Studies

in the History of Statistics and Probability

Vol. 15

Compiled by Oscar Sheynin

Berlin
2020

Contents

Except for one item, the author of all the papers is O. Sheynin

oscar.sheynin@gmail.com

Introduction by the compiler

Notation

Notation **S, G,** *n* refers to downloadable file *n* placed on my website www.sheynin.de which is being diligently copied by Google (Google, Oscar Sheynin, Home. I apply this notation in case of sources either rare or translated by me into English.

I

Oscar Sheynin

Adjustment of a trilateration figure by frame structure analogue

Survey Review, No. 127, vol. 17, 1963, pp. 55 – 56

Referring to the articles of Leung Kui-Wai [9] and F. Halmos [7], I would like to comment on the connection between the laws of mechanics and the adjustment of geodetic observations. Following the line of the article of Halmos, I present several facts from the history of the subject in question.

The abovementioned connection was no doubt understood very early, as is implied by the somewhat "mechanical" term "adjustment" This "mechanical" relationship is also present in the Russian language, and the German *ausgleichen* was derived from mechanics (Gerling, 1843, p. 18).

One of the first applications of statics to the adjustment of (direct) observations was, I understand, due to Rodger Cotes (1722). Robert Adrain, the co-founder of the method of least squares, after arriving at the arithmetic mean as an estimate of direct observations, noted the analogue of this mean with the corresponding centre of gravity [8]. As to the adjustment of networks (indirect observations), S. V. Vissotskij [16], whose article could well be considered as preceding [9], noted that the first application of mechanics was due to S. Wellisch [17], cf. also [7]. Other Soviet contributors are N. I. Tovstoljess [13], [14] and I. N. Temovskikh [12].

The application of mechanics led to the method of (geodetic) relaxation [11, 4]. Curiously enough, some modern contributors to the "mechanical adjustment" only mention relaxation in their references. Relaxation is known in two forms: (i) a non-cyclic one-step iterative process for solving the normal equations; (ii) a process of successive displacement of geodetic points when adjusting networks without the normals being worked out. The first form is due to Gauss [6], who applied it to a problem of station adjustment. Gauss considered his method as appropriate for calculating while half-asleep or reflecting on other matters. A point of special interest is that the system of normal equations was singular, the last normal being equal to the sum of all the previous normals. After Gauss, this method was recommended by L. Seidel [10], whose work appears to be insufficiently known. Besides this, Seidel was the first to use the second form, although only theoretically.

The second farm, more commonly called "method of minimization", appears as a practice] procedure in [3], [4], [15]. Some later contributions are [2] and [18]. This form was recently applied to the adjustment of large blocks of triangulation by an electronic computer [1].

Both forms are used as "group relaxation" and in this connection the article of K. Chow [5] (as well as of course [3]) ought to be mentioned.

References

1. A. A. Abramov, S. L. Khublarova, " On the Iterative Solution of Systems of Linear Algebraic Equations Originating in the Adjustment of Geodetic Networks ". Transactions Central Research Institute Geodesy, Air Survey and Cartography, No. 135 (1960). Also Journal of Abstracts " Geodesija ", No. 8, abstract 80.234, (1961).
2. K. Arnold, "Die Bestimmung der Gewichtsreziproken durch das Minimisieren einer gegebenen Funktion ". Z. für Vermessungswesen, 112 (1959).
3. A. N. Black, "The Method of Systematic Relaxation Applied to Survey Problems". Empire Statistical Survey, vol. IV, No. 29 (1938).
4. A. N. Block, R. V. Southwell, "Relaxation Methods Applied to Engineering Problems ", Proc. Roy. Soc., A, vol. 164, No. 919 (1938).
5. K. Chow, "The Great Scale Triangulation Adjustment by Groups ", Acta Geodetica et Cartographica Sinica, No. 1 (1958).
6. G. E. Forsythe, "Gauss to Gerling on Relaxation ". Malhematical Tables and Other Aids to computers, No. 36 (1951).
7. F. Holmes, "Adjustment of a Trilateration Figure by Frame Structure Analogue". Empire Survey Review, vol. 16, No. 125 (1962).
8. E. Hammer, "Beitrag zur Geschichte der Ausgleichungsrechnung ". Z. für Vermessungswesen, Bd. 29, H. 24 (1900).
9. Leung Kui-Wai, "Adjustment of a Trilateration Figure by Frame Structure Analogue ". Empire Survey Review, vol. 16, No. 123 (1962).
10. L. Seidel, " Über ein Verfahren, die Gleichungen durch Successive Annäherung Aufzulösen", Abh. Bayerische Akademie Wissenschaften, Math.-Phys. Kl., Bd. 11, Abt. 3 (1874).
11. R. V. Southwell, "Stress Calculations in Frameworks by the Method of Systematic Relaxations of Constraints". Proc. Roy. Soc. A, vol. 151, No. 872, vol. 153, No. 878 (1935).
12. I. N. Temovskikh, "Application of the Method of Structural Mechanics to the Adjustment of Trilateration Control Nets". Proc. Higher Educ. Inst., Series "Geodesy and Air Survey", No. 1 (1962). Also Journal of Abstracts Geodesija, No. 12, abstract 12G230 (1962).
13. N . I. Tovstoljess, "The Methods of Structural Mechanics as Applied to the Solution of Problems of Geodesy ". Trans. Institute Structural Mechanics Acad. sci. Ukrainian SSR, No. 19 (1954). Also Journal of Abstracts Geodesija, No. 2, abstract 797 (1955).
14. N; I. Tovstoljess, "Application of Graphical Statics to the Solution of Problems of Geodesy". Trans. Kiev Automobile and Highway Institute, No. 2 (1955).
15. J . Vignal,"Rapport sur la Compensation des Réseu de Nivellements par la. Méthode d'Approximations Successives de Southwell – Black ". Bull. géodésique, No. 62 (1939).
16. S. V. Viss)otskij, "Adjustment of Geodetic Observations by Methods of Engineering Structures ". Trans. Moscow Institute Engineers Transport, No. 4 (1927).
17. S. Wellisch, *Fehlerausgleichung nach der Theorie des Gleichgewichtes Elastischer Systeme* ". Wien, 1904.
18. H. Wolf, " Ausgleichen ohne Zuhilfenahme von Normalgleichungen ". Vermessungstechnische Rundschau, Bd.. 21, H. 12 (1959).

In the above, [l], [12], [13], [14] and [16] are in Russian and prior to 1963 the Journal of Abstracts Geodesija was a separate issue of the Journal of Abstracts Astronomia i Geodesija.
"

II

O. Sheynin

On the early history of the law of large numbers

Biometrika, vol. 55, 1968, pp. 459 – 467.
Reprint: E. S. Pearson, M. G. Kendall (1970),
Studies in History of Statistics and Probability. London, pp. 231 – 239

Summary

This paper is devoted to the early history of the law of large numbers. An outline of the prehistory of this law is given in § 1. The algebraic part of J. Bernoulli's theorem is presented in a logarithmic form and. the lesser known role of N. Bernoulli is described in § 2.

Comments on the derivation of the De Moivre – Laplace limit theorems by De Moivre, in particular, on the inductive character of his work, on the priority of De Moivre as to the continuous uniform distribution, on the unaccomplished possibility of Simpson having arrived at the normal distribution and on the role of Laplace are presented in § 3. The historical role of J. Bernoulli's form of the law of large numbers is discussed in § 4.

l. Prehistory of the law of large numbers

The most rudimentary form of the law of large numbers is to be credited to Cardano (Ore, 1963, p. 150), who held that *in the long run* the number of occurrences of an event in n independent trials is approximately equal to

$$\mu = np, \tag{1}$$

where p is the constant probability of the occurrence of this event in one trial. This *reasoning on the mean outcome*, as Ore called it (p. 145), was systematically used by Cardano and, again as Ore pointed out, led him in some cases to erroneous results.

Halley (1693, p. 484) stated that the reason for the frequencies of mortality of different age groups being irregular *seems rather to be owing to chance, as are also the other irregularities in the series of age, which would rectify themselves, were the number of years* (of observation on the studied population of Breslaw) *much more considerable ...*

Such assertions were possibly also made by other scholars, but what distinguishes Halley is that he adjusted the frequencies concerned so that they would more nearly correspond to their general trend and therefore be applicable to populations other than the population of Breslaw; see also Graetzer (1883, pp. 77 – 78). However, irregularities could have been produced by systematic influences. More comment on Halley is made in § 3.

A *reasoning on the mean*, differing from Cardano's, occurs in a sixteenth-century commentary by Ganésa on a still earlier Indian mathematical text. Commenting on the calculation of the volume of an irregular earth excavation considered to be equal to the product of the mean measures of the length, width and depth of the excavation, with the measures taken at different places, Ganésa pointed out (Colebrooke, 1817, p. 97) t*hat the greater the number of the places* (of measurement), *the nearer will the mean measure be to the truth and the more exact will be the consequent computation*.

The mean measures were introduced to compensate for inaccuracies in the mathematical model. The application of the arithmetic mean for the same purpose can be traced to ancient Babylonia where, for example (Veiman, 1961, p. 204), the area of a quadrangle was held to be equal to the product of the half-sums of its opposite sides. Though the commentary of Ganésa should be specially noticed because of his reference to the increase of the number of measurements, it is to be inferred that, strictly speaking, his reasoning is of a deterministic (not probabilistic) kind and its formalization would have led to integral sums.

Qualitative assertions for the preference of the arithmetic mean of several observations over a single observation in astronomy and geodesy are found in the works of the seventeenth to eighteenth centuries, notably of Cotes (1722/1768, pp. 57 – 58). But these assertions, although undoubtedly of a probabilistic nature, are concerned only with a given set of observations and at least until the second half of the eighteenth century no mention was made of the effect of increasing the number of observations.

The general' impression seems to be that the prehistory of the law of large numbers contains an understanding of the nature and the use of formula (1) and of the arithmetic mean of a given set of observations. The matching and further formalization of these separate ideas were however, due to Jakob Bernoulli.

2. Jakob and Nicholas Bernoulli

J. Bernoulli (1713, German version 1899) proved that as $n \to \infty$

$$\lim \text{prob}\,(|\frac{\mu}{n} - p| < \varepsilon) = 1.$$

His work is described notably by Todhunter (1868, §§ 123 – 124) and Pearson (1925). Bernoulli also gave an example from which it directly followed that the sum of $2n$ middle terms of the expansion of $(r + s)^{(r+s)n}$, $r = 30$, $s = 20$ (even excluding the middlemost term) will be more than $c > 0$ times the sum of the other terms of the expansion if the number of trials

$$nt = n(r + s) \geq 25{,}500 + 5758\log(c/1000) = 8226 + 5758 \log c. \quad (2)$$

After the death of J. Bernoulli, but prior to the publication of the *Ars Conjectandi*, N. Bernoulli (Montmort, 1713, p. 388) deduced an approximate estimate for the ratio of the middle part of a binomial series to its other parts and, taking Arbuthnot's (1712) data, used this estimate for probabilistic reasoning about the *constant regularity observed in the births of both sexes*.

N. Bernoulli's estimate was rather crude, but it seems that he was the first to study how the probability that a random quantity falls in an interval depends on the length of that interval. In more detail, let n be the annual number of male births, with the ratio of male and female births equal to m:f. Assuming a binomial distribution, N. Bernoulli gives the following estimate which, as is also the case with J. Bernoulli, is equivalent to a local limit theorem

$$P(|\mu - rm| \le s) \approx (t-1)/t,$$
$$t \approx [1 + s(m+f)/mfr]^{s/2} \approx \exp[s^2(m+f)^2/2mfn],$$
$$P(|\mu - rm| \le s) \approx 1 - \exp(s^2/2pqn),$$

$$P[|\mu - np|/\sqrt{npq} \le s] \approx 1 - \exp(-s^2/2).$$

Here, $p = m/(m+f)$, $q = n/(m+f)$.

The last lines of N. Bernoulli' s letter to Montmort (*Quand ce Livre paroîtra nous verrons si dans ces sortes de inatiéres j'ai trouvé une approximation aussi juste que lui*) prove that he, possibly more interested in problems of 'moral probabilities' did not at that time at least pay sufficient attention to Bernoulli's law of large numbers.

Later research proved that Nicholas had plagiarised J. B.

3. Derivation of the De Moivre – Laplace limit theorems

3.1. De Moivre

A further development of the law of large numbers leading to the De Moivre – Laplace limit theorems came with De Moivre. General information about De Moivre is in many places, all of which, nevertheless, draw on the two main sources (*Eloge*, 1759; Maty, 1760). His main work, *Doctrine of Chances*, appeared in 1718, 1738 and, posthumously, in 1756. The last two editions have recently been reproduced; this information was recently received from Dr C. Eisenhart. [Last edition reprinted: New York, 1967.] The relevant work of De Moivre (Method of Approximating the Sum of the Terms of the Binomial $(a + b)^n$ ...) has been sufficiently described by Pearson (1924, 1925) and Archibald (1926); I have only a few comments to offer.

(i) De Moivre drew heavily on his book *Miscellamea Analytica de Seriebus et Quadraturis* (1730), an English translation of which is long overdue. [French translation: Paris, 2009.] It was there, partly in its Supplement, that all the main algebraic deductions were made.

(ii) It was also there, again in the Supplement, that De Moivre first published his 14-digit table of log$n$! for n = 10(10)900. The table is correct to 11 – 12 digits with a single misprint in the fifth digit of log 380!. I have compared this table with several modern tables, notably that of Peters (1922).

Both Pearson and Markov (1924) consider that the approximate formula for n! was derived by Stirling and De Moivre and should therefore be named *the De Moivre – Stirling formula*, while Archibald attributes it to De Moivre. If De Moivre's table is taken into account, as it should be, the opinion of Archibald is substantially strengthened.

(iii) The important appearance of the estimate of accuracy, $n^{-1/2}$, in De Moivre's work was originally occasioned by a pure algebraic fact: $x = n^{1/2}/2$ was a bordering value for two different ways of integrating.

De Moivre employed a power series when $l \leq n^{1/2}/2$ and, because of the slow convergence of this series, otherwise, applying the approximate method of Newton – Cotes.

(iv) De Moivre experimentally checked the accuracy of his formula for $n = 900$ and 100 (*Method of Approximating* ..., Corol. 5). Although he gave no indication about the nature of his experiments and although these experiments seem to be well within mathematics (not exact sciences at large), the fact that he made these checks is interesting. The opinion (Walker, 1934, p. 320) that *even in his* (De Moivre's) *writings on the* Doctrine of Chances *his work is deductive and he does not set up experimental checks on the outcome* is hardly fair.

But a more important comment in this connexion is that the whole *Method of Approximating* was clearly intended for *experimental checks*, as is proved by the corollary to Problem 87 in the 1738 edition of the *Doctrine of Chances* (this is the problem after which immediately follows the *Method*):

…if after taking a great number of experiments, it should be perceived that the happenings and failings have been nearly in a certain proportion … it may safely be concluded that the probabilities of happening or failing at any one time assigned will be very near in that proportion, and that the greater the number of experiments has been, so much nearer the truth will the conjectures be that are derived from them.

But suppose it should be said, that notwithstanding the reasonableness of building conjectures upon observations, still considering the great power of chance, events might at long run fall out in a different proportion from the real bent (according to) *which they have to happen one way or the other; and that supposing for instance that an event might as easily happen as not happen, whether after three thousand experiments it may not be possible it should have happened two thousand times and. failed a thousand, and that therefore the odds against so great a variation from equality should be assigned, whereby the mind would be the better disposed in the conclusions derived from the experiments:*

In answer to this, I will take the liberty to say that this is the hardest problem that can be proposed on the subject of chance, for which reason I have reserved it for the last … I shall derive … some conclusions that may be of use to every body: in order thereto, I shall here translate (*from the Latin*) *a paper of mine which was printed November 12, 1733, and communicated to some friends, but never yet made public, reserving to myself the right of enlarging my own thoughts, as occasion shall require.*

This quotation means that De Moivre attempted to reconcile statistical and prior probabilities.

(v) A special comment on the theological reasoning of De Moivre's contemporaries is warranted. It is true that, as Pearson points out, De Moivre and other Fellows of the Royal Society were greatly influenced by Newton's theology. But at least one mathematician, Simpson, who, moreover, humbly refers to De Moivre in the introduction to *Nature and Laws of Chance* (1 740), did not follow Newton's theology.

3.2. Simpson

It is hardly appropriate to describe the works of Simpson in detail in this paper. But it ought to be said that Simpson arrived at a continuous triangular distribution, proved that with this distribution the arithmetical mean is preferable to a single observation and went on to deduce the probability of a given error in the arithmetical mean in the limiting case. From this he could have arrived at a normal distribution (as pointed out in a private communication from L. N. Bolshev) although later than De Moivre and could have been the first to draw a graph of this distribution. His failure is evident in that his limit curve (Fig. 20) does not behave as a normal one.

General information about Simpson is given by Clarke (1 929). Referring to his correspondence with the Royal Society, Penkov (1961, pp. 300, 302) states that there exists no portrait of Simpson.

[Pearson (1978, pp. 145 and 184) described Simpson's later attitude to De Moivre and called him *a most disreputable character* and *an unblushing liar and a thorough knave at heart*.]

3.3. Continuous uniform distribution credited to De Moivre

The tercentenary of the birth of De Moivre occurred in 1967 and it is opportune to add two short comments not connected with the law of large numbers.

(i) While considering one of his problems in games of chance De Moivre followed the same method which Poisson followed when he originally arrived at his distribution (Newbold, 1927, pp. 490 – 491).

(ii) While calculating annuities on lives De Moivre arrived at the continuous uniform distribution and used the corresponding first moment. (I have seen the 1756, posthumous edition of the *Treatise of Annuities on Lives*, incorporated in the *Doctrine of Chances*, 1756, and the 1743 edition of that treatise.) In problem 20, part 1 of the *Treatise*, he calculated the expectation of life:

$$\int_0^n x dx/n = n/2.$$

Only the result is given. Here n is the complement of life. The uniform distribution appeared as an empirical law corresponding to Halley's data (1693).

In Chapter 8 of Part 2, De Moivre calculates the probability of one person with a complement of life equal to n outliving another person whose complement of life is $p < n$:

$$P(\xi \ge x, \eta = x) = [(n - z)/n]dz/p, \; P(\xi > \eta) =$$
$$\int_0^p [(n - z)/n]dz/p = 1 - p/2n.$$

This result could have been arrived at geometrically. In problem 21, part 1, De Moivre calculated the expectation of two joint lives (again, only the result is given):

$$P(x \le \xi \le x + dx \text{ or } x \le \eta \le x + dx) =$$

$$[(n-x)/n]dx/p + [(p-x)p]dx/n,$$

$$E\zeta = \int_0^p \{[(n-x)/np] + [p-x)p/n]\}dx = p/2 - p^2/6n.$$

This reconstruction is due to Czuber, see Note 22 to the German translation of 1906 of De Moivre's work.

3.4. Laplace

De Moivre had proved the theorems which are now known as the De Moivre—Laplace limit theorems (§ 3.1), but even Laplace did not introduce the concept of uniform convergence which is of a later origin. Laplace actually repeated these proofs, and, characteristically for his works in which his own contributions are not readily separated from those of his predecessors, did not refer to De Moivre, giving only a barest possible outline of his work in the historical part of his *Essai philosophtque.*

It is worth noting that Laplace, who was really able, as no one else, to grasp and elaborate the ideas of his predecessors, in his first probabilistic memoir (1774) deduced an exponential distribution

$$\varphi(x) = \frac{1}{2}me^{-m|x|}.$$

A function of this kind might be deduced by reversing (2), the possibility of which was not mentioned by J. Bernoulli. But (2) is a deterministic not probabilistic formula and therefore this is an example, possibly superfluous, of the way Laplace extended the ideas of his predecessors.

Lastly, we note that the correspondence of Lagrange and Laplace (Lagrange, 1892; see letter to Laplace dated 30 December 1776 on p. 66) testifies that they both, independently, had contemplated translating De Moivre's *Doctrine of Chances* into French. Only Lagrange is mentioned by Todhunter in this connection.

The appearance of the normal law in the work of De Moivre was not noticed until Eggenberger (1894, especially p. 165) referred to it, possibly because this was not mentioned by Laplace. Although noted by Czuber (1899), Eggenberger's work seems to be little known, perhaps because Haussner called it unclear. Haussner's remark appears on pp. 158 – 159 of the *Ostwald Klassiker*, no. 108, see Bernoulli (1899). And Pearson (1924) independently from Eggenberger credited the normal law to De Moivre.

[De Morgan noted De Moivrte's discovery before Eggenberger, see Sheynin (2017, p. 66). On pp. 133 – 134 I put on record De Morgan's unimaginable statements.]

4. The role of the law of large numbers

4.1. Statistics

This law became the stepping stone between the theory of probability and statistics. The main problem of population theory after Quetelet, the problem of the stability of statistical frequencies, came down to the testing of the existence of preconditions for the law of large numbers in Bernoulli's form and was solved by Lexis and especially by Bortkiewicz (1898) . The latter used the Poisson form of the law. Bortkiewicz (1905, p. 140) praised Lexis as *the first to establish the integral connexion* between theoretical statistics and. the theory of probability. It seems, however, that initial connections were made as early as in the 1850's. For example, following Poisson, Davidov (1855b) tested the statistical significance of various empirical inequalities. In another paper Davidov (1855a) stated that the *excessive development* of statistics, and its deductions, *often unfounded*, can smear statistics and that a discussion of the initial statistical data with probabilistic verification is the most reliable tool for eliminating *immature deductions*. One of his probabilistic tools was the local De Moivre – Laplace limit theorem.

We note a special passage on the law of universal gravitation (Davidov, 1855b, p. 63) in a context devoted to philosophical problems of empirical proofs:

Who is in a position … to state that this law is an exact expression of the law of nature, and that it is not a particular case of a more general principle, only approximately correct?

For general information about Davidov (1823 – 1885) see Zhoukovsky et al (1890).

Similarly, Cournot (1843, or a German translation by Schnuse, 1849) whose book appears to be insufficiently known, reasons on the possibility of *smearing* statistics (§ 103), tests the statistical significance of empirical inequalities (§§ 108 – 110) and considers the law of large numbers in J. Bernoulli's form a reliable base for connecting statistics with probability (§ 115).

4.2. Classical theory of errors

The classical theory of errors originated in the middle of the 18th century, especially at the hands of Lambert (Sheynin, 1966) and Simpson. But the concept of random errors of observation which Simpson effectively introduced became divorced from the concept of random variables of the theory of probability.

A qualifying remark should be added: a definition of a random variable did not seem to appear in the classical theory of probability. But we hold that in the second half of the nineteenth century a definition of a random variable as *dependent on chance* and possessing a certain law of distribution had become so natural, even if not definitely formulated, that this definition, just as the *classical* definition of probability, ought to be called classical.

As to random errors, these were usually taken to be errors with certain probabilistic properties, their specific distribution especially following the *Theoria Combinationis* way of reasoning, being not so important.

It seems that Vassiliev (1885, p. 133) was the first who definitely held that random errors of observations are ranked among random variables. As far as the theory of probability is concerned, A. V. Vassiliev (1853 – 1929) is known primarily as one of Markov's correspondents. It was in a letter to Vassiliev that Markov (1898) originally described his reasoning on the method of least squares. According to Markov, references to the law of large numbers in the theory of errors suddenly became even too numerous and the law was misused (p. 249). Although also connecting random errors with random variables in his other works, Markov stated (1898, p. 250) that *nothing comes* from

$$\mathrm{Prob}(|\bar{x} - a| < \varepsilon) \to 1 \text{ as } n \to \infty$$

because other linear estimates of the constant *a* possess the same limit property of consistency as the arithmetic mean $\bar{x}$ and because, furthermore, a method optimal in the case of a finite number of observations is needed. Such a method, according to Markov (p. 246) was the second (*Theoria Combinationis*) method of Gauss as opposed to his first (*Theoria Motus*) method. On the work of Markov in the theory of errors, see also Plackett (1949).

The above description of the law of large numbers in the theory of errors seems somewhat inconsistent, but it reflects different points of view held at different times. In particular, it is our guess that the sudden interest in the law of large numbers which occurred at the turn of the 19th century, and to which Markov refers, was an inevitable result of the ideas of Lexis and other *continental* statisticians and of the work of Bienaymé and Chebyshev becoming generally known.

4.3. Probabilities proper: Markov versus Pearson

Markov (1924) studied Bernoulli's work on the law of large numbers simultaneously with or somewhat prior to K. Pearson (1924, 1925). Markov held a high opinion on this law; he edited a Russian translation (1913) of Part 4 of the *Ars Conjectandi*, was the originator of a special sitting of the Russian Academy of Sciences in commemoration of the bicentenary of the law of large numbers (where speeches were given by Markov himself, Chuprov (1914) and Vassiliev) and dedicated the third edition of his *Ischislenje veroyatnostey* (1913) to the memory of J. Bernoulli.

Pearson (1924) considered J. Bernoulli's estimate (the necessary number of trials) and, furthermore, the whole part 4 of the *Ars Conjectandi* as unsatisfactory. This opinion is hardly fair: the law of large numbers in Bernoulli's form was of great significance for the whole development of the theory of probability at least until the time of Laplace and Poisson, and the *crude values* of his estimate *with their 200 to 300%* excesses (Pearson) being no serious obstacle. Similarly, this law proved itself of utmost importance in applications of the theory of probability (§§ 4.1 and 4.2).

As opposed to Pearson, Markov, while modernizing Bernoulli's algebraic deductions and improving his estimate, did not use Stirling's theorem which remained unknown to Bernoulli. Consequently, Markov's estimate turned out to be worse than Pearson's, but because of this very reason his approach seems to be methodologically more correct. [Markov applied that theorem somewhat below but without warning the readers about that addition.]

A special feature of the Markov's review is that he eliminated J. Bernoulli's tacit condition that the exponent $(r + s)n$ is divisible by the sum of the terms of the binomial $(r + s)$.

No attempt is made here to describe the original work of Markov in the field of the law of large numbers.

References

ARBUTHNOT, J. (1712). An argument for divine Providence etc. *Phil. Trans. Roy. Soc.* 27, no. 328 for 1710. Also in M. G. Kendall & R. L. Plackett (1977), *Studies in the History of Statistics and Probability*, vol. 2. London, pp. 30 – 34.

ARCHIBALD, R. C. (1926). A rare pamphlet of De Moivre and some of his discoveries. *Isis* 8, no. 4 (28), pp. 671 – 684. Also in 1926 a discussion with K. Pearson in *Nature*, 117, pp. 551 – 552 entitled *A. De Moivre*.

BELLHOUSE, D. R., GENEST, CHR. (2007), Maty's biography … translated. *Stat. Sci.*, 22, pp. 109 – 136.

BERNOULLI, J. (1899). *Wahrscheinlichkeitsrechnung* (*Ars Conjectandi*). Ostwald Klassiker no. 107—8. Herausgegeben R. Haussner. Leipzig.

Bortkiewicz, L. (1898). *Das Gesetz der kleinen Zahlen*. Leipzig: Teubner.

--- (1905). On statistical regularity. (Russian). *Vestnik Prava* 35, No. 8, 124 – 154.

CLARKE, F. M. (1929). *Thomas Simpson and His Times.* A thesis submitted to the Columbia University. New York. Dr A. E. Ritchie (Ames, Iowa, U.S.) kindly sent to me its microfilm.

COLEBROOKE , H. T. (1817). *Algebra and Mensuration from the Sanscrit of Brahmegupta and Bhascara*. London: Murray. [Wiesbaden 1973.]

COTES, R. (1768). Aestimatio errorum, etc. In: *Opera misc*. pp. 10 – 58. London: Typis Meyerianis. Originally published in 1722.

COU'RNOT, A. A. (1843, 1984). *Exposition de la théorie des chances et des probabilités*. Paris. **S, G,** 54.

CZUBER, E. (1899). Die Enwicklung der Wahrscheinlichkeitstheorie, etc. *Jahresber. deustch. Mathematiker-Vereinigung* 7, no. 2 (separate pagination).

DAVIDOV, A. JU. (1855 a). Application of the theory of probability to statistics. (Russian). In: *Scientific and Literary Papers of the Professors and Instructors of the Moscow Univ*. Published on the Occasion of its Centenary Jubilee. (Separate pagination for each paper.) Moscow.

--- (1855b). Use of the deductions of the theory of probabilities in statistics. (Russian). *Journal Ministerstva Narodnogo Prosveschenia* vol. 88, No. 11, pp. 45 – 109 of section 2.

DE MOIVRE, A. (1906). *Leibrente*. Herausgegeben E. Czuber. *Sonderheft der Versicherungswissenschaftlichen Mitt*., Wien.

EGGENBERGER, J. (1894). Beiträge zur Darstellung des Bernoullischen Theorems, etc. *Mitt. Naturforsch. Ges. Bern* for 1893, No. 1305 – 1334, pp. 110 – 182. Also published as a separate edition (1906), Berlin: Fischer.

ELOGE DE DE MOIVRE (1759). *Hist. Acad. Roy. Sci. Paris* for 1754, pp. 175 – 184. (Incorporated in the same volume with the *Mém. Acad. Roy. Sci.*)

GBAETZER, J. (1883). *E. Bailey und G. Neumann.* Breslau: Schottlaender.

HALLEY, E. (1693). An estimate of the degree of mortality of mankind, etc; *Phil. Trans. Roy. Soc. 1665 – 1800 Abridged*, vol. 3. London, 1809. Baltimore, 1942.

LAGRANGE, J. L. (1892). *Oeuvr. Compl.*, t. 14. Paris: Gauthier.

LAPLACE, P. S. (1774). Mémoire sur la probabilité des causes par les événements. *Oeuvr. Compl.*, t. 8, pp. 27 – 65. Paris: Gauthier (1891).

MARKOV, A. A. (1898). The law of large numbers and. the method of least squares. (Russian). This is an extract of a letter to Vassiliev. Reprinted (1951) in Markov' s *Selected Works*, pp. 231 – 251.

