We are only interested in the time when the money is paid out to employees, contractors et al, and thus put into circulation. In short, *the money is issued by the government when coming into its bank, then kept there for some time, and, finally, when going out, is put into circulation.*

We will call all the money that comes into a government bank until some moment t the *grand issue* and denote it by u_t ; all the money coming into circulation until that moment, the *small issue*, m_t ; and all the money kept at that moment in government banks, w_t . Obviously,

$$u_t = m_t + w_t. \quad (1)$$

We certainly will not consider either money collected by taxes etc. or issued instead of worn-out banknotes believing that those operations are entirely separated from issuing it.

The magnitude denoting the multiplier by which the general level of prices had risen, as compared with the moment for which the value of the money is considered equal to 1, is called the index and denoted by i_t . The value of all the money put into circulation is then the quotient of the small issue divided by the index:

$$M_t = \frac{m_t}{i_t}. \quad (2)$$

We will understand the revenue of the government from the emission as the sum of the values of money obtained by it during a certain time interval at the moments the money had been taken from the banks and put into circulation. If the sums of the money issued were $\Delta_1 m, \Delta_2 m, \dots, \Delta_n m$ and the index at the respective moments was i_1, i_2, \dots, i_n , the revenue J will be

$$J = \frac{\Delta_1 m}{i_1} + \frac{\Delta_2 m}{i_2} + \dots + \frac{\Delta_n m}{i_n}. \quad (3)$$

Denote the value of all the money put into circulation at the beginning of a given period by M_0 , then, following Prof. Falkner (1923), we will call the ratio

$$\eta = \frac{J}{M_0} \quad (4)$$

the relative efficiency of the emission.

2. The Forms of the Movement of Emission

Money comes into banks in separate amounts and in general several times daily, and leaves them in separate amounts as well, and in countless separate payments across the nation. A rigorous investigation of the form of the movement of the grand issue is only possible by a detailed study of the ledgers of the appropriate central office of Credbil²; investigation of the small issue demands a still more difficult study of the ledgers of the local banks, only possible by sampling. This cannot yet be accomplished and we must restrict our goal to *hypotheses*. The following statement seems to be sufficiently close to reality.

Emission flows discontinuously, but its separate partial masses are random fluctuations around some smooth continuous curve which can replace up to a vanishing error the real process that we are impossible to study directly.

Further preliminary hypotheses based on considerations of the behaviour of appropriate curves are:

1) For sufficiently short periods of time a straight line can sufficiently precisely represent the logarithm of the amount of paper money put into circulation ($\lg m$).

2) For sufficiently short periods of time a straight line can sufficiently precisely represent the temporal change of the value of the entire amount of money put into circulation (M).

The former statement can to some extent be [additionally] based on O. Yu. Schmidt's study (1923); the latter can be checked by any graph of an appropriate function over a short interval of time and is partly based on the coincidence of the pertinent calculations below with reality. Hypothesis 1) means that

$$\lg m_t = c + gt. \quad (5)$$

Let us assume a month, say, as a unit of time and denote $m_t = m_0$ at $t = 0$ and $m_t = m_1$ at $t = 1$. Denote by n the multiplier showing how many times had the amount of money put into circulation risen during unit time and call it the *rate* of the (small) emission. Then

$$n = \frac{m_1}{m_0}. \quad (6)$$

By issuing from (5) it is not difficult to derive now that

$$m_t = m_0 e^{gt}, \quad g = \ln n. \quad (7; 8)$$

However, it is sometimes convenient to replace (7) by

$$m_t = m_0 n^t. \quad (9)$$

Then, it follows from hypothesis (5) that

$$M_t = a + bt \quad (10)$$

and, taking into account (2) and (7), we find that

$$i_t = \frac{m_0 e^{gt}}{a + bt} = \frac{m_0 n^t}{a + bt}. \quad (11)$$

For $t = 0$ and 1 formula (10) provides

$$a = M_0, \quad b = M_1 - M_0. \quad (12)$$

Had we assumed that the level of prices changes proportional to the amount of money put into circulation, i. e., that

$$\frac{i_t}{i_0} = \frac{m_t}{m_0}, \quad (13)$$

then, owing to (9), we would have had

$$i_t = i_0 n^t. \quad (14)$$

Now, equation (11) can be written in another way. Replace a and b by their expressions (12), then

$$M_0 = m_0/i_0 \quad \text{and} \quad M_1 = m_1/i_1. \quad (14^*a, b)$$

Then, denoting the rate of the rise in the price level by

$$k = \frac{i_1}{i_0}, \quad (15)$$

we will arrive after simple transformations at

$$i_t = \frac{i_0 n^t}{1 + \left(\frac{n}{k} - 1\right)t}, \quad (16)$$

which will be transformed into (14) if $n = k$ (if the rate of emission coincides with that of the price level).

3. Revenue from the Emission and Relative Efficiency

We begin by admitting that the level of prices changes proportionally with the emission. Then the elementary revenue will be $dJ = dm/i$. However, taking into account (13) and (14*a),

$$i = \frac{i_0 m}{m_0} = \frac{m_0}{i_0} = \frac{m}{M_0},$$

so that

$$dJ = M_0 \frac{dm}{m}. \quad (17)$$

Integrating from $m = m_0$ to $m = m_t$, we find that

$$J = M_0 \ln \frac{m_t}{m_0}. \quad (18)$$

The studied period can be assumed as a unit, and $m_t/m_0 = n$, the rate of the emission, so that

$$J = M_0 \ln n. \quad (19)$$

Recalling (4), we obtain for the relative efficiency the expression

$$\eta = \ln n. \quad (20)$$

For $n < 2$ we have the known formula

$$\ln n = \lg n(1+n-1) = (n-1) - \frac{1}{2}(n-1)^2 + \frac{1}{3}(n-1)^3$$

and for n near to 1 that efficiency will be near $(n-1)$, i. e., near to the *norm of the increment* of the emission³. However, if, for example, $n = 2$, that norm is 100% and the relative efficiency 69,3% since $\ln 2 = 0.693$.

If the change in the price level is not proportional to the emission, and admitting both hypotheses of § 2, we have from (7)

$$dm = gm_0 e^{st} dt = gm_0 n_t dt,$$

and, dividing it by i from equation (11),

$$dJ = \frac{dm}{i} = g(a+bt) dt,$$

$$J = g \int_0^t (a+bt) dt = g \left(at + \frac{1}{2} bt^2 \right) = gt \left(a + \frac{1}{2} bt \right).$$

Owing to (7) and (12) and noting that $b_t = M_t - M_0$ follows from (10), we arrive at

$$J = \frac{1}{2}(M_0 + M_t) \ln \frac{m_t}{m_0}, \quad (21)$$

an expression similar to (18), only M_0 is replaced here by the appropriate arithmetic mean. For unit time

$$J = \frac{1}{2}(M_0 + M_t) \ln n. \quad (22)$$

Removing M_0 from the brackets and replacing M_0 and M_1 within the brackets according to formula (2) and recalling also formulas (6) and (15) [M_0 certainly remains within the brackets in the denominator], we come to

$$J = \frac{1}{2} \left(1 + \frac{n}{k} \right) M_0 \ln n \quad (23)$$

and the relative efficiency will now be

$$\eta = \frac{1}{2} \left(1 + \frac{n}{k} \right) \ln n. \quad (24)$$

We see that if $n > k$, i. e., if the rate of emission is higher than that of the prices, the relative efficiency must be higher than $\ln n$, and otherwise lower. We may therefore somewhat tentatively call *normal relative efficiency* the magnitude expressing the relative efficiency when these ratios are equal to each other, see formula (20).

4. The Revenue from the Emission and the Relative Efficiency in 1922

In our literature devoted to studying emission, it is often assumed that the time that the money is kept in banks (and during journeys) is two weeks. Therefore, the figures of the *grand* issue for the first day of the month, – and until 1 Sept. 1922 only they were known, – are supposed equal to the figures of the small issue for the middle of the month and the revenue from the emission is calculated according to an approximate formula

$$J^1 = \frac{m_1 - m_0}{i_{1/2}}. \quad (25)$$

Here, m_0 and m_1 are the small issues for the middle of the first and second months respectively, and $i_{1/2}$ is the index for the middle of that period, i. e., for the first day of the next month.

In our previous work [vi] based on hypotheses 1) and 2) of § 2 we proved that formula (25) is erroneous not more than by 2% for the period when the emission is increasing less than twofold and mostly not more than by 1%

when applied to monthly periods; and if hypothesis 2) is valid, this holds independently from the rate of the index.

Here, I only provide as an illustration the revenue J^1 calculated according to formula (25) for 11 months of 1922 (from 15 Jan. to 15 Febr., from 15 Febr. to 15 March, ..., from 15 Nov. to 15 Dec.) and compare it with the revenue J calculated for those same periods according to formula (22).

Note that I calculated J separately for each halfmonth by issuing from the indices provided by *Statistika Truda* (Statistics of Labour) which I interpolated bringing them to the middle of the appropriate months. Table 1 thus provides figures calculated by adding up the appropriate results in pairs.

The proximity of the figures determined by such different methods is an empirical check of the initial hypotheses. An almost complete coincidence of the results provides a cause for thinking that the monthly fluctuations include random elements [components] smoothed when being summed up. Without a more detailed investigation, carefulness does not allow more categorical conclusions.

In Table 2 we offer the result of our calculations [vi, Table 3]: the rates of the emission and index (n and k), the normal relative efficiency ($\ln n$) and the theoretical and empirical relative efficiency, formulas (24) and (4). All the figures are calculated for monthly periods, from the first day of a month to the first day of the next one. The figures of the revenue J were calculated by the method described above, and as near to reality as was possible. The result is shown on Fig. 1.

Theoretical efficiency is very near to the empirical, and they both are situated higher than the curve of the normal relative efficiency when the rate of the emission is higher than the rate of prices (from May to September) and lower otherwise (January – April and October – November). When the rates coincide, the mentioned curves completely or almost coincide with the curve of the normal relative efficiency (April and December). The chosen period belongs to those with most sharp oscillations and the acceptability of the empirical check apparently testifies that our hypotheses reflect reality without considerably corrupting it.

5. The Pattern of the Movement of Money in the Banks

The fate of the banknotes in government banks can be thus sketched. Arriving in a bank, they move from entrance to exit as a current in a channel, but in such a manner that all the separate drops preserve their relative positions; those entering first, exit earlier and vice versa. This pattern does not corrupt reality since we are interested not in the banknotes in the technical sense, but rather understanding them as units of purchasing power, in how one banknote can replace another one as though transferring to it the role of a representative of the emission at a given moment.

Let a, b, \dots, r, s be banknotes ordered in time, each at the moment of its issue, i. e., of its entering the bank; k is the banknote exiting at some moment t , and s , is entering at the same moment. Then all the amount of money put into circulation up to a given moment, that is, m^4 , will be represented in our pattern by letters $a - k$, and all the money in the bank, by letters $l - s$ with letters $a - s$ being the total of all the issued banknotes, i. e., the grand emission u . Let the distance in time from k to s be τ , or the interval

of time between the banknote k entering the bank and exiting from it, that is, the time it is kept at the bank.

And so, if t is the moment of its entering, the moment of its exit, and simultaneously of the last banknote s entering the bank, will be $t + \tau$. At moment t all the issued banknotes (the grand emission) were exhausted by the series a, b, \dots, k ; and at moment $t + \tau$ those same banknotes represent the totality of money exiting the bank and put into circulation. At those moments their sum should be denoted by u_t and $m_{t+\tau}$, and we have thus arrived at the main equation of the movement of money in the bank, absolutely independent from any particular suggestion about the rapidity and rate of the emission:

$$u_t = m_{t+\tau}. \quad (26)$$

6. Determining the Period during which Money Is Kept in the Bank in the General Case

Let (Fig. 2) AB and $A'B'$ be the curves of the grand and the small issue, u and m . The amount of money entering the bank, u , and put into circulation, m , are marked off parallel to the vertical axis, and time, parallel to the horizontal. Let us take some point C on AB and draw a straight horizontal

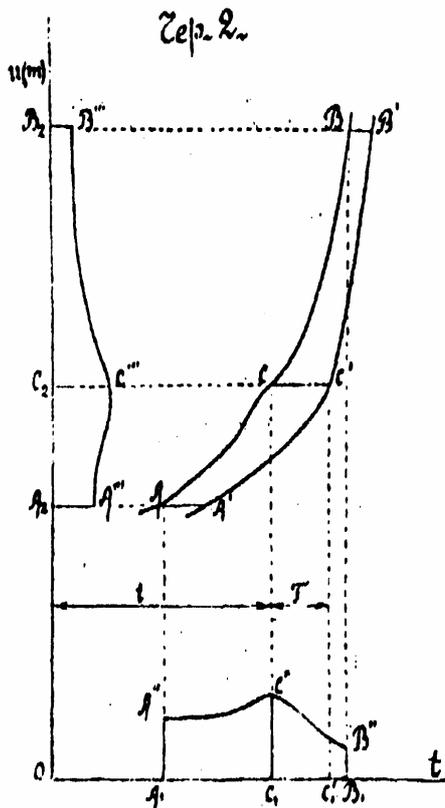


Fig. 2. Curves of the grand and the small issues

Horizontal axis: with letters $A_1, C_1, C'_1,$ and B_1

Vertical axis: with letters O, A_2, C_2, B_2 ; $u(m)$ means u or m

Curve near horizontal axis: with letters A'', C'', B''

Curve near vertical axis: with letters $A, B, C,$ each with three strokes

Main curves: with letters $A, C, B,$ and A', C', B'

Intervals $t = C_2C, \tau = CC'$

line passing through it until meeting $A'B'$ at some point C' . The segment CC' will be the small issue $m_{t+\tau}$ at a later moment when $m_{t+\tau} = u_t$. From the above, it follows that $CC' = \tau$, the time interval during which the elementary amount of money entering the bank at moment t is kept there.

For differing amounts of money entering the bank at different moments that magnitude will generally vary, but it is always equal to the distance as measured along the horizontal line between the indicated curves. When transferring these distances to the abscissa and ordinate axes as shown on Fig. 2, we will have two figures. The area of the first one divided by the base of the figure $A_1B_1 = t_2 - t_1$ is the mean time interval for the money entering the bank, then to be kept there. Equal weight will be here assigned to equal elementary time intervals of the entering. The area of the second figure divided by the base of the figure $A_2B_2 = u_{t_2} - u_{t_1}$ will also be the mean duration of the money kept in the bank, but weighted by the amounts of the entering money.

In other words, we will obtain the first mean by separating the period under discussion into equal intervals, determining τ separately for each of them taking the simple arithmetic mean of all those values and establishing the limit to which that magnitude tends when the number of the intervals is being increased. The second mean will be the arithmetic mean of all the values of τ corresponding to each entering rouble. Those means do not generally coincide, and their logical meaning differs; the first of them seems more suitable for characterizing various periods of the emission.

As an example, we will determine the mean duration (of the first kind) of the grand issue for the period from 1 Oct. 1922 to 1 July 1923. Owing to the kindness of the Currency Directorate, we have unpublished data for that time. Namely, we have the figures for the grand issue u for the first day of each month and of the small issue m on the 1st, 8th, 16th and 23rd day of each month. These figures are shown in Table 3. Shown in bold type is each least figure of the small issue out of the two, between which the corresponding figure of the grand issue of the previous first day of the month is situated.

The method of calculation is determined by the data. We have to interpolate for the moment when $m = u$; the intervals roughly equal to a week are rather short, and we have only applied simple interpolation. Thus, the moment when m is equal to the grand issue for 1 March (3652.7) is⁵

$$7 + 8 \frac{3652.7 - 3449.9}{3713.8 - 3449.9} = 7 + 8 \cdot 0.7685 = 13.148 \text{ days.}$$

I carried out similar calculations in all the cases allowing for the period being sometimes seven and sometimes eight days. I increased each of the numbers thus obtained by 3 days because m actually represents the amount of money not put into circulation, but entering the bank's working fund which may keep money needed for three days. We have no information about how the banks are using that possibility and presume that in the mean the money is kept in that fund exactly 3 days.

For τ calculated above we therefore have 16.148 days. All the magnitudes thus obtained are entered in column τ of Table 4 [but still without adding those 3 days]. The next column contains the arithmetic mean of the two

values of τ for the beginning and the end of the appropriate month representing the approximate values of the duration of the emission of the appropriate month in the banks.

The same figures are shown in the last column in mean “weeks” which we understand as 1/4 of the mean month during the period from 1 Oct. 1922 to 31 July 1923. That interval is 304 days, so that the mean month is 30.4 days, and the mean “week” is 7.6 days. We thus see that in cases known to us the time that money had been kept in the bank fluctuated between $\tau = 7.2$ and 38.5 days (column τ) or between 9.07 and 31.79 days (next column) or, finally, between 1.19 and 4.18 of the “mean” quarter of month. The general mean is 17.667 days = 2.32 mean “week” = 0.581 of mean month, that is, (1/2 month + 2 1/2 days).

It thus occurs that at least for the considered period in the mean the hypothesis, which several authors applied in our literature for various calculations concerning the half-month duration of time between issuing the money and its coming into circulation, apparently reflects reality. Whether that mean can be attributed to the entire period of 1918 – 1922 or not, we ought, however, to leave unanswered.

7. Determination of τ for the Normal Movement of the Emission

Had the movement of the emission been uniform and at the same rate for both the grand and the small issues, the ratio of the money in the bank to the entire amount of the emission would have remained constant, and the time τ that the money is kept in the bank could have been determined directly from equation (26), see below.

Under constancy of the rates of both emissions and their equality to each other we should have, according to equation (9),

$$m_t = m_0 n^t \text{ so that } u_t = u_0 n^t, \quad (26^*)$$

but then it would have followed that

$$\frac{u_t}{m_t} = \frac{u_0}{m_0} = \text{Const} (= \lambda). \quad (27)$$

Denote the ratio of the amount of money kept in the bank to all the money put into circulation by β ,

$$\frac{w_t}{m_t} = \beta, \quad (28)$$

and divide equation (1) by m_t , then

$$\frac{u_t}{m_t} = 1 + \frac{w_t}{m_t},$$

and, owing to (27),

$$\beta = \lambda - 1. \quad (29)$$

In this case the equation (26) is solved very easily. First of all, since any moment can be assumed as being initial, we choose $t = 0$, then it will take the form

$$u_0 = m_\tau. \quad (30)$$

Then, owing to (26*),

$$m_\tau = m_0 n^\tau, \quad u_0 = m_0 n^\tau. \quad (31; 32)$$

Dividing both parts of (32) by m_0 and recalling (27) we will have

$$n^\tau = \lambda, \quad \tau = \frac{\lg \lambda}{\lg n}. \quad (33; 34)$$

As an illustration, we apply this formula for determining τ by issuing from the data provided in § 6. However, since our empirical magnitudes do not obey the indicated conditions quite strictly, we ought to represent (33) in a somewhat modified way. Indeed, under the premises admitted above, the rates of both issues are the same, so that we may assume n equal either to m_1/m_0 or u_1/u_0 , and λ equal to u_1/m_1 or u_0/m_0 or u_τ/m_τ . For the empirical data, however, it is not so.

Returning to the strict equation (30), we consider expressions (9) or (31) as an approximate formula. Then, τ is not yet known and that expression cannot be applied for deriving n . We therefore assume that during the appropriate month (which is the unit of time) the rate of the small issue remains roughly constant. Denoting the value of that issue in the beginning and end of month by m_0 and m_1 , we thus approximately have $n = m_1/m_0$.

Equation (32) will then be

$$u_0 = m_0 \left(\frac{m_1}{m_0} \right)^\tau, \quad \tau = \frac{\lg(u_0 / m_0)}{\lg(m_1 / m_0)} = \frac{\lg u_0 - \lg m_0}{\lg m_1 - \lg m_0}. \quad (35)$$

For example, in Nov. 1922 $u_0 = 1217,6 \cdot 10^{12}$ nominal roubles, $m_0 = 1095,5$, $m_1 = 1479,2$,

$$\tau = \frac{\lg 1217,2 - \lg 1095,5}{\lg 1479,2 - \lg 1095,5} = 10,572 \text{ days.}$$

As indicated above, 3 more days should still be added to provide 13.572 days. After such calculations for all months from Oct. 1922 to July 1923 we obtain Table 5 similar to Table 4. When comparing the tables with each other, we see that our method provides rough, but not entirely unsatisfactory approximations, whereas in the mean we arrive at almost the same value of τ : 2.23 of a mean "week" instead of 2.32. Instead of $\tau = 1/2$ of month + 2 1/2 days, here $\tau = 1/2$ of month + 1 3/4 days, the difference only being 3/4 days.

The changes in time are absolutely similar as seen in the comparison of the two curves of τ on Fig. 3. This method can apparently be applied if the data are insufficient for making use of the more precise first method.

8. The Size of the Working Fund

We know absolutely nothing about it, and its estimate is only possible by assuming that the local institutions are fully enjoying the right to keep an amount of cash necessary for three days. It is hardly possible to ignore that sum when determining the amount of money in circulation since (see below) it can reach almost 10% of a monthly emission.

We considered the small issue (m) as money put into circulation. Now, we will take a stricter look and consider it, as it is in reality, the amount of money put into the working fund as into the final stage before being actually circulated. Now, we denote by m the amount of money really put into circulation, whereas the small issue, i. e. the money put into the working fund until the given moment, will now be u' . Then, the cash in the working fund is Δ , and in the bank (and en route) without the working fund, w' and the total cash, $w = w' + \Delta$. The grand issue will be denoted as previously by u , then

$$u = m + w, u = u' + w', u' = m + \Delta. \quad (36a, b, c)$$

It is not difficult to realize that all the equations (36) are similar to each other and the magnitudes occupying the same places can be mutually changed. For example, equation (9) will become

$$u = u_0 n^t, u' = u'_0 n^t, m = m_0 n^t,$$

with n being u_1/u_0 , u'_1/u'_0 , and m_1/m_0 , respectively. For (36c) equation (30) will be

$$u'_0 = m_\tau. \quad (38)$$

Since

$$m_\tau = m_0 n^\tau = (u'_0 - \Delta_0) n^\tau,$$

we have from (38)

$$u'_0 = (u'_0 - \Delta_0) n^\tau, \Delta_0 = u'_0 \left(1 - \frac{1}{n^\tau} \right). \quad (39)$$

We will derive from (39) a convenient approximation for a sufficiently small τ (for example, $\tau = 1/10$ of a month). Representing $1/n^\tau = n\tau$ as $e - \tau \ln n$, we will have⁶

$$1 - \frac{1}{n^\tau} = \tau \ln n - \frac{1}{2} (\tau \ln n)^2 + \frac{1}{6} (\tau \ln n)^3 - \dots \quad (40)$$

In our data $n \leq 2$, therefore $\ln n < 0.693$ and

$$1 - (1/2^{1/10}) = 0.06931 - 0.00240 + 0.00055 - 0.00001 + \dots$$

provides the upper boundary of error. Neglecting all terms of (40) beginning with the second one, we will approximately have

$$\Delta_0 = u'_0 \tau \ln n. \quad (41)$$

For a month, the small issue is $u'_1 - u'_0$ and we may write

$$u'_0 = \frac{u'_0}{u'_1 - u'_0} (u'_1 - u'_0) = \frac{u'_1 - u'_0}{n - 1}.$$

Inserting this expression into (41) we finally have

$$\Delta_0 = \tau (u'_1 - u'_0) \cdot \frac{\ln n}{n - 1}.$$

The first factor, $\tau (u'_1 - u'_0)$, is the part of the monthly issue proportional to τ , the second one is a correction, or coefficient equal, as it is not difficult to prove, to u'_0 / u'_m where the denominator is the monthly mean of the small issue⁷. The value of that coefficient is given in Table 6.

For a month of 30 days and the working fund then receiving $100 \cdot 10^{12}$ roubles at the beginning of the month that fund will have not $(1/10) \cdot 100 \cdot 10^{12}$ roubles, but, for $n = 2$ and 1.7 , only $6.93 \cdot 10^{12}$ and $7.58 \cdot 10^{12}$, etc.

Explanation of Tables and Figures

I am only explaining them; one exception is Fig. 2 inserted into the text above. O. S.

Table 1 (§ 4). Shows months of 1922, J^1 , J and ΔJ .

Table 2 (§ 4). See explanation in text.

Fig. 1 (§ 4). See explanation in text.

Table 3 (§ 6). Shows the months from Oct. 1922 to June 1923, u and m .

Table 4 (§ 6). Explanation in text.

Table 5 (§ 7). As explained in text, it is similar to Table 4.

Fig. 3 (§ 7). Shows almost coinciding broken lines of the mean duration of money in banks as calculated by both methods and summarized in Tables 4 and 5 respectively.

Table 6 (§ 8). Shows $n = 1.1(0.1)2.0$, $\ln n$ and $\ln n/(n - 1)$.

Notes

1. The first half of this paper is a brief description of some of my main previous findings [vi] and the second half (§§ 5 – 8) is a natural continuation of the first §§ 1 – 4. E. S.

2. This is obviously an abbreviation of *Creditnye Biletu* (Certificates of Credit, promissory notes, banknotes). O. S.

3. This explains the mistake made by Prof. Falkner (1923, p. 54) who called the norm of the increment the rate of the issue (which is hardly advisable) and considered it, i. e., our $(n - 1)$, so to say a normal measure of relative efficiency. E. S.

4. Slutsky wrote out all the intermediate letters except j . The letter m that appeared separately has nothing in common with banknote m . O. S.

5. Calculations with superfluous digits are frustrating the more so since Slutsky later adds 3 days very roughly estimated. Same comment is valid for § 7 below and will be made in more detail in Note 6 to [xiv].

3449.9 in the following calculation is the figure for 8 March (provided in bold type, see explanation in text) and 3713.8, the figure for 16 March. O. S.

6. The right side of (40) is the expansion of

$$1 - e^{-t \ln n} = 1 - \exp(\ln n^{-t}) = 1 - (1/n^t)$$

which is the expression contained in (39). Just above (40) something is wrong in “Representing ...”. O. S.

7. Indeed,

$$u'_m = \int_0^1 u'_0 n^t dt = u'_0 \left[\frac{n^t}{\ln n} \right]_0^1 = u'_0 \frac{n-1}{\ln n}, \quad \frac{u'_0}{u'_m} = \frac{\ln n}{n-1}. \text{ E. S.}$$

Bibliography

Falkner, S. A. (1923, in Russian), The past and the future of the Russian system of issuing paper money. *Sozialistich. Khosiasvo*, No. 2 – 3, pp. 40 – 68.

Schmidt, O. Yu. (1923, in Russian), Mathematical laws of the issue of paper money. *Vestnik Sozialistich. Akad.*, vol. 3.

VIII

On the Law of Large Numbers

K voprosu o "Zakone Bol'shikh Chisel". *Vestnik Statistiki*, No. 7 – 9, 1925, pp. 1 – 55

I

1. The very idea of the "law of large numbers" is based on a concept that had been until quite recently left unnamed in spite of our acquaintance with it for more than two hundred years, that is, from the time that Jakob Bernoulli's celebrated *Ars Conjectandi* was published. The matter concerns the relation between magnitudes owing to which one of them is the limit in the probability-theoretic sense, or, shorter, is a *stochastic limit*¹. That concept was not clearly defined or separated as such in proper generality and had not therefore been distinctly noticed in each necessary instance. As it always occurs in such cases, neither did it attract the deserved attention. Consequently, extremely elementary [easy to notice] gaps had been left even in the first chapters of the calculus of probability. They would have been certainly bridged long ago had the notion of stochastic limit been named in due time.

Thus, the courses in that discipline lack theorems similar to those of the doctrine of usual limits: the limit of the sum (of the product) is equal to the sum (the product) of the limits, etc. At the same time, an even much more general theorem can be extremely easily proved, viz., that the stochastic limit of a function is equal to the function of the stochastic limit provided that that function is continuous at least near the appropriate value of the argument and its parameters do not depend on the relevant argument².

In general, the sole formulation of the concept of stochastic limit suggests a number of theoretical issues, clearly distinguishes the corresponding kind of problems and as though all by itself reveals the lack of generality in a number of statements. A *name* is not an empty word; what is left in science without it cannot possess completeness of being.

I consider the mathematical theory of stochastic limit elsewhere. Here, I only aim at sketching its general concept so as to revise on that basis some aspects of the problem of the LLN³.

2. For revealing the essence of the stochastic limit I consider the situation necessary for the existence of the Bernoulli theorem and I allow myself to remind readers its premises. A number of *trials* is made in each of which some event can either occur or not. Its probability p is one and the same in each; it does not depend on the outcome of previous trials whose series can be continued indefinitely. This is illustrated by the known pattern of extracting white and black (say) balls from an urn, returning each extracted ball back and "properly" shuffling the urn's contents before each subsequent trial.

Suppose that n trials are made and the event happened m times. The magnitude m/n , i. e., the frequency of the event, ought to take some of the following values

$$0/n, 1/n, 2/n, \dots, (n-1)/n, n/n \quad (1)$$

with each of these cases possessing a certain probability that can be calculated according to known rules. Frequencies near the probability p will have maximal probabilities and the further they are from that p , the lower will be their probabilities.

It is possible to look for the probability that the frequency will be not less than $p - \varepsilon$ and not greater than $p + \varepsilon$, that is, that its deviation from probability will not exceed ε in absolute value. Denote that probability by

$$P_{(0)}^{(\varepsilon)} \left| \frac{m}{n} - p \right| \quad (2)$$

which means that these deviations will take some values between 0 and ε . For deriving it, we ought to separate cases corresponding to the condition

$$0 \leq |(m/n) - p| \leq \varepsilon \quad (3)$$

from the pattern (1) and to add up their probabilities.

The Bernoulli theorem is known to mean the following. When determining that probability given an infinitely increasing number of trials, n_1, n_2, \dots , the corresponding probabilities

$$P_{(0)}^{(\varepsilon)} \left| \frac{m}{n_1} - p \right|, P_{(0)}^{(\varepsilon)} \left| \frac{m}{n_2} - p \right|, \dots \quad (4)$$

will constitute a series of magnitudes boundlessly tending to unity whatever is the value of ε chosen beforehand. However small is that ε , it will be possible to determine such a number of trials n that, given any greater, the difference between the corresponding probability and 1 will remain less than any given beforehand ε .

Compressing all this into a short formula, we arrive at the following formulation of the Bernoulli theorem: Given a boundless series of independent trials in each of which the probability of some event is one and the same constant magnitude, *the probability of the deviation of the frequency m/n from the probability of its happening in a separate trial p will not exceed in absolute value any arbitrarily chosen and however small magnitude ε if the number of trials n tends to infinity, has a limit equal to 1, or*

$$\lim P_{(0)}^{(\varepsilon)} \left| \frac{m}{n} - p \right| = 1, n \rightarrow \infty. \quad (5)$$

And we ought to indicate that it seems highly inexpedient that no shorter expression by means of some suitable term and its symbol is in existence for such a complicated relation (5) between a number of magnitudes. It is just as inexpedient as having to repeat entirely every time the definitions of *limit*, *derivative*, *integral* instead of using these terms. Indeed, we are dealing with such a relation that plays an important role in a number of chapters of the calculus of probability and has to be discussed very often. Surprisingly this circumstance had been for so long unnoticed.

Formula (5) tells us (to express it more briefly) that the probability of deviations not exceeding a given value tends to 1. As a corollary, we may conclude that *the boundaries of the deviations to be expected with a definite probability however near to unity are ever nearer approaching each other.* That form allows us to perceive especially vividly the similarity of the considered relation with the notion of limit. Keeping for the time being to the same example, we will therefore say that under the conditions of the Bernoulli theorem the probability p of the occurrence of the event in a separate trial p is the limit in the probability-theoretic sense for the event's frequency m/n , given an infinitely increasing number of trials n .

Or, more briefly: probability p is the *stochastic limit* of frequency m/n as $n \rightarrow \infty$. And still more briefly:

$$p = \lim_B (m/n), n \rightarrow \infty. \quad (6)$$

The letter B distinguishes the new notation from the usual symbol of limit. I propose that notation in honour of Jakob Bernoulli as an abbreviated expression of a parallel term, stochastic or Bernoullian limit, *limes stochasticus vel Bernoullianus*. When a magnitude [a number] C is the usual limit of some magnitude z , we may write down an *approximate equality*⁴

$$z \neq C, t \rightarrow \infty \quad (7)$$

valid for sufficiently large values of the independent variable t and reminding that in the process of its unbounded increase the error of (7) unboundedly decreases.

From the above follows a relative [see below] rightfulness of the method of expression proposed by Prof. V. I. Romanovsky. Concerning the magnitudes involved in the Bernoulli theorem it offers a *stochastically-approximate equality*⁴

$$p \neq m/n, \text{ modo Bernoulliano}, n \rightarrow \infty. \quad (8)$$

It ought to be implied here that, given an infinitely increasing n , the boundaries of the errors to be expected with any given probability however near to 1 infinitely approach each other. Rigour of such notation will only be maintained when applying a special sign for approximate equality, as we have done it in (7) and (8), and indicating the infinitely increasing independent variable.

In any case, similar to the logically main symbol

$$\lim z = C, t \rightarrow \infty$$

being in the largest number of instances most convenient, only the symbol

$$\lim_B x = C, n \rightarrow \infty$$

directly corresponds to the essence of stochastic limit whereas the symbol of the formula (8) should be considered as derivative and supplementary. Indirectly this idea is corroborated by the fact that essentially the same symbolical notation as that, provided by V. I. Romanovsky, had occurred

long ago but disappeared with absolutely no trace. Thus, in Poisson (1837, p. 139, in passing) we meet more than once with the expression: such-and-such a magnitude is equal to another one *à très peu près et très probablement* (almost precisely and highly probably). Formally speaking, this is not quite clear, but in essence it certainly implies the same idea of a stochastically approximate equality.

3. The presence of some connection between the probability of an event, its frequency and the number of trials was certainly noted even before Jakob Bernoulli. It is unquestionable, however, that he was the first to bring to complete clarity the logical structure of the relevant relations. When discussing the empirical determination of probabilities of events by observing their frequencies, Bernoulli himself (1713/2005, Chapter 4 of pt. 4, p. 29) remarked that that method was “not new or unusual”, and he acknowledged that

Neither can anyone fail to note also that it is not enough to take one or another observation for such a reasoning about an event, but that a large number of them is needed. Even the most stupid person, all by himself and without any preliminary instruction, being guided by some natural instinct (which is extremely miraculous), feels sure that the more such observations are taken into account, the less is the danger of straying from the goal.

His next phrase throws light on how he perceives his aim:

Although this is known by nature to everyone, its proof, derived from scientific principles, is not at all usual and we ought therefore to expound it here. However, I would have estimated it as a small merit had I only proved that of which no one is ignorant. Namely, it remains to investigate something that no one had perhaps until now run across even in his thoughts.

When pondering this over, and especially bearing in mind the following, we will understand what had the author wished to say. His idea was undoubtedly that the situation as it existed before him was not such that “proof derived from scientific principles” had only to be found for some proposition.

The subject itself was not yet made clear and the relations between the appropriate magnitudes were perceived vaguely. Had the problem been formulated as a purely mathematical exercise in the theory of combinations⁵, some aptitude would yet be required for achieving the aim, but we would not have been then surprised by the author’s genius. He himself (Ibidem, pp. 30 and 31) turns our attention to the circumstance that his discovery consists in the formulation itself of the problem:

It certainly remains to inquire whether, when the number of observations thus increases, the probability of attaining the real ratio between the number of cases in which some event can occur or not, continually augments so that it finally exceeds any given degree of certitude. Or [to the contrary], the problem has, so to say, an asymptote; i. e., that there exists such a degree of certainty which can never be exceeded no matter how the observations be multiplied.

To avoid false understanding, it ought to be noted that the ratio between the numbers of cases which we desire to determine experimentally, is accepted not as precise and strict (because this point of view would have led to a contrary result and the probability of determining the real ratio would have been the lower the more observations we would have taken)⁶, but that this ratio be accepted with a certain latitude, that is, contained between two limits [boundaries] which could be taken as close as you like.

I allowed myself to include these long extracts so as to show to what extent did Bernoulli value his merit exactly in analytically revealing the problem and how clearly did he establish all the indications characterizing the connection between a magnitude and its stochastic limit. Historical fairness demands that we acknowledge not only the authorship of the celebrated theorem named after him, but also his discovery of the concept of stochastic limit. Exactly that term is too expressive and convenient for not being evaded and I thought it would be proper to attach to it the parallel name of Bernoullian limit by assuming the symbol \lim_B and establishing it with all the necessary degree of generality⁷.

4. When defining the notion of stochastic limit in general, it is of course necessary to abstract ourselves from the conditions of the particular case which we considered for the initial introduction of that stochastic formation. We ought to go in several directions beyond the boundaries within which that notion is usually applied.

a) First of all, it is clear that not only frequencies, not only arithmetic means but an immeasurable number of other magnitudes can also have a stochastic limit. For example, denoting expectation as usual by E we will have the correlation coefficient between random variables x and y

$$r_{x,y} = \frac{E(x - Ex)(y - Ey)}{\sqrt{E(x - Ex)^2 E(y - Ey)^2}} \quad (9)$$

and its empirical ersatz, the only calculable, given a series of empirical values of those variables in n trials,

$$\rho_{x,y} = \frac{(1/n) \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(1/n) \sum_{i=1}^n (x_i - \bar{x})^2 (1/n) \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (10)$$

Here, \bar{x} and \bar{y} are the arithmetic means of the corresponding magnitudes. The coefficient (10) is not an arithmetic mean, but it certainly is a random variable. Depending on the random choice of the elements of the appropriate samples, it can take various values with different probabilities for one and the same coefficient (9). It is proved (Chuprov 1922, p. 267/2004, beginning of § 4.2A; 1924, p. 42) that the expectation of (10) is not in general equal to (9). However, with a constant $r_{x,y}$, and an increasing number of trials, it tends to $r_{x,y}$ as to its limit.

The following relations will therefore take place. Suppose we have s samples of n trials each from one and the same urn and denote the arithmetic mean of the s empirical coefficients of correlation by $\bar{\rho}_{x,y}$. When increasing the number of samples s , it will *stochastically* approach $E\rho_{x,y}$ as its limit

$$\lim_B \bar{\rho}_{x,y} = E\rho_{x,y}, \quad s \rightarrow \infty. \quad (11)$$

If, according to another pattern, we will infinitely increase the number of cases in a sample,

$$\lim_B \rho_{x,y} = r_{x,y} = \lim E\rho_{x,y}, \quad n \rightarrow \infty. \quad (12)$$

I think that this example shows sufficiently clearly how wide can the notion of stochastic limit be applied. When formulating it, we should therefore mention not some separate kinds of random variables, but a random variable in general. That will be a problem not of defining stochastic limit, but of its theory; viz., of establishing conditions under which a random variable has a stochastic limit.

b) The magnitude that is a limit should not be burdened by superfluous restrictions either. We ought not to be tempted by the fact that until now in all known cases such limit was an expectation, or, as in the last example above, its limit. Even be that restriction always valid, it should not have been mentioned in the definition but rather constitute the subject of a special theorem. It is not difficult, however, to prove that it is not so.

Let the possible values of a random variable and their probabilities be

	I	II	III	IV	
$x:$	$-n$	$1-1/n$	$1+1/n$	n	(13)
$p:$	a/n	$(1/2)[1-(a+1)/n]$	$(1/2)[1-(a+1)/n]$	$1/n$	

This pattern can be illustrated by an urn having n balls, a of them of the first kind, one ball of the fourth and an equal number of balls of the two other kinds. Let the numbers of the first line represent the loss of a gambler in case I and his gains in the other cases when a ball of the corresponding kind is being extracted.

Suppose also that after each game one ball of the II and one of the III kind are added, but the number of balls of the other kinds remain constant. The sum of the probabilities in the two cases, II and III, will evidently become arbitrarily close to 1 so that the probability of winning will also behave the same way.

And so, the stochastic limit of gain is 1 whereas, when multiplying the corresponding numbers of both lines, we will obtain the expectation of gain equal to

$$Ex = -(a-2) - (a+1)/n. \quad (14)$$

For $a = 1$

$$\lim Ex = 1 - \lim_B Ex,$$

and, if $a > 2$,

$$\lim Ex = -(a - 2) < 0.$$

Its absolute value differs from the stochastic limit arbitrarily much if only a is taken sufficiently large. A banker, the partner of our gambler, will have a positive expectation of gain, but the gambler will win almost certainly.

c) Our last example also illustrates a third point that should be paid attention to when formulating the notion of stochastic limit. Namely, the consecutive values of probabilities taken in some arbitrarily narrow boundaries of deviations can be a function not of the number of trials, but of some other independent variable. In the last example above that variable was the number of balls in the urn. True, it increased with the number of trials, but that condition may be abandoned without changing anything in the result.

Or, here is another example. The probability of an event occurring all n times in n trials, given a constant probability p of its happening in a separate trial, is p^n and tends to unity if n is constant and p approaches 1. Denote the number of occurrences of the event in n trials by m , then

$$\lim_{(0)} P_{(0)}^{(e)} | m - n \approx 1, \lim_B (m) = n, p \rightarrow 1. \quad (15; 16)$$

What we are discussing now can be essentially important for correctly formulating many propositions in most various fields. Thus, for an unbounded space with matter spreading in it from a single centre, the probability of having a certain number of molecules in a unit volume situated at a given distance from that centre should lower. Certain premises are possible for that probability to lower unboundedly but never to vanish completely. It would have been wrong to say then that the limit of density of the matter is zero; rather we ought to state that the stochastic limit of that density equals zero as the distance from the centre increases unboundedly. Distance will be the independent variable.

d) Finally, the example covered by formulas (15) and (16) shows that the independent variable does not necessarily tend to infinity; it can take a number of values and, for example, approach some finite magnitude. Allowing for that circumstance as well, we may rigorously formulate the notion itself of that limit.

5. If for each value of a magnitude as well as for each interval of its values there exists a definite corresponding probability, that magnitude will be a *random variable*. It is possible to simplify that definition since condition $x = c$ is a particular case of $a \leq x \leq b$, namely when $a = b = c$. That formulation covers both the case in which the random variable can only take definite discrete values, so that the probability of its being situated in any interval not including any of them is zero, and the case in which it can take all values in some continuous interval or even all real values from $-\infty$ to $+\infty$.

The entire set of probabilities uniquely corresponding to intervals (or values) of a random variable x constitute its *distribution of probabilities*; denote it by Ω_x . Consider some independent variable ϕ and let each of its values uniquely correspond to a definite distribution of some random variable so that a definite sequence

$$\Omega_x^{(1)}, \Omega_x^{(2)}, \dots, \Omega_x^{(i)}, \dots \quad (17)$$

will correspond to each sequence of values of φ

$$\varphi^{(1)}, \varphi^{(2)}, \dots, \varphi^{(i)}, \dots \quad (18)$$

Now we may say that the distribution of a given random variable x is a function of a given independent variable φ , and x is *stochastically connected* with φ ⁸. If one and the same distribution corresponds to any values of x , we say that x is *stochastically independent* from φ and *stochastically dependent* otherwise. Let in addition a boundless sequence

$$\varphi_1, \varphi_2, \dots, \varphi_i, \dots \quad (19)$$

be given. For a definite value of φ , it is possible to determine the probability that the deviation of x from some constant will not exceed in absolute value some arbitrary ε . Then an unbounded sequence of probabilities

$$P_{(0)}^{(\varepsilon)} |x - c|_{\varphi = \varphi_1}, P_{(0)}^{(\varepsilon)} |x - c|_{\varphi = \varphi_2}, \dots, P_{(0)}^{(\varepsilon)} |x - c|_{\varphi = \varphi_k}, \dots \quad (20)$$

will correspond to sequence (19). If for any however small ε and α such φ_k can be found that for any $\varphi_n, n > k$, the corresponding probabilities (20) will be higher than $(1 - \alpha)$, the constant magnitude c will be the stochastic limit of x for the sequence (19).

Shorter, *if a random variable x is stochastically connected with an independent variable φ and*

$$\lim P_{(0)}^{(\varepsilon)} |x - c| = 1, \varphi = \varphi_1, \varphi_2, \dots \quad (21)$$

where c is some constant and ε some positive arbitrarily small magnitude, then

$$\lim_B (x) = c, \varphi = \varphi_1, \varphi_2, \dots \quad (22)$$

6. The notion of stochastic limit, whose rigorous definition is given just above, can still be extended in an extremely important direction. Instead of deviations of a random variable x from some constant c we will consider its deviations from some variable v being a function of the independent variable φ stochastically connected with x . After all these explanations, I hope that the following short definition will be absolutely clear:

If a random variable x is stochastically connected with an independent variable φ in such a way that

$$\lim P_{(0)}^{(\varepsilon)} |x - v| = 1, \varphi = \varphi_1, \varphi_2, \dots \quad (23)$$

where ε is a positive arbitrarily small magnitude and $v = f(\varphi)$ is some function of φ , then we call v a *stochastic (or Bernoullian) asymptote* of the random variable x :

$$\text{as}_B (x) = v \quad (24)$$

where $as_B(x)$ means asymptota Bernoulliana.

It is the analogy with non-stochastic concepts that compels us to introduce the term *asymptote* since *limit* in its usual sense is always only understood as a constant to which the relevant variable infinitely approaches. When two functions are approaching each other in that same way, so that the limit of their difference is zero, asymptotes are introduced, and this term is therefore suggested all by itself for describing the stochastic relations under consideration. And it is not difficult to prove that, if v_1, v_2, \dots, v_k are asymptotes of each other in the usual sense, and if one of these magnitudes is a stochastic asymptote of a random variable x , all the others will also be its stochastic asymptotes and I call them equivalent. A random variable having a stochastic asymptote will certainly have an incalculable number of them, but it cannot have two such asymptotes not equivalent to each other, i. e., not being usual asymptotes for one another. Therefore, if some stochastic asymptote has a limit, all the other ones have the same limit which will be the stochastic limit of the same random variable.

On the contrary, if a random variable has a stochastic limit, each of the latter's asymptotes, i. e., any variable having it as its usual limit, will be a stochastic asymptote. And, recalling the example in § 4, we see that the empirical coefficient of correlation ρ , given an unbounded increase in the number of empirical indications n on whose basis it was calculated, will have the theoretical coefficient r as its stochastic limit. At the same time, we know that the expectation $E\rho$ is a magnitude which varies as the number of trials n increases and infinitely approaches the theoretical coefficient of correlation:

$$\lim_B(\rho) = r, \lim E\rho = r, as_B(\rho) = E\rho \quad (25; 26)$$

which means that the expectation of the empirical coefficient of correlation is its stochastic asymptote.

Least interesting is the case in which a stochastic asymptote has a limit. The demand of introducing the concept of stochastic asymptote is necessarily mainly based on the circumstance that we can thus consider cases which do not yield to be studied by the idea of stochastic limit. Those are the instances in which a random variable stochastically approaches some magnitude which does not tend to any (finite) limit but either oscillates within some boundaries or infinitely increases (decreases).

We will at first offer a purely formal illustration. Let the distribution of probabilities be

$$\begin{array}{l} x: \frac{\varphi}{\varphi+2} \sin \varphi \quad \frac{\varphi}{\varphi+1} \sin \varphi \quad \sin \varphi \quad \frac{\varphi+1}{\varphi} \sin \varphi \quad \frac{\varphi+2}{\varphi} \sin \varphi, \dots \\ p: \quad 1/2^4 \quad \quad 1/2^3 \quad 1/2 \quad 1/2^3 \quad 1/2^4, \dots \end{array} \quad (27)$$

where φ is an independent variable. The term $\sin \varphi$ in the upper row has probability differing from 1 by $1/2$; together with both its adjacent terms they have probability differing from 1 by $1/4$, etc. It is thus possible to isolate a group of terms containing that middlemost term and an equal

number of them in each direction from it so that their probability however little deviates from 1.

Then, the difference between $\sin \varphi$ and each of the extreme terms of that group infinitely decreases with an infinite increase in φ ; therefore, it will be always possible to find such a large value of φ that, for a however small ε , the probability of the deviations lying within $[\sin \varphi - \varepsilon; \sin \varphi + \varepsilon]$ differed from 1 less than by an arbitrarily small magnitude α . Indeed, it is always possible to isolate such a group of terms with $\sin \varphi$ in its middle having probability higher than $(1 - \alpha)$. Then, increasing φ , we can always bring the extreme terms of the group however near one to another.

And so, it is proved that as φ infinitely increases, the probability that the value of the random variable will be however near to $\sin \varphi$ will approach 1. Nevertheless, that magnitude, $\sin \varphi$, is not constant but variable, continuously and periodically oscillating between -1 and 1 ; there can certainly be no mention of a stochastic limit, but it will be a *stochastic asymptote* of x ; and, for sufficiently large values of φ , x will *almost certainly deviate from it however little*.

7. This example only formally illustrates our idea, but we will see that the concept of statistical asymptote is very important in problems of another kind closely connected with issues of statistical theory and practice.

Suppose that a series of trials stochastically independent one from another is made. Let the probability of the occurrence of some event be p_1, p_2, \dots with a possible infinite extension of that series. The event occurred m times in n trials. It can be easily proved, see for example Markov (1900/1924, pp. 98 – 100), that, when the number of trials indefinitely increases, the probability of arbitrarily small in absolute value deviations of frequency m/n from the mean probability \bar{p} however closely approaches 1:

$$\lim P_{(0)}^{(\varepsilon)} \left| \frac{m}{n} - \bar{p} \right| = 1, \quad n \rightarrow \infty.$$

If \bar{x} is constant (for example, if in each group of n trials the given probabilities will be repeated equally often), \bar{p} will be the stochastic limit of the frequency:

$$\bar{p} = \lim_B \left(\frac{m}{n} \right). \quad (29)$$

And if \bar{x} is not constant but has a limit, it will be the stochastic limit of the frequency:

$$\lim \bar{p} = \lim_B \left(\frac{m}{n} \right). \quad (30)$$

However, if \bar{x} does not have a limit at all, but only oscillates within some bounds, which is the most ordinary case, if only the opposite is not assured by a special arrangement or some regularity dominating the given field of phenomena, then the formulations above are not applicable. However, in any case, owing to (28), the stochastic limit of the difference between the frequency and the mean probability vanishes

$$\lim_B \left(\frac{m}{n} - \bar{p} \right) = 0, \quad (31)$$

and the mean probability is a stochastic asymptote of the frequency

$$\bar{p} = as_B \left(\frac{m}{n} \right). \quad (32)$$

Similar conclusions are valid for the arithmetic mean of the values of a random variable. Let $x_1, x_2, \dots, x_n, \dots$ be the independent random variables in the respective trials, each with its own distribution of probabilities. If x_i can take values

$$x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(k)} \text{ with probabilities } p_i^{(1)}, p_i^{(2)}, \dots, p_i^{(k)},$$

then

$$Ex_i = \sum_{j=1}^k p_i^{(j)} x_i^{(j)}. \quad (33)$$

If only the mean square values of our random variables, i. e. the magnitudes

$$\sigma_i = \sqrt{E(x_i - Ex_i)^2}, \quad (34)$$

do not increase infinitely, then (Markov 1900/1924, pp. 116 – 118) the probability, that the difference between the arithmetic mean \bar{x} of the empirical values and that of their expectations will be however small, approaches 1 arbitrarily nearly.

Denoting the empirical values in consecutive trials by x'_1, x'_2, \dots we will therefore have

$$\lim P_{(0)}^{(e)} \left| \bar{x} - (1/n) \sum_{i=1}^n Ex_i \right| = 1, n \rightarrow \infty. \quad (35)$$

If the mean expectation is constant or tends to some limit, then either it or its limit will be the stochastic limit of the arithmetic mean. However, in the general case we can only say that \bar{x} has a stochastic asymptote, so that

$$as_B \bar{x} = (1/n) \sum_{i=1}^n Ex_i, n \rightarrow \infty. \quad (36)$$

The stochastic limit can be certainly included in the notion of stochastic asymptote as its particular limiting case so that (36) will be the general expression following from (35), or, finally, from the statement that, *given a sufficient number of independent trials and a restricted scatter of the*

appropriate random variables, the arithmetic mean will almost certainly however little differ from the arithmetic mean of their expectations.

We will only properly understand this important theorem after indicating that it does not depend on the constancy of the mean of expectations or the existence of its limit, that is, after perceiving that that limit is, generally, a stochastic asymptote. This result becomes all the more important when recalling that it can also be extended to cover a number of cases of stochastically connected random variables as shown by Markov⁹. Especially important is that kind of connections when each term of a series is only connected with a restricted, even if a very large number of adjacent terms, or if that connection, although spreading infinitely, weakens with the distance between the terms even according to a however slow geometric progression [with common ratio near but less than unity]. All series of events consecutively arranged in time unquestionably belong here.

By these investigations Markov had extremely widened the importance of the arithmetic mean for studying reality, but it is necessary to stress once more that such an almost universal covering of the appropriate theorems would have been practically reduced to insignificance had they been only applicable under “constancy of the relevant objective conditions”, – an often mentioned expression that should be understood as the constancy of the appropriate expectations.

Indeed, Markov (1900/1924, p. 117) did not assume such constancy. For example, one of his theorems ends in the following way:

We thus easily ascertain that under the condition stated above and for sufficiently large values of S the probability of the inequalities

$$-\varepsilon < \frac{X+Y+\dots+W}{S} - \frac{a+b+\dots+l}{S} < \varepsilon$$

will be higher than $(1 - \eta)$ however small are the given positive numbers ε and η .

Here, X, Y, \dots, W are random variables whose number S can increase infinitely and a, b, \dots, l are their expectations. The arithmetic mean of these expectations is not subjected to any restrictions, neither is it necessary for it to be constant or to have a limit. Therefore, in our notation Markov’s conclusion is

$$as_b \left[\frac{X+Y+\dots+W}{S} \right] = \frac{a+b+\dots+l}{S}, \quad S \rightarrow \infty. \quad (37)$$

The notion of stochastic limit was indirectly contained in the Bernoulli theorem and the further development represents a number of propositions implicitly containing the concept of statistical asymptote.

8. Markov, however, was not the first who applied it. Already Poisson proved propositions mentioned in the beginning of § 7 about stochastic asymptotes of frequencies and means in case of independent trials. We ought to dwell on that circumstance since otherwise it will be impossible to estimate properly the central idea of his entire concept. This is all the more necessary because exactly on this point Prof. V. I. Bortkevich, the most

authoritative interpreter of his ideas, explained Poisson's opinion in a way that I am absolutely unable to consider correct.

Leaving aside the arithmetic mean since everything stated below can be applied with appropriate changes to it as well, I will restrict the discussion to the case of frequency, and, correspondingly, to the mean probability, and allow myself to recall the sense of two terms coined by Bortkevich. Suppose we have an infinite series of independent trials with the probability of the occurrence of some event only taking values

$$p_1, p_2, \dots, p_k. \quad (38)$$

Suppose also that these values in turn have constant probabilities of taking place equal to $\pi_1, \pi_2, \dots, \pi_k$, independent both from the probabilities concerning previous trials and from the results of those trials. Then the probability of the occurrence of the event in each separate trial p_0 will be constant:

$$p_0 = \pi_1 p_1 + \pi_2 p_2 + \dots + \pi_k p_k, \quad (39)$$

which is a weighted mean of p_i .

Bortkevich called p_0 *mean probability in its proper sense*. We will have to deal with it if, for example, each time before an extraction from an urn it is selected by extracting tickets from a supplementary urn with an invariable set of those tickets.

Consider now another case. Suppose that an infinite series of trials consists of subseries with s trials each, that in each such subseries trials with probabilities (38) occur exactly s_1, s_2, \dots, s_k times and denote

$$g_1 = s_1/s, g_2 = s_2/s, \dots, g_k = s_k/s.$$

For each subseries and each series of trials consisting of a whole number of them we will have the mean probability

$$\bar{p} = g_1 p_1 + g_2 p_2 + \dots + g_k p_k. \quad (40)$$

Bortkevich calls it the *mean probability of a constant composition*. In the particular case in which all the numbers g_i are unity and $k = s$, it will be the arithmetic mean of p_i . Here, however, all the probabilities (38) must be repeated in each subsequent subseries of s trials even if in another order (Bortkevich 1894 – 1896, 1894, pp. 649 – 651)¹⁰. He (1917, pp. 49 and 54) believes that Poisson's LLN represents some proposition about the mean probability in the proper sense based on a lemma considering *mean probabilities of a constant composition*. This latter statement seems wrong to me. The mentioned lemma does not deal with such a mean probability, it rather concerns other probabilities for which I would not wish to provide here a special term, but which in this context could have been likely called *mean probability of an arbitrary composition*¹¹.

I have in mind the case discussed in the beginning of § 7; there, a series of independent trials was introduced with each of them having quite arbitrary probabilities of the occurrence of some event, $p_1, p_2, \dots, p_n, \dots$, in that order.

The arithmetic mean of the first n of them was indeed that which Bortkevich applied in the proposition which he called the lemma preceding the LLN.

Poisson (1837, § 52) mentioned that same proposition, then proved it in §§ 94 and 96, pp. 138 – 139 and 246 – 254. In neither place did he say that the same probabilities ought to be repeated in the next n trials, nor did he state that \bar{p} should be constant which is a most essential point. He formulated his result on p. 254:

Therefore, it is almost absolutely certain that, given a very large number n of trials, the [frequencies – E. S.] will very little deviate from the mean chances (probabilities) [...] which they will approach ever more nearer the more is n increased still further, and with which they would have coincided had it be possible for n to be infinite¹².

Therefore, we are not justified to introduce restricting conditions into the general theorem proved by Poisson. In our terminology, it can only be formulated thus:

The frequency of the studied event in a series of independent trials whose number n can be infinitely increased, always has the mean probability (of the relevant probabilities) as a stochastic asymptote:

$$as_B\left(\frac{m}{n}\right) = \frac{1}{n} \sum_{i=1}^n p_i, \quad n \rightarrow \infty. \quad (41)$$

Poisson also proves a similar proposition about arithmetic means. For the sake of brevity we will call both these theorems taken together Poisson's Theorem I; his Theorem II will be that concerning "mean probabilities in the proper sense". Their relative importance, as well as that of Theorem I in the absolute sense ought to be radically reappraised.