--- (1914). Bicentenary of the law of large numbers. (Russian). *Vestnik opitnoy physiki i elementarnoy matematiki*, No. 603, pp. 59 – 64.

MABKOV, A. A. (1924). *Ischislenie veroyatnostey*, 4th (posthumous) Russian edition. A German (1912) version of an earlier Russian edition: *Wahrscheinlichkeitsrechnung*. An English translation was being arranged by F. M. Weida for the *Ann. Math. Stat. Suppl.* (stated by F. N. David & J. Neyman (1938).

MATY, M. (1760). *Mémoire sur la vie de De Moivre*. La Haye. Unavailable. An anonymous obituary of De Moivre is in *J. Britannique* (vol. 18 for Sept. – Oct. 1755, pp. 1 – 51, La Haye, 1755), a journal edited by Maty. This source was found by A. De Morgan (1914), *Essays on the life and work of Newton*, Chicago and. London: Open Court publ., p. 189, who states that (i) the author is Maty and that (ii) the 1760 memoir of Maty is a second edition of this. Engl. translation: Bellhouse, Genest (2007).

MONTMORT, P. R. (1713). *Essay d'analyse sur les jeux de hazard.* Paris: Quilau. See letter of N. Bernoulli to Montmort dated 23 Jan. 1713.
NEWBOLD, E. N. (1927). Practical applications of the statistics of repeated events, etc. *J. Roy. Stat. Soc*. 90, pp. 487 – 547.
ORE, 0. (1963). *Cardano, the Gambling Scholar*. Princeton Univ. Press.
PEARSON, K. (1924). Historical note on the origin of the normal curve of errors. *Biometrika* 16, pp. 402 – 404.
--- (1925). James Bemoulli's theorem. *Biometrika* 17, pp. 201 – 210.
PEN'KOV, B. (1961) On the history of mathematics in 18th century England: Bayes and Simpson. (Bulgarian). *Phys.-Math. Spisanie* 4 (37), No. 4, pp. 292 – 303.
PETERS, J. (1922). *Zehnstellige Logarithmentafeln*, Bd. 1. Anhang, Taf. 6. 18-stellige log*n*! Berlin: Reichsamt für Landesaufnahme.
PLACKETT, R. L. (1949). A historical note on the method of least squares. *Biometrika* 36, pp. 458 – 460.
SHEYNIN, 0. B. (1966). Origin of the theory of errors. *Nature*, vol. 211, pp. 1003 – 1004.
SHEYNIN, 0. B. (2017), *Theory of Probability. Historical Essay*. Berlin. **S, G,** 10.
SIMPSON, T. (1757). An attempt to show the advantage arising by taking the mean, etc. In author's *Misc. Tracts on Some Curious Subjects*, etc., pp. 64 – 75. London: Nowurse. First version: 1756.
TODHU'NTER, I. (1863). *History of the Mathematical Theory of Probability*. Cambridge and London: Macmillan. New York, 1949, 1965.
TCHOUPROV (CHUPROV), A. A. (1914). The law of large numbers in modern science. (Russian). *Statistichesky Vestnik*, books No. 1 – 2, pp. 1 – 21. Another version of this appeared in *Nord. stat. tidskr*. 1, No. 1 (1922). It proved available only in a Russian translation (1960) in author's *Voprosy Statistiki*. Moscow.
VASSILIEV. A. V. (1885). *Theory of Probabilities*. (Russian, lithographic edition). Kazan.
VEIM.AN, A. A. (1961). *Shumero-Babylonian Mathematics of Third – First Millenia B. C.* (Russian). Moscow: Oriental liter. publ.
WALKER, H. M. (1934). A. De Moivre. *Scripta Math*. 2, No. 4, pp. 316 – 333. Reprinted in the 1967 edition of De Moivre's *Doctrine* …
ZHOU'KOVSKY, N. YE. et al (1890). The life and works of Davidov. (An obituary.) (Russian). *Mathematichesky Zbornik* 15, No. 1, pp. 1 – 57.

Later note (only inserted in 1970)

About Davidov's statement on the law of universal gravitation: the same idea and, roughly, in the same words, is in a Russian translation of Quetelet's *Social Physics* (1869, Book 1, § 8).

III

O. B. Sheynin

Daniel Bernoulli on the normal law

Biometrika, vol. 57, 1970, pp. 199 – 202.
Reprinted: Sir Maurice Kendall, R. L. Plackett,
Studies in History of Stat. and Prob., vol. 2. London, 1977, pp. 101 – 104

Summary

This paper discusses one of D. B. memoirs in which he deduced the De Moivre – Laplace limit theorems, nevertheless credited to De Moivre. The memoir is described in § 2 while § 1 attempts to sum up Bernoulli's contributions more generally.

1. General

Between 1738 and. 1778 Daniel Bernoulli (1700 – 1782) published seven probabilistic memoirs. Their essence except the memoir described in § 2, is given by Todhunter (1865). The memoirs contain solutions of important problems in demographic statistics (political arithmetic) and astronomy obtained with the help of probabilistic ideas and methods. As to probability and mathematical statistics proper, Bernoulli was the first to use systematically differential equations for deducing a number of formulas, one of the first to raise the problem of testing statistical hypotheses and the first to introduce *moral expectation* (due to Cramer) and to study random processes. He is also to be credited, after Lambert, for the second introduction of the maximum likelihood principle (Bernoulli 1961). In summary, it may be argued that D. Bernoulli's influence upon Laplace, especially concerning applications of probability, was comparable to that of De Moivre.

The account of Bernoulli's memoirs given by Todhunter could well be modernized but the present paper is restricted to the description of the 1770 – 1771 memoir, the second part of which remained unnoticed by Todhunter. For this and other reasons, Todhunter's account of the memoir is unsatisfactory and until now no one has remarked on the appearance in this memoir of the De Moivre – Laplace limit theorems and of the first published small table of the normal distribution. Had these limit theorems been noticed in Bernoulli before, they possibly would not now be called only after De Moivre and Laplace.

2. The normal distribution and the De Moivre – Laplace limit theorems

[It is too difficult to reproduce the numerous formulas contained here Instead, I recommend to study those which are inserted in Sheynin (2017, pp. 72 – 73). However, I retain here some texts.]

2.1. The formula of Wallis

The Wallis formula was known, Euler used it in 1748. It may be inferred that Bernoulli had forgotten it. It is more difficult to explain the total lack of references to De Moivre. The title of D. B.'s memoir includes the expression *Mensura Sortis*, which coincides with the title of De Moivre's memoir published in 1712. Furthermore, Bernoulli arrives at the normal distribution (see below), already known to De

Moivre, by starting from the same formula as the latter. And if the practice of Laplace is remembered, one really could infer that the lack of references to predecessors was characteristic of those times.

However, it is also possible that Bernoulli, although having read De Moivre, perhaps some thirty years before 1770, just did not pay due attention to the essence of his work.

2.2. The Normal Law

Bernoulli used the integral De Moivre – Laplace theorem with summation instead of integration. He did not note that with a larger value of *a/b* the form of his curve would have changed, i. e. that this curve is specified by an important parameter (standard deviation) and, as is also the case with De Moivre, paid little attention to the curve itself.

2.3. Table of normal distribution

In contrast with De Moivre, Bernoulli computed a small table of the normal curve, the first ever published. It is compiled for $\exp(-\mu^2/100)$ where μ is the excess of male births and $\mu = 1(1)5$ and $10(5)30$ with four significant digits. In three cases the error of the last digit is unity and in one case it equals two.

2.4. Testing statistical hypotheses

Bernoulli concludes his memoir with a table of male and female births in London, 1721 – 1730, compares it with values of $a/b = 1.055$ and 1.040. Although he doubted the constancy of a/b in time and. space, his goal was to find the 'real' value of this ratio.

He thus raised the question of choosing one or another value of a statistical parameter. He noted the signs of the deviations between computed and observed values of male births, singled out deviations with absolute values less than 47, noticed the prevalence of deviations of one sign with $a/b = 1.055$ and of the opposite sign with $a/b = 1.040$ but reasonably failed to make a definite selection. However, already the raising of the question of testing statistical hypotheses seems to be very important. It should be emphasized that he returned to this question in 1778 (Bernoulli, 1961) and that the method of differential equations, twice used in this memoir, had been extensively used by him elsewhere.

Acknowledgement is due to Dr L. N. Bolshev and to Prof. A. A. Yushkevich for advice and corrections on the first version of this paper.

References

BERNOULLI, Daniel (1770 – 1771). Mensura sortis ad fortuitam successionem rerum naturaliter contingentium applicata. *Werke*, Bd. 2. Basel, 1982, pp. 326 – 338, 341 – 360.

--- (1961). The most probable choice between several discrepant observations and the formation therefrom of the most like induction. *Biometrika* 48, pp. 3 – 13.

SHEYNIN, 0. B. (2017), *Theory of Probability. Historical Essay*. Berlin. **S, G,** 10.

TODHUNTER, I. (1865). *History of the Mathematical Theory of Probability*. Cambridge and London: Macmillan. New York, 1949, 1965.

IV

Oscar Sheynin

Lies, damned lies and statistics

Intern. Z. f. Geschichte u. Ethik der Naturwissenschaften, Technik u. Medizin, Bd. 11, 2003, pp. 191 – 193

In the 19th century, doubts about statistical data and/or conclusions were being expressed time and time again. Cournot (1843, p. 123) warned against "les applications prématurées et abusives" of statistics that can "la décréditer pour un temps". Quetelet and Heuschling (1865, p. LXV) concluded that a "document statistique" was not "certain" but only "probable" and John (1883, p. 672) complained that statistical data were excessive and unreliable.

Public figures apparently had similar misgivings. Thomas Carlyle (1795 – 1881) reported that "a witty statesman said, you might prove anything by figures" (Carlyle 1839, p. 122). And he continued on p. 125 that statistical inquiry "throws not light, but error worse than darkness". Saenger (1935, p. 452) maintained, although without proof, that Bismarck (1815—1898),"für die Statistik sehr wenig übrig hatte und sie eigentlich für entbehrlich hielt".

The first appearance of the maxim that I chose as the title of this note was long ago traced to Mark Twain (1959, Chapter 29), real name Samuel Langhorne Clemens (1835 – 1910) who testified that it was attributed to Benjamin Disraeli (1804 – 1881). In this connection I quote Arthur White (1993, p. 222) who remarked that Mark Twain was "fascinated" with Disraeli because the latter had been "balancing the romantic views with those of the hard headed politician". White had not mentioned statistics.

I intended to specify Twain's statement, but Disraeli was much too prolific for any direct confirmation (or refutation) to be possible. Indeed, he was not only a great statesman (Prime Minister of England for several years) but an eminent writer. Here are my indirect arguments that nevertheless testify that he was not responsible for coining that pithy saying.

1) The authors of collected sayings (Seldes 1966; Gaither and Cavazos-Gaither 1996) do not help.

2) Both Professor Stanley Weintraub, the author of a comprehensive study of Disraeli (Weintraub 1993) and Professor Melvin G. Wiebe, the Editor of several volumes of Disraeli (1987/1997), reported, in June 2001, that they have not found the celebrated maxim in Disraeli's writings.

3) Oskar Anderson (1887 – 1960) maintained that the maxim had been attributed to "verschiedenen englischen Staatsmännern, so z; B. Disraeli …" (Anderson 1962, p. 1). He continued, once more without substantiation:

"Dieser Ausspruch nur eine ab – bzw. fehlergeleitete Variante eines viel älteren englischen *geflügelten Wortes* darstellt, welches die höchste Steigerung der Lügner in den Juristen Erblickt …"

4) Fairley (1978, p. 799) provided a somewhat related story. Disraeli's reference, he wrote, "was to the testimony of expert witnesses, not statisticians (see Cook 1913, pp. 433 – 434)". Here, however, are Edward T. Cook's words:

"There is probably no department in human inquiry in which the art of cooking statistics is unknown, and there are sceptics who have substituted *statistics* for *expert witnesses* in the well-known saying about classes of false statements".

Cook did not mention Disraeli!

5) At the turn of the 19th century or later, neither Ladislaus von Bortkiewicz (1868 – 1931), nor Aleksandr Aleksandrovich Chuprov (1874 – 1926), both extremely well-read and likely conversant with everything that was going on in statistics, mentioned the maxim. Moreover, Chuprov (1903, p. 42) referred instead to Carlyle as stating that statistics was a carriage that took you wherever you wished. I have not found that particular passage, but perhaps Chuprov had freely translated the few lines that I quoted above.

As a fitting epilogue I adduce a statement attributed to Winston S. Churchill (1874 – 1965):

"When I call for statistics about the rate of infant mortality, what I want is proof that fewer babies died when I was prime minister than when anyone else was prime minister. That is political statistics".

This passage appeared in the February, 2002, issue of the Royal Statistical Society's newsletter without any exact references. Answering my inquiry, the Editor informed me that everybody believed that the attribution to Churchill was correct, but that no one knew the exact (likely oral) source.

Acknowledgement. I am grateful to Prof. Herbert A. David for helpful comments and for contacting Professors Weintraub and Wiebe, as well as to these scholars who agreed that their opinions be made public.

References

Anderson, Oskar: *Probleme der statistischen Methodenlehre in den Sozialwissenschaften* (1954). Physica: Würzburg, 1962.

Carlyle,Thomas: Chartism (1839), in author's *Historical and Political Essays.*Tauschnitz: Leipzig, 1916, pp. 115 – 214.

Chuprov, Aleksandr Aleksandrovich: Statistics and the Statistical Method. Their Vital Importance and Scientific Problems (1903), in author's *Voprosy Statistiki* (Issues in Statistics). Gosstatizdat: Moscow, 1960, pp. 6 – 42. In Russian.

Cook, Edward T.: *Life of Florence Nightingale* (1913). Reprint, Macmillan: New York, 1942.

Cournot, Antoine Augustin: *Exposition de la théorie des chances et des probabilités* (1843). Reprint, Vrin: Paris, 1984. **S, G,** 54.

Disraeli, Benjamin: *Letters*, vols. 1—5, edited by Melvin G. Wiebe. University of Toronto: Toronto, 1987 – 1997.

Fairley, William B.: Public Policy and Statistics. *Intern. Enc. of Statistics*, vol. 2. Edited by William H. Kruskal and Judith M.Tanur. Free Press and Collier Macmillan: New York – London, 1978, pp. 789 – 801.

Gaither, Carl C.; Cavazos-Gaither, Alina E.: *Statistically Speaking. A Dictionary of Quotations*. Institute of Physics: Bristol, 1996.

John, V. The term statistics. *J. Roy. Stat. Soc*., 46 (1883), pp. 656 – 679.

Quetelet, Adolphe; Heuschling, Xav. *Statistique internationale* (*population*). Commission centrale de statistique de Belgique: Bruxelles, 1865.

Saenger, K.: Das Preussische statistische Landesamt 1805 – 1934. *Allg. stat. Archiv* 24 (1935), pp. 445 – 460.
Seldes, George: *The Great Quotations*. Lyle Stuart: New York, 1966.
Twain, Mark, *Autobiography* (1906 – 1907). Harper & Brothers: New York, 1959.
Weintraub, Stanley: *Disraeli. A Biography*. Truman Talley-Dutton: New York, 1993.
White, Arthur W: Disraeli, Benjamin. *The Mark Twain Enc*., edited by J. R. Le Master and James D. Wilson. Garland: New York –London, 1993, pp. 222 – 223.
[In 2005 P. M. Lee stated, in the *Newsletter* of the Roy. Stat. Soc., that the real author was Lord L. H. Courtney and referred to Baines (1896). There, on p. 87, both Lord C. and the saying itself (treated as generally known) were indeed mentioned, but no definite source was cited.]
Baines J. A., Parliamentary registration in England etc. *J. Roy. Stat. Soc.*, 59, 1896, 38 – 118.

V

Oscar Sheynin

Sampling without replacement: history and applications

Intern. Z. f. Geschichte u. Ethik der Naturwissenschaften, Technik u. Medizin, Bd. 10, 2002, pp. 181 – 187

I dwell on the appearance, in the mid-19th century, of a formula describing random sampling without replacement under incomplete knowledge and discuss the statistical aspect of this formula and its applications. I also provide illustrations showing that the fairness of sampling without replacement was being questioned.

1. An Explanation and the initial problem

Sampling without replacement is essentially connected with opinion surveys as well as with the statistical inspection of mass manufactured commodities. Suppose that an electorate consists of two different parties which will support either of the two candidates for presidency. By collecting information on the preferences shown by a reasonably chosen random sample of the electorate, it is possible to forecast the outcome of the elections.

The study of the sample is tantamount to drawing a series of balls, all at once, from an um with an appropriate number of white and black balls contained in there in an unknown ratio. Nothing changes if we imagine that the balls are extracted without replacement, one by one.

It was Laplace who initiated scientific sample studies of the population; and Ostrogradsky, 1801 – 1862 (1848), studied statistical inspection of commodities. More precisely, he examined the sample inspection of foodstuffs supplied to the armed forces.

If not stated otherwise, I discuss random sampling without replacement from an urn containing a white balls and b black ones ($a + b = c$) and m and n balls, respectively ($m < a$, $n < b$, $m + n = s$), are supposed to be extracted.[1] When describing the reasoning of various authors, I do not distinguish between balls, tickets, etc., nor do I preserve their notation.

It is easy to verify that the probability of extracting the sample (m; n) is

$$P_1 = C_a^m C_b^n \div C_c^s \quad (1)$$

and the probability of drawing a white ball anew will then obviously be

$$P_2 = (a - m) \div (c - s). \quad (2)$$

When m and n remain unknown, P_2 will however remain as it was from the very beginning, see § 2:

$$P_1 = a/c \quad (3)$$

The expected value of the random variable n is

$$\mathrm{E}n = bs/c. \tag{4}$$

Below in this section, I recall the first appearance of a related problem; in § 2 I consider my main case of m and n remaining unknown and continue, in § 3, by discussing its statistical aspect. Finally, in§ 4 I provide some illustrations of this case.

In 1657, Christian Huygens, 1629 – 1695 [1888 – 1950, t. 14, Additional Problem No. 4] was the first to formulate a problem concerning sampling without replacement, and he himself solved it in a manuscript dated 1665 [Ibid., pp. 96 – 101; Sheynin 1977, p. 245]. It was required to determine the chances for $m = 3$, $s = 7$ when $a = 4$ and $b = 8$. His Additional Problem No. 2 was concerned with sampling with replacements, but, also in 1665, his correspondent Jonan van Waveren Hudde, 1628 – 1704 [Huygens 1888 – 1950, t. 5, p. 306] solved it believing that the extractions were done without replacement.

Here is this Problem No. 2: Once more, $a = 4$ and $b = 8$; calculate the chances of three gamblers who draw the balls in turn, one by one, until someone extracts a white ball and wins the game. Jakob Bemoulli, 1654 – 1705 [1713, T1. 1, pp. 63 – 66], and then De Moivre, 1667 – 1754 [1756, pp. 56 – 58] also solved this problem. They considered sampling both with and without replacement and obtained identical answers, in the latter instant the chances were as 77:53:35.[2]

2. Its Main specification

I am now concerned with the case in which m and n remain unknown. From among the contributions here discussed, Luchterhandt [1842] is forgotten, and Mondésir [1837] was only briefly mentioned by Jongmans and Seneta [1994][3].

2.1. Mondésir [1837]

Mondésir proved the following proposition. The probability of extracting q white or black balls (in succession and without replacement) out of the urn was the same, whether or not s balls were drawn out of it previously (with m and n remaining unknown).

Mondésir considered three cases: a) $s < a$, $s < b$; b) $a < s < b$; and c) $s > a$, $s > b$. In each of these, he calculated the probabilities of all the possible hypotheses on the composition of the sample, and of drawing the q balls under each hypothesis. He then multiplied these probabilities in pairs and summed up the products. In each case the obtained probability was indeed equal to

$$a(a-1)\,[a-(q-1)]/\{c(c-1)[c-(q-1)]\}.$$

Mondésir did not fail to note that Poisson had tacitly applied formula (3); he himself had not, strictly speaking, proved it.

2.2. Poisson

In a few cases, Poisson [1825—1826] tacitly issued from formula (3). He considered a game of chance in which two series of cards were extracted without replacement from the same set of six decks, and in one such case he treated the second series as though its cards were drawn from the initial untouched set.

Later Poisson [1837, pp. 231 – 234] returned to his assumption. Denote the probability of the sample (m; n) by $f(a; b; m; n)$ and suppose that it was obtained after a preliminary sample had provided g white and h black balls, $g + h = j$. Then

$$f(a; b; m; n) = \sum f(a - g; b - h; m; n) f(a; b; g; h)$$

where the summation extends over $g, h = 0, 1, 2, ..., g + h = j$. Poisson proved that the right side was independent of j so that it was possible to take $j = 0$. This meant that the preliminary sample, whose results remained unknown, did not influence the probability of obtaining the sample specified beforehand.

On the same p. 231, Poisson appropriately corrected his negligence made on p. 61. There he remarked that Mondésir, in a still unpublished memoir, had proved formula (3), see however my § 2.1. Poisson apparently thought it worthwhile to dwell on this point. He checked this formula against a simple numerical example; noted that the case $a = b$ obeyed it because then there was no reason to prefer either colour[4]; and referred to his own limit theorem for sampling without replacement. True, for large a, b, see formula (2), $(a - m)/(b - n) = a/b$.

Here, on p. 61, he did not cite either his earlier paper, nor his (perhaps not yet printed) pp. 231 – 234.

2.3. Luchterhandt [1842]

Having been unable to get hold of Mondésir's paper, Luchterhandt [1842] provided an independent (and direct) proof of formula (3). Issuing from formula (l), he multiplied its right side by $(a - m)/(c - s)$, see formula (2). The product thus obtained was the probability of the appropriate compound event, viz., of drawing a white ball after obtaining the sample (m; n). Luchterhandt then derived the sum of such products for $m = 0, 1, ..., s$, taking also account of the respective number of possible cases for each product. This was necessary because of the ignorance of m, and he thus arrived at formula (3).

During this last step he properly applied Poisson's calculations [1837, pp. 60 – 63]. For that matter, Poisson could have well derived there the same formula, and he possibly restrained himself because of Mondésir's manuscript.

2.4. Catalan

It was Eugene C. Catalan [Dale 1991; Jongmans and Seneta 1994] who directed his attention to the unexpected result discovered by Poisson and Mondésir; I shall only remark that at first he [1877] formulated an extremely general (and therefore, I would say, extra-mathematical) theorem; then [1884], without changing a single word, he promoted it to the rank of principle:

La probabilité d'un événement futur ne change pas 1orsque les causes dont il depend subsistent des modifications inconnues.

At about the same time, Bertrand [1888, p. xx], while dwelling on the regularity of mass random events, formulated a lucid, but not at all binding remark:

Le hasard, à tout jeu, corrige ses caprices. Les irrégularités même ont leur loi.

I would say that formulas (3) and (4) provide an especially remarkable quantitative illustration of Bertrand's idea (which, of course, goes back to the law of large numbers). Nowadays, however, at least the latter formula appears in the literature without any comment [Brownlee 1965, § 3.5][5].

3. The statistical point of view

Suppose that the extractions from an um provided the series white, black, black, … It is required to state whether the drawings were made with replacement or otherwise. Even a more general problem became topical after Wilhelm Lexis, 1837—1914, at the end of the 19th century, began his studies of the stability of statistical series [Bauer 1955; Sheynin 2011, § 14]. In my present context, it is sufficient to say that he attempted to distinguish whether a statistical series was, or was not composed of outcomes of independent Bernoulli trials.

One of the leading statisticians who developed (and finally largely refuted the Lexian criterion, but certainly not his influence on the development of statistics), was Tschuprow, or Chuprov, 1874 – 1926. Here is a passage from his letter of 1921 to his student Chetverikov, 1885 – 1973:

Not knowing the prior data it is impossible to distinguish a series of numbers obtained when extracting the tickets without replacement from a series obtained according to the usual way of replacing ... the ticket. It sounds like a paradox, but it is so! [Sheynin 2011, p. 145].

In a few years, he publicly repeated his statement [Tschuprow 1923, pp. 666 – 667; 1924, p. 490] concluding,‘ m the latter case, that *So paradox dies auf den ersten Blick erscheint, ist die Aufgabe unlösbar*.[6]

Seneta [1987] linked Chuprov's and his student Mordukh's [1923][7], studies with later important investigations of dependent random variables. According to my present viewpoint, Seneta should have also stressed that, almost from the beginning, Chuprov had indicated the relevance of his studies to sampling without replacement. Neither Chuprov nor Seneta mentioned formula (3).

4. Illustrations

I provide a few examples which show that the fairness of drawings without replacement was being questioned (Items 1, and, likely, 3); that formula (3) can be advantageously applied by a gambler (Item 4); and I also describe a very special case of such drawings (Item 2).

1. Redemption of the first-born [Jerusalem Talmud, Sanhedrin l[4]; French translation: Paris 1960; Sheynin 1998, pp. 192—193]. Moses intended to fill an urn with 22,000 white balls, and 273 black ones, and have 22,273 people to draw these without replacement. Those and only those who extracted a black ball would have had to pay five shekels each.

The Talmud implies that the proposed procedure was, however, (unduly) suspected.[8] Corroborating my earlier demonstration of the fairness of even the initial set-up, I remark now that, before the drawings were to begin, the expected losses were the same for each participant.

This episode was also mentioned, though only m a few words, in Numbers 3:46. Lots were also drawn, evidently again without replacement, in order to apportion the land among the tribes of Israel (Numbers 26:55 – 56 and 33:54), but details were not provided here either.

2. I left out of my earlier paper [1998] a curious statement which had appeared in the Rabbinic literature. Indeed, I thought that it was unworthy of discussion; now, I can at least link it with my present subject.

I bear in mind the opinion of Rabbi Shlomo ben Adret [Rabinovitch 1973, p. 40][9] about several pieces of meat all but one of them being kosher: *When he* (when somebody) *eats the* (first) *one, I say* This is not the forbidden one, *and similarly for each piece. When he eats the last piece, I say* The forbidden one was among those already consumed and this one is permitted *since by Biblical law, one in two is nullified*.[10]

Suppose that four out of five pieces of meat were kosher. Then a = 4, b = 1 and a/c = 4/5 so that after eating only three pieces of meat, the man would have consumed the forbidden piece with probability 0.6. The Rabbi likely had a clear idea about the chance of eating the forbidden piece at the very beginning, and a notion that the chances of the same event occurring later on were (at the time) impossible to calculate. And he hardly failed to notice that his conclusion was wrong the more so since the Talmud had set forth certain allowed ratios of forbidden/kosher food for various cases.

3. Draft lotteries. Several times during 1917 – 1970, young Americans were conscripted into the armed force by lot [Fienberg 1971]. At each locality, eligible men had to draw lots, and those who extracted numbers 1, 2, … (M – 1), M out of the first N ($N > M$) natural numbers were inducted.[11] Now, such lotteries are tantamount to drawings without replacement with $a = M$ and $c = N$.

At least in some of the cases an (unnecessary) double randomization was achieved by previously assigning a different number to each eligible man. It seems that the numbers in the urns were not sufficiently shuffled, but I leave this issue aside. Instead, I note that

The 1970 draft lottery has not helped to mitigate the doubts of many regarding the equity and fairness of random drawings [Fienberg].

Some of the doubts were possibly unconnected with the shuffling of the numbers.

4. Consider a game of blackjack with several decks of cards going on at a casino. When dealing out the cards for himself, the banker has to stop at 17 points; indeed, during the first game the probability of scoring more than 21 points then becomes 7/13 > 0.5, and for him it remains constant because he is unable to memorize the casted aside cards. However, a gambler endowed with an extra sharp memory will be able to apply advantageously the formula (2), and it seems that such instances did happen.

Acknowledgement. Prof. Dr. K. Dietz (Tübingen) offered useful methodological advice.

Notes

1 Already Emile Mondésir [1837, p. 10] remarked that a generalization onto balls of several colours in a related problem was possible by mentally combining all the colours but one. The case in which each of the *c* balls has its own colour leads to the celebrated Genoise lottery (beginning of the 17th century). It served as a point of departure for interesting stochastic studies, suffice it to mention Leonhard Euler.

2 Hudde, who had not provided any calculations, obtained a slightly different result, 232:159:104.

3 I cite the last-mentioned article once more in § 2.4. Its authors also discuss some related modern issues.

4 On p. 47 of the same source, Poisson even attempted to prove that the probability of drawing a white ball from an urn containing *n* balls, white and black, was 1/2, again because there was no reason to assume anything else (because of the same principle of insufficient reason, as it is called nowadays). If, in such a case, a probability might be thought to exist, it should indeed be equal to 1/2 so as to provide minimal information (in the sense of the theory of information) about the unknown contents of the urn. No other justification of such a result seems possible.

It is not amiss to quote Ellis, 1817 – 1859, [1850, p. 57]: *mere ignorance is no ground for any inference whatever, Ex nihilo nihil*.

5 Formula (4) is equivalent to $\mathrm{E}n/s = b/c$. This means that drawings without replacement, whether or not *m* and *n* remained unknown, indeed *corrige ses caprices*. Cf. § 3 below.

6 In 1925, Chuprov published a Russian version of his paper of 1924, and on p. 209 of its reprint [1960], he added a remark which I do not understand. Distinguishing between the two patterns, he stated, would be easy in the particular case of $a = b$.

7 Mordukh apparently left Russia without completing his education. He graduated from Uppsala University in 1921, and nothing is known about his further life. The only archival information which I have is that Jakob Mordukh was born in 1895 and graduated as Bachelor of Arts.

8 Those who would have begun the drawings were apparently thought to be luckier than the last ones; in our time, as I noted in the same article, such a misgiving was being directly voiced. In any case, Moses placed 273 additional white balls in the urn (which certainly did not meet the (imagined) issue and could have, with low probability, deprived him of some money).