Had the latter been, as Bortkevich thinks, a proposition about "the mean probability of a constant composition", then, indeed, we would have been compelled to state that its role in interpreting reality was comparatively modest. As described above, however, it is a proposition of great generality only recently being superseded by Markov's theorems. Its applicability is in any case much wider than that of Theorem II¹³.

II

9. Leaving aside the history of the interpretation of the LLN, a possible subject for a special investigation, it is still necessary to mention that no single concept which could be considered generally acceptable has even yet crystallized out of the discord generated by that term¹⁴. As it always occurs, in each meaningful debate about some term the essence certainly concerns not the word, but the nature of problems connected with it and not sufficiently made out.

In this case, the use of that expression remains essentially shaky and diffuse although some "centres" have been outlined long ago. I begin with one such tendency, namely with that, observable in writings of a number of authors having a definite mathematical inclination. I think that after specifying and modifying it, we may reach an acceptable concept.

Markov (1900/1913, p. 70n; 1924, p. 98) says that, according to his opinion, “we should understand the law of large numbers as the totality of all the generalizations of the Bernoulli theorem”, but this formulation certainly cannot be called successful. Already Bortkiewicz (1917, pp. 53 – 54, note 13) quite rightfully remarked that it was inconvenient to call a *totality* of theorems a law; that such a totality is usually called a theory.

Although Markov undoubtedly did not precisely interpret his idea as represented in a number of places in connection with describing various theorems¹⁵, it nevertheless contains a notion whose main kernel deserves much attention.

Prof. V. I. Romanovsky (1924, NNo. 4 – 6, p. 15n) apparently discovered an expression much more adequate to the essence of the matter:

Perhaps, however, it would be best of all in accordance with a long ago established custom, now almost impossible to change, to leave the term law of large numbers as a general name for many theorems of the calculus of probability in which a large number of some conditions or trials is essential.

It seems that our results stated in the previous Chapter allow to restrict in a definite way those propositions here implied. Any generalization of the Bernoulli theorem in which the number of trials is an independent variable has the form

$$\text{as}_B(x) = f(n), n \rightarrow \infty \quad (42)$$

where x is some definite random variable and $f(n)$ some one-valued function of the number of trials. The notion of function here means a definite relation between definite values of n and of $f(n)$ ¹⁶.

Each time when x and $f(n)$ are definite magnitudes (one of them is random, but not the other one)¹⁷ expression (42) will represent by its logical form some *proposition* which must therefore be either true or false. However, when understanding x and $f(n)$ as just *any* magnitudes, some random variable and some function of the number of trials, that expression ceases to be a logical proposition, a correlate of possible opinions, and only becomes a pure *form* into which many propositions can be included, namely all those to whom, given that form, correspond definite logical values of x and $f(n)$.

The difference is expressed in that (42) cannot now be either true or false, it is not a proposition (*propositio*) but only its form, a *propositional function*. From our point of view, any general stochastic proposition reducible to that form will be a LLN¹⁸. And so, we suggest to *consider as that law any (of course true) general stochastic proposition that a random variable having such-and-such property (or, more general, properties) must necessarily have a stochastic asymptote* (respectively, *a stochastic limit*).

And the random variable can be not only one-dimensional, with which case we only have dealt previously, but any stochastic formation that can be interpreted as that notion by its reasonable extension: a randomly variable vector in an n -dimensional continuum (Mises 1919a; 1919b), a randomly variable geometric figure, etc.

The suggested use of the word *law* is not anything new. Such expressions as law of nature, causal law, etc deal not with some definite law but with a

definite kind of laws, and the expression *law of large numbers* can also be applied as a generic term. If a random variable does not obey *any law* of large numbers, we may briefly express it as not obeying that *law*¹⁹.

10. The idea underlying our concept of the LLN would have been altogether corrupted had it been understood as only purely mathematical. Pure mathematics knows nothing, and cannot know anything about events, trials, probabilities. These things are as alien to it as electrons, light waves, celestial bodies, space and time. Mathematics is entirely oriented at the formal, in the strict sense, for instance *thing (something), ratio, set, number, the whole, a part*, etc. The doctrine of sets [set theory], the theory of groups, arithmetic, – these are the typical chapters of pure mathematics. From here also issues that surprising universality of mathematics, its applicability to any material contents absolutely independently from it if only the relevant forms of categories are in existence (Husserl 1900, pp. 247 – 254/1913, pp. 20 – 23). It seems to me that after all the work in justifying mathematics done during the latest decades at least its formal essence should be unquestionable.

The fate of geometry is especially instructive. Pure geometry, as first constructed by Hilbert (1899/1903), is already certainly not a doctrine of space forms. The meaning of *space* which we attach to it is there intentionally eliminated. Points, straight lines and planes are dealt with, but no material contents is connected with them, they are so-called points, straight lines and planes of that *geometry*, simply some “things” in the wide sense of logical “something”. No restricting premises about their essence are introduced.

Just the same, when further the relations such as “a point is situated on a straight line”; “a straight line passes through a point” are discussed, it is not at all demanded to attach there our usual (or any whatsoever) vivid ideas. A point could have been replaced by “some thing A” and the straight line by “some thing B” and the relations above formulated as some R_1 and some R_2 . The axiom that “if a point is situated on a straight line, then the line passes through that point” can describe in that system an absolutely formal assumption: “given a relation AR_1B , the relation BR_2A is also given”.

This, or any such proposition about unknown relations between unknown things cannot evidently be justified as being true. Instead, if admitting absolutely conjecturally that a number of such “axioms” is valid, an immeasurable set of theorems can be purely deductively derived from them. Given one system of premises, we obtain a system of propositions corresponding to the Euclidean geometry; issuing from another system, we arrive at the Riemann or Lobachevsky geometry, etc.

Correspondence should be understood in the sense that indefinite symbols of one or another system are transferred to the language of a known field of things by means of some dictionary (A = point, B = straight line, R_1 = lies on, etc); in this case, to the language of space forms²⁰. And it ought to be noted that in each case such a deductive hypothetical system in principle admits an infinite set of translations. If, concerning some system of things of a definite kind, we convince ourselves that those relations are formally admissible in some system of axioms, that system becomes a system of the *main laws* of the appropriate field and the entire construction of the deductive theorems when being translated into the appropriate language, immediately acquires a worth equal to *deductively established laws*.

Revising the axiomatic basis of geometry and separating a purely mathematical discipline from it creates therefore the necessary premises for studying space as such, and it would have been more proper to call that discipline not geometry, but a doctrine of manifolds (Mannigfaltigkeitslehre). In our time, the solution of that problem understood as a problem of physics is being looked for. We are unable to discuss it here, but it is evident that the problem *thus* formulated does not already belong to pure mathematics.

11. The situation with the “calculus of probability” is absolutely similar. A radical formalization of that discipline is necessary as a preliminary and indispensable condition for perceiving in essence the issue of its applicability to real life, to understand it up to its initial basis. After becoming a purely mathematical discipline, the calculus of probability will have to lose also its historical appearance. Indeed, the numbers that it has to consider as being in a one-to-one conjunction with terms of a disjunctive relation (or A, or B, ..., or S) are deprived of the meaning of *probabilities* and can only acquire it in some applications; moreover, in other applications that meaning can be different.

Conjecturally call those probabilities *valencies*, then the goal of the corresponding [of the emerged] calculus will be the revealing of some valencies by issuing from other valencies and making use of the appropriate axioms. In accordance with the main relation characterizing the connection between the considered elements, a provisional name of *disjunctive calculus* may be suggested for such a purely mathematical discipline²¹.

The doctrine of probabilities of random events can be named *stochastics*, see Note 1. For that materially determined field the axioms of the disjunction calculus as well as propositions deduced from them will become its laws. Whether they are prior or based on experience, is their significance unconditional, as it obviously follows from the nature of their contents, or they themselves just as all empirical laws are only more or less probable, – let all that be still debated. In any case, there certainly exists a logical abyss between those propositions and the formal statements of the disjunctive calculus.

As we understand it, the LLN is obviously not a purely mathematical theorem²²; mathematics itself cannot justify its applicability to one or another field. If that law is valid, it is only because and insofar as the known purely mathematical theorems are applicable to calculation of the *probabilities of random events*. The essence of those latter should include a basis enforcing the appropriate propositions and transforming them into laws of the corresponding field of reality²³.

It would be easy to formulate the Bernoulli theorem (say) in such a form that nothing is said about either events or probabilities. For attaining that aim, it is sufficient to consider infinite series of pairs of numbers

(1, 0); (1, 0); ...; (1, 0) conjugated
with *valencies* (p, q) , (p, q) , ..., (p, q)

($q = 1 - p$), with any sequence 1, 1, 0, 1, 0, 0, ..., 1 where 1 occurs m times and 0, $(n - m)$ times, conjugated with valency $p^m q^{n-m}$. Then, after considering all possible sequences of 1 and 0 having n terms, we can ask about the sum of valencies of those satisfying the condition

$$\left| \frac{m}{n} - p \right| \leq \varepsilon.$$

We will infer that for any however small ε and an infinitely increasing n that sum will have unity as its limit. This will indeed be the *Bernoulli theorem* in its purely mathematical kernel, i. e., a theorem of the disjunctive calculus. It is not however that same theorem in its historical appearance in the sense attached to it by its author; it is its shadow, its pattern devoid of all its stochastic contents.

In its real significance, the Bernoulli theorem assumes that natural events can have probabilities, that it does make sense to apply the notion of probability to nature (and therefore to opinions about it) although Bernoulli himself did not yet perceive what was, properly speaking, the relation between *probability* and *occurrence* of the corresponding event, – and neither is it understandable even to us in the 20th century, as indicated by radically diverging opinions.

In any case, the Bernoulli theorem assumes the existence of such a relation. But then, what it states is also a proposition about the connection of probability and frequency, and namely about what probability corresponds to known frequencies when the number of trials infinitely increases. At the same time, it is a proposition about the frequencies themselves. Indeed, had it been possible to establish that under such-and-such conditions frequencies ought to obey inequalities

$$p - \varepsilon \leq \frac{m}{n} \leq p + \varepsilon, \quad (43)$$

that restriction would be about frequencies, it would be a law concerning them. And the *Bernoulli theorem* states the same, only not in such an absolutely categorical form. Instead of saying *frequencies should not go out of any however tight boundaries*, it states that *they will almost certainly* not go out of them. Or, in another form,

Given an infinitely increasing number of trials, the deviations of frequency from probability will almost certainly be included in infinitely tightening boundaries.

It cannot be gainsaid that those were statements about frequencies, only not absolutely certain although approaching such a certainty when the number of trials is infinitely increasing. When saying that all such propositions concern not frequencies, but only their probabilities, exactly that intimate connection between probability and the object whose probability it is, is overlooked; missed, therefore, is the nature of the entire empirical knowledge. Husserl (1900/1913, Bd. 1, pp. 13 – 14; translated [by Slutsky] from its Russian translation of 1909, pp. 9 – 10) says that

The most perfect indication of truth is obviousness; for us, it is as though a direct mastering of truth itself. In the greatest majority of cases we are deprived of that absolute understanding of truth; its ersatz is [...] the obviousness of that higher or lower probability of the corresponding situation with which, providing an “essential” degree of probability is gained, a definite and decisive opinion is commonly connected.

[I delete further quotations. Slutsky concludes:] The situation described by Husserl is valid for any empirical knowledge.

12. I consider the defended [the substantiated] concept of the LLN as purely stochastic and oppose it to all the attempts of representing that regularity as a law of frequencies. The most direct method for achieving it [?] would have been to reduce probability itself to frequencies; however, the abyss between these notions cannot be bridged, and there is no essential difference between the more artless constructions of the empiricists of the English school (Ellis, Venn) and modern *asymptotism*. On the contrary, a clearer mathematical formulation allows to perceive more distinctly the radical unfeasibility of those intentions.

The issue is reduced to the question of whether it is admissible to determine probability as the limit of frequency as the number of trials infinitely increases, see for example Mises (1919b, p. 55). Its solution ought to be searched on the path recently well formulated by Cantelli (1917a, p. 40): *is such a definition compatible with the calculus of probability?* He himself did not apply his indication.

Suppose we have an infinite series of independent trials and p is the probability of the occurrence of some event in each separate trial, not depending, consequently, on the results of previous trials. As above, let the number of occurrences of the event in n trials be m and its frequency therefore m/n . The conditions of the Bernoulli theorem are satisfied, so that

$$\lim_B \frac{m}{n} = p, n \rightarrow \infty. \quad (44)$$

Is

$$\lim \frac{m}{n} = p, n \rightarrow \infty \quad (45)$$

compatible with (44)? What is a limit in mathematics? No indefiniteness is present here, the notion of limit is strict, rigorous and established. If equality (45) is really true, then, having chosen any arbitrarily small positive number ε , we always ought to find such a large finite value of n , n_0 say, that for any $n > n_0$ the inequality

$$\left| \frac{m}{n} - p \right| \leq \varepsilon \quad (46)$$

will be satisfied.

The matter, however, is that exactly this inequality is not compatible with the conditions of the problem. Indeed, if it is valid for any $n > n_0$, its probability will be

$$P\left\{ \left| \frac{m}{n} - p \right| \leq \varepsilon \right\} = 1 \quad (47)$$

and the probability of the opposite

$$P\left\{ \left| \frac{m}{n} - p \right| > \varepsilon \right\} = 0. \quad (48)$$

However, under conditions of the Bernoulli theorem either (47) or (48) do not take place given any finite number n of trials. Indeed, if the trials are independent, the occurrence of the event after any previous sequence of outcomes remains equally possible with the same probability p and its failure with the same probability q . Therefore, it is not allowed to consider impossible even its occurrence, as well as its failure all n times (with frequencies 1 and 0 respectively) however large is n . Calculus of probability provides the respective probabilities, p^n and q^n . It is wrong to state that the occurrence of the event more than n_0 times (say) in succession is impossible, that would be tantamount to saying that the probability of its occurrence in the $(n_0 + 1)$ -th trial was zero, i. e., that the trials are not independent. Absolutely the same concerns the impossibility of lesser deviations [?].

To state that probability is the limit of frequency means to assert that, given an arbitrarily small number ε , it is possible to find such a large number of trials, that, for any larger number of them, some series of occurrences (or failures) of the event, namely those for which the inequalities

$$p - \varepsilon \leq \frac{m}{n} \leq p + \varepsilon \quad (49)$$

do not hold, become impossible.

Thus, for independent trials the equality

$$\lim \frac{m}{n} = p, \quad n \rightarrow \infty \quad (50)$$

is a nonsense (in the strict sense of a formal logical contradiction). For defining probability as the limit of frequency it is therefore necessary to justify as some universal and prior law the impossibility of independent trials, but this, however, is nonsense²⁴.

13. Chuprov's construction based on some of Cournot's ideas is a logically incomparably more careful attempt in which, however, that same empirical tendency as previously cannot be missed. The subject of proof considered as the contents of the LLN is that, given a large number of trials, "the frequencies of events are keeping close to their probabilities" (Chuprov 1909/1959, p. 167; a similar statement is on p. 164). The proof is based on two lemmas the second of which representing a *mathematical theorem* of the Bernoulli or Poisson type, whereas the first one, constituting the peculiarity of that concept, is this: "Events whose probabilities are very low, will not occur often" (1909/1959, p. 166); or, elsewhere (1909, p. 229)²⁵ "Low probability conditions rarity". Then (1909/1959, p. 167),

Out of these lemmas we indeed come to the law of large numbers. At first we show that a low probability conditions rarity. Then we prove that, given a large number of trials, large deviations of frequencies of events from their

probabilities are unlikely. It follows that such deviations are rare and, as a rule, frequencies are keeping near their probabilities.

We have to discuss the first lemma since it plays here the main role. Those theorems, in which we see the true essence of the LLN, do not satisfy Chuprov (1900/1959, p. 168):

They state that, when there is a large number of trials, it is unlikely for frequencies of events and their probabilities to differ much. But a new premise is needed, – the first lemma, establishing the connection between low probability and rarity, – for moving from the world of probabilities, high or low, to the field of frequencies. It is that premise which serves as the base for the move from mathematics to statistics and constitutes the subject of the law of large numbers.

Let us return to propositions discussed above. In our understanding, the LLN (Chuprov's second lemma) states that, under such-and-such premises and having a large number of trials, the deviations of frequency from probability beyond any however tight boundaries become arbitrarily unlikely. In our opinion, the connection of probabilities with frequencies consists here in that

1) Almost certainly, and the nearer to absolute certainty the larger is the number of trials, that in any given trial such deviations will not occur. That *almost certainly* is based on actual data of the appropriate branch of knowledge in some peculiar way. We may say that any law of nature based on some empirical data is almost certainly valid. Here, however, it is possible that this statement does not represent the truth, it has a different essence. If we happen to be wrong, and the statement which we (*almost certainly*) considered as a law, is false, than it, as an ideal timeless entity, is refuted once and for all rather than rarely admitting exceptions.

It is ever clearer when considering instead of a law of nature a mathematical theorem whose complicated and difficult proof was attained and checked by a number of knowledgeable persons. In such a case, it is possible to assume that all of them would have hardly missed a mistake, and the proof will *almost certainly* be true. If, however, our assumption will not be justified, that theorem will be called false and deleted from the edifice of knowledge.

Concerning the LLN, that *almost certainly* has another meaning. It takes into account not that a given general proposition can as such occur to be false, it rather provides for possible *exceptions* in some separate case. In addition, the LLN, when applied to such very unlikely deviations, furnishes for us the following *almost certain* knowledge:

2) Those deviations about which we know *almost certainly* that they will not occur in any given trial, will *almost certainly* occur in some very long series of trials, usually with a very low frequency (whose measure will roughly correspond to that probability which distinguishes the former *almost certainly* from certainty). And it is this proposition included in the LLN that represents the logically rightful kernel of Cournot's first lemma. That proposition is indeed stating that

Very unlikely events will “almost certainly” occur very rarely, whereas the lemma throws out that specification and simply tells us that such events will not occur often.

This, however, is exactly something we are not at all empowered to assume, it is indeed an unlawful exit from the stochastic ground. It is unlawful since being a refusal of an absolutely certain statement that at best we only have an “almost” rather than absolutely certain knowledge of the rarity of unlikely events.

The Laplacean definition [justification] of probability by the principle of “insufficient reason” was repeatedly criticized; it was indicated that by means of the calculus of probability knowledge could have been then acquired out of ignorance. I think that no one will yet defend that viewpoint at face value; everyone agrees that the conclusions arrived at by that calculus assume some initial knowledge. However, it should not be missed that, when dealing with stochastically posed problems, we never acquire absolutely complete knowledge. It follows absolutely inevitably, as it seems to me, that absolutely certain knowledge cannot be elicited out of stochastically given premises since that would again be recoinning ignorance into knowledge.

We are only able to make practically harmless the indefiniteness rooted in the incompleteness of our knowledge. This aim is achieved by discovering the circumstances left *almost* intact by the mentioned indefiniteness and about which we can therefore know something, *almost certainly* augmenting our knowledge as that becomes necessary and possible by (again) *almost certain* knowledge of frequencies of exceptions, of exceptions from exceptions etc.

However, until our problem remains stochastic, we will never be able to throw away that *almost*. This radical demand made by the very essence of the stochastic is the true *main principle* of the *logic of stochastics*. I do not dwell on its exact definition since that will divert us too far from our main subject; see Höfler (1922, pp. 668 – 669 and 739).

14. In addition to these essential considerations which seem to me decisive, we can approach Cournot’s *first lemma* somewhat differently. First of all, it is permissible to ask about its logical essence: is it nomological or only ontological? The interpretation of the relevant passages is somewhat difficult and leaves room for both assumptions. Let us consider them.

When supposing that that lemma is nomological, we inevitably contradict the theory of probability, i. e., the main principles of stochastics and deductions made by it. Can we say that *events whose probability are very low, will not occur often* in the sense that they *cannot* occur often? The answer seems doubtless and negative if only we properly understand that proposition. It is yet possible to admit that *all* unlikely events cannot occur often, but that is self-evident.

If a billion numbered balls are in an urn, then, during a billion trials, each can be extracted *rarely*, for example only once, or, some of them will be extracted a million times each thus depriving millions of their companions of being extracted even once. The Cournot lemma certainly assumes not such cases, but supposes that *no* unlikely event will occur often. And so, to repeat, when interpreting that denial nomologically, as *cannot*, we arrive at a contradiction. Any event, having an arbitrarily low probability of occurrence, *can* occur however often in any series of trials. Thus, in a

roulette game, the probability of the outcome of *red* 10^{10} times in succession is not at all zero (i. e., it is not an impossibility, but a very small positive magnitude). In any sufficiently long (although practically unfeasible) number of series of 10^{10} trials each, we can expect with probability however near to certainty that some (although exceptionally small) part of them will only consist of *red*.

Even the most lowest probability essentially differs from impossibility and we are unable to bridge that gap however increasing our numbers [of trials]²⁶.

The remark just made in the Note should be restricted. Suppose that the probability of an event is $p = 1/2$. The probability of its occurring twice in succession is $1/4$, of three and four times, $1/8$ and $1/16$. All these numbers are of one and the same order if order is understood as a power of 10 ($1/2:1/16 = 8 < 10$). However, an occurrence of that event 10 times in succession has probability $(1/2)^{10} = 1/1024$, of a higher, although not excessively so, order. But then, if the probability of the event is itself very low, for example $1/1000$, its occurrence twice in succession already has probability $1/10^6$, of 10 times, $1/10^{20}$, of 100 times, $1/10^{300}$.

Even if such an event can occur each second, it will *almost absolutely certainly* fail during the entire history of mankind. Here, we therefore have yet another truth about [another property of] the *often* occurrence of unlikely events: they *can* occur however often but that will constitute an event whose probability is a small magnitude of a much higher order. Or, shorter: *an often occurrence of an unlikely event is a superunlikely event²⁷*. That *almost certainly* will happen unimaginably rarely.

As it seems to me, exactly from Cournot's point of view it cannot be upheld that such a statement only deals with probabilities but not at all with frequencies. (I am here abstracting myself from the considerations offered at the end of § 11 and in § 13.) Indeed, Cournot defended the view that each proposition about probabilities of frequencies deals with the latter:

Then mathematical probability becomes a measure of physical possibility [...] which directly indicates the existence of some relation [...] that takes place between the things themselves, a relation upheld by nature and revealed by observation when the trials are sufficiently repeated²⁸.

However, I cannot discuss here the complicated concept of objective possibility which Cournot touches in the passage just above (Kries 1886)²⁹.

15. Since the nomological interpretation of the Cournot lemma is so resolutely contradicting, as it seems to me, the principles of the theory of probability, the issue of its possible ontological interpretation becomes more pressing³⁰.

The statement that *events whose probabilities are very low will not occur often* can not only express some law necessarily taking place always and everywhere, but a *fact*, an *actual structure*, an actual constellation of elements of our world or of its part surrounding us, and, again, not always, but only during a given epoch observed by us.

It is impossible to dispute that such an approach is logically lawful; along with laws, initial facts are necessary for understanding and explaining

events. Already Mill (1843/1898, Book 3, Chapter 16, § 4, p. 310) stated that *collocations cannot be reduced to any law*. What is true concerning particular cases and their explanation, is true for the universe as a whole³¹. Thus, in addition to laws of the causal type, the essence of the space of our world is a fact, as *random* in principle as any other, only of an immeasurably greater importance.

The necessary connections between phenomena following from it will be laws of our world, although not causal but structural. So what does the discussed proposition mean according to such an approach? Suppose that among all generally possible arrangements of the elements of the universe there are such in which most unlikely (when considered from our usual viewpoint) events are taking place not because of a regular connection, but only owing to that *chance*, actually to a given constellation, as though contrived by a demonic force, occur such results as warming of some bodies by cooling of other bodies, as some parallelism between the aspects of heavenly bodies and the fate of men, etc.

According to our modern understanding, such a world is not impossible, but only unlikely. Unlikely? Too weak an expression! There are no means to explain how unlikely it is. Perhaps the following image will provide some idea. If, wishing to write down the number $(10^{10}$ to the power of $10)^{10}$ as a unity with a number of zeros, an inkwell as big as the observable universe will not hold enough ink for achieving that. And the insignificant probability of that extraordinary monstrous world seems to be immeasurably less than unity divided by the number mentioned. Even comparatively modest approximations to that extraordinariness are expressed by absolutely unimaginably small numbers.

It would be possible to attempt understanding Cournot's lemma as follows: the actual world is ordinary rather than most rare and imagined. This is an expression of the actual structure of the world as far as it is known to us from both daily and more subtle scientific experience. That, however, is not a law, but a fact just as the fact of our Earth having a satellite.

That our world is *ordinary* can be also presented otherwise. Even the ancients noted two principles in the world: as a whole, it is the *cosmos*, an orderly unity subjected to unshakeable laws; *chaos*, an element of disarrangement and lawlessness, is taken prisoner, restricted to certain boundaries, and dashing about in its entrails.

These ideas are also included in our science. Each cell of the cosmos, down to its last atom, is subjected to strict and stable laws leading to well-proportioned complicated edifices: solar systems, climatic belts with their flora and fauna of highly developed creatures. Basically, however, along with the laws, chaos is also governing: separate tiniest elements of the world form gatherings where the regular life of an individual is going on beside but unconnected to the similar life of another one. The incoherent irregularity of gas molecules appropriately follows from their previous, just as incoherent and random arrangement. That chaos is a prisoner of cosmos and serves its aims. Exactly owing to irregularity the laws, for example, the second law of thermodynamics, are realized³².

What would follow from the existence of one of those extraordinary arrangements of elements mentioned above? Chaos will remain chaos, its cells will be somehow arranged not because of some law but *actually so*. That *random* arrangement will create an illusion of order, compelled to

suppose an existence of a law, of definite causes. Submitting to laws and being ordered, chaos will become a cosmos. *We cannot distinguish an ordered chaos from cosmos.*

Here then is an important principle of knowledge. Indeed, whereas we cognize the laws of numbers, logical connections, some laws of space forms, etc, owing to their obviousness since they are more or less observable, the laws of nature are understandable without such inner comprehension. Repetitions of the same under the same conditions, – that is the main and irreplaceable guide of the practice of empirical knowledge. Had random rarest and most unlikely in principle “lawless” repetitions sometime occurred, we would be unable to distinguish that fact from the action of a regular cause.

Suppose that some deviation with probability 1/1000 is repeated twice, three, four or five times, not to mention a larger number. Our knowledge of nature and its laws, our penetration into the depth of facts is yet so insignificant, that we will believe and firmly keep on believing that there exists a lawful explanation of such an occurrence and will decidedly reject the hypothesis of having encountered a rarest randomness.

The hypothesis that our world is ordinary can therefore be expressed as an assumption of an *unorganized chaos*. *The principle of unknowable of the too unlikely* exists alongside as a principle of stochastic cognition. Not being the main or unprovable, it still deserves to be specially mentioned. It tells us that, even had unlikely events occurred often, that fact would have remained unknowable because we would have no reason to consider the corresponding events really unlikely³³.

And, finally, the last series of considerations. Being empirically justified, the hypothesis of the ordinariness of the world is only substantiated for the past. In principle, it does not allow any conclusions concerning the future. Indeed, since we deal with chaos, that is, with unconnected phenomena, the probability of any of their combinations beginning exactly at this moment, is the same for any moment.

No exceptional anomalies in proper dealings out of arbitrarily large numbers of games of cards was observed up to now, which does not at all secure against their occurrence³⁴. Let us provide an example. Knowing the yearly number of games taking place in the world and their distribution over the methods of dealing out the cards, it will be possible to calculate the probability that during the next year the cards will be always randomly distributed among the gamblers so that each has only one suit. This probability, almost immeasurably insignificant, nevertheless differs from zero and the suggested occurrence ought to be admitted as possible.

However, since each dealing is independent from the others, that probability will not depend on the outcomes of former games [dealings]. Therefore, *the totality of observations of previous unconnected events does not in principle have anything in common with their future*. At the same time, Cournot’s lemma is peculiar exactly in its orientation towards future events. Not without reason it is formulated accordingly: *events whose probabilities are very low, will not be often repeated*. Without that property it would have been lacking any value as a basis for the LLN.

It is thus impossible to equate the Cournot lemma and the hypothesis of the ordinariness of the world. The latter can only be justified as an empirical proposition reflecting the ontological structure of the past whereas the

former claims to be important for the future. As a nomographic proposition, that lemma contradicts the principles of the theory of probability, as an ontological proposition it lacks in principle empirical substantiation. Indeed, to repeat, it claims to throw light on the field of future stochastically independent events where the past is not a law for the future, where the future remains in principle obscure, where ontology is in principle inapplicable.

There is another standpoint allowing to state that the hypothesis of the world's ordinariness can also be formulated for the future. If we, being continually threatened with the destruction of the habitual cosmos by the rebelled forces of chaos, pay no attention to that danger as though it does not exist, it is only because chaos as such is radically blind and its unconnected forces can only by chance constitute an army dangerous for the cosmos. That case is however so incredibly unlikely that we sleep peacefully. Thus, that hypothesis cannot justify the LLN.

On the contrary, only the latter in its purely stochastic understanding essentially clarifies it when the past is concerned and justifies its applicability to the future. As far as the Cournot lemma is true, it is only true in the purely stochastic sense as a corollary or a particular case of the LLN, again understood purely stochastically. However, that nullifies it as a lemma and negates the deductions which it should have served.

16. If I am not greatly mistaken, the method applied above consisted in earnestly taking into account the terms of probability theory, that is, to consider the words *probability* and *certainty* not as though unchangeable paper money in whose face value only sometimes trust naïve people, but as reliable promissory notes in which every word is law.

For us, *probability*, *certainty*, *trial* and *event* were coins of standard weight and full value just as points and straight lines of the usual geometry rather than in Hilbert's sense of pure mathematics. The idea of disjunctive calculus all by itself therefore allows to perceive the difference between pure mathematics and stochastics.

It also seems that the above should not be understood as an attempt at creating new logical theories about *probability* or at substantiating the theory of probability (= *stochastics*). The approach was quite different and in a sense logically more primary. Before *justifying*, it is necessary to know what really should be justified, i. e., to *ascertain the real sense* of the relevant propositions whereas it seems to me, although proving it will be likely difficult, that many statements were made and debates carried out about the theory of probability without perceiving each time that sense sufficiently distinctly.

How could have otherwise emerged the opinion that the *Bernoulli theorem* and other similar propositions only represent purely mathematical theorems? Had the sense of the statement that, given a sufficiently large number of observations, we know *almost* absolutely certainly that the frequencies will behave in such-and-such a way, been allowed for most seriously, – how was it then possible to miss that something is stated here about frequencies in general and therefore about frequencies of real events in real time and space? This I am unable to understand.

That proposition can be considered false and be rejected, – such a viewpoint is understandable, but to consider that we only have here a statement about probabilities, but not frequencies, it is necessary, as it

seems, not to consider seriously the sense of the words *probability* and *certainty*. We rejected some viewpoints about probability and some attempts to justify the LLN, and we only did it by issuing from the sense of the notions of the theory of probability itself.

We conclude that it is impossible to justify probability by that approach, and neither are we able to substantiate the propositions based on it: here, we encounter an inner contradiction. He, who wishes to insist on the propositions criticized by us, ought to reject the theory of probability (but perhaps not the *calculus* of probability understood in the sense of our disjunctive calculus). And who wishes to justify the theory of probability by objectively and impartially studying its basis rather than accomplishing a task formulated beforehand, must consider its real sense and the true structure of its propositions.

17. Let us consider now one more circumstance connected with the LLN at least in the sphere of causality but without inquiring whether that sphere is the only one here relevant. As a concrete example, I imagine two cubic containers of equal volumes, each divided by an imagined wall into left and right halves. There is one molecule, whose velocities certainly change with the temperature of the walls, in each container. One of these is here, on the Earth, the other one on Mars or even further.

At definite moments separated by equal intervals of time, very long as compared with the velocity of the molecules, we observe whether the molecules are situated in the same or in differing halves of their containers. Call the first case *coincidence* and count the number of those in n observed trials. Independence of trials and a prior probability of coincidence equal to $1/2$ are apparently secured, so that, according to the LLN, after a sufficiently large number of trials the frequencies [observed after various numbers of trials] will be almost absolutely certainly situated within boundaries specified by known rules. Let us somewhat more attentively study all this.

The movement of each molecule is determined by the laws of nature (A) and some initial conditions (B) (by collocations, according to Mill). Since the latter are given, the fate of each molecule is determined uniquely and necessarily from the beginning of time to eternity, as is consequently the frequency of *coincidences*. Knowing *almost certainly* the boundaries within which the frequencies *will be* situated, we thus know *almost certainly* that they will *necessarily* be there. Having known A and B, we would have been able to do away with the restriction *almost certainly*, but the conditions *known* to us (independence of trials and the probability of the separate events equal to $1/2$) do not allow that because they do not quite determine the future. The unknown will, however, become the less important the more is the number of trials.

Considering an indefinite set of sufficiently long series of trials rather than one such series with conditions $(A_1; B_1), (A_2; B_2), (A_3; B_3)^{35}, \dots$ we will know *almost certainly* that *almost all initial conditions* only satisfying simple stochastic premises *necessarily* determine certain circumstances concerning frequencies. Abstracting ourselves from special conditions and restrictions introduced for obviousness of illustration, we may formulate our idea in the following way.

If the premises of the LLN are satisfied, then, given a sufficiently large number of trials, *almost certainly, almost in all cases* some behaviour of

frequencies (and means) must necessarily occur, accompanied by a corresponding very low frequency of exceptions, etc.

This aspect of the LLN therefore allows us to perceive its deep similarity with absolute certain (at least in principle) strict *laws* establishing *necessary* rather than only *actual* relations. The more is the number of trials the nearer are the prior indications of the LLN to absolute certainty and to absolute strictness of necessary connections. Such an approximation to certainty and strictness can become arbitrarily near so that the particular cases of the LLN are thus themselves arranged by the steps of their approximation and have limiting absolutely certain and absolutely strict laws.

This, as I believe, justifies the application of the very term *law* in the expression LLN since the idea of law, when properly and strictly used, suggests the notion of necessity, if not revealed unconditionally and obviously, then at least reasonably conjectured as in the case of inductive laws of nature³⁶. It seems that there are no appropriate causes to reject the understanding of the LLN in the sense just described and especially for transforming that term to the *fact* of approximate stability of statistical frequencies (and means) if those are based on sufficiently large numbers and constant or weakly varying general conditions as Bortkevich (1917, pp. 56 – 57) had recently suggested³⁷.

Apart from all possible logical objections to apply *law* for denoting *fact*, in our case, as it seems to me, neither are objections terminologically necessary since *facts* can be here conveniently integrated into statistical terminology as the *fact of statistical stability*³⁸. Some possible misunderstandings are easily excluded just as all of them are by converting a usual polysemantic word or phrase into a scientific term with a fixed meaning. And a general fact is also easily specified: stability of mean frequencies, of *large* or *small numbers*³⁹, etc.

When eliminating the word *fact* from the Bortkiewicz definition, and determining the LLN as an empirical and inductively established “law of stability”, we encounter another, and as it seems to me, insurmountable difficulty. Being ignorant of the theory of probability, we would have certainly established and suitably formulated such a law. Experiments with tossing coins and dice, extracting lots etc, and, most importantly, the behaviour of large gatherings of molecules would have provided a sufficient inductive base.

The transition from the *fact* of the observed *approximate regularity* (the only kind of regularity observed in any field of the empirical) to a *strict law* with its underlying idea of necessity would have been a natural step incessantly taken by science in all the branches of studying nature. In essence, little would have changed even if we formulate a law as an *approximate* or *limiting* as the number of repetitions increases rather than as a strict regularity.

Unfortunately, it would not be true in any of the three cases, and the theory of probability would have compelled to abandon that law and replace it by the LLN in the stochastic sense. Indeed, as we saw above, *any proposition about the behaviour of frequencies of independent events can only be formulated when essentially connected with some statement about the corresponding probability, that is, at best, almost certain.*

18. The expression LLN is known to be due to Poisson. What contents did he himself insert in that notion? Here is his statement (Poisson 1837, p. 7; translated by Hald 1998, p. 576):

Things of every kind are subject to a universal law that we may call the law of large numbers. It consists of this: If we observe a very considerable number of events of the same nature, depending on constant causes and on causes varying irregularly, sometimes in one way, and sometimes in another, that is, without their variation being progressive in a deterministic sense, we will find that the ratios between these numbers are very nearly constant. For every kind of things these ratios will have a special value from which they deviate less and less the more the series of events increases, and which they would reach if it were possible to prolong this series to infinity.

If (see pp. 7 – 8) the observations were continued sufficiently, then, by comparing empirical deviations one with another, it will be possible to calculate according to known rules

The probability that that special magnitude to which those ratios tend to converge, is comprised in boundaries however near to each other. And, if, when making new experiments, we will discover that those same ratios notably deviate from their final value determined by the previous observations, it will be possible to conclude that the causes, on which the observed facts are depending, had experienced a progressive variation or even some sharp change during the time interval between the two series of experiments.

Poisson then indicates that an interpretation of the results of such comparisons should be based on the calculus of probability since otherwise wrong conclusions are possible. He provides a number of examples of the action of the LLN and concludes (p. 12):

Those examples taken from most various fields indicate that for us the universal LLN, being the result of never questioned observations is already a general and incontestable fact⁴⁰. [...] However, owing to its importance it was necessary to prove it directly, and that I have attempted to accomplish. And I believe that I have finally succeeded, as will be seen in the further exposition.

All those passages are contained in the *Préambule* and should be therefore only considered as the author's preliminary explanations. However, some definite conclusions can be already made. First, it is seen that Poisson regarded the LLN as a law of nature rather than an algebraic theorem. I note in passing that the Bernoulli theorem, as he (pp. 12 – 13) believed, "coincides with that LLN in the particular case in which the chances of the event remain constant during a series of trials".

Second, he thought that the LLN is corroborated by experience. Third, it can also be derived deductively, and he promises to accomplish that. See similar remarks on pp. 137, 143 and 246. Fourth, the LLN has some tight connection with the calculus of probability, but it is not yet clear what kind of connection that is. Finally, a reader inclined to epistemology could have

concluded that the LLN is apparently a prior law. Indeed, how can that law be refuted by experience if the author clearly indicates that a divergence of the empirical results not agreeing with the LLN points out that the conditions whose constancy is a premise of that law had changed in a way inaccessible to direct observations. The author himself later agrees with that conclusion⁴¹.

These simple observations concerning Poisson's text decidedly forbid me to agree with Bortkevich' opinion that Poisson's LLN

According to its literal sense (seinem Wortlaut nach), *is none other than a statement concerning the stability of the corresponding statistical numbers without specific indications about the degree of that stability*⁴².

Stability as a fact and a deductively justified universal law of nature are certainly logical categories infinitely remote from each other.

19. In the sequel we find a more precise formulation of the LLN. Suppose, says Poisson, some event occurred m_1 times in a very large number of trials μ_1 , then m_2 times in μ_2 also very large number of trials. Then (p. 139), "almost precisely and highly probably" (à très peu près et très probablement)

$$\frac{m_1}{\mu_1} = \frac{m_2}{\mu_2}. \quad (51)$$

After explaining the sense of a similar equality of the arithmetic means, s_1/μ_1 and s_2/μ_2 , Poisson (p. 143) proves the following definition:

And so, the LLN is contained in these two equations

$$\frac{m_1}{\mu_1} = \frac{m_2}{\mu_2}, \quad \frac{s_1}{\mu_1} = \frac{s_2}{\mu_2} \quad (52)$$

applicable to all cases of physical and moral things.

(Maintenant la loi des grands nombres réside dans ces deux équations [...] applicables à tous les cas d'éventualité des choses physiques et des choses morales.)

Taking into account his explanation that those equalities should be understood as approximate and only probable rather than certain, and that both the degree of approximation and the probability unboundedly heighten with the increase in the number of trials, we may write them down as

$$\lim_B \left(\frac{m_1}{\mu_1} - \frac{m_2}{\mu_2} \right) = 0, \quad \lim_B \left(\frac{s_1}{\mu_1} - \frac{s_2}{\mu_2} \right) = 0. \quad (53)$$

That is, the stochastic limit of the difference between the frequencies (the arithmetic means) is zero as the number of trials increases infinitely.

Poisson (p. 246) refers to these definitions and explanations when directly commencing to prove [the LLN]. Note that in the indicated places he

determines that law as a stochastic connection between empirically knowable magnitudes, means and frequencies, rather than between these latter and empirically (in principle, always, practically, in most cases) unknowable expectations of magnitudes and probabilities of events. That difference is not, however, essential since the connection between frequencies or means, which Poisson calls the LLN, is not direct but accomplished through the same unknowable in principle stochastic magnitude. Indeed, the stochastic limiting relations between frequencies or means (52; 53) will be valid only with a *constant* mean probability (expectation of a random variable).

A violation of these equalities does not yet testify to the inobservance of law but to an essential difference of mean probabilities (expectations) in the appropriate series of trials. Poisson repeatedly says so both in his *Préambule* and the main text. For my part, I note that both theoretically and, likely, practically, the stochastic connection between frequencies (means) and their mean probabilities (or, in general, expectations) should be given priority. Even to say nothing about their being logically more primary, the LLN is practically important for us mostly not because of its possible conclusions, given that the premise of constant conditions (expectations) *constituting its basis* is never strictly fulfilled, but owing to the provided possibility, when there is a sufficient number of observations, to estimate both the magnitude and the constancy or degree of change of exactly those principally unknowable quantities.

It was repeatedly stressed how important were experiments with urns, tossing of coins, etc for convincing us in the empirical reality or applicability of the LLN. Not so often was the attention possibly turned to the fact that only that law allows us to conclude by issuing from those experiments that the practical means invented for maintaining constancy and independence of chances (shuffling of cards, rotating lottery wheels, etc) are with a rather good approximation *practically sufficient* for that goal⁴³.

That is an empirical fact, impossible to establish without experimenting. No prior considerations can replace here the experiment, but to understand it and formulate appropriate conclusions is only possible on the basis of the prior LLN. Its formulation as applied to somewhat narrowed problems will be therefore

$$\lim_B \frac{m}{\mu} = p_0, \quad \lim_B \frac{s}{\mu} = E \frac{s}{\mu}. \quad (54)$$

Here, Poisson's notation is made use of in order to compare this formula with (52) and (53).

His own formulation is, however, only a particular case of our more general formula

$$\lim_B x = c \quad (55)$$

because, for constant p_0 and expectation $E(s/\mu)$, formulas (53) can be written as

$$\lim_B \left(\frac{m_1}{\mu_1} - \frac{m_2}{\mu_2} \right) = E \left(\frac{m_1}{\mu_1} - \frac{m_2}{\mu_2} \right) = 0 \quad (56)$$

and similarly for arithmetic means.

Poisson himself was not a stranger to the generalization of the notion of LLN to relations of the type of (54). At least in the *Préambule* he formulates it in exactly that sense. Thus, see the beginning of our § 18 [and end of § 8], he says:

These ratios will have a special value from which they deviate less and less the more the series of events increases, and which they would reach if it were possible to prolong this series to infinity.

He certainly had in mind the relation of the type

$$\lim_B x = c, n \rightarrow \infty, \quad (57)$$

cf. (55), rather than

$$\lim_B (x - x_1) = 0. \quad (58)$$

Our interpretation is corroborated since somewhat further Poisson (pp. 12 – 13) calls the Bernoulli theorem a particular case of the LLN, the theorem that again, being expressed by

$$\lim_B \frac{m}{n} = p, n \rightarrow \infty \quad (59)$$

is of the type of (57).

Since the author himself of the LLN thus hesitated over two formulations, it seems natural to conclude that he did not attach such an essential importance to the difference between them so that we, remaining true to his main ideas, may choose that which is apparently more important, more simple and general. Below, we will call them formulations I (given in the *Préambule*) corresponding to formulas (54) and (57) and II.

20. Turning now to the proof of the LLN provided by Poisson, we can only indicate his main idea. First of all, he *proves* it not in the more general formulation I, but in the narrower formulation II. For the sake of brevity, we only consider the case of frequencies, and we will be able to understand his point of departure by contemplating that formulation itself, cf. formulas (56):

$$\lim_B \left(\frac{m_1}{\mu_1} - \frac{m_2}{\mu_2} \right) = 0. \quad (60)$$

What are the conditions for that equality to be valid? When abstracting ourselves from the idea of a variable asymptote about which Poisson apparently did not think (although, as we remarked above, it was the mathematical logic of the computation itself that led him to formulate the

asymptotic approximation in the stochastic sense), and it is only possible to admit as a premise that, cf. (56),

$$\lim_B \frac{m_1}{\mu_1} = \lim_B \frac{m_2}{\mu_2}, \quad (61)$$

which in turn assumes the equality of the appropriate expectations, i. e., in this case, of the mean probabilities.

Constant probability, that is the Bernoulli case, is only a particular instance; mean probability of a constant composition will hardly occur anywhere beyond artificially arranged experiments (see however below). Mean probability of an arbitrary composition will most likely differ from series to series. Only one case, *mean probability in the proper sense*, is apparently left for being admitted as a premise for explaining observed statistical regularities.

Consequently, choosing that case as a basis for a deductive justification of the LLN in its formulation II, Poisson acted absolutely properly. However, he missed two circumstances. First, the case he considered was not the only one; second, his justification only concerned the formulation II. We will consider both points.

When discussing the Bernoulli theorem, Poisson (pp. 137 – 138) indicated that it most often happens in various applications of the calculus of probability that the probability of the occurrence of an event changes from trial to trial, and, once more most often, extremely irregular. That theorem is therefore not sufficient for studying such problems:

Nevertheless, there also exist other more general propositions which are valid however the consecutive chances [probabilities] of events are changing, and which are the basis of the most important applications of the theory of probability.

Then Poisson adds that he will prove them in the next chapters but that now he preliminarily describes them and, by issuing from them, will derive the LLN “as a universal fact following from observations of phenomena of most various nature”.

Propositions mentioned above are Poisson’s theorems about *mean probability of arbitrary composition* and *mean probability in the proper sense*. Curiously, he contrasts them with the LLN as with a proposition *not coinciding* with, but only *derivable* from them. This is the *literal sense* of Poisson’s words that cannot be ignored; if not contradicting the context or his other statements, it should be seriously allowed for. This is indeed the case. Poisson (pp. 137 and 139) states that the LLN is *derived* (dédire) from definite theorems which he indicates (including that about the mean probability in the proper sense) but it “almost precisely and highly probably” *resides* (p. 143) in the equality

$$\frac{m_1}{\mu_1} = \frac{m_2}{\mu_2}, \quad (62) = (51)$$

i. e., in

$$\lim_B \left(\frac{m_1}{\mu_1} - \frac{m_2}{\mu_2} \right) = 0. \quad (63)$$

Poisson repeatedly mentions the LLN and most decidedly expresses its universality and applicability to phenomena of most various nature, and it therefore seems unquestionable that the indicated difference between the *formulation* of that law and its *proof* means none other but his proper feeling (if not definite conviction) that the former is logically more general than the latter. I would have therefore never agreed with the opinion that Poisson clearly regarded the name LLN as referring to the case of *mean probability in the proper sense* and that to replace this by [mean] probability of a constant composition would be tantamount to “absolute misunderstanding” (eine gänzliche Verkennung) of his viewpoint, see Bortkevich (1917, pp. 53 and 54).

Defending his views, Prof. Bortkevich (pp. 54 – 55) indicates that for Poisson, *the task consisted in constructing a probability-theoretic pattern adequate to real, namely to “irregular change of random causes”, whereas the pattern of mean probability of constant composition is its exact antithesis since the appropriate probabilities are included in the mean in fixed proportions.*

However, a complete constancy of probabilities is a still more extreme special case whereas Poisson clearly considers the Bernoulli theorem as a particular case of the LLN; he (pp. 12 – 13) only believes that it does not cover “continually varying chances”⁴⁴. It follows that “an irregular change of random causes” is not a constitutional indication of the notion of the LLN. However, if nevertheless asking what was Poisson’s “task” when he pronounced the idea of his law, it would be certainly necessary to dwell on somewhat indefinite formulations in his *Préambule*, or still deeper on some vivid image of order in a chaotic change appearing in his mind in one, then in another formulation, gradually acquiring ever more definite mathematical forms until finding such in which the LLN can become the subject of a strict deduction.

There is one more place where, as it seems to me, the motif of introducing the idea of *mean probability in the proper sense* is seen absolutely clearly. Directly after the last-quoted passage above formula (62), Poisson reports the result of his investigations of the mean probability of arbitrary composition. Suppose that the probabilities of the occurrence of an event in consecutive trials are p_1, p_2, \dots, p_μ . Then, when having a sufficiently large number of observations, the frequency m/μ will be as near as desired “almost precisely and highly probably” equal to the mean probability. And,

given a sufficiently long series of observations, it will be an approximate expression of that mean probability, Poisson says. But, he (p. 138) continues, for that magnitude to determine approximately the frequencies in another series of trials, it is necessary, or at least very probable, for the mean probability in that series to be the same or almost the same. And exactly that he wishes to prove by his second theorem on the mean probability in the proper sense.

Suppose we tell Poisson that the same equality of the mean probabilities will be achieved under the assumption of mean probability of a constant composition, he certainly would not have formally rejected that indication by saying that it is not his LLN, but only noted that it is [its] particular and likely rare case as that of the Bernoulli theorem. And it is of no consequence that that theorem is directly derived from the theorem about the *mean probability in the proper sense* as a particular case under some definite assumption whereas the case of mean probability of constant composition is not thus derivable. It is more important that that latter case is a particular instance for which the equality (63) is valid. Poisson understands it as a form of expressing the LLN.

Suppose that molecules of s kinds are mixed in some substance proportionally to g_1, g_2, \dots, g_s with $g_1 + g_2 + \dots + g_s = 1$. If the probabilities of the explosion of one molecule in unit time are p_1, p_2, \dots, p_s , the mean probability will be

$$g_1p_1 + g_2p_2 + \dots + g_sp_s. \quad (64)$$

Then, if there are n molecules in a studied piece of that substance, unit time will be equivalent to n trials with the same mean probability, the mean probability of constant composition. The frequencies of molecular explosions and decompositions per unit time *almost precisely and highly probably* will be equal to (64) and to each other. I think that it will be absolutely unlikely that Poisson, having acknowledged the Bernoulli theorem as a particular case of the LLN, would have differently regarded that or a similar example rather than welcoming it as a still new confirmation of the universality and scope of his law.

However, the issue concerning mean probabilities of constant composition is secondary as compared with the widening of the notion of LLN occurring when moving from Poisson's formulation II to formulation I. And we saw that he, in his *Préambule*, kept exactly to that wider sense. The narrowing [that occurred in the main text] can only be considered as his oversight. Perhaps he thought that in its more special form that law covered all the conditions which can happen in reality but in any case we are unable to follow him here.

When formulating the notion of LLN as a general name for any proposition stating that the stochastic limit of the difference between such-and-such random variable and some other magnitude is zero as

the number of trials infinitely increases, we cover all cases answering both formulations. And, almost without deviating even from the literal expression of his most general formulations, we thus extend the notion of the LLN to many cases that Poisson had not envisaged.

Anyway, the most essential, constituting the basis of his intention, remains invariable: the idea of statistical limit and of the prior essence and widest applicability [of the LLN] to most various fields of reality. Even the greatest man is never able to formulate such an idea that later generations will not have to restrict it in some directions and widen it in other directions. To reject a name initially provided by the author of an idea means tearing the living connection which that name was bound to keep so as to prevent “ungrateful oblivion from penetrating the scroll of time”, which is an excellent expression due to St. Augustine.

Notes

1. The term *stochastics* is due to Jakob Bernoulli who applied it as a synonym of *ars coniectandi*, of the *art to measure the probability of things as precisely as possible*. His treatise thus entitled is, however, a systematic exposition of the doctrine of probabilities issuing from principles and art (in our understanding, *and from practice* only dealing with applications). It would be therefore more correct to say that Bernoulli equated *stochastics* with the [not yet existing] *theory of probability*. It is in this sense that I would consider it advisable to apply that term but to distinguish *stochastics* and *calculus of probability* as a purely mathematical discipline which it can and should become after the final demarcation, see my §§ 10 and 11.

Bernoulli himself did not consider his celebrated theorem as a purely mathematical proposition like the theorems of the theory of combinations etc, since, when separating the purely algebraic structure of the proof, he stressed that logical step by a clear declaration (Bernoulli 1713/2005, Chapter 5 of pt 4, p. 33): “I will attempt to reduce everything to abstract mathematics”.

Prof. Bortkevich had recently revived the term *stochastics*, and it has all chances to become firmly rooted. He himself understands it as

consideration of empirical totalities oriented toward the theory of probability and therefore based on the “law of large numbers”.

Chuprov says that he applies *stochastics* as a synonym of the expression *based on the theory of probability* and he also provides here another, narrower meaning which is hardly expedient bearing in mind the principle of definiteness of terminology. See Bernoulli (1713); Bortkevich (1917, [p. x and] p. 3); Chuprov (1923, p. 461; 1924, pp. 6 – 7 and 32). E. S. [See also Sheynin (2009, Note 1 to Chapter 3.). E. S.

2. When considering also that in an overwhelming number of cases the stochastic limit is expectation, and ascertaining the conditions for that to take place necessarily, the theorem mentioned easily leads to a proposition similar to that recently proven by Bohlmann (1913) on the limit of expectation of a function only based on a lesser number of premises. E. S.

In a letter of 1923, Chuprov (Sheynin 1990/1996, § 7.5.1) introduced the expression (never used by anyone) *convergence modo Bohlmann* (1913) for $Ef(x) \rightarrow f(Ex)$. The stochastic limit, as Slutsky called it, corresponds to convergence in probability. See also next Note. O. S.

3. I had come to the idea of stochastic limit by thoughts about the reports of Prof. Romanovsky delivered in 1922. He had applied the expression

$$x = c, \text{ modo Bernoulliano}$$

and thus, when formulating the stochastic relation now under discussion, he used an analogy with an approximate equality which can under certain conditions become ever

more precise. There, as it seemed to me, I perceived a stricter and more direct similarity with the [general] mathematical notion of limit. E. S.

[See Chuprov's response to Romanovsky's report in Sheynin (1990/1996, pp. 50 – 53). O. S.]

It turned out, however, that, being completely cut off from foreign literature, I did not then know that the Italian statistician Prof. Cantelli had expressed that idea already in 1916. At the end of 1924 Prof. Chuprov kindly brought it to my notice; already in summer 1923 he had friendly acquainted himself with the initial sketch of my paper (1925) whose relation with current literature could not have been clear to me owing to the conditions then being experienced [by Russian citizens] and I take this opportunity to express him my most heartily felt gratitude for that and for extremely valuable critical indications.

See Romanovsky (1922) and Cantelli (1916a, a paper I was still unable to see; 1916b; 1917a; 1917b; 1923). Cantelli's priority is doubtless; however, I have obtained a number of results not covered by his works with the very notion of stochastic limit, as far as I can judge, being more generally and wider developed. E. S.

4. The unusual notation for approximate equality was perhaps due to Markov (1900/1924, p. 62). O. S.

5. See Bernoulli's phrase quoted in Note 1 *I will attempt to reduce everything to abstract mathematics*. E. S.

6. The maximal term of the binomial $(p + q)^n$ is approximately equal to $1 / \sqrt{2\pi npq}$ and therefore decreases with an increasing n as $1 / \sqrt{n}$, see for example Feller (1950, § 3 of Chapter 6). O. S.

7. Cantelli (1916b, p. 339) suggests leaving symbol *lim* for the *limit* in the usual sense and applying *Lim* for the stochastic limit. It seems to me that such difference in notation is here not sufficiently expressive, and, in addition, that the custom of applying both symbols in the same sense is too rooted. E. S.

8. See Chuprov (1924, pp. 12 – 17/1960, pp. 167 – 174). As I see it, he made an extremely important logical step by distinguishing stochastic connection and stochastic dependence, cf. Chuprov (1922, pp. 241 – 242/2004, § 1.2). The former is the premise both for the latter and for independence. It only seems to me that his concept should be somewhat modified as I did it in my text for quite ascertaining his idea. Chuprov defines the notion of stochastic connection by opposing it to the concept of functional dependence.

If the value X being determined, the value Y will be random, the connection is stochastic, otherwise functional. First of all, that alternative is not complete. Chest measures of inhabitants of London and Paris are not connected either functionally or stochastically until it is somehow ascertained which individual should be compared with which. Such a comparison is a logical premise for further questions about the kind of connection, about dependence or independence, etc. In my definition, the relation between the notions of stochastic and functional connections is restricted to considering a *random variable*, that, unlike an independent variable, has a *distribution of probabilities* and therefore remains random when the values of the independent variable are fixed. In this aspect my definition does not differ from that of Chuprov. However, considering all the logical difficulties connected with the notion of function, and especially when comparing the two definitions, I intentionally left that notion alone. E. S.

9. In a number of memoirs beginning with (1906), and summarized in the posthumous edition of his treatise (1900/1924, p. 119ff). See also Chuprov (1918 – 1919, 1919, pp. 199 – 211/1968). E. S. [The page number is obviously wrong. O. S.]

10. I did not repeat the definitions literally but translated them into the language of other terms; nevertheless, I hope that their essence is provided quite correctly. E. S.

11. It seems to me that Bortkiewicz' terminology should be revised, but here I am unable to consider that task. Mean probability in the proper sense is in essence not at all mean in the *statistical* sense, it is the expected value of probability p , itself a random variable with possible values p_1, p_2, \dots, p_k , with their own probabilities $\pi_1, \pi_2, \dots, \pi_k$. E. S.

That terminology only describes the behaviour of the separate probabilities when they are constant (mean of constant composition). The other case is the only main alternative, but it can be further separated into sub-cases. Then, nothing is said about the possibility of unknown probabilities which Poisson certainly also envisaged. For that matter, it was Bortkiewicz himself who noted that point (Sheynin 2009, § 8.7). O. S.