9 Rabinovitch mentions the Rabbi seven times. On p. 5 he lists Rabbi Shlomo (ben Abraham) ben Adret (1235 – 1310) along with several other scholars as a Talmudic commentator, and his p. 40, from which I have just quoted, belonged to a section entitled *Talmudic Acceptance Rules*. To my mind, however, the Rabbi's reasoning is a specimen of a hardly known stochastic variety of sophisms.

10 These last words likely mean that probability 0.5 is *nullified*.

11 This is at least my understanding; for my purpose, the authorities could have just as well chosen the alternative. The author likely believed that the rules were generally known and did not sufficiently explain them.

References

Bernoulli, Jakob: *Wahrscheinlichkeitsrechnung* (1899), this being a translation of *Ars Conjectandi* (1713). Reprint: Harri Deutsch: Frankfurt/Main, 1999.

Bertrand, Joseph (**Louis Francois**): *Calcul des probabilités* (1888). Second, practically identical edition, 1907. Reprint, Chelsea: New York, 1970 and 1972.

Brownlee, K. A.: *Statistical Theory and Methodology in Science and Engineering*. Wiley: New York, 1965.

Catalan, Eugene C.: Un nouveau principe de probabilités. *Bull. de l'Acad. roy. des scien ces, des lettres et des beaux-arts de Belgique*, 2me sér., 46e année, t. 44, 1877, pp. 463 – 468.

---:Application d'un nouveau principe de probabilités. Ibidem, 3me sér., 53e année, t. 3, 1884, pp. 72 – 74.

Chuprov, Aleksandr Aleksandrovich: The main problems of the stochastic theory of statistics (1925). Reprinted m author' s *Voprosy Statistiki* (Issues in Statistics). Editors B. I. Karpenko and N. S. Chetverikov. Gosstatizdat: Moscow, 1960, pp. 162 – 221. In Russian.

Dale, Andrew I.: *History of lnverse Probability*. Springer: New York, 1991 and 1999.
De Moivre, Abraham: *Doctrine of Chances* (1718, 1738, 1756). Reprint of the last edition, Chelsea: New York, 1967.
Ellis, Robert Leslie: Remarks on an alleged proof of the method of least squares (1850). In author's *Mathematical and Other Writings.* Editor W. Walton. Deighton & Bell: Cambridge, 1863, pp. 53 – 61.
Fienberg, S. E.: Randomization and social affairs: the 1970 draft lottery. *Science* 171 (1971), pp. 255 – 261.
Huygens, Christian: *Oeuvres Complétes*, tt. 1 – 22. Nijhoff: La Haye, 1888 – 1950. Editor of t. 5, D. Bierens De Haan, of t. 14, D. J. Korteweg (named only in t. 22).
Jongmans F. and Seneta, E.: A *Probabilistic new principle* of the 19th Century. *Arch. for Hist. of Ex. Sci*. 47 (1994), pp. 93 – 102.
Luchterhandt, A. R. Über einen Lehrsatz aus der Wahrscheinlichkeitsrechnung. *Archiv der Math. u. Phys.*, Bd. 2, 1842, pp. 65 – 67.
Mondésir, Emile. Solution d'une question qui se présente dans le calcul des probabilités. *J. de math. pures et appl.*, t. 2, 1837, pp. 3 – 10.
Mordukh, Jakob. On associated trials corresponding to the condition of stochastic commutativity. *Tr. Russk. Uchenykh Zagranitsei* (Trans. of Russ. Scientists Abroad), vol. 2. Slovo: Berlin 1923, pp. 102—125. English translation: **S, G,** 6.
Ostrogradsky, Mikhail Vasilievich: Sur une question des probabilités. *Bull*. (*Izvestia*) Fiziko-matematicheskikh nauk *Imp. Akademii Nauk*, t. 6, No. 21—22, 1848, pp. 321 – 346.
Poisson, Siméon Denis:Sur l'avantage du banquier au jeu de trente-et-quarante. *Annales de math. pures et appl.*, t. 16, 1825 – 1826, pp. 173—208.
---: *Recherches sur la probabilité de jugements.* Bachelier: Paris, 1837, 2003. **S, G,** 53.
Rabinovitch, Nachum L.: *Probability and Statistical Inference in Ancient and Medieval Jewish Literature*. Toronto University: Toronto, 1973.
Seneta, Eugene: Chuprov on finite exchangeability, expectation of ratios and measures of association. *Hist. Math*. 14 (1987), pp. 243 – 257.
Sheynin, Oscar: Early history of the theory of probability. *Arch. for Hist. of Ex. Sci*. 17 (1977), pp. 201 – 259.
--- Chuprov. Vandenhoeck & Ruprecht: Göttingen, 1996. V&Runipress, Göttingen, 2011.
--- Stochastic thinking m the Bible and the Talmud. *Annals of Sci.* 55 (1998), pp. 185 – 198.
Tschuprow (**Chuprov**), **Aleksandr Aleksandrovitsch**: On the athematical expectation of the moments of frequency distributions in the case of correlated observations. *Metron* 2 (1923), pp. 461 – 493 and 646 – 683.
--- : Ziele und Wege der stochastischen Grundlagen der statistischen Theorie. *Nordisk Statistisk Tidskrift*, 3 (1924), pp. 433 – 493.

VI

Oscar Sheynin

Kepler as a statistician

Silesian Stat. Rev., No. 15/21, 2017, pp. 227 – 232

Summary: Drawing on my previous publications (see Bibliography), I describe Kepler's work on the mathematical treatment of observations and astrology. In particular, I investigate how he rejected the Ptolemaic system of the world and note that his astrology had the features of qualitative correlation.
Key words: reformation of astronomy, astrology, qualitative correlation, minimax method, Monte Carlo method

1. Mathematical treatment of observations

This is my main subject. Modern astronomers are not anymore interested in it, and even historians of astronomy are ignorant of it. William H. Donahue, who translated Kepler's great work (1609) into English and thus made an excellent contribution to the history of astronomy, did not comment on Kepler's treatment of observations. This, however, is just what I will do in this section; and I quote Kepler (1609) by only mentioning the page numbers of its translation.

1.1. The arithmetic mean. Kepler (p. 200) collected four astronomical observations of the right ascension of Mars and, without any explanation, remarked: *The mean, treating the observations impartially* (*medium ex aequo et bono*), is …

Actually (Eisenhart 1976, p. 356) Kepler had chosen a weighted arithmetic mean (and had to assign subjectively the weights). But the main point here is that his Latin expression had occurred in Cicero (*Pro A. Caecina oratio*, § 65) whom Kepler likely read. It connoted *rather than according to the letter of the law*. (I have found this connotation in a Russian textbook of the Latin language for student lawyers.)

So now we know that at the very beginning of the 17th century or somewhat earlier the arithmetic mean became *the letter of the law*.

1.2. The Monte Carlo method. When adjusting observations, Kepler sometimes corrupted them by small arbitrary magnitudes. Thus (p. 334), *One might hold suspect such license* since then we will be able *to change whatever we don't like in the observations*. He reasonably added that the changes *ought to remain within the limits of observational precision*. And he certainly had to take into account the properties of *usual* random errors: an approximate equality of those changes of both signs and a larger number of changes smaller in absolute value.

Actually, Kepler applied elements of the Monte Carlo method.

1.3. Reformation of astronomy. And now the main point, Kepler's rejection of the Ptolemaic system of the world (p. 286):

Since the divine benevolence has vouchsafed us Tycho Brahe, a most diligent observer, from whose observations the 8' error in this Ptolemaic computations is shown, it is fitting that we … acknowledge and honour this benefit of God … They could not be ignored, these *eight minutes alone will have led the way to the reformation of all the astronomy*.

This passage has been quoted a thousand times, but no one thought of investigating it. Two questions have to be answered: why Kepler was sure that the error of Tycho's observations was less than 8′; and how did he arrive at that estimate?

Kepler gave an indirect answer to the first question by stating that the error of his own observations was of the order of two or three minutes (pp. 215, 621 and 611). But the main question is the second one, and I ought to go into details about the adjustment of observations.

Given, a system of n equations with k unknowns, $n > k$

$$a_i x + b_i y + \ldots + w_i = 0, \; i = 1, 2, \ldots, n \qquad (1)$$

Here, the coefficients are provided by the appropriate theory and the free terms are the observations or their functions. The observations, and therefore the equations are mutually physically independent (linear independence was not yet known) and systems (1) had no solutions. Astronomers (and geodesists) had to be satisfied by any set of numbers $\hat{x}, \hat{y}, \ldots$ approximately satisfying (1), i. e. such that the residual free terms, call them v_i, were small enough and more or less satisfying the properties of *usual* random errors, cf. § 1.2. In other words, an additional restriction had to be imposed on those residuals. One of those restrictions was the condition of least squares

$$v_1^2 + v_2^2 + \ldots + v_n^2 = \min.$$

Petrov (1954) is apparently still the best investigation of the optimal properties of the method of least squares.

Other methods had been earlier introduced, and, among them, the minimax method or rather its elements since only Laplace offered an algorithm for applying it properly. This method meant that the v_i, maximal in absolute value, is minimal among all the possible "solutions" of system (1); in the period before Laplace, minimal only among some reasonable "solutions".

The method of minimax is not optimal in any sense, but it answers an important question: if the derived maximal v_i is unacceptably large, then either the theory justifying the system (1) was wrong, or the observations (the w_i) were too bad.

And I believe that Kepler had indeed applied elements of the minimax method to a system corresponding to the Ptolemaic picture of the world and decided as stated above. This, however, was not enough! I also believe that Kepler had then repeated such calculations for the Copernican system and likely arrived at a maximal v_i of the order of 3′, see above the estimation of the precision of his own observations. He had not regrettably said anything about that likely second calculation, but in principle it can be repeated now.

Interestingly, the minimax method is tantamount to generalized least squares:

$$\lim (v_1^{2k} + v_2^{2k} + \ldots + v_n^{k}) = \min, \; k \to \infty.$$

An important circumstance here is that in astronomy systems of equations are not linear and not even algebraic, but they can be linearized. Suppose that such a system involves x^2 (a similar conclusion will apply, for example, to $\sin x$). It is then possible to solve any subsystem with an equal number of unknowns and equations. The value x_0 will be calculated and

$$x = x_0 + \Delta x$$

with a comparatively small $|\Delta x|$. Then

$$ax^2 = a(x_0 + \Delta x)^2 \approx ax_0^2 +2\, x_0\Delta x$$

and the system will be linear in Δx.

Kepler had to linearize his systems, otherwise he would have been obliged to obtain reasonable solutions by solving non-algebraic systems many times over.

1.4. Systematic influences. Kepler (1634/1967, p. 142) formulated recommendations for observers of solar eclipses. Actually, he insisted that systematic influences ought to be excluded (as far as possible).

2. Other topics

2.1. Randomness. Kepler (1606/2006, p. 163) rejected it: *What is chance*? *An idol* … Nevertheless, he had to find room for randomness, see Sheynin (2014, § 2). There also, in § 3, I have followed the subsequent views of Kant and Laplace likely borrowed from Kepler. See also § 3 below.

2.2. An embryo of the law of large numbers. Kepler (1627) stated that the total weight of many coins (more precisely, the mean weight of a coin selected from them) is constant.

3. Astrology

From a modern point of view, astrology is a pseudoscience. There were, however, astrologers, scholars of the highest calibre included, who strove to discover connections between heaven and earth. They sincerely believed in the existence of such connections the more so since heaven does influence earth; thus, ocean tides are occasioned by the sun and the moon.

Astrologers singled out the *aspects*, i. e., remarkable mutual positions of the sun, the moon and the planets visible by the naked eye. Without any criteria they somehow separated randomness and regularity, a problem which is still remaining a fundamental challenge for modern mathematics. Kepler (1601/1979, p. 97) *added* three aspects to those recognized by ancient astrologers, so he also participated in the solution of that perennial problem.

Ptolemy (1956, I 2 and I 3) believed that the influence of heaven was a tendency rather than a fatal drive, and I understand his astrology as qualitative correlation. Indeed, ancient science was qualitative, witness Hippocrates (1952, vol. 10, No. 44):

Persons who are naturally very fat, are apt (!) *to die earlier than those who are slender*.

Kepler contributed to this direction of astrology. He stated that the influence of heaven at the moment of his birth was only a tendency, and, what is more interesting, he (1610/1941, p. 217; 1619/1939, pp. 256, 263) introduced intermediate causes (climate, geographical location, political structure of the land etc.) which were able to corrupt the influence of heaven. This was another step towards qualitative correlation since correlation analysis involves the isolation of the essential factors and a decision about the other influences (to disregard them, or to take them somehow into consideration). On the other hand, such intermediate causes pave the way for deception by quacks.

Kepler (1619/1997, book 4, chapter 6) considered himself the founder of a scientific astrology based on tendencies, but, even disregarding ancient scholars, Tycho Brahe had forestalled him (Hellman 1970, p. 410).

Now, Kepler was mostly interested in studying the general destiny of nations according to the tendency of the prevailing aspects. As a *Landschaftsmathematiker*, he also had to compile yearly astrological almanacs, see M. Casper, p. 22* of his commentary on Kepler's *Welt-Harmonik* (1619/1939). He was dissatisfied by them since, as he (1610/1941, p. 253) stated, ordinary men were only interested in impossible precise predictions about their lives, and he decided to quit those compilations (but had to continue owing to financial difficulties).

Bibliography

Eisenhart C. (1976), [Discussion of invited papers on history of statistics]. *Bull. Intern. Stat. Inst.*, vol. 46, pp. 355 – 357.

Hellman C. D. (1970), Brahe. *Dict. Scient. Biogr.*, vol. 2, pp. 401 – 416.

Hippocrates (1952), Aphorisms, No. 44. *Great Books of Western World*, vol. 10, pp. 131 – 144.

Kepler J. (1601, in Latin), On the most certain foundation of astrology. *Proc. Amer. Phil. Soc.*, vol. 123, 1979, pp. 85 – 116.

-- (1606, in Latin), *Über den Neuen Stern im Fuß des Schlangenträger*. Würzburg, 2006.

--- (1609, in Latin), *New Astronomy*. Cambridge, 1992. Translated by W. H. Donahue.

--- (1610, in Latin), *Tertius interveniens*. *Ges. Werke,* Bd. 4. München, 1941, pp. 145 – 258.

--- (1619, in Latin), *Weltharmonik*. München – Berlin, 1939. *Harmony of the World*. Philadelphia, 1997.

--- (1627), An den Senat von Ulm. Brief 30 Juli 1627. In Caspar M, von Dyck W. (1930), *Kepler in seinen Briefen*. München – Berlin, Bde 1 – 2. Bd. 2, p. 248.

--- (1634, in Latin), *Somnium*. München – Berlin, 1967. In English.

Petrov V. V. (1954, in Russian), On the method of least squares and its extreme properties. *Uspekhi Matematich. Nauk*, vol. 9, No. 1, pp. 41 – 62.

Ptolemy (1956, in Greek), *Tetrabiblos*. London. In Greek and English.

Sheynin O. (1973), Mathematical treatment of astronomical observations (a historical essay). *Arch. Hist. Ex. Sci.*, vol. 11, pp. 97 – 126.

--- (1974), On the prehistory of the theory of probability. Ibidem, vol. 12, pp. 97 – 141.

-- (1978), Kepler, Johannes. In Kruskal W. H., Tanur Judith M., Editors, *Intern. Enc. of Statistics*, vols 1 – 2. New York – London, pp. 487 – 488.

--- (1993), The treatment of observations in early astronomy. *Arch. Hist. Ex. Sci.*, vol. 46, pp. 153 – 192.

--- (2014), Randomness and determinism. Why are the planetary orbits
elliptical? *Slaski przeglad statystytczny, Silesian Stat. Rev.*, No. 12 (18), pp. 57 – 74.

VII

Oscar Sheynin

The inverse law of large numbers

Math. Scientist, vol. 35, 2010, pp. 132 – 133

Denote the probability of a studied event by p, and the statistical probability of its occurring in ν independent trials by $\hat{p}$. Jakob Bernoulli (1713, pt. 4) proved, although without introducing probabilities, that

$$\lim P(|\hat{p} - p| < \varepsilon) = 1, \nu \to \infty, \varepsilon > 0. \quad (1)$$

He also studied the rapidity of the limiting process with less success, largely because Stirling's theorem was still unknown to him. Several authors improved his estimate; one of them, Pearson (1925, p. 202), considered Bernoulli's result too crude and rejected it altogether, obviously disregarding existence theorems.

Bernoulli (chapter 4 of pt. 4), however, sought to prove that $\hat{p}$ can replace an *unknown p*, so that (1) was not really what he needed, but, he nevertheless claimed to have proved that induction was not worse than deduction. Even more: in his examples he described cases in which no probability was known to exist including that of being taken ill with a certain disease. Graunt (1662) made use of such statistical probabilities for compiling his life table, certainly faulty and not deserving attention from the present viewpoint but extremely important at the time.

Mathematicians have been dealing with entities not existing in nature (e. g., imaginary numbers) and natural scientists (and statisticians) deal with entities of the same kind including true values of estimated measures of precision. Fourier (1826/1890, p. 533) defined such objects as the limit of the appropriate arithmetic means [viii].

De Moivre (1733) proved the first version of the central limit theorem and, in an extended version of his memoir, he (1756, p. 251) stated that "conversely", if $\hat{p}$ tends to some magnitude, that "ratio" will express p. It was Bayes [xi] who investigated that converse case, the inverse law of large numbers, as I am now calling it, and for this reason I think that he can be credited with completing the first version of the theory of probability and that Mises, when introducing his frequentist definition of probability, could have been (but apparently was not) inspired by Bayes.

References

Bernoulli, J. (1713). *Ars Conjectandi*. English translation of pt. 4: Bernoulli (2005).
--- (2005). *On the Law of Large Numbers.* Berlin. **S, G,** 8.
De Moivre, A. (1733), A method of approximating the sum of the terms of the binomial $(a + b)^n$ expanded into a series from whence are deduced some practical rules to estimate the degree of assent which is to be given to experiments. Translated by author from its Latin text privately circulated about 12 years earlier. Incorporated

into author’s *Doctrine of Chances.* London, 1738 and 1756; see pp. 243 – 254 of latter edition.
Fourier, J. B. J. (1826), Sur les résultats moyens déduits d’un grand nombre d’observations. *Oeuvres*, t. 2. Paris, 1890, pp. 525 – 545. **S, G,** 88.
Graunt, J. (1662), *Natural and Political Observations Made upon the Bills of Mortality*. London. Reprint: Baltimore, 1939.

VIII

Oscar Sheynin

The true value of a measured constant and the theory of errors

Hist scientiarum 17, 2007, pp. 38 – 48

Abstract

The theory of errors is a discipline indispensable to experimental science at large, and *true value* of a measured constant is one of its main notions. I reject a modern statement which claims that the true value "syndrome" is left behind. I dwell on the history of that notion, – on its heuristic use, informal connection with the arithmetic mean of the pertinent observations, and on its formula (Laplace, Fourier), forgotten perhaps up to the mid-20th century. Mises, although not really interested in the theory of errors, effectively connected *true value* with his *frequentist* definition of probability as the limit of the corresponding statistical frequency. Mathematical statistics largely but not completely moved from the true value to the estimation of parameters of functions. Condorcet hesitatingly introduced an intermediate theory of means which studied the determination of both true values and abstract mean values but which became divided between statistics and the theory of errors.

Key words: Experimental science; Frequentist theory of probability; Theory of errors; Theory of means; True value of constant

1. Introduction

From the most ancient times astronomers have been measuring the coordinates of the fixed stars, i.e., of presumably constant magnitudes. Actually, however, this supposition, as will be seen in the sequence, is not really true.

The concept of *true value of a measured constant* had always been inseparably linked with the measurements themselves; only mathematical statistics (almost) changed this situation. Thus, Al-Biruni (1967, p. 83): "Now all the testimonies that we have adduced point out collectively that the [obliquity of the ecliptic] is …" And here is Cotes (1722/1768, p. 22), also without using the term *true value*: "The place of some object defined by observation[s] …"

My second concept is *theory of errors* which I define as the statistical method (statistics) applied to the treatment of observations in experimental science. I only deal with its *stochastic branch*; its *determinate* branch might be related to experimental design.

2. The Arithmetic mean and the true value

The first to connect directly these two notions was possibly Picard (1693, pp. 330, 335, 343) who called the arithmetic mean the true (*véritable*) value (of the angle measured in triangulation). The next, and much more outspoken author was Lambert. First, he (1760, § 286) stated:

Da nun Fehler um so häufiger auftreten, je kleiner sie sind, folgt daraus, das in einem beliebigen gegebenen Fall nach wiederholten Versuchen die häufiger auftretenden Größen dem Mittelwert oder auch dem wahren Wert näher liegen.

[*Since errors happen the oftener, the smaller they are, it follows that in any given case of repeated experiments the more frequently occurring quantities are situated nearer to the mean value, or, also, to the true value*.]

And in § 290 he added that the error of the arithmetic mean was much smaller than that of a single observation and that consequently the mean was nearer to the true value. Then, Lambert (1765, § 3) argued that, if, in modern terms, the density curve of the observational errors was even,

Das Mittel aus mehrern Versuchen dem wahren desto näher kommen müsse, je mehr der Versuch ist wiederholt worden. Denn unter allen Fällen, die man sich dabey gedenken kann, ist derjenige am möglichsten, wobey gleich große Abweichungen auf beyden Seiten gleich ofte vorkommen.

[*The mean of a large number of experiments ought to move the nearer to the truth, the more is the experiment repeated. Because, among all the cases which might be imagined, the most possible is that in which equally large deviations to both sides occur equally often.*]

He, as well as some later authors, see below, tacitly (but almost directly in his previous case) assumed that the density was unimodal and not *bad* (cf. for example the Cauchy distribution under which a single observation is not worse than the mean) and he certainly had not proved his statements. Thus, only Thomas Simpson, in 1756, proved the essence of Lambert's § 290, and, for that matter, only for two distributions.

That the mean tends to the appropriate theoretical parameter is now called, in statistics, the limit property of consistency which holds for linear estimators in general. In my context, however, this remark is hardly of consequence.

My next author here is Laplace. He (1795/1912, p. 161) stated that with an unrestricted increase in the number of observations their mean converged to a certain number, so that

Si l'on multiplie indéfiniment les observations ou les expériences, leur résultat moyen converge vers un terme fixe, de manière qu'en prenant de part et d'autre de ce terme un intervalle aussi petit que l'on voudra, la probabilité que le résultat moyen tombera dans cet intervalle finira par ne différer de la certitude que d'une quantité *moindre que toute grandeur assignable. Ce terme est la vérité même si les erreurs positives et négatives sont également faciles* …

[*If we multiply observations or experiments indefinitely, their mean result will tend to a fixed term, so that, taking on both its sides an interval as small as you wish, the probability that the mean result finds itself there will finally differ from certitude by a quantity less than any assigned magnitude. This term is the truth itself provided that positive and negative errors are equally likely*]

He repeated this statement word for word (1810a/1898, p. 303), and he also repeated it elsewhere, either a bit later, or a bit earlier (1810b/1979, p. 110/272), writing *se confond avec le vérité* [*merges with the truth*] instead of *est la vérité même* [*is the truth itself*].

And in his *Essai philosophique* (1814/1886, p. LVI) which originated from the *Leçons* of 1795, we find:

Plus les observations sont nombreuses et moins elles s'écartent entre elles, plus leurs résultats approchent de la vérité.

[*The more observations there are and the less they deviate from one another, the more their results approach the truth.* From the translation of 1995, p. 43.]

He added that the optimal mean results were determined by probability theory. Now, it is generally known that he strongly advocated (and furthered) the method of least squares; hence, when discussing the case of one unknown, as above, he certainly meant the arithmetic mean. In the fifth edition of the *Essai* (1825) Laplace also left a similar pronouncement concerning the general case (p. 44 of the English translation (1995) of that edition).

I hasten to add that Gauss had not left anything comparable. When providing his first justification of the method of least squares, he (1809/1887, §177) issued from the hypothesis that the arithmetic mean was the most probable value of the constant sought, or very close to it.

Understandably, Poisson (1811, p. 136; 1824, p. 297; 1829, pp. 12 and 19) followed his predecessors in that he used the term *vraie valeur* and indirectly stated that this value was the mean of infinitely many observations.

3. The definition

Fourier (1826/1890, pp. 533 – 534) provided the still lacking formal definition:

Supposons donc que l'on ait ajouté ensemble un grand nombre m de valeurs observées, et que l'on ait divisé la somme par le nombre m, ce qui donne la quantité A pour la valeur moyenne; nous avons déjà remarqué que l'on trouverait presque exactement cette même valeur A, en employant un très grand nombre d'autres observations. En général, si l'on excepte des cas particuliers et abstraits que nous n'avons point à considérer, la valeur moyenne ainsi déduite d'un *nombre immense d'observations ne change point; elle a une grandeur déterminée H, et l'on peut dire que le résultat moyen d'un nombre infini d'observations est une quantité fixe, où il n'entre plus rien de contingent, et qui a un rapport certain avec la nature des faits observés. C'est cette quantité fixe H que nous avons en vue comme le véritable objet de la recherche.*

[*Suppose therefore that a large number of observations are added together, and their sum is divided by* [their] *number, m, which provides the quantity A for the mean value. We have already remarked that almost exactly the same value A will be found when taking a very large number of other observations. In general, excepting particular and abstract cases which we will not consider at all, the mean value thus derived from an immense number of observations does not change at all. It has a certain magnitude H, and it is possible to say that the mean result of an infinite number of observations is a fixed quantity which never contains anything accidental anymore, and which is in a certain relation to the nature of the observed events.*

It is this fixed magnitude H that we have in mind as the veritable object of research.]

I doubt that his formula was widely noticed and in any case I was unable to find even a single reference to it; perhaps it was thought to be hardly needed. Nevertheless, a number of later authors repeated the same definition independently one from another, and likely, from him, see below. First, however, I turn to Markov (1924, p. 323) who cautiously, as was his wont, began the chapter on the method of least squares of his treatise by remarking that

It is necessary in the first place to presume the existence of the numbers whose approximate values are provided by observations.

A similar statement concerning an unknown probability is on p. 352; his first pronouncement was inserted in the edition of 1908 (perhaps even in the first edition of 1900), the second one appeared in the edition of 1913. Several remarks are in order.

1) Before and after Markov many scholars either expressly mentioned, or indirectly referred to the true value without bothering to define it (Gauss, in all of his writings pertaining to the treatment of observations; Markov himself 1899/1951, p. 250; Poincaré 1912, p. 176; Kolmogorov 1946, title of § 7).

2) Probability (Markov, p. 352) is not an entity of the real world, at least not in the usual sense. This generalization of the concept under my study is an important point for a natural scientist, although not for Markov the mathematician. Incidentally, already Gauss (1816/1887, §§ 3 and 4), a mathematician *and natural scientist*, repeatedly considered the true value of a measure of precision of observations. See also Fisher's relevant statement in my § 4.

3) I also note Markov's reluctance to step out of the field of mathematics: he had not mentioned true values at all which was

hardly accidental. Recall that he never provided any applications of his *chains* to natural sciences.

Fourier's definition heuristically resembles Mises' celebrated formula for probability; strangely enough, no one saw fit to mention this fact except Mises himself. Here is what he (1919/1964, pp. 40 and 46) actually stated, largely repeating Fourier:

Der "wahre" Wert der Beobachtung (d. i. derjenige, der sich als Durchschnitt bei einer ins Unendliche fortgesetzten Beobachtungsreihe ergeben müsste) ... Der "wahre" Mittelwert ist nicht anderes als die Größe, die nach der Definition des Wahrscheinlichkeitsbegriffes als arithmetisches Mittel einer ins Unendliche fortgesetzten Ziehungsserie sich ergeben müsste.

[*The "real" value of the observation (that is, such that ought to occur as the mean value when the series of observations continues to infinity). ... The "real" mean value is nothing but the magnitude that ought to occur by the definition of the concept of probability as the arithmetic mean when the series of drawings continues to infinity.*]

In 1919, the corresponding page numbers were 80 and 87, and it was in that contribution that Mises first introduced his frequentist theory. In other words, the *concept of probability* [*Wahrscheinlichkeitsbegriff*] could have only been his frequentist definition of probability. But to explain the *drawings*. Suppose that an urn contains white and black balls and that *m* white balls and *n* black ones are extracted and returned back one by one. Then, as Mises stated, the ratio *m/n* approached the unknown ratio of the balls contained in the urn. This was his illustration of the connection of the true value and frequentist probability but he had not directly offered it as a formula.

My next author is also interesting because he (Eisenhart 1963/1969, pp. 30 – 31) deals with metrology, an important scientific discipline which statisticians hardly ever discuss when they (also on rare occasions) recall the theory of errors:

The "true value" of the magnitude of a quantity ... is the limiting mean of a conceptual <u>exemplar process</u> ... The mass of a mass standard is ... specified ... to be the mass of the metallic substance of the standard plus the mass of the average volume of air adsorbed upon its surface under standard conditions. I hope that the traditional term "true value" will be discarded in measurement theory and practice, and replaced by some more appropriate term such as "target value"

And so, first, Eisenhart largely repeated Fourier. Second, here, as had always implicitly been the case before, he clearly stated that the residual systematic error was inevitably included in the true value. Third and last, Eisenhart's *hope* had not materialized, see below, but he was quite right when stating, in addition, that it was impossible to obtain any true value.

To conclude, I mention that Whittaker & Robinson (1924/1958, p. 215n) largely repeated the Fourier definition:

True mean is expectation although different values of a random variable reflect its intrinsic property of change whereas different values of observations of a measured constant are in the first place the result of our helplessness.