12. On p. 139 I find Poisson's statement: the mean chance (probability) of E (of the event) should be considered the same for two or more series each of them consisting of a very large number of trials. O. S.

13. When describing that theorem, Bortkiewicz (1917, p. 55) notes that “the corresponding probabilities p_k enter into the mean in fixed (feststehenden) proportions”. We saw that Poisson did not assume that, so that neither does his theorem quite correspond to his task of establishing the pattern for studying phenomena influenced by absolutely arbitrarily changing causes, cf. Bortkiewicz (1894 – 1896, p. 655; 1917, p. 55). Poisson’s examples to which Bortkiewicz refers, illustrate his another theorem (on *mean probability in the proper sense*) and are therefore not conclusive, see Bortkiewicz (1894 – 1896, Ibidem). After all, for understanding what should be called such-and-such Poisson theorem, it is only important to note what did he actually prove rather than what he thought or possibly thought about it. In that sense, the problem as it seems to me is solved absolutely unquestionably. E. S.

14. See Chuprov (1909/1910), Bortkiewicz (1917) and Keynes (1921), a source containing valuable studies including investigations on the history of the LLN. I was still unable to acquaint myself with it. E. S.

15. See the remark extremely typical of him (Markov 1900/1924, p. 2n):

I think that various notions are defined not so much by words each of which can in turn demand definition, but rather by our attitude towards them which is being ascertained gradually. E. S.

Here is Markov’s no less telling statement (1911/1981, pp. 149 – 150):

I shall not defend these basic theorems connected with the basic notions of the calculus of probability [...] since I know that one can argue endlessly on the basic principles even of a precise science such as geometry [...]. O. S.

16. *The modern notion of function coincides with that of correspondence (Den moderne Begriff einer Funktion deckt sich mit dem einer Zuordnung);* Carathéodory (1918, p. 71). E. S.

17. If x and y are random variables stochastically connected with one and the same independent variable n , then, if also

$$\lim_B(x - y) = 0 \text{ as } n \rightarrow \infty,$$

x and y will be random statistical asymptotes for y and x respectively. I do not dwell on them here; incidentally, representing a considerably different stochastic formation, they ought to be otherwise symbolically noted. E. S.

18. Propositions mentioned in Note 17 are also included here since the difference of random variables is generally itself a random variable. E. S.

19. Above, see my remark connected with Note 15, I have indicated that Markov’s unfortunate formulation did not adequately represent his own idea. Indeed, after having proved that, under such-and-such assumptions the arithmetic mean of expectations was a stochastic asymptote of the arithmetic mean of the corresponding number n of values taken by a random variable, see above formula (37), Markov (1900/1924, p. 117) says: “in that does the LLN consist; its applicability to the series of magnitudes X, Y, Z, \dots considered by us we wished to establish”. This is nearer to our formulation than to his own. See also his pages 116, 121, 134, 173, 174.

Without referring to other authors, I only indicate as an illustration, the typical title of § 6 of Mises (1919b): *The Laws of Large Numbers*. Chuprov (1918 – 1919, 1919, p. 208) [the page number is obviously wrong – O. S.] says that “under such conditions, the LLN finds no applications for itself”, and as far as I can judge, he thus expresses the same, or almost the same idea. See also Cantelli (1916b, p. 343). E. S.

20. Weber & Wellstein (1913) properly ascertained, and, for that matter, in a comparatively very popular way, the essence and sense of the entire upheaval. Their philosophical commentary, on the other hand, leaves much to be desired. E. S.

21. I [iii] discussed the *calculus of alternatives*. Owing to various considerations, the term *disjunctive calculus*, kindly suggested later to me by Prof. Bortkiewicz in a private letter, seems to be more convenient. Regrettably, in 1922 I was not acquainted with Bernstein (1917) where the indicated ideas had already been realized, and more formally at that. That contribution deserves to be most seriously studied. E. S.

22. Chuprov (1909/1959, p. 168) stated that

All these mathematical constructions beginning with Bernoulli's initial theorem [as well as in Laplace's elegantly dressed form] up to the more general Poisson's law of large numbers, the even more general Chebyshev theorem, and the constructions by Nekrasov and Bruns, the most general of all of them, are separated by an abyss from the law of large numbers that establishes the connection between probabilities of phenomena and their frequencies. They are only theorems from the field of the theory of combinations.

Kries (1886, p. 91) expressed himself similarly. E. S.

Above, I have inserted the statement about Laplace which Slutsky had omitted. Note that Chuprov did not mention De Moivre.

Markov was "astonished" that Nekrasov's name thus appeared together with Chebyshev's, but Chuprov reasonably answered that he did not at all compare their scientific merits. Thus began their correspondence, extremely profitable for both scientists (Ondar 1977, Letters 1 and 2 of 1910). O. S.

23. As it seems, Bortkiewicz (1894 – 1896, p. 665, Note 2) too severely admonishes Poisson for insufficiently understanding the *physical* interpretation of the LLN in the epistemological sense. It should be borne in mind that exactly in that sense the problem cannot up to now be considered sufficiently ascertained and the best testimonial is the huge volume of investigations, Meinong (1915). There, on p. 599, we read (my own underscoring): "The basis of the LLN is still as obscure as the Bernoulli theorem is clear". E. S.

24. The attempts made in this direction were not serious. See D'Alembert (1767, pp. 275ff, 298) as referred to by Czuber (1903/1908, p. 145); Marbe (1899, pp. 30 – 39) as quoted by Bortkiewicz (1913, p. 145ff; 1903). On p. 82 in the last-mentioned source Bortkiewicz remarks that Mill had apparently noted that he would have to abandon the theory of probability if "seriously admitting that, so to say, purely statistical interpretation of the notion of mathematical probability". See also Bortkiewicz (1923, pp. 14 – 15). My criticism in the main text does not essentially represent anything new as compared with Kries' (1886, Kap. 1, NNo. 5 – 6) thoughtful considerations. I think that Meinong (1915, pp. 597 – 598) absolutely unconvincingly defends Marbe's viewpoint. E. S.

25. I have not found the corresponding page in the edition of 1959. O. S.

26. Kries (1886, p. 21). The previous example is also his; I only replaced 1000 by 10^{10} . In principle, there is no difference between series of repetition twice, or 10, or a thousand times, or even $(10^{10})^{10}$ times. E. S.

27. This expression is due to N. S. Chetverikov. E. S.

28. Cournot (1851/1912, p. 45 [§§ 33 – 35, 38, 51 – 52]). Only the first underlining is due to him. I note in passing that he was hardly successful when contrasting *physical* or *factual impossibility* and *mathematical* or *metaphysical*, also *rational* or *absolute impossibility*. There is *no transition* from one to another, and that is a significant acknowledgement. Physically impossible is an event having one chance against an infinity of all the chances; *mathematically* or *absolutely* it is possible, but not *physically*, believes Cournot (1843, §§ 43 – 44 and 240.5 – 240.8), but his explanations are insufficiently clear. On infinitely low probability see Ch. Lagrange (1901). E. S.

29. Even such a "subjectivist" as Stumpf (1890, p. 110, in passing) was unable to manage without the notion of "real chances". In spite of its being one-sided, his work is subtle and transparent and, as everything penned by him, it represents an important contribution to the logic of probability. E. S.

30. Chuprov's own standpoint should apparently be understood exactly in that sense. This is being hinted both by his careful wording of the "lemma" without a direct indication of the nomological sense, and a clear statement that the LLN "represents a synthesis of nomographic and ontological elements" (1909/1959, p. 168). E. S.

31. So where is ontology here? O. S.

32. Here, the difference between our outlook from the views of the ancients does apparently exist. Teophratus sympathetically quotes Heraclitus of Ephesus and believes that it is absurd to think that the most beautiful order of the world is a randomly scattered rubbish heap (Makovelsky 1914, Heraclitus, Fragment 124, p. 167). E. S.

33. Cf. Zilsel (1921). It is surprising that an author, after formulating propositions tantamount to the two just explicated, can believe that his construction makes the stochastic viewpoint unnecessary. E. S.

Kolmogorov's remark (1972) directly bears on the issue of revealing the meaning of information obtained, although in a very special imagined case. He stated that messages

sent by much higher developed creatures living somewhere in the universe will be attributed to random noise. O. S.

34. Matthiesen(1867) reported that in a game of whist each of the four participants got cards of one and only one suit. His testimony certainly cannot be checked. O. S.

35. How could have the laws of nature changed? O. S.

36. When Boutroux [perhaps Boutroux (1908)] mentions randomness of the laws of nature themselves, his idea belongs to quite another order of ideas and can only be understood in the light of the concept of absolute randomness [?]. E. S.

37. [Bortkiewicz'] reference to the established usage of words in statistics is not convincing. Statisticians have sinned too much by confusing *facts* and *laws*. He himself (1918, p. 115n), when arguing against Bienaymé, wrote: "That he did not object to consider empirical coefficients as laws of nature is not anymore surprising". E. S.

38. Not law of stability, as Romanovsky (1924, No. 4 – 6, p. 15n) suggests. E. S.

39. Slutsky indirectly refers to the then still popular law of small numbers; see, however, Sheynin (2008). O. S.

40. Poisson left many of his examples without any proof; in such cases, he simply stated that the pertinent phenomenon was stable. O. S.

41. "Owing to the importance of that law, it would have been good to see it proved as a prior proposition" (p. 138). E. S.

I do not see any connection of this Note with the text. It rather has to do with the reference from pp. 12 – 13 above. O. S.

42. Bortkiewicz (1917, p. 58 with his emphasize). We should agree that some of Poisson's phrases can be interpreted in that sense, but his exposition is not at all a specimen of condensed strictness and is written contrary to the style of strict mathematical, logical or legal formulations where each word and phrase are thought out. There are only a few authors who could have repeated what Lipps (1893, p. viii) said about himself: "I am only asking always to bear in mind that each word in this book is properly considered".

As to Poisson (1837), a general rule is that each phrase should be understood not only literally, but in the general context, together with other relevant places. Regarding the degree of stability, Poisson, already in his *Préambule*, refers in a general sense to his deductions in connection with the LLN, i. e., to his long calculations in Chapter 4. E. S.

43. This is questionable. Thus, many Americans had been doubting that the extraction of tokens from an urn was not "uniformly" random (Fienberg 1971). See Muller (1978, p. 841) for further sources. I also note that Slutsky's (and Chuprov's) problem about the applicability of the LLN should have been specified: actually, it was necessary to find out whether the *premises* of that law were obeyed in reality. O. S.

44. I translated the last phrase from Poisson's own text. O. S.

Bibliography

Abbreviation: VS = *Vestnik Statistiki*

Bernoulli, J. (1713), *Ars Conjectandi*. Reprint in author's *Werke*, Bd. 3. Basel, 1975, pp. 107 – 259. English translation of pt. 4 by O. Sheynin in Bernoulli's book *On the Law of Large Numbers*. Berlin, 2005. Also at www.sheynin.de

Bernstein, S. N. (1917), An essay on an axiomatic justification of the theory of probability. *Sobranie Sochinenii* (Coll. Works), vol. 4. No place, 1964, pp. 10 – 60. Translated in Sheynin (2005, pp. 49 – 111).

Bohlmann, G. (1913), Formulierung und Begründung zweier Hilfsätze der mathematischen Statistik. *Math. Annalen*, Bd. 74, pp. 341 – 409.

Bortkiewicz, L., Bortkevich, V. I. (1894 – 1896), Kritische Betrachtungen zur theoretischen Statistik. *Jahrbücher f. Nationalökonomie u. Statistik*, 3. Folge, Bd. 8, pp. 641 – 680; Bd. 10, pp. 321 – 360; Bd. 11, pp. 701 – 705.

--- (1903), Wahrscheinlichkeitstheorie und Erfahrung. *Z. f. Philosophie u. philos. Kritik*, Bd. 121, pp. 71 – 86.

--- (1913), *Die radioaktive Strahlung als Gegenstand wahrscheinlichkeitstheoretischer Untersuchungen*. Berlin.

--- (1917), *Die Iterationen*. Berlin.

--- (1918), Der mittlere Fehler des zum Quadrat erhobenen Divergenzkoeffizienten. *Jahresber. d. deutschen Mathematiker-Vereinigung*, Bd. 27, pp. 71 – 126.

--- (1923), Wahrscheinlichkeit und statistische Forschung nach Keynes. *Nordisk Statistisk Tidskrift*, Bd. 2, pp. 1 – 23.

- Bortkevich, V. I., Chuprov, A. A.** (2005), *Perepiska (Correspondence) 1895 – 1926*. Berlin. Also at www.sheynin.de
- Boutroux, P.** (1908), Les origines du calcul des probabilités. *Rev. du Mois*, 3^e année, t. 5, No. 6, pp. 641 – 654.
- Cantelli F. P.** (1916a), La tendenza ad un limite nel senso del calcolo delle probabilità. *Rendiconti del circolo matematico di Palermo*, t. 41, pp. 191 – 201.
- (1916b), Sulla leggi dei grandi numeri. *Atti della Reale Accademia dei Lincei*, Cl. di sci. fis., mat. e nat., ser. 5, t. 11, pp. 330 – 349.
- (1917a), Sulla probabilità come limite della frequenza. *Ibidem*, t. 26, pp. 39 – 45.
- (1917b), Su due applicazioni di un teorema di G. Boole alla statistica matematica. *Ibidem*, pp. 295 – 302.
- (1923), Sulla oscillazione delle frequenze intorno alla probabilità. *Metron*, t. 3, No. 2, pp. 167 – 174.
- Carathéodory, C.** (1918), *Vorlesungen über reelle Funktionen*. Leipzig. [New York, 1948.]
- Chuprov, Tschuprow, A. A.** (1905), Die Aufgabe der Theorie der Statistik. *Schmollers Jahrb. f. Gesetzgebung, Verwaltung u. Volkswirtschaft im Dtsch. Reich*, Bd. 29, No. 2, pp. 421 – 480.
- (1906), Statistik als Wissenschaft. *Arch. f. soz. Wiss. u. soz. Politik*, Bd. 5 (23), No. 3, pp. 647 – 711.
- (1909), *Ocherki po Teorii Statistiki* (Essays on the Theory of Statistics). Moscow, 1910, 1959.
- (1918 – 1919), Zur Theorie der Stabilität statistischer Reihen. *Skandinavisk Aktuarietidskrift*, Bd. 1, pp. 199 – 256; Bd. 2, pp. 80 – 133.
- (1922, in Russian), On the expectation of the ratio of two mutually dependent random variables. *Trudy Russkikh Uchenykh Zagranitsej*, vol. 1, pp. 240 – 271. Translated in author's book (2004, pp. 120 – 157).
- (1923), On the mathematical expectation of the moments of frequency distributions in the case of correlated observations. *Metron*, t. 2, pp. 461 – 493, 646 – 683.
- (1924, in Russian), The main issues of the stochastic theory of statistics, *VS*, No. 10 – 12, pp. 5 – 67; in author's book (1960, pp. 162 – 221).
- (1960), *Voprosy Statistiki* (Issues in Statistics). Moscow. Coll. reprints and translations.
- (2004), *Statistical Papers and Memorial Publications*. Berlin. Also at www.sheynin.de
- Cournot, O.** (1843), *Exposition de la théorie des chances et des probabilités*. Paris, 1984.
- (1851), *Essais sur les fondements des nos connaissances*. Paris, 1975.
- Czuber, E.** (1903), *Wahrscheinlichkeitsrechnung und ihre Anwendungen*, Bd. 1, 1908. New York, 1968.
- D'Alembert, J. Le Rond** (1767), Réflexions philosophiques et mathématiques sur l'application des calcul du probabilité. *Mélanges de littérature, d'histoire et de philosophie*, t. 5. Amsterdam, pp. 267 – 367.
- Feller, W.** (1950), *Introduction to Probability Theory and Its Applications*, vol. 1. New York, 1957.
- Fienberg, S. E.** (1971), Randomization and social affairs: the 1970 draft lottery. *Science*, vol. 171, pp. 255 – 261.
- Hald, A.** (1998), *History of Mathematical Statistics from 1750 to 1930*. New York.
- Hilbert, D.** (1899), *Grundlagen der Geometrie*. Second edition: Leipzig, 1903.
- Höfler, A.** (1922), *Logik*, second edition. Leipzig.
- Husserl, E.** (1900), *Logische Untersuchungen*. Halle, 1913. [Halle, 1928; English translation: London, 1973. Russian translation: 1909.]
- (1913), Ideen zu einer reinen Phänomenologie und phenomenologischen Philosophie. *Jahrbuch f. Philos. u. phän. Forschung*, Bd. 1, Tl. 1. Separate edition with detailed Subject Index by Herda Walther: 1922 – 1923.
- Keynes, J. M.** (1921), *Treatise on Probability*. *Coll. Woks*, vol. 8. London, 1973.
- Khinchin, A. Ya.** (1928, in Russian), The strong law of large numbers and its significance for mathematical statistics. *Vestnik Statistiki*, No. 1, pp. 123 – 128.
- Kolmogorov A. N.** (1972, in Russian), Probability is not equal to zero. Newspaper *Literaturnaia Gaset*a, 20 Sept., p. 12.
- Kries, J.** (1886), *Die Principien der Wahrscheinlichkeitsrechnung*. Freiberg i. B. [Tübingen, 1927.]

- (1888), Über den Begriff der objektiven Möglichkeit. *Vierteljahrsschrift f. wiss. Philosophie*, 12. Jg., pp. 179 – 240, 287 – 323, 393 – 428.
- Lagrange, Ch.** (1901), Etude du principe de la limite. *Bull. Acad. Roy. de Belg.*, classe de sciences, NNo. 9 – 10.
- Lipps, Th.** (1893), *Grundzüge der Logik*. Bonn.
- Makovelsky, A.** (1914), *Dosokratiki* (Presocratics), pt. 1. Kazan.
- Marbe, K.** (1899), *Naturphilosophische Untersuchungen zur Wahrscheinlichkeitslehre*. Leipzig.
- Markov, A. A.** (1900), *Ischislenie Veroiatnostei* (Calculus of Probability). Later editions: 1908, 1913 and Moscow, 1924. German translation: Leipzig – Berlin, 1912.
- (1906, in Russian), Extension of the law of large numbers on mutually dependent magnitudes. In author's book (1951, pp. 339 – 361). Translation: Sheynin (2004, pp. 143 – 158).
- (1911, in Russian), On the basic principles of the calculus of probability and on the law of large numbers. In Ondar (1977/1981, pp. 149 – 153).
- (1951), *Izbrannye Trudy* (Sel. Works). No place.
- Matthiesen, L.** (1867), Vermischtes aus dem Gebiete der Wahrscheinlichkeitsrechnung. *Archiv f. Math. u. Phys.*, Bd. 47, pp. 457 – 460.
- Meinong, A.** (1915), *Über Möglichkeit und Wahrscheinlichkeit*. Leipzig.
- Mill, J. S.** (1843), *System of Logic*. London, 1886, 1898. [*Coll. Works*, vol. 8. Toronto, 1974.]
- Mises, R.** (1919a), Fundamentalsätze der Wahrscheinlichkeitsrechnung. *Math. Z.*, Bd. 4, pp. 1 – 96.
- (1919b), Grundlagen der Wahrscheinlichkeitsrechnung. *Math. Z.*, Bd. 5, pp. 52 – 99.
- (1964), *Mathematical Theory of Probability and Statistics*. New York – London.
- Muller, M. E.** (1978), Random numbers. In Kruskal, W., Tanur, J. M. (1978), *Intern. Enc. of Statistics*, vols 1 – 2. New York, pp. 839 – 847.
- Ondar, Kh. O., Editor** (1977, in Russian), *The Correspondence between A. A. Markov and A. A. Chuprov on the Theory of Probability and Mathematical Statistics*. New York, 1981.
- Poisson, S.-D.** (1837), *Recherches sur la probabilité des jugements* etc. Paris. [Paris, 2003.]
- Romanovsky, V. I.** (1922, in Russian), On the linear correlation of two magnitudes. *VS*, No. 9 – 12, pp. 23 – 28.
- (1924, in Russian), Theory of probability and statistics. *VS*, No. 4 – 6, pp. 1 – 38; No. 7 – 9, pp. 5 – 34.
- Sheynin, O.** (1990, in Russian), *A. A. Chuprov. Life, Work, Correspondence*. Göttingen, 1996.
- (2004), *Probability and Statistics. Russian Papers*. Berlin. Coll. translations. Also at www.sheynin.de
- (2005), *Probability and Statistics. Russian Papers of the Soviet Period*. Berlin. Coll. translations. Also at www.sheynin.de
- (2008), Bortkiewicz' alleged discovery: the law of small numbers. *Hist. Scientiarum*, vol. 18, pp. 36 – 48.
- (2009), *Theory of Probability. Historical Essay*. Berlin. Second edition. Also at www.sheynin.de
- (2010), The inverse law of large numbers. *Math. Scientist*, to appear.
- Slutsky, E. E.** (1925), Über stochastische Asymptote und Grenzwerte. *Metron*, t. 5, No. 3, pp. 3 – 89.
- (1960), *Izbrannye Trudy* (Sel. Works). Moscow.
- Stumpf, C.** (1890), Über den Begriff der mathematischen Wahrscheinlichkeit. *Sitz. Ber. Kgl. Bayer. Akad. Wiss. zu München*, philos.-philolog. u. hist. Cl., Jg. 1892, pp. 37 – 120.
- Weber, H., Wellstein, I.** (1913), *Enziklopedia Elementarnoi Matematiki* (Enc. of Elementary Math.), Book 1. Second edition. Odessa. This is exactly how Slutsky referred to it.
- Zilsel, V.** (1921), Versuch einer neuen Grundlegung der statistischen Mechanik. *Monatshefte f. Math. u. Phys.*, Bd. 31, pp. 153 – 154.

IX

Al. A. Tschuprov

Z. f. angew. Math. u. Mech., Bd. 6, 1926, pp. 337 – 338

Aleksandr Aleksandrovich Tschuprow died in Geneva on 19 April of the present year at the age of 52 years after a serious and protracted heart disease. The statistical science has thus lost a theoretician of the very first rank, a man who devoted his life to revise essentially the logical and probability-theoretic foundations of statistics.

How early had A. A. Tschuprow come to perceive his scientific aim is proved by his entering the mathematical faculty of Moscow University at the age of 18 with a quite definite intention of mastering mathematics and then of applying it to social sciences. After graduating, he studied those sciences in Berlin and Strasbourg, mostly under L. von Bortkiewicz¹ and G. Knapp. The former helped the young scholar, who had already developed many of his early ideas in his Moscow [student] dissertation, to delve deeper in the new Lexian direction. To Knapp's school Tschuprow owed the perfection of his ability to assimilate facts.

His teaching began already in 1902 as chair of statistics at the Economic faculty of the then newly established Petersburg Polytechnic Institute and lasted very successfully until 1917. In summer of that year he went holidaying abroad for the last time not ever to return.

It is impossible to honour here the great wealth of his scientific merits. His *Ocherki* (1909) whose main ideas became also accessible to German readers owing to his papers (1905; 1906), already was an achievement in a big way. He synthetically and brilliantly described a number of the main problems of statistical theory and organically fused together the ideas of Windelbrand, Rickert, Cournot and Knies with the Lexis – Bortkiewicz direction [of statistics] and his own deep thoughts².

Tschuprow's study (1916) showed him as a real master of empirical investigation. However, what earned him a great reputation and ensured him a place in the history of statistics as one of its creative thinkers was a number of logical and mathematical studies which he accomplished during his last decade. A clear separation of prior and posterior elements of statistical science based on probability theory; a definite grasp of various problems occurring during examination of the former by means of the empirical data concerning the latter; the further construction of the Chebyshev – Markov method of expectations and its application to the most difficult issues of the stability of statistical series³; moments of densities and problems of the correlation theory, – these were the subjects of the scholar carried off from us by such an early death amid work yet envisaged to last for a long number of years. The unification of the English and Continental directions in the theory of statistics can be already considered successful and in many main features an accomplished fact⁴.

High appraisals of Tschuprow's scientific merits are not lacking either. From long ago he was Member of the International Statistical Institute; in 1917 he was elected Corresponding Member of the Petersburg [Petrograd] Academy of Sciences, and in 1923, Honourable Fellow of the Royal

Statistical Society. During his journeys to Scandinavia in 1924 when he delivered reports on correlation theory, one in Copenhagen, and several in Oslo, he was happy to perceive a full and solemn acknowledgement of his ideas.

Tschuprow was deeply knowledgeable and subtly felt for arts and at the same time he was gifted in displaying his real sympathy and intimate understanding when dealing with those surrounding him. Although the entire way of his life was with great skill and endurance adapted to scientific studies, his correspondence during his last years with his [former] students and colleagues had been occupying an essential part of his working time. And he, who knew how to satisfy skilfully all his requirements, did not wish to cut down here. And thus will he live further in our memory: not only as a great scientist and distinguished teacher, but also as a great noble person.

Notes

1. This is doubtful. In 1897 Chuprov wrote to his father (Sheynin 1990/1996, p. 37) that Bortkiewicz “cannot be my mentor [...], the difference in knowledge between us is insufficiently large [...]”. O. S.
2. See Foreword. O. S.
3. Chuprov did apply the method of moments, but hardly constructed it “further”. O. S.
4. An explanation lacking here is that that unification was Chuprov’s main goal. An author, acting on the request of Oskar Anderson, Chuprov’s former student, decided in 1955 that the application of the analysis of variance was going side by side in both schools but did not tend to unification and reasonably remarked that, unlike their British counterparts, Continental statisticians had been concentrating on nonparametric statistics. See Sheynin (2009a, § 15.3). I think that also in general the situation was similar. O. S.

Bibliography

A. A. Chuprov

- (1905), Die Aufgabe der Theorie der Statistik. *Schmollers Jahrb. f. Gesetzgebung, Verwaltung u. Volkswirtschaft im Dtsch. Reich*, Bd. 29, No. 2, pp. 421 – 480.
- (1906), Statistik als Wissenschaft. *Arch. f. soz. Wiss. u. soz. Politik*, Bd. 5 (23), No. 3, pp. 647 – 711.
- (1909), *Ocherki po Teorii Statistiki* (Essays on the Theory of Statistics). Moscow, 1910, 1959.
- (1916), Zur Frage des sinkenden Knabenüberschusses unter den ehelich geborenen etc. *Bull. Intern. Stat. Inst.*, t. 20, No. 2, pp. 378 – 492.
- (1923), On the mathematical expectation of the moments of frequency distributions in the case of correlated observations. *Metron*, t. 2, pp. 461 – 493, 646 – 683.

Other Authors

- Bortkevich, V.I., Chuprov, A. A.** (2005), *Perepiska* (Correspondence) 1895 – 1926. Berlin. Also at www.sheynin.de
- Khinchin, A. Ya.** (1928, in Russian), The strong law of large numbers and its significance for mathematical statistics. *Vestnik Statistiki*, No. 1, pp. 123 – 128.
- Markov, A. A.** (1911, in Russian), On the basic principles of the calculus of probability and on the law of large numbers. In Ondar (1977/1981, pp. 149 – 152).
- Materialy** (1991), *Materialy o V. A. Steklove* (Materials about V. A. Steklov). *Nauchnoe Nasledstvo*, vol. 17. Leningrad.
- Ondar, Kh. O.**, Editor (1977, in Russian), *The Correspondence between A. A. Markov and A. A. Chuprov* etc. New York, 1981.
- Romanovsky, V. I.** (1930), *Matematicheskaja Statistika* (Math. Statistics). Moscow – Leningrad.
- Sheynin, O.** (1982), On the history of medical statistics. *Arch. Hist. Ex. Sci.*, vol. 26, pp. 241 – 286.
- (1990, in Russian), *A. A. Chuprov. Life, Work, Correspondence*. Göttingen, 1996.
- (2003), Nekrasov’s work on the central limit theorem. *Arch. Hist. Ex. Sci.*, vol. 57, pp. 337 – 353.

- Translator (2007), *Chetvertaia Khrestomatia po Istorii Teorii Veroiatnostei i Statistiki* (Fourth Reader in the History of Probability and Statistics). Berlin. Also at www.sheynin.de
- (2008), Bortkiewicz' alleged discovery: the law of small numbers. *Hist. Scientiarum*, vol. 18, pp. 36 – 48.
- (2009a), *Theory of Probability. Historical Essay*. Berlin. Also at www.sheynin.de
- (2009b), *Studies in the History of Statistics and Probability*. Coll. translations. Berlin. Also at www.sheynin.de

X

On the Distribution of Errors [On the Law of Distribution] of the Correlation Coefficient in Homogeneous Connected Series

O raspredelenii oshibok koeffizienta korreliatsii v odnorodnykh svyazannykh riadakh.
Zurnal Geofiziki, vol. 2, 1932, pp. 66 – 98

Summary [in its original English]¹

1. *The Technical terms.* We shall call the chance values x_0, x_1, \dots, x_n associated with the values $0, 1, 2, \dots, n$ (or $0, h, 2h, \dots, nh$) of an independent variable (time, etc) a *chance series* or a *chance function*. It will be called a *random series* if its terms are mutually independent, and a *coherent series* in the contrary case. The chance series will be called a *homogen[e]ous* one, if the law [law] of probability for a single value is constant, and if the law of probability for every set of s values of $x_\alpha, x_\beta, x_\gamma, \dots, x_\nu$ is depending only on the differences of indices $\beta - \alpha, \gamma - \alpha, \dots, \nu - \alpha$.

In many cases it is sufficient to suppose a *relative homogen[e]ity* of the order m defined by the analogous suppositions relating to the moments of any order $k < m$.

2. *Thema probandum.* In a former paper I (1930) have deduced the formulae for the standard error of the correlation coefficient in the case of homogen[e]ous coherent chance series and developed a method for their practical application to the empirical series. In this paper I propose to give an empirical verification of the hypothesis *that in the case of homogen[e]ous coherent chance series with the terms, the values of which have a distribution not very different from the normal one, the distribution of the Fisher's function $z = \text{Arctanh } r$ obeys very nearly to the normal law [law] with (I) the centre in a point tending to the theoretical (a priori) value $\text{Arctanh } (v)\rho$ [?] when the number (n) of terms of the resp. series indefinitely increases and with (II) a standard deviation σ_s being a function of σ_r, ρ and n, ρ being the theoretical correlation coefficient.*

3. *The series used.* To prove our hypothesis, the following series were used. (I) Model I' consisting of two series: $x_i, y_i, i = 1, 2, \dots, 1050$, described in the paper cited above, p. 68. The correlation coefficient $\rho_{xy}(0) = 0$. (II) Sir Beveridge's Index Numbers for wheat in Western Europe [1922], $N = 360$. The correlation coefficients $\rho_x(t)$ between x_i and x_{i+t} for $t = 10$ are presumably 0, this assumption being based on an investigation which will be published on another occasion.

(III) The model imitating Beveridge's series having the same moments m_1, μ_2, μ_3, μ and the same serial correlations from $r_x(1)$ till $r_x(10)$. $N = 1251$. The method of construction will be described on another occasion. (IV) Model V described in the paper cited above (pp. 72 – 73) consisting of two series $x_i, y_i, i = 1, 2, \dots, 1000$. $\rho_{xy} = 0.814$.

(V) Model V' constructed according to the scheme practically coinciding with the scheme of model V, the corresponding chance variables being consequently also practically identical. The details see above § 8. This model consists of 280 partial models mutually independent and consisting each of two series $x_i, y_i, i = 1, 2, \dots, 20$ the a priori correlation coefficient being $\rho_{xy} = 0.816$.

4. *The verification of the hypothesis.* Let n be the number of the pairs of values from which the correlation coefficient r is determined and m the number of the correlation coefficients obtained. In some cases cited below several partial sets of the values $z_i^{(k)}$, $k = 1, 2, \dots, n_s$; $i = 1, 2, \dots, s$, corresponding to the given $r_i^{(k)}$ for different values of $n_1 = n, n_2, \dots, n$ [?] were reduced to one set of values by calculating the values of $(z - \bar{z}) / \sigma_s$ the number of which is also designated by m . N is the total number of values or pairs of values in the corresponding series, ρ is the theoretical correlation coefficient. χ^2 is the Pearsons criterion and P the corresponding probability. The results are given in the table below (Table 13). One remarks that the results are very favourable for our hypothesis in all cases but one, namely that under number 8 (model V).

5. *The negative instance.* In the first stage of this work the case of the model V was the last one. It came to mind that all the correlation coefficients being found not especially for this paper they could be regarded in every case as a set of chance values mutually correlated in the manner not consisting with the theory of χ^2 for usual distributions. To elucidate this point a coherent chance series was obtained consisting of 1280 terms being a realization of a chance variable with the possible values 0, 1, 2, 3, 4 the corresponding probabilities 1/16, 4/16, 6/16, 4/16, 1/16 and the serial correlations $r(1) = 1/4$, $r(2) = 1/2$, $r(3) = 1/4$, $r(t) = 0$ for $t \geq 4$.

Table 8 shows the values of χ^2 for one set of 1280 values, for two series of 640 values and so on. It is evident that the results are incompatible with the usual theory of χ^2 . In absence of the adequate formula for χ^2 it seems however probable that the correlation between the terms of a series must in the most cases rather exag[er]ate the value of χ^2 than to diminish it. To prove strictly at least the case of model V (a negative instance) the model V' (practically identical with V) was constructed in such a manner that the correlation coefficients should be mutually independent. The result being a positive one our hypothesis can be regarded as non-contradicted by all the material involved in the investigation.

6. *On the standard deviation of z .* The theory of the standard deviation of r being given in the paper cited above, we need the means to find σ_2 [σ_s] as a function of σ_r , ρ and n . Assuming that the error of centring the distribution of z in $\text{arctanh } \rho$ can be (for n great) neglected, we obtain at once the distribution function of z given by the formulae 20 – 22 whence the distribution of r will be given by (24) and the mathematical expectation of $F(r)$ by (25) or by (26). By rather a lengthy algebra we get the semiconvergent series for σ_r (see formula 56) the inversion of which gives the semiconvergent series for σ_s , see formulae (62; 63; 64) and Table 12. For practical purposes one can use the *Abac* of Fig. 1.

Table 13 provided the summary lists of ρ with indices xy , $x_{i,i+t}$, $x_{i,i+1}$, $x_i y_i$, then N , n , m , χ^2 and P for the different models. O. S.

1. Introductory Remarks

It is only possible to discuss the correlation coefficient in the true sense with regard to magnitudes about which we certainly know or hypothetically presume on some grounds that it is worth applying to them the concepts and patterns of the theory of probability. Assuming that condition, we will begin

by recalling several main definitions, either generally known, or introduced by me in my previous papers.

Let x be a random variable, i. e., a magnitude taking definite numerical values x_1, x_2, \dots, x_s with definite probabilities p_1, p_2, \dots, p_s ; here, we do not need a more general definition. The sum $\sum p_i x_i$ is called the expectation of x and denoted Ex . The mean square deviation of x is

$$\sigma_x = \sqrt{E(x - Ex)^2}, \quad (1)$$

and

$$\rho_{xy} = \frac{E(x - Ex)(y - Ey)}{\sigma_x \sigma_y} \quad (2)$$

is the coefficient of correlation between random variables x and y .

For calculating ρ_{xy} , it is obviously necessary to know not only the distributions of probabilities [laws of distribution] of x and y , but also the probabilities of a joint occurrence of each pair of their possible values. A simplest case, in which the exact value of the correlation coefficient can be known beforehand, is for example the following.

There are n urns and tickets contained in them are somehow numbered in the same way. They are extracted with replacement and n series of numbers, of the results of independent trials made on n identical random variables, are thus obtained. Separate all these series into three parts of m , g and g series ($m + 2g = n$) and form two series, x_1, x_2, \dots and y_1, y_2, \dots by combining the first part with the second, then with the third. Then any x_i and y_i will have ($m + g$) terms, m of them being common, and it is not difficult to show that the coefficient of correlation between x and y will be $m/(m + g)$.

Lacking such a favourable situation, we cannot calculate the correlation coefficient, but in many cases we will be able to derive its statistical analogue, the *empirical coefficient of correlation*, or (L. March's suggestion) *coefficient of covariation*². Suppose that series

$$x_1, x_2, \dots, x_n \text{ and } y_1, y_2, \dots, y_n \quad (3)$$

are given. Then the empirical mean square [deviations] will be

$$s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}, \quad s_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n}} \quad (4)$$

and, according to the known Pearson formula, the *covariation (the empirical correlation coefficient)* will be

$$r_{xy} = \frac{(1/n) \sum (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}. \quad (5)$$

If the parameters of probability and connection, or, otherwise, the stochastic conditions for the appearance of the appropriate magnitudes remain constant over each of the series (3), then, calling such series

homogeneous, we state that under some additional conditions the so-called law of large numbers will take place. And, when n increases unboundedly, the magnitudes

$$\bar{x}, \bar{y}, s_x, s_y, \text{ and } r_{xy} \text{ will have } Ex, Ey, \sigma_x, \sigma_y, \text{ and } \rho_{xy}$$

as their stochastic limits. In other words, for a sufficiently large n each of those magnitudes will differ from its limit less than by an arbitrarily small magnitude ε except cases whose probability is less than a no matter how small magnitude η^3 .

An investigator is interested in the stochastic conditions of a phenomenon; for example, not in the actual mean, but in the level about which the mean is fluctuating; not the actual coexistence, but the conditions of connection, etc. since it is important to know the mean square error of the statistical summary characteristics and in particular that of the correlation coefficient. For the case of *unconnected* series x_1, x_2, \dots when any x_i (or y_i) and any other x_j (or y_j) are independent in the stochastic sense; when in addition any x_i is only connected with y_i , but not with $y_j, j \neq i$, – then the mean square error of the correlation coefficient is expressed by the known Pearson formula

$$\sigma_r = \frac{1-\rho^2}{\sqrt{n}}. \quad (6)$$

For a more or less considerable n we may obtain its approximate estimate when replacing ρ by the empirical magnitude r (or by any other hypothetical value, for example 0) if wishing to find out how likely is it that ρ differs from 0. It is important to note, first, that formula (6) only represents the first term of some expression and that therefore it is only suitable for a sufficiently large n (say ≥ 30 if ρ is not large and $n \geq 100$ if ρ , although large, is not very near to 1).

And, second, that formula supposes that x and y obey the so-called normal law; third, that, as stated above, the series x_i and y_i are internally unconnected and only connected by terms with identical indices⁴. Considering point 2, it is possible to remark that experience shows its practical harmlessness since in an overwhelming number of cases the deviation from the *normal* law is not large enough for the formula (6) to become useless. With respect to the second assumption we ought to say that for small values of n or in dubious cases that formula should be replaced by a method due to Fisher, see below. *As to the third condition, it is necessary to stress that its violation does not at all exclude the use of the Pearson formula (6) when n is large, or the Fisher method otherwise.*

2. On the Mean Square Error of the Correlation Coefficient for Homogeneous Connected Series

I have deduced the approximate formula in a previous paper (1927). It is simplified under the assumption of normal correlation but even then it still remains considerably more complicated than the Pearson formula. Indeed, instead of one magnitude ρ_{xy} , it involves three functions, $\rho_x(t)$, $\rho_y(t)$, and $\rho_{xy}(t)$ which express the values of the coefficient of correlation between x_i and x_{i+t} , y_i and y_{i+t} and x_i and y_{i+t} . I will only provide my most important results.

For normal correlation we have

$$\sigma^2 r_{xy} = (1/n) \sum_{t=-n+1}^{n-1} \{\rho_x(t)\rho_y(t) + \rho_{xy}(t)\rho_{xy}(-t) - 2\rho_{xy}(0)[\rho_x(t) + \rho_y(t)]\} + \quad (7)$$

$$(1/2n)\rho_{xy}^2(0) \sum_{t=-n+1}^{n-1} \{\rho_x^2(t) + \rho_y^2(t) + 2\rho_{xy}^2(t)\} + \text{ terms of order higher than } 1/n.$$

For terms of one and the same series separated by distance τ that formula becomes

$$\begin{aligned} \sigma^2 r_{xy}(\tau) = (1/n) \sum_{t=-n+1}^{n-1} \{\rho_x^2(t) + \rho_x(\tau+t)\rho_x(\tau-t) - 4\rho_x(\tau)\rho_x(t)\rho_x(\tau+t)\} + \\ (1/n)\rho_x^2(\tau) \sum_{t=-n+1}^{n-1} \{\rho_x^2(t) + \rho_x^2(\tau+t)\} + \dots \end{aligned} \quad (8)$$

Now, if the terms of (3) are not at all connected with each other, i. e., if no x_i depends on any y_j , the correlation coefficient $\rho_{xy} = 0$ and the empirical coefficient of correlation r_{xy} will fluctuate about 0 with mean square error⁵

$$\sigma^2 r_{xy} = (1/n) \sum_{t=-n+1}^{n-1} \rho_x(t)\rho_y(t) \quad (9)$$

where the limits of summation can be replaced by $-\omega$ and ω if at $t > \omega$ at least one of the functions $\rho_x(t)$ or $\rho_y(t)$ vanishes. It can also be shown that if $\rho_x(t) = 0$ at $t > \omega$, then, for $\tau > 2\omega$ (for $\tau > \omega$ if the correlation is normal) the mean square error of $r_x(t)$ is expressed by a simple formula

$$\sigma^2 r_x(\tau) = (1/n) \sum_{t=-\omega}^{\omega} \rho^2(t). \quad (10)$$

For applying formulas of the type indicated it is obviously not sufficient to find a number of empirical correlation coefficients $r_x(1), r_x(2), \dots, r_y(1), r_y(2), \dots$ since we do not yet know where each of those series should be terminated and instead of the vanishing terms $\rho_x(\omega + 1), \rho_x(\omega + 2), \dots$ included in the formula we will have $r_x(\omega + 1), r_x(\omega + 2), \dots$ generally not equal to 0 and the result will be corrupted.

The method proposed in the cited work consists in that the given series are partitioned into intervals having n_1, n_2, \dots terms. We will obtain m_1 , say, correlation coefficients calculated from n_1 pairs of magnitudes, m_2 coefficients from n_2 pairs etc. Suppose the magnitudes in each pair will be x_i and $x_{i+\tau}$, then, for a sufficiently large τ , it can be admitted from general considerations about the appropriate phenomena that the correlation coefficients $\rho_x(\tau), \rho_x(\tau + 1), \dots, \rho_x(2\tau), \dots$ should vanish or be near zero.

Then, calculating the empirical mean square magnitudes s_1^2, s_2^2, \dots by issuing from the correlation coefficients of the first, the second, ... series, we should obviously obtain by formula(10) the approximate equalities

$$n_1 s_1^2 \approx n_2 s_2^2 \approx \dots \approx \sum_{t=-\omega}^{\omega} \rho_x^2(t) = 1 + 2 \sum_{t=1}^{\omega} \rho_x^2(t). \quad (11)$$

It will not be difficult now to derive the approximate value of ω by summing consecutively the squares of the empirical correlation coefficients⁶ and ω_y and ω_{xy} are calculated in a similar way⁷.

Examples easily show that in the sense interesting to us an essential difference can exist between unconnected and connected series. Suppose that the random variable x_i is obtained by taking m moving sums of two terms at a time of some unconnected homogeneous series. Then (Slutsky 1927, p. 54) it is not difficult to determine the correlation coefficient

$$\rho_x(t) = C_{2m}^{m+t} : C_{2m}^m \quad (13)$$

and for $t > 2m$ according to formula (10) the square error of the empirical correlation coefficient $r_x(t)$ will be

$$\sigma_r^2 = \sum_{i=0}^{2m} [C_{2m}^i]^2 : n [C_{2m}^m]^2. \quad (14)$$

Applying the Stirling formula, it is not difficult to find an approximate expression

$$\sigma_r = \frac{\sqrt[4]{\pi m / 2}}{\sqrt{n}} \quad (15)$$

instead of $1/\sqrt{n}$ as it turns out by the Pearson formula at $\rho = 0$. For example, if $m = 12$, $\sigma_r = 2.08/\sqrt{n}$, twice larger than provided by the usual formula. Or, σ_r is the same as for a four times shorter unconnected series.

3. The Issue of the Distribution of Errors of the Empirical Correlation Coefficient

The above sufficiently clearly describes the case of the correlation coefficient for connected series. According to the very essence of the phenomenon no simple formula such as the Pearson expression is here thinkable. A further development of the problem will perhaps lead to some simplification of the pattern of calculation, but in any case at present we have to agree that the determination of the mean square error should be eliminated from usual practice, and can only be included as a special investigation when having sufficient data and prompted by important incentives.

I think that such cases will inevitably occur in most various applications of statistics to geophysics; indeed, a critical appreciation of results obtained is, and will continue to be a vital demand without whose implementation scientific work cannot be imagined. And practical applications under the immense and [still] ever increasing scale of socialist construction and planning will in turn raise the issue of reliability of the estimates and the practical boundaries of their suitability. In spite of the difficulties

encountered, the problem of the error of the correlation coefficient in connected series cannot be therefore abandoned.

If so, we have to recall that the calculation of the mean square error does not in general solve the issue of *estimating* the possible error. Indeed, when is the knowledge of the square error sufficient? Then, and only then, when the errors are distributed according to the normal (Gauss – Laplace) law, so that it is possible to find out how probable is any deviation exceeding the square error⁸ by such-and-such a factor just by looking at a table of the integral of probabilities.

If the number of cases made use of when calculating the empirical correlation coefficient was large, and if that coefficient was not too near to its extreme values – 1 or 1, then it is possible to assume that the distribution of errors was more or less normal, but otherwise such an assumption will lead to considerable mistakes, to blunders. It is easy to understand why. The normal curve of probabilities extends to infinity in both directions whereas the correlation coefficient cannot be less than – 1 or exceed 1.

If the square error σ_r is not very small, then, and especially when $|r|$ is near to 1, the distribution of errors becomes sharply asymmetric. This circumstance is especially felt when the number of observations is small, and we must remember that this is indeed what we mostly have when dealing with connected series. And even if our series are comparatively long, owing to their connection the error of the correlation coefficient can be of the same order as in the case of unconnected series sometimes several times shorter.

Let us recall Fisher's solution of the problem given a small number of observations (for unconnected series whose terms themselves obey the normal law). Fisher introduces a supplementary variable denoting it by z for the empirical, and by ζ for the theoretical correlation coefficient, such that

$$r = \tanh z = \frac{e^{2z} - 1}{e^{2z} + 1}, \quad z = \operatorname{arctanh} r = \frac{1}{2} \ln \frac{1+r}{1-r}. \quad (16; 17)$$

Now, z is distributed very close to the normal law even for a so small number of observations as $n = 9$ or 10 , and its mean square error is expressed extremely simply:

$$\sigma_z = \frac{1}{\sqrt{n-3}}. \quad (18)$$

When wishing to find out whether r essentially differs from 0, and choose as a practical boundary 2σ or 3σ , it is only necessary to establish the corresponding r as provided by the appropriate table. Suppose we derived $r = 0.90$ and need to estimate the likely boundaries within which the true value of the correlation coefficient ρ is supposed to be, for example to find out the $3\sigma_z$ boundaries (and the corresponding quite definite probability). We find 1) The value of z corresponding to the given r ; 2) Then σ_z ; 3) And then $z_1 = z - 3\sigma_z$ and $z_2 = z + 3\sigma_z$; 4) Finally, the magnitudes r_1 and r_2 corresponding to those z_1 and z_2 . More precisely, the magnitudes ρ_1 and ρ_2 the deviation from which in either direction the given z could have been

provided with a definite probability. The problem is thus completely solved⁹.

4. The Hypothetical Law of Distribution of the Errors of the Empirical Correlation Coefficient for Connected Series

It seems that Fisher's solution of the problem for unconnected series becomes very complicated if the series are supposed not to be unconnected. I at least was still unable to overcome the difficulties. But it is possible to recall that experiment proved to be essential in the history of the problem of errors for a small number of observations.

Time and time again it indicated beforehand the correct or nearly correct solution, then corroborated [how?]. I have collected not quite a small amount of empirical material that can be used for solving the problem of the error of the correlation coefficient and I therefore allow myself to formulate here a hypothesis corroborated by it. Namely, there are grounds for thinking that also *in the case of connected series the magnitude z that I calculated by formula (17) is normally distributed if the corresponding random variables (3) themselves obey the normal law*¹⁰. And the mean square error can be determined as a function of the mean square error of r and the theoretical value of the correlation coefficient ρ ; in practice, we replace ρ by r if we do not have to test other hypotheses. And the centre of the distribution of z is some magnitude generally differing from ζ but more or less rapidly approaching it with the increase of the number of terms of the series so that this circumstance is apparently not practically important.

If that hypothesis is valid, and I think that it is possible to admit it as an approach to the truth, z is calculated by formula (17) just as in the case of unconnected series after σ_r is found by means of the indicated methods. Then σ_z is determined by the formula

$$\sigma_z = \frac{q\sigma_r}{1-\rho^2} \quad (19)$$

where q is a function of ρ and σ_r , see formulas (20) and (21) below. In very many cases q is near 1 and in general it can be obtained with practically sufficient precision from an applied diagram (Fig. 1)¹¹. Otherwise, the Fisher method remains valid.

5. Testing a Hypothesis

Ia [Model I']. I borrow the data for model I from my previous work (1929). The model consists of two independent from each other connected series x_i and y_i each of which is obtained by moving sums, ten terms at a time, of a supplementary series whose terms take the values 0, 1, 2, ..., 9 with equal probabilities. It is easy to determine that each such series will have correlation coefficients

$$r_x(1) = 0.9, r_x(2) = 0.8, \dots, r_x(9) = 0.1, r_x(10) = r_x(11) = 0, \text{ etc.}$$

and the distribution of probabilities will be characterized by the Pearson coefficients $\beta_1 = 0$ and $\beta_2 = 2.878$, not very remote from $\beta_2 = 3$ as in the case of the normal law.

We had 1050 pairs of $(x_i; y_i)$; by separating their series into intervals containing 10 pairs, it was possible to obtain 105 correlation coefficients, see Slutsky (1929, pp. 99 – 100). Table 1 indicates the values of the function z taken as the boundaries of the intervals (column 1). In the next columns, are the corresponding values of r ; the actual number of cases n'_i in the appropriate interval if the absolute values of the correlation coefficient are issued from; the theoretical number of cases n_i for a suitable normal curve; and, finally magnitudes $(n'_i - n_i)^2/n_i$ needed for calculating the criterion χ^2 which occurred to be 6.68. For the number of groups $n' = 10$ the probability of the same and less probable random deviations is $P = 0.73$. The concordance between the distribution of z and the normal law of distribution is quite satisfactory. The distribution itself is shown on Fig. 2 where the distribution of r is also provided. This latter, as the reader will see, sharply differs from [has nothing in common with] the normal law.

Ib. We separate the numbers in the same model in intervals containing 20 pairs partly overlapping each other so that we obtain a new series of 207 correlation coefficients. Since the theoretical distribution is strictly symmetric (because the true correlation coefficient $\rho = 0$), we again restrict our investigation to the distribution of the absolute values of z , and, correspondingly, of r . See the values of r for these groups of 20 and the data for the next section ($n = 30$ etc) in Slutsky (1929, p. 100). We obtain Table 2, and the concordance is again better than satisfactory, see also Fig. 3A.

Ic. To exhaust the material of the discussed model, we ought to consider all the other groupings. We have (Slutsky 1929, p. 100) 69 correlation coefficients for groups of 30; 51, 41, 34, 29, 25, 22 and 20 for groups of 40, 50, ..., 100. These groups are comparatively scanty, and we will apply an appropriate artificial trick. For each of these 291 coefficients we calculate the corresponding z (we round off both r and z to 2 decimal points); determine the boundaries of each group according to the pattern $0 - 0.3s_z$; $0.3s_z - 0.6s_z$; etc, count the number of the coefficients in each interval and compile Table 3, see also Fig. 3B.

Once more we find that the concordance is quite satisfactory. Thus we conclude our investigation of the discussed series. It is, however, worthwhile to mention in addition that the square error of the correlation coefficient, since the two series are independent from each other so that their correlation functions coincide, is calculated according to formula (10) and is equal to

$$\sigma_r = \sqrt{6.70/n} + \dots \approx 1/\sqrt{n/6.70} \approx 2.59/\sqrt{n}.$$

This is more than 2.5 times the error for unconnected series for which $\rho = 0$ with the same n : $\sigma_r = 1/\sqrt{n}$. We also see that, with regard to the fluctuations of the correlation coefficient, the investigated connected series are equivalent to two unconnected 6.7 times shorter. For example, series of 40 pairs of $(x_i; y_i)$ are equivalent to the case of two unconnected series of 6 terms. This fact very expressively stresses the sharp peculiarity of connected series and the impossibility of instinctively transferring to them the rules borrowed from another statistical region.

One more remark. Formula (10) is only approximately valid since it omits terms of the type B/n^2 , C/n^3 , ... [see Note 6]. We (1929) have nevertheless shown that for the discussed model it provided practically quite suitable

results even at $n = 40$. This conclusion should be considered quite favourable because $n = 40$, as we just saw, corresponds to $n = 6$ for unconnected series, i. e., to an extremely small number of observations. When $n < 40$, we have to determine σ_r by another method briefly indicated in § 2.

6. Testing a Hypothesis (Continued)

II [second model]. In this, our second example, we will consider the distribution of the coefficient of correlation between terms of an empirical series considerably remote one from another, – the series of indices of prices of wheat in Western Europe for the period 1500 – 1869 compiled by Beveridge (1921, p. 429; 1922, p. 412) who studied them in detail and attempted to discover periodicities there.

If we discard the beginning of the series based on too scanty data and the final terms where the homogeneity of the series was corrupted, as the author himself (1921, p. 432) remarked, by the influence of industrial cycles, there will still remain a rather considerable amount of material covering ca. 1550 – 1800.

Beveridge analyzed the interval 1545 – 1844, and I had to follow him for ensuring comparable results. I will devote a special paper to that investigation; here, I only study the fluctuations of the correlation coefficient and I am considering the entire series without discarding its beginning or end which is admissible because in any case the tested hypothesis does not thus become more favoured.

We investigate the correlation coefficients for distances 20 – 84 between the terms and groups of $n = 20, 30, \dots, 120$. As I will justify elsewhere, correlation connection hardly extends for more than 10 terms so that the values of the correlation coefficient for all the distances studied can be considered to be very likely either equal to zero or some small practically vanishing number. For each group of $n = 20, 30, \dots$ taken separately I calculated the values of the empirical mean square deviation s_z of z and the distribution of the latter over the intervals expressed in fractions of s_z [and the distribution of z/s_z] just as it was done in the last of the examples above.

The results for all the groups are summarized in Table 4 and we see that, judging by the χ^2 criterion, they should once more be considered quite satisfactory, see also Fig. 4B. I can additionally note that this example belongs to those which corroborate my statement that z can be near-normally distributed even when the correlated magnitudes themselves rather considerably deviate from normality. Thus, for the terms of the Beveridge series the Pearson coefficients, which are criteria of normality ($\beta_1 = 0, \beta_2 = 3$), are $\beta_1 = 1.2767$ and $\beta_2 = 4.8613$. The distribution itself is shown on the same Fig. 3A and, as we see, it really differs from the normal law.

III [third Model]. As our third example we study the connected random series which I will call a model of the Beveridge series because it was compiled as its imitation. Detailed information about this model will be contained in the paper devoted to the Beveridge series; here, it is sufficient to indicate that its first four moments almost exactly, and the next two moments in somewhat crude terms coincide with the respective moments of the real Beveridge series.

In addition, the correlation coefficients of the model almost exactly coincide with those hypothetically assumed for the Beveridge series;

according to some considerations on which I am unable to dwell here, they somewhat differ from their empirical values because of a certain smoothing procedure.

The series under discussion has 1251 terms. Multiplying x_1 by x_2 , x_3 by x_4 etc. we obtain 1250 products; separating them in groups of five gets us 250 coefficients $r_x(1)$; and, forming groups of 10, – 125 such coefficients. Then, groups of 20 partly overlapping each other, 124 coefficients. It is these partitions that we will consider, but first of all I note that this case provides a curious illustration of what is called the systematic error of the correlation coefficient. In general, even for the case of unconnected series, the expectation of the empirical correlation coefficient Er_{xy} is not equal to the theoretical coefficient ρ_{xy} . The empirical coefficient r_{xy} for very long homogeneous series should therefore approach its stochastic limit ρ_{xy} as close as desired. If, however, the series is separated into small intervals of n terms each, say, and the coefficient r_{xy} is calculated for each of them, their arithmetic mean will not coincide with ρ_{xy} however large is the number of those intervals, but will have its own limit Er_{xy} , a function of n . For unconnected series the difference between ρ_{xy} and Er_{xy} is represented by a magnitude of the order of $1/n$ and may be almost always ignored because generally the mean square error of the correlation coefficient considerably exceeds it having order $1/\sqrt{n}$ (Chuprov (1925/1926, pp. 91, 105 – 106; Slutsky 1923).

In our case, the stochastic pattern of the series is such that $\rho_x(1) = 0.458$ whereas the mean value of 250 correlation coefficients each calculated from 5 pairs of numbers $(x_i; x_{i+1}), (x_{i+1}; x_{i+2}), \dots, (x_{i+4}; x_{i+5})$, almost vanishes (= 0.091). The mean of 125 coefficients for $n = 10$ is 0.313, and only for $n = 20$ it reaches 0.405 (Fig. 5). It is thus *easily seen how large can be the errors of the correlation coefficients calculated for very small numbers of observation*, as recommended, for example, by Schmauss (Tikhomirov 1930, p. 288)¹².

Returning now to our main problem, we consider the next table (Table 5). Numbers n_i are calculated for the Gaussian curve corresponding to the arithmetic mean $z = 0.1028$ and $s_z = 0.54836$ and obtained from the given distribution with the Sheppard corrections being applied. The extreme groups of the n_i column, as it is usually done when applying this method, were combined to avoid numbers less than 1. The reader will see that the concordance is quite satisfactory.

Now, however, let us attempt to interpret these just mentioned low probabilities [?] constituting the weak link of the method discussed. A simple calculation for 11 and 12 groups provides the theoretical number of cases 0.24 and 0.76 for 1 and 2 groups. Actually, we obtained 1 and 3 respectively and we estimate the appropriate probability according to the Poisson formula

$$P = \frac{e^{-\alpha} \alpha^m}{m!}$$

where α is the expected number of cases for a given number of trials np , and m is their actual number. Assuming at first $\alpha = 0.24$ and $m = 0$, we have $P = e^{-0.24} = 0.787$.

The occurrence of one or more such cases has probability 0.213 and should therefore happen rather often. Taking now $\alpha = 0.76$, $m = 0, 1$ and 2 , we find

$$P_0 = 0.468; P_1 = 0.355; P_2 = 0.135; 1 - (P_0 + P_1 + P_2) = 0.042.$$

The probability of the occurrence in the considered interval of three or more cases is 0.042 and, under the hypothesis of normality can generally happen approximately once in 25 instances. This is not yet sufficiently rare for rejecting the hypothesis. At most, it empowers us to assume that the true law of distribution of the values of z perhaps *somewhat* differs from the normal law. However, the tested hypothesis is indeed such an assumption so that our observations do not contradict it.

The conclusion is worse with regard to a circumstance not yet discussed although the reader could have noticed that we mentioned above 250 correlation coefficients whereas the Table only contains 249. We have rejected one of them because it essentially damaged the entire picture. This is the case of $r_x(1) = 0.996$ and the corresponding $z = 3.1063$ is separated from its mean 0.1028 by 5.48 of its mean square deviation which is extremely unlikely for a normal distribution.

It seems, however, that this single exception cannot lead to the rejection of our hypothesis. *It is very likely that with such a small number of terms ($n = 5$) a deviation from normality should be greater than when the correlation coefficient is calculated for any reasonable number of observations.* It is rather more surprising that even in such a case the main mass of data had obeyed the normal law to such an extent. Moreover, it is also possible that we were *unlucky*, that we had encountered a comparatively rare case although not as rare as it would have been for the normal law.

If our idea stressed above is correct, we ought to obtain a better result when studying the correlation coefficient for the same series at $n = 10$ and 20 . My idea is indeed corroborated by considering Tables 6 and 7 and Fig. 5B, C.

7. On the Application of the χ^2 Criterion To Connected Series. Testing a Hypothesis by the Example of Model V

Without dwelling on the theory of the χ^2 criterion it is necessary to indicate that we have applied it above in the only possible form but that it is not quite suitable for the conditions of our problem. One of the assumptions of its theory is *independence* of trials from which the separate elements of the studied totality were derived. *Strictly speaking, without being additionally adapted, the criterion is therefore not applicable to totalities whose elements form connected sequences.*

Here is an example. I examined a series of absolutely unconnected zeros and unities to which, because of the stochastic pattern of their occurrence, we had to assign equal probabilities¹³. Call these numbers ξ_1, ξ_2, \dots . I applied them to form a new series (Model B) by moving summations

$$x_i = \xi_i + \xi_{i+1} + \xi_{i+2} + \xi_{i+3}$$

and obtained 1280 terms. It is easy to find out that the new series is connected with a correlation function

$$r_x(0) = 1, r_x(1) = 3/4, r_x(2) = 1/2, r_x(3) = 1/4, r_x(t) = 0 \text{ at } t \geq 4$$

and that its terms take values

0, 1, 2, 3, 4 with probabilities 1/16, 4/16, 6/16, 4/16, 1/16.

Then I compiled 1 distribution of all the 1280 terms, 2 distributions of 640 terms, 4, 5, 10 and 20 distributions of 320, 256, 128 and 64 terms. I am only providing a table (Table 8) of the values of χ^2 obtained by comparing the empirical numbers with theoretical equal to 1/16, 4/16, ... of the respective numbers of terms.

It is seen at once that the results are bad. According to the appropriate table of probabilities we find¹⁴ that in the column for the number of groups $n' = 5$ the probability of random deviations leading to $\chi^2 \geq 7$ is 0.135888, or approximately 1/7. So, certainly for independent trials, out of 42 cases we may expect to have 6 such for which $\chi^2 \geq 7$ whereas actually the Table shows us 19! Then, $\chi^2 = 11$ corresponds to probability 0.026564 or ca. 1/40. But, instead of one, the Table contains 7 such cases. It is worthwhile to show the probabilities of the most considerable values of χ^2 (rounded off):

$$P = 14.08; 16.81; 18.32; 23.29; 41.87; 52.89$$

$$\chi^2 = 1/10^2; 2/10^3; 1/10^3; 1/10^4; 1/10^8; 1/10^{10}$$

It is absolutely doubtless that either the hypothesis on the distribution of the terms of Model B does not conform to reality, or that the theory of the χ^2 is not applicable here. Actually, it is the second alternative that is true. Yes, the numbers 0, 1, 2, 3, 4 must occur in 1/16, 4/16, ... of all the cases if *those are independent one from another*. This will happen if, for example, we select each fourth term of the model, and the same will be true for each of the 4 such partial series. It is therefore not difficult to conclude by simple stochastic considerations that *the same law also persists for a connected series* at least if the connection between the terms indefinitely weakens with the distance. And the difference between both cases consists in that *equal deviations* from 1/16, 4/16, ... will have there differing probabilities. How should this influence χ^2 ? Let us write its formula in the following way:

$$\chi^2 = \sum_{i=1}^{n'} \frac{(n'_i - n_i)^2}{n_i} = N \sum_{i=1}^{n'} \frac{[n'_i / N - (n_i / N)]^2}{(n_i / N)}$$

Here, N is the number of elements in the totality, n' , the number of groups, n'_i , the actual, and n_i , the theoretical number of terms in each group. The expression above shows that if different totalities have *one and the same relative distribution*, the value of χ^2 will be proportional to the total number of the elements in the totality.

Suppose now that we have series x_1, x_2, \dots, x_N , and y_1, y_2, \dots, y_N , both unconnected but such that each x_i is connected with y_i with the correlation coefficient almost equal to 1. Then the actual distribution, in absolute and respective numbers, will be almost the same for both series; therefore, when combining them to form a single totality, χ^2 will be almost twice greater. If the inadmissibility of that procedure is not noticed, even a quite likely

deviation can, on the contrary, be considered very unlikely when judging by the value of χ^2 .

The correlational connection between members of connected series should produce a similar effect; moreover, it seems that an exaggeration of the value of χ^2 must occur in practice much oftener than its underestimation. I think that the following example can be explained exactly in this way.

Example IV. This is Model V (Slutsky 1929, pp. 72 – 73 and Table 6 on p. 100) consisting of two connected series x_1, x_2, \dots and y_1, y_2, \dots of 1000 terms each. The correlation coefficient $\rho_{xy}(0) = 0.814$. We have 40 correlation coefficients calculated for $n = 25$; 20 coefficients for $n = 50$, and, for intervals partly overlapping each other, 18, 16, 14 and 12 coefficients for $n = 75$ and 100; 125 and 150; 175 and 200; 225 and 250.

We see at once that this example is barely suited for testing a hypothesis. First, the intervals are too wide for a sufficient number of non- or more or less weakly correlated coefficients to be obtained out of the total 1000 terms. Second, the number of not overlapping intervals is too small for considering them only, and, beginning with $n = 125$, the other intervals overlap each other more than by a half of their length. Then, a high correlation should exist between coefficients of various respective values of n close to each other.

Similar objections could have been partly formulated to all the previous examples, and notably to some of them, but the present example seems to be especially unfavourable. The stochastic structure of the respective series doubtlessly influences the matter as well, but I was compelled to consider also this example since the correlation coefficients here discussed are provided in my previous work and the reader himself can attempt to apply it for testing the formulated hypothesis.

Table 9 is compiled after the previous summary tables which means that at first z was found for each r , then the mean z and mean square magnitudes s_z were calculated for each group of coefficients and after that the number of cases in the intervals was counted in the terms of its mean square values.

The conformity of the given data with the normal law is doubtlessly very bad, and the discussed example should have testified against our hypothesis had the considerations provided above not discredited the χ^2 criterion. Perhaps that example should not have been provided at all, but it confirms our reasoning about χ^2 by the nature of the actual distribution as shown in Table 9¹⁵.

Let us consider the numbers in the column n'_i of that table and the corresponding picture on Fig. 6. Suppose that the Gaussian distribution does not suit that material and the distribution of z obeys quite another law. However, none of what we know about laws of distribution and no similarities out of those we were able to apply allow us to think that that unknown law is reflected in the numbers n'_i in a likely way. These numbers fluctuate too sharply; in neighbouring intervals doubtlessly situated near the middle of the distribution they experience such leaps as 21, 19, 10, 31, 10, 17 etc.

Whichever is the law of distribution of z , it is likely to be smooth here also whereas the leaps just mentioned correspond, as far as it is possible to judge without exact calculations, not to the total 142 cases, but to some other number, at least twice or even thrice smaller. This is a doubtless indication that the mutual correlation of the elements of the totality led here

to an exaggeration of the value of χ^2 . The example may be abandoned since it does not testify either for, or against our hypothesis. However, we reject such a way out of a tricky situation as insufficiently convincing, and, in the next section, we will check the applicability of our hypothesis to Model V with all rigour.

8. Checking the Hypothesis on Independent Correlation Coefficients of Model V

For checking the previous considerations it would be necessary to extend the series of Model V to such an extent that it will be possible to choose a sufficient number of intervals separated by spaces long enough for the terms of one interval not to correlate with the terms of another one. Then the correlation coefficients obtained for such independent intervals will also be independent from each other, the application of the χ^2 criterion will become legitimate and we will see whether our hypothesis is satisfied here.

However, such a plan would have demanded too much work since Model V is compiled in a rather complicated way. Without dwelling on the pertinent details, see Slutsky (1929), I will only describe how that difficulty was overcome. An essential part of the pattern of Model V consisted in obtaining the numbers of some connected series by calculating moving sums of some unconnected series multiplied by certain *weights* and subsequent rounding off. Now, we adopt another pattern by *rounding off* the weights themselves. A model was thus obtained whose numbers differed here and there by a unity from those of Model V. We consider it as a very close analogue of that latter, denote it Model V' and apply it for checking the studied hypothesis.

Model V' was based on unconnected series of numbers ξ_i and η_i . Both (at first the ξ_i , then the η_i) were chosen from the vertical columns of Tippet's table of random numbers [1927] beginning with its first page. Each column constituted a separate series unconnected during treatment with other series. We obtained ξ_i by replacing the even numbers of the Table by 0 and odd numbers by 1 and the values of ξ_i were therefore 0 and 1 with equal probabilities. For η_i we left numbers 0, 1, 2, 3, 4 intact and replaced numbers 6, 7, 8, 9 by -4, -3, -2, -1 and we either did not change number 5 or replaced it by -5 if, respectively, an even or an odd number occurred in the same row and neighbouring column. The probabilities of 5 and -5 were therefore 1/20 with 1/10 being the probability of all the other numbers. Then we determined

$$u_i = \sum_{k=0}^{10} \psi_k \xi_{i-k}, \quad \psi_k = -1, -2, -4, -6, -4, 0, 4, 6, 4, 2, 1$$

and finally the numbers of Model V' were

$$x_i = u_i + \eta_i, \quad y_i = u_i + \eta_{i+1}.$$

The coefficient of correlation between x and y is

$$r_{xy}(0) = \frac{E x_i y_i}{\sigma_x \sigma_y} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\eta^2} = 0.816$$

which is very close to 0.814 as in Model V.

We have thus compiled 280 absolutely independent one from another series of numbers x , y , each consisting of 20 such pairs. The 280 correlation coefficients were also therefore absolutely independent. Table 10 compiled after the previous tables provided the distribution of the corresponding values of z . The mean $z = 1.0864$ and the appropriate $r = 0.796$ was very near to the true value (0.816) of the correlation coefficient. The agreement with the normal distribution is very good, see also Fig. 7, if judged by the criterion χ^2 whose application is here methodologically impeccable.

The example just considered refutes the sole unfavourable result (see § 7) and we may consider that our hypothesis had stood all the trials to which I was able to subject it.

9. Derivation of the Formulas

Necessary for Applying the Suggested Hypothesis

I will represent the formulated hypothesis (§ 4) in a somewhat simplified way by assuming that the number of the elements of the totality, from which the empirical correlation coefficient is derived, is so great that the systematic error of z may be neglected.

I only reproduce Slutsky's final results derived after a very long discussion involving hyperbolic functions which he assumes to be known. For the coefficient q in formula (19) Slutsky gets

$$q = \sqrt{1 + \alpha_1 \left(\frac{\sigma_r}{1-\rho^2}\right)^2 + \alpha_2 \left(\frac{\sigma_r}{1-\rho^2}\right)^4 + \dots}, \quad (20)$$

$$\begin{aligned} \alpha_1 &= 2(1 - 4\rho^2), \quad \alpha_2 = 21/3 - 8\rho^2 + 59\rho^4, \quad \alpha_3 = 4 - 56\rho^2 - 12\rho^4 - 496\rho^6, \\ \alpha_4 &= 1/5 - 160\rho^2 + 309\rho^4 + 1888\rho^6 + 4094\rho^8, \quad (21) \\ \alpha_5 &= 242/3 - 1472\rho^2 + 3056\rho^4 - 185862/3\rho^6 - 23518\rho^8 - 38352\rho^{10}. \end{aligned}$$

Slutsky also provides a table of α_i for $\rho = 0(0.05)0.95$. He continues:

By means of this table I calculated a series of values of q and constructed a graph (Fig. 1). Its lines are terminated at points in which, owing to the properties of semi-convergent series, the error of q became approximately equal to 0.01. The error of the graph apparently never exceeds that magnitude.

Explanation of Tables and Figures

Fig. 1, § 5. Model I'. Diagram of coefficient q from formula (19).

Table 1, § 5. Model I'. Explanation in text.

Fig. 2A, 2B § 5. Model I'. Distribution of $|z|$ and $|r|$ for $n = 10$.

Table 2, § 5. Model I'. According to explanation in text, the first columns indicate $|z|$ and $|r|$, but in the Table itself these columns are labelled (apparently wrongly) z and r .

Fig. 3A, 3B, § 5. Model I'. Distribution of $|z|$ for $n = 20$ and of $|z|/s_z$ for $n = 30, 40, \dots, 100$.

Table 3, § 5. Model I'. Shows the same magnitudes as Table 2 although $|z|$ is replaced here by $|z|/s_z$. Then, according to the text, the isolated groups

are $0 - 0.3s_z$; $0.3s_z - 0.6s_z$ etc whereas the Table apparently shows $0 - 0.3|z|/s_z, \dots$

Fig. 4A, 4B, § 6. The Beveridge series. Fig. 4A shows the distribution of its terms, explanation insufficient. Fig. 4B shows distribution of z/s_z for correlation between its terms separated by long distances, $n = 20, 30, \dots, 120$.

Table 4, § 6. The Beveridge series. Shows the same magnitudes as Table 3.

Fig. 5A, 5B, 5C, § 6. The Beveridge series. Shows the distribution of its terms for various values of n . Legend contains obvious mistake and is difficult to understand.

Tables 5, 6, 7, § 6. The Beveridge series. Shows the distribution of same magnitudes as previous tables for $n = 5, 10$ and 20 , but one of those magnitudes is z rather than $|z|$ or $|z|/s_z$.

Table 8, § 7. Model B. Explanation in text but difficult to understand, see Note 14.

Fig. 6, § 7. Model V. Shows z/s_z for various values of n .

Table 9, § 7. Model V. Explanation in text. Necessary addition: one of the magnitude shown is $(z - \bar{z})/s_z$.

Fig. 7, § 8. Model V. Shows distribution of z for $n = 20$.

Table 10, § 8. Model V'. Compiled "after the previous tables" with one of the shown magnitudes being z , but the corresponding column in the table itself is not labelled at all.

Notes

1. The author wishes to convey his deep gratitude to Mrs. Helene Iliinskaia-Pomerantseva for her valuable help on all stages of the elaboration of this paper. E. S.

That Note was attached to the summary. A similar Note in main text adds:

I am sincerely thankful to N. V. Korotneva for her help with calculations. E. S.

2. Now, we simply say *covariance*. O. S.

3. In my previous work (1930), I denoted the empirical correlation coefficient by ρ , and the empirical mean square deviation by σ^1 . Here, I follow Fisher's notation which is likely to become widely accepted. See for example Fisher (1925/1930). E. S.

Slutsky misspelled Fisher's name; see similar mistake mentioned in Note 9. O. S.

4. It was possible to state as a fourth assumption the invariable behaviour of probability and constancy of connections. I do not discuss this issue since I hope to return to it in one of my next contributions. E. S.

5. The mean square error is σ rather than σ^2 . O. S.

6. Here, I can only mention the main idea of the applied method, actually somewhat more complicated. We have to bear in mind the more general formula

$$\sigma_r^2 = \frac{A}{n} + \frac{B}{n^2} + \frac{C}{n^3} + \dots$$

and calculate

$$A = 1 + 2 \sum_{t=1}^{\infty} \rho_x^2(t)$$

by least squares. E. S.

7. A more direct method can be based on the expansion of the given [finite] series into a Fourier series, see my report at the 1st All-Union Congress of Mathematicians in Kharkov, 1930 [1936]. E. S.

8. Here and somewhat below: a loose expression meaning mean square error. O. S.

9. See Fisher (1921, p. 3) and Romanovsky (1928). Hayashi (1926) is the most detailed table of hyperbolic functions. E. S. Slutsky misspelled Fisher's name. O. S.

10. This qualification remark is of no practical consequence since in an apparently overwhelming majority of cases the necessary degree of approximation is present. E. S.

11. The graphs of that diagram are terminated since, as I explain below, it was necessary to apply semi-convergent series. As it seems, this should not be practically important but still regrettable. It is certainly possible to calculate further, but the necessary time and work will be more expedient to spend after my hypothesis undergoes thorough critical analysis. E. S.

12. My remark does not at all exclude the application of the correlation coefficient in cases of very small numbers of observation, but it demands distinct understanding of what this can, and cannot provide and of the relations between empirical and theoretical correlation coefficients. I hope to return to this subject in another connection. E. S.

13. That is my *main* series No. III continued by the numbers of the main series II, see Slutsky (1927, pp. 36 Note and 57 – 61). E. S.

14. The explanation is certainly inadequate and all the six alleged *probabilities* of some values of χ^2 (below) exceed 1! O. S.

15. Theoretical numbers n_i were calculated for a normal curve with centre at point 0 and assuming $\sigma = 1$. The actual mean square value of n' is very close to 1. The probability P , as also in other cases in which the scope of the tables is insufficient, was calculated by the known Pearson formulas (1900). E. S.

Bibliography

E. E. Slutsky

(1927, in Russian), The summation of random causes as a source of cyclic processes. *Voprosy Konjunktury*, vol. 3, No. 1, pp. 34 – 64, 156 – 160. Reprinted in Slutsky (1960, pp. 99 – 132). Translation: *Econometrica*, vol. 5, 1937, pp. 105 – 146.

(1929), Sur l'erreur quadratique moyenne du coefficient de corrélation dans le cas des suites des épreuves non indépendantes. *C. r. Acad. Sci. Paris*, t. 189, pp. 612 – 614. [Slutsky referred to his paper (1930), and I think it therefore admissible to include here this note as well. O. S.]

(1930, in Russian), On the mean [square] error of the correlation coefficient in the case of homogeneous coherent chance series. *Trudy Konjunktur. Inst.*, vol. 2 for 1929, pp. 64 – 101, 150 – 154.

(1936, in Russian), On connected random functions of one independent variable. *Trudy Pervogo Vsesoiuznogo S'ezda Matematikov* (Proc. 1st All-Union Congr. Mathematicians) (1930). Moscow – Leningrad, pp. 347 – 357.

(1960, in Russian), *Izbrannye Trudy* (Sel. Works). Moscow.

Other Authors

Beveridge W. H. (1921). Weather and harvest cycles. *Econ. J.*, vol. 31, pp. 429 – 452.

--- (1922), Wheat prices and rainfall in Western Europe. *J. Roy. Stat. Soc.*, vol. 85, pp. 412 – 459.

Chuprov A. A. (1925, in German), *Osnovnye Problemy Teorii Korreliatsii* (Main Issues of the Theory of Correlation). Moscow, 1926, 1960. English edition (1939): *Principles of the Mathematical Theory of Correlation*. London.

Fisher R. A. (1925), *Statistical Methods for Research Workers*, 1930. Many later editions.

Hayashi K. (1926), *Sieben- und mehrstellige Tafeln der Kreis- und Hyperbelfunktionen*. Berlin.

Hawkins D. L. (1987), Using U statistics to derive the asymptotic distribution of Fisher's Z statistic. *Amer. Statistician*, vol. 43, pp. 235 – 237.

Ondar Kh. O. (1977, in Russian), *The Correspondence between A. A. Markov and A. A. Chuprov on the Theory of Probability and Mathematical Statistics*. New York, 1981.

Pearson K. (1900), On a criterion that a given system of deviations etc. *Phil. Mag.*, vol. 50, pp. 157 – 175.

---, Editor (1914), *Tables for Statisticians and Biometricians*. London, 1923.

--- (1921), On the *probable error* of a coefficient of correlation deduced from a small sample. *Metron*, vol. 1, No. 4, pp. 3 – 32.

Romanovsky V. I. (1928, in Russian), *Elementy Teorii Korreliatsii* (Principles of the Theory of Correlation). Tashkent.

Tikhomirov E. (1930, in Russian), On the application of the method of correlation in meteorology. *Meteorologich. Vestnik*, No. 9 – 12.

Tippet L. H. C. (1927), *Random Sampling Numbers (1st series)*. *Tracts for Computers*, No. 15.

XI

On the Existence of Connection between the Solar Constant and the Temperature

K voprosu o sushchestvovanii svyazi mezdu solnechnoi postoiannoii i temperaturoi.
Zurnal Geofiziki, vol. 3, 1933, pp. 263 – 281

Summary [in its original English] On the Existence of Connection between the Solar Constant and Temperature

Abbreviation: CC = correlation coefficient
MT = max. temperature
SC = solar constant

1 – 3. The daily Montezuma values of the SC which have been used here, were obtained by the critical examination of the following data:

1) The values found by measuring the ordinates on the enlarged photocopy of the C. G. Abbot's diagram in Smiths. Misc. Coll., Publ. 3114, p. 2 – 3, covering the period 1924 – 1930;

2) Ten day SC values for the same period (l. c., p. 12);

3) The daily values of the SC published in the Daily Weather Map of the United States Weather Bureau for the period from 24 July 1927 till 31 Oct. 1931.

The errors found by the comparison of our values with the Annals of the Astro-Phys. Obs. of the Smiths. Inst. (vol. 5), which came to us when this study was rather finished, are given in Table 1. Only on one case they are to be imputed to the misreading of the Abbot's diagram, in ten cases to errors in the Daily Weather Map, in the remaining 65 cases to the errors which are to be found in the Abbot's diagram republished now without alteration in the Annals (vol. 5, p. 246). The mean frequency of the errors being less than 1:30 and their influence being found quite negligible in the one of the most doubtful cases, it is to be hoped that the results of this study cannot be substantially vitiated by the said errors.

4. As we intended to prove the existence of the correlations between the SC and the MT found by H. H. Clayton, the deviations of the ten-daily means from the thirty-daily means of the SC and the MT for Cordoba (Argentina) have been computed. Then we have found the st. d. for every three-months period of the each year and the analogous st. d. based on the data for the whole period 1924 – 1931.

A glance on the Fig. 1 tells us that these st. d. are to be considered as periodic time-functions. Having calculated 3 (resp. 2) harmonics (see the full lines, Fig. 1), we reduced them by the due factors. The momentaneous st. d. having thus been found, the original deviations were standardized by dividing them with the values proportional to these standard deviations.

5. From the series of the MT thus obtained we have chosen the partial series corresponding to the 56-th till 155-th and to the 156-th to the 255-th day of each year and we have thus correlated them with the SC values 1) for the same year and 2) for the two preceding resp. the two following years with the additional lags from 0 to 15 days (see Tables 3a & 3b). The all 16

correlational functions thus obtained for the corresponding years and a specimen containing 4 functions for the different years are shown on the Fig. 2¹.

After the second partial series had been divided in the two equal parts, the same combinations of years have been considered and for each combination the largest from the CC corresponding to the additional lags from 0 to 15 days was found. They were found thus 8 + 8 + 16 CC between SC and MT values relating to the same year and 26 + 26 + 52 CC between the values relating to the different years, each CC being the largest (as to its absolute value) from the 16 CC corresponding to 16 additional lags from 0 to 15 days which were tried for each combination. These CC are shown on the Fig. 3. Thus, it is fairly evident that there is no significant difference between the CC found for the data relating to the same year and the CC found for the data relating to the different years, whence it follows that *the true CC between the values of the SC of radiation and the MT in Cordoba must be quite negligible the correlation c. which can be empirically found being nothing else but the errors of the random sampling.*

6. Table 4 gives the values of the momentaneous st. d. of the deviations of the ten-daily from the 30-daily means of the SC calculated for the middle points of respective months. In discussing these values the author comes to the conclusion that during six months from the twelve the errors there involved constitute presumably the greater and the true value the lesser part of the values of the said deviations.

7. The distributions of the CC relating to cases when SC precedes MT and to cases when MT precedes SC (in both cases with the lags from 1 to 2 years) cannot be considered as significantly different, the value of chi-square being 16.84 and the corresponding probability $P = 0.3$. Combining both we find $\sigma_r = 0.2672$ instead of 0.1 given by the Pearson's formula (for $r = 0, n = 100$), this formula being inapplicable to connected series, i. e., to the series composed by the casual values which are not mutually independent.

Applying further the theory of the R. Fisher's function z for the connected series developed in our paper (J. Geophysics, vol. 2, No. 1(3)), we find $\sigma_z = 0.2873$ which leads to the theoretical distribution of the CC ($\chi^2 = 12.84, n' = 15, P = 0.5$). The values of z being thus normally distributed, it is possible to find, for instance, the probability of the deviation 0.65, this being the largest CC in the case of the correlation of values of the SC and MT relating to the same year. This probability being 0.007 the mathematical expectation of the number of such cases in the universe of 256 cases will be 1.8 the actual value, as a matter of fact, being only 1. The same theoretical distribution has been compared (see Table 6 and 7) with the distribution of the CC between the values of the SC and of the MT relating to the same year.

The distributions of the Table 6 being at the first sight significantly different, the author analyses the discrepancies and comes to the conclusion that there is probably no significant divergency, the discrepancies being enlarged by the correlation between the CC constituting the set of values under consideration. [See also the paper of the present author in the Journal of Geophysic vol. 2, No 1(3)]. This point of view is confirmed by the distribution of the absolute values of the CC (Table 7), and by the value of the standard deviation for the distribution of the Table 6 (0.250) being not

substantially different from the value (0.267) of the st. d. of the CC for the case of different years.

8. There were found further 4·192 CC between the SC values with the lags equal, or nearly equal, to one and to two years and $n = 40, 60, 80, 100$. The empirical st. d. of these CC are shown in Table 8 where the last column gives the theoretical values according to the formula

$$\sigma_r^2 = \frac{A}{n} + \frac{B}{n^2} + \frac{C}{n^3},$$

A, B, C being found by the method of least squares. As we know (see the paper of the author cited above) the coefficient

$$A = 1 + 2 \sum_1^{n-1} r^2(t)$$

whence it follows that

$$\sum_1^{n-1} r^2(t) = 0.5(A - 1) = 4.14.$$

As it was found (see Table 9)

$$\sum_1^{31} r^2(t) = 3.45$$

it must be admitted that the values of $r(t)$ for $t > 31$ cannot be regarded as negligible. As it has been necessary to postpone the further study of the serial correlations, the theoretical value of the st. d. between the SC and the MT which (under the supposition of the zero-correlation) is given by

$$\sigma_r^2 = \frac{1}{n} \left[1 + 2 \sum_1^{n-1} r_x(t)r_y(t) \right]$$

could not be evaluated. Nevertheless it is to be noted that the substitution of the sum

$$\sum_1^{31} r_x(t)r_y(t)$$

in the preceding formula gives us the value of $\sigma_r = 0.30$ not substantially different from the value 0.27 found above by the direct computation based on 832 CC.

As it follows from the values of the serial correlations for the SC and for the MT given in the Table 9, there is a great similarity between the serial correlations for the periods 1924 – 1927 resp. 1928 – 1931, the relatively small differences being probably of the casual provenience. This fact cannot be underestimated and deserves further studies.

1. Introductory Remarks. The Subject of Study

Abbreviation: see Summary

The solar constant is the amount of energy received [in 1 minute – not mentioned by Slutsky] from the sun by a surface perpendicular to the sun's rays, 1cm^2 in area and located outside the earth's atmosphere at the earth's mean distance from the sun.

The remarkable investigations of Abbot and his collaborators (*Annals* 1932) have apparently definitively proved that this magnitude is actually not constant but fluctuates from year to year, from month to month, and perhaps even from day to day. Not so is it with the Clayton – Abbot (Abbot 1931, p. 1) theory of weather which maintains that exactly those alterations in the intensity of the solar emanation constitute the most essential cause of all meteorological changes which in their totality compose that which is called weather.

The provided justification of that proposition does not seem convincing to us and we aim here to report about the work done for at least partly checking it. Clayton's study that went on year after year led him to conclude that each alteration in the SC produces changes of temperature, of the same sign in the equatorial and polar zones, and of the opposite sign in the temperate zones, and that first of all those alterations are reflected in the equatorial zone and in the high latitudes of the temperate zones. These perturbations move in waves towards the equator and shift eastward travelling at speeds inversely proportional to the length of their periods, and, in the tropical regions, are superimposed on the waves generated in the equatorial zone (Clayton 1923, pp. 215 – 269).

Clayton took into account a large number of stations, ensured a geographical coherence of the entire picture, and, last but not least, his separate, masterly selected illustrations are inspiring. At first, this creates an impression of reliable validity; only after having a closer look you begin to notice that the edifice of Clayton's constructions is not so robust.

First of all, it is necessary to remark that the number of stations indicating a correspondence between the course of meteorological processes and the changes in the SC cannot be especially significant. Since those processes are interconnected, such parallelism observed at one station will almost certainly be revealed in a number of other stations. It is much more important to cover the longest possible period and exactly in this respect Clayton's work leaves too much to be desired.

Clayton, to be truthful, determines CCs many times exceeding their mean square errors. Thus, at Sarmiento in Argentina after two days the CC between the SC and the temperature in winter of 1916 reached 0.82, see Clayton (1923, p. 224); on p. 269 he expressly mentions a small probable error. For 77 days of observation the CC elevenfold exceeded its mean square error.

These data would have provided a reliable guarantee had he issued from series consisting of mutually independent terms. But, when this condition is lacking, as it always does when dealing with wavy series, the usual formula for the mean square error becomes absolutely unsuitable (Slutsky 1929; 1933) and its application can lead to most deplorable blunders.

Indeed, Clayton compares series mostly representing deviations of moving decadic averages from similar monthly averages. Suppose that rhythms of about the same length occur in the series of temperatures as well

as in that of the SC. That resemblance will be strengthened by averaging and it is not difficult to imagine that when the series are suitably shifted one with respect to the other intervals of 3, 4 and 5 wavelengths will quite often provide sufficiently high CCs.

That, however, is just what Clayton does when he calculates those coefficients after each shift up to 15 days. Shoot the flight of a crow in Moscow and of another one in New York. Measure the ascent of the wing on each film and calculate the CC. If your series are not too long, after a suitable shift [of one film relative to the other one] you will likely find a high coefficient, but does it mean that the flights of those two crows were causally connected?

And so, we decided to restrict our investigation by considering one station, but to take into account the entire period covered by the data on the SC, i. e., the eight years from 1924 to 1931. It was necessary to establish whether Clayton's results pertaining to the country which he especially studied and for which they, the results, occurred most striking were corroborated².

2. The Data

When beginning our work, the *Annals* (1932) had not yet appeared whereas (Abbot, no reference provided) it was already known that a large part of the previously published values of the SC should now be considered dated because the methods [of measurement] had been since improved and a number of new corrections introduced. We could therefore only base our study on the following sources.

1. The diagram of the daily values of the SC at mount Montezuma in Chile for 1924 – 1930 (Abbot 1931).

2. The Table of the mean decadic and monthly values of the SC (Abbot 1931, p. 12).

3. The *Daily Map* (no date) containing the same data on Montezuma for the period from 24 July 1927 to the end of October 1931³.

Here is how we proceeded. The ordinates on a photo of the Abbot diagram (22.5·17.5cm) enlarged 2.5 times were measured twice and all the doubtful cases thoroughly considered. A number of values of the SC was thus established. Abbot distinguished satisfactory, almost satisfactory and unsatisfactory data by differing symbols (S, S– and U) and we were therefore able to determine decadic and monthly means in which he neglected those of the last-mentioned type. A comparison of our means with his was satisfactory; namely, for all eight years the decadic means of the CC were 0.990 with fluctuations in separate years from 0.977 to 0.994 and the monthly means for all that period, 0.9998. Deviation of the former from the latter, 0.963 with fluctuations in separate years from 0.946 to 0.986.

We compiled the series of values of the SC selected for the further work in three parts: from Jan. 1924 to 23 June 1927 (obtained from the Abbot diagram), from 24 June 1927 to 31 Dec. 1930 (the data corrected by critical comparison with the *Daily Map*), and for 1931 (*Daily Map*, the only source here). The *Unsatisfactory* data were neglected.

For comparing the SC with MT, we selected the data pertaining to Cordoba (Carta del Tiempo) in Argentina⁴. They only had a few essential gaps (40 days in succession from 1 Jan. 1929, and 15 days both in Dec. 1928 and Dec. 1931); other gaps were not longer than two days in

succession (in the mean, missing was a little less than one day monthly) and we decided that it was permissible to fill those [shorter] gaps by linear interpolation.

3. Comparison of Our Series of the Solar Constant with Abbot's Final Data

Those final data (*Annals* 1932, Table 31, Montezuma 1920 – 1930, pp. 195 – 213) only became available after we concluded our work. We may certainly ignore the deviations concerning the *Unsatisfactory* cases, the rejection of those cases or the change from gaps to *Unsatisfactory* or vice versa as well as the change from *Satisfactory* to *Almost Satisfactory*.

There were 76 deviations left (Table 1), 10 of them (with symbol W attached) based on the *Daily Map*, one of those caused by an unfortunate reading of the Diagram (28 Oct. 1927). The rest 65 cases, as we ought to state regretfully, were mistakes of the Diagram itself, reprinted without change in the *Annals* (1932, p. 246). Concerning their influence on the results, the number of mistakes can be thought unimportant.

The worst case concerns Jan. – March 1925 (16 mistakes). Ten of the other mistakes, each amounting to not more than 1 or 2 units [of the last digit] were absolutely insignificant; 51 that had occurred during 81 month are left, 1 mistake per 48 days, and they certainly cannot discredit our conclusions.

As to the worst case mentioned above, we made the necessary calculations anew. For 100 days of the comparison of MT with the SC (from the 56th to the 155th day of the year) we obtained the highest in absolute value CC of 0.39 for a shift of 10 days instead of 0.40 for a shift of 11 – 12 days according to the previous calculation. Thus, even for the worst case, the error turned out to be absolutely inessential.

4. The Treatment of the Series

For the sake of convenience we adopted the following artificial calendar (Table 2) considering that each year had 365 days. That assumption would not have been possible to make for a longer period, but for eight years the inaccuracy thus introduced may apparently be neglected.

We bear in mind the study of periods lasting 100 days: from the 56th to the 155th day and from the 156th to the 256th day of the year. The latter approximately corresponds to the period for which Clayton had considered the connection between SC and MT in Argentina, and we indicate the appropriate calendar dates in Table 2. For the calculations below, months were thought to be 30 days long except for December (35 days), and an artificial trick explained below was introduced for ensuring intervals of equal duration.

Following Clayton, we had to study the correlation between the decadic and monthly mean deviations, so we began by calculating the appropriate series; the means were taken with respect to the fifth and the fifteenth days of the appropriate moving time intervals. For the MT, because of the filling of the random gaps in the data (§ 2), the number of consecutive terms was always the same (10 and 30); for the SC, we calculated the arithmetic mean for the data at hand in those decadic and monthly intervals; following Abbot, we did not exclude cases in which even only one observation was

available. The units adopted were 0.001cal/cm^2 and 1°C and the means and the deviations were calculated to one decimal point.

The numbers in the first series were rounded to integral units; the same was done with those of the second series after multiplying them by $10/3$. We then calculated the sums and the squares [of those numbers?] for the moving twelve three-monthly periods of each year (January – March etc.). The lacking data on the SC for Nov. and Dec. 1931 were filled up by the means calculated for the same months of the other years [of all other years?]; and, when calculating the sums for the first three months of 1924 and the last three months of 1931, we replaced Dec. 1923 by Dec. 1931 and Jan. 1932 by Jan. 1924. For each three months we denoted the square of the mean square deviation $\sigma_{3,ij}^2$ where i denoted the month, and j stood for the year.

Then, separately adding up the appropriate numbers of each month for all the years, we called the 12 numbers $\sigma_{3,i}^2$, $i = 1, 2, \dots, 12$, which described the mean fluctuation of each three months for all the eight years. These numbers are shown on Fig. 1 by small circles, separately for SC and MT. There also, are the $\sigma_{3,ij}^2$ shown by points for each year.

Becoming thus convinced in the presence of a yearly course of fluctuations, we expanded each empirical function $\sigma_{3,i}^2$ in a Fourier series. It occurred that they can be satisfactorily represented by three (SC) or two (MT) first harmonics shown on Fig. 1 by continuous curves. Their parameters were (A_0 – arithmetic mean; A_i and B_i – coefficients of cosines and sines of harmonic i respectively):

$$\begin{aligned} \text{SC: } A_0 &= 11.191, & A_1, A_2, A_3 &= 2.015, & 0.947, 0.777 \\ & & B_1, B_2, B_3 &= 0.153, & 3.031, 0.239 \\ \text{MT: } A_0 &= 38.958, & A_1, A_2 &= -17.255, & 4.830 \\ & & B_1, B_2 &= -2.800, & -0.136 \end{aligned}$$

For three-months periods the arithmetic means of SC and MT are very near to zero, and we will therefore insignificantly violate reality by replacing them below by expectations and by considering those latter equal to zero. And so, let there be m series of random variables

$$x_{j1}, x_{j2}, \dots, x_{jN}, j = 1, 2, \dots, m,$$

$$Ex_{jt} = 0, Ex_{jt}^2 = \sigma_t^2 = f(t).$$

Let

$$s_{2n}^2(t+1/2) = \frac{1}{2mn} \sum_{j=1}^m \sum_{k=-n+1}^n x_{j,t+k}^2$$

be the square of the mean square [literal translation] for the appropriate parts of all the series with centres at $[t + (1/2)]$. Then, obviously,

$$Es_{2n}^2(t+1/2) = \frac{1}{2n} \sum_{k=-n+1}^n f(t+k) = F(t+1/2).$$

If $2mn$ is sufficiently large, then, according to the law of large numbers, the mean value will be approximately equal to its expectation. But in our case $2mn$ is indeed sufficiently large as can be supposed on the basis of the smooth course of the magnitudes $\sigma_{3,i}^2$, which, owing to their meaning, ought to coincide with $s_{2n}^2(t+1/2)$. Let us call $f(t) = \sigma_t^2$ the *instantaneous*, and $\sigma_{3,i}^2$, the *mean three-month variability*.

As proved above, we will have an approximate equality

$$\sigma_{3,i}^2 = \frac{1}{90} \sum_{k=-44}^{45} \sigma_{t+k}^2$$

where t is the fifteenth day of month i . Supposing that σ_t^2 is a sum of several sine curves, we recall a well known fact: $\sigma_{3,i}^2$ will then be equal to the sum of the same number of sine curves having the same periods and phases, but altered amplitudes. Knowing the coefficients of the harmonics for $\sigma_{3,i}^2$ and wishing to determine the coefficients of the harmonics comprising the instantaneous variability σ_t^2 , it is only necessary to multiply them by

$$Q = \frac{2n \sin(h/2)}{\sin(hn)},$$

where, in our case, $2n = 90$, $h = 1, 2, 3^\circ$ for harmonics 1, 2 and 3 respectively, Note that its 12 ordinates, when expanding $\sigma_{3,i}^2$ into a Fourier series, were treated as being equally spaced in spite of the 35-day long December. This means that December was squeezed into 30 days so that at that stage of our work a year consisted of 360 days. This is exactly why the abovementioned values of h were obtained. Now, after calculating the coefficients of the expansion of σ_t^2 , and shifting the origin of the system of coordinates from mid-January 15.5 days back, we multiplied the coefficients of the appropriate harmonics by 4 (for the SC) and divided them by 2.25 (for the MT). Here are their final values.

The solar constant

$$a_0 = 44.764, a_1 = 8.445, a_2 = 4.710, a_3 = 4.819$$

$$b_1 = 3.048, b_2 = 19.387, b_3 = 9.704$$

The maximal temperature

$$a_0 = 17.315, a_1 = -7.839, a_2 = 2.939$$

$$b_1 = -3.608, b_2 = 1.655$$

Now, calculating the appropriate sine curves for each day of the 360-day year, then increasing the days of December up to 35 by interpolation, we compiled a table of the values of $10/k\sigma_t$ with $k = 2$ and $2/3$ for SC and MT. The deviations of the decadic means from the monthly means (see above) were multiplied by those values and the results rounded off to integers. Thus we obtained final series of standardized deviations. The multipliers k were selected so that the absolute values of numbers in the final series will not

exceed 21 or 22 which provided sufficient precision and essentially simplified further calculations.

5. Lack of Correlated Connection between the Solar Constant and the Maximal Temperature in Cordoba

That correlational connection was studied according to the following pattern. For MT, two intervals of 100 terms each were selected for each year, – from the 56th to the 155th and from the 156th to the 256th day, and two more of 50 terms each were obtained by dividing that second interval into halves. By comparing the SC with the MT of the day having the same number or a number less by 1, 2, ... we were able to obtain CCs with differing “shifts”. For the larger intervals CCs with shifts of 0, 1, 2, ..., 15 were calculated, and for the shorter intervals, only the CCs maximal in absolute value among the same shifts. When determining these maxima, we were guided by the maximal values of the products, partly by superimposing graphs and we checked our work by calculating a few CCs around the supposed maxima.

As ascertained above, it was impossible to apply in our case the usual formula of the mean square error, but the use of the suitable theory encountered some difficulties (see below), so that we applied the following method. First, we calculated the CC between the values of SC and MT for the same year, i. e., by combining our series in pairs (1924, 1924), ..., (1931, 1931). Second, we did the same for differing years, i. e. correlating MT of some year with the SC one or two years apart in either direction (Table 3) [call them combinations A and B].

The course of the CCs for combinations A and both large intervals is shown on Fig. 2. As an illustration, there also we show 4 correlation functions for the second interval and 4 combinations B. Our attention is at once arrested by the lack of any essential difference between combinations A and B. And it is also seen that even for the former combinations it is hardly possible to say that regularities are clearly discerned either in magnitude, sign or the shift corresponding to the maximal in absolute value CCs.

We now take a look at Fig. 3 where all the maximal in absolute value CCs are seen in a decreasing order; horizontal lines separate the larger and the lesser CCs and we clearly see that CCs of the same magnitude appear in both types of combinations and not rarer in the mean in group B. Thus, for the period between the 156th and the 255th day there are 8 [and 26] CCs in groups A and B; a half of those groups is not less than 0.49 and 0.36 respectively. However, we still ought to indicate that almost a quarter among group B reaches 0.49 whereas only 5 CCs from group A are higher than 0.39. It thus occurs that the difference only depends on one CC out of the eight which can well be a random occurrence.

Then, the insignificant superiority of group A in the series 156 – 255 is compensated by a superiority of B over A both in the interval 56 – 155 (the medians almost coincide, but considerably larger CCs are in group B) and in the shorter intervals (superior in both respects).

From all the above it follows that in Cordoba, if judging by the deviations of the decadic from the monthly means, correlational connection between SC and MT either does not exist at all, or is quite insignificant and the comparatively high CCs are simply maximal values of random errors.

We will confirm this conclusion by another method (§ 7) whereas § 6 is devoted to a slight digression.

6. On the Error of Determining the Solar Constant

When calculating the instantaneous variability σ_t^2 for the middle of each month (see Table 4), we clearly see the magnitude of errors from which the determination of the SC was yet unable to get rid of. Represent the deviation of the mean decadic from the mean monthly [values] x as the sum of the real deviation ξ and its error ε and denote the squares of their mean square deviations by σ^2 , α^2 and β^2 respectively. For any two months we will have

$$\sigma_1^2 = \alpha^2 + \beta_1^2, \sigma_2^2 = \alpha^2 + \beta_2^2.$$

If $\sigma_1^2/\sigma_2^2 = p$, then

$$\frac{\beta_2^2}{\alpha^2} = p - 1 + p \frac{\beta_1^2}{\alpha^2} \geq p - 1.$$

Comparing now all the months in Table 4 with November we find that for 6 months out of 12 $p \geq 2$. It follows that for these months not less than half of the magnitude of the deviations which we are studying are errors of observation. The deviations of the separate values from the monthly means are certainly corrupted by errors even more. It is hardly necessary to note that these conclusions, being a by-product of our work on which we cannot dwell anymore, should be specified by studying the probable errors of the numbers in Table 4.

7. The Mean Square Error of the Coefficient of Correlation between the Solar Constant and Maximal Temperatures

When shifting the series of SC and MT with regard to each other by 1 or 2 years and some days, from 0 to 15, we obtained, as stated above, 832 CCs, each of them for the two series consisting of 100 terms. Separating them into two groups depending on whether the SC precedes MT (a) or vice versa (b), we obtain two distributions of the CCs (Table 5, columns a and b). For estimating the homogeneity/heterogeneity of those distributions, we can apply Pearson's formula; in our case it will be

$$\chi^2 = \sum \frac{(a_i - b_i)^2}{a_i + b_i}.$$

We obtain $\chi^2 = 16.84$; for $n' = 15^5$, we have $P = 0.3$ which shows a sufficient correspondence between those distributions. This circumstance confirms our assumption that in any case when the shift is 1 year or larger, the CCs between SC and MT vanishes, and the empirical CCs are nothing but "errors". Considering now both groups together (Table 5, column c), we calculate the mean square error of those CCs: $\sigma_r = 0.2672$. Had our series been lacking internal connections, such an error for ($r = 0$) would have taken place if the number of terms $n = 1/(0.2672)^2 = 14$. Or, the presence of such

connections influences the square error and the number of terms is lessened from 100 to 14.

Supposing after Fisher that

$$z = \operatorname{arctanh} r = \frac{1}{2} \ln \frac{1+r}{1-r}$$

and, taking into account that in our case we may suppose that the real CC is zero, we find that $\sigma_z = 0.2873^6$. Assuming that z is normally distributed, we calculate the theoretical numbers corresponding to the group in Table 5 (column m)⁷. If, as it is done after Pearson, the extreme groups having theoretical numbers less than 1 are combined with the neighbouring groups, we will have $n' = 15$

$$\chi^2 = \sum \frac{(m-c)^2}{m} = 12.84$$

and the probability $P = 0.5$ of a random deviation of the empirical distribution from the theoretical.

This fact is not devoid of interest since it again confirms my hypothesis formulated in the abovementioned contribution⁸. In addition, and it is here certainly more important, we become able to estimate the most considerable CCs which occur when comparing SC and MT for the same years. In Table 3 we see that out of 256 CCs of that group not a single one exceeds 0.65. And since

$$z = \operatorname{arctanh} 0.65 = 0.7753,$$

which exceeds the calculated $\sigma_z = 0.2873$ only by a factor of 2.7, it means that not a single CC out of those 256 deviates from zero by three mean square errors. At the same time, according to the tables of the integral of probability, the theoretical number of deviations $\geq 2.7\sigma$ is $256 \cdot 0.00693 = 1.78 > 1$.

These considerations, as it seems, decidedly confirm the conclusion which we reached by following quite another approach, i. e., that there are no grounds for believing that the CC between the SC and MT in Cordoba appreciably differs from 0.

We will now check this conclusion in yet a different way by comparing the distribution of 256 CCs of group A with the theoretical obtained by studying the 832 CCs for pairs of different years (Table 6). It is not necessary to calculate χ^2 here: we see at once that it ought to be very considerable and the corresponding probability, very low. We ought to recall, however, that, as I had discovered in the quoted above paper, the χ^2 test is suitable, strictly speaking, only for totalities comprised of independent elements. It can be applied to totalities of dependent magnitudes⁹, if at all, only tentatively since an entirely adequate criterion is yet lacking.

It seems that dependence has a stronger influence when the number of terms is comparatively small which is well illustrated by Tables 5 and 6. Indeed, a close look at the latter rather sharply brings home that the deviation between the empirical and theoretical distributions occurs owing

to the essential accumulation of few cases in which the smoothness of the empirical distribution is grossly corrupted in a way that always takes place exactly in distributions of an insufficient number of elements.

In our case it is easy to explain this. Table 5 consists of 832 CCs, 52 groups of 16 terms each (shifts from 0 to 15 days) whereas only 16 such groups are in Table 6. At the same time the CCs in each separate group between certain series of the SC and MTs provide a series of 16 terms corresponding to shifts of 0 – 15 days closely correlated with each other; this is indeed revealed by the smooth wavy course of the relevant series (Fig. 2).

Therefore, if the maximal range of such a wave is about 0.55, say [?], and the wave forms a smooth stretched peak, a few consecutive CCs will at once be placed in the same cell. Two such waves are sufficient for 6 – 8 superfluous unities to occur, and they very considerably augment the value of the chi-square. Thus, for example, occurred the deviation between empirical and theoretical numbers in Table 6, third cell from above (15 and 7.7). This is easy to become convinced of when having a look at Tables 3a and 3b.

If these considerations are valid, an essential improvement will happen at once when the number of groups is decreased by combining symmetric categories, see Table 7. We get $\chi^2 = 9.61$ and $P = 0.2$. In other words, not more probable deviations occur roughly once in five cases of independent elements. There are therefore no grounds for concluding that that distribution essentially differs from those indicated by the theory when independence is assumed.

Calculation of σ_r by issuing from data of Table 6 provides 0.250 which almost coincides with the case of different years. The conclusion is obvious.

8. Some Preliminary Results of Analysing Series of the Solar Constant and Maximal Temperatures and Derivation of the Mean Square Error of the Correlation Coefficient

If SC and MT are really not correlated, the mean square error of the empirical CC should be represented by a comparatively simple formula

$$\sigma_r^2 = \frac{1}{n} \sum_{t=-n+1}^{n-1} \rho_x(t)\rho_y(t)$$

in which $\rho_x(t)$ and $\rho_y(t)$ are the true CCs between x_t and x_{i+t} and y_t and y_{i+t} . The difficulty in applying that formula consists in that, instead, we have to make do with the statistical CCs, $r_x(t)$ and $r_y(t)$; for more details, see my paper [x?] quoted above. The errors of these CCs can essentially corrupt the results because a large number of terms are being added up. In that previous paper the problem was really solved, at least in principle, for the case of $\rho(t) = 0$, $t > \omega$ and not large values of ω as compared with n . An example of a more difficult case is apparently encountered with the SC. We will assume an obviously highly probable hypothesis that the CCs between the values of SC separated by a year or more are either zero or negligible.

Comparing segments of the series of MT with numbers 156 – 255 taken either entirely ($n = 100$) or by parts with 40, 60 and 80 terms with the corresponding segments of the series of SC differing in time by one or two years in either direction and additionally shifted by 0 – 15 days we have

calculated 112 CCs for shifts of about 1 year, and 80 CCs for shifts of about 2 years for each of the cases $n = 40, 60, 80, 100$. Table 8 contains empirical mean square errors of the CCs calculated accordingly and we note that for shifts of about 2 years all the σ 's are somewhat smaller which perhaps argues for the presence of some remaining correlation (in any case, quite insignificant) at shifts of about 1 year. This can be checked by a similar study extended to shifts of 3 and 4 years. Anyway, the indicated differences can be neglected in the first approximation, and this is what we do.

Issuing from the known expansion

$$\sigma_r^2 = \frac{A}{n} + \frac{B}{n^2} + \frac{C}{n^3} + \dots$$

and restricting it to three terms, we determine by least squares that

$$A = 9.28, B = -164, C = 2190.$$

The *theoretical* (i. e., the adjusted) values of σ_r^2 are shown in the last column of Table 8.

We consider the satisfactory adjustment as a testimonial that the number of terms allowed for in the formula above was sufficient and that, as I have shown in the paper quoted above, the value of A should therefore satisfy the approximate equality

$$\frac{A-1}{2} = \sum_{t=1}^{\omega} \rho_x^2(t). \quad (*)$$

Replacing here ρ by empirical CCs r , we can determine an approximate value of ω which is calculated by taking $\rho_x(t) = 0$ for $t > \omega$; if $\omega > (n - 1)$ it should be replaced by that difference.

The next table (Table 9) provides the values of the serial CCs for SC and MT with shifts of 1, 2, ..., 31 days and for the 156th – 255th days of each year when correlated for shift t with the segment $(156 - t; 255 - t)$. All these CCs were calculated for the first and the second half of the 8-year period, and for that period as a whole.

The following remark suggests itself first of all: the first and the second 4-year period both for SC and MT provide sufficiently close correlational functions at least when the CCs are still more or less considerable; the discrepancy between them can be certainly explained by random errors¹⁰. A curious conclusion is that both SC and MT, after eliminating the 30-day level [?], and a suitable standardization of the fluctuations can be considered homogeneous, at least in the first approximation. If the future confirms and extends that inference to other geophysical series, it will be quite an important step in their statistical studies.

We have found the value of the coefficient A , $A = 9.28$. Therefore, the right side of (*) is equal to 4.14. We do not know the true CCs or values of ρ_x , but when replacing them by their approximate values r_x , the sums of the squares of the CCs calculated by means of Table 9 provide

$$\sum_{t=1}^{31} r_x^2(t) = 3.45$$

and it is obvious that, since the further CCs are doubtless small, a large number of them are needed for coming near to 4.14, so that ω should be considerably greater than 31.

However, bearing in mind that the squared sum of all the rest CCs in the series of SC from $t = 32$ to infinity is a magnitude of the order of 0.5 (approximately equal to the difference $4.14 - 3.45$), we may hope that the sums of the products of serial CCs for the SC multiplied by the same CCs for the MTs can also be established although somewhat roughly. Multiplying the appropriate values taken from Table 9, we find for $n = 100$ the approximate equality

$$\sigma_{r_{xy}} = \sqrt{\frac{1}{100} [1 + 2 \sum_{t=1}^{31} r_x(t)r_y(t)]} = \sqrt{0.0897} = 0.30$$

which is very near to its empirically determined value 0.27.

In all probability, the further CCs (for shifts $t > 31$) are not important and, in addition, the error made by neglecting them was possibly compensated by dropping the term of order $1/n^2$. In any case, it is hardly accidental that the values of the mean square error of the CCs between SC and MT derived by such different methods are so close.

Explanation of Tables and Figures

Table 1. It lists the values of SC both adopted by Slutsky and either published in the *Annals* (1932) indicating categories *satisfactory* (S), *almost satisfactory* (S-) and *unsatisfactory* (U), or included with symbol W in the *Daily Map*, and the differences between them.

Table 2. Lists the month and day for the 1st, 56th, 155th, 255th and 365th day of an artificial calendar. Example: the 155th day of 1927 = 3 June 1927.

Fig. 1. Cordoba, SC and MT, separately. Shows by points their mean variability σ_{3ij}^2 over three months (Jan. – March, Febr. – April, etc) for 1924(1)1931. Their mean variability (the deviations of the decadic means from the monthly means) over those eight years σ_{3i}^2 shown by small circles. Continuous curves show the sum of three or two harmonics for SC and MT respectively. Translation of legend partly tentative owing to difficult original text.

Table 3a. Lists CCs between SC and MT for period 56th – 155th day, years 1924(1)1931, shifts 0(1)15 days; separately shown are combinations of same year and of different years.

Table 3b. Same for period 156th – 255th day.

Fig. 2. CCs between SC and MT for same year (two upper series) and different years (the lower series), shifts 0(1)15 days. Additional curves shown with inadequate explanation moreover only given in text.

Fig. 3. Maximal in absolute values CCs between SC and MT for same year (A) and different years (B) for series of 100 and 50 days and shifts of 0(1)15 days.

- Table 4.** Lists magnitude σ_r^2 for each of 12 months, year not indicated. Explanation lacking; explanation in text (§ 6) only states that SC is meant.
- Table 5.** Frequency table of CCs between SC and MT for different years, separately for SC preceding MT and vice versa and combined. Theoretical magnitudes additionally provided.
- Table 6.** Frequency table of CCs between SC and MT for same year, empirical (m') and theoretical (m) values.
- Table 7.** Same for absolute values of those CCs. Magnitude $[(m' - m)/m]^2$ additionally provided leading to $\chi^2 = 9.61$ and $P = 0.2$, see end of § 7.
- Table 8.** Lists empirical mean square errors of coefficients of serial correlation for SC, σ_r^2 , shifts of about 1 year and about 2 years, and both these shifts combined, periods of 40, 60, 80 and 100 days. Theoretical values of σ_r^2 additionally provided.
- Table 9.** Lists coefficients of serial correlation for SC and MT, shifts of 1(1)31 day, periods 1924 – 1927, 1928 – 1931 and 1924 – 1931, interval 156th – 255th day.