4. Mathematical statistics and the theory of errors

Purportedly, mathematical statistics had done away with true values and introduced instead parameters of densities (or of distribution functions). Fisher (1922, pp. 309 – 310) was mainly responsible for this change; indeed, he introduced there the notions of consistency, efficiency and sufficiency of statistical estimators without any reference to the theory of errors or to true values. But then, on p. 311 we read that a

Purely verbal confusion has hindered the distinct formulation of statistical problems; for it is customary [for the Biometric school] *to apply the same name, <u>mean, standard deviation, correlation coefficient, etc</u>., both to the true value which we should like to know, but can only estimate, and to the particular value at which we happen to arrive by our methods of estimation.*

So the true value was still alive even in mathematical statistics. A few other examples. The *Dictionary* (Aleksandrov 1962) cites *true* correlation; mean; and value. Bolshev (1964, p. 566) dwells on the "true value of a parameter". His was a commentary on Bernstein (1941/1964) who mentioned a "true probability" of an inequality (in § 5, p. 390 in 1964). Then, Smirnov & Dunin-Barkowski (1959/1973, pp. 16 and 17) had chosen to say *true value*.

But what about our contemporaries? Here is an opinion which I oppose (Chatterjee 2003, p. 264): the methods of the theory of errors "were rarely applied outside these narrow fields" [of astronomy and geodesy] and "the true value syndrome" "was ultimately left behind".

First, I object to the *narrow fields* and note the author's failure to recognize metrology. And how about measurements in geophysics (of magnetism, or of the acceleration of gravity), or in physics (of the velocity of light in vacuum, or of the mass of electron), etc.?

Then, *syndrome* is usually connected with some abnormal condition. Second, since Chatterjee (pp. 248 – 249) still believes in the existence of the mysterious "well-known" Gauss-Markov theorem, I doubt that he is proficient in the history of statistics (and especially of the treatment of observations).

I am also dissatisfied with Chatterjee's statement (p. 273) that Quetelet was "mentally bound by … the true-value syndrome" and that, implicitly, for Quetelet variations were "of secondary importance". Even excluding meteorology, his important field of research beyond social statistics, Quetelet (Sheynin 1986) studied the change of the probability of conviction for differing groups of defendants (my § 4.4 there), held that the *tables de criminalité pour les différents ages* [*tables of criminality for different ages*] merited full attention (p. 304n 45) and declared that the normal law was *une les plus générale de la nature animée* [*one of the most general of the animate nature*] (p. 313), – especially in anthropometry. More about Quetelet in § 5 where I also dwell on the study of mean values (conditions).

Third, I cite Hald (1998) who described the *History of mathematical statistics from 1750* (when it did not yet exist) *to 1930* on the present-day level. Thus, when discussing the work of Gauss he (p. 353) introduced without explanation the recent notation for a function with an unspecified argument: $f(\cdot)$. He mentions the true value many times, for example in Chapters 5 and 6, and here is how he begins this latter chapter (p. 91): " … we have discussed … the estimation of the true value, the location parameter, in the … model".

I conclude that the term itself, and the notion of true value are still applied to a certain extent even in mathematical statistics.

I defined the theory of errors in § 1. According to its "official" mathematical definition (Bolshev 1984/1989), it is a branch of mathematical statistics beyond whose confines is the *processing of observations* (Bolshev (1982/1991) which studies systematic errors. I do not agree. First, the theory of errors is just unable to divorce itself from such studies. Second, systematic errors are a feature of the structure of statistical data, and their absence or presence should therefore be verified by *exploratory data analysis*, an important chapter of theoretical, even if not mathematical statistics (Sheynin 1999/2006). Third and last, Bolshev's description of the processing of observations is somewhat indefinite and does not mention data analysis at all.

5. The intermediate stage

It is usual to credit Galton with breaking away from true value (and the theory of errors in general). In 1908 he (Eisenhart 1978, p. 382) wrote:

The primary objects of the Gaussian Law of Error were exactly opposed, in one sense, to those to which I applied them. They were to get rid of, or to provide a just allowance for errors. But these errors or deviations were the very things I wanted to preserve and to know about.

Deviations together with their respective probabilities, i. e., their densities.

But the intermediate stage between the theory of errors and mathematical statistics began much earlier with Condorcet (1805/1986, p. 604) who introduced

Théorie des valeurs moyennes … un préliminaire de la mathématique sociale … dans toutes les sciences physico-mathématiques, il est également utile d'avoir des valeurs moyennes des observations ou du résultat d'expériences.

[*The theory of mean values … a preliminary to social mathematics … in every physical and mathematical science is equally useful to have mean values of observations or of the results of experiments.*]

On the same page he definitely separated this proposed theory from the "théorie du calcul des probabilités". Nevertheless, he had not elaborated, had not offered a formula of the theory of means. On pp. 555 – 559 Condorcet reasoned on the connection between the arithmetic mean (only in the case of a finite number of observations) and the *vraie valeur inconnue* [*true unknown value*], noted, on p. 555, that *On peut distinguer deux espèces de valeurs moyennes* [*It is possible to distinguish two kinds of mean values*], but still had not explained himself clearly enough, cf. Quetelet's statement below.

Anyway, the emerged theory of means (hardly separated from probability!) was more general than the theory of errors in that it also dealt with mean states; for example, with the mean stature of draftees (Quetelet, his celebrated study). It was Lambert (Sheynin 1971, pp. 254 – 255), who, in 1765, introduced the term theory of errors (*Theorie der Fehler*), but it had not taken root until the mid-19th century; Gauss and Laplace, for example, had not applied it.

I repeat now my quotation (Sheynin 1986, p. 311) from Quetelet (1846, p. 65):

En prenant une moyenne, on peut avoir en vue deux choses bien différentes: on peut chercher à déterminer un nombre qui existe véritablement; ou bien à calculer un nombre qui donne l'idée la plus rapprochée possible de plusieurs quantités différentes, exprimant des choses homogènes, mais variables de grandeur.

[*When taking a mean, it is possible to bear in mind two quite different things. We can attempt to determine a number that really exists; or, we can indeed calculate a number that provides the nearest possible idea of many differing quantities expressing uniform objects varying however in magnitude.*]

In the same article I have also cited or mentioned several other pertinent sources from 1830 to 1874.

The study of mean values or states rather than laws of distribution (Galton, see above) had been a necessary stage in the development of natural sciences. Humboldt (Sheynin 1984b, p. 68, n 36), in 1850, mentioned *die einzig entscheidende Methode, die der Mittelzahlen* [*the only decisive method, that of the mean numbers*], and Buys Ballot (Ibidem, p. 55), also in 1850, stated that the study of the mean state of the atmosphere had begun with Humboldt and constituted the first period of the new history of meteorology.

Finally, I refer to Hilbert (1901/1935, § 6) who was perhaps one of the last scholars to mention the *Methode der mittleren Werte* [*method of mean values*]. That the theory of means does not exist anymore is understandable: being an intermediate entity, it became divided between statistics (to which already Quetelet, see the quotation above, had attributed it) and the theory of errors.

Without turning to meteorology anymore, I am giving word to the astronomer, who, in that branch of natural sciences, originated the change from means to frequencies (Kapteyn 1906, p. 397):

Just as the physicist … cannot hope to follow any one molecule in its motion, but is still enabled to draw important conclusions as soon as he has determined the mean of the velocities of all the molecules and the frequency of determined deviations of the individual velocities from this mean, so … our main hope will be in the determination of means and of frequencies.

6. Conclusion

It is generally known that the development of mathematics has always been connected with its moving ever away from Nature (for example, from natural numbers to real numbers in general to imaginaries) and that the more abstract it was becoming, the more benefit accrued to natural sciences. In particular, the general transition from the true value to estimating parameters of functions in mathematical statistics was also very useful.

I stress however that the science of measuring real objects and treating the collected data does not at all abandon the true value. That Mises (§ 3) also saw fit to define (not formally) the true value and to link it (indirectly) to his theory certainly lends it some additional support. Of course, in spite of his own opinion, his frequentist theory of probability belongs to natural sciences [xiii], but, after all, the theory of errors does not belong entirely to mathematics either. The statements of Chatterjee (§ 4) and possibly other likeminded statisticians ought to be modified accordingly and *the theory of errors must remain to be seen as a worthy scientific discipline*. Together with its *true value*, alive and kicking, it continues to service experimental sciences at large.

To a certain extent, the ideas and methods of mathematical statistics ought to be applied there. Primarily I bear in mind the estimation of precision, which, after all, is not inseparably connected with true values. I ought to mention correlation theory and analysis of variance as well, but these subjects are beyond my scope now. Nevertheless, it is opportune to note that Kapteyn (1912), who was dissatisfied with that theory as having been developed then, introduced his own *astronomical version* of correlation. Without knowing it, he thus quantified Gauss' pertinent ideas and, although his contribution had never been cited (perhaps because of this very fact), geodesists have always kept to his (to Gauss') concepts of dependence and correlation, see Sheynin (1984a, pp. 187 – 189). This does not, however, mean that the "statistical" correlation has no place in the theory of errors.

Acknowledgement. It is a pleasant duty to thank the reviewers who indicated some shortcomings both in my own exposition and in my translations from French and German sources.

References

Al-Biruni (1967), *The Determination of the Coordinates of Positions for the Correction of Distances between Cities.* Beirut.

Aleksandrov, P. S., Editor (1962), *English – Russian Dictionary of Mathematical Terms*. Moscow.
Bernstein, S. N. (1941, in Russian), On the Fisherian "confidence" probabilities. In author's book (1964, pp. 386 – 393).
--- (1964, in Russian), *Sobranie Sochineniy* (Coll. Works), vol. 4. N. p.
Bolshev, L. N. (1964, in Russian), Commentary on S. N. Bernstein paper on the Fisherian "confidence" probabilities. In Bernstein (1964, pp. 566 – 569).
--- (1982, in Russian), Processing of observations. *Enc. of Mathematics*, vol. 7. Dordrecht, 1991, pp. 314 – 315.
--- (1984, in Russian), Errors, theory of. Ibidem, vol. 3, 1989, pp. 416 – 417.
Chatterjee, S. K. (2003), *Statistical Thought: a Perspective and History*. Oxford.
Condorcet, M. J. A. de Caritat de (1805), Elémens du calcul des probabilités, et son application aux jeux de hasard, à la loterie, et aux jugemens des hommes. In author's book *Sur les élections et autres textes*. No place, 1986, pp. 483 – 623.
Cotes, R. (1722), Aestimatio errorum in mixta mathesi per variationes partium trianguli plani et sphaerici. In author's *Opera misc*. London, 1768, pp. 10 – 58.
Eisenhart, C. (1963), Realistic evaluation of the precision and accuracy of instrument calibration systems. In Ku, H. H., Editor (1969), *Precision Measurement and Calibrations.* Washington, pp. 21 – 47.
--- (1978), Gauss. In Kruskal, W., Tanur, J. M:, Editors, *International Encylopedia of statistics*, vols 1 – 2. New York, single paging, pp. 378 – 386**.**
Fisher, R. A. (1922), On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc*., vol. A222, pp. 309 – 368.
Fourier, J. B. J. (1826), Sur les résultats moyens déduits d'un grand nombre d'observations. *Œuvres*, t. 2. Paris, 1890, pp. 525 – 545. **S, G,** 88.
Gauss, C. F. (1809, in Latin), Theoria motus … German translation in author's *Abhandlungen zur Methode der kleinsten Quadrate* (1887). Editors, A. Börsch & P. Simon. Vaduz (Lichtenstein), 1998, pp. 92 – 117.
--- (1816), Bestimmung der Genauigkeit der Beobachtungen. Ibidem, pp. 129 – 138.
Hald, A. (1998), *History of Probability and Statistics and Their Application from 1750 to 1930*. New York.
Hilbert, D. (1901), Mathematische Probleme. *Ges. Abh*., Bd. 3. Berlin, 1970, pp. 290 – 329.
Kapteyn, J. C. (1906), Statistical methods in stellar astronomy. [Reports] *Intern. Congr. Arts & Sci. St. Louis – Boston 1904*. N. p., vol. 4, pp. 396 – 425.
--- (1912), Definition of the correlation-coefficient. *Monthly Notices Roy. Astron. Soc*., vol. 72, pp. 518 – 525.
Kolmogorov, A. N. (1946, in Russian), Justification of the method of least squares. *Sel. Works*, vol. 2. Dordrecht, 1992, pp. 285 – 302.
Lambert, J. H. (1760, in Latin), *Photometria*. Augsburg. Its German translation in the *Ostwald Klassiker* series does not include the pertinent section. The German quote in my text is from Schneider (1988, p. 228).
--- (1765), Theorie der Zuverlässigkeit der Beobachtungen und Versuche. In author's *Beyträge zum Gebrauche der Mathematik und deren Anwendung*, Tl. 1. Berlin, pp. 424 – 488.
Laplace, P. S. (1795), Leçons de mathématiques. *Œuvr. Compl.*, t. 14. Paris, 1912, pp. 10 – 177.
--- (1810a), Sur les approximations des formules qui sont fonctions de très grands nombres et sur leur application aux probabilités. Ibidem, t. 12. Paris, 1898, pp. 301 – 345.
--- (1810b), Notice sur les probabilités. In Gillispie, C. C. (1979), Mémoires inédites ou anonymes de Laplace. *Revue d'histoire des sciences*, t. 32, pp. 223 – 279.
--- (1814), *Essai philosophique sur les probabilités*. *Œuvr. Compl.,* t. 7. No. 1. Paris, 1886. Separate paging. English translation (1995):
Philosophical essay on Probabilities. New York.
Markov, A. A. (1899, in Russian), The law of large numbers and the method of least squares. *Izbrannye Trudy* (Sel. Works). N. p., 1951, pp. 231 – 251.
--- (1924, in Russian), *Ischislenie Veroiatnostei* (Calculus of Probability). Moscow. First edition: 1900. German translation of the edition of 1908: Leipzig – Berlin, 1912.
Mises, R. von (1919), Fundamentalsätze der Wahrscheinlichkeitsrechnung.

Math. Z., Bd. 4, pp. 1 – 97. Partly reprinted in author's *Selected Papers*, vol. 2. Providence, Rhode Island, 1964, pp. 35 – 56.
Picard, J. (1693), Observations astronomiques faites en divers endroits du royaume en 1672, 1673, 1674. *Mém. Acad. Roy. Sci*. 1666 – 1699, t. 7, 1729, pp. 329 – 347.
Poincaré, H. (1912), *Calcul des probabilités.* Paris. First edition, 1896.
Poisson, S. D. (1811), Review of a memoir of Laplace. *Nouv. bull. sciences, Soc. philomatique Paris*, t. 2, No. 35, pp. 132 – 136.
--- (1824), Sur la probabilité des résultats moyens des observations. *Connaissance des tem[p]s* pour 1827, pp. 273 – 302.
--- (1829), Second part of same. Ibidem pour 1832, pp. 3 – 22 of second paging.
Quetelet, A. (1846), *Lettres sur la théorie des probabilités*. Bruxelles.
Schneider, I., Herausgeber (1988), *Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfängen bis 1933*. Darmstadt.
Sheynin, O. (1971), Lambert's work on probability. *Arch. Hist. Ex. Sci*., vol. 7, pp. 244 – 256.
--- (1984a), On the history of the statistical method in astronomy. Ibidem, vol. 29, pp. 151 – 199.
--- (1984b), On the history of the statistical method in meteorology. Ibidem, vol. 31, pp. 53 – 95.
--- (1986), Quetelet as a statistician. Ibidem, vol. 36, pp. 281 – 325.
--- (1999), Statistics, definitions of. In Kotz, S., Editor (2006), *Enc. of Statistical Sciences*, 2nd ed., vol. 12. Hoboken, New Jersey, pp. 8128 – 8135.
Smirnov, N. V., Dunin-Barkovski, I. W. (1959, in Russian), *Mathematische Statistik in der Technik*. Berlin, 1973.
Whittaker, E. T., Robinson, G. (1924), *The Calculus of Observations*. London – Glasgow, 1958.

IX

Oscar Sheynin

Poisson and statistics

S.-D. Poisson. *Les mathématiques au service de la science*. Palaiseau, 2013. Editor Yvette Kosmann-Schwarzbach
Poisson et la statistique, pp. 357 – 366
Translated from English by Editor

1. General information

Poisson introduced the concepts of random variable and distribution function. He contributed to limit theorems and brought into use the law of large numbers proving it for the case of *Poisson trials*. He devoted much attention to the study of juridical statistics and systematically determined the significance of empirical discrepancies which proved essential for the development of statistics. Poisson stressed the difference between subjective and objective probabilities. Cournot (1843) kept to the same attitude and even introduced non-numerical probabilities. They as well as the subjective probabilities are being applied as expert estimates.

Arago (1854) discussed Poisson's work in various branches of natural science including important issues concerning our Solar system. Some of those investigations had to be based on statistical data and their treatment which I did not study.

From my viewpoint, Poisson (1837a) is his main contribution. Below, I am repeatedly applying the findings of my previous paper (1978).

2. Statistics

The statistical method is usually understood as applied statistics, and mostly narrower, as applied to natural sciences rather than to human populations or activities. Juridical statistics does not therefore belong to the statistical method, but I am separating it because of its importance for my subject. Some branches of the statistical method have specific names, for example, stellar statistics.

Poisson left several statements concerning statistics and its essential need for the theory of probability. Thus, Quetelet (1869, t. 1, p. 103) testified that Poisson had *parfois* derisively expressed himself *pour les statisticiens qui prétendaient subsister leurs fantaisies aux véritables principes de la science*. Somewhat more definite was the pronouncement (Libri Carruci et al 1834, p. 535): *The most sublime problems of the arithmétique sociale can be only resolved with the help of the theory of probability*.

(*les questions les plus élevées d'arithmétique sociale ne pouvaient être résolues d'une manière complète qu'ù l'aide du calcul des probabilités.*) I only left the French text.

Social arithmetic was a short lived term possibly coined by him that denoted demography, medical statistics and actuarial science. And, finally (Double et al 1835, p. 174) with Poisson as co-author:

La statistique mise en pratique, qui est toujours en définitive le mécanisme fonctionnant du calcul des probabilités, appelle nécessairement des masses infinies, un nombre illimité de faits …

This was likely one of the first pronouncements linking statistics with a large amount of data.

Since he (1837a) consistently demanded to check the significance of empirical discrepancies, for example between results of different series of observations, Poisson, along with Bienaymé, can be called the Godfather of the Continental direction of statistics (Lexis, Bortkiewicz, Chuprov, Markov, Bohlmann) that studied population and social statistics. True, his approach was definitely restricted as it became apparent in medicine.

His generally known formula (1837a, p. 206)

$$P \approx e^{-\omega}(1+\omega+\omega^2/2!\ +\ldots+\omega^n/n!),\ \omega = \mu p$$

for an event having probability $q = 1 - p \approx 0$ to occur not more than n times in a large number μ of Bernoulli trials. This formula had been all but ignored until Bortkiewicz (1898) introduced his law of small numbers, allegedly a breakthrough extremely important for statistics. However, in 1954 Kolmogorov had identified it as the Poisson formula. He did not justify that statement, and I (2008) proved it.

Poisson's law of large numbers (LLN) is his best known innovation. The first version of the LLN was due to Jakob Bernoulli who proved that, given a series of (independent) trials with a constant probability p of the occurrence of the studied event (of *success*), the frequency υ of that success tended to p, and he also estimated the rapidity of that process. In 1733, De Moivre discovered a new and much better form of the LLN by proving the first version of the central limit theorem.

Poisson (1837a) generalized the LLN on the case of variable probabilities p_i of success in different trials although many authors have reasonably noted that his proof was not rigorous.

There is a lesser known aspect of the LLN. All three scholars, Bernoulli, De Moivre and Poisson, alleged that their findings were just as applicable for the inverse case, in which the probability p (or probabilities p_i) was (were) unknown and had to be estimated by the observed frequency. Even more: Bernoulli and Poisson (1836; 1837a) thought that even the existence of p (or p_i) was not necessary. The former provided an example of an individual being taken ill by an infectious disease, Poisson mentioned several such cases such as stability of the mean sea level, of the mean interval between molecules of a body, and (1837a, § 59) of the sex ratio at birth.

Nevertheless, it was Bayes (1765) who investigated the inverse case. In both the direct and inverse cases the behaviour of a centred and normed random variable was studied, of $(\upsilon - E\upsilon)/\text{var}\,\upsilon$ and $(p - Ep)/\text{var}\,p$, for any p_i, respectively. The concept of variance was unknown to Bayes, but he proved that $\text{var}\,p > \text{var}\,\upsilon$ which was reasonable since in the inverse case we have less information. In other words, for achieving the same precision, the inverse case demanded more observations. Bayes had thus completed the first version of the theory of probability but his finding was only noted by the Editor of the German translation (1908) of his memoir. Earlier, Chebyshev (1879/1880, pp. 186 – 192), see also Sheynin (1994, p. 333) repeated Bayes' finding but did not cite anyone.

Many authors have stated that statisticians had been happily justifying their investigations by the LLN. Actually, they only recognized the LLN for the case of Bernoulli trials, and only when the probability of the studied event existed and (was constant), otherwise they refused to turn to the theory of probability at all. Even worse, as a rule, they only understood the LLN in a loose sense. Thus, Maciejewski (1911, p. 96) introduced *la loi des grands nombres des statisticiens* which only stated that the fluctuations of statistical numbers decreased as the number of observations increased.

Poisson published a memoir (1830) on the sex ratio *m:f* at birth. He noted that it was roughly the same over the whole country and stated that that ratio was lower for births out of wedlock.

His programme of probability calculus and social arithmetic (1837b) devoted serious attention to that subject. I quote the appropriate part of the programme:

Des tables de population et de mortalité. De la durée de la vie moyenne dans diverses contrées. Partage de la population suivant les âges et les sexes. De l'influence de la petite vérole, de l'inoculation et de la vaccine sur la population, et la durée de la vie moyenne.

Inoculation of smallpox meant communicating (a mild form of) the disease from an ill to a healthy person, not quite safe but very beneficial when considered for a large number of people. The most celebrated pertinent statistical study was due to Daniel Bernoulli. Inoculation was practised before (and somewhat longer than) vaccination became available.

Statistics of financial institutions can be dealt with separately. It was also prominently present in the same programme:

Des bénéfices et des charges des établissements qui dépendent de la probabilité des événements. Des rentes viagères, des tontines, des caisses d'épargne, des assurances, des annuités, des fonds d'amortissement, des emprunts.

Tontines were groups of annuitants of about the same age considered by the entrepreneurs (usually by the appropriate state) as single entities. A tontine distributed yearly payments among still living members, and those living long enough came to enjoy quite considerable moneys. The term stemmed from the name of the Italian banker Lorenzo Tonti (1630 – 1695).

Many scholars contributed to this subject, suffice it to mention De Moivre, Euler and Markov. Poisson, however, apparently only participated in reviewing the desirability of establishing a tontine (Fourier 1826). The reviewers opposed that proposal and mentioned the negative properties of tontines.

2.1. Theory of errors

The stochastic theory of errors is the particular case of the statistical method as applied to the treatment of observations. From the mid-18th century to ca. 1930 it remained the main field of application of probability theory whereas statistics borrowed its principles of maximum likelihood and least variance.

In 1805, Legendre introduced without justification the principle of least squares (known to Gauss from 1795) which belongs to the theory of errors. Gauss provided two substantiations of the *method* of least squares, in 1809 and 1823, the second one based on the principle of least variance, and suited for adjusting a finite (and even small) number of observations and never publicly acknowledged Legendre's priority.

Laplace offered his own version of the method which demanded a large number of observations and least absolute expectation of error. The second condition meant that calculations were only possible for the case of normal distribution, and both demands taken together led to practical uselessness.

Poisson followed Laplace and never mentioned the Gauss version partly since French mathematicians had been reasonably angered by Gauss' attitude towards Legendre. Here is what he (1833, p. 361) stated at the funeral of Legendre:

Notre confrère est auteur d'une méthode pour le calcul des orbites des comètes. [...] *C'est à lui qui les sciences d'observation sont redevables d'une règle de calcul qu'il a nommée Méthode des moindres carrés des erreurs, et dont Laplace à montré tout avantage probable sous le rapport de la précision des résultats ...*

A wrong and harmful attitude! *These words did not appear in the French translation.*

I do not consider the determinate error theory which now ought to be included in experimental design and with which Poisson did not deal.

One additional point is interesting. When discussing the precision of firing, Poisson (1837c, p. 73) stated that the less was the scatter (the appropriate variance) of hit-points, the better was the gun. He thus made a step towards recognizing Gauss' choice of least variance as the main criterion for adjusting observations according to his mature version of the method of least squares.

2.2. Juridical statistics

Laplace and Poisson studied the ideal case of independent decisions reached by jurors; Laplace (1812/1886, p. 523) mentioned this restriction only in passing, Poisson did not say anything about it. Unlike Laplace, Poisson introduced the prior probability of the defendant's guilt, a magnitude certainly not to be applied in individual cases. Poisson desired to test the stability of the rate of conviction and to compare different legal proceedings with a view to minimize the number of possible unjust verdicts. The issue of independence of the votes was here hardly important.

One of Poisson's statements (1837a, pp. 375 – 376) is debatable: he thought that the rate of conviction should increase with crime. At the same time he (p. 21) recognized that such numbers represented *l'état moral de notre pays*. Concerning convictions, Poisson (p. 6) likely followed Laplace who had stated that accusing an innocent person should remain more dangerous than that of acquitting a guilty man. The pertinent findings could have been guiding the establishment of the proper number of witnesses and jurors etc. That was Gauss's statement as reported by W. E. Weber in a letter of 1841 and published since then (Gauss 1929, pp. 201 – 204).

Quetelet devoted much attention to the same subject; his first pertinent contributions were published even before those of Poisson, but in the long run he undoubtedly profited from the very fact that that outstanding *géomètre* (like Laplace before him) linked his name with juridical statistics. Although he did not pursue his work on a mathematical level, Quetelet was able to make a valuable contribution.

Nevertheless, the application of probability theory to jurisprudence had been criticized time and time again. Poinsot, who participated in the discussion of Poisson (1836), called it, on p. 380, *une fausse application de la science mathématique* and unwisely quoted Laplace (1814/1886, p. XI) who had remarked that the theory of probability was very delicate.

I say *unwisely*, because the same *Essai philosophique* contains a page (p. LXXVIII) entitled *Application du calcul des probabilités aux sciences morales* where Laplace declared that such applications are the *effets inévitables du progrès des lumières*. Moreover, the same *Essai* contained three chapters devoted to such applications to say nothing of Laplace's own work on juridical statistics.

Then, Mill (1843/1886, p. 353) had called the application of probability to jurisprudence an *opprobrium* [disgrace] *of mathematics*. In 1899, Poincaré (Sheynin 1991, p. 167) appraisingly cited him in connection with the notorious Dreyfus case. Later he (1896/1912, p. 20) stated that people *régissent les uns sur les autres* and act like the *moutons de Panurge*.

Heyde & Seneta (1977, pp. 28 – 34) devoted some attention to juridical statistics and noticed, on p. 31, that *there was a surge of activity stimulated by Poisson*. Regrettably omitting Cournot and Quetelet, they described the relevant work of Bienaymé, Ostrogradsky, Buniakovsky and put on record the recent resurgence of interest in the application of probability and statistics to jurisprudence coupled with an increased understanding of the importance of interpreting background information. True, many authors (Leibniz, in his letters to Jakob Bernoulli in the very beginning of the 18th century, Mill, see below) always kept to the same viewpoint concerning the pertinent circumstances.

Gelfand & Solomon (1973) reviewed Poisson's study and included information about the French legal system of his time. They (p. 273) somewhat softened the issue of the interdependence of jurors:

There is much evidence [a reference to a source published in 1966 is provided] *to show that immediate voting by jurors before any deliberation produces essentially the same result as after deliberation*.

And they add that the jurors can possibly submit their verdicts by secret ballot. At the very best, however, there still remains the interdependence of jurors caused by their likely similar upbringing and social standing.

2.3. Statistical physics

Poisson qualitatively connected his law of large numbers with the existence of a stable mean interval between molecules (Gillispie 1963, p. 438; Sheynin 1978, p. 271). Clausius, Maxwell and Boltzmann could have well mentioned this opinion as also his important related considerations, but nothing of the sort actually happened.

2.4. Medical statistics

Is it possible to reconcile the individual approach to a given patient with an abstract statistical point of view? This question is the same here as in juridical statistics, and the answer is the same. When reviewing a contribution at the Paris Academy of Sciences Double et al (1835, p. 173, 174 and 176) with Poisson as co-author stated:

En matière de statistique, c'est-à-dire dans les divers essais d'appréciation numérique des faits, le premier soin avant tout c'est de perdre de vue l'homme pris isolément pour ne le considérer que le comme une fraction de l'espèce. [...]

En médecine appliquée au contraire, le problème est toujours individuel [...]

La condition des sciences médicales, à cet égard n'est pas pire, n'est pas autre que la condition de toutes les sciences physiques et naturelles, de la jurisprudence, des sciences morales et politiques. etc.

Anyway, the statistical method did gnaw its way into medicine. First, population statistics was closely connected with medical problems as it happened in the pioneer work of Graunt. Leibniz busied himself with demography (Sheynin 1977, p. 225). He did not collect statistical data but he urged practitioners to record their observations and he also proposed to compile an encyclopaedia of medical science and to establish a special *Collegium Sanitatis*. Halley compiled the first (after Graunt's not really reliable finding) mortality table for a closed population and estimated populations from data on births and deaths. Daniel Bernoulli, Lambert and Euler studied mortality, birth rates and sicknesses and their work belongs to the history of probability and of medicine.

Second, the range of application of the statistical method greatly widened after the emergence, in the mid-19th century, of public hygiene (largely a forerunner of ecology) and epidemiology. Third, about the same time surgery and obstetrics, branches of medicine proper, yielded to the statistical method.

Fourth and last, in 1825 a French physician P. Louis introduced the so-called numerical method (actually applied much earlier in various branches of science) of studying symptoms of various diseases. His proposal amounted to the use of the statistical method without involving stochastic considerations. Discussions about that method lasted at least a few decades. Thus, d'Amador (1837) attacked Louis wrongly attributing to him a recommendation to use the theory of probability.