Notes

1. In § 5 of the main (Russian) text, Slutsky wrote: *We show* [on Fig. 2] *4 correlation functions* etc. Anyway, it is difficult to understand what exactly is shown there. In the context of this paper, *correlation function* means *values of the CCs*. O. S.
2. Abbot (*Annals* 1932, p. 277 and 255ff) has recently put forward a new concept concerning the connection between SC with the weather. He assumes that each periodic component of that constant is reflected in the phenomena of weather with differing shifts moreover variable in time. Separate waves are superimposed upon each other and the connection can be lost in the general picture. The material he adduced for proving this thesis is still too scanty for being convincing but it is extremely interesting, suggests ideas and for the time being compels us to abstain from a final judgement. A check of that new theory was not included in our aims. E. S.
3. Abbot (*Annals* 1932, Table 31, pp. 195 – 213) had since essentially corrected the values of the SC published there before the indicated date. E. S.
4. For Cordoba, Clayton derived one of his best results, $CC = -0.74$. True, the CC was even higher for some stations in Argentina, – up to -0.82 in Sarmiento, – but upon revealing that there were so many missing days we preferred Cordoba. E. S.
5. When being increased by 1, there will be 16 (groups) – 2 (connections) + $1 = 15$ degrees of freedom, as Fisher called it. E. S.
6. By applying the formula

$$\sigma_z = \sigma_r \sqrt{1 + 2\sigma_r^2 + 2\frac{1}{3}\sigma_r^4 + 4\sigma_r^6 + \frac{1}{5}\sigma_r^8 + 24\frac{2}{3}\sigma_r^{10} + \dots}$$

see my paper on the distribution of the errors of the CCs, quoted above, pp. 95 – 96. E. S.

[Slutsky twice (here and in § 8) refers to his previous paper *quoted above*. Actually, he did not mention any previous paper, but the page numbers here stated allows to conclude that he meant either [x] or Slutsky (1929). O. S.]

7. I took the values of z corresponding to $r = 0.5$ [0.05?], 0.15, 0.25 etc from Romanovsky's table (1928, p. 147). E. S.
8. Apparently, Slutsky (1929). O. S.
9. It was Fisher, who, in 1925, showed that the chi-squared test was not suited for studying dependent trials, see Hald (1998, p. 201). O. S.
10. We saw that for sufficiently large values of n and $t > 2\omega$ we may take

$\sigma_r = \sqrt{9.28/n}$ for the CC between SC and MT. According to the above calculations, we have $\sigma_r = 0.267$ at $n = 100$ and we may therefore approximately assume that $\sigma_r = 0.13$ at $n = 400$. Although all the necessary formulas are available, we are not yet able to calculate σ_r for serial CCs at lesser shifts, but the indicated magnitudes probably provide sufficiently correct indications about their order. E. S.

Bibliography

Abbot C. G. (1927), Corrected solar constant values, Montezuma, Chile, from May 27 to August 24, 1927 inclusive. *Monthly Weather Rev.*, September.

--- (1931), Weather dominated by solar changes. *Smithsonian Misc. Coll.*, vol. 85, No. 1 Publ. No. 3114, Washington.

--- (1932), In *Annals Astrophys. Obs. Smithsonian Instn*, vol. 5.

Clayton H. H. (1923), *World Weather*. Washington.

Daily Map (no date), *Daily Weather Map of the United States Weather Bureau*.

Hald A. (1998), *A History of Mathematical Statistics from 1750 to 1930*. New York.

Romanovsky V. I. (1928), *Elementy Teorii Korreliatsii* (Elements of the Theory of Correlation). Tashkent.

Slutsky E. (1929, in Russian), On the [mean] square error of the correlation coefficient for homogeneous connected series. *Trudy Konjunkt. Inst.*, vol. 2, pp. 64 – 101.

--- (1933, in Russian), On the distribution of the errors of the correlation coefficient for homogeneous connected series. *Zurnal Geofiziki*, vol. 2, No. 1, pp. 66 – 98. Corrections in No. 2. Translated here [x].

XII

On the Solar Constant

K voprosu o solnechnoi postoiannoii. *Zurnal Geofiziki*, vol. 4, 1934, pp. 392 – 399

Summary [in its original English]

The Problem of the Solar Constant

1. Serial correlations found by C. G. Abbot for the solar constant values showing discordant features for the different years scarcely can be considered as really significant owing to the relative paucity of the data constituting the separate yearly series. The formula of the probable error employed by the same author is unapplicable to the series of this art, the consecutive terms forming the series being not independent from each other. The serial correlations found by the same author for two groups of three years each (see Fig. 1) must also be discarded being biased by the method of their formation (the similarities and dissimilarities of the serial correlations for separate years being the ground of the unification or of the rejection of the data).

2. The serial correlations published in the present note are the correlational functions for the deviations of the ten-daily from the thirty-daily means of the solar constant of radiation standardized by the factors inversely proportionate to the momentaneous standard deviations (for more details see [xi]). The said deviations relating to 156 – 255 days of each year (1924 – 1931) were multiplied by the respective values t days before and the correlation coefficients were then formed 1) for the first four years ($n = 400$), 2) for the second four years ($n = 400$), and 3) for all eight years ($n = 800$). Each series contains the correlation coefficients from r_0 to r_{143} (see Table 2 and Fig 2).

3. In analysing the results the method of the formation of the series under consideration must be accounted for.

Let x_1, x_2, \dots be some series of the mutually independent random values taken at random from the same general population. Then the deviations of the art used here [see formula (1) in the text] will be intercorrelated, the serial correlations being given by the formula (3) leading in our case to the values of the Table 2 (see also the little crosses line, Fig. 2). The values of this function for $T > 30$ being 0, the striking similarity between the correlational functions for two consecutive four-years groups must be therefore regarded as probably significant. The positions of the maxima and minima suggest the hypothesis that the approximate regularity observed therein may probably be occasioned by the revolution of the Sun. Whether there is a strong period in the solar constant values, or the cycles occasioned by the Sun's revolution are of the pseudo-periodical character we cannot say as yet. The problem evidently deserves further studies.

[The Main Russian Text]

Abbreviation: CC = correlation coefficient

SC = solar constant

1. Abbot (1922) published the results of his study of the serial correlational connection of the SC. He separately investigated each year from 1908 to 1916, only leaving out 1912 due to bad conditions for observations caused by the eruption of Katmai [a volcano in Alaska]. Multiplying the values of SC by their values 1, 2, 3, ... days earlier, he thus determined the relevant CCs for r_1 to r_{40} .

It is not necessary to reproduce his graphs; Abbot himself, when commencing his study, remarked first of all that the appropriate curves were dissimilar. I will only provide the mean course of the CCs for two groups of three years each (Fig. 1). I selected the first three years (1908, 1911 and 1913) because of some similarity in those courses; I entirely rejected two years (1915 and 1916) owing to the sharp peculiarity of their correlational functions, and I combined the remaining years (1909, 1910 and 1914) into the second group. The reader will see that the two graphs indeed indicate quite different courses and in many features they are even contrary.

If periodic components are present, the correlational function must reveal the appropriate periods, and Abbot concludes that not a single clearly visible periodicity in the fluctuations of the CCs had been preserved over all the eight years of his study: *Each season is a law unto itself*. That conclusion, generally speaking, would not at all been unlikely, but the foundation that led Abbot to it ought to be questioned.

The main point is that he considers that the discord between the results for separate years was essential because the CCs calculated by him often exceeded their probable error many times over. However, he determined that error (as it is regrettably done very often) by means of a formula only suited when connections between the terms of a series are lacking. His reasoning therefore falls down and all of his other arguments are up in the air. Indeed, whether the discrepancies between the calculated results are significant or not; could they be occasioned by a random coincidence of circumstances or not, – judging that by the eye without any chances of checking yourself by a rigorous calculation is certainly impossible¹.

When considering the graph of the course of the SC we indeed convinced ourselves in that that magnitude can by no means be disconnected, be such whose consecutive values do not at all depend on each other in the stochastic sense. No calculations are even needed for reaching such conclusions since the wavy fluctuations in the course of the SC are seen too strikingly. These waves are very diverse. Some are short, lasting a few days, others cover months and there also are waves, that is, regular lowerings and rises, going on for years on end.

Under such conditions, if the studied series does not last a large number of years, the determination of the serial CCs for the SC seems to be rather helpless.

For coherent series, the number of observations is only enough if they cover sufficiently many waves. When there are very few longest waves, they have to be treated individually rather than statistically, to be separated as a secular component by some statistical method. True, none of these latter can be considered quite satisfactory for an objective analysis; it is much more rightful to see them as practical tricks for arbitrarily treating a given numerical series and providing *preparations* rather than its real components.

When mentioning *preparations*, I conscientiously wish to recall biological analogies, for example microscopic sections treated by various

chemicals. Such preparations are not real but artificially created parts of the studied organism. And what we discern then represents corrupted pictures of reality. Nevertheless, they are known to be useful provided we are familiar with the properties of the operations made in the process and precisely understand the essence of the inserted corruptions.

With regard to the statistical methods of making preparations or at least to some of them, we possess such knowledge. Series can be treated in a way that neither the periods, nor the phases of harmonic functions, into which it can be expanded, will change, only the amplitudes will be corrupted which can be easily taken into account. Reducing long-period waves to insignificant amplitudes we obtain series sufficiently long compared to the essentially important for them shorter fluctuations so that hopefully they can be successfully treated.

2. Clayton applied one of such preparations (deviations of the decadic means from the monthly means) when studying the connection of SC with temperature. In the paper indicated above¹ I thoroughly analysed his conclusions by examining one of his examples (Cordoba, in Argentina). I naturally had to apply his methods of smoothing series and thus obtained as a preparation from a number of values of the SC the deviations of the decadic means from the monthly means. Since my main aim was to study the connection between SC and temperature, some issues concerning the SC remained unascertained.

In particular, I only determined the serial correlation coefficient for shifts of up to 31 day although it seems almost unquestionable that the connection does not vanish there. Naturally I wished to fill that gap. Concerning the SC we now have CCs for shifts from 1 to 143 days (Tables 1 and 2). Owing to lack of time they were only calculated for 800 days rather than for the whole material at hand, namely only for 100 days (from the 156th to the 256th day) of each of the 8 years 1924 – 1931 for which we had the data on SC.

The CCs were calculated separately for both 4-year periods and for the 8-year period as a whole. A glance at the diagram (Fig. 2) is sufficient exactly now, when we have series of the SC from r_1 to r_{143} , for becoming convinced in the reality of connection. The following reasoning will show why it was by no means possible to be satisfied by the previous data, i. e., by series ending with the shift of 31 days.

We have to do not with the SC itself, but with a preparation. So how was it constructed? Denote the values of SC by x_1, x_2, \dots , then the numbers in our series will be represented by the formula

$$y_{i+15} = \frac{1}{10}(x_{i+11} + x_{i+12} + \dots + x_{i+20}) - \frac{1}{30}(x_{i+1} + x_{i+2} + \dots + x_{i+30}), \quad (1)$$

$$y_{i+15} = - \left[\frac{1}{30}(x_{i+1} + \dots + x_{i+10}) + 2(x_{i+11} + \dots + x_{i+20}) + (x_{i+21} + \dots + x_{i+30}) \right].$$

Suppose that the values of SC, x_1, x_2, \dots , are random numbers not connected with each other. We know that, when forming moving sums from the terms of such a series according to the formula

$$y_i = \sum_{k=1}^s A_k x_{i-k}, \quad (2)$$

the CCs will be

$$r_t = \frac{\sum_{k=1}^{s-t} A_k A_{k+t}}{\sum_{k=1}^{s-t} A_k^2}. \quad (3)$$

Above, see formula (1), we have provided the values of A_k for our case and now we calculate the CCs by formula (3), see Table 2; on Fig. 2 they are shown by crosses. Generally speaking, they are so close to the actually obtained CCs that, until we restricted our investigation to series up to r_{31} , we could have apparently considered the obtained picture to an essential degree as a sole result of smoothing and would have thus *completely* explained the coincidence of the series of the CCs for the first and the second 4-year period.

It turns out, however, that the issue is not at all as simple as that. Suppose that all the coefficients of the serial correlation are zeros, then, up to n_{30} their course will be such as shown on Fig. 2 and vanish after r_{30} [?]. However, it would be absolutely incomprehensible why the further courses of our series for both 4-year periods will then be so similar. In both series we have

minima at shifts 11, 40 – 41, 68, 83 – 85 [rather 93 – 95],
and 115 – 116
maxima at shifts 30 – 31, 49 – 54, 83 – 85, 103 – 105
and 124 – 125

Consider that the maxima for the curve describing the entire 8-year period are $t = 31$, then 226 , $327\frac{2}{3}$, $426\frac{1}{4}$, $524\frac{4}{5}$, and allow for the possible influence of random errors. Then the hypothesis that the course of the correlation function reflects the rotation of the Sun about its axis becomes very likely since the figures above are close to the synodic period [close to those that would follow from the synodic period] of that rotation.

This would have provided a material cause for the presence of the main wave revealed above in the correlation function. Deviations, if becoming real after analysing more complete materials, could have possibly be explained as the result of interference with other periodic or pseudo-periodic components.

We also note that one of the latest contributions of Abbot (Abbot & Bond, Publ. 3172), even if not yet proving that strictly periodic components of SC do exist, had at least made their existence highly likely. Most convincing seems to be the agreement between the phases of the waves of different parts of the series and their concord with the phases of the wave established for the series as a whole, see waves C_1 , C_2 and C_3 with period of 8 months and waves D_1 , D_2 and D_3 with period 11 months on Fig. 3 of their p. 5. This issue certainly deserves further study.

Explanation of Tables and Figures

Fig. 1. The mean course of the CCs along the series of SC for shifts of 1(1)40 days for two groups of three years each (Abbot).

Table 1. Lists CCs in a smoothened series of SC. The coefficients are shown for $t = 1(1)143$ separately for 1924 – 1927, 1928 – 1931 and for the period 1924 – 1932 as a whole.

Fig. 2. Correlation function for SC separately for 1924 – 1927, 1928 – 1931 and for 1924 – 1931 as a whole. Crosses indicate the course of that function for a smoothened disconnected series.

Table 2. Lists the values of the CCs (of $60r$) for a disconnected series smoothened according to formula (1) and $t = 0(1)30$. The text makes it clear that this table deals with SC.

Note

1. Curves shown on Fig. 1 are also unconvincing since Abbot combined the years in a group not consecutively, but according to similarity/distinction of the correlation function. He thus introduced an element of selection that entirely compensated the increase in the number of observations and utterly corrupted the independence of the series. E. S.

Bibliography

Abbot C. G., Bond Gladis T. (Publ. 3172), Periodicity in solar variation. Smithsonian Misc. Coll., vol. 87, No. 9.

Annals (1922), *Annals of Astrophys. Obs. Smithsonian Instn*, vol. 4.

Annals (1923?), *Annals of Astrophys. Obs. Smithsonian Instn*, vol. 5.

XIII

On the Eleven Year Periodicity of Sunspots

Comptes Rendus (Doklady) Acad. Sci. URSS,
t. 4 (9), No. 1 – 2 (70 – 71), 1935, pp. 37 – 40,
English version

Communicated by I. M. Vinogradov

The result of most attempts hitherto made to unravel the riddle of the 11-year sunspot cycle being rather uncertain (Stumpf 1930, pp. 39 – 41), there is no doubt that a real progress in this matter can scarcely be attained without a considerable widening of the research field. The most promising data seem to be the information about the north [northern] lights recorded by historians of past ages, which subject has so far been treated rather inadequately.

We shall start from the observation that only epochs of the most intensive north lights can be considered as fairly trustworthy indices of the epochs of sunspot maxima, other cases being scattered through the whole extent of the solar cycle. Our set of data is therefore only an excerpt from the full list of observations contained in the well-known catalogue (Fritz 1873).

To avoid the superposition of “personal equations”¹, we have made up our series solely from the epochs underlined by Fritz (1878) himself, this series beginning with 397 AD and stopping at 1605 to prevent its overlapping with the epochs of sunspots min/max (Brunner 1930, p. 77), see Table 1, Columns 2, 3. For the time before 397AD we have adopted (with a few amendments²) all the data given by Fritz (1878, pp. 37 – 38, 40), the reason for this being that the north lights observed in Greece or Italy must be considered to be the most prominent ones. Finally, all the epochs have been represented by the middle moments of the respective years, 1AD, 1BC, 2BC, ... being designated as 1.5, 0.5, – 0.5, ... (Table 3, Column 2)³.

Turning to the sunspot min/max, we have the condition

$$\sum[(T_0 + \lambda k - t_k)^2 + (T'_0 + \lambda k - t'_k)^2] = \min,^4$$

λ being the period, T_0 and T'_0 the initial theoretical, and t_k, t'_k the empirical epochs of minima, resp. maxima (Wolfer 1902, pp. 95 – 96). We obtain $\lambda = 11.13724$, $T_0 = 1766.67$, $T'_0 = 1771.68$ ⁵. Since further calculations based on these values lead to the deviations – 4.4 and 5.9 (maxima for $k = 1, 2$) which seem to be exorbitant, it was judged that by cancelling the four epochs corresponding to $k = 1, 2$, more reliable results should be obtained. Finding thus new normal equations, we have $\lambda = 11.14435$, $T_0 = 1766.912$, $T'_0 = 1772.064$, the respective errors ($t_k - T_k, t'_k - T'_k$) being given in Table 1 (Columns 4, 5)⁶.

The following observations can now be made. A) The standard errors being $\sigma_{\min} = 1.135$, $\sigma_{\max} = 1.345$, the epochs of the minima appear to be more stable than those of the maxima. B) The distribution of the \pm signs being evidently not random, an application of classical formulas of probable errors to our case could be considered illegitimate. C) Dividing the series into two parts ($k \leq 0, k \geq 3$) we find that α) the arithmetic means of the

errors are -0.03 and 0.05 (min) and 0.24 and -0.30 (max); β) the mean absolute errors are 0.93 and 0.78 (min), 1.08 and 0.95 (max); γ) the error distribution corresponding to the first and to the second part of the series do not seem to be substantially different. For instance, the errors of the epochs of the minima which do not exceed 1.4 are $12/15 = 80\%$ in the first, and $10/12 = 83\%$ in the second case. D) the results obtained seem to be consistent with the hypothesis of a unique period, the series, however, being not large enough to preclude every possibility of doubt.

If we now consider the polar [the northern] lights data, we find that the numeration [enumeration] of the epochs not being given a priori, the problem is to be treated by successive approximations. *First stage.* Let us round up the epochs of polar lights (Table 3, Column 2) by subtracting throughout $1/2$, and the theoretical epochs of the sunspot maxima in the 19th century by cancelling the decimals. The last two figures of these epochs

05 16 27 38 50 61 72 83 94

shall now be considered as epochs of maxima for every century, thus putting $\lambda = 111/2$ as our first approximation. Taking the terms of the last series (τ_k) and those of the former one (t_k) to be mutually correspondent if $|t_k - \tau_k| < 1/2\lambda$, we find $\sum(t_k - \tau_k) = 27$. Then, the middle term of the second series being 550 , we find $\tau_k = 550.75$ for the corrected origin.

Second stage. We have to compare now the suppositions $9\lambda = 100 + \eta$, $\eta = 0, \pm 0.1, \pm 0.2, \pm 0.3, \dots$ To this end the above series of the polar lights is to be confronted against every set of values given by the equation

$$\tau = 550.75 + k\lambda$$

for each of the values of λ just named with the subsequent rounding up of the values of τ in the manner described above. Shifting then the origins so as to make the respective sums of the errors vanish, we find as a matter of fact that, with any λ , the correspondence between the terms of the respective sets has nowhere been upset by this operation. The corrected sums of the squares of the errors (ϵ^2) have then been found to be as follows [Table 2]. The numeration of the epochs of the north lights (Table 3, Column 1) remaining identical in the interval $-0.2 \leq \eta \leq 0.1$.

The circumstance just mentioned allows us to take the *last step* by solving the normal equations corresponding to the condition

$$\sum[(\tau_0 + \lambda k - t_k)^2] = \min$$

k being the ordinal numbers found above and t_k the values of the empirical epochs of the north lights (Table 3, Column 2). Hence we obtain $\lambda = 11.10266$, $\tau_0 = 550.783$, the deviations of the empirical epochs from the theoretical ones being given in Table 3, Column 3.

It can easily be shown that the probability of a random provenance of the deviations found by us must be held for a very slight one. We shall not, however, discuss this topic, there being another test of a considerably more conclusive character. Let us extrapolate the theoretical values on the basis of the parameters just found and compare the values obtained with the empirical values of the sunspot maxima. The respective deviations (Table 1,

Column 6) exceeding on the average very little the errors of the direct method (Table 1, Column 5), the excellence of the agreement would appear obvious. There would scarcely remain any doubt as to the significance of the result, should we not remember that some sources of error may be eventually inherent in the “personal equation” of the author to whom we owe the data. The problem undoubtedly deserves further study⁷.

Notes

1. Even if in inverted commas, this expression is applied in a greatly generalized meaning; in astronomy, it is a precise term, here it means something like personal inclination. An additional explanation was desirable: why did Slutsky believe that the previous data were reliable and why no new information was not needed? O. S.
2. 208BC has been included in the group 202 – 199BC, the epoch 204BC being taken as representative of the whole group. We have also substituted 217BC for 216BC, which is taken by Fritz to represent the group 218 – 215BC. E.S.
3. Years 1AD, 1BC and 2BC are “designated” 1.5, 0.5 and – 0.5. How come? O. S.
4. Slutsky wrote down the condition of the method of least squares (but that expression is lacking) and below he mentioned normal equations, but did not provide them. And what did he mean by the “initial theoretical epochs” of sunspot extrema, and where did he find them? He did not explain the meaning of k ; apparently it is the number that Fritz (1873) had attached to a group of consecutive years. O. S.
5. Here and below, all decimals actually used in the calculations are given, the question of the significance [of the final result] being postponed to a further stage of the study. E. S.
Slutsky issued from numbers with six significant figures and calculated the periodicity with seven figures which was absolutely senseless and kept to the same approach in other contributions as well, see for example [xiv, Note 7]. O. S.
6. Slutsky rejected the cases $k = 1$ and 2 and Table 1 contained new calculations. Nevertheless, those cases are included there and the deviations did not diminish and neither did the periodicity change (both times it actually was 11.1 years). Perhaps the error of the new determination was less, but he did not estimate it which in itself was an essential shortcoming. O. S.
7. All three Tables explained in text. O. S.

Bibliography

- Brunner, W., Editor** (1930), *Astronomische Mitteilungen*. Eidgenössische Sternwarte Zürich, Bd. 124.
- Fritz H.** (1873), *Verzeichnis beobachteter Polarlichter*. Wien.
- (1878), *Die Beziehungen der Sonnenflecken zu den magnetischen und meteorologischen Erscheinungen der Erde*. Haarlem.
- Sheynin O.** (2002), Simon Newcomb as a statistician. *Hist. Scientiarum*, vol. 12, pp. 142 – 167.
- Stumpf K.** (1930), *Prager geophys. Studien*, Bd. 4, 33.
- Wolfner A.** (1902), *Eidgenössische Sternwarte Zürich*, Bd. 93.

XIV

Statistical Experiment As a Method of Investigation. Critical Notes on the Problem Earth – Sun

Statisticheskii eksperiment kak metod issledovaniia.
Kriticheskie zametki k probleme “Zemlia – Solnze”.
Zhurnal Geofiziki, vol. 5, No. 1, 1935, pp. 18 – 38

Summary [in its original English]

On the Statistical Experiment As a Method of Empirical Research. Critical Essays to the “Earth – Sun” Problem

It is highly probable that by far the greatest part of the statistical problems in geophysics cannot be reasonably treated under the supposition of the independent probabilities. The decisive question of the significance cannot be therefore settled in such cases with the aid of the usual formulae of the standard errors. Some illustrations of the fallacious [fallacious] conclusions reducible to this source are treated in §§ 2 and 3 (F. Baur, C. G. Abbot, H. Clayton).

The difficulties do not belong however to those which could be easily obviated with the aid of an adequate theory (see the papers by the present author [in the Bibliography]), the respective formulae leading for the most part to calculations of a rather prohibitive art to be used in ordinary practice, not to say that the number of observations may be often not large enough to secure the significance of the results. These are therefore the grounds why in some cases the method of statistical experiment can be reasonably applied in the empirical research and the approximative solutions of the significance questions obtained, more apt however to reject than to accept definitely a hypothesis in question.

The paper contains some specimens of the application of this method to the critical examination of results obtained by several students in the problem of the influence of the fluctuations of the sun’s activity on the terrestrial phenomena.

Central Institute of the Experimental Hydrology and Meteorology

Introductory Remarks

The statistical experiment (extraction of tokens from an urn, tossing coins, etc) had played an important part in the development of statistical thought. By illustrating various propositions of the theory of probability, experiments showed that it was possible to create conditions under which the premises of independence of trials and constancy of [the appropriate] probability actually took place. The discussion of the patterns and results of various experiments repeatedly prompted discussions and led to deeper penetration of the problems studied.

However, statistical experiments had and have to fulfil some essential functions even beyond the field of the most general issues of the theory of probability. When the demands of practice outstrip theory, exactly they have to provide the solution that in essence should have been obtained purely mathematically. For example, experiments have been repeatedly applied for estimating the degree of approximation of some formulas only strictly valid

under limiting and often practically unattainable conditions. In some cases the correctness, or, more precisely, the applicability of a formula derived by non-rigorous considerations was ascertained experimentally hoping that it will prove at least approximately correct.

The above can be illustrated by many examples but it would have, however, diverted us far from our goal since that function of statistical experiments, whose significance we would like to show by means of a small number of examples, essentially differs from those indicated and was until now barely practically applied. That goal again certainly has to do with such problems that either had not been yet theoretically solved, or if for some reason such solutions cannot be applied. However, unlike the described above, the statistical experiment must investigate some concrete material problem rather than check formulas.

A typical case, as it appears, will be such that either some statistical magnitudes are calculated or some actual coincidences established; had we been sure that these latter were not random, that the appropriate magnitudes are near to each other not accidentally, or do not accidentally deviate from zero, then definite conclusions about the essence of some known processes (for example, about their mutual connection) should have followed.

It is for cases in which at present we do not know any formulas for solving such problems, that we provide the method of statistical experimentation¹. That method can be understood both in its narrow, proper sense as a real experiment creating sequences of random numbers corresponding to the conditions of the problem at hand, and in its somewhat wider sense. At present, we will not yet separate the two possibilities; later, after collecting sufficient knowledge for logical generalizations, that will perhaps happen.

We may preliminarily denote the second case as the method of fictitious propositions. Its logical aspect approaches that of the first version, although the data for comparisons are taken not from the results of specially carried out experiments, but from observing processes that really occur independently from us. The examples below illustrate both cases which we can show in a general way by the following patterns. Imagine two graphical registrations of some wavy processes, A and B, with a striking coincidence of maxima and minima. After somehow measuring the degree of their closeness, we ask ourselves, how probable is it that such closeness had occurred accidentally. If that problem cannot be solved theoretically, we have two possibilities open. The first one is what we called a statistical experiment in its narrow sense. It will consist of

- 1) Constructing a theoretical pattern of an experiment capable of reproducing an unbounded number of specimens of random functions² whose statistical characteristics coincide in the mean with those of one of the processes;

- 2) Actually providing such specimens;

- 3) Superimposing those on the second process and approximately calculating the probability with which the observed closeness would have occurred purely randomly.

For being the most convincing, the second [version of the] method presumes that the graphical representation of one of the processes, A, is much shorter than that of B. Various intervals of B can then be applied in the same way as the experimentally reproduced curves were made use of in

the previous pattern; in other respects, the logical deliberations are absolutely the same.

Namely, in both cases we may think thus. Suppose there is no connection between A and B, then there would be no reason for the coincidence of the extrema. Had both curves [of the processes] been strictly periodic, had their period been 10 years (say), with observations only available for each year rather than registered continuously, they, the curves, could have only coincided once in ten years. That probability, however, should not be considered as an argument for the hypothesis of connection because our curves, as it usually happens, were discovered by the researcher among many other pairs and in all probability turned his attention to them owing exactly to their coincidence.

The investigation can be based otherwise. Our processes, as it occurs almost always, are not strictly periodic. The intervals between the [consecutive] points of maximum (and minimum) are sometimes longer, sometimes shorter and therefore, when superimposing curve A on curve B so that it coincides in the best possible way with another segment of B, the degree of closeness will generally change. And, when superimposing the first maximum of A not on the first, but on the second, the third, ... maximum of B, we will obtain a series of different degrees of closeness. Supposing that there really is no causal connection between the two processes, so that their coincidence is random, there will be no reason for that degree to be minimal in their actual arrangement rather than in some other imagined case. If, on the contrary, for example from all 20 fictitious possible comparisons the same tight, or even tighter closeness is observed in 10 of them as in reality, the hypothesis of connection will be abandoned, but if the real closeness will be the tightest of all the 20, it will testify for that hypothesis.

Here, the method is restricted by the length of the series of observations which can be increased by various means. Thus, it is possible to join the ends of a series and transform it into a ring with the first maximum coinciding with the last one. Or, to change the direction of time for one curve and superimpose AA' at first on BB', then on B' B. And, for attaining another goal, cut one of the series into intervals (from one maximum or minimum to the next one etc) and create many new series by randomly shuffling them (Marvin 1930). Here, we return to statistical experimentation in its narrow sense.

Once the problem is solved theoretically, the experimental method is certainly superfluous³. However, new unsolved problems will appear and a field for applying the statistical experiment is apparently secured for a long time if only we will learn how to use it in the most effective way.

I

1. On the Essence of Difficulties in Applying Statistics to Geophysics

Statistical series with which we have to deal in geophysics have a number of peculiar features hampering their stochastic treatment.

1) First of all, independence of "trials" occurs not at all frequently, consecutive terms of the series are usually mutually correlated. I have called such series connected. The formulas of the classical theory of errors are certainly inapplicable to them.

2) That circumstance leads to the waviness of the appropriate processes, often so distinct that “latent periodicities” are unintentionally suspected. In most cases this is an illusion, and even if there are periodic components, their discovery is much impeded by those pseudo-periodic components of the process⁴.

3) Geophysical processes are often non-stationary: their level and standard deviation, their correlation functions change in time. Those changes can occur either periodically, being for example connected with the seasons, or having the essence of secular movements. In either case they give much trouble not yet overcome by theory.

4) Even if rudiments of a theory are already created, a number of practical difficulties can occur. In case of connected series, all the usual formulas for estimating errors are replaced with other ones, much more complicated. Their application often demands cumbersome calculations; still, if the investigator resolves to go ahead, the results can be doubtful owing to the available series being insufficiently lengthy for ensuring reliable conclusions. This is why it is necessary to continue searching for new, more convenient solutions not forgetting about the methods of statistical experimentation but applying and developing them by practical animated investigations. See my related papers of 1927 – 1934 listed [in the Bibliography]⁵.

2. Examples of Mistakes

It is impossible to warn too often how usual are in geophysics mistakes occasioned by ignoring the peculiarity of the statistical structure of its processes. Examples can be multiplied ad infinitum, and I provide a few illustrations at random.

In essence, Baur (1928) published quite a good little book, but a number of formulas contained there cannot be applied to connected series without his even mentioning that fact. For example, he (pp. 23 – 24, 49 – 50) makes use of the Pearson formula for the square error⁶ of the correlation coefficient in a case in which it is not at all difficult, at least for one of the considered series, to reveal its connectedness.

One of the variables being compared there is the [set of] deviations (x_i) of November mean values of the difference in atmospheric pressure between [...] from the corresponding means for 1874 – 1923 (y_i). For providing an idea about the possibility of considering that our series lack connectedness, we will count the number of complete iterations [runs], i. e. sequences of numbers of equal signs ending on both sides by numbers of the opposite sign (the first element [term] being considered as being adjacent to the last one). The results are provided in Table 1. When comparing the actual and theoretical numbers (Bortkiewicz 1917, p. 85)

$$v_n = \frac{50}{2^{n+1}},$$

we see that the numbers of the first series (v_x) are behaving here as random and independent, but those of the second series (v_y) reveal a strikingly different picture. The mean length of the iterations in the series of (v_y) is $\lambda = 50/16 = 3.125^7$ whereas theoretically (Bortkiewicz 1917, p. 90)

$$E\lambda = \frac{1}{2pq}, \quad \sigma_\lambda = \sqrt{\frac{1-3pq}{4Np^3q^3}}.$$

Here, p and q are the probabilities of the cases of the first and second kind [?] respectively and N is the total number of cases. Suppose that $p = q = 1/2$, then, for $N = 50$, $E\lambda = 2$, $\sigma_\lambda = 0.2828$ so that the actual mean length differs from its theoretical value by exactly its fourfold standard error. It follows that Baur should not have applied here the formulas derived for independent trials⁸.

When considering now another work of the same author (Baur 1932, pp. 15 – 18), we find the same method applied to the problems directly interesting us. He provides coefficients of correlation between the solar constant and sunspots; yearly mean temperatures of tropical island [meteorological] stations, and, again, sunspots, and each time accompanies them by their “square errors” calculated by means of the Pearson formula. However, all the series he deals with are composed of terms certainly connected with each other, and his arguments founded on an inapplicable formula are obviously unjustified.

In our third example (Abbot 1922, pp. 183 – 184), we consider the author’s proof that the oscillations of the solar constant are real. He compares them with the changes in the distribution of radiation along the diameter of the solar disc; in the centre, the disc is brighter than at the edges (that *contrast* slightly changes from day to day, differently for each wavelength). Those changes cannot be attributed to the influence of the Earth’s atmosphere so that establishing a connection between the oscillations of the contrast and of the solar constant is tantamount to those latter also being “real”; i. e., it testifies to the changes in the Sun itself rather than to being only caused by atmospheric influences.

For my goal, it is not necessary to go into details; suffice it to say that Abbot puts forth a rather complicated hypothesis concerning the physical essence of the connection studied. He decomposes the complex of causes into three factors two of which act in one direction, and the third one, in the opposite sense. Depending on the prevailing direction, the contrast can augment either with the increase or decrease in the solar constant.

That hypothesis was certainly formulated for explaining the facts, and the power of the proof was only based on the comparison of the calculated correlation coefficients with their probable errors. Because [?], as the author states,

According to the theory of probability, such a high degree of correlation, positive or negative, lasting during several long intervals of time, especially in 1913 and 1916, would have hardly occurred randomly even once in several thousand cases.

The author then provides examples of correlation coefficients and their probable errors:

$$- 0.363 \pm 0.097 \text{ and } 0.601 \pm 0.067.$$

In accord with the above considerations, I am compelled to repeat that Abbot in vain swears on the theory of probability and that his arguments in

the form furnished are absolutely unjustified. And it is possible to go on drawing suchlike examples for an infinitely long time⁹.

3. On Some Other Mistakes in Applying the Correlation Method

I (1927b) studied in detail an example of a wrong application of the correlation method by Clayton (1923). This point is important, and I briefly indicate the essence of his main mistakes.

Clayton believes that the changes in the solar constant directly or otherwise cause temperature changes. A certain time period passes between act and effect, and, naturally, it can differ depending on various circumstances (latitude and geographical location in general, etc), so he calculates the correlation coefficients between that constant and the temperature at some stations each 0, 1, 2, ..., 15 days.

The magnitude of the time shift corresponding to the maximal in absolute value correlation coefficient indicates some geographic regularities, and the maximal coefficients are so large (of the order of 0.60 – 0.70 and larger) that the author does not at all doubt that the calculated coefficients are real. At the same time, the formula for the square error which he uses is inapplicable because of the waviness of the series. Another circumstance that should be especially stressed, and would have been important even for a correct formula for the errors of the correlation coefficients, is to be accounted for.

The point is that in connected series those coefficients themselves corresponding to differing shifts also constitute a wavy function. The square error of the correlation coefficient for a certain shift is not at all the same as for the maximal coefficient of some set of shifts. When observing the flight of two crows, one of them in Moscow, the other one in New York, is it not evident that, although the real correlation coefficient between their movements equals zero, the maximal coefficient will always be near unity if only a small number of flaps is considered and the series are suitably shifted one from another.

Even without having the proper formula for the error, our author could have come near to the correct result had he not restricted himself to correlating series for one year or a few years but rather made use of all his data and studied the alleged regularity over 10 to 15 years. As a check, it could have been possible to correlate the solar constant for a series belonging, for example, to a winter of some year and the temperature shifted by one, two or three years forward or backward.

I have done it and became convinced that the maximal correlation coefficients attain the same large values about as often when comparisons are both “absurd” and normal¹⁰. In spite of the coefficients of correlation being rather large, the power of their testimony for the existence of a connection was absolutely insignificant.

Even when only considering “normal” comparisons, it is seen that Clayton obtained large coefficients of correlation for roughly 100 days in 1917 only because he either accidentally came across a year favourable for his hypothesis, or selected it from a number of less (and certainly much less) favourable years. In geophysics, an investigation of material as complete or extensive as possible is even more necessary than in other branches [of natural sciences] characterized by a simpler inner structure of series.

Investigators should be warned against one more danger. Suppose we have series

$x_1, x_2, \dots, y_1, y_2, \dots, z_1, z_2, \dots, u_1, u_2, \dots, v_1, v_2, \dots, \text{etc.}$

If y, z, u, v, \dots are correlated one with another, they will often also correlate with some x . If the parallelism between the courses of the processes x and y provides an idea of connection, the parallelism between the x series and a more or less sufficiently considerable number of other series will certainly essentially strengthen our opinion. And this is what the comparison of the graphs of the courses of the solar constant and of the temperature for a rather considerable number of stations furnished by Clayton are doing.

However, that impression is illusory. Many processes taking place on our planet are, owing to more or less understandable causes, connected with each other and are going on roughly in parallel. A process taking place in the Sun, when being considered for a restricted period of time, coincides with each of them or with neither, the coincidence can sometimes be random, and, generally speaking, the number of coinciding series does not in itself augment the power of proof.

The geographical connection of some regions with correlation of the same sign should be quite similarly considered as a false, or at least as a doubtful argument. That fact is also based on the link of meteorological processes taking place over large portions of the globe so that the coefficient of correlation between some solar and terrestrial processes (concerning various stations) cannot be chaotically distributed over the Earth's surface. Without studying the stability of the regions of positive and negative correlation over time, nothing follows only from regularity over space.

A number of researchers (Clayton 1923; Helland-Hansen & Nansen 1917 et al) indicate that positive parallelism between a solar and a terrestrial processes at some station is often replaced by an opposite course, then, sometimes returns back, etc. It ought to be said that such a picture always speaks rather for an absence of connection than for its presence. If we deal with waves (on the Earth and the Sun) of approximately the same order, a parallelism lasting for a few of them can readily be accidental; and when non-coincidence starts, it will be natural for it, again, to last for a few periods. It would be different had we seen a transition from high positive to high negative values of correlation coefficients with a more or less sharp decrease of the portion of coefficients near zero, but I have not seen such cases in the literature.

Consider also that the chances of a random parallelism between processes strongly heighten when the two appropriate series were previously smoothed in the same way which is also usually done.

The indicated causes are already more than sufficient for extremely hampering both an appraisal of separate investigations of the discussed issue and a formulation of a sensible opinion about the degree to which they really confirm each other. A thorough criticism is needed for which a simple or even the most thoughtful reading is often absolutely not sufficient without checks by calculations, sometimes up to treating the data anew.

I allow myself to provide an example of a personal nature. The reading of Clayton, especially of his work (1923), resulted in their being very convincing. However, their critical study indicated a large number of doubtful links in his argumentation, but the measure of the revealed doubts remained absolutely unknown until treating the material anew made it

possible to conclude that at least one point in his proofs does not hold. What I was unable to recalculate even now remains only not worse than doubtful.

II

4. The Temperature over the Whole Globe and the Sunspots: an Example of Estimating the Reality of the Correspondence between Two Curves

A parallelism between two processes was by no means always established by calculation. More often, especially in the older literature, the curves were simply compared with each other by subjectively appraising the degree of their mutual correspondence, i. e., by experience, tact and feeling.

The rightfulness of such a method cannot be summarily denied if only its application can be properly restricted and checked by more rigorous means. Practically, however, it is often natural to suspect that in essence you are dealing not with a method, but with its absence. Moreover, it seems necessary to make every effort for discovering and developing methods and tricks for at least partly checking if not completely replacing subjective considerations and appraisals, and to regard all this as one of the most vital issues of geophysics.

I borrowed my next example from Helland-Hansen & Nansen (1917, p. 170), see Fig. 1. These authors obtained temperatures of the globe by issuing from mean temperatures over various regions (Köppen 1914 from data collected by Mielke, 1914). The authors themselves (pp. 184, 185) conclude that

The correspondence between these curves is striking (ist ja auffalend), and there is no reason to doubt the presence of the periodicity of the sunspots in the variations of the air temperatures of the terrestrial globe.

They calculated the coefficient of correlation between the indicated curves (but did not provide its value) and constructed a regression equation that allows to calculate the temperature from [data on] sunspots, and graphs of those calculated temperatures and of the deviations of actual temperatures from them.

According to our estimation, the calculated temperature curve and the curve of deviations, when being appraised by the naked eye, are characterized by an equal measure of fluctuation so that the correlation coefficient should be near to $\sqrt{0.5} = 0.707$ ¹¹ which rather corresponds with the impression from the two curves that are here correlated. For independent trials, the square error would have been $(1-0.5)/\sqrt{0.98} = 0.05$ ¹², so that the correlation coefficient being 14 times larger [$14 \cdot 0.05 = 0.70$] would have been quite justified. Even for $r = 0.5$ (in this case, it could not at all be less) the ratio would have been 6.6, also more than enough for proving the reality of the connection.

That approach is, however, banned owing to the lack of independence of the observations; also, for calculating the error of our coefficient according to the proper formula the series are insufficiently long, so that we will consider this point in the following way.

The two given series doubtless have approximately equal mean periods of oscillation which already somewhat testifies to the presence of a connection. However, it is extremely difficult to estimate the weight of this argument

since it indeed concerns only one of very many series studied by the authors mentioned. It would not at all be surprising if some of those series had a mean period near to that of the sunspots and we will therefore only apply that result for analysing rather than considering it as a proof of connection.

Fig. 1 shows us that the maxima and minima of temperature almost always either coincide or are very near to those of the sunspot curve turned upside down. Had the length of the mean period of one curve been only accidentally equal to the same length of the other curve, their observed mutual arrangement would not have at all followed. The coincidence of the extrema, if sufficiently tight, indeed testifies for a connection. Since we are dealing with their coarse structure seen by the naked eye and neither curve is at least strictly periodic, that coincidence means that the longer (the shorter) waves and half-waves of one curve correspond with the longer (the shorter) waves and half-waves of the other one.

It is necessary to estimate, at least approximately, to what extent is the actual mutual arrangement of the curves distinguished from the totality of all the other possible arrangements which, supposing that connection is lacking, can nevertheless be seen as equally probable. This idea can be more definitely realized by different means. Let us imagine that, for example, both curves are wound around a cylinder with the first minimum [of each?] coinciding with the last one. Then, turning one ring relatively to the other by one year, two years, etc, and each time calculating the squared sum of the distances between the corresponding extrema of the curves, we can determine for how many arrangements is that sum not larger than for the actual arrangement.

I have indeed done that. First of all, I determined the years when the sunspot curve had maxima by issuing from the data provided by Brunner (1930, p. 77). Since the mean yearly temperature should correspond on the graph with the middle of the appropriate yearly interval, I assumed the same epoch for the sunspots; for example, the year 1905 means the middle of that year, 1905.5. Therefore, when rounding off the decimal parts of the epochs provided by Brunner, the epochs 1883.9, 1889.6, 1894.1 etc became 1883.5, 1889.5, 1894.5 in his notation and 1883, 1889 and 1894 in ours.

This, then, is the origin of the first column in Table 2. The second column shows the years of extreme temperatures. Since Helland-Hansen & Nansen only provided a graph whereas we were mainly testing a method, we allowed ourselves to determine the epochs of the extrema more or less by naked eye. The introduction of some subjective moments was of course unavoidable since (Fig. 1), apart from the main extreme points corresponding to those of the sunspots, the temperature curve has additional smaller bends. For example, I am not sure whether rejecting a "loop" in the beginning of that curve and attributing the first minimum to 1915 was correct.

I have then proceeded thus. I prepared a band divided into equal intervals 5mm long and marked on it the extreme years of sunspots from 1816 to 1906. I did the same for the temperature curve, only the band was longer and the years, from 1815 to 1905, were repeated with the year 1905 once more shown as 1815. It is easily seen that these bands replace the rings on the indicated [imagined] cylinder.

The total number of all mutual arrangements is 90, or 180 if the left end of one band is [additionally] considered as its right end and vice versa. It is

not necessary to test all of them since the squared sum of the distances between the nearest in the mean corresponding points takes minimal numerical values 16 times in each of the two versions. All the other squared sums will be certainly larger, and we may disregard them.

Calculations are extremely simple. First, we determine the deviations of the extrema of the temperature curve from the same nearest points on the sunspots curve at its main (actual) arrangement. They are

$$\begin{aligned}\varepsilon_1 &= -1 (= 1915 - 1916); \varepsilon_2 = 0 (= 1923 - 1923); \\ \varepsilon_3 &= 1 (= 1830 - 1829), \text{ etc.}\end{aligned}$$

The mean deviation is $(1/17)\sum\varepsilon_i = 4/17 < 0.5$. We conclude that $\sum\varepsilon_i^2 = 39.5$ for that arrangement cannot be lessened by shifting the temperature curve one or two years in either direction. The next minimal $\sum\varepsilon_i^2$ will only be achieved by a more considerable shift.

Now we move the temperature band to the left so that its nearest extremum, 1823, coincides with the initial extremum of the sunspots band, 1816. We will have

$$\begin{aligned}\varepsilon_1 &= 0 (= 1823 - 1816 - 7)^{13}; \varepsilon_2 = 0 (1830 - 1823 - 7); \\ \varepsilon_3 &= -3 (= 1833 - 1829 - 7), \text{ etc.}, (1/17) \sum\varepsilon_i = -1.\end{aligned}$$

It follows that $\sum\varepsilon_i^2$ can be lessened by shifting the temperature curve by one year to the right [instead of the above]; all the ε_i will increase by a unity so that

$$\varepsilon_1 = 1; \varepsilon_2 = 1, \varepsilon_2 = -2, \text{ etc.}, \sum\varepsilon_i^2 = 36.5$$

which again is the least possible; it cannot be lessened by further shifting by one or two years in either direction. Had that sum exceeded or been equal to 39.5, see above, we would not be interested by its values at neighbouring arrangements. That is not, however, the case, and we calculate those values. They are 53.5 and 53, both larger than 39.5.

In that way, we find all the results shown in Table 2, and it is seen that in 7 arrangements including the actual out of 180 the squared sum of deviations [differences] between the corresponding extrema do not exceed its value for the actual arrangement.

I should certainly admit that the above calculations only provide some tentative rather than rational determination of the probability. Remembering that reservation, the "probability" that the observed tight proximity of the observed extrema of the curves could have occurred randomly is equal to $7/180 = 0.039$. For the normal distribution, such a probability corresponds to deviations 2.1 times exceeding the standard deviation and in such cases it is admissible to suspect that the corresponding result was not random, but that cannot be practically certain.

Our trick can be modified in different ways. Thus, the calculation of $\sum\varepsilon_i$ can be replaced by determining the coefficients of correlation between the two curves; or, in other words, by determining the number of mutual arrangements of the curves for which those coefficients would not be less than for their actual arrangement. And instead of the cyclic movement along

closed curves it is possible to apply the known data for the extrema of the sunspots from 1610 to 1933, i. e., for 324 years.

The temperature curve of the previous example with the time interval of 98 years could have then be superimposed on the first curve in 226 positions in each direction; when forgetting that for some interval of time the epochs for the sunspots were somewhat less precisely established and considering it insignificant that until 1749 we only know these epochs rather than the numbers themselves of the sunspots, that version seems to be even preferable to the previous.

5. The Arrangement of a Small Number of Events according to the Phases of the Solar Cycle

Here is the next example in which precise methods of calculating probabilities seems to be difficult to apply. According to data given me owing to the kindness of Prof. V. M. Obukhov, the years of the worst harvests in Russia/the European part of the Soviet Union for the period 1883 – 1931 were¹⁴

1891, 1906, 1921 ($\geq 25\%$ less than the mean harvest)
1889, 1911, 1920, 1924 (15 – 20% less)

After arranging those seven years according to the phases of the sunspot cycle, we have (see Table 3). Or, in words: *With one exception, those events took place not more than three years before, and not more than two years after a sunspot minimum.*

Direct calculation convinces us that between 1883 and 1931 there were 24 such years covering $24/49 \approx 1/2$ of the entire interval. It is not difficult to calculate the probability that not less than 6 events out of 7 will occur during those phases. Its approximate value can be obtained by neglecting that two bad harvests cannot happen in one and the same year. Then, supposing that $24/49 = 1/2$, we will have

$$P = 7 \cdot \left(\frac{1}{2}\right)^6 \left(\frac{1}{2}\right) + \left(\frac{1}{2}\right)^7 = \frac{1}{16}.$$

Such a probability certainly is not so low for being practically ensured that the coincidence is not accidental. However, it is sufficiently low for being a more or less weighty argument for non-randomness and it can be thought that the immediate impression provided by Table 4 on an unbiased mind corresponds to an instinctive estimation of the probability just calculated.

An unbiased mind ... I should have added *not versed in probability* since both the impression and the above calculation are in essence wrong [or irrelevant]. The point really is that not the hypothesis that *the years of very bad harvest mostly coincide with the period 3 years before – 2 years after a minimal number of sunspots* should be tested. That approach would have only been proper if the indicated phase were singled out before being acquainted with the relevant data. Otherwise, a somewhat different hypothesis should be put forward, namely that the years of bad harvest are arranged compactly on *some* part of the sunspot period; we ought to ask for the probability that at least 6 out of 7 events will occur during a continuous

stretch [during continuous stretches] of solar phases 24 years long out of the 49 (1883 – 1931).

Such a problem, especially after being properly formulated, can certainly be also theoretically solved. The time I devoted to this study did not, however, allow me to accomplish that, and I am therefore only providing an approximate trial solution of the problem. I denote¹⁵

...– a, + b/ = from the a-th year before a maximum to the b-th year after it
 /– a, + b = from the a-th year before a minimum to the b-th year after it
 + a/, – b = from the a-th year after a maximum to the b-th year before a minimum
 +a\, – b = from the a-th year after a minimum to the b-th year before a maximum

The above set of 7 years of bad harvest with one exception corresponds to .../– 3, + 2. When calculating the number of such years during 1883 – 1931, as we have done it, the compactness of that arrangement will be 24/49, and, for a hundred years, 1826 – 1924¹⁶, it will be 54/100. Our method consists in obtaining a number of random combinations of 7 years, determining the proper version of each [out of the four possible] and the compactness of the corresponding arrangement. There are cases which can be described in two ways; we will then choose the version of greater compactness.

We found random numbers by means of a booklet Tippett (1927) useful for any such experiments. We begin with the first row of [Tippet’s] Table 1, separate them [the random numbers] into consecutive pairs and write them down in their order if only they suit us, i. e., if they belong to the intervals (83, 99), (00, 15), and (20, 31)¹⁷. Unsuitable pairs are not considered and also rejected are pairs repeating two digits already included in the given seven digits. We have thus obtained 20 groups of 7 two-digit numbers each, wrote them down in a table similar to Table 3 above but extended to cover 5 years both before and after a maximum (a minimum). This allowed us to write down some years (i. e., some pairs of digits) twice; for example, the tenth “year” was the third before a minimum and, the second time, shown in brackets [no brackets there], the fourth after a maximum. It was thus easier to see which “year” can be rejected as an exception for the remaining six to compose a most compact combination and to find the correct version for it. I only provide the results (Table 4) rather than that [entire] table.

There, the appropriate versions for all the combinations are given as also the compactness of each, separately for 1883 – 1931 and 1826 – 1925; instead of 31/49 or 61/100 only 31 or 61 are entered. Our table tells us that cases as probable as the studied occurred 6 times out of 20 during the first period, and 7 times during the second. Now, 20 trials is certainly a small number, but sufficient for showing that the considered facts provide no grounds for the tested hypothesis.

We would have concluded otherwise when testing instead a hypothesis of bad harvests mostly occurring exactly during given phases. Indeed, out of the 20 experiments the version .../–a, +b appeared only once (No. 12, .../–3, 2). Such a rare occurrence, had it been confirmed by a larger number of trials, would have provided that hypothesis with a noticeable presumption. My next section will show that considerations testifying for the coincidence

of minimal harvests with minimums of solar spots, whose convincingness is difficult to deny, can indeed be provided.

6. Discussion of Semenov (1922)

Our attempt below to appraise Semenov's results (1922, p. 57) is apparently an instructive example of applying the methods here considered. These results seem at first surprising; they were rather met sceptically but no serious estimation of them was yet done.

Semenov studied the connection between sunspots and harvest. Having determined a suitable parabola of the second degree to represent the known curve of the yield of rye in European Russia compiled by Mikhailovsky, he separated the entire period (1801 – 1915) into five such intervals during each of which there occurred two maxima and two minima of sunspots. For each interval he constructed a parabola of the fifth degree most suitable in the sense of the method of least squares to represent the deviations of the harvest from the general level indicated above [by the parabola of the second degree]. The fifth degree was necessary for the parabola to have two maxima and two minima.

Semenov quite properly indicated that the arbitrariness involved in such decisions did not predetermine the results. Indeed, parabolas of the fifth degree can possibly have no extrema, or have them beyond the appropriate intervals, or, finally, even when having them within, they, the extrema, do not thus ensure any connection with the epochs of sunspots.

In all, during the studied years there were almost exactly 10 "periods" of sunspots with mean length near to 11 years. During the same time, the fluctuations of the harvest, after being adjusted as stated above, had 9 minima and maxima. Considering for the time being only the minima (we suspect that it is too risky to believe that the epochs of both are independent from each other) we established that the epochs of minimal bad harvests¹⁸ occurred in the following years, see Table 5.

Thus, all the 9 numbers are without exception situated on a comparatively narrow interval, [- 3 before a minimum of sunspots – + 1 after it]. Assuming the solar cycle equal to 11 years, which seems to be not largely erroneous, we conclude that only 5 years out of 11 are occupied. Now the problem admits a theoretical calculation of probabilities. Namely, similar to the previous example, we will ask not for the probability of such a concentration of random events around a minimum, but of same in some phases of the solar cycle.

Formulas and tables (Table 8 in the Supplement) provide $P = 0.00467 = 1/214$, the probability that 9 points randomly thrown on a circumference divided into 11 equal intervals will continuously occupy not more than 5, and fail to appear not less than in 6 such intervals. That probability is already so low that the hypothesis of a connection at once acquires an essential weight. Such a conclusion would have been wrong if a certain periodicity or clearly expressed pseudo-periodicity¹⁹ was noticeable.

However, our investigation (1930) had shown that the correlation function of the series of rye harvest in Russia is weakly pronounced, the correlation coefficients are small, and, what is the main point, they vanish over rather near distances. Coefficients between harvests separated more than by four years can be considered non-existent. Lacking this

circumstance, we would not have risked applying the considered method at all since it presumes independence of random variables.

We would have nothing to object to, had the answer above been against the hypothesis. However, since it was apparently confirmed, special considerations become valid. Namely, it seems natural to suspect whether it is really allowed to deal with the epochs of minimal harvests as variables independent from one another since their course is parallel to that of the sunspots minima which are although not strictly, but nevertheless more or less periodic. And we may recall, first, that lack of correlation is not identical with lack of any stochastic dependence; second, that the epochs of minima were determined by parabolas whose parameters were calculated by issuing from 22 – 26 consecutive harvests whereas any such connection of the terms of a series tends to lead to statistically connected results.

It is thus necessary to conclude that the described method of calculating the probability is here only very tentative. We consider the answer only as an indication for the mentioned hypothesis, as a demand not to reject, but deeply check it from many sides, rather than as a proof that Semenov was in the right.

Let us attempt to apply here the method considered in § 4, and more precisely its second version, briefly there indicated, but not tried out. Prepare two bands divided into intervals of equal length, 0.5cm each, mark the years and the special points interesting us. On the shorter band we will mark the maximal and minimal harvests according to Semenov; on the other one, maxima and minima of sunspots according to Brunner (1930) beginning with the minimum of 1610.8 and ending with the maximum of 1928.4.

The first band will be fixed, the second one, movable. For each mutual arrangement of these bands we determine the distance from Semenov's maxima (minima) to the nearest maxima (minima), or, in the second case, to the nearest minima (maxima) of sunspots. In both cases the squared sums of the distances, changing with every position of the movable band, take minimal values at some arrangements of the bands, and we search for them by means of the same trick as in § 4. The results are shown in Table 6.

The column T indicates the year of the movable band corresponding to the year 1804 of the fixed one and we thus determine the "natural" arrangement in the second row: year 1804, minimal squared sum of distances both in the first and the second case. We see that from the 201 (202) possible arrangements in the first (the second) case the least squared sum both times occurred when the bands were arranged synchronously. It is therefore difficult to imagine that the correspondence between harvests and sunspots discovered by Semenov was purely accidental. Here, our goals are mostly methodological, and we may stop now. A final check of the considered hypothesis is a problem of quite another kind.

Supplement

Suppose that some event is periodically repeated after t years. If the line of time, s complete periods long, be rolled up into a ring, we will obtain a circumference along which, beginning in any point, we can cover either direction exactly 9 times. After dividing the line of time into s , then further into t parts, we will call the obtained intervals arranged the same way with regard to their appropriate periods *phases of the same name*.

On the ring, they will be situated one on top of the other and the circumference will be divided into t parts. For the sake of clarity we imagine that each of these is divided into s “nests”. Suppose then that some event can occur once yearly, any year with the same probability. Two such events will be independent from each other and the problem is formulated thus.

Each of c points falls one after another on a circumference divided into t intervals each containing s nests, and occupies one of the free nests with the same probability independently from one another. Required is the probability $p_{m,c}$ that m consecutively situated nests will be left free after the fall of all c points.

Suppose that $p_{k,c-1}$ is known for all values of k and that $(c - 1)$ points are already on the circumference. For m consecutive nests to remain free after the fall of the last point it is necessary that a free interval not less than m nests long had still remained, and we are only interested in values

$$m \geq t/2 \quad (*)$$

so that if such an interval exists, it is unique.

If $(c - 1)$ points occupied $(t - m)$ nests, there will still exist

$$s(t - m) - (c - 1)$$

free nests, and the probability that the last point will fall in one of them and leave m nests free, will be

$$\frac{s(t - m) - c + 1}{st - c + 1} = 1 - \frac{ms}{st - c + 1}.$$

If the $(c - 1)$ points occupy less than $(t - m)$, that is, $(t - m - k)$, $k > 0$, nests, then the last point should fall on one of the two free nests for only m consecutive nests to remain free. The corresponding probability will be

$$\frac{2s}{st - c + 1}.$$

The probability for $(m + 1)$, or $(m + 2)$, ..., or $(t - 1)$ nests to remain free after the fall of the $(c - 1)$ points is, for $(c - 1) \neq 0$,

$$p_{m+1,c-1} + p_{m+2,c-1} + \dots + p_{t-1,c-1}$$

and we thus arrive at the recursion formula²⁰

$$p_{m,c} = \frac{s(t - m) - c + 1}{st - c + 1} p_{m,c-1} + \frac{2s}{st - c + 1} \sum_{k=m+1}^{t-1} p_{k,c-1}.$$

Note that if all the $(c - 1)$ points cannot be concentrated on one nest, i. e., if $(c - 1) > s$, then $p_{t-1,c-1} = 0$, etc. It is also obvious that if $m = t - 1$, the sum in the formula above vanishes.

Suppose now that $c = 1$, then $(t - 1)$ nests will certainly remain free, therefore

$$p_{t-1,1} = 1, p_{m,1} = 0 \quad (m < t - 1). \quad (**)$$

For $c = 2$ both points can occupy the same nest with probability $(s - 1)/st$ or the second one can occupy one of the two nests situated at an equal distance to the right or to the left of the first point with probability $2s/(st - 1)$.

Therefore

$$p_{t-1,2} = \frac{s-1}{st-1}, p_{t-k,2} = \frac{2s}{st-1}, k = 2, 3, \dots,$$

and, if $t = 2n + 1$ or $2n$, the last term of the series will be, respectively,

$$p_{n,2} = \frac{2s}{st-1}, p_{n-1,2} = \frac{s}{st-1}.$$

It is not difficult to establish that all these formulas except the last one can be obtained from the recursion formula with the values $(**)$ being substituted there. The last formula is an exception because, when deriving the recursion formula, we supposed that the length of the free interval obeyed condition $(*)$.

Now, if $c = 3$ or $m < t - 1$, we derive after some algebraic work, respectively,

$$p_{t-1,3} = \frac{(s-1)(s-2)}{(st-1)(st-2)}, p_{m,3} = \frac{6s[s(t-m-1)-1]}{(st-1)(st-2)}.$$

After that the formulas rapidly become complicated. For large values of c an approximate expression can be found, but, for not very large values, it is perhaps simpler to calculate directly the corresponding probabilities by applying the main formula consecutively. Thus, for $t = 11$ and $s = 10$ we find (Table 7) and then the probabilities (Table 8).

Explanation of Figure and Tables

Table 1 (§ 2). The “strikingly different picture” is provided by the sequence 6, 4, 1 (five times), 0 (three times), 1 whose terms indeed do not possess the pertinent property of random deviations. The Table provides theoretical and actual lengths of iterations (runs).

Figure 1 (beginning of § 4). Relation of number of sunspots with the temperature of the globe, 1810 – 1910. Two pertinent curves are shown (Helland-Hansen & Nansen 1917, p. 185, Figure 67).

Table 2 (middle of § 4). Years of extrema of both sunspots and temperatures rounded off as explained in text. Two positions of the movable band as explained in text were studied and the coincidence of the two bands estimated here by $\sum \varepsilon_i^2$.

Table 3 (inserted in the beginning of § 5 but mentioned in the middle of that section as “Table of the type of 3”). Explanation only in text and insufficient.

Table 4 (end of § 5). Explanation in text.

Table 5 (beginning of § 6). Shows deviations in years of “minimal bad harvests” from minima of sunspots, period 1810 – 1911.

Table 6 (end of § 6). Explanation in text.

Table 7 (end of Supplement). Explanation in text, insufficient. Provides magnitude $109 \cdot 108 \dots (111 - c) p_{m,c}$ for $m = 10, 9, 8, 7, 6$ and $c = 3(1)9$.

Table 8 (first mentioned in § 6 after Table 5, then at the very end of the Supplement). Provides probabilities $p_{m,g}$ and, as Slutsky formulated it,

$$P_{m,g} = \sum_m^{10} P_{m,g}.$$

Notes

1. I have only met a rudiment of a similar method in Marvin (1930, p. 490). See also his remarks about a paper of H. W. Clough (Ibidem, vol. 52, 1924, p. 439). I would be very grateful for indications about other examples of applying that method. E. S.

2. I do not know whether that term had appeared earlier, perhaps not. O. S.

3. For a modified pattern of the experiment, when composing the experimental series from pieces of the actual series randomly selected according to the extraction of tokens with replacement, a theoretical solution replacing the experiment would have in principle presented no difficulties. At present, it is not, however, possible to say whether that solution will be practically convenient. E. S.

4. Pseudo-periodic is now understood otherwise [xix, foreword]. O. S.

5. Some of my other contributions are devoted to issues connected with the continuity of stochastic processes, also really essential for geophysics. They do not, however, directly bear on the subject of the present article. E. S.

6. Also called mean square error. In several contributions Slutsky himself, see for example [iv, end of § 4] applied that more specific term. O. S.

7. Here and below, the large number of digits is certainly unwarranted; true, that had been a tradition followed by Gauss and Fisher (Sheynin 1994, p. 255n). O. S.

8. Now, however, we know that if one of the series is unconnected, and no correlation between them is supposed, the standard error of the coefficient of correlation will be expressed, as a first approximation, by the same Pearson formula, $1/\sqrt{N}$. This follows from the general formula (Slutsky 1930)

$$\sigma_r^2 = \frac{1}{N} \sum_{t=-\omega}^{\omega} \rho_x(t) \rho_y(t)$$

since all of its terms vanish except at $t = 0$ when

$$\rho_x(0) = \rho_y(0) = 1.$$

In the present case (Baur 1928, p. 38) $r = 0.69$; supposing that $\rho_{xy} = 0$, we arrive at a deviation five times greater than the standard deviation ($1/\sqrt{50}$) which is extremely unlikely. Concerning the other formulas applied by Baur, we do not yet know whether they are correct even if supposing that one serie is unconnected. Hardly they are. E. S.

9. I wish to stress that, when adducing some examples, I did not at all intend to discredit the work and merits of the pertinent eminent investigators. Mistakes are spread everywhere like an endemic disease being a chronic scourge. When everyone is sinning, the choice of examples cannot be just. E. S.

10. Examples of qualitative nonsense (spurious) correlation have been collected since mid-19th century. Slutsky provided an example above (the two crows) borrowed from his earlier book (1912, part 2, § 21/2009, § 31). O. S.

11. The fluctuation of a magnitude calculated from a regression equation is $\sigma_y^2 r_{yx}^2$ and that of the error, $\sigma_y^2 (1 - r_{yx}^2)$. Equating these expressions, we arrive at the result provided in the main text. E. S.

12. No explanation provided. O. S.

13. The 7 is a mystery; it does not stand for (1823 – 1816) as is evident in the case of ε_3 below. O. S.

14. Suchlike investigations covering periods with greatly differing social and economic conditions are meaningless. I especially note the famine of 1921 – 1922 and the great (and thoroughly concealed) famine of 1932 – 1933. Those years had been significantly left out. Slutsky perhaps hinted as much in the last lines of his § 4. O. S.

15. The notation is decidedly unfortunate. O. S.

16. No explanation provided. O. S.

17. Having no information about the harvest during 1916 – 1919, I omitted the pair (16, 19). E. S.

18. In Semenov's table minima are wrongly stated as 1854 and 1912, and in column U the year 1839 is accompanied by number – 0.02 (apparently a misprint) instead of 0.02. E. S.

19. See Note 4. O. S.

20. Explanations in this derivation were again insufficient, but I will restrict my comments to solving the discussed problem at least for $s = 1$ in quite another way. At first, the probability of m free places left can be determined by elementary combinatorial considerations (Feller 1950/1957, formula (11.7)). In Slutsky's notation

$$p_m(c, t) = C_m^t \sum_{i=0}^{t-m} (-1)^i C_{t-m}^i \left(1 - \frac{m+i}{t}\right)^c.$$

Condition (*) should be added; it was apparently necessitated by Slutsky's concrete data: 6 events should have occurred during less than a half of the sunspot period and $6 < 11/2$.

Then, for calculating the required probability, it only remains to multiply the formula above by r / C_r^m . O. S.

Bibliography

E. E. Slutsky

(1912), *Teoria Korreliatsii* (Theory of Correlation). Kiev. *Izvestia Kiev Kommerchskii Institut*, Book 16. Also published separately (Kiev, 1912). Translation: Berlin, 2009, also at www.sheynin.de Cf. [i].

(1927a, in Russian), Addition of random causes as a source of cyclic processes. *Voprosy Kon'iunktury*, vol. 3, No. 1, pp. 34 – 64.

(1927b), Sur un théorème limite relative aux series des quantités éventuelles. *C. r. Acad. Sci. Paris*, t. 185, pp. 169 – 171.

(1930, in Russian), On the square error of the correlation coefficient for homogeneous connected series. *Trudy Kon'iunkturnogo Instituta*, vol. 2 for 1929, pp. 64 – 101.

(1934), Alcune applicazioni dei coefficienti di Fourier all'analisi delle funzioni aleatorie stazionarie. *Giornale dell' Istituto Italiano degli Attuari*, anno 5, No. 4, pp. 3 – 50.

Other Authors

Abbot, C. G. (1922), In *Annals Astrophys. Obs. Smithsonian Instn*, vol. 4.

Baur, F. (1928), Korrelationsrechnung. *Math.-Phys. Bibl.*, Bd. 75.

--- (1932), Zur Frage der Realität der Schwankungen der Solarkonstante. *Meteorolog. Z.*, No. 1.

Bortkiewicz, L. (1917), *Die Iterationen*. Berlin.

Brunner, W., Editor (1930), *Astronomische Mitteilungen Eidgenossische Sternwarte Zürich*, No. 124.

Clayton, H. (1923), *World Weather*. New York.

Feller, W. (1950), *Introduction to Probability Theory*, vol. 1. New York, 1957. [New York, 1968.]

Helland-Hansen, B., Nansen, F. (1917), *Temperaturschwankungen des Nordatlantischen Ozeans und in der Atmosphäre*. Kristiania.

Köppen, W. (1914), Lufttemperaturen, Sonnenflecke und Vulkanausbrüche. *Meteorolog. Z.*, Bd. 31, No. 7, pp. 305 – 328, as quoted by Helland-Hansen & Nansen (1917).

Marvin, C. F. (1930), Are meteorological sequences fortuitous? *Monthly Weather Rev.*, vol. 58, No. 12.

Mielke, J. (1914), *Die Temperaturschwankungen 1870 bis 1910 in ihrem Verhältnis zu der elfjährigen Sonnenfleckenperiode*. Hamburg.

- Semenov, M. I.** (1922, in Russian), On the regularity of the fluctuations of harvests. *Vestnik Statistiki*, Book 11, No. 5 – 8, pp. 57 –.
- Sheynin, O.** (1984), On the history of the statistical method in astronomy. *Arch. Hist. Ex. Sci.*, vol. 29, pp. 151 – 199.
- (1994), Gauss and geodetic observations. *Ibidem*, vol. 46, pp. 253 – 283.
- Tippet, L. M. C.** (1927), *Random Sampling Numbers. Tracts for Computers*, No. 15. London.

XV

The Correspondence between E. E. Slutsky and V. I. Bortkevich

G. Rauscher, O. B. Sheynin, C. Wittich

Perepiska E. E. Slutskogo i V. I. Bortkevicha.
Finansy i Biznes, No. 4, 2007, pp. 139 – 154

Letter No. 1. Slutsky – Bortkevich, 20.7.1923, Kiev

Highly respected Vladislav Iosifovich!

I have received reprints of two of your papers (1918b; 1921) and hasten to thank you. You have veritably helped me: here, at the Institute for National Economy, I am delivering lectures on theoretical statistics and M. V. Ptukha had informed you how much we are lacking in the newest literature.

I wish to hope that, should I ask you to keep sending me reprints of your future contributions, I will not abuse your kindness too much. Incidentally, it is very important for me to have your paper (1920) since I am much interested in its subject. I myself have recently discovered an expansion of the hypergeometric series in terms of some parameters; regrettably, I do not know whether it is new.

I am sending you a reprint of my paper [iii]. Accept it as a token of the gratitude and deep respect from the sincerely devoted to you E. Slutsky.
17.7.1923

P. S. I am unable to resist the temptation to inform you about the solution of a modest problem. A biologist who prompted me asked about the probability of a random conjugation of two identical chromosomes arranged in a ring. It occurred that that problem was an elegant illustration of the law of small numbers².

Assume that $2s$ elements, s pairs aa, bb, \dots, ll of identical elements among them, are randomly arranged in a ring. Let $\Pi_{m,s}$ denote the probability that in each of some m pairs their elements will occur alongside and that the other elements will not be so arranged. Then

$$\Pi_{m,s} = \frac{2s}{m(2s-1)} \left(\frac{2s-(m+1)}{2s-2} \Pi_{m-1,s-1} + \frac{m}{2s-2} \Pi_{m,s-1} \right), \quad (1)$$

$$\begin{aligned} \Pi_{m,s} = & \frac{1}{(2s-1)(2s-2)} \{ 2[2s-(m+1)]\Pi_{m-1,s-1} + \\ & \{ [2s-(m+2)][2s-(m+3)] + 2m \} \Pi_{m,s-1} + \\ & 2(m+1)[2s-(m+3)]\Pi_{m+1,s-1} + (m+2)(m+1)\Pi_{m+2,s-1} \}. \quad (2) \end{aligned}$$

Each of these recurrence formulas allows us to calculate consecutively the probability sought by issuing from

$$\Pi_{2,2} = 2/3, \Pi_{1,2} = 0, \Pi_{0,2} = 1/3.$$

And, when increasing s unboundedly, we will find that

$\lim \Pi_{m,s} = \lim \Pi_{m,s-1}$, see formula (2),

$\lim \Pi_{m,s} = (1/m)\lim \Pi_{m-1,s}$, see formulas (1 and 2).

Consequently, if $\lim \Pi_{m,s} = \Pi_m$, $s = \infty$, then

$\Pi_1 = \Pi_0$, $\Pi_2 = (1/2)\Pi_1$, $\Pi_3 = (1/2 \cdot 3)\Pi_2$ etc, so that

$$\Pi_m = (1/m!)e^{-1}, \quad (3)$$

which is a particular case of the law of small numbers.

When issuing from (1) and (2), it is not difficult either to derive approximate expressions for Π_m , Π_{m-1} , and then for Π_0 as well, which show how rapidly are the probabilities tending to their limit. We have

$$\Pi_{m,s} = (1/m)\{1 + [(3 - m)/2s] + [(10 - 4m)/(2s)^2] + \dots\}\Pi_{m-1,s}, \quad (4)$$

$$\Pi_{0,s} = e^{-1} - (0.5518/2s) - [0.8124/(2s)^2] + \dots \quad (5)$$

I have not yet checked definitively the numbers thus occurring, but here is the approximate and exact results for $s = 6$:

$\Pi_{0,6}$	0.316	$3326/10395 = 0.320$
$\Pi_{1,6}$	0.382	$3948/10395 = 0.380$
$\Pi_{2,6}$	0.210	$2190/10395 = 0.211$
$\Pi_{3,6}$	0.069	$740/10395 = 0.071$
$\Pi_{4,6}$	0.015	$165/10395 = 0.016$
$\Pi_{5,6}$	0.002 ₃	$24/10395 = 0.002_3$
$\Pi_{6,6}$	0.000 ₂	$2/10395 = 0.000_2$
Σ	0.994 ₅	1

Regrettably, I do not know whether this problem had been also already solved³. Allow me to thank you once more: you cannot imagine how glad I am to receive your reprints.

Deeply respecting you E. Slutsky

Letter No. 2. Bortkevich – Slutsky, 31.7.1923, Berlin

Deeply respected Evgeny Evgenievich!

I am very thankful for the reprint of your report [iii] and your letter [No. 1]. I quite agree with you in that the theory of probability, being a branch of pure mathematics⁴, should be constructed absolutely independently from the logical problems connected with the notion of probability in its proper sense. I do not gainsay, however, that much may be expected from a change of the name. Your construction seems to adjoin that of Lange (1877) who had issued from the concept of disjunctive judgement (Disjunktionsurteil).

I was glad to perceive that you refuse to identify probabilities with limiting frequencies. You will find something relevant to this in my review (1923) of Keynes (1921) which I have sent you yesterday together with three other reprints. Regrettably, I have to ask you to return on an occasion that review as well as the paper (1920). My other paper (1921) can to a

certain extent serve as an ersatz of the second part of (1922a). [...] Owing to disagreements between the publisher and the printing establishment I did not receive any reprints of that paper.

In due time, I have sent Ptukha my paper (1918a). I regret that only one reprint is left. There, in Notes to pp. 108 – 110, you will find some remarks of principle which will perhaps be somewhat interesting to you⁵. Two books have recently appeared: Czuber (1923) and Urban (1923).

If you happen to see Ptukha, please thank him on my behalf for sending me four copies of his mortality table for Ukraine (1928). Your problem and its solution are very amusing. I do not know whether anyone dealt with it previously. At present, owing to lack of time I did not yet quite make out your initial formulas and their interrelation, and the coefficients of $\Pi_{m-1,s-1}$ in formula (1) are not completely understandable. As to formula (3), it can be derived easily and directly.

We have the expectation of m

$$s \cdot 2s \cdot (1/2s) \cdot [2/(2s - 1)] = 2s/[2s - 1].$$

For $s = 2$ and 6

$$2\Pi_{2,2} + 1\Pi_{1,2} + 0\Pi_{0,2} = 4/3,$$

$$0 \cdot 0.320 + 1 \cdot 0.380 + \dots + 6 \cdot 0.0002 = 1.091 = 12/11,$$

$$\lim [2s/(2s - 1)] = 1, s = \infty,$$

so that for a rare event whose expectation is unity to occur m times we indeed obtain (3).

I think that Prof. Mises will not refuse to publish a paper on that problem if only you send it to him.

Sincerely respecting you V. Bortkevich

I append Mises' address.

Letter No. 3. Slutsky – Bortkevich, 25.9.1923, Kiev

Highly respected Vladislav Iosifovich!

Please excuse me for being late with replying your letter, so kind and hearty. During the latest weeks I was busy over my head but wished to write you without hurrying. And I have only received your letter about four weeks ago after returning to Kiev from the countryside partly spending the summer there.

I cannot say how grateful I am for receiving your papers (1920; 1922b; 1923; 1921); all of them are extremely interesting for me. I am now going for three weeks on a scientific trip to Moscow to study in the libraries; upon returning, I will first of all make necessary extracts from those papers which you have asked me to return and send them back immediately.

I like very much your term *Disjunctive calculus*; it did not enter into my head, but now seems so natural. A *name* is certainly something secondary which does not at all mean unimportant. Not without reason so many events occurred only owing to the *iota*⁶. A name is a great thing as a mystic and metaphysicist would have said. Anyway, this probably is yet music for the

future although that transition of the calculus of probability about which I dreamt is perhaps not far off, and perhaps I will yet see yours *Disjunktionsrechnung* published⁷.

Concerning that problem about which I wrote to you, I will avail myself with gratitude of your good advice and send Mises my paper as soon as preparing it.

During summer I made 3000 trials changing conditions after each thousand. I was interested in finding out whether the size and the form of small objects, after being shaken together and arranging themselves in a ring, influence the equal probability of all the arrangements. I used a round box with a domed rise at its bottom so that my peas should have formed cyclically all by themselves. Before experimenting, I thought that the differences in size or form will be of no consequence, but it was not exactly so.

For greatly differing peas (two very small and round, two somewhat larger and quite flat, two still more elongated and roundish, and four almost spherical), deviations from theory were absolutely unquestionable (two experiments with a thousand peas of that composition). The third thousand, with peas of *roughly* the same form, although much more differing from each other in size than dice used in stochastic experiments, provided results remarkably coinciding with the theory.

In general, I think that for my arrangement of experiments the form and size of the small objects must indeed influence the results much less than the findings of those other kinds of experiments about which I was able to discover. Incidentally, I never heard or read about the use of automatic self-registering devices for *stochastic experiments*, but it seems to me that for them to be scientifically important their results should not depend on human patience, very cruelly tempted as judged by my own experience.

I had been shaking my box and inventing out of boredom a device that could have replaced me. It seems that I succeeded (certainly arriving only at its pattern). My device would be able to shake and count and it would not be apparently difficult to construct a suchlike device for the Buffon problem⁸.

As to the theory of my problem, I did not write down the derivation of the formulas not wishing to burden you by considerations possibly quite uninteresting for you. And now, I am also afraid of dragging out this letter. However, since you, as I understood, had wished to perceive how I obtained them, I venture to describe briefly the idea of the derivation; it would be too risky to send you the very long derivation in full.

Call a definite sequence of elements an arrangement; denote their total number out of s pairs by N_s and the number of those which include m pairs whose elements are adjacent, by $N_{m/s}$. Then

$$N_s = (2s - 1)!, \quad P_{m/s} = N_{m/s}/(2s - 1)!$$

Let $N'_{m/s}$ be the number of cyclical arrangements including a definite pair whose elements are taken in a definite order such as a_1a_2 or b_2b_1 or etc. Compiling a complete list of all the arrangements including a_1a_2 , then of those including a_2a_1 , then b_1b_2 etc, we will obtain $2sN'_{m/s}$ in which, as it is easy to see, each arrangement will occur m times. Therefore

$$N_{m/s} = (2s/m)N'_{m/s}.$$

Let us derive $N'_{m/s}$. Each arrangement including for example a_1a_2 can only belong either to $b_1a_1a_2b_2$ or $b_1a_1a_2c_1$. In the first case, after excluding aa , we will have, as before, an arrangement including m pairs with adjacent elements, but only $(m - 1)$ such pairs in the second instance. Out of each arrangement of $(s - 1)$ pairs with $(m - 1)$ of them with adjacent elements an arrangement of s pairs with m such pairs can be obtained by including aa in each interval except those between the adjacent elements, i. e., in $2s - 2 - (m - 1) = 2s - (m + 1)$ various ways.

Out of each arrangement of $(s - 1)$ pairs having m pairs with adjacent elements an arrangement of s pairs again with m such pairs can only be obtained by including aa in one of the intervals between the adjacent elements, i. e., by m various ways. Consequently,

$$N'_{m/s} = [2s - (m + 1)]N_{m-1/s-1} + mN_{m/s-1}.$$

Substituting this expression in the previous formula and including P instead of each N , for example, including $P_{m/s}(2s - 1)!$ instead of $N_{m/s-1}$ we will indeed derive formula (1).

The derivation of the other formula is so lengthy that I do not venture to repeat it here. In essence, I assume that all elements are taken in a definite order but included randomly and that $(s - 1)$ pairs are already thrown in the ring so that only one (a_1a_2 , say) is left⁹. I consider all four cases leading to m/s with $(m - 1)$, m , $(m + 1)$ and $(m + 2)$ pairs with adjacent elements, given those $(s - 1)$ pairs.

In each of those instances I determine how should a_1 and a_2 be situated for m pairs with adjacent elements to occur out of the s pairs. For example, in the last instance they should not fall alongside each other but each of them ought to occur in the interval between pairs with adjacent elements so that each of the two pairs will be thus separated and no new pair with adjacent elements will occur. The further derivation is not difficult.

Please excuse me for being too diffuse and be assured of my most sincere devotion.

Deeply respecting you E. Slutsky

Letter No. 4. Slutsky – Bortkevich, 24.2.1924, Kiev

Deeply respected Vladislav Iosifovich!

I was at last able to send you back those two papers (about Keynes and Laplace – Eggenberger) which you had asked me to return. Once more I thank you most heartily for sending them to me, but please do not blame me for having detained them: there was no way of sending them earlier.

In the same registered parcel I have also sent my meanwhile published papers [iv], 2 copies; [v], 5 copies; [...] the same in Ukrainian, [vi], 1 copy; [vii], 1 copy). [...] The last-mentioned work is a summary of the last-but-one and completes it by making use of unpublished material. I wrote [vi] at the request of my friend, Prof. L. N. Iasnopolsky, as a supplement to his own paper (1923). I had to compile it more rapidly than I wished and it turned out lengthier than necessary.

With M. V. Ptukha returning from Germany, vivid westerly impressions are disseminating; a few more threads are restoring the torn contacts. Books are appearing and we are ordering the volumes of periodicals going ten

years back. Thus, in a few months we will to a certain extent become Europeans.

Keynes interested me very much; when reading your paper, I was extremely glad to feel myself being at one with you. However, concerning a certain particular point: I would not have reprimanded Keynes of an *überraschend engherzige Auslesung der Ausdrucks "Form"* (of a sudden petty interpretation of the expression "form"), see Bortkevich (1923, p. 6).

Denote the statement about the number (m) of combinations of k elements of a given kind (A) taken n at a time by $F(m; n; k; A)$. Then F will be a logical function of those four variables, and you are certainly in the right about that function. However, when assuming definite values for three (say) variables and denoting

$$F(m_1; n_1; k_1; A) = f_1(A), F(m_2; n_2; k_2; A) = f_2(A),$$

f_1 and f_2 will then be different functions of A . In that sense Keynes is apparently in the right.

I would wish to talk to you about a subject that has been interesting me for a long time although I am yet unable to study it as deeply as necessary. Even when writing a review of Kaufmann [ii], I expressed the idea that each method is based on some theory, so that the statistical method is based on applying either the statistical theory or some other theor. science. I had chosen the first alternative, and now, after pondering over your *Iterationen* (1917), I do not feel myself wavering, and the more I think about it, the more I become convinced of the same.

Allow me to issue from your objections to the expressions *statistical physics*, etc. You indicate (p. 4) that physicists apply the designation *statistical* to such conditions in which no *actual* counting of elements is meant at all. But is that essential? A triangle remains a triangle both when we find and apply it in the empirical reality, and when we study it in the imaginary *reality*.

A physicist deals with physics both when experimenting and when solving an abstract problem formulated by hypothetical assumptions such as "Assume that (masses, forces, electrons) are given ...". The logical essence is obvious: the nature of the thing (*Wesen*) does not depend either on the existence or non-existence, or on admitting it in our judgement or premise. Thus, when a physicist says: Assume that n molecules with such-and-such velocities are given in some volume, etc, it means that their enumeration would have provided n items having some distribution according to a certain indication. If the actual counting is a statistical operation, then the imagined enumeration is also a statistical operation, only indeed imagined, just as an imagined murder or theft are murder or theft, only indeed imagined.

The designation of the operation can be transferred to the subject of study. In geometry, we study the forms of extents abstracting ourselves from the material; and similarly, in statistics, we study the numerical content of totalities or sets abstracting ourselves from everything determining one or another kind of things to which the counted items are belonging.

The subject of (theoretical) statistics is thus the numerical content (in abstracto), this being its constitutional subject with all the other subjects of study representing its logical derivatives (Husserl¹⁰). Totalities are studied by statistics since numerical contents always imply them. *But totalities are*

only studied concerning their numerical contents and the logical derivatives of the latter. Numerical structures as relations of the numerical contents of the whole and its parts; various means and relative numbers, are logical derivations of the *main notion* of statistics, that is, indeed, *numerical content* rather than totality.

Subdivisions of theoretical statistics are determined by further indications which could be conjugated with the constitutional indication and its logical derivations, but which leave the species of the elements comprising the totality indefinite. These indications are, in turn, *order, time and chance* and we thus obtain the following subdivision¹¹:

Statistics: 1. Sylleptics. 1.1. Sylleptics in its narrow sense. 1.2. Sylleptical kinematic? 1.1.1 Horistics. 1.1.2. Syntagmatics. 2. Stochastics.

I am not sure whether it would not be better to restrict the term *sylleptics* to its narrow sense and am unable to devise a term for *sylleptical kinematic* (I took the first ad hoc appellation that came to hand).

I would resolutely object to *Bevölkerungssylleptics* (sylleptics of population) since that term is not logically pure; *Bevölkerung* can only be applied as a terminus technicus, as British statisticians apply *population*. Even this is apparently not good enough because population, as they understand it, should not at all bear relation to time.

I do not consider the subdivision of stochastics. In the logical sense, it is clearly separated from the calculus of probability; and, once you admit your name, *disjunctive calculus*, it will also be separated in the terminological sense. Then the notions of stat. method, stat. technique, applied statistics (of population, fixed stars, etc) follow quite naturally from the concept of statistics as a theor. science that, as such, serves as a basis for a special method and, together with a number of applied disciplines, constitutes the foundation of a special technique, etc.

It is least of all clear to me what is it that justifies the separation between sylleptics and Mengenlehre (set theory). You mention this point [the separation] as something absolutely unquestionable; I, however, am regrettably unable to say the same about myself. I will be much obliged to you for somewhat explaining this to me even by hints.

One more consideration. If my point of view is rejected, I will insist to call the usual statistics (without including stochastic viewpoints) applied sylleptics. This seems to be the only logical attitude toward terminology. However, since the term *statistics* is thrown overboard, then, not to abandon habit, it can be applied as a common term for sylleptics and stochastics, and we thus have returned to the same conclusion.

Accept assurances of my deep respect and devotion. Yours E. Slutsky

P. S. I will gratefully avail myself of your advice to send my paper on the probabilities of cyclic arrangements of pairs of identical elements to Prof. Mises. The manuscript is quite ready, but it is necessary to wait for some more time.

Letter No. 5. Slutsky – Bortkevich, 24.7.1925, Kiev

Deeply respected Vladislav Iosifovich!

Allow me to thank you wholeheartedly for the sent reprints of the paper (1924). I intend to study it with great interest in summer. Although not

excusing my belated response to your kindness, I ought to say that all this spring and summer, until the latest day, I had been feverishly working at a rather large paper on the theory of probability¹² (1925) about six lists [1 list = 16 typed pages] long. I obtained a few new results to say nothing about the treatment of a number of problems from the viewpoint of the notion of limit in the stochastic sense apparently not devoid of some interest. As I learned later, the notion of st.[ochastic] limit is due to Cantelli, but of st.[ochastic] *asymptote* seems not to have been formulated by anyone.

Let the probability of condition A being satisfied be $P(A)$ and the distribution of the probabilities of a random variable¹³ x be a function of some independent variable φ . If, for ε as small as desired,

$$\lim P\{|x - f(\varphi)| \leq \varepsilon\} = 1, \varphi = \varphi_1, \varphi_2, \dots$$

then $f(\varphi)$ will be the stoch. asymptote of x , or

$$\text{as}_B(x) = f(\varphi), \varphi = \varphi_1, \varphi_2, \dots$$

and, in the particular case of $f(\varphi) = C(\text{const})$, $\lim_B(x) = C$ ($\text{as}_B = \text{asymptota Bernoulliana}$, $\lim_B = \text{limes Bernoullianus}$)¹⁴.

Suppose we have a series of independent trials with the probability of the occurrence of some event being $p_1, p_2, \dots, p_n, \dots$, and the number of the occurrences of that event in n trials being m . Then

$$\lim_B[(m/n) - p_{(n)}] = 0$$

or

$$\text{as}_B(m/n) = p_{(n)}, n \rightarrow \infty,$$

and $p_{(n)}$ is the arithmetic mean of $p_i, i = 1, 2, \dots, n$.

After having another look at Poisson [1837], I see that in Chapter 4, §§ 94 – 96, *his theorem* is expressed exactly by those equalities. He did not assume that the limit of $p_{(n)}$ as $n \rightarrow \infty$ does exist, nor did he subject the probabilities to any restrictions. Therefore, I am compelled to refuse to consider that theorem as a particular case of a mean probability of a constant composition¹⁵. To the best of my understanding, I am unfortunately differing from you here, as also in interpreting Poisson's general attitude to the law of large numbers, but it would be too long to write that down.

Apart from the German paper, I am also writing about that, and in much detail [viii], and I will have the pleasure to send you its reprints rather soon. I would like to inform you about a few theorems not to be published so soon.

Let

$$g_r(x; v) = \text{El}|x - v|^r \quad (r \geq 0).$$

Denote by C_r (central value, *Zentralwert*) the value of v corresponding to minimal g_r ; in general, I call that v *Bezugswert* (initial value).

1) If some even moment of $g_r(x; v)$, $r \geq 0$, has a fixed upper boundary when the number of trials (or, more generally, when the sequence of values

of the independent variable $\varphi_1, \varphi_2, \dots$) increases/continues indefinitely, and x has a stochastic asymptote, then all the moments lower than that r , centred around any stoch. asymptote, tend to vanish:

$$\lim g_{r-\alpha}(x; v) = 0, \varphi = \varphi_1, \varphi_2, \dots \text{ if } v \text{ is some stochastic asymptote}$$

and all the central magnitudes C_k ($k < r$) will be stochastic asymptotes. And if $r \geq 1$, the same will be also valid for C_r ; I do not know whether that persists for $0 < r < 1$.

2) It follows that if the condition

$$\lim g_r(x; C_k) = 0, \varphi = \varphi_1, \varphi_2, \dots (r > 0; k \leq r)$$

does not hold, and x has stoch. asymptotes, the moments higher than r ought to be infinite.

3) Therefore, if g_s is bounded from above at $s > r$ and g_r does not vanish, x cannot have stochastic asymptotes (and, correspondingly, does not obey the law of large numbers).

Applying Chuprov's notation¹⁶, we are known to have for the arithm. means

$$\mu_{2(n)} = (1/n)\mu_{[2;n]} + [(n-1)/n]\mu_{[1; 1; n]}.$$

If

$$\lim \mu_{[1; 1; n]} \neq 0, n \rightarrow \infty, \quad (1)$$

then, in the limit, the variance of $x_{(n)}$ does not vanish either. Excluding the case in which x only takes finite values, it cannot be, however, inferred that $x_{(n)}$ does not obey the law of large numbers since the condition

$$E[x_{(n)} - Ex_{(n)}]^2 \rightarrow 0, n \rightarrow \infty$$

is only established as being sufficient. My theorem proves that if

$$\mu_{4(n)} = E[x_{(n)} - Ex_{(n)}]^4$$

is bounded from above, for which, as it is easy to prove, it is sufficient that

$$\mu_{(4; n)} = (1/n) \sum_{i=1}^n \mu_4^{(i)}$$

has a fixed upper bound, then the random variable

$$x_{(n)} = \frac{1}{n} \sum_{i=1}^n x'_i$$

will not obey the law of large numbers (will not have stoch. asymptotes).

The question presents itself, however: will it not be sufficient for that conclusion if (a) $\mu_{[2; n]}$ is bounded from above and (b) the condition

$$\lim E|x_{(n)} - Ex_{(n)}|^{2-\alpha} = 0, 0 < \alpha < 2, n \rightarrow \infty$$

is impossible for at least one single value of α , for $\alpha = 1$, say, or for an arbitrary small α . I was, however, unable to solve it.

Be confident in my perfect respect and devotion. E. Slutsky

Letter No. 6. Slutsky – Bortkevich, 31.12.1925, Kiev

Deeply respected Vladislav Iosifovich!

Accept my apologies for sending you so belatedly my latest paper [viii]. This, however, happened mostly owing to alien circumstances. Since cherishing each rare contact with you, I read your letter with heartfelt joy. It was impossible to change anything in a work then being printed, but note that, when allowing myself to criticize your point of view, I am mainly issuing from your *Iterationen* (1917) rather than *Krit. Betr.* (1894 – 1896).

There (1917), as it seemed to me, a certain point of view was expressed with full clarity. However, in such difficult issues it is incredibly hard to find a quite suitable formulation, and I readily admit that certain nuances had escaped me.

Rest assured in my perfect respect and devotion. Evgeny Slutsky

Letter No. 7. Slutsky – Bortkevich, 16.5.1926, Moscow

Deeply respected Vladislav Iosifovich!

Your letter did not find me in Kiev and was forwarded to Moscow to me. I have moved here because of *some discords with the Ukrainian language*¹⁷. I wish to hope that you will generously excuse me for my so belated answer. In a new place, amid new duties, it was difficult to collect thoughts, then urgent tasks had occurred, etc.

I am a consultant at the Conjuncture Institute, work together with N. S. Chetverikov¹⁸ and for the time being am living at his place until getting the promised apartment. In addition, I was compelled to take up a *consultative job at Gosplan* [State Planning Committee]¹⁹. *I do not teach*. My situation and state are very unusual and seem transitional, and only God knows what will actually happen.

It is difficult and painful to write about A. A. Chuprov's death staggering us in spite of there lately being barely any hope. Chetverikov had certainly informed you already how our statistical family endured it and what we are supposing to do. Although I had not experienced the happiness of being close to A. A., I cannot forget his refined tact and incessant readiness to help with [my] scientific work.

With warmest gratitude I recall his attitude to my [future] paper (1925) which I sent him in 1923 as a very brief sketch asking his advice about its publication since being entirely cut off from foreign literature. Without his insistent advices I would have hardly transformed it into that more complete version now in press. However, only now, while having a look at his letters, I perceived the infinite delicacy with which he had avoided, in his critical comments, any hints at possible continuation and extension of the subject, at anything which should have suggested itself but which he did not wish to touch so as not to prompt me, to allow me to arrive independently at necessary conclusions²⁰.

Meanwhile, I have been thinking time and time again about the subject of our disagreement and especially so in connection with the proofs of my paper (1925) to appear in the next issue of *Metron*. There, Chapter 1 more

briefly covers roughly the same range of problems that is embraced in my Russian paper [viii]. [...] I was unable to introduce any essential changes since it proved impossible for me to study Poisson's text (1837) anew and perhaps to change my point of view. Concerning the second point of our disagreements, namely about my understanding of your concept of *mean prob. of a const. composition* [see Letter No. 5 and Note 15], I could have written much more, but am extremely afraid of abusing your attention. I will only say the following.

It seems that the problem has two aspects. A) How much do your conclusions about the cases covered by the *mean prob. of a const. comp.* extend into those instances [having to do with] what I call mean prob. of arbitr. composition. B) Are there any indications in your text itself that the author [Poisson?] allowed for the *mean prob. of arbitr. composition*?

In his last letter *Chuprov* wrote me in autumn [of 1925] that he *did not agree* with me, that he thought that the problem was solved by the expression of the mean square [error]. His remark is partly quite correct, but it is off the mark because of being covered by item (A). Indeed, the mean prob. of a const. composition in its narrow sense and the *mean prob. of an arbitr. composition* [see Note 15] have much in common especially when compared with the *mean prob. in its proper sense*.

The point is, however, as I am convinced, that it seems quite impossible to find out from your text (1894 – 1896 or 1917) that you had been stipulating and allowing for the case considered by me. A number of places in both sources just mentioned objectively contradict that possibility. I allow myself to indicate, for example, only one of them (1917, pp. 54 – 55):

Poisson, however, wished to construct a probability-theoretic pattern corresponding to real events, namely, to irregular changes of random causes. Nevertheless, the pattern of mean probability of a constant composition absolutely contradicts that [intention] since the values of the probabilities involved (p_n) included here [in the latter case] enter the mean in fixed proportions.

[*Galt es doch für Poisson, ein wahrscheinlichkeitstheoretisches Schema zu konstruieren, das dem wirklichen Geschehen, nämlich dem regellosen Wandel der zufälligen Ursachen adäquat wäre. Das Schema der konstant zusammengesetzten Durchschnittswahrscheinlichkeit ist aber das gerade Gegenteil davon: denn hier gehen die betreffenden Wahrscheinlichkeitswerte (p_k) in feststehenden Proportionen in den Durchschnitt ein].*

Fixed proportions mean that some restriction is imposed on the choice of the values of p_1, p_2, \dots . The logical sense of the phrase absolutely excludes the idea that those values are any whatsoever. Indeed, it is impossible to consider *fixed proportions* in case of a series whose terms are varying without being subjected to any rule, or, for example, to this rule:

1/10, 1/10, ..., 1/10 (m times); 1/2, 1/2, ..., 1/2 (m^2 times);
1/10, 1/10, ..., 1/10 (m^4 times); 1/2, 1/2, ..., 1/2 (m^8 times), etc.

And I am therefore allowing myself to think that the problem of the objective sense of a text, of *what* is objectively included in it and can be found there by any objective investigator, is beyond any doubt. The quotation above seems to be decisive. Please excuse my categorical expressions if I am mistaken, and you perceive that I am missing something. I always readily admit my mistakes, both in letters and publicly.

I am also sending you two of my papers. One of them is that which you had in due time so kindly helped me to find its place in the Mises' periodical (1926), – its turn occurred only now. The other one dates back (1915)²¹; its reprints sent to me during the [Great] War did not arrive, and only now have I acquired five copies for myself one of which I am sending to you.

I believe I was able to add something essential to Irving Fisher, Edgeworth and Pareto. I do not know when will I be able to return to those subjects if at all. It is all the more annoying since I have shelved almost completely prepared manuscripts, but *almost* is here the decisive factor.

During Easter I made a short trip to Kiev and fetched back my reprints so that now I am able to send all I can to Altschul²² about whom you had written me. I will do it with great pleasure but will be obliged to apologize for the impossibility of sending everything.

Sincerely devoted to you E. Slutsky

Letter No. 8. Slutsky – Bortkevich, 19.5.1926, Moscow

Deeply respected Vladislav Iosifovich!

I am allowing myself to add a few considerations to my previous letter since I wish very much to explain to you my idea as clearly as possible while leaving aside both Poisson and the entire history of the problem in general.

Suppose we have an unbounded sequence of urns with probabilities [of extracting a white ball] being

$$p_1, p_2, \dots, p_n, \dots$$

One ball is extracted from each; let the frequency (relat. Häufigkeit) of extracting a white ball in n trials be α_n . Then

$$E\alpha_n = (1/n) (p_1 + p_2 + \dots + p_n) = p_{(n)}.$$

When $p_{(n)}$ will be, so to say, the ideal norm for α_n ?

Suppose that s series of trials are made with the same first n urns in each trial, one trial with each urn, and that the frequencies obtained were

$$\alpha_n^{(1)}, \alpha_n^{(2)}, \dots, \alpha_n^{(s)}, \dots$$

Then, if you allow me to apply my notation,

$$\lim_B \left[\frac{1}{s} \sum_{i=1}^s \alpha_n^{(i)} \right] = p_{(n)}, s \rightarrow \infty.$$

This, then, is the pattern suitable for the idea of mean prob. of a constant composition. Here, $p_{(n)}$ is the stoch. limit not of $\alpha_n^{(i)}$, but of $\sum_{i=1}^s \alpha_n^{(i)}$ and for the independent variable s rather than for indep. variable n .

And here is another arrangement of the experiment. An unbounded series of trials is made with the same urns taken in turn and we consider the sequence of frequencies

$$\alpha_1, \alpha_2, \dots, \alpha_n, \dots, \alpha_N, \alpha_{N+1}, \alpha_{N+2}, \dots$$

Consider N as a variable magnitude, then

$$\lim_B [\alpha_N - p_{(N)}] = 0, N \rightarrow \infty, \text{ as } \alpha_B(\alpha_N) = p_{(N)}, N \rightarrow \infty.$$

This arrangement will indeed be the case which, as it seems to me, we ought to consider separately from the previous instance as the case of mean probability of an arbitrary (or arbitrarily variable) composition. We encounter the same case if two unbounded series of trials are made with the same series of urns providing frequencies α'_N and α''_N . Then

$$\lim_B (\alpha'_N - \alpha''_N) = 0, N \rightarrow \infty$$

however the urns are changed [replaced?] in the series.

Accept assurances of my perfect respect and sincere devotion. E. Slutsky

Letter No. 9. Bortkevich – Slutsky, 4.6.1926, Berlin

Highly respected Evgeny Evgenievich!

I have received both your letters of May 16 and 19 [NNo. 7 and 8]. Concerning the difference between the mean prob. in the proper sense and of a const. composition, I am keeping to my former opinion and do not find any contradictions in my writings. Indeed, I consider that difference in connection with the dispersion of statistical series, which is interesting so far as in the first case the measure of the variance²³, that is, the sum of the squared deviations of the number of occurrences [of the event] from its expectation [here, Bortkevich crossed out “the square of the mean square error”] = npq , whereas in the second instance

$$n \sum_{\lambda=1}^m g_{\lambda} p_{\lambda} q_{\lambda} (< npq), \tag{1}$$

where m is the number of *different* values of p_{λ} , and g_{λ} is propor. [?] in both series. Instead of (1) it is of course possible to write

$$\sum_{i=1}^n p_i q_i, \tag{2}$$

where n is the number of trials.

I applied the first of these two expressions to stress that the order of the probabilities p_i was of no consequence. If separate series of trials are not

connected by any conditions, we cannot even discuss any measure of variance. For expressions (1) or (2) to be measures of variance it is necessary for one series almost to generate the composition of the other one. If, on the contrary, we have N trials made with probabilities having nothing in common, and we separate that series into s parts of n trials each ($sn = N$), then the measure of variance will certainly be not $\sqrt{p_0(1-p_0)}$ which is not less but greater than $p_0(1-p_0)$. Poisson, however, does not consider that case since it means heterogeneous variance which is more general than the case of supernormal dispersion (Lexis 1879) that he did not touch [either]²⁴.

Regrettably, I ought to restrict my remarks to these previous comments since I have absolutely no time for a more detailed explanation of my point of view. I do not know when will I manage to read your paper (1915) although the problem there does interest me. Recently, I returned to it but considered it in a much less intricate setting.

I am very glad that you were able to place one of your investigations in Mises' periodical. He wishes to publish a note on the late Chuprov not longer than one page (two columns). [Bortkevich crossed out here: I allowed myself to name you since I thought that you will be highly successful.] And I would be very grateful for that. I hope that you will not refuse, and, according to Mises' wish, will submit the manuscript in the *nearest* future²⁵. And Bresciani will write [an obituary] for the *Giornale degli Economisti*. A friendship lasting thirty years connected me with the late A. A., and for me, each meeting with him was a festive occasion. It is difficult to become reconciled with the idea that he is gone. Thank Chetverikov²⁶.

Appended here are Bortkevich' draft calculations of variance made in connection with Slutsky's letter. In a covering text he mentioned his unpublished manuscript of 1914.

Letter No. 10. Slutsky – Bortkevich, 14.6.1926, Moscow

- Deeply respected Vladislav Iosifovich!

I will consider it my duty to write about A. A. Chuprov for Mises. I am informing him that my note [ix] will be ready not later than in a fortnight.

I have read your remarks about the subject of our discussion with greatest interest but I do not want to abuse your attention anymore. Perhaps it will be able for us to meet some day and discuss much. I wish very much that that will really happen.

If nevertheless your interests and pursuits will some day turn to the subject of my paper (1915), and you will glance at it, certainly you will not refuse to drop a line to me.

I would have now ended it in an essentially different manner. A supplement suggests itself here. Namely, for uniqueness (to an additive constant) of the definition of the function of utility it is not necessary to demand that on each hypersurface of indifference there exists a pair of such benefits that

$$\frac{\partial^2 U(x_1, x_2, \dots, x_n)}{\partial x_i \partial x_j} = 0.$$

It is sufficient to be able to draw a line cutting a number of such hypersurfaces along which the marginal utility $\partial U / \partial x_k$ remains constant,

and this is in principle always possible. This result can also be obtained by elementary considerations as I do in one not yet published manuscript where I consider the entire problem of the measurable in general and of the measurement of the so-called *subjective* value in particular [Slutsky 1927]. The tasks still to be done do not regrettably lead to the final elaboration of that manuscript.

I have already read your paper, – Nik. Serg. Chetverikov's copy, – with most vivid interest. The copy for me did not yet arrive, but will certainly reach me, and I am heartily thanking you [in advance] for it.

Devoted to you Evgeny Slutsky

Letter No. 11. Slutsky – Bortkevich, 29.9.1928, Frankfurt/Main

Deeply respected Vladislav Iosifovich!

After the Bologna congress²⁷ and a short trip over Italy, I came to Germany to stay here for about three weeks. Here in Frankfurt I learned that I may, provided you will not consider it immodest, congratulate you on the occasion of your sixtieth birthday and express my very best wishes. I would be very happy if you allow me to visit you when I come to Berlin. This will happen, as I think, in the middle of next week [...]. For me, to meet you in person would be greatly delightful.

Deeply respecting you, and sincerely devoted to you Evgeny Slutsky

Notes

1 (Note to Foreword). Mikhail Vasilievich Ptukha (1884 – 1961), demographer and historian of statistics. On Chetverikov see Notes 18, 26, Sheynin (1990, § 7.7) and [xix].

Nikolai Nikolaevich Volodkevich, or Nikolaus Wolodkewitsch, born 1888, was a brother of Slutsky's wife, Iulia Nikolaevna. He remained in Germany, and in 1932 earned a doctorate in physics at the Technical University of Darmstadt and later worked in the field of food technology and testing (in Turkey for a period in the 1930s, then again in Germany). German publications in his name appeared at least until 1959.

2. For his expression *law of small numbers* introduced by Bortkiewicz and then in vogue, read *Poisson distribution*. See Sheynin (2008).

3. Vilenkin (1969/1971, pp. 127 – 130) solved a particular case of Slutsky's problem for $m = 0$. After simple calculations, his answer for $s = 6$, given in another form, provide the same figures as Slutsky's table did.

Later Slutsky (1926) published the solution of this problem which he also discussed in Letters 2 and 3. In his paper, Slutsky named the biologist who prompted him to solve the described problem. His name (in German) was M. W. Tschernojarow, but the first who had considered the same problem was, as Slutsky believed, S. Navashin who had offered its solution in 1912, in a paper published by the Imperial Academy of Sciences (Petersburg). Slutsky, however, expressed reasonable doubts about the result of his predecessor. Slutsky's formulas from his letter to Bortkiewicz are repeated in his paper of 1926, but formula (4), which is there numbered (16), see p. 153, is corrected as is, rather insignificantly (see same page of the paper) his table.

In Letter 3 Slutsky several times wrote *combination* translated here as *arrangement* (*Anordnung* in his published paper).

4. Probability had indeed entered the domain of pure mathematics, but only after its axiomatization.

5. Those remarks specified the sense of equally possible favourable chances. In particular, Bortkevich indicated that *uniform* randomness not necessarily occurs when tickets are extracted from an urn and returned back.

6. At least, the letters *i* and *j* owe their origin to iota.

7. See Slutsky [iii] where he used this term (disjunctive calculus) and referred to Bortkevich.

8. The celebrated Buffon problem of 1777. A needle falls upon a set of parallel lines equally distant one from another; required was the probability of its intersection with one of

the lines. This problem decisively introduced geometric probability into the theory of probability.

9. Slutsky (1926, p. 151) supposed that chance only decided the arrangement of the last pair of elements.

10. Edmund Husserl (1859 – 1938), a German philosopher, founder of the philosophical school of phenomenology. C. Wittich, in Note 1 (p. 371) of his translation of Slutsky, see Chipman (2004), remarks that Husserl (to whom Slutsky did not refer there) juxtaposed primary perceptions and their interpretation by mind.

11. It was Bortkiewicz (1917, pp. 4 – 5) who (unsuccessfully) proposed the terms *Sylleptik*, *Horistik* and *Syntagmatik*, deriving them from the Greek. It is now generally known that, after referring to Jakob Bernoulli, he also reintroduced *Stochastik*. Already Wallis, in 1685, had applied the expression *stochastic* (iterative) *process* and Prevost & Lhuillier, in 1799, had used it in a probability-theoretic context (Sheynin 2009, Note 1 in Chapter 3).

12. The Russian term is theory of *probabilities*; here, however, Slutsky used the singular number.

13. The Russian term became *random magnitude*, which seems to be worse than its English counterpart. Slutsky, however, twice applied the term random variable (here and below in the same letter and in earlier contributions as well).

14. *Modo Bernoulliano* was an expression coined by Romanovsky in 1922 (Sheynin 1990/1996, pp. 50 – 51). Slutsky himself [viii, Note 3] mentioned Romanovsky in connection with the notion of *stochastic limit* (see above). There also, on his next pages, he explained the difference between it and the concept of limit in analysis and quoted a relevant although not altogether distinct (as he himself remarked) statement by Poisson. However, it was Laplace who expressly noted that difference in 1786 and, less definitely, in the beginning of Chapter 3 of his *Théorie analytique* (Molina 1930, p. 386).

Slutsky [viii, § 6] also explained that he adopted the term *stochastic asymptote* since the pertinent notion resembled the concept of asymptote in analysis as describing the behaviour of two functions.

15. See Bortkiewicz (1894 – 1896, 1894, p. 650). There also, on the next page, he introduced *mean probability in the proper sense*, see Letter 7.

16. In the sequel, Slutsky explained the meaning of the first two symbols whereas the last one, as the reader will see, can actually be left without explanation. For this reason, after unsuccessfully scanning Chuprov [Tschuprow] (1918 – 1919) and Chuprov (1918 – 1919 and 1921), we prematurely abandoned here our attempt at finding it.

17. Slutsky did not master the Ukrainian language sufficiently whereas a compulsory decree of the time stipulated it for all the lectures offered in academic institutions of that republic (Chetverikov [xix, beginning of § 6]).

18. Nikolai Sergeevich Chetverikov (1885 – 1973), Chuprov's student especially close to him. Worked in agricultural statistics, and on index numbers. Spent four years (apparently in 1931 – 1935) in prison as a *saboteur* and in 1937 or 1938 was subjected to new repressive measures and in any case was banned from living in big cities (Anonymous 1995).

19. Nothing is known about Slutsky's work there. Gosplan always was a highly prestigious institution.

20. This statement somewhat contradicts the previous description of Chuprov's advice concerning the same writing. Slutsky (1925, § 1.1, in a Note) had also publicly expressed his gratitude to Chuprov. There also, in another Note, he favourably remarked that Chuprov (contrary to Markov's opinion!) applied the term *random magnitude* (as it is called in Russian) "as the basis of the whole construction of theoretical statistics".

21. Slutsky (1915) is of course the paper on rational consumer behaviour on which Slutsky's fame in economic theory is based. It was published in Italian in one of the few European economic journals open at the time for contributions with mathematical content. In this work Slutsky developed further some ideas from his 1910 master thesis, as well as earlier contributions by Francis Y. Edgeworth (1845–1926) and Vilfredo Pareto (1848 – 1923). Slutsky's main achievement was to prove mathematically that under certain assumptions the consumer's reaction to a price change (*price-effect*) can be separated into two independent and additive effects: (a) an *income-effect*, related to the level of consumption and (b) a *substitution effect*, pertaining to changes in the structure of consumption. The so-called *Slutsky Decomposition* has become an integral part of every economics syllabus today.

Owing to its appearance in Italy in the middle of WWI, the essay remained unnoticed at the time – even the author, as this letters shows, received reprints only in 1926, and then only five. While one of these rare items went to Bortkevich, Slutsky sent another one almost simultaneously to Ragnar Frisch (1895 – 1973), the Norwegian economist (in 1969 the first laureate of the Nobel Prize in Economics) with whom he corresponded between 1925 and 1937 (this copy was recently found among Frisch’s papers in Oslo). Although both recipients were pioneers of mathematical economics, it took another ten years before Slutsky’s merits were finally recognized by various European and US-American scholars, who had derived the same results partially independently and who all were significantly involved in the further development of modern consumer theory, among them Sir John Richard Hicks (1904 – 1989) and Henry Schultz (1893 – 1938). Even then, the first translation of Slutsky’s paper into English did not appear until 1952, and the first Russian translation had to wait another decade until 1963. The story of the discovery and impact of Slutsky’s paper in Western economic literature during the 1930s is related in Chipman & Lenfant (2002).

Slutsky’s master thesis, *Theory of Marginal Utility* (in Russian) is kept at the manuscript section, V. I. Vernadsky National Library (Kiev), Fond I, No. 44850. Its Ukrainian translation appeared in Kiev in 2006 and an English translation is to be published. There, on p. 56, Slutsky’s letter of 27 March 1919 to the Rector of the Kiev Commercial Institute is reprinted stating that he submitted his article of 1915 in English.

22. Eugen S. Altschul (1887 – 1959), a scholar of Latvian origin. Chuprov (1922) mentioned him in passing in one of his reviews. In 1925 Altschul was living in Berlin and his main occupation was somehow connected with banking (Bortkiewicz & Chuprov 2005, Letter 199). In 1926, in a conversation, Chuprov (Letter 211) favourably referred to Altschul the statistician.

Altschul had remained in Germany after his studies in Freiburg, Leipzig and Strasbourg and a 1912 doctorate. After a long period of work in property administration, banks (see Chuprov’s remark above) and economic journalism, in Berlin in 1923 – 1926 (where Bortkevich might have known him), Altschul was in mid-1926 called to head the newly-founded Frankfurt *Gesellschaft für Konjunkturforschung*, where from 1927 he also taught conjunctural research methods at the university. Slutsky may thus have been asked to provide information about the Moscow Conjunctural Institute. Altschul was dismissed from his Frankfurt appointments after the Nazi seizure of power in 1933, emigrated to England in the same year (William Beveridge helped him to a research appointment at LSE) and then to the US, where he worked until 1939 at the National Bureau of Economic Research (supported by Wesley Mitchell, whose *Business Cycles* he had translated and published in German in 1931) and later taught at various universities, including U. of Minnesota and the University of Kansas-City, Missouri. He died in 1959 in Kansas-City. See Hagemann & Krohn (1999, Bd. 1, pp. 4 – 7).

23. Lexis (1879) had introduced dispersion of statistical series. Below, however, the variance (not dispersion) characterizes not a series, but the number of the occurrences of the studied event; p is the probability of its occurrence in a single trial, $q = 1 - p$, and n is the number of independent trials.