Gavarret (1840) definitely noted the shortcomings of the numerical method and introduced two formulas necessary for the application of probability theory, the formulas for the normal approximation of the binomial distribution and for the Poisson estimate of the permissible difference between frequencies of the occurrence of an event in two series of binomial trials with variable probabilities.

He adduced examples on the use of the second formula and, in particular, on the comparison of competing methods of medical treatment as also an advice on the check of the null hypothesis (as it is now called), see p. 194:

Le premier travail d'un observateur qui constate une différence dans les résultats de deux longues séries d'observations, consiste donc à chercher si l'anomalie n'est qu'apparente, ou si elle est réelle et accuse l'intervention d'une cause perturbatrice; il devra ensuite [...] *chercher à déterminer cette cause*.

Thus, apart from popularizing probability theory, Gavarret's main achievement was the introduction of the principle of the null hypothesis and of its check into medicine (actually, in natural science in general). His contribution became generally known and many authors repeated his recommendations. The time for mathematical statistics or for its application in medicine was not yet ripe, but at least the Poisson – Gavarret tradition led to the existence, in medicine, of a lasting drive towards the use of probability based on numerous observations (and the skill of the physician).

Before taking to medicine, Gavarret had graduated from the Ecole Polytechnique where he studied under Poisson. He (1840, p. XIII) sincerely acknowledged Poisson's influence:

Ce n'est qu'après avoir long-temps médité les leçons et les écrits de l'illustre géomètre, que nous sommes parvenu à saisir toute l'étendue de cette question [...] *de régulariser l'application de la méthode expérimentale* (!) *à l'art de guérir*.

A large number of observations! However, at least from the mid-18th century (Bull 1959, p. 227) valuable medical conclusions had been based on very small numbers of them, but it was Liebermeister (ca. 1877) who vigorously opposed Gavarret and Poisson. He argued that it was impossible, in therapeutics, to collect vast observations and, anyway, recommendations based on several (reliable) observations should be adopted as well. Statisticians have only quite recently discovered his paper written as though by a specialist in mathematical statistics.

Bibliography

d'Amador R. (1837), *Mémoire sur le calcul des probabilités appliqué à la médecine*. Paris.
Arago F. (1854, lu 1850), Poisson. *Oeuvres, Notices biographiques*, tt. 2. Paris, pp. 593 – 671.
Bayes T. (1765), A demonstration of the second rule in the essay [of 1764] towards the solution of a problem in the doctrine of chances. *Phil. Trans. Roy. Soc*. for 1764, vol. 54, pp. 296 – 325.
--- (1908), *Versuch zur Lösung eines Problems der Wahrscheinlichkeitsrechnung.* Hrsg. H. E. Timerding. Leipzig. Ostwald Klassiker No. 169.
von Bortkiewicz L. (1898), *Das Gesetz der kleinen Zahlen*. Leipzig.
Bull J. P. (1959), The historical development of clinical therapeutic trials. *J. Chronic Diseases*, vol. 10, pp. 218 – 248.
Chebyshev P. L. (1879/1880), *Teoria Veroiatnostei* (Theory of probability). Lectures as written down by A. M. Liapunov. Moscow – Leningrad, 1936.
De Moivre A. (1733, in Latin), A method of approximating the sum of the terms of the binomial $(a+b)^n$ etc. Included in translation in the subsequent editions of the author's *Doctrine of Chances* (1738, 1756); in 1756, an extended version is on pp. 243 – 254.
Double F. J, rapporteur, Dulong P. L., Larrey F. H., Poisson S. D. (1835), Review of Civiale, Recherches de statistique sur l'affection calculeuse. *C. r. Acad. Sci. Paris*, t. 1, pp. 167 – 177.
Fourier J. B. J., rapporteur, Poisson S. D., Lacroix S.-F. (1821, publ. 1826), Rapport sur les tontines. In Fourier (1890), *Oeuvres*, t. 2. Paris, pp. 617 – 633.
Gauss C. F. (1929), *Werke*, Bd. 12. Göttingen. All 12 volumes of the *Werke* reprinted: Hildesheim, 1973 – 1981.
Gavarret J. (1840), *Principes généraux de statistique médicale*. Paris.
Gelfand A. E., Solomon H. (1973), A study of Poisson's models for jury verdicts in juridical and civil trials. *J. Amer. Stat. Assoc.*, vol. 68, pp. 271 – 278.
Gillispie C. (1963), Intellectual factors in the background of analysis by probabilities. In: *Scientific Change*. Ed., A. C. Crombie. New York, 1963, pp. 431 – 453.
Heyde C. C., Seneta E. (1977), *I. J. Bienaymé.* New York.
Laplace P. S. (1812), *Théorie analytique des probabilités. Oeuvr. Compl.*, t. 7. Paris, 1886.
--- (1814), *Essai philosophique sur les probabilités*. Ibidem, separate paging.
Libri-Carruci G. B. I. T., rapporteur, Lacroix S. F., Poisson S. D. (1834), Report on Bienaymé's manuscript. *Procès verbaux des séances Acad. Sci. Paris*, t. 10, pp. 533 – 535.
Liebermeister C. (ca. 1877), Über Wahrscheinlichkeitsrechnung in Anwendung auf therapeutische Statistik. In *Sammlung klinischer Vorträge*. Innere Medizin, NNo. 31 – 61. Leipzig, n. d., No. 39 (No. 110 of the whole series), pp. 935 – 962.
Maciejewski C. (1911), *Nouvaux fondements de la théorie de la statistique.* Paris.
Mill J. S. (1843), *System of Logic*. London, 1886. Also *Coll. Works*, vol. 8. Toronto, 1974.
Poincaré H. (1896), *Calcul des probabilités*. Paris, 1912, 1987.
Poisson S.-D. (1824), Observations relatives au nombre de naissances des deux sexes. *Annuaire de Bureau des longitudes* pour 1825, pp. 98 – 99.
--- (1830), Sur la proportion des naissances des filles et des garcons. *Mém. Acad. Sci. Paris*, t. 9, pp. 239 – 308. Preceded by the note of 1824.
--- (1833), Discourse prononcé aux funéralles de M. Legendre. *J. für d. reine u. angew. Math.*, Bd. 10, pp. 360 – 363.
--- (1836, April 11 and 18), Note sur la loi des grandes nombres. *C. r. Acad. Sci. Paris*, t. 2, pp. 377 – 382, 395 – 400.
--- (1837a), *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*. Paris. Also Paris, 2003. **S, G,** 53.
--- (1837b), Elements du calcul des probabilités et arithmétique sociale, this being a part of the *Programmes de l'enseignement de l'Ecole Polytechnique* [...] *pour l'année scholaire 1836 – 1837*. Paris.
--- (1837c), Sur la probabilité du tir a la cible. *Mémorial d'artillerie*, No. 4, pp. 59 – 94.
Quetelet A. (1846), *Lettres sur la théorie des probabilités*. Bruxelles.
--- (1869), *Physiqe sociale*, tt. 1 – 2. Bruxelles.

Sheynin O. (1977), Early history of the theory of probability. *Arch. Hist. Ex. Sci.*, vol. 17, pp. 201 – 259.
--- (1978), Poisson's work in probability. Ibidem, vol. 18, pp. 245 – 300.
--- (1982), On the history of medical statistics. Ibidem, vol. 26, pp. 241 – 286.
--- (1991), Poincaré's work in probability. Ibidem, vol. 42, pp. 137 – 172.
--- (1994), Chebyshev's lectures on the theory of probability. Ibidem, vol. 46, pp. 321 – 340.
--- (2008), Bortkiewicz' alleged discovery: the law of small numbers. *Hist. Scientiarum*, vol. 18, pp. 36 – 48.

X

Oscar Sheynin

Where are Kolmogorov's posthumous papers?

Math. Intelligencer, vol. 39, No. 4, 2017, p. 46

In a worthwhile tradition, the Archive of the Russian Academy of Sciences (RAN) collects and keeps the posthumous papers of the late Russian academicians. Kolmogorov died in 1987, so I asked the Academy for permission to look at his papers. I found that the RAN did not have Kolmogorov's papers. Staff at their Archive advised me to inquire at the Archive of Moscow University where Kolmogorov had been a staff professor.

I twice asked Moscow University for information but did not receive an answer. So I asked the Presidium of the RAN. An anonymous representative from the Class of Mathematical Sciences answered that nothing was known about Komogorov's papers. Period! As though they were not interested at all ...

A colleague told me that Albert Shiriaev, professor at Moscow University, perhaps kept those papers. I wrote to Shiriaev but he did not answer. I hope that some members of the scientific community will write to their Russian colleagues and ask them to help remedy the situation by providing any available information regarding Kolmogorov's papers.

XI

Oscar Sheynin

On the history of Bayes's theorem

Math. Scientist, vol. 28, 2003, pp. 37 – 42

Abstract

This paper reconsiders the history of Bayes's theorem, and analyses several possibilities with regard to its authorship. The conclusion reached is that Bayes was the author of the results in the memoir on *the doctrine of chances* presented to the Royal Society of London by his friend Richard Price in 1763 – 1764.

_ 1. Introduction

A few years after Thomas Bayes (1702 – 1761) had died, the Royal Society of London published his memoir on *the doctrine of chances* with additions compiled by Richard Price (in two parts, Bayes (1763) and Bayes (1764)). This paper considers the recently expressed doubts about the priority for that theorem, for which, at least for the sake of convenience, I retain the traditional term *Bayes's theorem*. For biographical information about Bayes and Price, see Pearson (1978, pp. 355 – 362, 370 – 377), Edwards (1993) and Kruskal (1978).

The memoir is still a source of endless debate. One of the issues is why Bayes had apparently never attempted to publish it. Being a perspicacious mathematician, he could well have understood that his work was not quite perfect (but even so, extremely important). Indeed, Hald (1998, p. 135) suggests that Bayes was too modest; could it be, however, that he made known his discovery privately but did not dare publish it (see § 6)? Two relevant facts are the presumption of prior ignorance (as in the celebrated problem inserted by Price on the probability of the next sunrise) and the need for a better estimation of the integral in the formula (l) below, cf. Hald (1998, pp. 135, 144).

I am unable to answer this question. Note, however, that De Moivre had for some reason postponed for twelve years the appearance of his *De Moivre – Laplace* limit theorem. He mentioned this delay in the original 1733 Latin text of De Moivre (1756), although he had carried out all the main algebraic deductions needed for deriving his theorem in 1730; see [ii, § 3.1].

The debate to which I am now contributing began when Stigler (1983) quoted a curious statement, see Hartley (1749, pp. 338—339) and interpreted it as a testimony against Bayes's priority. After referring to De Moivre, Hartley wrote, in part:

An ingenious friend has communicated to me a solution of the inverse problem of determining the probability of an event given the number of times it happened and failed.

He then specifically mentioned the case of a large number of trials.

Later, Stigler (1986, pp. 98, 132) recalled Hartley (1749) and a paper of his own, Stigler (1983), but did not definitively repeat his previous inference. Strangely enough, Stigler (1999) reprinted the above paper and added a tiny footnote brushing aside any criticism. I now begin the main story by describing some findings made by Bayes and earlier scholars; see also Sheynin (1968; 1971).

2. Bayes's theorem

Hald (1998, p. 141) noted that John William Lubbock and John Elliot Drinkwater-Bethune used (1), see below, in 1830 to express the probability $P(b \le r \le c)$, which is about to be specified.

Here is Bayes's reasoning; for a detailed description, see Hald (pp. 133 – 154). Roughly speaking, Bayes considered the fall of an object on some point *r*, belonging to a square *ABCD*, on either side of some straight line *MN* parallel to *AB* and *CD* and situated between them. Without loss of generality, assume that *AB* = 1. All the positions of *MN* as well as of the points of fall with respect to *AB* and *CD* are equally probable. Let *bc* be a segment situated inside *AD* and suppose that, after $n = p + q$ trials, the point *r* occurred *p* times to the right of *MN* and *q* times to the left of it. Then

$$P(b \le r \le c) = \frac{\int_b^c x^p (1-x)^q \, dx}{\int_0^1 x^p (1-x)^q \, dx}. \qquad (1)$$

For the denominator of (1) Bayes obtained

$$\mathrm{B}(p + 1;\, q + 1) = \frac{1}{(q+1)C_{p+q+1}^p}$$

and exerted considerable efforts in estimating its numerator. The right-hand side of (1) is, in modern terms, the difference of two values of the incomplete beta function:

$$P(b \le r \le c) = I_c\,(p+1;\, q+1) - I_b(p+1;\, q+1).$$

Bayes did not study the case of $n \to\infty$ whereas Price, in his covering letter (see Bayes (1763, p. 135)), indicated that De Moivre's *rules* are not pretended to be rigorously *exact* in the case of a finite *n*. Nevertheless, by applying a clever nick, H. T. Timerding, the editor of the German translation of the Bayes memoir, inserted the limiting case. He proved that, when estimating the probability that point *r* is to the right of *MN*, then, as $n \to\infty$,

$$\lim P(b \le r \le c) = (-z \le \frac{\hat{p} - a}{[pq/n^{3/2}]^{1/2}} \le z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-x^2/2} dx, \qquad (2)$$

$$a = p/n = \mathrm{E}\,\hat{p},\; pq/n^{3/2} = \mathrm{var}\,\hat{p}, \qquad (3)$$

where the interpretation (3) is my own.

The term *Bayes's theorem* is, however, usually understood otherwise, namely as

$$P(A|\mathrm{B}) = \frac{P(A_i)P(B \mid A_i)}{\sum_{j=1}^{n} P(A_j)P(B \mid A_j)}. \tag{4}$$

Both the conditions for this formula and the notation used are generally known. Bayes himself only proved that

$$P(A|\mathrm{B}) = \frac{P(AB)}{P(B)},$$

a formula known to De Moivre (1756, p. 7).

In this second sense, the term *Bayes's theorem*' originated with Cournot (1843), who, nevertheless, felt some doubts about its attribution; see his § 88.

Thus, the term *Bayes's theorem* is ambiguous, at least historically speaking. For me, however, Bayes's main results are (1) and (2).

3. Jakob Bernoulli, De Moivre and Bayes

3.1. Jakob Bernoulli

Jakob (or Jacob or Jacques) Bernoulli established the correspondence between the theoretical (p) and the estimated ($\hat{p}$) probabilities in the limiting case (as the number of *Bernoulli trials* increased). The last lines of his unfinished classic Bernoulli (1713, p 265) hinted at a solution of the inverse problem, namely determining p given $\hat{p}$. If all the events are (I would say, if the studied event is) observed forever, he wrote, probability will become certainty. Elsewhere (p. 248), Bernoulli expressed a similar opinion concerning the case of a large finite number of observations. The ratio of the appropriate chances, he maintained, would then be probably (wahrscheinlicher Weise) determined.

3.2. De Moivre

De Moivre (1756) determined the probability of the frequency of the random number μ of successes in n Bernoulli trials. His *De Moivre – Laplace* limit theorem might be written as

$$\lim \mathrm{P}\left(a \le \frac{\mu - np}{\sqrt{npq}} \le b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx, \text{ as } n \to \infty \tag{5}$$

with $q = 1 - p$, $np = \mathrm{E}\mu$, $npq = \mathrm{var}\mu$ and where p is the probability of success in each trial. De Moivre (1756) studied the particular case of $p = 1/2$ but he concluded (p. 250) that the general case *will be solved with the same facility*. Indeed, the title of his memoir refers to the binomial $(a + b)^n$, rather than to the binomial $(1/2 + 1/2)^n$.

Laplace (1812, pp. 280 – 286) improved its derivation and, in addition, considered the case of a large finite n. Neither De Moivre nor Laplace knew about uniform convergence or variance (introduced by Gauss in 1823), nor did they distinguish between strict and non-strict inequalities.

De Moivre also thought about inverting his theorem. Enlarging on his memoir of 1733, he (1756, p. 251) mentioned the inverse problem although did not study it:

Conversely, if from numberless observations we find the Ratio of the Events to converge to a determinate quantity, … then we conclude that this Ratio expresses the determinate law according to which the Event is to happen …

To clarify, instead of *Ratio of Events* read *frequency of the studied event*. De Moivre thus called the implied binomial distribution determinate; only the empirical deviations from it remained random. He then went on to discuss natural laws from such a viewpoint.

The same source (De Moivre 1756) contains an *Advertisement* (p. xi) inserted by the anonymous editor who stated that the *whole* book was finally compiled *according to the plan concerted with the Author, above a year before his death* (in 1754).

But when did De Moivre approve the final text of his book? Late in life, his eyesight failed him (as the anonymous editor testified), although possibly not completely, and, in general, he withdrew from scientific work. However, it is still possible that he read Hartley's remarkable statement or came to know about it from someone else. If so, he would have had a personal reason for mentioning the inverse problem.

3.3. Bayes (addendum to § 2)

The functions on the left-hand sides of both (2) and (5) might at present be written in the same way as

$$\frac{\xi - E\xi}{\sqrt{\operatorname{var}\xi}}$$

where ξ is a random variable. It is remarkable that Bayes apparently understood that (5) did not adequately describe the inverse problem. In any case, here again is Price, in Bayes (1763, p. 135)):

I know of no person who has shewn how to deduce the solution of the converse problem … What Mr De Moivre has done … cannot be thought sufficient to make the consideration of this point unnecessary.

4. Bayes's priority is doubted

After drawing on various sources, Stigler (1983) concluded that

1. Hartley had in essence completed the text of his book (Hartley 1749) in 1739;

2. He was well acquainted with Nicholas Saunderson (1682 – 1739), a broad-minded blind mathematician;

3. Both Roger Cotes and De Moivre held a high opinion of Saunderson, who, in particular, was acquainted with De Moivre's 1733 memoir;

4. Nothing is known about the possible acquaintance of Bayes with De Moivre or Hartley.

Stigler inferred that Saunderson was the author of Bayes's theorem and, by applying (4), he even found that his conclusion was three times more probable than the former opinion.

However, the prior probabilities of the two hypotheses cannot be equal, as Stigler assumed. Not only a honest personality like Saunderson, but almost any pretender will be able to claim equal prior rights with an established author (or a politician of the past).

It followed that the posterior probabilities of the identity of the anonymous friend, as assigned by him, are doubtful. For example, items 2 and 4 led, in his opinion, to favouring Saunderson in the ratio 3:1. Three such posterior considerations provided a single probability (accidentally) equal, again, to 3:1, and the equality of the prior probabilities meant that the same ratio in favour of Saunderson persisted in Stigler's final answer.

Stigler admits that Bayes could have been, after all, the real author if only he had read Hartley (1749), but then Bayes would probably have referred to him. Granted, the circumstances are somewhat confusing since the introduction to the Bayes memoir is lost; Price presumably disregarded it and provided instead his own covering letter. The quotation from it in § 3.3 seems to show, however, that the notion of the existence of a Bayes's predecessor, as implied by Stigler, is hardly justified. [This means that the extra-mathematical arguments (for example, the evidence of Price, a close friend of Bayes) are not considered at all. And it is opportune to recall the opinion of Gauss: applications of the theory of probability can be greatly mistaken if the essence of the studied object is disregarded.]

5. The ensuing debate

1. Edwards (1986) believed that Bayes was indeed acquainted with Hartley's statement and that the latter had described de Moivre's finding. His conclusion is probably wrong because Hartley mentioned both De Moivre and somebody else. Recall also (§ 3.2) that De Moivre only declared that he solved the inverse problem, whereas the comparison of (5) and (2) is evidence against him.

2. Gillies (1987, p. 329) considered it essential that, after the Bayes memoir was published, no one complained about loss of priority. However, both Saunderson and Hartley died before 1764. Gillies also supposed that Bayes could have been finally prompted to compile (or conclude?) his work in 1748 by David Hume's reasoning on the method of induction in his *Enquiry Concerning Human Understanding*. Bayes, as Gillies is prepared to believe, had indicated his finding to Hartley, who was then able, at the last minute, to supplement his manuscript accordingly.

3. Dale (1988, p. 358) thinks that the Hartley statement could quite possibly have influenced Bayes. It seems that he also believes, as Edwards did, that Hartley bore De Moivre in mind.

4. Hald (1990, p. 400) described the discussion but did not formulate his own opinion.

5. Dale (1991) described a manuscript and a notebook that, according to the handwriting, had formerly belonged to one and the same person (p. 313). He provided arguments (pp. 313, 322) showing that that person was Bayes, and, on p. 322, he remarked that the notebook contained the proof of 'Rule 2' from the Bayes memoir. For me, the essential point here is that Bayes had apparently not dated that proof.

6. Hald (1998, p. 132) again referred to the previous authors, and, without presenting any arguments, concluded that Bayes himself was the author of the Bayes theorem.

7. Dale (1999, p. 8) repeated (cf. items 1 and 3 above) that Hartley had thought about de Moivre, but he said nothing about the authorship;

[Zabell (1975, p. 316) concluded that Stigler's opinion *cannot be seriously credited*. His paper, as I see now, is extremely important and includes much material unconnected with Bayes.]

6. Conclusion

It remains unknown when Bayes completed his work, but it is perhaps relevant that he effectively retired from his ministry in 1749, not later than when Hartley's book was published. It is possible that Bayes, being excessively modest (see § 1), then chose to convey (2) (here and below, I mean the Bayes version of the formula) privately to Hartley who, as Gillies believes (see item 2 of § 5), was able to insert a few appropriate lines in his still unpublished book.

In any case, the existence of some predecessor of Bayes seems unlikely (and, anyway, the proof of (2) is in the Bayes memoir). Bayes understood that (5) did not properly estimate the studied event, given a large number of Bernoulli trials; on the other hand, he hardly thought about deriving (4), and it follows that in this sense the term *Bayes's theorem* is a misnomer.

For Bayes, the importance of (2) could have been evident because, taken together with the then known elements of the new mathematical discipline (certainly including (5)) it amounted to a complete contemporary doctrine of chances; Even had Bayes proved (4), it provided nothing comparable.

Acknowledgement. The referee noticed a serious non-mathematical mistake in a previous version of this paper. His further work was also helpful.

References

BAYES, T. (1763). An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Roy. Soc. London* 53, 370 – 418. Printed in 1764. German translation (edited by H. T. Timerding): Leipzig, 1908. Reprints of the original edition (with a biographical note by G. A. Barnard): *Biometrika* 45 (1958), pp. 293 – 315 and *Studies in the History of Statistics and Probability*, vol. 1, eds. E. S. Pearson and M. G. Kendall, Griffin, London, 1970, pp. 131 – 153. Quotations in text are from the last-mentioned source.

BAYES, T. (1764). A demonstration of the second rule in the essay towards the solution of a problem in the doctrine of chances (communicated by R. Price). *Phil. Trans. Roy. Soc. London* 54, pp. 296 – 325. Printed in 1765. German translation (edited by H. T. Timerding): Leipzig, 1908.

BERNOULLI, J. (1713). *Ars Conjectandi*. Basel. German translation (its latest edition): *Wahrscheinlichkeitsrechnung*, Thun, Frankfurt am Main, 1999.

COURNOT, A. A. (1843). *Exposition de la theorie des chances et des probabilités*. Hachette, Paris. Reprinted in: *Eeuvr. Compl*., t. 1. Ed. B. Bru,J. Vrin, Paris, 1984. **S, G,** 54.

DALE, A. I. (1988). On Bayes' theorem and the mverse Bernoulli theorem. *Hist. Math.* 15, pp. 348 – 360.

--- (1991). Thomas Bayes' s work on infinite series. Ibidem. 18, pp. 312 – 327.

--- (1999). *A History of Inverse Probability From Thomas Bayes to Karl Pearson*, 2nd edn. Springer, New York.

DE MOIVRE, A. (1756). A method of approximating the sum of the terms of the binomial $(a + b)^n$ expanded into a series. Privately printed in 1733 (in Latin). Incorporated, with additions, in author' s *Doctrine of Chances*, London, 1738 and 1756. Reprint of 1756 edn: Chelsea, New York, 1967 (an English translation of the memoir of 1733 occupies pp. 243 – 254). Quotations in text are from this translation.
EDWARDS, A. W. F. (1986). Is the reference to Hartley (1749) to Bayesian inference? *Amer. Statistician* 40, pp. 109 – 110.
--- (1993). Bayes. In *Dictionary of National Biography: Missing Persons*, Ed. C. S. Nicholls, Oxford Univ. Press, pp. 50 – 51.
GILLIES, D. A. (1987). Was Bayes a Bayesian? *Hist. Math*. 14, pp. 325 – 346.
HALD, A. (1990). *A History of Probability and Statistics and Their Applications Before 1750*. John Wiley, New York.
--- (1998). *A History of Mathematical Statistics from 1750 to 1930.* John Wiley, New York.
HARTLEY, D. (1749). *Observations on Man, His Frame, His Duty and His Expectations.* London. Reprint of 1791 edn: Woodstock Books, Poole, 1998.
HUME, D. (1748). *An Enquiry Concerning Human Understanding*. Reprinted: Oxford University Press, 2000.
KRUSKAL, W. H. (197 8). Price, Richard. In *International Enc. of Statistics*, eds W. H. Kruskal and J. M. Tanur, Free Press, New York, pp. 733 – 734**.**
LAPLACE, P. S. (1812). *Théorie Analytique des Probabilités*. Paris. Reprinted with supplements: *Oeuvr. Compl*, t. 7, Gauthier-Villars, Paris, 1886 and Editions Jacques Gabay, Paris, 1995.
PEARSON, K. (1978). *The History of Statistics in the 17th and 18th Centuries Against the Changing Background of Intellectual, Scientific and Religious Thought* (Lectures given at University College, London, during the academic sessions 1921 – 1933. ed. E. S. Pearson). Griffin, London.
SHEYNIN, 0. B. (1971). On the history of some statistical laws of distribution. *Biometrika* 58, pp. 234 – 236.
STIGLER, S. M. (1983). Who discovered Bayes' s theorem? *Amer. Statistician* 37, pp. 290 – 296.
--- (1986). *The History of Statistics.* Harvard University Press, Cambridge, MA.
--- (1999). *Statistics on the Table. The History of Statistical Concepts and Methods*. Harvard Univ. Press, Cambridge, MA.
ZABELL S. L. (1975), The rule of succession. *Erkenntnis* 32, No. 2/3, pp. 283 – 321.

XII

Oscar Sheynin

Mises on mathematics in Nazi Germany

Historia Scientiarum, vol. 13, No. 2, 2003, pp. 134 – 146

Abstract

I am publishing a manuscript on mathematics in Nazi Germany surely written by Mises. It is interesting in itself and the more so since Mises' political and social activities seem to be completely forgotten.

1. Introduction

1.1 The Source. I am making public a manuscript from the Richard von Mises (1883 – 1953) papers kept at Harvard University Archives (Cambridge, Mass.). It is listed in Frank et al (1964, p. 566) as consisting of 11 typed pages and written in 1934. What follows below, in § 2, was handwritten on nine pages, undated and signed *R. S.* Upon obtaining the manuscript (code HUG 4574.22.5), I asked for an explanation of the discrepancy, and here, in part, is the answer written Febr. 14, 2002, by Kyle Carey, Assistant at the Archives:

A search of Mr Mises' papers did not identify a typed manuscript … some of the unpublished manuscripts listed [by Frank et al, see above] *are not in the same form as described* [by them].

It is difficult to name any appropriate person (a mathematician as witnessed by the nature of the manuscript, and apparently Mises' likeminded colleague) who would have thus initialed his note. My conclusion is that *R. S.* is not a signature, but stands for Roh Stoff[1]. The year, 1934, as stated by Frank et al, seems correct as all the events described in the manuscript occurred either in 1933 or early in '1934. And its author is Mises himself. I have compared the handwriting of the manuscript with that of Mises' handwritten materials, a letter to Einstein, of 1919 (Bernhardt 1993, pp. 61 – 62), and two notes from the same *Archives* (code HUG 4574. 8). Although not an expert in graphology, I feel that my conclusion is correct[2].

I also indicate that Mises remained in Germany at least until the very end of 1933 (with possible short visits back m 1934). He renounced his position at Berlin University in October 1933 (see a photocopy of Mises' letter of resignation published by Bernhardt (1993, p. 60). On Dec. 21 of the same year he appealed to the Kulturministerium for a pension and asked for a speedy decision so that he *die mir* [him] *angebotene Stelle in der Turkei entgültig annehmen kann*. Siegmund-Schültze (1998, p. 72), who quoted that document, did not say, however, whether Mises had written it in Germany or in Turkey.

Then, Begehr (1998, p. 227) states that *am 1.1.1934 tritt er* [Mises] *seine neue Stelle in Istanbul*, a fact confirmed by a private communication to him (Celebi 2002).

1.2 Gumbel. Since I firmly believe that Mises is indeed the author of the manuscript in question (of a political pamphlet, as I would say), it is not amiss to add a few words about his previous attempts to help Emil Julius Gumbel (1891 – 1966), a German (later, American) statistician and active fighter against the rightist movement in his native country[3].

On Jan. 2 1932 Einstein, in a letter to an unknown person (code 50110) wrote that *nicht nur seine* [Gumbel's] *Position, sondern auch sein Leben bedroht ist*. Gumbel had indeed made many attempts to find a job outside Germany, both before and after 1932. On May 12 1928 he informed Karl Pearson (Pearson Papers, code 709) that Mises *as proposer* will apply for a fellowship for him at the European office of the then existing International Educational Board so as to secure a position at the Galton Laboratory. Both Einstein and another German statistician of Polish ancestry, born and educated in Russia, Ladislaus von Bortkiewicz (1868 – 1931), also helped Gumbel, but the plan failed.

Later, on Apr. 22, 1931, a Professor Holde, in a letter to Einstein (Code 46545), listed several individuals including Mises who were prepared to support Gumbel by a public *Erklärung*.