Bortkiewicz discussed the subject-matter of this part of his letter not only in 1894 – 1896, but also in his contribution (1917, §2.2). Concerning Bortkiewicz’ notation g_λ (below), Slutsky [viii § 8] explained it thus. A number of series of s trials is given. In each series trials having probabilities of *success* p_1, p_2, \dots, p_k are repeated s_1, s_2, \dots, s_k times and $g_i = s_i/s$. Bortkevich called the sum of the terms $p_\lambda g_\lambda$ the mean probability of a constant composition.

24. The initial Russian phrase was wrongly constructed and its translation is only conjectural.

25. Slutsky [ix] is his obituary of Chuprov that Mises had indeed published.

26. In 1908, C. Bresciani (Bresciani-Turroni) objected to Gini (Bortkiewicz & Chuprov 2005, Letter 88) who denied the law of small numbers. He then translated into Italian at least one of Bortkiewicz’ manuscripts on the same subject (Letter 91) which appeared in Gini’s *Giornale* in 1909. Later, he thought of reviewing Chuprov’s *Ocherki* (1909), see Letter 123 of 1913, and, finally, in 1925 he helped Chuprov to obtain a visa for travelling to Italy (Letter 210).

Concerning Chetverikov: he corresponded with Bortkiewicz and, in September 1926 (Bortkiewicz & Chuprov 2005, Note 178.2) informed him that Maria Smit (a notorious hard-liner) became the leading figure at the *Vestnik Statistiki* periodical, and he added: “The

conclusions are obvious". In other words: the era of obscurantism had *in general* set in. A Black Sun had risen, as Mikhail Sholokhov wrote somewhere on quite another occasion.

27. In Bologna, Slutsky participated in the work of the Congress of Mathematicians, see Chetverikov [xix, § 7].

Bibliography

V. I. Bortkevich (L. von Bortkiewicz)

For a rather comprehensive list of his works
see Bortkevich & Chuprov (2005, pp. 309 – 314)

(1894 – 1896), Kritische Betrachtungen zur theoretischen Statistik. *Jahrbücher f. Nat.-Ökon. u. Statistik*, Bd. 8 (63), pp. 641 – 680; Bd. 10 (65), pp. 321 – 360; Bd. 11 (66), pp. 671 – 705.

(1917), *Die Iterationen*. Berlin.

(1918a), Der mittlere Fehler des zum Quadrat erhobenen Divergenzkoeffizienten. *Jahresber. Deutschen Mathematiker-Vereinigung*, Bd. 27, pp. 71 – 126 of first paging.

(1918b), Homogenität und Stabilität in der Statistik. *Skand. Aktuarietidskr.*, Bd. 1, pp. 1 – 81.

(1920), Das Laplacesche Ergänzungsglied und Eggenbergers Grenzberichtigung zum Wahrscheinlichkeitsintegral. *Arch. Math. Phys.*, Bd. 20, pp. 37 – 42.

(1921), Variationsbreite und mittlerer Fehler. *Sitzungsber. Berliner math. Ges.*, Bd. 21, pp. 3 – 11.

(1922a), Die Variationsbreite beim Gaußschen Fehlergesetz. *Nord. Statistisk Tidskr.*, Bd. 1, pp. 193 – 220.

(1922b), Das Helmerzsche Verteilungsgesetz für die Quadratsumme zufälliger Beobachtungsfehler. *Z. f. angew. Math. u. Mech.*, Bd. 2, pp. 358 – 375.

(1923), Wahrscheinlichkeit und statistische Forschung nach Keynes. *Nord. Statistisk Tidskr.*, Bd. 2, pp. 1 – 23.

(1924), Zweck und Struktur einer Preisindexzahl. *Ibidem*, pp. 369 – 408, Bd. 3, pp. 208 – 251 and 494 – 516.

(1926), Chuprov. An obituary. *Ibidem*, Bd. 5, pp. 163 – 166. In Swedish. Translated in Chuprov, A. *Statistical Papers and Memorial Publications*. Berlin, 2004; also at www.sheynin.de

Bortkevich, V. I., Chuprov, A. A. (2005), *Perepiska* (Correspondence), 1895 – 1926. Berlin; also at www.sheynin.de

E. E. Slutsky

For an almost complete list of his publications see his *Selected Works* (1960).

Items in Russian are immediately detected

(1915), Sulla teoria del bilancio del consumatore. *Giornale degli Econ.*, vol. 51, pp. 1 – 26. Translated in 1952, see below.

(1925), Über stochastische Asymptoten und Grenzwerte. *Metron*, Bd. 5, No. 3, pp. 3 – 89.

(1926), Über die zufällige zyklische Anordnung paarweise gleicher Elemente. *Z. f. angew. Math. u. Mech.*, Bd. 6, pp. 150 – 159.

(1927, in German), A critique of Böhm-Bawerk's concept of value and his theory of the measurability of value. *Structural Change and Econ. Dynamics*, vol. 15, 2004, pp. 357 – 369. Cf. Chipman (2004).

(1952), On the theory of the budget of the consumer. In *Readings in Price Theory*. Editors, G. J. Stigler, K. T. Boulding. Homewood, Ill., pp. 27 – 56.

(1960), *Izbrannye Trudy* (Sel. Works). Moscow.

Other Authors

Anonymous (1995, in Russian), Jubilee dates and anniversaries. *Voprosy Statistiki*, No. 11, p. 77.

Chipman, J.S. (2004), Slutsky's praxeology and his critique of Böhm-Bawerk. *Structural Change and Econ. Dynamics*, vol. 15, pp. 345 – 356. On pp. 357 – 380 Slutsky's pertinent German paper of 1927 is translated by C. Wittich.

Chipman, J. S., Lenfant, J. S. (2002), Slutsky's 1915 article: How it came to be found and interpreted. *History of Political Economy*, vol 34, No. 3, pp. 553 – 597.

Czuber, E. (1923); *Die philosophischen Grundlagen der Wahrscheinlichkeitsrechnung*. Leipzig.

- Hagemann H., Krohn C.-D.**, Hrsg (1999), *Biographisches Handbuch der deutschsprachigen wirtschaftswissenschaftlichen Emigration nach 1933*, Bde 1 – 2. München.
- Iasnopolsky L.** (1923, in Russian), Our money circulation at the epoch of the revolution. *Mestnoe Khoziastvo* (Kiev), No. 2, pp. 3 – 37.
- Keynes, J. M.** (1921), *Treatise on Probability. Coll. Writings*, vol. 8 (the whole volume). London, 1973.
- Lange, F. A.** (1877), *Logische Studien*. Isorlohn, 1894.
- Lexis, W.** (1879), Über die Theorie der Stabilität statistischer Reihen. *Jahrbücher f. Nat.-Ökon. u. Statistik*, Bd. 32, pp. 60 – 98. Reprinted in author's *Abhandlungen*. Jena, 1903, pp. 170 – 212.
- Molina, E. C.** (1930), The theory of probability: some comments on Laplace's Théorie analytique. *Bull. Amer. Math. Soc.*, vol. 36, pp. 369 – 392.
- Poisson, S. D.** (1837), *Recherches sur la probabilité des jugements*. Paris. [Paris, 2003.]
- Ptukha, M. V.** (1928), *Smernost v Rossii i na Ukraine* (Mortality in Russia and the Ukraine). Kharkov – Kiev.
- Sheynin, O.** (1990, in Russian), *A. A. Chuprov. Life, Work, Correspondence*. Göttingen, 1996.
- (2001), Anderson's forgotten obituary of Bortkiewicz. *Jahrbücher f. Nat.-Ökon. u. Statistik*, Bd. 221, pp. 226 – 236.
- --- (2008), Bortkiewicz' alleged discovery: the law of small numbers. *Hist. Scientiarum*, vol. 18, pp. 36 – 48.
- (2009), *Theory of Probability. Historical Essay*. Berlin. Also at www.sheynin.de
- Tschuprow (Chuprov), A. A.** (1909, in Russian), *Ocherki po Teorii Statistiki* (Essays on the Theory of Statistics). Moscow, 1910, 1959.
- (1918 – 1919), Zur Theorie der Stabilität statistischer Reihen. *Skand. Aktuarietidskr.*, t. 1, pp. 199 – 256; t. 2, pp. 80 – 133.
- (1918 – 1919, 1921), On the mathematical expectation of the moments of frequency distributions. *Biometrika*, vol. 12, pp. 140 – 169 and 185 – 210; vol. 13, pp. 283 – 295.
- (1922), Lehrbücher der Statistik. *Nordisk Statistisk Tidskr.*, Bd. 1, No. 1, pp. 139 – 160 and No. 2, pp. 329 – 340.
- Urban, F. M.** (1923), *Grundlagen der Wahrscheinlichkeitsrechnung*. Leipzig.
- Vilenkin, N. Ya.** (1969, in Russian), *Combinatorics*. New York, 1971.

XVI

Autobiography [1939]

Zizneopisanie. *Ekonomich. Skola*, vol. 5, No. 5, 1999, pp. 18 – 21

My grandfather on my father's side, Makary Mikhailovich Slutsky, served in Kiev in the Judicial Department. He began his career already before the Judicial Reform [of 1864] and stood out against the civil service estate of those times because of his exceptional honesty. He died in poverty, but he had nevertheless been able to secure higher education for my father, Evgeny Makarievich, who graduated in 1877 from the Natural-Scientific Department of the Physical and Mathematical Faculty at Kiev University.

From the side of my mother, Yulia Leopoldovna, I descend from Leopold Bondi, a physician of French extraction who, together with others [?], moved to Russia under circumstances unknown to me. A part of his numerous descendants from two marriages established themselves as Russians. Thus, his son Mikhail, who joined the Russian Navy, was the father of the well-known Pushkin scholar S. M. Bondi. However, some of his children regarded themselves as Poles, and became Polish citizens after Poland was established as an independent state.

Soon after my birth my mother adopted Orthodoxy and, under the influence of my father, became an ardent Russian patriot in the best sense of that word and the Polish chauvinism of our relatives always served as a certain obstacle to more close relations. For about 30 years now, I have no information about these, absolutely alien [to me] representatives of our kin. After the death of my grandmother all the contacts between me and my relatives [in Russia] with them have been absolutely broken off¹.

I was born in 1880 in the village Novoe, former Mologsky District, Yaroslavl Province, where my father was a teacher and tutor-guide in the local teacher's seminary. In 1886, not willing to cover up for his Director, who had been embezzling public funds, he lost his job. For three years we were living in poverty in Kiev after which my father became the head of a Jewish school in Zhitomir. There, he had been working until his resignation in 1899, again caused by a clash with his superiors.

But then, in 1899, I had just graduated from a classical gymnasium with a gold medal and entered the Mathematical department of the Physical and Mathematical Faculty at Kiev University and earned my livelihood by private tutoring. In January 1901, I participated in a [student] gathering demanding the return to the University of two of our expelled comrades, and we refused to obey our superiors' order to break up. In accordance with the then current by-laws of General Vannovsky², I, among 184 students, was forcibly drafted into the Army. Student unrest broke out in Moscow and Petersburg and in the same year the government was compelled to return us to the University.

However, already in the beginning of 1902 I was expelled once more because of [my participation in] a demonstration against the Minister Zenger and this time prohibited from entering any higher academic institution of the Russian Empire. My maternal grandmother, whom I mentioned above, had helped me to go and study abroad. From 1902 to 1905 I studied at the

Machine-Building Department at Munich Polytechnic School. I had not graduated from there. When, in the fall of 1905, owing to the revolutionary movement in Russia, it became possible for me to enrol in a university in Russia, I entered the Law Faculty at Kiev University.

Munich was a turning point in my development. Circumstances imposed the machine-building speciality on me; it oppressed me, and, as time went on, I liked it ever less. I was forced to analyze my situation and discovered that my visual memory was very weak. Therefore, as I understood, I could not become a good mechanical engineer. And, by the same reason, I very badly memorized people by sight and mistook one person for another one even if having met them several times so that I was unable to be a political figure either. A further analysis of my abilities confirmed this conclusion. I studied mathematics very well and everything came to me without great efforts. I was able to rely on the results of my work but I was slow to obtain them. A politician, a public speaker, however, needs not only the power of thought but quick and sharp reasoning as well. I diagnosed my successes and failures and thus basically determined the course of my life which I decided to devote exclusively to scientific work.

I became already interested in economics during my first student years in Kiev. In Munich, it deepened and consolidated. I seriously studied Ricardo, then Marx and Lenin's *Development of Capitalism in Russia* [1899]³, and other authors. Upon entering the Law Faculty, I already had plans for working on the application of mathematics to economics. I only graduated from the University in 1911, at the age of 31. The year 1905 – 1906 [the revolutionary period] was lost since we, the students, barely studied and boycotted the examinations, and one more year was lost as well: I was expelled for that time period because of a boyish escapade. At graduation, I earned a gold medal for a composition on the subject *Theory of Marginal Utility*⁴. However, having a reputation as a Red Student, I was not left at the University and [only] in 1916/1917 successfully held my examinations for becoming Master of Political Economy & Statistics at Moscow University.

In 1911 occurred an event that determined my scientific fate. When beginning to prepare myself for the Master examinations, I had been diligently studying the theory of probability. Then, having met Professor (now, academician) A. V. Leontovich and obtaining from him his just appeared book on the Pearsonian methods, I became very much interested in them. Since his book did not contain any proofs and only explained the use of the formulas, I turned to the original memoirs and was carried away by this work. In a year, – that is, in 1912, – my book (*Theory of Correlation*) had appeared. It was the first Russian aid to studying the theories of the British statistical school and it received really positive appraisal.

Owing to this book, the Kiev Commercial Institute invited me to join their staff. I worked there from January 1913 and until moving to Moscow in the beginning of 1926 as an instructor, then Docent, and, from 1920, as an Ordinary Professor. At first I took courses in mathematical statistics. Then I abandoned them and turned to economics which I considered my main speciality, and in which I had been diligently working for many years preparing contributions that remained unfinished. Because, when the capitalist economics [in the Soviet Union] had been falling to the ground, and the outlines of a planned socialist economic regime began to take shape, the foundation for those problems that interested me as an economist and

mathematician disappeared. The study of the economic processes under socialism, and especially of those taking place during the transitional period, demanded knowledge of another kind and other habits of reasoning, other methods as compared with those with which I had armed myself.

As a result, the issues of mathematical statistics began to interest me, and it seemed to me that, once I return to this field and focus all my power there, I would to a larger extent benefit my mother country and the cause of the socialist transformation of social relations. After accomplishing a few works which resulted from my groping for my own sphere of research, I concentrated on generalizing the stochastic methods of the statistical treatment of observations not being mutually independent in the sense of the theory of probability.

It seemed to me, that, along with theoretical investigations, I ought to study some concrete problems so as to check my methods and to find problems for theoretical work in a number of research institutes. For me, the methodical approach to problems and the attempts to prevent deviations from the formulated goal were always in the forefront. In applications, I consider as most fruitful my contributions, although not numerous, in the field of geophysics.

I have written this in December 1938, when compiling my biography on the occasion of my first entering the Steklov Mathematical Institute at the Academy of Sciences of the Soviet Union. I described in sufficient detail the story of my life and internal development up to the beginning of my work at Moscow State University and later events are sufficiently well outlined in my completed form. I shall only add, that, while working at the University, my main activity had been not teaching but work at the Mathematical Research Institute there.

When the Government resolved that that institution should concentrate on pedagogic work ([monitoring] postgraduate studies) with research being mainly focussed at the Steklov Institute, my transfer to the latter became a natural consequence of that reorganization.

Notes

1. It had been extremely dangerous to maintain ties with foreigners, and even with relatives living abroad, hence this lengthy explanation. A related point is that Slutsky passed over in silence his work at the Conjunction Institute, an institution totally compromised by the savage persecution of its staff. O. S.

2. Vannovsky as well as Bogolepov mentioned in the same connection by Chetverikov in his essay on Slutsky (also translated here) are entered in the third edition of the *Bolshaiia Sov. Enz.*, vols 4 and 3 respectively, whose English edition is called *Great Soviet Encyclopedia*. It is not easy, nor is it important, to specify which of them was actually responsible for expelling the students. O. S.

3. Yes, it is possible that Slutsky read Lenin (and in any case it was necessary to mention him), but he probably also read Tugan-Baranovsky whereas Chetverikov [xix, § 1] mentioned Ricardo and “classics of theoretical economy”. See also [xv, Letter 7]. O. S. These latter were obviously representatives of the Austrian school. C. W.

4. This unpublished composition is kept at the Vernadsky Library, Ukrainian Academy of Sciences. A Ukrainian translation (Kiev, 2006) is available. O. S. An English translation will hopefully soon appear. C. W.

XVII

Autobiography [1942]

Zizneopisanie. *Ekonomich. Skola*, vol. 5, No. 5, 1999, pp. 21 – 24

I was born on 7(19) April 1880 in the village Novoe of the former Mologsky District, Yaroslavl Province, to a family of an instructor of a teacher's seminary. After graduating in 1899 from a classical gymnasium in Zhitomir with a gold medal, I entered the Mathematical Department of the Physical and Mathematical Faculty at Kiev University. I was several times expelled for participating in the student movement and therefore only graduated in 1911, from the Law Faculty. Was awarded a gold medal for my composition on political economy, but, owing to my reputation of a Red Student, was not left at the University for preparing myself for professorship. I passed my examinations in 1917 at Moscow University and became Master of Political Economy and Statistics.

I wrote my student composition for which I was awarded a gold medal from the viewpoint of a mathematician studying political economy and I continued working in this direction for many years. However, my intended [summary?] work remained unfinished since I lost interest in its essence (mathematical justification of economics) after the very subject of study (an economic system based on private property and competition) disappeared in our country with the revolution. My main findings were published in three contributions ([6; 21; 24] in the appended list [not available]). The first of these was only noticed 20 years later and it generated a series of Anglo-American works adjoining and furthering its results.

I became interested in mathematical statistics, and, more precisely, in its then new direction headed by Karl Pearson, in 1911, at the same time as in economics. The result of my studies was my book *Theory of Correlation*, 1912, the first systematic explication of the new theories in our country¹. It was greatly honoured: Chuprov [xviii, § 3] published a commendable review of it and academician Markov entered it in a very short bibliography to [one of the chapters of] his *Calculus of Probability*.

The period during which I had been mostly engaged in political economy had lasted to ca. 1921 – 1922 and only after that I definitively passed on to mathematical statistics and theory of probability. The first work [8] of this new period in which I was able to say something new was devoted to stochastic limits and asymptotes (1925). Issuing from it, I arrived at the notion of a stochastic process which was later destined to play a large role. I obtained new results, which, as I thought, could have been applied for studying many phenomena in nature. Other contributions [22; 31; 32; 37], apart from those published in the *C. r. Acad. Sci. Paris* (for example, on the law of the sine limit), covering the years 1926 – 1934 also belong to this cycle. One of these [22]² includes a certain concept of a physical process generating stochastic processes and recently served as a point of departure for the Scandinavian [Norwegian] mathematician Frisch and for Kolmogorov. Another one [37], in which I developed a vast mathematical apparatus for statistically studying empirical stochastic processes, is waiting to be continued.

Indeed, great mathematical difficulties are connected with such investigations. They demand calculations on a large scale which can only be accomplished by means of mechanical aids the time for whose creation is apparently not yet ripe. However, an attempt should have been made, and it had embraced the next period of my work approximately covering the years 1930 – 1935 and thus partly overlapping the previous period. At that time, I had been working in various research institutions connected with meteorology and, in general, with geophysics, although I had already begun such work when being employed at the Central Statistical Directorate.

I consider this period as a definitive loss in the following sense. I aimed at developing and checking methods of studying stochastic empirical processes among geophysical phenomena. This problem demanded several years of work during which the tools for the investigation, so to say, could have been created and examined by issuing from concrete studies. It is natural that many of the necessary months-long preparatory attempts could not have been practically useful by themselves. Understandably, in research institutes oriented towards practice the general conditions for such work became unfavourable. The projects were often suppressed after much work had been done but long before their conclusion. Only a small part of the accomplished during those years ripened for publication. I have no heart for grumbling since the great goal of industrializing our country should have affected scientific work by demanding concrete findings necessary at once. However, I was apparently unable to show that my expected results would be sufficiently important in a rather near future. The aim that I formulated was thus postponed until some later years.

The next period of my work coincides [began] with my entering the research collective of the Mathematical Institute at Moscow State University and then [and was continued], when mathematical research was reorganized, with my transfer to the Steklov Mathematical Institute under the Academy of Sciences of the Soviet Union. In the new surroundings, my plans, that consumed the previous years and were sketchily reported above, could have certainly met with full understanding. However, their realization demanded means exceeding any practical possibilities. I had therefore moved to purely mathematical investigations of stochastic processes [43; 44]; very soon, however, an absolutely new for me problem of compiling tables of mathematical functions, necessary for the theory of probability when being applied in statistics, wholly absorbed my attention and activity.

Such tables do exist; in England, their compilation accompanied the entire life of Karl Pearson who during three decades published a number of monumental productions. Fisher's tables showed what can be attained on a lesser scale by far less work. Nevertheless, a number of problems in this field remained unsolved. The preparation of Soviet mathematical-statistical tables became topical and all other problems had to be sacrificed. The year 1940 – 1941 was successful. I was able to find a new solution of the problem of tabulating the incomplete gamma-function providing a more complete and, in principle, the definitive type of its tables. The use of American technology made it possible to accomplish the calculations during that time almost completely but the war made it impossible to carry them through.

I described all the most important events. Teaching had not played an essential part in my scientific life. I had been working for a long time, at

first as a beginning instructor, then as professor at a higher academic institution having a purely practical economic bias, at the Kiev Commercial Institute, which under Soviet power was transformed into the Kiev Institute for National Economy. I had been teaching there from 1912 to 1926. The listeners' knowledge of mathematics was insufficient which demanded the preparation of elementary courses. I do not consider myself an especially bad teacher, but I had been more motivated while working as professor of theoretical economy since my scientific constructions conformed to the needs of my listeners. During a later period of my life the scientific degree of Doctor of Sciences, Physics & Mathematics, was conferred on me as an acknowledgment of the totality of my contributions and I was entrusted with the chair of theory of probability and mathematical statistics at Moscow State University. However, soon afterwards I convinced myself that that stage of life came to me too late, that I shall not experience the good fortune of having pupils.

My transfer to the Steklov Mathematical Institute also created external conditions favourable for my total concentration on research, on the main business of my scientific life. A chain of events, which followed the war tempest, took me to Uzbekistan. But it is too soon to write the pertinent chapter of my biography. I shall only say that I am really happy to have the possibility of continuing my work which is expected to last much more than a year and on which much efforts was already expended, – of continuing it also under absolutely new conditions on the hospitable land of Uzbekistan.

Notes

1. See translation of its Introduction [i]. The book is entirely translated (Berlin, 2009, also www.sheynin.de). O. S.
2. Its new version [42] was prepared on the request of *Econometrica*. E. S.

XVIII

Oscar Sheynin

Slutsky: Commemorating the 50th Anniversary of His Death

E. E. Slutsky: k 50-letiu so dnia smerti.
Istoriko-Matematich. Issledovania, Book 3 (38), 1999, pp. 128 – 137

Note. The original Russian text lacks §3.3 and only lists rather than quotes archival sources published in my booklet Sheynin (2004).

1. Introduction

Many authors (Kolmogorov 1948; Smirnov 1948; Allen 1950; Chetverikov [xix]; Gnedenko 1960; Youshkevich 1975; Konüs 1978; Seneta 1988)¹ described the life and work of Evgeny Evgenievich Slutsky (1880 – 1948), an outstanding mathematician, statistician, and economist, and his most important writings are available in a one-volume edition (Slutsky 1960). I am therefore restricting my main goal to publishing or describing a few archival letters either written by, or having to do with him (§3). In addition, I say a few words about Slutsky's life (below) and throw light on the events which apparently compelled him to abandon economics (§2)². In 1920 Slutsky became Professor at Kiev Commercial Institute. However, he had not mastered the Ukrainian language which was then made compulsory for academic institutions, and in 1926 he had to move to Moscow and to start working there at the Central Statistical Directorate [xix, p. 268], and, at the same time, at the Conjunction Institute under the Finance Ministry (Gnedenko 1960, p. 8).

Already then Slutsky busied himself in real earnest with applying his statistical research to geophysics. Being forced to abandon his activities in economics (§2), he [xix, § 2], for a few years,

*Went over to working in institutes connected with geophysics and meteorology where he [...] hoped to find application for his discoveries in the field of pseudo-periodic waves.*³

He had not found suitable conditions for theoretical research (Ibidem), and in 1934 he moved to the Moscow State University, then (in 1939) going over to the Steklov Mathematical Institute. The University conferred on him the degree of Doctor of Physical and Mathematical Sciences *honoris causa* [xix, § 2].

Slutsky was an original and deep researcher. He is mostly known as a cofounder of the purely mathematical theory of probability and the theory of stochastic processes, and remembered for his application of stochastic ideas and methods in economics and geophysics (especially in studying solar activity) and as a compiler of important mathematical tables which constituted “a masterpiece of the art of calculation” (Smirnov 1948, p. 417).

Slutsky's contribution to the theory of consumer's demand is very valuable (Allen 1950, p. 210). For a very long time before his death he (Ibidem, pp. 213 – 214) remained

Almost inaccessible to economists and statisticians outside Russia. [...] His assistance, or at least personal contacts with him would have been invaluable.

2. Withdrawal from Economics

In 1927, N. D. Kondratiev, the Director of the Conjuncture Institute, published a critical article concerning the first Five-Year-Plan. Soon he was elbowed out of science, arrested (1931) and then (1939!) shot (Makasheva 1988). N. S. Chetverikov, Kondratiev's assistant, served four years in prison, and, in 1937 or 1938, was subjected to new "repressive measures" (Anonymous 1995). Slutsky apparently had not suffered,⁴ but the general situation in statistics became unbearable. Later Chetverikov [xix, § 2] warily remarked that in 1930 "*The Conjuncture Institute ceased to exist and the Central Statistical Directorate underwent radical change*".

I myself add that, also in 1930, the leading statistical journal, *Vestnik Statistiki*, was closed down and only reappeared in 1948;⁵ during that period only a meagre number of statistical papers had been published in *Planovoe Khoziastvo*.

Under the changed social conditions, Maria Smit (more correctly, Falkner-Smit), a statistician of the new wave, became especially useful in spite of her crass ignorance (and in 1939 she was even elected Corresponding Member of the Soviet Academy of Sciences). Pearson, she (1934, p. 228) wrote,

Does not want to subdue the real world by a single curve [of distribution] as ferociously as it was attempted by Gaus [Gauss!]. [...] His system [of curves] nevertheless only rests on a mathematical foundation, and the real world cannot be studied on this basis at all.

She (1930, p. 168) also declared that Marxist statisticians should help the state security service in exposing the "saboteurs". Iastremsky (Ibidem, p. 153) effectively agreed and mentioned D. F. Egorov (who died soon afterwards in his exile in Kazan):

I had recently an occasion to hear out [...] the speech of Prof. Egorov, the then not yet exposed saboteur.⁶ He came out with a programme of sorts saying so ardently, even with a cry in his voice, What are you harping here on sabotage? [...] There are no saboteurs worse than you yourselves, comrades, since you standardize reasoning by popularizing Marxism.

Also see Sheynin (1990; 1998).

3. Archival Sources

Before adducing the promised letters to Pearson, I quote similar and already published archival materials concerning Slutsky in Kolmogorov (1948) and my own booklet (1990).

1) In three of his letters to Chuprov, Markov, in 1912 (Ondar 1977/1981, pp. 53 – 58) criticized Slutsky's book (1912). In the same source (p. 143) the Editor, in his review of the Markov – Chuprov correspondence, quoted a

passage from a letter written by Slutsky to Markov. I translated and published this letter almost in full, see below.

2) I myself (1990/1996, pp. 43 – 50) made known a few other archival or hardly known materials:

a) Chuprov's review of Slutsky (1912) published in 1912 in a newspaper, Slutsky's answer and his correspondence with Markov. I continue treating Slutsky's encounter with Markov in my §§ 3.1 and 3.2.

b) Slutsky's scientific character written by Chuprov in 1916.

c) Passages from the correspondence of these scholars with each other.

3) Seneta (1992) published English translations of two of Slutsky's letters to his wife concerning the author's appraisal of the comparative contribution of Borel and Cantelli to the discovery of the strong law of large numbers.⁷ After comparing Seneta's translation with Slutsky's letter now published in its original Russian (Eliseeva & Volkov 1999, pp. 116 – 118), I see that some hardly significant changes ought to be made there; the deviations were possibly present in Chetverikov's copy. I am now providing Seneta's translation (p. 29) of Chetverikov's covering note.

Dear Oscar Borisovich,

I enclose the promised letters of E. E. Slutsky transcribed from the originals and received by me for this purpose from his wife Iulia Nikolaevna Slutsky (née Volodkevich – in Kiev).

The lively controversy, which flared up between E. E. Slutsky and the mathematician Cantelli on the question of the strong law of large numbers, after examination of the opposing theses concluded in favour of E. E. who succeeded not only in defeating but also convincing his opponent. The whole episode so clearly characterizes E. E. – his relation to colleagues in his discipline, his meticulous consideration of all related authors and problems, his uncompromising rigour of scientific thinking – that it is best to let him speak for himself and present that graphic description of his dispute with Cantelli which has survived in letters to his wife – Iulia Nikolaevna – written in the very heat of the “battle”. From the letter of Thursday 6 September, 1928: [the text of the letter followed].

I am now quoting the materials mentioned above in Item 2.

2a) Chuprov's review:

In a short book the author described, shortly and distinctly, the theoretical constructions created by English statisticians-mathematicians and devoted to one of the most interesting problems of statistical theory, viz., to the measurement of the closeness of connection between phenomena. In developing Galton's methodological ideas, Pearson and his school gradually worked out an entire system of diverse and delicate quantitative characteristics of connection; [...] Slutsky had gained a good understanding of the vast English literature belonging to this subject and, upon becoming quite proficient in the material, described it intelligently. There is not much scientifically original in the book, but the author never aimed at anything of the sort. However, he successfully fulfilled his problem of compiling a manual [...] bringing together the findings made. Slutsky's book may be most energetically recommended to those Russian statisticians who possess at least some knowledge of higher mathematics. Even those, to whom

foreign literature is available, will not make a mistake by turning to this Russian manual for initial acquaintance.

On 22 Nov. 1912 Slutsky sent a letter to Chuprov:

I read your review [...] and express my most sincere and deep gratitude to you for it. To me, as a beginner in this field, it is extremely gratifying to realize that my first published work, in spite of all its shortcomings, may still be considered generally useful for the Russian public. [...]

Markov gave me a good dressing down. I received a letter and a postcard from him, and Professor Grave received three letters on the same subject with a request to show them to me. If my latest answer appears to him more or less satisfying, the correspondence, to all indications, will continue. Grave actively participates in the dispute, adjusting it, so to say (otherwise I would have been hard put to adapt myself to such an unusual manner of writing as possessed by Markov).

From the point of view of a rigorous mathematician it was of course easy for Markov to discover a number of weak points, but at the same time [his] resolute attack affected a number of fundamental problems in which I had to defend Pearson.

Dmitry Aleksandrovich Grave (1863 – 1939) was a mathematician then working at the Kiev Commercial Institute and one of his letters to Markov is quoted in § 3.1. Markov's "unusual" manner, blunt and often rude, is generally known.

Slutsky's letter to Markov of 13 Nov. 1912:

Highly respected Andrei Andreevich: // I have learned from D. A. Grave that you had written to me; however, I have not yet received the postcard which you had addressed to the Commercial Institute. Nonetheless, allow me to answer it the more so since D. A. Grave found it possible to acquaint me with the contents of your letter to him.

I begin by taking up [...]. This is a purely editorial shortcoming. [...] Your other remarks are so indefinite that it is too difficult to comment on them. My work was a result of studying the Pearson methods as described in his original memoirs. I experienced a direct impetus from Leontovich's book (1909 – 1911) (which is absolutely unsuitable for studying these methods) as well as from information reaching me about the awakening of certain statistical circles to the need of using these methods and the method of correlation in particular. I thought that I have no right to postpone the publishing of a contribution whose improvement was hindered by various personal circumstances and decided to restrict myself to a simple concise description; this, as it seemed to me, will help those statisticians who, either because lacking mathematical knowledge or of other reasons, are unable to read the original memoirs.

Your words that you do not understand the proofs [in my book] I can only interpret figuratively. I dare to believe that the mathematical competence of my work cannot be denied; incidentally, I am convinced of this since (according to my private information) A. A. Chuprov has recommended my book to his students. And so, I think that you consider my book obscure for the same reason that you believe that Pearson's works are anti-

mathematical. [Slutsky apparently repeated Markov's words from his letter to Grave.] The crucial point is obviously the lack of rigorous mathematical substantiation of the basis of [his] theories and methods.

In spite of my deep and sincere admiration for your knowledge and authority, I allow myself to differ. I believe that the shortcomings of Pearson's exposition are temporary and of the same kind as the known shortcomings of mathematics in the 17th and 18th centuries. A rigorous basis for the work of geniuses was only built post factum, and the same will happen with Pearson. I took upon myself to describe what was done. Sometime A. A. Chuprov will set forth the subject of correlation from the philosophical and logical point of view, and describe it as a method of research. An opportunity will present itself to a ripe mathematical mind of a pure mathematician to develop the mathematical basis of the theory.

My modest expectations will be satisfied if my work turns the attention of Russian mathematicians and statisticians to Pearson. Although I consider it possible to develop all the Pearsonian theories by issuing from rigorous abstract assumptions, I do not consider myself in a position to trouble you by imposing on you my opinion. In any case, I shall see it as my pleasant duty to discuss, either in private or publicly, any perplexities and answer any objections expressed in connection with my book.

When Nekrasov's book (1912) had appeared, I began to think that my work was superfluous; however, after acquainting myself more closely with his exposition, I became convinced that he did not even study sufficiently the relevant literature. Thus, I continue to believe that my book written without claims to originality (except for the additions to the Pearson theories in §§ 15, 18 and 33) is not superfluous for the Russian literature.

2b) A scientific character written by Chuprov:

Slutsky, a young and promising representative of mathematical statistics, addressed me with a request to formulate my opinion on his contributions in order to append it to his application submitted to the educational department [of Kiev Commercial Institute] for his approval as senior instructor.

*I know about three Slutsky's works worthy of attention. 1) Theory of corr. [1912]; 2) [A blank in Chuprov's document; he certainly meant Slutsky's paper of 1914 in the *J. Roy. Stat. Soc.*]; 3) *Sir William Petty* [of 1914]. Slutsky's English paper devoted to an important special problem of statistical theory attracted attention and caused debate in the literature in which Pearson, the head of the contemporary English mathematical school in statistics, took part [in 1916, in *Biometrika*].*

The Russian book [...] should be considered not only extremely useful, but, up to now, the only Russian treatise dwelling on these complicated problems with a full knowledge of literature and a quite correct understanding of the subject. The author thoroughly summarized a vast quantity of material chiefly of English investigations. Being scattered among special journals and mathematically expounded and thus demanding at times considerable study of extremely complicated branches of higher mathematics, these investigations are hardly available to most Russian statisticians. Slutsky treated everything he could find in the literature perfectly well, developed some topics on his own and intelligently, clearly

and coherently described the entire subject for Russian readers. This work testifies to a faculty important for a teacher, of converting dispersed materials of special scientific research into a harmonious system and the ability [no less important for a teacher] to describe clearly and systematically even extremely involved constructions. I often have the occasion to recommend the book to those of my listeners who possess some schooling in higher mathematics and I see how useful it is for them.

In their totality, both abovementioned works testify that in Slutsky's person Russian science possesses a serious force especially valuable since in Russia a researcher occupying himself in the field of social science rarely has mathematical training.

The contribution devoted to Petty [a historical study complete with translations of fragments from Petty's writings] is of a more modest nature. It is a good lecture based on a careful study of Petty's own work and an honest acquaintance with the literature. A good and wide schooling in economic theory and history of economic teachings allows [...] the author to sketch with a sure hand those historical and dogmatic [theoretical?] perspectives in which the teaching of Petty, that outstanding economist of the 17th century, should be considered.

2c) Even before 1916 Chuprova invited Slutsky to deliver a course of lectures on some subject, see below. The latter agreed adding, in his reply of 27 June 1914:

I construct my own course in mathemat. statistics at the [Kiev] Commerc. Inst. on a purely mathematical basis devoting about a month (12 hours plus classes) to an introduction to analyt. geometry and different. calculus; I teach the method of least squares in its interpolation aspect⁸ (at first without considering the theory of probab.). After that the theory of correlation goes on easily enough. [...] The experience of the previous year [...] which I carried out with five of my regular listeners satisfied me more than the experience of the year before that when I had started with the elementary, the easy. While having a few dozens of students in the beginning, I lost all of them when passing on to more dry material and exercises.

Somewhat previously, Slutsky wished to decline that invitation, so that Chuprova argued (20 June 1914):

During Lent, [...] courses on the theory of st-cs will be organized at Shaniavsky University in Moscow⁹ [...]. Nik. Serg. [Chetverikov] told me that you do not want to take upon yourself the reading of the course in corr. which we thought to entrust to you. Pity indeed! It fetters our plans. We cannot do without correlation – correlation and interpolation must certainly occupy the central place. And whom can we charge with the task? [...] Me? However, out of what will be left after deducing corr., namely interpolation, sample invest., some want to include for good measure stability [of statistical series], I can only take a part since I am unable to come to Moscow for more than a week. [...] Again, from your own personal point of view I do not know whether it is correct to decline this offer. Of course, it is better to not to delay with the master's exam. [The letter ends unexpectedly.]

Chuprov's correspondence with Slutsky during 1923 – 1925 testifies that these scholars remained close to each other. In particular, Chuprov then favourably mentioned two of Slutsky's papers. With regard to the first one he wrote Chetverikov on 9 March 1924:

For me, the work [iv] is very interesting; both in its approach and in the results obtained it fully accords with what I arrived at for the coefficient of correlation.

And, about the second one [viii], in a letter of 3 Aug. 1925 to Slutsky: "I consider your analysis perfect".

3.1. D. A. Grave – A. A. Markov, 4 Nov. 1912, Kiev

Archive of the Soviet Academy of Sciences, Fond 173, Delo 5, No. 1

Highly respected Andrei Andreevich, – I got to know E. E. Slutsky under the following circumstances. I was invited to a sitting of the Society of Economists at K. Comm. [Kiev Commercial] Inst. to attend a report on applying the Pearson theory to statistics. The report was delivered by Slutsky, a young man who had recently graduated from the [Kiev] University with a gold medal awarded for a work on political economy, but, because of some reasons, was not left at the University [to prepare himself for professorship].

I inquired directly of Slutsky's professor of political economy the reasons for this, and his answer surprised me by the justification unusual for a mathematical ear. According to his words, Slutsky is quite a talented and serious scientist, but the professor had not ventured to nominate him for being left at the University because of his distinct sympathy with social-democratic theories. And when I was unable to refrain from stating that at the mathematical faculty the author is not usually asked about his political views, the professor advised me to leave Slutsky at the mathematical faculty. I was naturally obliged to say that I have absolutely no desire to intervene in the business of the law faculty and that I am therefore asking him to leave the mathematical faculty alone. After this encounter Slutsky became my student and protégé. Although I am not at all acquainted with his works and had not understood the mathematical part of his report.

The lawyers, professors at the K. Comm. Inst., who did not understand Slutsky's book (1912) but desired to acquaint themselves with the Pearson theory, have asked me to explicate it properly in my course in insurance mathematics (1912). I do not know how to find a way out of this difficult situation: it is simply repulsive to read all this

[The sequel has no bearing either on Slutsky or probability and/or statistics. As also below, I myself inserted or specified the bibliographic information provided. For Grave, it was "repulsive" to read Pearson; cf. the now published letter of Slutsky to Markov (below).]

3.2. The Extant Part of the Unsigned and Unaddressed Letter
(obviously, from Slutsky to Markov; no date)

Same Archive, Fond 173, delo 18, No. 5

are not independent in magnitude from the sum of the already accumulated deviations or that the probabilities of equal deviations are not constant, we shall indeed arrive at the formula

$$(1/y)dy/dx = x/F(x).$$

In an infinite number of cases (naturally, not always!) F can be expanded into a Taylor series, and the first few (e.g., three) terms will ensure a sufficient approximation. These qualification remarks should have certainly been made.

Only experience can show how often do empirical polygons of distribution, which could with a sufficient approximation be interpolated by a Pearson curve, appear in practice. Much material is already collected for answering this question in the positive. In many cases the Gauss curve will not do since asymmetric polygons are often encountered in practice. Interpolation by parabolic curves

$$y = a_0 + a_1x + a_2x^2 + \dots$$

is unsuitable since these curves do not give an adequate picture at the edges of the figure: it is impossible to ensure their asymptotic approximation to the X axis; in addition, they lead to many superfluous inflexions. Pearson curves constitute the type that occurred to be practically the most suitable.

Since the Gauss curve in very many cases is well suited for representing statistical facts, especially in anthropology [anthropometry], it seems desirable also for the asymmetric Pearson curves not only to indicate that they are corroborated by practice, but in addition to provide a theoretical derivation that would put this curve [these curves] in the same line as the Gauss curve on the basis of the theory of probability (hypergeometric series).

The derivation on pp. 16 – 17 only serves to make the striking practical suitability of these curves less incomprehensible by means of the hypothesis on the action of infinitely many causes combining semi-randomly one with another.

2) The method of moments. Here, I allow myself to remark that neither Pearson, nor Lakhtin (1904) say that they proved that the method of moments brings

$$\int (y - Y)^2 dx$$

to its minimal value. They only prove that the method ensures an approximation. It would have been interesting to investigate this problem and to indicate precisely when is the method of moments applicable, and when it is not. Lakhtin does it, but is he not mistaken?

I think that, quand même, approximate formulas should not be objected to. Indeed, you yourself (Markov 1908, p. iv) admit that such formulas might be used in probability theory even “without estimating their error” since “the aims of applied mathematics” demand this. You also state that approximate formulas should in addition be created for ensuring the calculations (Markov 1912, p. 77).¹⁰ At the same time, the method of moments is very convenient; and, since it is proved to provide an approximation for a large number of types of functions, its critical investigation is desirable. In many cases it is simply indispensable since the method of least squares sometimes leads to intolerable or even unrealizable calculations. If desired, I shall next time illustrate this proposition.

3) *The theory of correlation. Here, I shall allow myself for the time being ...*¹¹

3.3. Slutsky's Letters to Karl Pearson

I (Sheynin 1990/1996, pp. 46 – 47) published Slutsky's letter of 31 March 1913 to Chuprov. It occurred that Slutsky sent Pearson two manuscripts for publication in the *Biometrika*. Pearson had, however, returned both of them, and Slutsky, considering that he was treated improperly, asked Chuprov's advice. Chuprov recommended that Slutsky submit his work to the Royal Statistical Society, and one of these manuscripts was indeed published by it (1914); the other one, on a modification of the difference method, had not appeared anywhere.

Now, I am able to make known three letters from Slutsky to Pearson;¹² Pearson's letters are lost. Slutsky invariably gave his address as the Volodkevich Commercial "Schoole" in Kiev. Volodkevich was the name of his wife, and I am sure that since 1917 Slutsky never mentioned this private enterprise of his father-in-law.

3.3.1. Slutsky – Pearson, 23 April 1912

University College London, Library, Pearson Papers 856/4

Dear Sir, – I am sending for your approval a paper concerning a correction to be made in the theory of contingency. If you find no fallacy in chief results, will not the paper be of some interest to the readers of the Biometrika? [!] Should you find any fault making idle the whole of my reasoning, I hope you will not refuse to communicate me your kindly criticism. It is a pleasure to acknowledge beforehand my great debt to you for the slightest of hints on the fallacies possibly made in my work. I am,

Yours faithfully E. Slutsky

P.S. The summary of the results is to be found at the end of the paper.

3.3.2. Slutsky – Pearson, 6 May 1912

Kept at the same place, 856/7

Dear Sir, – I had the pleasure to receive your honored letter on the 3^d May and I must excuse myself for answering so late – the reason is that I wanted much time for translating my letter in English. I thank you very much for your long and very interesting letter and for the proof which I am sorry not to have got yet, probably because it must be censured before I get it. Being you really very thankful for your suggestiv and very valuable criticism and agreeing with you in many points, I fear nevertheless that I shall not be able to agree with you about their bearing concerning my main thesis. I think I can keep my ancient opinion about the best method of determining the probability we have in view, though after your letter I feel compelled to change its foundation. I take the liberty to begin with some general considerations and then I shall continue with the question in which we disagree.

1. There is not a single method for the determination of the probability that a given system of frequencies has arisen from random sampling.

A) The theoretical frequencies being known à priori, we can determine the probability of the given system of errors:

$$e_1 = m_1 - \mu_1, e_2 = m_2 - \mu_2, \dots P = Q(\chi_{\mu}^2; n') -$$

in the notation of my paper – where n' is the number of groups,

$$\chi_{\mu}^2 = \Sigma[\mu R_{ii}e_i^2 / \mu R \sigma_{\mu_i}^2] + 2\Sigma[\mu R_{ij}e_i e_j / \mu R \sigma_{\mu_i} \sigma_{\mu_j}],$$

$$R = \begin{vmatrix} 1 & r_{\mu_1\mu_2} & r_{\mu_1\mu_3} & \dots \\ r_{\mu_2\mu_1} & 1 & r_{\mu_2\mu_3} & \dots \\ \dots & \dots & \dots & \dots \end{vmatrix}. \quad (1)$$

Now it is to be remarked that the method, even when applied to the same material, gives us very different results, the value of n' being arbitrary. As you have shown (Pearson 1900, p. 160), by infinitesimal grouping $P = 1$ for any value of χ^2 will appear. There is thus a number of groups n'_m which brings the value of P to the minimum, and I think you will agree that this minimal value of P is that really significant for the probability in question. "Really significant" means but this: we cannot assume a value greater than this P_{minim} to the probability that the given system of frequencies has arisen by random sampling from the supposed theoretical population.

B) Let

$$\theta_1 = f_1(m_1; m_2; \dots; m_n), \theta_2 = f_2(m_1; m_2; \dots), \theta_q = f_q(m_1; m_2; \dots)$$

be functions of empirical frequencies such that

$$f_1(\mu_1; \mu_2; \dots) = 0, f_2(\mu_1; \mu_2; \dots) = 0, \dots, f_q(\mu_1; \mu_2; \dots) = 0$$

and let $\sigma_{\theta_1}, \sigma_{\theta_2}, \dots, r_{\theta_i\theta_j}, \dots$ be their standard deviations and correlations.

Then the probability of our frequency distribution being a random sample of the theoretical population $(\mu_1; \mu_2; \dots; \mu_n)$ can be judged

α) From the probability of the deviation of any θ_i from its zero value. In this case

$$P = \sqrt{\frac{2}{\pi}} \int_{\theta_i}^{\infty} \exp\left(-\frac{\theta_i^2}{2\sigma_{\theta_i}^2}\right) d\theta_i.$$

β) From the probability of the set of deviations from their zero values of a correlated system of functions $\theta_1; \theta_2; \dots; \theta_q$

$$P = Q(\chi_{\theta_1; \theta_2; \dots; \theta_q}^2; q+1)$$

where q is the number of independent values $(\theta_1; \theta_2; \dots; \theta_q)$,

$$\chi_{\theta_1; \theta_2; \dots; \theta_q}^2 = \Sigma[\theta_i R_{ii} \theta_i^2 / \theta_i R \sigma_{\theta_i}^2] + 2 \Sigma[\theta_i R_{ij} \theta_i \theta_j / \theta_i R \sigma_{\theta_i\theta_j}], \quad (2)$$

and R is the same as (1) but with θ_i replacing μ_i .

The question of the relations between the results obtained by different methods seems to me to be a very difficult one. I think, however, that the following propositions hardly can meet objections.

Proposition 1. From all the values $\chi_1, \chi_2, \dots, \chi_s$ that is really significant which gives the least value for P . For ex. (Pearson 1902, p. 280 & 283 –

284): In the case (1 – 3) – Motion of bright Line – the probability of the frequency distribution being a random sample from the general population distributed normally equals 1/23 if judged from the value of the criterion χ^2 and it is < 1/1000 if the probable error of the skewness will be taken into account.

Proposition 2. Should we take a great number of random samples from the general population and evaluate all values

$$\chi^2 \text{ with indices } \mu, \theta_1, \theta_2, \dots, \theta_q, \theta_i\theta_j, \theta_i\theta_j\theta_k, \dots, \theta_i\theta_j\dots\theta_s, \dots$$

for each random sample, the distribution of each χ^2 must be that indicated by the theory within the errors of random sampling.

Proposition 3. Let us have χ_1^2 (for n_1 independent values $\theta_i; \theta_j; \dots; \theta_k$) and χ_2^2 (for n_2 independent values θ with other indices) and let n_1 not be equal to n_2 . Then it is impossible that for all random samples $\chi_1^2 = \chi_2^2 = \chi^2$ say. Indeed, the theoretical distribution of χ_1^2 as given by $Q(\chi^2; n_1 + 1)$ differs from the theoretical distribution of χ_2^2 as given by $Q(\chi^2; n_2 + 1)$ whereas χ_1 being identical with χ_2 their distributions must and will be also identical.

2. I come now to consideration of the point of our divergence and I confess that “if I write

$$1e_p = \int f_p - N(1f_p + 2f_p)/(N' + N'')$$

I vary the constitution of the general population for each pair of samples I take, whereas it must really be a constant, as we take all pairs of samples”.

For consequence χ^2 proposed by me as the criterion of divergency cannot be regarded as your criterion for goodness of fit as worked out in your paper (Pearson 1900, pp. 160 – 163). In the notation of this letter it is not χ^2_μ . But nevertheless it is significant. Let us have a contingency table [Table 1] and let us look upon the values like

$$m_{ij} - N_i'N_j''/N = \varepsilon_{ij}$$

as on the functions of the group frequencies, varying from sample to sample, and becoming all zeros for the general population. Then my criterion of divergency χ^2_ε [Slutsky wrote out the right side of (2) with ε replacing θ]; the corresponding value of

$$P = Q[\chi^2_\varepsilon; (s - 1)(t - 1) + 1]$$

measures the probability “that a given system of deviations from the probable ($\varepsilon_{ij} = 0$) in the case of a correlated system of variables (ε_{ij}) is such that it can be reasonably supposed to have arisen from random sampling”.

It is quite analogous with my $\chi^2_{\theta_1; \theta_2; \dots; \theta_q}$ and it is easily to be subsumed under your general theory in Pearson (1900, p. 157 – 160).

Let us suppose there is no correlation in the general population and let a great number of random samples be taken from it. Then the distribution of values of χ^2_ε will be that given by $Q[\chi^2_\varepsilon; (s - 1)(t - 1) + 1]$.

I have shown in my paper that my criterion of divergency (χ^2_ε) for a fourfold table is identical as to its numerical value with your square continugency χ^2_μ . If so both theories cannot be valid as it is shown in the proposition 3 above.

I am not able now to see any error in my reasoning and it seems me the divergence in our views resolves as follows: We do not know the theoretical frequencies and we use “the best available values”, i.e. $N_i'N_j''/N$ as it occurs in many other cases.

(A) I think that they are not the best, and it seems to me you will agree that we should obtain far better values if we have had a theory of skew surfaces. Then fitting such a surface to the system of values like $N_i'N_j''/N$ and integrating its volume for the base elements of the subgroups we have had indeed the best available values.

(B) Yet supposed the values like $N'N''/N$ be “the best available”, there is still no ground that they are sufficiently good, for we can safely use the theoretical values deduced from the sample itself instead of the unknown quantities relating to the general population only if their probable errors are sufficiently small. That is the case with the standard deviation, when used to determine the probable error of the mean. In determining the goodness of fit we bring into the comparison the empirical frequencies with the theoretical ones deduced from the sample itself. But in using the method of moments for fitting the curves we reduce largely the probable errors of the theoretical group frequencies so that they become small as compared with the empirical frequencies.

For Ex. the frequency in Gaussian distribution, the base element being h , is $\mu_x \approx yh$ whence $\sigma_\mu/\mu = \sigma_y/y$. But in this case

$$\delta y/y(x^2/\sigma^3)\delta\sigma, \text{ so that } \sigma_y/y = (x^2/\sigma^2)\sqrt{2N}.$$

For the empirical frequency m_x we have

$$\sigma_m = \sqrt{m[1 - (m/N)]}, \sigma_\mu/\mu = \sqrt{(1/m) - (1/N)} \approx 1/\sqrt{m}.$$

Let $x = (1/2)\sigma$, $h = (1/8)\sigma$, $N = 450$, $m = \mu$. Then $\sigma_\mu/\mu = \sigma_y/y = 0.008$ and $\sigma_m/m = 0.224$ exceeding by 28 times the former value of procentual error of theoretical frequency. Let us take now a fourfold table [Table 2] and suppose the values a, b, c, d be proportional to the values in the general population. Let $a' = (a + b)(a + (1)c)/N$. Then

$$\begin{aligned} \sigma_a &= [a(1 - (a/N))]^{1/2}, \\ \sigma_{a'} &= (1/N)[(a + c)^2\sigma_{a+b}^2 + (a + b)^2\sigma_{a+c}^2 + \\ &\quad 2(a + b)(a + c)\sigma_{a+b}\sigma_{a+c}r_{a+b, a+c}]^{1/2} \end{aligned}$$

where

$$\begin{aligned} \sigma_{a+b}^2 &= (a + b)[1 - (a + b)/N], \\ \sigma_{a+b}\sigma_{a+c}r_{a+b, a+c} &= a - (a + b)(a + c)/N. \end{aligned}$$

For $a = b = c = d = 12$, $\sigma_a = 3$, $\sigma_{a+b}^2 = \sigma_{a+c}^2 = 12$, $r_{a+b, a+c} = 0$, $\sigma_{a'} = 2.45$.

For $a = b = c = d = 12$, $\sigma_a = 30$, $\sigma_{a'} = 24.5$.

Thus, taking for the theoretical frequency $(a + b)(a + c)/N$ as determined by any random sample and dealing with every possible random sample we shall have our errors measured from the point the position of which is subject to errors of random sampling almost so great as the values we are measuring thereof. In consequence we shall obtain the values of χ^2 on the average largely reduced as compared with the case we knew the à priori frequencies in the general population. In my paper are given the values of χ^2_ε evaluated for random samples obtained by the experiment. The values of ε which correspond to the ε in the notation of this letter were measured from the theoretical frequencies deduced from the data. If we measure them from the frequencies known in my case à priori: $a = b = c = d = 12$, we obtain, as a matter of fact, much greater values (given in the table here apart). If we use the same grouping as before we obtain [Table 3].

This seems to me to confirm my views that your theory is to be applied in the cases where we know the à priori frequencies but that in the cases we do not know them your χ^2_μ must be replaced by my χ^2_ε which is numerically identical with it, so that the whole difference in the results touches only the value of n' being in the case we use χ^2_ε , $(s - 1)(t - 1) + 1$.

It seems to me I have found now more stronger grounds for the proposed modification in the theory and I will be immensely grateful to you if you let me know your views on the matter. Again thanking you for your courtesy I am

Yours very faithfully E. Slutsky

Table 1				Table 2		
	m_{ij}	...	N_1'	a	b	$a + b$
			N_2'	c	d	$c + d$
			...	$a + c$	$b + d$	N
N_1''	N_2''		N			

Table 3

Values of χ^2_μ (on à priori grounds)	Theory μ	Statistics m	e^2 / μ
0 – 0.25	1.54	2	0.14
0.25 – 0.50	2.51	1.5	0.41
0.50 – 1	5.88	3.5	0.96
1 – 2	11.44	13	0.21
2 – 3	9.04	5.5	1.39
3 – ∞	19.58	24.5	1.24

whence $P = 0.50$, $\chi^2 = 4.35$

Table 4

The values of χ^2_μ (criterion for goodness of fit) – for the experiment described in my paper – if we use the à priori probabilities: $a = 12$, $b = 12$, $c = 12$, $d = 12$.

N	χ^2_μ										
1	1 1/3	10	5 1/6	19	1 1/3	27	1 2/3	35	1/6	43	2 1/2
2	1 2/3	11	4 1/6	20	4 1/2	28	1 1/3	36	1 2/3	44	2 1/6
3	1 1/6	12	3 1/2	21	4 1/2	29	2 1/2	37	5 5/6	45	3 2/3
4	2 1/6	13	1 1/2	22	1/2	30	1	38	4 1/6	46	3 1/6
5	4 1/6	14	4 1/6	23	5/6	31	6 1/2	39	3	47	5 1/6
6	3 5/6	15	3 1/2	24	9 1/2	32	3 1/6	40	3 1/2	48	4 1/2
7	1 1/6	16	1 5/6	25	1/2	33	0	41	5 5/6	49	3 2/3
8	3 1/6	17	1	26	2 1/6	34	3 1/6	42	1 1/6	50	5 5/6
9	1/2	18	1 5/6								

3.3.3. Slutsky – Pearson, 18 May 1912

Kept at the same place, 856/4

Dear Sir, – I take the liberty to write you again, before I have your answer on my previous letter. I am printing now a treatise (or a text-book) on the theory of correlation and I would be very gratefull to you if you let me know whether the probable error of the partial correlation coefficient can be reduced to the same form as the probable error of the total one, as m^r Yule says.¹³

I have also brought fast [replace this German word by the proper English almost – O. S.] to the end a paper on a General test for Goodness of Fit of the Regression Curves. To keep your valuable time I do not send it to you and I take the liberty only to communicate you an idea of it you will easily appreciate. It is very simple but I am not able to refer to any previous mention of it.

In the notation of your memoir on Skew correlation (1905) the criterion will be simply

$$\chi^2 = \frac{S(Y - y_x)^2}{\sigma_{n_x}^2 / n_x}$$

n' = number of arrays + 1 for there is no correlation between the means of the x -arrays and the probability of a deviation is

$$C \exp\left[-\frac{(Y - y_x)^2}{2[\sigma_{n_x}^2 / n_x] \delta(Y - y_x)}\right].$$

Quite analogous will be a criterion which can be applied in the physical sciences to test the probability that a given system of measurements can reasonably be supposed to correspond to the proposed functional relationship. If you will agree with this I can send you a more elaborate – but still a short paper – with the illustrations taken from your memoir on skew correlation (1905).