1.3 Differing approaches to mathematics. Hamel (1933; 1934)[5], whom Mises extensively quoted in his manuscript, had attempted to define the subject of mathematics and to stress its special role in the Third Reich. When considering his first point, he (1934, p. 12) stated that *Mathematik handelt vornehmlich vom Raum und von der Zahl* but at the same time he (p. 15) emphasized its connection with the spirit: *Mathematik als Lehre vom Geiste, vom Geiste als Tat* and declared, on p. 14, that *die Mathematik an sich rein ideell ist ...*

It is interesting to compare these statements with the thoughts formulated by Engels (1877 – 1878, p. 93) and certainly adopted in the Soviet Union:

Die reine Mathematik hat zum Gegenstand die Raumformen und Quantitätsverhältnisse der wirklichen Welt dieser Stoff in einer höchst abstrakten Form erscheint.

Kolmogorov (1977) quoted that definition and enlarged on, and actually moved away from it by describing how mathematics was becoming ever more abstract. Definitely lacking in Soviet mathematics was however the notion of *Geist als Tat* and the history of mathematics conclusively proves that quite abstract notions without any visible applications might well acquire practical importance.

Recent authors tend to underscore ever more the abstract, Platonic nature of mathematics. Thus, Bochner (1987, p. 522) stated that this science is *a realm of knowledge entirely unto itself*. Even Soviet authors had begun to differ from Engels. Almost agreeing with Bourbaki in that mathematics is a system of hierarchy or structures, Youshkevich & Rosenfeld (1972, pp. 475 – 476) diplomatically concluded:

After all, is it possible or necessary to offer a rigid and frozen definition of a science that is in a state of permanent lively development and dialectical interrelation with the entire complex of other areas of knowledge?[6]

It is instructive to note that already Cournot (1847, p. 355), as noticed by Bru (1991, p. 6), considered mathematics as a science of the abstract:

Sous le nom collectif de Mathématiques, on désigne un systéme de connaissances scientifiques, étroitement liées les unes aux autres, fondées sur des notions qui se trouvent dans tous les esprits, portant sur des vérités rigoureuses que la raison est capable de découvrir sans les secours de l'expérience, et qui néanmoins peuvent toujours se confirmer par l'expérience, dans les limites d'approximation que 1'expérience comporte.

For his part, Hamel stressed the practical importance of the science: *Geist als Tat*, see above; and, as Mises saw it (see my § 2), his main definition of mathematics: *In der Mathematik ist Denken und Handeln eins* (p. 12). This meant that applied mathematics (on which Hamel had indeed mostly dwelt) was much more important than its pure branch.

1.4 Conditions in Germany. It is hardly known whether Hamel really believed in his own words. But what was the real situation concerning pure mathematics in Germany? Gumbel (1937b) did not say that it had been forgotten. He reviewed the first two issues of *Deutsche Mathematik* published in 1936 and edited by Bieberbach and noted ideological attempts to govern mathematics made by an *ewiger Student* as well as postulates of *eine neue rassische Geschichte der Mathematik* and an attack against Einstein *von einem Heidelberger Studenten* [!]. The work of the great scholar had allegedly been *eine Kampfansage mit dem Ziel der Vernichtung des nordisch-germanischen Naturgefühls*. At the same time, however, Gumbel remarked that *jüdische, sogar emigrierte Mathematiker werden in Gemütsruhe zitiert, und vielfach wird auf ihren Arbeiten weiter gebaut*[7].

Two other authors (Pinl & Fortmüller 1973, p. 138) testified that, apparently in 1938, there existed an opposition between pure and applied mathematicians:

Modern algebra was suspect and so was modern quantum mechanics. At the same time the racialist mathematicians headed by Bieberbach were pained to see that some of their friends and allies among the Nordic experimental physicists would have liked to do away with mathematics altogether.

It is extremely difficult, however, to describe the situation comprehensively. First, the boundaries between pure and applied mathematics are fuzzy and change in time. Second (Mehrtens 1986), from the end of the 19th century onward, the climate of opinion in Germany had been varying and the emphasis on applied mathematics was not put for the first time in 1933; for example (Mehrtens 1987, p. 166), in 1926 Bieberbach, as a partisan of a down-to-earth mathematics, attacked Hilbert. Moreover (Mehrtens 1986, p. 324; 1987, p. 162), even in Nazi Germany much depended on local conditions and on the power struggle going on between different branches of the Establishment. One general and hardly original conclusion is nevertheless possible: under the new regime, science in general and mathematics in particular had essentially weakened;

however, Mises' opinion (see end of his manuscript) proved too pessimistic. Here are some figures illustrating the first part of my statement. Up to the end of 1936, 1145 professors and 539 assistants and others were dismissed from the universities, new wholly uneducated staff had been installed and students were being admitted in accord with non-intellectual criteria (Hartshorne 1937 as reviewed by Gumbel 1938b). Siegmund-Schültze (1998, p. 15) provided other figures: during 1933 – 1945, there were 980 natural scientists and technologists among the emigrés, 130 of them mathematicians. Also see Gumbel (1936; 1937a; 1938a). More to the point was that (Mehrtens 1987, p. 177) *the social system of mathematics was able to survive with the loss of some 20 – 30% of its members*". These figures were not, however, substantiated. Hanshorne (1937, p. 170) provided another estimate, also without justification: the *sheer loss* of the scientific manpower amounted to 15 – 20%.

1.5 Conditions in the Soviet Union. There was much in common in the conditions of life and work for mathematicians in Nazi Germany and Soviet scientists during the Stalinist period. Litten (1996) provided an example of a scholar's tragic life in Germany[9] and Rozhanskaia, in her Foreword to the translation of that paper, described the situation in her country:

Total shadowing, … especially of those suspected of non-conformism, the Party spirit in science, … a permanent yoke of the Party organs, the striking community of tragic fates … fright, leading to meanness and treachery.

Already in 1921, 15 academics (including Markov) from the Petrograd University declared that students should be admitted on the strength of their knowledge rather than in accord with political considerations (Grodsensky 1987, p. 137). The condemned practice persisted, however, perhaps not continuously, until the break-up of the Soviet Union. In the 1920s, vicious attacks against non-complying mathematicians were successfully launched (Ford 1999). And the notorious Luzin case is now generally known (Demidov et al 1996).

Way back, an *Appeal* (Anonymous 1923) protested against the Soviet *murderous and shameful system*, *the total lack of personal immunity*, and against *the execution, or, rather, the murder* of scientists.

I also note that the Soviet regime had for a long time been hostile to the theory of relativity. A certain Vislobokov (1952) denied it, and a high ranking Party official (and a petty mathematician) Kolman (1939) stated that velocities can exceed 300,000 *km*/*sec* since the contrary would have contradicted dialectical materialism. Both these authors published their papers in leading Party periodicals. About ten years later, Einstein was denounced as a Zionist. Novikov (1995) testified that in the 1970s the same attitude had persisted and reflected the directives of the highest organ of the Communist party; for that matter, all this could not have happened without the party's relevant decisions. The attitude with respect to Einstein, as Novikov states (see end of Note 9), was shaped in accordance with official antisemitism.

2. The Manuscript
Die Mathematik und das dritte Reich

Vielfach ist die Ansicht verbreitet, als müsste der Totalitätsanspruch des national-sozialistischen *Geistes* doch irgendwo seine Schranken finden, wenn nicht früher, so doch vor den Pforten der abstraktesten und objektivsten aller Wissenschaften, die wahrhaftig keinen Spielraum für politische Extratouren zu bieten scheint. Aber wer so denkt, rechnet nicht mit dem deutschen Urphänomen der *Maßlosigkeit*, die keinerlei Grenzen anerkennt, keine aber mit solcher Wollust überspringt wie eine, die vom guten Geschmack gezogen zu sein scheint.

Es soll im folgenden nicht etwa von personellen Maßnahmen im Sinne des famosen *Beamtengesetzes*[10] die Rede sein, durch die mindestens ein Drittel der mathematischen Produktivität Deutschlands über die Reichsgrenzen hinausgeschafft wurde. Hier ist ja nur ein gewisser Starrsinn wirksam, der zwingt, einen einmal zum Prinzip erhobenen Akt der Rohheit und Unvernunft auch dort durchzuführen, wo niemand daran zweifelt, dass er ins eigene Fleisch schneidet.

Auch die Unterordnung der Mathematik durch Rassenkunde neudeutscher Prägung interessiert uns hier weniger. Es war immerhin ein Glück, dass es dem Berliner Ordinarius Ludwig Bieberbach noch gelang, den Titel einer für das Wintersemester angekündigten Vorlesung über *Große deutsche Mathematiker* durch den Zusatz *vom Standpunkt der Rassenkunde* rechtzeitig (für den Nachtrag des Vorlesungsverzeichnisses) zu ergänzen.

Aber ganz ungetrübt blieb die Freude nicht, als er sich genötigt sah, um die Überlegenheit der arischen Rasse zu dokumentieren, mehr als hundert germanische Leistungen anzuführen, die der Schöpfung der Mengenlehre durch Georg Cantor gleichwertig sind, oder mehr als dreihundert einwandfreie Namen zu nennen, die sich mit denen Jacobis, Kroneckers, Minkowskis messen können. Dass er etwa den leichteren Ausweg gewählt haben sollte, einiges von den nicht-arischen Leistungen zu unterschlagen, wird man bei einem Mann nicht voraussetzen, der sich bis zum 30. Januar[11] nicht genug tun konnte, seine Ablehnung des Nationalsozialismus zu betonen, jetzt aber im braunen Hemdchen herumläuft, dass es nur so eine Freude ist. Doch das alles gehört ja heute zu dem *rein Menschlichen* in Deutschland und man müsste eher die Ausnahmen – die es immerhin gibt – registrieren, wenn man damit nicht die wenigen Anständigen ins Verderben brächte.

Nun aber zu dem eigentlichen Greuelmärchen, das wir heute erzählen wollen, dem von der *Geistesverbundenheit* der Mathematik mit dem Dritten Reich. Am 17. Oktober 1933, 20 Uhr, im Auditorium Maximum der von Humboldt begründeten Berliner Universität, gab der frühere Vorsitzende, jetzt *Führer* des mathematischen Reichsverbanden, der ordentlicher Professor der Mathematik Georg Hamel die Parole aus. Seine Ausführungen liegen seit kurzem in einer von ihm selbst redigierten Veröffentlichung in den *Unterrichtsblättern für Mathematik und Naturwissenschaften* vor[12]. Die *Vossische Zeitung* [Anonymous (1933)] hatte in einem unmittelbaren Bericht über die denkwürdige Versammlung unter dem Titel *Mathematik des faustischen Menschen* u.a. erwähnt[13], dass die *typischen Unendlichkeitsbegriffe* der Mathematik besonders hervorragende schöpferisch-faustische Leistungen seien. Das ist nun in der endgültigen Formulierung für alle Fälle beiseite gelassen, da sich inzwischen herausgestellt hat, dass der Schöpfer der transfiniten Zahlen mindestens zwei nicht-arische Großeltern hatte[14] dass also dieser Teil der Unendlichkeitsbegriffe vermutlich zur logisch-semitischen Seite der Mathematik gehört, die der intuitiv-arischen diametral gegenübersteht. Hamel gibt sich die größte Mühe, klar zu machen, was Mathematik eigentlich sei und leisten könne. Keine Definition könnte treffender sein als diese: *In der Mathematik ist Denken und Handeln eins*[15]. Aber man darf nichts übertreiben:

Der kritische strenge Geist der Mathematik muss und kann sich selbst die Grenzen setzen. Was einmalig ist, ist außer der Ordnung und damit frei. Und trägt selbst die Verantwortung vor sich und Gott.

Daraus folgt offenbar, dass man dem Einmaligen, das der Nationalsozialismus und sein oberster Führer darstellt, nicht mathematisch beikommen kann.

Aber das weitaus wichtigere ist der Erziehungswert, der aus der Geistesverbundenheit der Mathematik mit dem Dritten Reiche folgt. Die Grundhaltung beider ist die Heroische ... beide verlangen den Dienst; ... Beide sind antimaterialistisch ... Beide wollen Ordnung, Disziplin, beide bekämpfen das Chaos, die Willkür ... Beide sind streng, aber nicht kalt.

Offenbar ist die Strenge der semitische, die Wärme der arische Einschlag. Schließlich mündet des Ganze in den leidenschaftlichen Appell:

Neben die Lehre vom Blut und vom Boden[16] *gehört deshalb als allgemein verbindlich bis ans Ende der Erziehung die Mathematik als Lehre vom Geiste, vom Geiste als Tat*.

Diese hochtönende Schluss reuet dem Kundigen viel. Der Reichsverband kämpft seit Jahren gegen die Herabsetzung der Mathematikstunden in den Oberklassen der höheren Schulen. Während der schmachvollen vierzehn Jahre[17] hatte man andere Argumente, jetzt entdeckt man die Verwandtschaft mit *Blubo*[18], um die Stundenzahl *bis ans Ende der Erziehung*, nämlich bis Oberprima, zu bewahren. Dass ein Mann von wissenschaftlichen Bildung und einigem persönlichen Anstandsgefühl sich zu solchen Auftreten hergibt, ist nur aus der grotesken Überspannung der deutschen *Sachlichkeit* zu erklären. Wo man sich für ein als nützlich erkanntes, unpersönliches Ziel einsetzt, ist jedes Kampfmittel, folglich auch das der tiefsten geistigen Prostitution, erlaubt.

Keinen Anspruch erhebt der *Geist* des dritten Reiches mit solch unerschütterlicher Beharrlichkeit wie den, schöpferisch und schaffend[19] in bisher nie geahntem Maße zu sein. Er vermag es, mindestens sieben Welten an einem Tage zu erschaffen, so den semitischen Gott weit übertrumpfend, der, offenbar in logisch-formalistische Fesseln verstrickt, zu einer Welt sieben Tage brauchte. Da veröffentlicht ein Professor in Darmstadt, Hugo Dingler[20], Direktor des *wissenschaftstheoretischen Instituts* der technischen Hochschule, ein Buch über eines der abstraktesten Gebiete der Mathematik, die Grundlagen der Geometrie [Dingler 1933]. Der Verleger (F. Enke in Stuttgart) muss ihn gegen den Verdacht in Schutz nehmen, etwa ein unfruchtbarer Bolschewist von der Art Einsteins zu sein[21]. Das sieht so aus:

Das rein formalistisch-rechnerische Denken, welches den Kalkul nicht als vielfach nützliches Hilfsinstrument, sondern als die Sache selbst, als das Absolutum betrachtet (*Einstein u. a.*), *und das eine so starke Analogie zur sinnlosen Verabsolutierung von Organisationsformen im politischen Bolschewismus* (*auch in soziologischer und personeller Richtung*) *zeigt, wird hier in seiner vollen Unfruchtbarkeit und Hohlheit nachgewiesen und ihm gegenüber dem wirklich schaffenden Tun und schöpferischen Denken des Menschen in der Idee wieder sein volles Recht gegeben.*

Diese Verteidigung wird nicht viel helfen; die einzig zugelassene, parteiamtliche Auffassung über neue *schöpferische Werte* und über die Grundlagen der Geometrie ist inzwischen in anderer Weise festgelegt worden.

Da erfindet ein armer Musiklehrer, dem einige unverdaute Brocken von mathematischen Schulstoff Beschwerden machen, zum 999000 male eine *Quadratur des Kreises*. So etwas gibt es immer und überall und in den Zeiten der Weimarer Republik hatte der oben erwähnte Professor Bieberbach als Leiter des mathematischen Seminars der Berliner Universität eine Postkarte drucken lassen, die solche unglückliche Dilettanten schonend über die Sachlage unterrichtete. Jetzt aber lebt ein anderer Geist in Deutschland. Der schöpferische Erfinder hat es überhaupt nicht mehr nötig, für seine Anerkennung selbst zu kämpfen, die Mühe wird ihm von der *zuständigen* Stelle abgenommen. Wir lesen am 7. März in der *Kurhessischen Landeszeitung* die folgende, auch sprachlich wertvolle Verlautbarung[22]:

Durch die Kurhessische Landeszeitung, gibt der Leiter des Kampfbundes für deutsche Kultur, Max Köhler, der Öffentlichkeit folgendes Forschungsergebnis von Willi Oberle, Kassel-Niederzwehren, bekannt, das besagt, dass dieser durch die Musikgeometrie zur Lösung der Quadratur des Kreises gelangt ist:

Dem Musikforscher, … Mitarbeiter des Kampfbundes für deutsche Kultur, Landesleitung Kurhessen, ist es gelungen, aus den Ergebnissen seiner Forschungen auf dem Gebiete der Musikgeometrie die Quadratur des Kreises aufzudecken. Außerdem …

Nun folgt ein langer Bericht, unterbrochen von schreienden Zwischentiteln wie *Forschungserfolg von unaussprechlicher Bedeutung* u. ähnl.

Aber selbst im dritten Reich finden sich noch Leute, die den Mut aufbringen, gegen den äußersten Unfug, den der Schandbund der deutschen Kultur hervorbringt, aufzutreten. Ein Einsender, der sich einen Rest von Studienkenntnissen bewahrt hat, klärt in bescheidener Form darüber auf, dass zum mindesten seit den Arbeiten des Mathematikers [Ferdinand] Lindemann vom Jahre 1882 jeder Zweifel an der Unmöglichkeit einer Quadratur des Kreises mittels Zirkel und Lineal geschwunden ist. Er weist nach, dass die Oberlesche Lösung darauf hinauskommt, den Kreisumfang dem dreifachen Durchmesser des Kreises gleichzusetzen. Um seine Zuschrift aufnahmefähig zu machen, fügt er hinzu, dass die schlechteste

Näherungslösung des Problems von den Juden herrührt: in einer Beschreibung des Salomonischen Tempels wird nämlich erwähnt, dass ein Becken 10 Ellen Durchmesser und 30 Ellen Umfang habe[23], also das, was zweitausend Jahre später den höchsten Triumph nationalsozialistischer Forschung begründet. Doch mit dem Kampfbund ist nicht zu spaßen.

Schon am 10. März wird der vorlaute Einsender vernichtend geschlagen; die parteiamtlich zuständige, maßgebende Stelle spricht:

Der Landesleiter ... hat die Verurteilung der Oberleschen Erkenntnisse wahrgenommen, um sich grundsätzlich ... wie folgt auszusprechen: ... Jedes Gesetz ist stets auf intuitivem Wege, d. h. auf dem Wege einer seelisch-geistigen Schau entdeckt worden niemals, aber niemals errechnet oder konstruiert. ...

Zu Sache selbst sei kurz bemerkt, dass die noch herrschende exakte Wissenschaft liberalistischer Herkunft[24] *am allerletzten dazu berufen ist, ausschlaggebende Stellung zu nehmen zu neuen schöpferischen Werten, die heute in umgeahnter Fälle im Schoße des jungen Dritten Reiches der Auferstehung harren, weil der schöpferische Begriff nationalsozialistischer Weltanschauung die Polarität von Seele und Vernunft bedeutet, was aber bekanntlich die exakte Wissenschaft liberalistischer Prägung ablehnt, und damit sehr richtig und konsequent durch Professor Lindemann 1882 die Unmöglichkeit erklärte, durch Errechnung der Quadratur des Kreises beizukommen und damit die Lösung des Problems aus dem Aufgabenkreis der Wissenschaft liberalistischer Prägung endgültig ausstieß. Der Bankerott der liberalistischen Wissenschaft ist hiermit schon damals ganz exakt ausgesprochen worden ... Max Kohler*[25] *, Leiter des Kampfbundes für deutsche Kultur, Gau Kurhessen.*

Noch gibt es in Deutschland viele Gelehrte, selbst unter den Anhängern des Dritten Reiches, denen dieses offizielle Dokument der *Geistverbundenheit* die Schamröte ins Gesicht treibt. Vielleicht wird sogar den Führer des Reichsverbandes[26] ein Zweifel darüber beschleichen, ob es ganz das Richtige war, öffentlich zu verkünden, dass die *vollkommene Logisierung der Mathematik nicht möglich ist*, und unter Berufung auf Blut und Boden den *frei-schöpferischen Menschengeist* zur Tat aufzurufen. Aber die, die heute Professoren der Mathematik sind, haben alle noch im Zeitalter der *liberalistischen Wissenschaft* ihre Ausbildung erhalten und daraus eine Geisteshaltung bezogen, die sich nicht ganz verleugnen lässt. Vom kommenden Herbst an wird kaum noch ein Student die Universität betreten, der nicht sichere Gewähr dafür bietet, dass er den geistigen Anforderungen des *Kampfbundes für deutsche Kultur* genügt. Die Ausfüllung der Studienzeit durch Wehrsport, die Ausschaltung eines großen Teiles der besten Universitätslehrer werden das übrige tun. So kann man sicher sein, dass die Nation, die zur Zeit ihrer liberalistischen Verirrung in *Carl Friedrich Gauss* den princeps mathematicorum hervorgebracht hat, recht bald keine anderen Mathematiker besitzen wird als solche, die Mathematik in enger *Geistverbundenheit mit dem Dritten Reich* betreiben und noch ganz andere Blamagen als die *Quadratur des Kreises mittels der Musikgeometrie* in die Welt herausschreien werden". R. S.

Acknowledgements. I am grateful to the scientific bodies mentioned in Note 3 for permission to quote/publish the relevant materials. It is also my pleasant duty to thank Professors K. Dietz and J. Pfanzagl who read a preliminary version of this paper and helped me to decipher Mises's handwriting; Professor H. A. David and Frau Dr. Hannelore Bernhardt for indicating some pertinent literature and the latter also for the second possible interpretation of the letters R. S. in Note 1; Professor H. Begehr for confirming the date of Mises's departure from Germany (§ 1.1); and Frau Dr. Bärbel Schäfer, see Note 22.

Notes

1. Although the correct spelling is of course *Rohstoff*, a single word. Then, I can hardly rule out that the letters R and S indicated **R**ichard Mise**s**.

2. These notes are undated; one is in English, the other one, entitled R. Carnap. *Logische Syntax der Sprache*. [Springer,] Wien, 1934, is in German and English. The only difference in the way of writing between the document of 1919 and the manuscript concerns the capital a, but the same difference exists also between the former and the German text of the note on Carnap's book.

3. I published a booklet (Sheynin 2003) on Gumbel's impressions of the Soviet Union and based on archival sources kept at the Albert Einstein Archives, Hebrew University of Jerusalem, and on the Pearson Papers from the University College London. I quote several passages from my booklet and, as before, provide the appropriate codes.

4. Bortkiewicz was also instrumental in securing Gumbel, in 1924, a position at Heidelberg (letter of an eminent economist Lederer to Einstein of 27.1 1.1930, Code 46522).

5. Georg Hamel (1877 – 1954). Sisma (2002) described his previous sympathetic attitude towards Mises. Below, Ludwig Bieberbach (1886 – 1982) and Erhard Tornier (l894 – l982) are also mentioned. The latter was a card- carrying Nazi. In 1939 he was relieved of his post because of unbecoming behaviour (possibly caused by bad psychological health), see Hochkirchen (1998).

6. Youshkevitch (1994, p. 13) repeated that he was unable to define mathematics. He also noted that Kolmogorov had understood the insufficiency of the Engels formula and attempted to stretch it. In a milder form, the last-mentioned statement appeared much earlier (Youshkevitch 1983, p. 387).

7. I adduce the exact references lacking in Gumbel's note; their order follows my exposition. The Heidelberg student, Kubach (1936); Türing (1936), who also opposed Einstein to Kepler and Newton; Schönhardt (1936), who referred to Jewish mathematicians. In Note 23 I mention Tornier from the same source.

8. David Hilbert stated, in 1934, that, after that, the Mathematical Institute at Göttingen became non-existent, see for example Fraenkel (1967, p. 159). I did not find a reference to a witness of Hilbert's oral utterance.

9. She had not mentioned the official antisemitism in the Soviet Union, never openly acknowledged by the authorities. Describing the situation there after 1945, Novikov (1995) testified:

With stealing [of state property] and corruption occupying the leading role in the basic Party line, antisemitism had advanced to the second place there.

10. This is a reference to the law of 7.4.1933 called *Zur Wiederherstellung des Berufsbearntentums* (Hattenhauer 1993, p. 408) . On his next page the same author says that with some exceptions the law stipulate the *Versetzung in den Ruhestand für solche Beamten von die nicht arischer Abstammung waren*. Then, however, he (p. 410) added: *Eine Ausnahmeregelung für hervorragende jüdische Wissenschaftler kam dagegen entgegen ursprünglichen Plänen nicht zustande*.

The professorial staff, and, from 6.5.1933, even the Privatdozenten, *die nicht beamtet waren* (Siegmund-Schültze 1998, p. 57), was thus included in the Beamtentum and the number of scholars relieved of their position was indeed great (§ 1.3). Being a Jew, Mises, however, was not (yet) involved (he participated-in the world war), but he reasonably chose to renounce his position and emigrate to Turkey.

A tiny episode (Sheynin 2001, p. 228, based on archival sources), as though crowning the general picture, might be described. Bortkiewicz, Professor at Berlin University (cf. § 1.2), doubled at the Berlin Handelshochschule. In 1938 his portrait disappeared from the School's Hörsaal. The Secretariat suspected that it *von einem Unbefugten in der irrtümlichen Annahme, Herr von Bortkiewicz sei nicht deutschblutig gewesen, entfernt worden ist*. A lame explanation!

11. On Jan. 30 1933 Hitler came to power.

12. Hamel (1933). Hamel (1934), without reference to 1933, is its somewhat expanded version, an *ausführliche Wiedergabe des Vortrags … auf der Berliner Kundgebung des Deutschen Vereins zur Förderung des mathematischen und naturwissenschaftlichen Unterrichts*, as stated in a note on its p. 10. The essence of the changes made in the previous version becomes obvious when comparing the titles of both papers with each other.

13. Both Hamel (1934, p. 11) and the anonymous author mentioned Spengler' s *Mathematik des Faustischen Menschen*. Oswald Spengler (1880 – 1936) enumerated eight *cultures*, and among them the Faustian, or Western European culture. He was close to Nazism, and the new regime adopted his ideas, but he refused to collaborate with it (Averintsev 1982). Petropoulus (2000, p. 4) stated that a Faustian bargain was something immoral or amoral leading to self-advancement.

14. The first *grandfather* was obviously Georg Cantor. Later appropriate scientists of the 19th century were perhaps E. Borel (in 1895) and R. L. Blaire (in 1898), see Medvedev (1965, p. 127).

15. This quotation as well as a few others below are from Hamel (1933). I mention some passages from the same source in § 1.3.

16. The Nazi regime inaugurated *a new farm program accompanied by much sentimental propaganda about Blut und Boden (Blood and Soil) and the peasant's being the salt of the earth and the chief hope of the Third Reich* (Shirer 1990, p. 257). The slogan *Blut und Boden* first appeared in 1930 (Eidenbenz 1993, p. 3).

17. The years of the Weimar Republic: after World War I and until 1933.

18. *Blubo* was apparently an abbreviation of *Blut und Boden*, see Note 16.

19. Unlike the previous underscored expressions, this one is not in inverted commas, and neither had it, or the *frei-schöpferischen Menschengeist* (below), occurred in Hamel (1933).

20. Incidentally, he was the author of a study of the history of Jewish culture (1919). Already in 1911 he published his *Grundlagen der angewandten Geometrie* (mentioned on p. 144 in 1919).

21. For a Nazi follower (real or sham), Bolshevism, twice mentioned, being a dictatorship of a single party, and even of its leaders, should not have been at all meaningless. On the contrary, for such authors the adjective *fruitless*, when applied to Einstein, was quite proper.

22. Mises had not provided the appropriate year. Its knowledge is important for dating his manuscript and I have attempted to ascertain it. Here are my findings.

a) Until November 1933 the newspaper *Kurhessische Landeszeitung* was called *Hessische Volkswacht*.

b) Its issues for 7 – 10 March 1933 do not carry the materials described by Mises.

c) The issues for the same dates (7 – 10 March) of 1934 are not extant in any library.

d) Again, the appropriate dates of the *Kurhessische Landeszei tung* (subtitle: *Hessische Volkswacht*) for 1935 do not contain the pertinent materials.

I conclude that Mises had in mind the year 1934. For items a) and c) I am indebted to Frau Dr. Bärbel Schäfer, Universitätsbibliothek Marburg (her letter of 21.10.2002). I also note that, since Mises had mentioned only the year 1933 (in the beginning of his manuscript), he thus mistakenly led his readers to believe that all the events he described had occurred during that year.

23. Loewy (1935, p. 224) provided the exact references (3Kings 7:23 and 2Chr 4:2). He (pp. 224 – 225) argued that what was actually meant there was the ratio of the diameter to the side of an inscribed regular hexagon.

24. *Wissenschaft liberalistischer Herkunft* or *Prägung*, or (below) *liberalistische Wissenschaft*: I note that Tornier (1936) denounced the *jüdisch-liberalistische Vernebelung* of mathematics. He applied the same adjective three times more connecting it with *Denken, These*, and *Illusionstechnik*.

25. Above, Mises called him Köhler.
26. That is, Hamel, see above.
27. A similar episode occurred in the Soviet Union about 1928. Groman, a leading official, attempted to predict the yield of cereals, which he assumed random, given its previous values. This was not better than squaring the circle, but in 1929 his prediction somehow came true. Then, in 1930, it failed, and, unlike the Musiklehrer, he perished after being arrested. See Sheynin (1998, p. 533, Note 4).

References

Abbreviation: (R) = In Russian

Anonymous: "An appeal to the scientists of all countries and to the entire civilized world". *Trudy Russkikh Uchenykh za granitsei* [Proc. Russ. Scientists abroad], vol. 2. Slowo, Berlin, 1923, p. 340. (R)
Anonymous: "Mathematik des faustischen Menschen". *Vossische Zeitung,* Abendausgabe. 18. Okt. 1933. No paging (p. 9)
Averintsev, S. S.: "Spengler". *Great Soviet Enc.*, vol. 29, 1982, p. 672. This source is a translation of the third edition of *Bolshaia Savetskaia Enziklopedia* whose vol. 29 appeared in 1978.
Begehr, H.: "Mathematik in der Akademie der Wissenschaften und in den Universitäten in Berlin". In *Mathematik in Berlin*, Hlbbd 1. Hrsg. H. Begehr. Aachen, 1998, pp. 3 – 449.
Bernhardt, Hannelore: "Skizzen zu Leben und Werk von Richard Mises". In collected articles *Österreichische Mathematik und Physik*. Hrsg. Zentralbibl. für Physik in Wien. Wien, 1993, pp. 51 – 62.
Bochner, S.: "Mathematics". In *McGraw – Hill Enc. of Science and Technology*, vol. 10. New York, 1987, pp. 522 – 527.
Bru, B.: "A la recherche de la démonstration perdue de Bienaymé". *Mathématiques, informatique et sciences humaines*, 29e année, No. 114, 1991, pp. 5 – 17.
Celebi. Okay: Private communication to Prof. Heinrich Begehr, 15 Dec. 2002. In a second communication dated 20 Dec. 2002 he allowed me to refer to him. Celebi obtained his information at Istanbul University.
Cournot, A. A.: "De l'origine et des limites de la correspondance entre l'algébre et la géométrie". Paris, 1989. First edition 1847.
Demidov, S. S., Ford, Ch. E.: "N. N. Luzin and the affair of the *National Fascist Centre"*. In *History of Mathematics: States of the Art*. J. W. Dauben et al, Editors. San Diego, 1996, pp. 137 – 148.
Dingler, H.: *Die Kultur der Juden* Leipzig, 1919.
---: *Die Grundlagen der Geometrie*. Stuttgart, 1933.
Eidenbenz, M.: *Blut und Boden*. Bern, 1993.
Engels, Fr.: "Herrn Eugen Dührings Umwältzung der Wissenschaft". In *Marx – Engels Gesamtausgabe*, Abt. 1, unnumbered volume containing this contribution (pp. 33 – 335) and *Dialektik der Natur*. Glasshütten im Taunus, 1970. First edition 1935.
Ford C. E.: "The Great Change at the Moscow Mathematical Front". *Istoriko-Matematicheskie lssledovania*, vol. 3 (38), 1999, pp. 74 – 92. (R)
Frank, Ph., Goldstein, S., Kac, M., Prager, W., Szegö, G., Editors: *Selected Papers of Richard von Mises*, vol. 2. Providence, Rhode Island, 1964.
Fraenkel, Abr Ad.: *Lebensreise. Aus der Erinnerungen eines jüdischen Mathematikers*. Stuttgart, 1967. This is the main source for studying my subject. I discovered it too late.
Grodzensky, S.Ya.: *Markov.* Moscow, 1987. (R)
Gumbel, E. J.: Die Gleichschaltung der Universität Heidelberg (1936). Reprinted: Vogt (1991, pp. 207 – 217).
---: "University of Heidelberg and New Conceptions of Science". *Nature*, vol. 139, 1937a, pp. 98 – 100. Appeared anonymously.
---: "Arische Mathematik"(l937b). Reprinted in Vogt (1991, pp. 218 – 221).
---: "Die Gleichschaltung der deutschen Hochschulen". In *Freie Wissenschaft*. Hrsg. E. J. Gumbel. Strassbourg, 1938a, pp. 9 – 28.
---: Review of Hanshorne (1937). *Annals Amer. Acad. Political and Soc. Sci.*, vol. 200, 1938b, p. 307.

Hamel, G.: “Die Mathematik im Dritten Reich”. *Unterrichtssblätter f. Math. u. Naturwissenschaften*, 39. Jg.. 1933, pp. 306 – 309.
---: “Die Mathematik im Dienste des Dritten Reiches”. *Z. f. math. u. naturwissenschaftlichen Unterricht aller Schulgattungen*, 65. Jg., 1934, pp. 10 – 15.
Hanshome, Ed. Ya. Jr: *The German Universities and National Socialism* (1937). New York, 1981.
Hattenhauer, H.: *Geschichte des deutschen Beamtentums*. Köln, 1993. Erste Ausg., 1980.
Hochkirchen. Th.: “Wahrscheinlichkeitsrechnung im Spannungsfeld von Mass- und Häufigkeitstheorie – Leben und Werk des *Deutschen* Mathematikers Erhard Tomier“. NTM. *Intern. Z. f. Geschichte und Ethik der Naturwissenschaften, Technik und Medizin*, N. F., Bd. 6, 1998, pp. 22 – 41.
Kolman, E.: “The Relativity Theory and Dialectic Materialism”. *Pod Znamenem Marksisma*, No. 10, 1939, pp. 129 – 145. (R)
Kolmogorov, A. N.: “Mathematics”. *Great Soviet Enc*., third edition, vol. 15, 1977, pp. 573 – 585. This source is a translation of the same volume of the *Bolshaia Sovetskaia Enz*. (vol. 15, 1974). The article was not written by, but based on Kolmogorov’s contribution of the same title from the second edition of the *Enz.* (vol. 26, 1954, pp. 464 – 484).
Kubach, F.: “Studenten, in Front!” *Deutsche Math.*, Bd. 1, 1936, pp. 5 – 8.
Litten, F.: “Ernst Mohr – Das Schicksal eines Mathematikers”. *Jahresbericht Deutschen Mathematiker-Vereinigung*, Jg. 98, 1996, pp. 192 – 212. Russian translation by M. M. Rozhanskaia: *Istoriko-Matematicheskie Issledovania*, vol. 3 (38), 1999, pp. 221 – 248.
Loewy, A.: “Zur Mathematik in Bibel und Talmud”. *Monatsschrift für Geschichte und Wissenschaft des Judentums*, N. F., Bd. 79, 1935, pp. 224 – 238.
Medvedev, F. A.: *Razvitie Teorii Mnozhestv v 19m Veke* (Development of the Set Theory in the 19th Century). Moscow, 1965. (R)
Mehrtens, H.: “Angewandte Mathematik und Anwendungen der Mathematik im nationalsozialistischen Deutschland”. *Geschichte und Gesellschaft*, Bd. 12, 1986, pp. 317 – 347.
---: “The social system of mathematics and National Socialism: a survey”. *Sociological Inquiry*, vol. 57, 1987, pp. 159 – 182.
Novikov, S. P. “The Academy’s mathematicians and physicists of the 1960s – 1980s”. *Voprosy Istorii Estestvoznania i Tekhniki*, No. 4, 1995, pp. 55 – 65. (R)
S, G, 78
Petropoulus, J.: *The Faustian Bargain. The Art World in Nazi Germany*. Oxford, 2000.
Pinl, M., Furtmüller, L.: “Mathematics under Hitler”. *Publ. of the Leo Baeck Inst.*, vol. 18, 1973, pp. 129 – 182.
Schönhardt, C.: “Alexander von Brill”. *Deutsche Math.*, Bd. 1, 1936, pp. 17 – 22.
Sheynin, O.: “Statistics in the Soviet epoch”. *Jahrbücher f. Nationalökonomie u. Statistik*, Bd. 217, 1998, pp. 529 – 549.
---: “Anderson’s forgotten obituary of von Bortkiewicz”. Ibidem, Bd. 221, 2001, pp. 226 – 236.
---: *Gumbel, Einstein and Russia*. Moscow, 2003. In English and Russian. **S, G,** 12
Shirer, W. L.: *The Rise and the Fall of the Third Reich*. New York, 1990. First edition 1959.
Siegmund-Schültze, R: *Mathematiker auf der Flucht vor Hitler*. Wiesbaden, 1998.
Sisma, P.: “Georg Hamel and Richard von Mises in Brno". *Hist. Math.*, vol. 29, 2002, pp. 176 – 192.
Tornier, E.: “Mathematik oder Jongleur mit Definitionen”. *Deutsche Math.*, Bd. 1, 1936, pp. 8 – 9.
Türing, B.: “Deutscher Geist in der exakten Naturwissenschaft". *Deutsche Math.*, Bd. 1, 1936, pp. 10 – 11.
Vislobokov, A.: “Against modern *energytism*, a variety of *physical idealism*. *Bolshevik*, No. 6, 1952, pp. 43 – 54. (R)
Vogt, Annette: *E. J. Gumbel: Auf der Suche nach Wahrheit. Ausgew. Schriften*. Dietz, Berlin, 1991.
Youshkevich, A. P.: “Kolmogorov: historian and philosopher of mathematics. On the occasion of his 80th birthday". *Hist. Math.*, vol. 10, 1983, pp. 383 – 395.
---: “Kolmogorov on the nature of mathematics and its division into periods".

Istoriko-Matematicheskie Issledovania, vol. 35, 1994, pp. 8 – 16. (R)
Youshkevitch, A. P., Rosenfcld, B. A.: "Conclusion". In Youshkevitch, A. P., Editor: *Istoria Matematiki s Drevneishikh Vremen do Nachala I9go Stoletia* (History of Mathematics from Most Ancient Times to Beginning of 19th Century), vol. 3. Nauka, Moscow, 1972, pp. 472 – 476. (R)

XIII

A.Ya. Khinchin

R. Mises' frequentist theory and the modern concepts of the theory of probability[1]

Voprosy Filosopfii, No. 1, 1961, pp. 92 – 102, No. 23, pp. 77 – 89
Posthumous paper published by B. V. Gnedenko
Translated by Oscar Sheynin, notes by R. Siegmund-Schültze

Foreword by B. V. Gnedenko

The paper presented here was written as early as sometime between 1939 and 1944[2] by the eminent mathematician Aleksandr Yakovlevich Khinchin, who is well known for his contributions to the theory of probability, statistical physics, number theory, and theory of functions. For reasons unknown to me, it remained unpublished, although I remember that Khinchin had submitted it to the periodical *Uspekhi matematicheskikh nauk*.[3] After he died, while I was putting in order his scientific and literary heritage, I recalled this work and began looking for it. Regrettably, I was unable to find any copies of a final version and the editorial office of *Uspekhi* did not have any record of the article. So I decided to make use of a copy that had been retyped in 1946 by my students, E. L. Rvacheva and D. G. Meyzler,[4] even though it had some lacunae. I am convinced that even in this state, Khinchin's work is of considerable interest.

Indeed, year after year, methods relating to the theory of probability are gaining importance in various fields of knowledge. Therefore, ascertaining the nature of random events, and discovering an approach for defining the fundamental concept, the probability of a random event, is an important matter. The concepts of Mises, outwardly attractive and convincing at first sight, continue to find many supporters, especially among members of schools of thought outside mathematical research.

Therefore, a logical and philosophical analysis of these ideas should still be considered topical today. This is especially true since Mises' original methodological viewpoints are all retained without modification in the comparatively recent (already posthumous) English edition of his well-known book, *Probability, Statistics, and Truth* (London 1957).

The publication of Khinchin's paper shows once again the need for Soviet mathematicians and philosophers to work out and develop their opinions on the nature of probability, and on the interrelation between the theory of probability and the world of real phenomena. The time for this is ripe, and Khinchin's paper can serve as an appropriate starting point for such a debate. In fact, when it was read aloud and discussed at two joint meetings of the philosophical seminar and the seminar on the history of mathematics at Moscow State University, it was decided unanimously that the paper should be published. I have just added a few notes in order to facilitate its reading; in addition, I adduce a short description of Khinchin's work in the field of the modern theory of probability.

Aleksandr Yakovlevich Khinchin (1894 – 1959) is deservedly considered one of the founders of modern probability theory. His name is connected with the important period when the set-theoretic approach to constructing the foundations of probability theory was developed. He is also remembered for establishing the basis for the general theory of stationary random processes[5]. These subjects were part of a wider programme that he conceived for ascertaining the role of statistical regularities in various fields of mathematics and the natural sciences. This methodological viewpoint stimulated his research in the metric theory of numbers, in statistical physics, in the summing of independent random variables and queueing theory (in these [two] last-mentioned fields he was directly connected with practical telephony) and, during the last period of his life, in problems concerning the transmission of information.

It was of course impossible to carry out this program without discussing problems of a philosophical nature; the logic of the development of science inevitably compelled this great scientist to investigate them. And, indeed, Khinchin invariably turned to the philosophical interpretation of the central problems of probability theory and its applications to the natural sciences. He discussed this subject in special philosophical articles and in monographs devoted to solving concrete mathematical problems, as well as in his lectures to students and his talks in seminars on methodology.

Khinchin was among those Soviet scientists who take a systematic interest in the methodology of their science and stand firmly on the ground of dialectical materialism while solving burning philosophical problems of modern natural science. He repeatedly expounded his viewpoint in print and also in talks and in seminars on methodology. L. E. Maistrov[6] reminded me of one of these talks, which Khinchin gave in 1951 at the methodological seminar of the V. A. Steklov Mathematical Institute, where he sharply criticized idealistic conceptions in the modern theory of probability. In Khinchin's papers a notebook is preserved with the handwritten title *On some idealistic tendencies in probability theory*, where he registered his fragmentary ideas on that topic.

For more than forty years, Khinchin was connected with Moscow University as a student, post-graduate, and professor. In 1939 he was elected Corresponding Member of the Academy of Sciences of the USSR. Since 1944, when the Academy of Pedagogic Sciences of the Russian Federation was established, he was a full member and acted as a member of its presidium.

The main text[7]

Introduction

About twenty years ago, the German scholar Richard von Mises began criticizing the generally accepted foundations of the theory of probability.[8] His sharp criticism appeared consistently in the pages of mathematical and philosophical periodicals. At the same time he offered his own, new basis for this science, which he still tirelessly continues to advocate in mathematical and philosophical articles, in popular monographs,[9] and in his well-known course on the theory of probability.[10] His so-called *frequentist* theory of probability found a large number of followers among mathematicians and especially among representatives of natural and applied sciences, in particular, physicists.

The literature devoted to the problems of probability theory is immense and discussions are heated; in a word, the frequentist theory became so essential for the life of the modern doctrine of probability that every representative of our science feels obliged to adopt some definite stand with respect to it. It is already becoming impossible to ignore or disregard the frequentist theory, or brush it aside.

Nevertheless, we still do not know of any attempt to analyse critically and exhaustively the role of his theory in the historical development of our science and its place within an array of different attempts at founding the theory of probability.[11]

It is typical that the most authoritative proponents of the modern probabilistic ideas (Bernstein and Kolmogorov in the USSR, Borel, Lévy and Fréchet in France, Cantelli and Finetti in Italy, Cramér in Sweden, and others) have hardly ever expressed their views on this problem.[12] The overwhelming majority of the relevant statements are made either by second-rate specialists[13] or by philosophers and physicists. In addition, these statements are almost always of a rather particular kind: they discuss this or that isolated feature, criticize a specific proposition, and suggest this or that change or improvement. A scornful, and almost ironic attitude prevails among mathematicians with respect to the frequentist theory. In private talks you will almost always hear that, undoubtedly, not everything is in order with it, that it suffers from incurable logical flaws, and that, from the mathematical point of view, it therefore cannot even be seriously approached. A physicist will usually object by stating that, even if this were true, it is the business and the duty of mathematicians to remove these formal defects; and that it is inadmissible to reject on principle, because of its inherent temporary and purely formal imperfections, this theory which so brilliantly conforms to the essence and requirements of scientific practice.

However, the broader questions, such as, whether the frequentist theory truly meets the requirements of the applied sciences better than all the other theories? and if so, why? or what are the theory's formal imperfections? These wider questions have never been considered in sufficient generality and completeness. Almost always the discussion has been restricted in a polemical way to some isolated features. It is interesting to note that in this polemic, the so-called classical theory of probability based on considering equally possible cases, against which the main accusations made by the adepts of the frequentist theory and Mises himself were and are directed, was hardly defended by anybody. Were it not that the classical conception still occupies a prominent place in textbook literature lagging considerably behind modern science, it could have been said that Mises, who even in 1936 continued his vigorous attacks against it,[14] had in essence been forcing an open door.

We must concede that in the current fight between the partisans of the frequentist theory and its opponents, the former have one obvious advantage: they at least are proposing something positive,[15] whereas their opponents, while pouncing on these proposals and noting the single shortcomings thereof, are as a rule unable to put forth suggestions of their own. How did this situation arise? We believe that until now the cause was that the development of the theory of probability has been considerably lagging behind other mathematical sciences. As a rule, a mathematical discipline begins an analysis of its foundations only after it has reached a certain scientific maturity. For probability theory, this moment occurred only in the most recent years; only lately has the theory firmly realized its place among the other disciplines of mathematics, its specific features, and its own scientific procedure. And only from the summit that it has presently attained can the broad and distinct picture of the development and the modern state of the theory of probability be grasped. Only now within this picture can the status of the frequentist theory be perceived with full clarity. Thus it becomes understandable why ten, or even five years ago, the most thoughtful representatives of our science had abstained from expounding in detail their views on the problems posed around the frequentist theory, and why the opponents of this doctrine were unable to counterbalance it with some sufficiently substantiated positive suggestions.

In particular, the author of these lines has to admit that in the light of our present-day level of the doctrine of probability, his critical paper written several years ago[16] and devoted to the frequentist theory, should be judged as unsatisfactory[17] even though some of the propositions defended there remain true.

The considerations put forth above compel us to believe that the time has come for a comprehensive critical elucidation both of the role of the frequentist theory and of its place in science. Nowadays the theory of probability is already shaped to such an extent, and has assimilated its logical grounds to such a degree, that this critical interpretation may be undertaken not from the subjective point of view of some scholar, as it was necessarily done several years ago, but from the objective and principled position firmly secured by our science during these last years. Our paper is mostly devoted to this purpose; our critical analysis is all the more topical since exactly this year the founder of the firequentist theory, Richard von Mises, has published a detailed survey[18] of the main objections and additions to it along with an exposition of his own thorough answers to them.

1. The Merits of the frequentist theory

We begin by ascertaining some of the most important merits of the frequentist theory not in order to follow the traditional maxim, *First the achievements, then the shortcomings,* but solely because it is by way of those merits easiest to acquaint the reader with the historical situation in which this theory was set up and developed.

In the first of his works, Mises some twenty years ago raised the alarm in connection with the scandalously unhappy situation concerning the foundations of the theory of probability. This single fact already constitutes an historical merit of such importance that one can forgive much. In those days the old system of principles, adopted without changes from Laplace, completely dominated all handbooks and treatises on probability; there is no doubt that its unfitness, its incompatibility with the level to which the mathematical science had arisen since Laplace's lifetime, were very visible to every thinking researcher. It was even possible to encounter, now and then, isolated and incidental pertinent statements; and yet, each author, while beginning his treatise, invariably spoke about equally possible and favourable cases, thereafter attempted to leave this annoying subject as soon as possible and to pass on to the subsequent calm course of the theory concealing no more reefs. In those days, the theory of probability instinctively turned its back on and avoided an unbiased and relentlessly critical revision of its foundations. Such a situation was conditioned, as it always is, by the theory's insufficient [logical perfection].[19] On the other hand, however, it became ever less tolerable in the face of all the other mathematical sciences that persistently worked on revising and rebuilding their foundations.

Of course, one could nevertheless not content oneself with merely disclosing the trouble indicated above, even by raising a vigorous and permanent alarm. And Mises, even in his first articles, did not restrict himself in this way. In order to establish a firm basis for the theory of probability, i.e. to lay a foundation that would make it a worthy member of the family of modern mathematical disciplines, it was first necessary to describe with absolute clarity all the shortcomings of the existing system for founding probability, and to show, convincingly, excluding any doubt, that no satisfactory foundations could be construed along the old lines. In a series of investigations, Mises accomplished this task with an exhaustive completeness, and here lies the second substantial merit of his doctrine. Many authors had indicated, even before Mises did, that the definition of probability by means of equally possible (i.e. in essence, equally probable) cases, amounted to a certain extent to an empty tautology. But, as Mises absolutely correctly states, this is the most harmless among the sins of the classical idea; moreover, in a certain sense, it can even be justified.[20] We may consider this definition as a reduction of the problem of finding a quantitative measure of probability in the general case, to a preceding notion of equiprobability of events; the vicious circle thus disappears and the definition itself acquires some scientific meaning.

Mises was the first to reveal, systematically and convincingly, the more essential flaws in the classical notion of probability, which are much more difficult to get rid of. The first, and the main one of these defects is the extremely restricted sphere of application. Having originated and been developed due to games of chance and simplest insurance operations, the old theory of probability had built for itself a basis fit to a certain degree for treating these simplest problems; however, once the sphere of its problems had extended in connection with the requirements of the physical and the social statistics, and later on, of biology and technology, the initially adopted foundations became too narrow. In problems reaching beyond the realm of games of chance, those equally possible cases, without which the classical concept cannot even speak about probabilities, just do not exist. Mises' celebrated example of an irregular die is unsurpassed in validity and simplicity of argumentation.

Moreover, Mises was the first to show, systematically and convincingly, the inaptitude of the classical basis for forecasting the real course of phenomena. He absolutely correctly indicates that, without new special assumptions, the conclusions of the theory of probability, built on the classical definition of its main notions, offer in essence no grounds for any, even the most unpretentious opinion about how the pertinent real processes should go on. Imagine for example twenty throws of a regular coin. For conclusions concerning the real course of this process, an additional definition of the concept of equipossibility, connecting it with experiment, is required, but the classical idea of probability is unable to provide it. Consequently, new principles, not following with logical rigour from the definitions of the main concepts, are needed for applying the classical theory practically. We see thus that the author of the frequentist theory has the important historical merit of consistently exposing a number of fundamental defects of, and unsubstantiated claims made by the classical doctrine of probability. Along with this negative, so to say, merit, it is necessary to credit him now for his most important constructive achievement. No matter how we assess the frequentist theory and its future possibilities, we must admit that exactly its main principles reflected, for the first time ever, the opinion playing a basic role in the modern probabilistic outlook: the opinion, namely, that the theory of probability is a doctrine of mass phenomena. Of course, it was well known even before Mises that mass processes constitute the sphere of application of probability; yet, the demand that this feature of probabilistic teachings be represented already in their first principles, that all their formalism should be set forth in this spirit, – this demand was first put forward and carried out only within the frequentist theory.[21]

Being abstract, any mathematical theory must of necessity draw itself away from some properties of the studied objects. However, the classical conception of the theory of probability attempted to segregate its main principles not from statistical populations or recurring processes, which represent the true subject matter of its study, but from the specific properties of isolated objects participating in these processes. On the contrary, modern ideas recognize the very notion of probability as meaningful only in connection with mass phenomena, and therefore consider it desirable that this feature play the main part also and already in laying down the foundations of probability theory. As we have already remarked, this demand was first distinctly formulated and carried out precisely within the frequentist theory.

When speaking about probability, the physicist, biologist, technician or social statistician invariably has in mind some relative frequency. Furthermore, even a mathematician, during those special moments of his work when, interrupting the chain of formal deductions, he is compelled to turn his intuition to the material content of his concepts, – even he is in most cases apt to imagine each probability exactly as relative frequency. This does not at all mean that probability as a concept of a mathematical theory should include in itself all the totality of the properties and features peculiar to real frequencies; not even the frequentist theory does so. It only means that the theory of probability ought to be a sufficiently precise, formal image (hence an image gained by way of abstraction) of that structure, of those possibilities that take place in the world of real frequencies.

It is precisely this thesis whose necessity is recognized by all modern probabilistic schools regardless of their specific orientation that also constitutes the foundation of the frequentist theory. It was first formulated by this theory and we ought to recognize this fact as one of the theory's most important merits. We shall see below that Mises' doctrine considerably diverged from the main direction of development of the modern theory of probability in its opinion on an expedient realization of the demand contained in that thesis, i.e., on how and to what extent the abstraction and formalization should be carried out. This, however, should not obscure the cardinal fact that the credit both of formulating this demand and of being the first in attempting to carry it out, doubtless completely belongs to the frequentist theory.

2. A Natural-scientific or a mathematical discipline?

The fundamental divergent paths that exist between the frequentist theory and the predominant direction of the modern theory of probability are very deeply rooted. In the first place, they are caused by the unremovable paths in the opinions concerning the theory of probability as a scientific discipline. We must attentively dwell on these differences because, without completely ascertaining them, the role or the situation of the frequentist theory in the modern doctrines of probability cannot at all be determined with sufficient clarity.

For Mises' school, the theory of probability is a natural-scientific discipline[22] which widely uses mathematical methods. Mises persistently defends this thesis against two other fundamental viewpoints that differ absolutely from one another. On the one hand, being in a philosophical sense a consistent positivist of a Machian[23] persuasion, he struggles against the aprioristic, metaphysical approach of the classical theory. This campaign, no matter how interesting and instructive it was in itself, cannot be elucidated in our study since it was directed against the past, whereas we are concerned, first and foremost, with the distinction between the frequentist theory and modern progressive ideas. The main antagonism can be formulated quite simply. Contrary to Mises' main thesis, according to which the theory of probability is a discipline of the natural sciences, the modern theory has defined itself as a branch of mathematics. This basic discord determines in a decisive and, moreover, in an almost exhaustive manner all the subsequent concrete discrepancies, which we discuss below. Let us pay attention to this point.

As we see, the main criterion by which to distinguish a natural-scientific from a mathematical discipline consists in the typical manner of defining the respective field of research. Each natural-scientific discipline is determined by the material specificity of its subject matter, by the real features of the studied domain of the existing world. It is in this way only that physics, biology, and psychology define their subject matter. One and the same subject matter can be studied by very different methods, mathematical ones included; however, when passing from one method to another one, we always remain within the boundaries of the given (natural-scientific) discipline since its main specific feature is its subject matter as it exists in reality rather than the method of its study. (Examples: phenomenological and statistical thermodynamics [remaining within] the mechanical theory of heat; the corpuscular and the electromagnetic theories of light, the quantum theory of light [are all] theories of light.)

In contradistinction, the determining indicator of any mathematical discipline is always some formal method potentially admitting most various material realizations and, consequently, also practical applications. Whether or not some subject matter, some phenomenon of the world in reality can be studied by means of a given mathematical method, is determined not by the concrete material nature of this subject matter or this phenomenon, but exclusively by their formal structural properties and, above all, by those quantitative relations and spatial forms (Engels)[24] in which they live or proceed. (One example: the method of differential equations in physics, chemistry, and biology; because for its applicability it is sufficient that there exist two continuously changing magnitudes, whose changes have a definite relative velocity.)

To which class of scientific disciplines should we then ascribe the theory of probability according to the criterion just described? What constitutes the basis of the unity of its method – the material, or the formal structural properties of the subject matter it studies? It suffices merely to pose the question in this manner in order to perceive with full clarity the one possible answer. The theory of probability is a doctrine of mass phenomena. Its methods are applied if, and only if, a large number of more or less equivalent ingredients participate in an actual phenomenon; its main concept is the relative number of those components which possess one or another given attribute. What is the material content of the studied phenomenon? or the real nature of these ingredients? or the nature of the attribute applied for classifying them? None of these questions hear any relation to the judgment about whether a given process is within the capacity of the theory of probability.

Defined by the formal features of the aspect of reality being studied, the theory of probability can be, on the basis of our main criterion expounded above, only a mathematical theory, but by no means a natural-scientific one. For all the simplicity and cogency of this conclusion, we were led to it merely by the considerable development of the theory of probability during the latest years, allowing it nowadays to appear as an actually shaped mathematical discipline. How then does the author of the frequentist theory substantiate his persistent assertion that the theory is, and must be, a branch of the natural sciences, an assertion that in this context links his doctrine to the old metaphysical idea? He often formulates that view without substantiating it by arguments and apparently does not even suppose that there exist other points of view.

On those occasions when he has to defend his outlook against the opinions described by us, Mises' arguments become quite primitive, not to say naive. He argues that each probabilistic problem relates certainly to some real repetitive process or some real mass phenomena; that the theory of probability is consequently a science addressing reality and hence a natural-scientific discipline. He does not want to notice that, on the basis of such an indicator, any mathematical discipline, e.g. the theory of differential equations, may be ascribed to the natural sciences; however, it would not have crossed anyone's mind to call into question the proposition that this theory is a typical branch of mathematics.

The cause of Mises' delusion is his Machian, and therefore, idealistic philosophical standpoint that engendered and still nourishes the foundations of the frequentist theory. An idealist, keeping, as Mises and his school do, to a positivistic stand, is always in a state of fear of mathematics[25] no matter how he extols its merits in words. For him, giving away some doctrine to mathematics always means its alienation from actual contact with reality. He does not want to, and cannot, admit that mathematics and the natural sciences alike study the real world, and only it, although mathematics studies its other aspects, and by means of other methods. This is why he labels as nihilists[26] those who want to perceive a mathematical doctrine in the theory of probability; this is why, filled with disgust and horror-stricken, he struggles against the proposition, accepted without hesitation by all advanced scientists of our day, that the theory of probability is a part of the general doctrine of functions and sets.

Thus, he says, *the theory of probability is by no means a part of set theory; it is the theory of certain observable phenomena … and merely makes use of theorems put forth by set theory*.[27] And in another passage: *To a logical mind this identification of two things belonging to different categories, this confusion of task and tool is something quite unbearable.*[28]

From our point of view, the theory of probability as a science of mass phenomena can be linked only to set theory as doctrine about totalities of the most general kind. All concrete mathematical facts constituting the theory of probability retain their significance irrespective of which of the two viewpoints described above is chosen; however, with respect to the problem that especially interests us here, the problem of establishing the foundations of the theory of probability, the indicated disagreement, as we shall see now, is of crucial importance.

3. Idealization and formalization

In order to construct the basis of a natural-scientific or of a mathematical discipline, a certain idealization of the subject matter to be studied is necessary. Within the boundaries of one and the same natural-scientific domain, the degrees of idealization can be quite different, depending on the adopted method of investigation. Thus, in the theory of turbulent motion of liquids we may idealize the investigated subject matter in different ways depending on whether we intend to study it by means of classical or statistical methods.

In exact conformity with this tradition, Mises, who recognizes the doctrine of probability as a natural-scientific discipline, distinctly indicates his proposed idealization of the subject which the theory of probability is called upon to study; he believes the problem of establishing the foundation of probability theory thus to be solved. The idealization he proposes is well known and extremely simple. Keeping to it, we ought to imagine the entire totality of the ingredients of a mass phenomenon or of a repetitive process as some infinite sequence of trials or observations called collective.[29] Then, two main properties are ascribed to each of these:

1) *The existence of the limits of relative frequencies* of those elements of the collective that possess one or another of some definite group of attributes.