I excuse myself, dear sir, for my very imperfect English and for the trouble I give you and remain very faithfully yours E. Slutsky

3.3.4. Slutsky's Letter to Aleksandr Nikolaevich Shchukarev, a specialist in physical chemistry (1928)

Archive of the Moscow State Univ., Fond 276, Inventory 1, No. 114

Slutsky made known his opinion about Shchukarev's unnamed paper, perhaps answering the latter's request. This paper (1928), which I located without much difficulty, was written extremely carelessly. In essence, Shchukarev vainly attempted to derive the Maxwellian law without introducing any stochastic ideas and it is therefore sufficient to say only a few words about Slutsky's reply.

Slutsky indicated that Shchukarev had not nevertheless managed without stochastic considerations; admitted (perhaps too modestly) that he "hardly understands" physics but "somewhat catches" the logical structure of "suchlike theories"; and offered concrete remarks (unnecessarily since the paper was beyond repair).

Notes

1. Short anonymous and hardly differing articles on Slutsky are included in the 2nd and 3rd editions of the *Bolshaia Sovetskaia Enziklopedia*; the 3rd edition is available in an English translation (entitled *Great Sov. Enc.*). My references do not at all exhaust the literature on him. Sarymsakov (1948) praised his work in geophysics, and the authors of several sections of Stokalo (1970) described his mathematical achievements. Romanovsky (1935) indicated that Slutsky was chairman of a commission on applying statistical methods in industry (as a young man he studied for a few years at the machine-building department of the Munich polytechnic school [xix, § 2]). It seems, however, that, because of the negative attitude of the Soviet establishment towards statistics in general (§2), that commission was unable to be of essential use.

2. For a background to this section see Sheynin (1998).

3. Slutsky had been applying these discoveries mostly to economics, and his transition to other branches of knowledge was painful: disallowing a report that appeared in 1932 but was delivered by Slutsky in 1928, he had not published anything during 1930 – 1931. I also note that an English translation of his paper of 1927 was published in 1937. It found important application in investigating time series in economics (Allen 1950, pp. 209 – 210).

4. In 1990 the eminent mathematician Konüs told me that at the time he had also worked at the Conjecture Institute. He was left alone; as Könus explained the attitude of those responsible for the decision-making, they had decided: "He is only a mathematician, not responsible for anything..."

5. In 1929 a paper by the mathematician and statistician N.V. Smirnov appeared in the *Vestnik* and Slutsky even before his move to Moscow had published four articles there.

6. Smit (1931, p. 4) clumsily declared that "the crowds of arrested saboteurs are full of statisticians". Anderson, a student of Chuprov, testified (1959, p. 294):

I could have listed many highly reputed in Russia statisticians and many young and very promising students [...] of Chuprov whose names had suddenly entirely disappeared after 1930 from the Soviet scientific literature.

7. Also see [xix, § 7]. In 1970 Chetverikov had given me (Russian) typed texts of these letters which I turned over to Seneta who acknowledged my help in obtaining “important materials” but had not elaborated and deleted my name from his translation of Chetverkov’s covering note (see below). He was concerned that I could have had problems with the Soviet authorities.

8. In § 3.2, Slutsky mentioned interpolation applied for representing empirical points by suitable curves.

9. Called after A. L. Shaniavsky (1837 – 1905). The University was established in 1908 and closed in 1918 (see *Great Sov. Enc.*, 3rd edition, vol. 29; also its English translation).

10. Slutsky obviously referred not to the paper itself as put out in the *Matematich. Sbornik*, but to its previously published offprint. Indeed, he mentioned the year 1911 and p. 4 neither of which agree with the periodical. The appropriate page numbers in the translation (see References) are 77 and 78.

11. Slutsky discusses the Pearson curves. At the time (and even in 1928, in his letter to Shchukarev, see §3.3.4, which I only describe but do not quote) he sometimes wrote “theory of probability” instead of the correct Russian “... of probabilities”.

Slutsky derived the equation (see beginning of letter) in his book (1912, p. 17/2009, § 5, formula 5.4). Also there (1912, pp. 15 – 17 rather than 16 – 17) he obtained the normal distribution as the limiting law for the binomial distribution. Assume the unknown law (Y) as, for example, a polynomial of the n -th degree, then, in principle, its $(n + 1)$ parameters can be determined given the appropriate moments. If the class to which Y belongs is not restricted, its unique determination is impossible even if “all” the moments are given. Slutsky’s question apparently touched on this *problem of moments*.

12. For some reason the pressmarks of two of the letters are identical.

13. Slutsky’s reference is *J. Roy. Stat. Soc.*, 1907, pp. 6 and 47. In both these cases Yule was a participant in discussing the contributions of other authors. Just below Slutsky actually mentions an initial version of his future paper (1914).

Bibliography

Allen, R. G. D. (1950), The work of Eugen Slutsky. *Econometrica*, vol. 18, pp. 209 – 216.

Anderson, O. (1959), Mathematik für marxistisch-leninistische Volkswirte. *Jahrb. f. Nat.-Ökon. u. Statistik*, Bd. 171, pp. 293 – 299.

Anonymous (1995, in Russian), Anniversaries and memorable dates. *Voprosy Statistiki*, No. 11, p. 77.

Gnedenko, B. V. (1960), Slutsky. In Slutsky (1960, pp. 5 – 11).

Grave, D. A. (1912, In Russian), Insurance mathematics. *Izvestia Kievsk. Kommerch. Inst.*, Book 16, pp. i – iv + 1 – 88 of second paging.

Eliseeva, I. I., Volkov, A. G. (1999, in Russian), E. E. Slutsky: Life and work. *Izvestia Sankt-Peterburgsk. Univ. Ekonomiki i Finansov*, No. 1, pp. 113 – 121.

Kolmogorov, A. N. (1948, in Russian), Slutsky. Obituary. *Math. Scientist*, vol. 27, 2002, pp. 67 – 74.

Könus, A. A. (1978), Slutsky. *Intern. Stat. Enc.*, vol. 2. Editors, W. H. Kruskal, Judith M. Tanur. New York – London, 1978, pp. 1000 – 1001.

Lakhtin, L. K. (1904, in Russian), On the Pearson method etc. *Matematich. Sbornik*, vol. 24, pp. 481 – 500.

Leontovich, A. (1909 – 1911), *Elementarnoe Posobie k Primeneniu Metodov Gaussa i Pirsona ...* (Elementary Manual on Application of the Methods of Gauss and Pearson for Estimating Errors in Statistics and Biology). Kiev.

Makasheva, N. (1988, in Russian), N. D. Kondratiev. Brief biographical essay. *Mirovaia Ekonomika i Mezhdunarodn. Otnoshenia*, No. 9, pp. 59 – 61.

Markov, A. A. (1908), *Ischislenie Veroiatnostei* (Calculus of Probability), 2nd edition. Petersburg.

--- (1912, in Russian), A rebuke to P. A. Nekrasov. *Matematich. Sbornik*, vol. 28, pp. 215 – 227. Translated in Nekrasov, P. A. (2004), *The Theory of Probability*. Coll. Papers, pp. 73 – 79. Berlin. Also at www.sheynin.de

Nekrasov, P. A. (1912), *Teoria Veroiatnostei* (Theory of Probability). Moscow. Second edition.

Ondar, Kh. O., Editor (1977, in Russian), *Correspondence between Markov and Chuprov*. New York, 1981.

- Pearson, K.** (1900), On a criterion that a given system of deviations from the probable etc. *Phil. Mag.*, vol. 50, pp. 157 – 175.
- (1902), On the mathematical theory of errors of judgement etc. *Phil. Trans. Roy. Soc.*, vol. A198, pp. 235 – 299.
- (1905), *On the General Theory of Skew Correlation etc.* Drapers' Co. Res. Mem., Biometric Ser., 2, 1905.
- Romanovsky, V. I.** (1935), On the application of math. statistics and the theory of probability in the industries of the Soviet Union. *J. Amer. Stat. Assoc.*, vol. 30, pp. 709 – 710.
- Sarymsakov, T. A.** (1948, in Russian), Statistical methods and problems in geophysics. *Vtoroe Vsesoyuznoe Soveshchanie po Matematicheskoi Statistike* (Second All-Union Conf. on Math. Statistics). Tashkent, pp. 221 – 239.
- Schükarev, A. N.** (1928), Ein Versuch der Ableitung des Maxwellischen Verteilungsgesetzes. *Phys. Z.*, Bd. 29, No. 6, pp. 181 – 182.
- Seneta, E.** (1988), Slutsky. *Enc. Stat. Sci.*, vol. 8. Editors, S. Kotz, N. L. Johnson. New York, pp. 512 – 515. Second edition of that source: Hoboken, New Jersey, 2006. Seneta's note is reprinted in vol. 12, pp. 7794 – 7796.
- (1992), On the history of the strong law of large numbers etc. *Hist. Math.*, vol. 19, pp. 24 – 39.
- Sheynin, O.** (1990, in Russian), *Chuprov. Life, Work, Correspondence*. Göttingen, 1996.
- (1998), Statistics in the Soviet epoch. *Jahrb. f. Nat.-Ökon. u. Statistik*, Bd. 217, pp. 529 – 549.
- (2004), *Russian Papers on the History of Probability and Statistics*. Coll. Translations of author's papers. Berlin. Also at www.sheynin.de
- Slutsky, E. E.** (1912), *Teoria Korreliatsii* (Theory of Correlation). *Izvestia Kievsk. Kommerch. Inst.*, Book 16. Later published separately (Kiev, 1912). Translation: Berlin, 2009. Also at www.sheynin.de
- (1914), On the criterion of goodness of fit of the regression lines etc. *J. Roy. Stat. Soc.*, vol. 77, pp. 78 – 84.
- (1960), *Izbrannye Trudy* (Sel. Works). Moscow.
- Smirnov, N. V.** (1948, in Russian), Slutsky. *Izvestia Akad. Nauk SSSR*, ser. matematich., vol. 12, pp. 417 – 420.
- Smit, M.** (1930, in Russian), Planned sabotage and the statistical theory. *Planovoe Khoziastvo*, No. 10, pp. 139 – 168. Incorporates several reports including that of B. S. Iastremsky.
- (1931), *Teoria i Praktika Sovetskoi Statistiki* (Theory and Practice of Soviet Statistics), 2nd edition. Moscow - Leningrad.
- (1934, in Russian), Against the idealistic and mechanistic theories in the theory of Soviet statistics. *Planovoe Khoziastvo*, No. 7, pp. 217 – 231.
- Stokalo, I. Z.**, Editor, (1970), *Istoria Otechestvennoi Matematiki* (History of National Mathematics), vol. 4, pt. 2. Kiev.
- Youshkevich, A. A.** (1975), Slutsky. *Dict. Scient. Biogr.*, vol. 12, p. 461.

XIX

N. S. Chetverikov

The Life and Scientific Work of Slutsky

Zizn i nauchnaia deiatelnost E. E. Slutskogo (1959).
Statisticheskie Issledovania (Statistical Investigations). Coll. Papers.
Moscow, 1975, pp. 261 – 281
First published 1959

The sources for this paper were Slutsky's biography written by his wife (manuscript [location not provided; discovered by G. Rauscher in the Russian State Archive for Literature and Art (RGALI), Fond 2133, Inventory 2, No. 52 – 53]); Kolmogorov (1948) and Smirnov (1948); Slutsky's autobiographies the first of which he presented when joining the Steklov Mathematical Institute in 1939, and the second one which he compiled for submitting it to the Uzbek Academy of Sciences on 3 December 1942 [xvi; xvii]; Slutsky's note [27]; his letters to his wife and to me; and my personal recollections.

[1] A historical perspective and a long temporal distance are needed for narrating the life and work of such a profound researcher as Evgeny Evgenievich Slutsky (7/19 March 1880 – 10 March 1948). Time, however, is measured by events rather than years; in this case, first and foremost, by the development of scientific ideas.

Only a little more than ten years have passed since E. E. had died, but the seeds of new ideas sown by him have germinated and even ripened for the first harvest, – I bear in mind the rapid development of the theory of random functions. [To repeat,] however, a comprehensive estimation of his total rich and diverse heritage will only become possible in the future.

The description of Slutsky's life presents many difficulties occasioned both by complications and contradictions of his lifetime and the complexity of his spiritual make-up: a mathematician, sociologist, painter and poet were combined in his person. In essence, his life may be divided into three stages: the periods of seeking his own way; of passion for economic issues; and the most fruitful stage of investigations in the theory of probability and theoretical statistics. The fourth period, when he moved into pure mathematics¹, had just begun and was cut short by his death.

E. E. grew up in the family of a teacher and educator of the Novinsk teachers' seminary (former Yaroslavl province). His father was unable to get along with the Director who had not been averse to embezzle state property, and, after passing through prolonged ordeals, his family settled in Zhitomir. There E. E. had learned in a gymnasium which he was later unable to recall without repugnance. His natural endowments enabled him to graduate with a gold medal and his exceptional mathematical abilities and the peculiar features of his thinking had been revealed already in school. Having been very quick to grasp the main idea of analytic geometry, he successfully mastered its elements all by himself and without any textbooks.

After graduating in 1899, he entered the physical and mathematical faculty of Kiev University. There, he was carried away by the political wave

of the student movement, and already in 1901, for participating in an unauthorized gathering (*skhodka*), he was expelled, together with 183 students, and drafted under compulsion into the Army on the order of Bogolepov, the Minister of People's Education. Because of vigorous public protests coupled with disturbances at all higher academic institutions, that order was soon disaffirmed. Nevertheless, already next year, for participating in a demonstration against the Minister Senger, E. E. was again thrown out of the University, and this time banned from entering any other Russian higher institution.

Only fragmentary information about Slutsky's active political work at that time, including the performance abroad of tasks ordered by a revolutionary group, is extant, but even so it testifies to the resolve and oblivion of self with which he followed his calling as understood at the moment. Owing to financial support rendered by his grandmother, E. E. became able to enter the machine-building faculty of the Polytechnic High School in Munich. Being cut off from practical political activities, he turned to sociology and was naturally enthralled by its main field, political economy [economics].

He had begun by studying the works of Ricardo, then Marx' *Kapital* and Lenin's *Development of Capitalism in Russia* [1899], and turned to the classics of theoretical economy. Although technical sciences provided some possibilities for his inclination to mathematics to reveal itself, he felt a distaste for them. He mostly took advantage of the years of forced life abroad for deep studies of economic problems. At the end of 1904 E. E. organized in Munich a group for studying political economy and successfully supervised its activities.

[2] After the revolutionary events of 1905 [in Russia] he became able to return to his homeland. He abandoned technical sciences and again entered Kiev University, this time choosing the law faculty whose curriculum included political economy. His plans contemplating long years of studying theoretical economy with a mathematical bias have ripened.

Slutsky's mathematical mentality attracted him to the development of those economic theories where the application of mathematics promised tempting prospects. However, now also his scientific activities and learning went on with interruptions. The years 1905 and 1906 were almost completely lost [because of revolutionary events] and in March 1908 he was expelled from the University for a year. As E. E. himself admitted, that disciplinary punishment followed after a "boyish escapade" resulting from his "impetuous disposition". Nevertheless, in 1911, being already 31 years old, he graduated from the law faculty with a gold medal awarded for his diploma thesis *Theory of marginal utility*, a critical investigation in the field of political economy², but his firmly established reputation of being a "red student" prevented his further studies at the university.

Such were the external events that took place during Slutsky's first stage of life. They should be supplemented by one more development, by his marriage, in November 1906, to Yulia Nikolaevna Volodkevich. Before going on to his second stage, let us try to discuss what were the inner motives, the vital issues, the inclinations that had been driving E. E. at that time. Those years may be called the period when he had been searching his conscience. An indefatigable thinker was being born; a person who criticized everything coming from without, who avidly grabbed all the novelties on which he could test his own ripening thoughts. He looked for

his own real path that would completely answer his natural possibilities and inclinations. [He withdrew from practical revolutionary activities because he had soon understood that the path of a revolutionary was alien for him: in dangerous situations he was unable to orient himself quickly enough, he had no visual memory and he lacked many more of what was necessary for a member of an underground organization. N. C.]

E. E. was attracted by creative scientific work and he examined himself in various directions, – in technology and economics, in logic and the theory of statistics. In any of these domains, however, he only became aware of his real power when becoming able to submit his subject of study to quantitative analysis and mathematical thought. In one of his letters he wrote:

The point is not that I dream of becoming a Marx or a Kant, of opening up a new epoch in science, etc. I want to be myself, to develop my natural abilities and achieve as much as is possible for me. Am I not entitled to that?

He aimed at finding his place in science that would be in keeping with his natural gifts. In 1904, he wrote:

A man must certainly be working [only] in that field to which his individuality drives him. [...] He must be living only there, where he is able to manifest it more widely, more completely, and to create, i.e., to work independently and with loving care.

The word “independently” was not chosen randomly; it illuminated his creative life-work.

When taking up any issue, he always began by thinking out the initial concepts and propositions. He always went on in his own, special manner, and the ideas of other authors only interested him insofar as they could serve for criticisms. This originality of thought deepened as the years went by and gradually led Slutsky to those boundaries after which not only new ways of solving [known] problems are opening up, but new, never before contemplated issues leading the mind to yet unexplored spaces, were discovered.

The most remarkable feature of Slutsky’s scientific work was the selfless passion with which he sought the truth and which he himself, in a letter to his wife, compared with that of a hunter:

You are telling me of being afraid for my work, afraid of the abundance of my fantasy [...]. Is it possible to work without risk? And is it worthwhile to undertake easy tasks possible for anyone? I am pleased with my work [published] in Metron exactly because it contains fantasy. For two hundred years people have been beating about the bush and no one ever noticed a simple fact, whereas I found there an entire field open for investigation [...]. It is impossible to avoid wrong tracks. Discovery is akin to hunting. Sometimes you feel that the game is somewhere here; you poke about, look out, cast one thing aside, take up another thing, and finally you find something [...]

My present work [17] [of 1925 – 1926 on pseudo-periodic waves created by the composition of purely random oscillations – N. C.] is, however,

absolutely reliable. I am not pursuing a chimera, this is absolutely clear now, but it does not lessen the excitement of the hunt. In any case, I found the game, found it after the hunt was over [...]. I am afraid that the purely mathematical difficulties are so great as to be insurmountable for me. But neither is this, after all, so bad.

[3] After graduating from the University, Slutsky plunged into the work of numerous scientific societies, and, at the same time was compelled to earn money and wished to pass on his views, knowledge, and achievements, into teaching. It seemed that he had left little time and strength for scientific work, but his creative initiative overcame every obstacle, and even during that difficult and troublesome period E. E. was able to publish his first, but nevertheless important investigations.

Already the list of the scientific societies whose member he was, shows how diverse were his interests and how wide was the foundation then laid for future investigations. In 1909, still being a student, he was corresponding member of the Society of Economists at Kiev Commercial Institute; in 1911, he became full member, in 1911 – 1913, he was its secretary, and, in 1913 – 1915, member of its council. In 1912 E. E. was elected full member of the [Kiev?] Mathematical Society; later on he joined the Sociological Society at the Institute for Sociological Investigations in Kiev, and in 1915 became full member of the A. I. Chuprov³ Society for Development of Social Sciences at Moscow University.

Owing to his disreputable political reputation, Slutsky's pedagogical work at once encountered many obstacles. In 1911 he was not allowed to sit for his Master's examinations at Kiev University⁴ and in 1912 he was not approved as teacher. The same year his father-in-law, N. N. Volodkevich, an outstanding educationalist of his time, took him on as teacher of political economy and jurisprudence at the [commercial] school established and headed by himself, but the Ministry for Commerce and Industry did not approve him as a staff worker. Only Slutsky's trip to Petersburg and his personal ties made it possible for him to remain in that school and to be approved, in 1915, in his position. Yulia Nikolaevna taught natural sciences at the same school. The apartment of the young married couple was attached to the school building and it was there that his life became then "mostly tied to the desk and illuminated by the fire of creative life" (from his biography written by his wife).

Slutsky first became acquainted with theoretical statistics in 1911 – 1912 having been prompted by Leontovich's book (1909 – 1911). It is impossible to say that that source, whose author later became an eminent physiologist and neurohistologist, member of the Ukrainian Academy of Sciences, was distinguished by clearness or correct exposition of the compiled material. Nevertheless, it was there that the Russian reader had first been able to learn in some detail the stochastic ideas of Pearson and his collaborators, and there also a list of the pertinent literature was adduced. That was enough for arousing Slutsky's interest, and we can only be surprised at how quickly he was able to acquaint himself with the already then very complicated constructions of the English statisticians-biologists by reading the primary sources; at how deeply he penetrated the logical principles of correlation theory; and at how, by using his critical feelings, he singled out the most

essential and, in addition, noticed the vulnerable spots in the Pearsonian notions.

It is almost a miracle that only a year later there appeared Slutsky's own book [1] devoted to the same issues, explicated with such clearness, such an understanding of both the mathematical and logical sides, that even today it is impossible to name a better Russian aid for becoming acquainted with the principles of the constructions of the British school of mathematical statistics. And less than in two years the *Journal of the Royal Statistical Society* carried Slutsky's paper [5] on the goodness of fit of the lines of regression criticizing the pertinent constructions of the English statisticians. A short review published in 1913 [3] shows how deeply E. E. was able even then to grasp such issues as Markov chains [a later term] and how ardently he defended Markov's scientific achievements against the mockery of ignoramuses.

During those years, economic issues had nevertheless remained in the forefront. Even as a student, E. E. decided not to restrict his attention there to purely theoretical constructions and contemplated a paper on the eight-hour working day. He buried himself in factory reports, established connections with mills, studied manufacturing and working conditions. Issuing from the collected data, he distributed the reported severe injuries in accord with the hours of the day and established their dependence on the degree of the workers' tiredness. Earnestly examining economic literature, he connected his studies with compilation of popular articles on political economy as well as with his extensive teaching activities which he carried out up to his move to Moscow in 1926.

[4] E. E. remained in Volodkevich's school until 1918 although school teaching was difficult for him. In the spring of 1915 he became instructor at the Kiev Commercial Institute. There, he read courses in sampling investigations and mathematical statistics, and, after the interruption caused by the World War, both there and at the Ukrainian Cooperative Institute, the history of economic and socialist doctrines. In 1917 Slutsky began teaching the discipline most congenial to him at the time, – theoretical economy. After the October revolution he taught in many newly established academic institutions: an elementary course in political economy at the Cooperative Courses for the disabled; introduction to the logic of social sciences at People's University. The Commercial Institute remained, however, his main pedagogical place of work, and there he also read courses in theoretical economy and political economy (theory of value and distribution).

The listing above, even taken in itself, clearly shows that in those years Slutsky concentrated on issues of theoretical economy and, more specifically, on those that admitted a mathematical approach. After his diploma thesis and a small essay on Petty [4], E. E. published an investigation about "equilibrium in consumption" [6] that only much later elicited response and due appreciation in the Western literature⁵. Less important research appeared in Kiev and Moscow respectively [10; 11]. However, two considerable contributions to economics [14; 15] were still connected with Kiev.

By 1922 Slutsky had already abandoned theoretical economics and afterwards devoted all his efforts to statistics. He himself, in his autobiography, explained his decision in the following way:

When the capitalist economics had been falling to the ground, and the outlines of a planned socialist economic regime began to take shape, the foundation for those problems, that interested me as an economist and mathematician, disappeared⁶.

This is a very typical admission: the decisive significance for E. E., when choosing a field for scientific work, was the possibility of applying his mathematical talent. His inclination was caused not as though by an artisan's joy of skilfully using his tools, – no, he was irrepressibly attracted to abstract thinking, be it mathematics, logic, theory of knowledge or his poetic creativity.

[5] We already know that E. E. began his investigations in the theory of statistics in 1911 – 1914, his first contributions having been the book on correlation theory [1] and a paper on the lines of regression [5]. In the beginning of September 1912, in Petersburg, where E. E. had come to plead for being approved as teacher, he became acquainted, and fruitfully discussed scientific and pedagogic issues with A. A. Chuprov, who highly appraised his book⁷. During 1915 – 1916 Slutsky's name regularly appeared in the *Statisticheskyy Vestnik*, a periodical issued by the statistical section of the A. I. Chuprov Society at Moscow University. There, he published thorough reviews [7] or short notes [8] directed against wrong interpretation of the methods of mathematical statistics.

In 1922, after an interval of many years, Slutsky returned to the theory of statistics. He examined the logical foundation of the theory of probability and the essence of the law of large numbers from the viewpoint of the theory of knowledge. In November 1922, at the section on theoretical statistics of the Third All-Russian Statistical Conference, he read a report of great scientific importance. It touched on the main epistemological issue of the theory of probability, was published [9] and then reprinted with insignificant changes.

In 1925 he issued another important paper [12] introducing the new notions of stochastic limit and stochastic asymptote, applied them for providing a new interpretation of the Poisson law of large numbers and touched on the logical aspect of that issue by critically considering the Cournot lemma as formulated by Chuprov⁸. Also in 1925, he published a fundamental contribution [13] where he defined and investigated the abovementioned notions, applied them for deriving necessary conditions for the law of large numbers, which he, in addition, generalized onto the multidimensional case. Later on this work became the basis of the theory of stochastic functions.

[6] By 1926, Slutsky's life in Kiev became very complicated. He did not master Ukrainian, and a compulsory demand of the time, that all the lectures be read in that language, made his teaching at Kiev higher academic institutions impossible. After hesitating for a long time, and being invited by the Central Statistical Directorate, he decided to move to Moscow. However, soon upon his arrival there, he was attracted by some scientific investigations (the study of cycles in the economy of capitalist countries) made at the Conjecture Institute of the Ministry of Finance. E. E. became an active participant of this research, and, as usual, surrendered himself to it

with all his passion. Here also, a great creative success lay ahead for him. In March of that year he wrote to his wife:

I am head over heels in the new work, am carried away by it. I am almost definitively sure about being lucky to arrive at a rather considerable finding, to discover the secret of how are wavy oscillations originating by a source that, as it seems, had not been until now even suspected. Waves, known in physics, are engendered by forces of elasticity and rotary movements, but this does not yet explain those wavy movements that are observed in social phenomena. I obtained waves by issuing from random oscillations independent one from another and having no periodicities when combining them in some definite way.

The study of pseudo-periodic waves originating in series, whose terms are correlatively connected with each other, led Slutsky to a new important subject, to the errors of the coefficients of correlation between series of that type. In both his investigations, he applied the “method of models”, of artificially reproducing series similar to those actually observed but formed in accord with some plan and therefore possessing a definite origin.

The five years from 1924 to 1928, in spite of all the troubles, anxieties and prolonged housing inconveniences caused by his move to Moscow, became a most fruitful period in Slutsky’s life. During that time, he achieved three considerable aims: he developed the theory of stochastic limit (and asymptote); discovered pseudo-periodic waves; and investigated the errors of the coefficient of correlation between series consisting of terms connected with each other.

[7] In 1928, E. E. participated at the Congress of Mathematicians in Bologna. The trip provided great moral satisfaction and was a grand reward deserved by sleepless nights and creative enthusiasm. His report on stochastic asymptotes and limits attracted everyone. A considerable debate flared up at the Congress between E. E. and the eminent Italian mathematician Cantelli concerning the priority to the strong law of large numbers. Slutsky [16] had stated that it was due to Borel but Cantelli considered himself its author.

Castelnuovo, the famous theoretician of probability, and other Italian mathematicians rallied together with Cantelli against Slutsky, declared that Borel’s book, to which E. E. had referred to, lacked anything of the sort attributed to him by the Russian mathematician, and demanded an immediate explanation from him. E. E. had to repulse numerous attacks launched by the Italians and to prove his case.

The point was that Slutsky, having been restricted by the narrow boundaries of a paper published in the *C. r. Acad. Sci. Paris*, had not expressed himself quite precisely. He indicated that Borel was the first to consider the problem and that Cantelli, Khinchin, Steinhaus and he himself studied it later on. However, he should have singled out Cantelli and stressed his scientific merit. Borel was indeed the first to consider the strong law, but he did it only in passing and connected it with another issue in which he was interested much more. Apparently for this reason Borel had not noticed the entire meaning and importance of that law, whereas Cantelli was the first to grasp all that and developed the issue, and his was the main merit of establishing the strong law of large numbers. E. E. was nevertheless

able to win. Understandably, he did not at all wish to make use of his victory for offending Cantelli.

He appreciated the Italian mathematician; here is a passage from his letter to his wife (Bologna, 6 September 1928)⁹:

[He is] *not a bad man at all, very knowledgeable, wonderfully acquainted with Chebyshev, trying to learn everything possible about the Russian school (only one thing I cannot forgive, that he does not esteem Chuprov). In truth, he has brought fame to the Russian name in Italy, because he doesn't steal but honestly says: that is from there, that is Russian, and that is Russian [...]. Clearly one must let him keep his pride.*

After a prolonged discussion of the aroused discord with Cantelli himself, and a thorough check of the primary sources, E. E. submitted an explanation to the Congress, agreed beforehand with Cantelli. The explanation confirmed his rightness but at the same time had not hurt Cantelli's self-respect. After it was read out, Cantelli, in a short speech, largely concurred with E. E. This episode vividly characterizes Slutsky, – his thorough examination of the problems under investigation, an attentive and deep study of other authors, and a cordial and tactful attitude to fellow-scientists. He was therefore able not only to win his debate with Cantelli, but to convince his opponent as well.

[8] In 1930, the Conjecture Institute ceased to exist, the Central Statistical Directorate was fundamentally reorganized, and Slutsky passed over to institutions connected with geophysics and meteorology where he hoped to apply his discoveries in the field of pseudo-periodic waves. However, he did not find conditions conducive to the necessary several years of theoretical investigations at the Central Institute for Experimental Hydrology and Meteorology. [These lines smack of considerable sadness but they do not at all mean that Slutsky surrendered. N. C.] In an essay [27] he listed his accomplished and intended works important for geophysics. He also explicated his related findings touching on the problem of periodicity, and indicated his investigation of periodograms, partly prepared for publication [26]. Slutsky then listed his notes in the *C. r. Acad. Sci. Paris* [16; 18; 20 – 23] where he developed his notions as published in his previous main work of 1925 [13].

To the beginning of the 1930s belong Slutsky's investigations on the probable errors of means, mean square deviations and coefficients of correlation calculated for interconnected stationary series. He linked those magnitudes with the coefficients of the expansion of an empirical series into a sum of (Fourier) series of trigonometric functions and thus opened up the way of applying those probable errors in practice. Slutsky himself summarized his latest works in his report at the First All-Union Congress of Mathematicians in 1929 but only published (in a supplemented way) seven year later [30].

Owing to the great difficulties of calculation demanded by direct investigations of the interconnected series, Slutsky developed methods able to serve as an ersatz of sorts and called by a generic name "statistical experiment". Specifically, when we desire to check the existence of a connection between two such series, we intentionally compare them in such a way which prevents a real connection; after repeating such certainly

random comparisons many times, we determine how often parallelisms have appeared in the sequences of the terms of both series. They, the parallelisms, create an external similarity of connection not worse than the coincidences observed by a comparison of the initial series. Slutsky developed many versions of that method and applied it to many real geophysical investigations of wavy oscillating series.

E. E. did not belong to those statisticians-mathematicians for whom pure mathematics overshadowed the essence of studied phenomena. He thought that the subject of a methodological work should be determined by its material substance:

It seemed to me that, along with theoretical investigations, I ought to study some concrete problems so as to check my methods and to find the problems for theoretical work,

he wrote in his autobiography submitted in 1939. Bearing in mind such aims, he studied the series of harvests in Russia over 115 years (compiled by V. G. Mikhailovsky), those of the cost of wheat over 369 years (Beveridge), series of economic cycles (Mitchell), etc. Passing on from economic to geophysical series, Slutsky then examined the periodicity of sunspots checking it against data on polar aurora for about two thousand years (Fritz¹⁰) and studied the peculiar vast information stored as annual rings of the giant sequoia of Arizona (mean data for eleven trees covering about two thousand years)¹¹.

[9] And yet fate directed Slutsky to the domain of pure mathematics. In 1934 he passed on to the Mathematical Institute of Moscow University and in 1935 abandoned geophysics. In 1939 he established himself at the Steklov Mathematical Institute of the Soviet Academy of Sciences. At the same time, having been awarded by Moscow University the academic status of Doctor of Mathematical and Physical Sciences honoris causa on the strength of his writings, and entrusted by the chair of mathematical statistics there, Slutsky apparently resumed the long ago forsaken teaching. [However, because of the situation that took shape at the University in those years, N. C.] teaching demanded more strength than he possessed at that time. As he himself wrote,

Having been entrusted with the chair of the theory of probability and mathematical statistics at Moscow University, I have convinced myself soon afterwards, that that stage of life came too late, and I shall not experience the good fortune of having pupils.

It seemed that, having consolidated his position at the Mathematical Institute, E. E. will be able to extend there his work on the theory of statistics. But his plans were too extensive, they demanded the establishment of a large laboratory, and, therefore, large expenses. That proved impossible, and Slutsky had to concentrate on investigations in the theory of stochastic processes and to plunge ever deeper into pure mathematics.

At the end of October 1941, when Moscow was partly evacuated, Slutsky moved with his family to Tashkent. A part of his [unpublished] works was lost. And still he [considered the year 1940/1941 as lucky and N. C.] wrote about that period:

I was able to find a new solution of the problem of tabulating the incomplete Γ -function providing a more complete, and, in principle, the definitive type of its table. The use of American technology allowed to accomplish the calculations almost completely in one year but the war made it impossible to carry them through.

The work had indeed dragged on, and even after his return to Moscow three more years were required for their completion. I cannot fail to mention the selfless help rendered by N. V. Levi, a woman employee of the Mathematical Institute, who accomplished that task when Slutsky had already begun feeling himself ill. He developed lung cancer, and it was a long time before the disease was diagnosed although E. E. himself never got to know its nature. He continued to work on the Introduction to the tables where he was explaining the method of their compilation, but it was N. V. Smirnov who wrote the definitive text. On 9 March 1948 Slutsky was still outlining the last strokes of the Introduction, but next day he passed away.

[10] Already in Kiev Slutsky had been deeply interested in the cognitive and logical side of the problems that he studied, especially concerning his investigations in mathematical statistics. His first independent essential writings were indeed devoted to these general issues. Later on, he essentially cooled down for them; he either solved them to a required by him degree, or his great success in more concrete investigations overshadowed philosophical problems. In any case, in the middle of the 1940s, E. E. even with some irritation refused to discuss purely logical concepts although he had been unable to disregard the then topical criticism levelled by Fisher against the problem of calculating the probabilities of hypotheses (of the Bayes theorem).

First of all Slutsky took it upon himself to ascertain the relations of the theory of probability to statistical methodology. To this aim, he singled out the formal mathematical essence of the theory itself by expelling from it all that, introduced by the philosophical interpretation of the concept of probability. So as to achieve complete clearness, he proposed to abandon the habitual terms and to make use of new ones: disjunctive calculus, valency (assigned to events), etc. To assign, as he stated, meant to establish some relation R between an event and its valency in accord with only one rule: if event A breaks down into a number of alternatives, the sum of all of their valencies must be equal to the valency of A . The valency of the joint event AB , that is, of the occurrence of the events A and B , was determined just as formally. These relations between valencies were included in the axiomatics of the disjunctive calculus, sufficient for developing it as a mathematical discipline. Its applications depended on the contents which we might introduce into the term valency and which can be probability, frequency, or, as a special notion, limiting frequency.

To what extent will these interpreted calculuses coincide and cover each other, depends on the contents of their axiomatics, which, under differing interpretations, can be distinct one from another. However, these distinctions cannot concern the purely mathematical discipline, the disjunctive calculus, because its axiomatics is constructed independently from the interpretation of the subject of valency [9]. When explaining his understanding of the logic of the law of large numbers, Slutsky issued from

those considerations, and he also made use of the notions of stochastic asymptote and stochastic limit. [Chetverikov describes here Slutsky's paper [13]: I advise readers to look up that contribution itself. O. S.]

Slutsky also criticized the purely logical Chuprov – Cournot [Cournot – Chuprov] construction that aimed at connecting probabilities with frequencies of the occurrence of phenomena in the real world, at throwing a “bridge” between them. He thought that the essence of the so-called Cournot lemma consisted in attaching to the law of large numbers the importance of a law of nature without any qualifying remarks about the probability of possible, although extremely rare exceptions. The notion of probability cannot be removed from the Cournot lemma, so, as he concluded, the logical value of the “bridge” itself is lost [12] ¹².

Having been especially prompted by the need to work with time series and issuing from the concept of stochastic limit (asymptote), E. E. also constructed a theory of random functions. [A description of Slutsky's pertinent findings follows. O. S.]

An important discovery made by Slutsky in the mid-1920s consisted in that he connected wavy oscillations with random oscillations and showed how the latter can engender the former [...]. Wavy oscillations are extremely common (for example, in series occurring in economics and meteorology), whereas unconnected randomly oscillating series are met with not so often. A practically important problem is, therefore, to derive the errors of the various general characteristics, – of the mean, the standard deviation, the correlation coefficient, – for connected series ¹³.

E. E. devoted much effort to the solution of that problem. His formulas are bulky, see for example the expression for the error of the correlation coefficient [24, p. 75]. Simpler formulas for particular cases are in [27]. Later Slutsky examined the possibility of applying the χ^2 test and its distribution to connected series as well as of determining the required magnitudes through the Fourier coefficients [25; 26].

By issuing from his theory of connected series, and allowing for the course of stochastic processes, Slutsky was able to provide a methodology of forecasting them, including sufficiently long-term forecasting, with given boundaries of error [29].

We ought to dwell especially on his *method of models* (of *statistical experimentation*) for discovering connections between phenomena. His idea was as follows. When studying many problems not yet completely solved by theory, it is possible to arrange a statistical “experiment” and thus to decide whether the statistical correspondence between phenomena is random or not. For example, when selecting a number of best and worst harvests in Russia from among the series collected by Mikhailovsky for 115 years, we can compare them with the series of maxima and minima of the number of sunspots for more than 300 years. If such comparisons are [if the correspondence is] only possible after shifting one of the series with respect to the other one, then, obviously, the coincidences will be random. However, since the sum of the squares of the discrepancies¹⁴ is minimal when those series are compared without such shifting, we may be sufficiently convinced in that the coincidences are not random [28].

I am unable to appraise Slutsky's purely mathematical studies and am therefore quoting most eminent Soviet mathematicians. Smirnov (1948, pp. 418 – 419), after mentioning Slutsky's investigation [13], wrote:

The next stage in the same direction was his works on the theory of continuous stochastic processes or random functions. One of Slutsky's very important and effective findings here was the proof that any random stochastically continuous function on a segment is stochastically equivalent to a measurable function of an order not higher than the second Baire class. He also derived simple sufficient conditions for a stochastic equivalence of a random function and a continuous function on a segment, conditions for the differentiability of the latter, etc. These works undoubtedly occupy an honourable place among the investigations connected with the development of one of the most topical issues of the contemporary theory of probability, that [issue or theory?] owes its origin to Slutsky's scientific initiative.

The next cycle of Slutsky's works (1926 – 1927) was devoted to the examination of random stationary series, and they served as a point of departure for numerous and fruitful investigations in this important field. Issuing from a simplest model of a series obtained by a multiple moving summation of an unconnected series, he got a class of stationary series having pseudo-periodic properties imitating, over intervals of any large length, series obtained by superposing periodic functions. His finding was a sensation of sorts; it demanded a critical revision of the various attempts of statistical justification of periodic regularities in geophysics, meteorology, etc. It occurred that the hypothesis of superposition of a finite number of regularly periodic oscillations was statistically undistinguishable from that of a random function with a very large zone of connectedness.

His remarkable work on stationary processes with a discrete spectrum was a still deeper penetration into the structure of random functions. In this case, the correlation function will be almost periodic. Slutsky's main result consisted here in that a random function was also almost periodic, belonged to a certain type and was almost everywhere determined by its Fourier series.

These surprisingly new and fearlessly intended investigations, far from exhausting a very difficult and profound problem, nevertheless represent a prominent finding of our science. With respect to methodology and style, they closely adjoin the probability-theoretic concepts of the Moscow school (Kolmogorov, Khinchin), that, historically speaking, originated on a different foundation. The difficult to achieve combination of acuteness and wide theoretical reasoning with a quite clearly perceived concrete direction of the final results, of the final aim of the investigation, is Slutsky's typical feature.

Proving that Slutsky's works were close to those of the Moscow school, Kolmogorov (1948/1962, p. 70) stated:

In 1934, Khinchin showed that a generalized technique of harmonic analysis was applicable to the most general stationary processes considered in Slutsky's work [...]. The modern theory of stationary processes, which most fully explains the essence of continuous physical spectra, has indeed originated from Slutsky's works, coupled with this result of Khinchin.

After E. E.'s interest in applications had shifted from economics to geophysics, it was quite natural for him to pass from considering connected series of random variables to random functions of continuous time. The

peculiar relations, that exist between the different kinds of continuity, differentiability and integrability of such functions, make up a large area of the modern theory of probability whose construction is basically due to Slutsky [19; 20; 26; 30 – 33]¹⁵. Among the difficult results obtained, which are also interesting from the purely mathematical viewpoint, two theorems should be especially noted. According to these, a ‘stochastically continuous’ random function can be realized in the space of measurable functions [31; 33]; and a stationary random function with a discrete spectrum is almost periodic in the Besikovich sense with probability 1 [32].

Kolmogorov then mentions the subtle mastery of Slutsky’s work on the tables of incomplete Γ - and B -functions that led him to the formulation of general problems. The issue consisted in developing a method of their interpolation, simpler than those usually applied, but ensuring the calculation of the values of these functions for intermediate values of their arguments with a stipulated precision. For E. E., this, apparently purely “technical”, problem became a subject of an independent scientific investigation on which he had been so enthusiastically working in his last years. He was able, as I indicated above, to discover a new solution of calculating the incomplete Γ -function, but that successful finish coincided with his tragic death.

Notes

1. Chetverikov thus separated the theory of probability from pure mathematics. O. S.
2. Still extant at the Vernadsky Library, Ukrainian Academy of Sciences, Fond 1, No. 44850 (Chipman 2004, p. 355). Translated into Ukrainian (Kiev, 2006). O. S.
3. A. I. Chuprov, father of the better known A. A. Chuprov. O. S.
4. He only held them in 1918, after the revolution, at Moscow University. O. S.
5. Slutsky made the following marginal note on a reprint of Schults (1935): “This is a supplement to my work that began influencing [economists] only 20 years after having been published”. N. C.
6. This explanation would have only been sufficient if written before 1926. Below, Chetverikov described Slutsky’s work in theoretical economics during 1926 – 1930 at the Conjecture Institute and then implicitly noted that in 1930 the situation in Soviet statistics had drastically worsened. I [xviii, § 2] stated that Slutsky had abandoned economics largely because of the last-mentioned fact. On the fate of the Conjecture Institute see also Sheynin (1990/1996, pp. 29 – 30). Kondratiev, its Director, who was elbowed out of science, persecuted, and shot in 1938 (Ibidem), had studied cycles in the development of capitalist economies. In at least one of his papers, he (1926) had acknowledged the assistance of Chetverikov and Slutsky, a fact that Chetverikov naturally had to pass over in silence. Three papers devoted to Kondratiev are in *Ekonomika i Matematich. Metody*, vol. 28, No. 2, 1992. O. S.
7. I (1990/1996, p. 44) reprinted Chuprov’s review originally published in a newspaper. I also made public Slutsky’s relevant letters to Markov and Chuprov and Slutsky’s scientific character compiled by Chuprov (pp. 44 – 50), see [xviii]. Slutsky’s correspondence with Chuprov discussed, among other issues, the former’s encounter with Pearson. Three letters from Slutsky to Pearson dated 1912 are now available [xviii]. Chuprov was six years older than Slutsky, had much more teaching experience, and was the generally accepted head of the [Russian] statistical school. In the [Petersburg] Polytechnic Institute, he laid the foundation of teaching the theory of statistics. O. S.
8. Chetverikov repeated the mistake made by Chuprov (1909/1959, pp. 166 – 168). The latter stated that Cournot had provided a “canonical” proof of the law of large numbers. In actual fact, Cournot did not even formulate that law (and did not use that term), and his “Lemma” (a term only used by Chuprov himself) had simply indicated (after D’Alembert!) that rare events did not happen (Cournot 1843, §43). Chuprov, however, interpreted that statement as “did not happen often”. Chetverikov was translator of Cournot (Moscow, 1970). Note that Slutsky [12, p. 33] followed Chuprov. O. S.

9. The translation of the passage below is due to Seneta (1992, p. 30) who published the letter (as well as another relevant one from Slutsky to his wife) in full. In 1970 Chetverikov had given me copies of these letters and about 1990 I sent them to Seneta who acknowledged my help in obtaining “important materials” but, being concerned that I could have problems with the Soviet authorities, did not elaborate. I (1993) explained all that and provided additional material concerning Chuprov, Slutsky and Chetverikov.

No one involved including Slutsky knew that the main merit of discovering the strong LLN belonged to Hausdorff. O. S.

10. Hermann Fritz (1830 – 1893), see the appropriate volume of Poggendorff’s *Handwörterbuch*. O. S.

11. Slutsky’s large work on those annual rings including all the pertinent calculations got lost during his evacuation from Moscow. N. C.

12. Chuprov and Slutsky formulated the “Cournot lemma” not as Cournot himself did, see Note 8. O. S.

13. These errors are usually many times greater than the respective errors in unconnected series. N. C.

14. A loose but understandable description. O. S.

15. I changed the numbering, here and below, to conform to that in the present paper. O. S.

Bibliography

E. E. Slutsky

Kolmogorov (1948) adduced a complete list of Slutsky’s contributions compiled by Chetverikov. For this reason, I only provide references to those writings, which the latter mentioned in his text.

Abbreviation: GIIA = *Giorn. dell’Istituto Italiano degli Attuari*

1. *Teoria Korreliatsii i Elementy Uchenia o Krivykh Raspredelenia* (Theory of Correlation and the Elements of the Theory of Distribution Curves). Kiev, 1912. See [i].

2. Essence and form of cooperatives. *Spravochnik-Kalendar Zemledeltsa* (Farmers’ Reference Calendar) for 1913. Kiev, 1912, pp. 1 – 15. In Russian.

3. Review of Markov (1913). Newspaper *Kievskaya Mysl*, 30 March 1930, p. 5. In Russian.

4. In Russian. *Sir William Petty*. Kiev, 1914.

5. On the criterion of goodness of fit of the regression lines and on the best method of fitting them to the data. *J. Roy. Stat. Soc.*, vol. 77, 1914, pp. 78 – 84.

6. Sulla teoria del bilancio del consumatore. *Giorn. del Economisti*, vol. 51, 1915, pp. 1 – 26. Translation: On the theory of the budget of the customer. In: *Readings in Price Theory*. Editors, G. J. Stigler, K. E. Boulding. Homewood, Ill., 1952, pp. 27 – 56.

7. Statistics and mathematics, this being a review of Kaufman (1912). *Statistich. Vestnik* No. 3/4, 1915 – 1916, pp. 1 – 17. In Russian. Translation [ii].

8. On an error in the application of a correlation formula. *Ibidem*, pp. 18 – 19. In Russian.

9. On the logical foundations of the theory of probability. Read 1922. Translated in this collection [iii].

10. Calculation of state revenue from the issue of paper money. An appendix to an article of another author. *Mestnoe Khoziastvo* (Kiev) No. 2, Nov. 1923, pp. 39 – 62. In Russian. Translated in this collection [vi].

11. Mathematical notes to the theory of the issue of paper money. *Ekonomich. Bull. Koniunkturn. Inst.* No. 11/12, 1923, pp. 53 – 60. In Russian. Translated in this collection [vii].

12. On the law of large numbers. *Vestnik Statistiki* No. 7/9, 1925, pp. 1 – 55. In Russian. Translated in this collection [viii].

13. Über stochastische Asymptoten und Grenzwerte. *Metron*, t. 5, No. 3, 1925, pp. 3 – 89.

14. Ein Beitrag zur formal-praxeologischen Grundlegung der Ökonomik. *Acad. Oukrainienne Sci., Annales Cl. Sci. Soc.-Écon.*, t. 4, 1926, pp. 3 – 12. Also in Ukrainian in same year. Translation: An enquiry into the formal praxeological foundations of economics. *Structural Change and Econ. Dynamics*, vol. 15, 2004, pp. 371 – 380.

15. Zur Kritik des Böhm-Bawerkschen Wertbegriffs und seiner Lehre von der Messbarkeit des Wertes. *Schmollers Jahrb. Gesetzgeb., Verwalt., Volkswirt. im Dtsch. Reich*, Bd. 51, No. 4, 1927, pp. 37 – 52. Translation: A critique of Böhm-Bawerk’s concept

of value and his theory of the measurability of value. *Structural Change and Econ. Dynamics*, vol. 15, 2004, pp. 357 – 369.

16. Sur un théorème limite relatif aux séries des quantités éventuelles. *C. r. Acad. Sci. Paris*, t. 185, 1927, pp. 169 – 171.

17. Revised translation of Russian paper of 1927: Summation of random causes as the source of cyclic processes. *Econometrica*, vol. 5, 1937, pp. 105 – 146.

18. Sur un critérium de la convergence stochastique des ensembles de valeurs éventuelles. *C. r. Acad. Sci. Paris*, t. 187, 1928, pp. 370 – 372.

19. Sur les fonctions éventuelles compactes. *Atti del Congresso Intern. dei Matematici* (Bologna, 1928), t. 6. Bologna, 1932, pp. 111 – 115.

20. Sur les fonctions éventuelles continues, intégrables et dérivables dans les sens stochastiques. *C. r. Acad. Sci. Paris*, t. 187, 1928, pp. 878 – 880.

21. Quelques propositions sur les limites stochastiques éventuelles. *Ibidem*, t. 189, 1929, pp. 384 – 386.

22. Sur l'erreur quadratique moyenne du coefficient de corrélation dans le cas des suites des épreuves non indépendantes. *Ibidem*, pp. 612 – 614.

23. Sur l'extension de la théorie de périodogrammes aux suites de quantités dépendantes. *Ibidem*, pp. 722 – 725.

24. On the standard error of the correlation coefficient in the case of homogeneous connected chance series. *Trudy Koniunkturn. Inst.*, vol. 2, 1929, pp. 64 – 101. In Russian. English summary on pp. 150 – 154.

25. On the distribution of the error of the correlation coefficient in homogeneous connected chance series. Translated in this Collection [x].

26. Alcune applicazioni dei coefficienti di Fourier all'analisi funzioni aleatorie stazionarie. *GIIA*, t. 5, No. 4, 1934, pp. 3 – 50.

27. (1924 – 1933). *Desiat Let Raboty dlia Statisticheskoi Geofiziki* (Ten Years of Research in Statistical Geophysics ...). Published by Central Inst. Experimental Hydrology and Meteorology as a mimeographed manuscript. Moscow, 1934.

28. Statistical experiment as a method of research. Critical essays on the Earth – Sun problem. *Zhurnal Geofiziki*, vol. 5, 1935, pp. 18 – 38. In Russian. Translated in this Collection [xiv].

29. On extrapolation in connection with the problem of forecasting. *Ibidem*, pp. 263 – 279. In Russian.

30. On connected random functions of one independent variable. *Trudy Pervogo Vsesoiuznogo S'ezda Matematikov* (Proc. First All-Union Congress Mathematicians). Kharkov, 1929. Moscow – Leningrad, 1936, pp. 347 – 357. In Russian.

31. Qualche proposizione relativa alla teoria delle funzione aleatorie. *GIIA*, t. 8, No. 2, 1937, pp. 3 – 19.

32. Sur les fonctions aléatoires presque périodiques et sur la décomposition des fonctions aléatoires stationnaires en composantes. In *Colloque consacré à la théorie des probabilités*, 5^{ième} pt. (Actualités scientifiques et industrielles No. 738). Eds, S. Bernstein, E. Slutsky, H. Steinhaus. Paris, 1938, pp. 33 – 55.

33. Some propositions on the theory of random functions. *Trudy Sredneaziatskogo Gos. Univ.* (Proc. Sredneasiatsk. State Univ.), ser. Math., No. 31. Tashkent, 1939, pp. 3 – 15. In Russian. Several issues collected, with separate paging, in *Sbornik Posviashchenny 30-letiu Deiatelnosti V. I. Romanovskogo* (Collection Honouring 30 Years of Romanovsky's Work).

34. *Izbrannye Trudy. Teoria Veroiatnostei, Matematicheskaiia Statistika* (Sel. Works in Theory of Probability and Math. Statistics). Moscow, 1960. Includes, among others, reprints or translations of [5; 9; 18 – 24, 26; 30; 32; 33] and a biography compiled by B. V. Gnedenko.

Other Authors

Chetverikov, N. S. (1963), *Statisticheskie i Stokhasticheskie Issledovania* (Statistical and Stochastic Investigations). Coll. Works. Moscow.

--- (1975), *Statisticheskie Issledovania* (Statistical Investigations). Coll. Works. Moscow. Includes B. I. Karpenko's paper (pp. 5 – 19) about him.

Chipman, J. S. (2004), Slutsky's praxeology and his critique of Böhm-Bawerk. *Structural Change and Econ. Dynamics*, vol. 15, pp. 345 – 356.

Chuprov, A. A. (1909), *Ocherki po Teorii Statistiki* (Essays on the Theory of Statistics). References in text from third (last) edition: Moscow, 1959.

Cournot, A. A. (1843), *Exposition de la théorie des chances et des probabilités*. Paris, 1984.

- Kaufman, A. A.** (1916), *Teoria i Metody Statistiki* (Theory and methods of statistics). Moscow. First published in 1911. Several later editions and German translation of 1913. Translation of Slutsky's review [ii].
- Kluikin, P. N.** (2009, in Russian), The RGALI [Russ. State Arch. of Arts and Liter.] documents about E. E. Slutsky's work. *Otechestvennye Arkhivy*, No. 2, pp. 73 – 82.
- Kolmogorov, A. N.** (1933), *Grundbegriffe der Wahrscheinlichkeits-Rechnung*. Berlin.
 --- (1948, in Russian), Obituary. E. E. Slutsky. *Math. Scientist*, vol. 27, 2002, pp. 67 – 74.
- Komlev, S. L., Manellia, A. I.** (1990, in Russian), N. S. Chetverikov. *Voprosy Istorii Narodnogo Khoziastva i Ekonomicheskoi Zisni*, Book 2.
- Kondratiev, N. D.** (1926, in Russian), The problem of forecasting. *Ekonomika i Matematich. Metody*, vol. 24, 1988, pp. 245 – 268.
- Leontovich, A.** (1909 – 1911), *Elementarnoe Posobie k Primeneniu Metodov Gaussa i Pirsona* etc. (Elementary Textbook on the Use of the Gauss and Pearson Methods), pts 1 – 3. Kiev.
- Manellia, A. I.** (1998, in Russian), Life and work of N. S. Chetverikov. *Voprosy Statistiki*, No. 10, pp. 94 – 96.
- Markov, A. A.** (1913), Essai d'une recherche statistique sur le texte du roman «Eugène Onegin». *Izvestia Imp. Akad. Nauk*, ser. 6, t. 7, No. 3, pp. 153 – 162. In Russian with additional French title.
 --- (1924), *Ischislenie Veroiatnostei* (Calculus of Probability). Moscow. Previous editions: 1900, 1908, 1913. German translation: 1912.
- Schulz, H.** (1935), Interrelation of demand, price and income. *J. Polit. Econ.*, vol. 43, pp. 433 – 481.
- Seneta, E.** (1992), On the history of the strong law of large numbers and Boole's inequality. *Hist. Math.*, vol. 19, pp. 24 – 39.
 --- (2001), Slutsky. In Heyde, C. C., Seneta, E., Editors, *Statisticians of the Centuries*. New York, pp. 343 – 345.
- Sheynin, O.** (1990, in Russian), A. A. Chuprov. *Life, Work, Correspondence*. Göttingen, 1996.
 --- (1993), Chuprov, Slutsky and Chetverikov: some comments. *Hist. Math.*, vol. 20, pp. 247 – 254.
 --- (1998), Statistics in the Soviet epoch. *Jahrbücher f. Nationalökonomie u. Statistik*, Bd. 217, pp. 529 – 549.
 --- (2005), Probability and Statistics. Russian Papers of the Soviet Period. Coll. translations of various authors. Berlin. Also www.sheynin.de
 --- (2009), *Studies in the History of Statistics and Probability*. Coll. translations of various authors. Berlin. Also www.sheynin.de
- Smirnov, N. V.** (1948), Slutsky. *Izvestia Akad. Nauk SSSR*, ser. Math., vol. 12, pp. 417 – 420. In Russian.

XX

B. V. Gnedenko, N. V. Smirnov

Foreword

Predislovie. Slutsky E. E. (1960), *Izbrannye Trudy* (Sel. Works). Moscow

The contents of the scientific heritage of the outstanding Soviet mathematician Evgeny Evgenievich Slutsky are very diverse. In addition to mathematics and mathematical-statistical investigations proper, a number of his works are devoted to problems in mathematical economics, some problems in genetics, demography, physical statistics, etc. It seems unquestionable, however, that Slutsky will enter the history of our national mathematics as one of the founders of the theory of stochastic processes, of that branch of the theory of probability which is the main current channel of research stimulated by ever widening demands made by contemporary physics and technology.

Being absolutely specific both in their final goal and approach to solving problems, and distinctively combining these qualities with rigour of mathematical treatment, Slutsky's fundamental contributions to the theory of random functions are an excellent introduction to this topical subject.

These *Selected Works* include all Slutsky's main writings on the theory of random functions and his most important investigations on statistics of connected series. Commentaries adduced at the end of the book trace the numerous links between his work and modern research. A complete [an almost complete] list of his scientific publications is appended. We take the opportunity to express our thanks to Yulia N. Slutsky and N. S. Chetverikov for the materials that they gave us.