2) The so-called *irregularity*, i.e., the invariance of the just mentioned limits with respect to the selection, according to any definite law, of some subsequence from a given collective; the law of selection, furthermore, should not rest upon the distinction between the elements of the collective concerning their relation to the studied attribute.[30] The limits of the relative frequencies are called probabilities of the respective attributes.

Thus, in this case, the idealization consists in that we have, first, to replace a real statistical population, of necessity finite, by some infinite series. Second, we have to attribute to this series two properties that a real population cannot possess, because, anyway, they have a definite meaning only for infinite series. The construction of the foundations for the frequentist theory is now concluded, and it is now possible to begin posing and solving concrete problems and establishing general regularities.

If, however, one considers the theory of probability to be a mathematical discipline, as we have done, then the construction of its foundations cannot come to a stop here. A modern mathematical discipline requires for its basis more than a mere idealization of a real subject matter. It demands complete formalization, or, what amounts to the same thing, an axiomatization of its domain. This means establishing some group of principal initial propositions, the so-called axioms, that are destined to describe precisely all the relations holding between the basic concepts of the given discipline, while both concepts and relations are considered to be defined by exactly the list of axioms. All the subsequent concepts required by the given theory must be consecutively determined in a formal way in terms of these basic concepts; and all the subsequent propositions (theorems) have to be formally proved issuing from the axioms. The arguments that insist on this sort of establishment of the foundations as the only possible one in modern mathematics are well known. Only when underpinned in such a manner can it be guaranteed that a theory's formal backbone has really been filtered out without any rest left unaccounted for. Only thus can we perceive with complete distinctness the formal logical connections and interrelations between the various mathematical disciplines.

It is hardly necessary to indicate that such a formalization does not substitute for the real subject-matter a new (speculative) one having nothing in common with the former, as idealistic philosophy would like to imagine.[31] Nor is there any doubt that quantitative relations and spatial forms of the external world[32] continue to be the subject matter of investigation and that the only change concerns the method of studying them.[33]

Strictly speaking, Mises nowhere offers such a formalization of his theory; quite intentionally, his collective always remains infinite and, consequently, becomes an idealized sequence of real trials or observations preserving all the concrete properties and peculiarities of natural objects of the given reality.[34] In his debate with the *nihilists*, the author of the frequentist theory frankly stresses the impossibility and the undesirability of abstracting the theory from these concrete properties. And it should be admitted that he acts here quite consistently: if the theory of probability is considered a natural-scientific rather than a mathematical discipline, then there are no grounds for putting higher demands on it than it is necessary to make with respect to the majority of the natural-scientific disciplines.

In contradistinction, the modern theory of probability, which considers itself a mathematical discipline, creates its axiomatics (i.e., a formal scheme of the studied real domain which represents it as precisely as possible); a scheme where the two opposing tendencies, the striving for maximal abstraction so as to achieve utmost simplicity and the aspiration for retaining a maximal number of the essential concrete properties of the studied subject, should be harmoniously synthesized.

We assume the existence of some observation or trial that at least theoretically admits of an unrestricted recurrence. Depending on chance, an isolated trial can have one or another outcome. The totality of all these possible outcomes constitutes the basic set E, the first among the basic notions of our axiomatics. Any of its subsets, i.e., any totality of possible outcomes, is, as we call it, an event. Mathematical research shows that it is inexpedient to demand here the introduction of all possible events. We restrict our description to considering some field[35] of events.[36] This we shall henceforth denote by Φ. It determines the formal structure of the totality of those events to which we, when considering a given problem, intend to ascribe definite probabilities (example with the die).

In order to accomplish our next step, we turn to a real phenomenon. The theory of probability is interested in those repetitive trials where the frequency of the occurrence of some studied event remains stable. This gives us an occasion to perceive the most important feature of the considered phenomenon, that is, that each event included in our examination is assigned some positive number called its probability. It is expedient here to abstract from everything that accompanies the real phenomenon in the shaping of that concept, in particular, its emergence as a frequency.

Thus: 1) The probability of the event E is unity; and 2) The probability of the sum of two non-intersecting (incompatible) events is equal to the sum of their probabilities. It is self-evident that our segregation from the frequentist picture is by no means its replacement by some other and, especially, by an aprioristic metaphysical view on the nature of probability; a correct abstraction never indeed eliminates the subject so that the possibility of returning, at any moment, from the formal scheme to the real phenomenon is assured.

Thus, in performing the described series of abstracting acts on the real phenomena that are studied by the theory of probability, we arrive at the following system of axioms:

1) Φ is some field of sets.
2) Φ includes the set E.
3) Each set A of field Φ is assigned some non-negative number *P*(A) called its probability.
4) *P*(E) is equal to unity.
5) If sets A and B of field Φ do not intersect, then the probability of their sum is equal to the sum of their probabilities.[37]

The subsequent development of the theory proves that the elementary theory of probability can indeed be completely constructed on the basis of these five axioms. It goes without saying that, by means of a reverse transition from the abstract pattern to concrete reality, all the conclusions of this theory can be interpreted without any hindrance in terms of that frequentist picture, which served as our point of departure for constructing our axiomatics. It is also obvious that the problem of formalization posed by us can be solved in more than one way, that other systems of axioms satisfying the stated demands are also possible.

Let us now return to the position held by Mises' theory. One might suppose that, since this theory does not recognize the theory of probability to be a mathematical science, its method of founding the theory cannot at all rival one or another axiomatic formalization. Indeed, Mises aims at quite another goal. It seems that he never established any system of axioms; the basic theoretical propositions formulated by him cannot by any means be called axioms in the mathematical acceptance of this term. The presence of such notions as trial, observation, phenomenon of nature in the initial propositions of his theory is not a result of carelessness or of an insufficiently distinct scientific language. On the contrary, Mises is undoubtedly apt to insist on preserving those terms. For him, they are a pledge of that indissoluble connection with the actual reality that constitutes the main pathos of his doctrine. But if this is so, may we not say that the frequentist theory and the axiomatic theory can by no means be in a relationship of competition, that their opposition must be the result of a misunderstanding, and that it is necessary to axiomatize the theory of probability, thus leaving to the frequentist theory the task of describing only the connections taking place between the mathematical theory of probability and those real phenomena to which it is applicable?[38]

The frequentist theory itself is not at all inclined to restrict its aims to that modest task that we have just suggested. Following Mises, it is called the new theory of probability. It wants to compete on equal footing with any axiomatic foundation of the theory. Of course, it can secure this right to compete only by completely formalizing itself, by banishing every mention of observations and trials from its propositions and by appearing in the role of a pure axiomatics. And it actually is embarking on this path at the hands of a number of Mises' followers (Dörge, Tornier, and others),[39] whose outlook, by the way, he himself does not always share. The problem consists in that, to render possible such a full formalization, the frequentist theory will apparently be obliged to abandon the initial simplicity of its basic propositions and replace them by other, much more complicated principles.[40] Although Mises' followers willingly agree with this, he himself is not yet inclined to such concessions.

What does he say about all this? It should be noted that it is sometimes difficult to understand this thinker who can be clear to the utmost in all concrete details, but fails to be as clear when it comes to general principles. Mises never formulated his opinion on the problem that interests us now, never offered a complete formalization of his theory. In addition, it is evident that this simply was not his intention. At the same time, however, he fights for his theory as though it were formalized, and when he believes it necessary, he fearlessly opposes it to other formal theories. By the way, the path along which a complete formalization of Mises' theory could have been attempted is fairly obvious. Thus we hardly run the risk of being accused by its author of misunderstanding his ideas, when we, in fact, not only attempt to describe that path, but also discuss it critically.

We must try to solve the two main problems that appear here: Is it possible to formalize the frequentist theory? And, if the answer is positive, will the thus constructed formal theory offer any advantages as compared with the one that was described in this paragraph?

4. Is the formalisation of the frequentist theory possible?

To focus all our attention on the main difficulties of the problem now facing us, we limit our investigation to the simplest probabilistic scheme, i.e., to the so-called simple alternative where a trial admits of only two outcomes. We denote these outcomes by 0 and 1 respectively.

Formalising this pattern in terms of the axiomatics explicated in the previous paragraph is extremely simple. The field Φ consists here of only four events:[41] E (either 0 or 1); 0; 1; (both 0 and 1) and M. Their probabilities are $P(\mathrm{E}) = 1$; $P(0)$; $P(1)$; and $P(M) = 0$. All the axioms are evidently fulfilled. The formalization of this situation in terms of the frequentist theory is much more dificult. This is not surprising because the theory deliberately does not agree with such a high degree of abstraction as provided by axiomatics. It wants to preserve more features of the actual reality, so that its schemes ought naturally to be more complicated.

In the frequentist theory, a formalized pattern is not abstracted from the image of a long series of trials, but preserves that image within itself in an idealized form. The idealization consists here, first, in the fact that this series, having been finite, becomes infinite. Accordingly, the formal scheme of the frequentist theory in our simple case looks like this. Given an infinite sequence, called collective, whose elements are symbols 0 and 1; it has to possess two properties:

1) *The existence of limits*. Denote by $\varphi(n)$ the number of zeros among the first n elements of the collective. Then, as n goes to infinity, the fraction $\varphi(n)/n$ tends to a definite limit $P(0)$ of event 0, which is to be called probability of the event 0. This requirement of course involves the existence of a similarly defined probability $P(1)$ of event 1.

2) *Irregularity*. Let us select, in accordance with an arbitrary law, any subsequence out of the given collective and denote by $\varphi(n)$ the number of zeros among the first n terms of this subsequence. Then $\lim \varphi(n)/n = P(0)$ as n goes to ∞. The mentioned principle for selecting the subsequence can be absolutely arbitrary, if only the choice of each of its elements is independent of whether it be 0 or 1.

In providing examples of such selections, Mises usually indicates purely arithmetical principles, i.e., those for which the choice for an element to belong to the subsequence is completely determined by its number [position] in this subsequence (select all the even numbers or all the absolutely[42] prime numbers, etc.). It is obvious that any other method of choice is here banned because the elements of the collective differ one from another only in their number within the series, and in their values (0 or 1), which are not allowed to be made use of. To avoid complication we shall not dwell here on such rules of choice as, e.g., *each element preceded by a zero* … which Mises nevertheless had to admit during the latest stage of the development of his theory.

Thus, we have here a formal scheme free from any concrete fittings such as *trials* or *observations.* However, before we can compare it with the pattern offered by the axiomatic theory, we must consider the problem of its intrinsic consistency. This will bring us, at once, into the region of the deepest and sharpest points of the modern debates on the foundations of mathematics; therefore, it demands a thorough analysis.

It is clear above all that the first property of the collective, the existence of limits, cannot in itself be doubted; sequences consisting of symbols 0 and 1, where the relative frequency of the zeros (and, consequently, of the unities) tends to some definite limit, are the most usual objects of mathematical research, and any number of them can be constructed.[43] The difficulty, if it exists, should therefore be contained either in the second property of the collective or in jointly postulating both properties.

And, indeed, here we at once encounter considerable difficulties. Every mathematician knows well enough the impediments connected with the idea of *arbitrary* law of selection that plays a fundamental part in formulating the property of irregularity. No matter what kind of a collective is given, its zeros form one of its subsequences which we shall denote by φ. The property of irregularity evidently includes in itself, in particular, the demand that no admissible rule of selection will lead to φ being the selected subsequence. But, since any law is acceptable, the sequence φ must be *lawless*; the zeros in our collective should be arranged in a manner that will not admit of any, even a very involved arithmetical description; only in this case can our sequence be recognized as a collective.

Hence it follows that it is impossible on principle (as Mises admits) to construct an individual collective. Indeed, since we are evidently unable to indicate for each element of the collective whether it is zero or unity (the infinity of the set of elements prevents this to be done), we can only define an individual collective by means of a rule establishing the mutual arrangement of the zeros and the unities. This, however, would have led to determining the sequence φ by means of a law; which is, as shown above, ruled out by the demand of *irregularity.*

There are mathematicians who, merely because of this fact, consider the motion of collective as untenable.[44] Yet, since their point of view is not generally accepted, we must nevertheless go further on, attentively keeping in mind the fact just established.

So, if we are unable, and shall on principle never be able, to indicate any individual collective, let us ask ourselves: what sense can the notion of collective have? and can it become a subject of mathematical investigation? While discussing this problem, Mises indicates a concept introduced during the last years into mathematics by the so-called intuitionist school, the concept of *the sequence of free choice.*[45] Indeed, there is a close connection with the sequences studied by the theory of probability. Concerning our case, the intuitionist school regards the sequence of empty places, each arbitrarily occupied by a zero or a unity, as the sequence of free choice. The act of filling up the empty space is considered an engendering (a making) of continuum. That the sequence of free choice can, and must, be a subject matter of mathematical research[46] is proved to a sufficient degree already by the existence of the theory of binary fractions and of the theory of probability.

Let us see now where we shall find ourselves when we understand a collective as a sequence of free choice. Above all we have to be aware that a sequence of free choice really satisfies the demand of irregularity.

To return, however, to the first demand, i.e., to the existence of limits. For anyone who had to deal with the concept *sequences of free choice*, this demand now rings extremely strange. It is clear, to begin with, that there exists only one sequence of free choice. What sense, then, can there be in separating such sequences into those having, and those not having, limits?

Moreover, the very notion of limit in its usual understanding is applicable only to an individual sequence determined by some regularity. If there are no such regularities, and if they do not exist on principle, the question about the existence or non-existence of limit may not even be posed.

What does Mises say in this connection? He admits that if a given sequence is irregular, the question about the existence or non-existence of limit cannot actually be answered in any concrete case. He adds, however, that this does not yet mean that the existence of limits is a demand contradicting the postulate of irregularity.

We may try to prove the mutual consistency of these two demands and if our attempt succeeds, the definition of the collective will become mathematically appropriate, since, in mathematics, any object whose definition yields no formal contradiction is recognized as existing.[47]

We do not know whether anyone will be able to prove the mutual consistency of those demands that constitute the definition of the collective. This, however, is not the point at all; the point, rather, is that in their substance they are not applicable to the same concept. If we understand a collective as an individually (and therefore regularly) defined sequence, then, as we see, it cannot satisfy the demand of irregularity; if, however, a collective is an irregular (and therefore *lawless*) sequence, then the notion of the existence of limit is not applicable to it. Note that the concepts of non-applicability and inconsistency are completely different. If some demand contradicts a given motion, then exactly the opposite statement is true with regard to this motion. Thus, the demand that a sine of a given angle be greater than unity contradicts the definition of the concept of sine. It follows that the sine is either less than, or equal to unity. If, however, we shall demand that the sine of a given angle should weigh more than one gram, then it will be impossible to say that this condition contradicts the notion of sine; the demand is simply inapplicable to it. It is thus natural that, assuming that *the sine of an angle weighs more than a gram*, we cannot encounter any contradiction.

We are absolutely sure that the disjunction between the existence and non-existence of a limit is inapplicable to the concept of irregular sequence in the same sense as the disjunction between heavy and light cannot be applied to the notion of sine. We think that it is evident that the application of that disjunction is based here on an illusion: in actual fact, mathematicians always have to do with regularly defined sequences and apply that disjunction rightfully. Never encountering other cases, they unwittingly accustom themselves to the rule, according to which that disjunction is applicable to any sequence; and thus, having met with an irregular sequence, they suppose that the appropriate limit either exists or does not exist without noticing that they behave like someone who intends to weigh sines.

Thus, proving the mutual consistency of the two demands made on the collective is not the point at all; the point is, to show what it can mean that the limit exists with respect to irregular sequences. The entire previous history of our science does not provide us with a slightest pertinent indication, and, naturally enough, the revealing of the meaning of the appropriate demand is the duty of the person who first advances it.

We may understand an irregular sequence only as an uncompleted object, as an object in the making. We shall not be able to connect it with any other concept and the demand concerning the existence of limits may be applicable only with respect to completed, to utterly fixed sequences. It follows inevitably that the frequentist theory, at least in that shape in which Mises himself insists, cannot at all be satisfactorily formalized. In any case, we do not now have a satisfactory formalization.

5. Is the formalization of the frequentist theory expedient?

While Mises himself did not give up hope of fully including both main postulates of his theory within a common formal system, his followers have apparently already realized the despair of such attempts at present. At the very least, some partisans are prepared to formalize the frequentist theory at the cost of abandoning one or another of its requirements, and many endeavours have been already made in this direction, viz., replacing infinite collectives by finite ones, completely abandoning the demand of irregularity (Kamke)[48], partially giving up this condition, i e., requiring the invariance of the limits with respect not to any choice of the subsequence, but only to some previously restricted set of such selections (Dörge, Tornier, Copeland and others).[49] Concerning most of these attempts, it is necessary to note that Mises himself adheres to an uncompromisingly negative position. In particular, he believes that, once the postulate of irregularity is even partially abandoned, the formal scheme by necessity acquires features sharply contradicting reality and that this approach cannot therefore be admitted.[50]

From a formal point of view, hardly any of these endeavours of founding the frequentist theory is objectionable. More: after modifying Mises' main postulates in one or another manner, this theory can undoubtedly be irreproachably formalized. As such, it may be admitted as a rival of the axiomatic theory, and we ought to perceive the outcome of this competition.

From the formal point of view, the mutual relations between the axiomatic and the frequentist theories are characterized above all by the former's higher degree of abstraction. Indeed, the foundation of the frequentist theory (in the simplest case) consists of number sequences, i.e., collectives. To each event corresponds a definite sequence of this collective, having a definite limit of the relative frequency, and this limit is called the probability of the event. Event E is defined as such to whom the entire collective corresponds as the chosen subsequence; the equality $P(E) = 1$ is proved as a theorem. The relation $P(A + B) = P(A) + P(B)$ for incompatible events A and B is another theorem. The axiomatic theory abstracts itself from the number sequences which, in the frequentist theory, define the events and their probabilities. It retains only a small number of properties of these concepts necessary for the further development of the theory. For this theory, events are merely elements of some field; their probabilities are simply some numbers made to correspond to these events and to satisfy a small number of simple demands stated in the axioms. The axiomatic theory results by way of abstraction from all the rest that is included in the frequentist theory.

Closely linked with this distinguishing feature is the fact that the theory of probability can be developed on the basis of the axiomatic theory in an incomparably simpler and easier way than when it is founded on the frequentist theory. The latter requires, at least initially, that the unwieldy concept of collective be permanently borne in mind. Especially when the restricted postulate of irregularity is introduced, this is indeed a construction of great formal complexity. And it is natural that in the formal respect the more abstract theory is always simpler.

When two formal theories are in competition, one of them being an abstraction of a higher degree than the other, then we ought to consider the following points to determine which one is to be preferred: **1)** The more abstract theory is always simpler in the formal respect. **2)** Contrariwise, the more concrete (i.e., the less abstract) theory is always richer in features which draw it nearer to reality, and it is therefore able to nourish creative intuition to a larger extent. In each given case the problem is: which of the two advantages might play a greater part?

We may consider it established by the entire modern development of the theory of probability that the high degree of abstraction peculiar to the founding axiomatic system never actually led to emasculating the richness or the diversity of creative intuition. Historically speaking, the comparatively poor equipment of this system with concrete features, if considered a shortcoming, was in any case negligible as weighed against those benefits that are provided by its greater formal simplicity. This is strikingly testified already by the fact that the entire many-sided building of the modern theory of probability was created by people thinking *axiomatically* rather than in the frequentist way; suffice it to mention M. Fréchet and P. Lévy; F. Cantelli and B. Finetti; H. Cramér, S. N. Bernstein, A. N. Kolmogorov, and E. E. Slutskiy.[51] But what indeed did Mises' followers, who were brought up on the frequentist theory, achieve? Can they boast of at least one considerable' finding, or at least one essential discovery enriching the theory of probability?

While discussing the scientific value of the act of abstraction, Mises adduces a vivid and apt example in one passage of his popularizing book *Probability and Statistics*.[52] He notes that geometry, while studying the forms of real space by means of an abstract method, abstracts in particular from the width and the thickness of those objects that we are inclined to call straight lines. Then he recounts that attempts were made to construct a geometry that would have treated stripes of small but differing-from-zero width, instead of straight lines.[53] This lowered the degree of abstraction and, obviously, made the closeness to reality greater. However, these attempts came to nothing and were abandoned. The formal complexity necessary for their realization was not compensated by approximating to reality. And, strictly speaking, there was no need for such an approximation. Experience shows that our mind has excellently learned how to work with the concept of a straight line having no width, and to originate for itself, on this basis, appropriate visual ideas necessary for the functioning of creative intuition.

We consider it obvious that the situation in our case is absolutely similar, and that all attempts at a formalization of the *frequentist* type will sooner or later be abandoned, just as the geometrical theories working with *wide* straight lines were, and for the same reason: They introduce an incomparably greater formal complexity without offering in exchange any practical benefit except for a simple, and so to speak, unselfish satisfaction from sensing that the formal scheme has come closer to actual reality.

When moving to a higher degree of abstraction, we narrow down, in a formal logical way, the content of the concepts in our theory, depriving them of some of their properties. According to the laws of formal logic, this should be accompanied by a certain widening of the extension of these concepts by making possible the inclusion of additional objects into our new and more abstract formal scheme. Practically speaking, in our case it means that, when passing from the frequentist to the more abstract axiomatic formalization, we may envisage such new situations whose interpretation in terms of the *collectives* was altogether impossible. That this is really true cannot be doubted at all.

One of the very simple patterns of the axiomatic theory is this. Let the events be all the Lebesgue-measurable sets of the interval [0, 1]. The probability of each event is then the measure of the corresponding set. This is a well-known interpretation of the problem concerning a mass point thrown *at random* on some unit interval. It is easy to see that the frequentist theory has no place for such a scheme. Indeed, here, the collective should have been the sequence of the abscissas of the fallen point when the series of the throws is unbounded. For any of these abscissas, their totality M is a countable set, its measure is therefore zero and, in agreement with the conditions of the scheme, the probability of the point falling on set M should also be zero. At the same time, however, all the elements of our collective belong to set M so that the probability sought, calculated according to the rules of the frequentist theory, should be equal to unity. The contradiction encountered shows that this pattern of the axiomatic theory cannot indeed be realized under the frequentist system as a basis of probability theory. The representatives of the frequentist theory infer, however, that the distribution described above cannot on principle be experimentally revealed since a statistical experiment amounts only to an empirical determination of frequencies in repeated trials. In accord with their Machian philosophical outlook, they regard their statement equivalent to proving that such a distribution cannot exist in reality.

Thus, the partisans of the frequentist theory claim that the axiomatic system of basing the theory of probability necessarily leads to considering formal schemes that have no prototypes in the real world studied by our science. They wish to regard the described fact as an inherent defect of the axiomatic theory. Several centuries ago the forefathers of those knights of *realism* have been battling against the introduction of negative, and later on, of imaginary numbers. They believed that these had no real prototypes in the actual world. Nowadays these battles seem excusable since we take into account the insufficiently clear understanding, in that early period, of the nature and the methods of mathematics. What should be said, however, about a mathematician of today who would insist on the need of constructing the algebra of polynomials, or the theory of analytic functions in the real, rather than in the complex field since no real magnitudes can correspond to complex numbers?

Meanwhile, however, what is the attitude of Tornier, that champion of the frequentist theory? Striving to *avoid the degeneration of animated essential concepts into lifeless mathematical formalism*,[54] he forbids the usage of schemes that are not confined to the fiequentist interpretation in the theory of probability. For the sake of attaining his goal he constructs a formal and incomparably bulkier structure and is obliged to give up not only the solution, but also the posing of a number of problems, quite elementary from the viewpoint of the wider axiomatic theory. The Lebesgue definition of measure is of course ousted, being replaced by Jordan's unwieldy and stiff definition.[55]

6. The frequentist theory and scientific forecasts

As shown in detail in § 1, Mises believes that one of the main flaws of the classical concept of probability theory is its incapacity to forecast something about the real course of events without introducing new special assumptions. We have recognized this critical remark as absolutely true. In contradistinction, he maintains that the conclusions reached by the frequentist theory contain explicit forecasts about how the real world will be proceeding under certain conditions. We ought to examine now to what extent he is correct.

It is easy to convince ourselves above all that the propositions of the frequentist theory do not on principle admit of experiential checks without the introduction of some new particular assumptions.[56] This is due to the special kind of idealization to which that theory resorts. Consider, indeed, the statement that the probability of throwing a six with a given die is equal to 1/8. According to the idealized interpretation, this means that the relative frequency of this outcome tends to 1/8 as long as the experiment is continuing infinitely. Turning now from the ideal to the real situation, we ought to understand this in the following way: no matter how small a positive number δ is, the relative frequency of the occurrence of a six when the number of throws is sufficiently large, will be contained between $1/8 - \delta$ and $1/8 + \delta$.

Assume now that, wishing to check this proposition by experience, we made a very large number of throws and that the relative frequency of "6" occurred to be equal to 1/4. Issuing from the principles of the frequentist theory, may we now admit, with any justification, that our proposition was wrong? We believe that we have no cause at all for such an inference. The established experiential fact (the relative frequency amounts to 1/4 is evidently quite compatible both with the frequency tending, or not tending to 1/8 as the number of throws increases.[57]

To this argument, which is often put forward, Mises usually objects by the remark that we have to do with the same situation in any physical theory since an experiment always provides only an approximate value of the examined magnitude.[58] Let us see to what extent his opinion is correct. Suppose that our theoretical calculation shows us that the specific weight of some substance must be 1.5, whereas an experiment that uses a device yielding specific weights with a precision of 0.01, furnishes a specific weight of 1.57. It follows that the theoretical conclusion is refuted as incompatible with the experiment. But had the specific weight measured in the experiment been 1.497, we would have been quite justified in assuming that the theoretical conclusion was corroborated, in the sense, of course, that it differed from the truth not more than by some small magnitude known beforehand.

There is nothing similar in the applications of the frequentist theory. No matter how many times we throw our die (i.e., no matter how much we increase the *precision* of our experiment), the result will always remain compatible with any assumption concerning the probability of a six, and no principle of the frequentist theory will tell us which of the possible values of this probability is preferable. Mises likes to compare a repeated statistical trial with a physical experiment; we see that the former is carried out by means of an instrument whose precision is not only unknown, but cannot on principle be determined. So how can the indications of such a device have any meaning when corroborating theoretical conclusions? And may we seriously declare that, these circumstances being given, the situation is here not worse than in any physical theory?

If the frequentist theory itself does not admit of experimental checking of its inferences, is it nevertheless able, as Mises maintains, to provide some forecasts about the actual course of phenomena? The law of large numbers is known to be the cornerstone of practically formulated probabilistic propositions on the proceeding of mass phenomena. Mises quite correctly indicates that, assuming the classical notion of probability, this law is never able without additional assumptions to state anything about the actual course of a given phenomenon. He believes, however, that in the context of the frequentist theory the law of large numbers characterizes the real course of events. Let us see whether this is really so.

Suppose, for example, that the probability that an insured person, belonging to a given category, dies during a certain period of time, is 0.016. According to the law of large numbers it follows that, with probability p rather close to unity, from among 10,000 such insured persons the number of deaths will be close to 160. Following Mises' incessant appeals, let us understand probability in its scientific (in the sense of the frequentist theory) rather than in its everyday meaning and let us ask ourselves what we can infer about the fate of the given real totality of 10,000 persons when issuing from the law of large numbers and its conclusion. In the sense of the frequentist theory, the answer given above by this law literally means that *If an experiment with the fate of the 10,000 persons is repeated infinitely, the relative frequency of those cases, where the number of deaths is contained between 155 and 165, will have the given number p as its limit.* But what relevance does this have to our question? Indeed, we do not intend to repeat the insurance of 10,000 persons a great number of times. We are only interested in the fate of the given totality. What can the frequentist theory tell us about it? Strictly nothing. It has made its conclusion formulated above and it is unable to say a single word more.[59] How does Mises object to this argument? He declares, again and again, that all this is of course true, but that the situation is exactly the same in any physical theory as well.

We do not at all believe it compulsory for a mathematical or a natural-scientific theory to contain in itself all the principles of its practical application. On the contrary, these are usually formulated beyond the given theory. In particular, neither the classical concept, nor the modern axiomatic foundation of probability theory taken in themselves, are capable without additional assumptions to state anything about the real course of events. Therefore the frequentist theory cannot be blamed for being incapable of such prediction; in this respect it fares in no way better or worse than the other methods of founding the theory of probability. The point is, however, that Mises asserts the contrary and believes that one of the theory's main advantages lies exactly in its capability of making such predictions. We have shown that this claim is entirely based on a misunderstanding.

In actual fact, whichever system of founding probability we adopt, the classical, the frequentist, or the axiomatic, the situation will remain invariably the same. To connect the theory with practice, additional assumptions are needed. For example, if the probability of an event is very low, you can be sure that, practically speaking, it will not happen. No matter how simple and clear this principle is, it cannot at all be derived by issuing from the theory of probability itself.

7. Conclusions

1) It is necessary to admit that Mises' most important historical merit is his systematic criticism of the *classical* foundation of the theory of probability revealing its most significant defects: the vicious circle in its definition of probability, the extremely restricted applicability of this definition, and the unsubstantiated claims of the classical theory to a direct practical applicability of its conclusions.

2) Mises' second merit is the systematic and persistent indication that the theory of probability is a doctrine of mass phenomena and that, consequently, it is necessary to derive its main notions and propositions by abstracting[60] it from real statistical populations or repetitive processes rather than to determine them by issuing from the properties of single objects.

3) Mises recognizes the theory of probability not as a mathematical, but as a natural-scientific discipline making use of mathematical methods. This is the reason why, above all, he never and nowhere carries out a complete formalization of his theory; he does not fashion it as a purely axiomatic structure restricting his eflorts with idealizing real processes to a certain degree and leaving intact such notions as *trials*, *observations*, *process*, etc., when formulating his main propositions.

4) In contradistinction, the approach that recognizes the theory of probability as a mathematical science believes that its complete formalization is necessary and that a system of its axioms should be constructed just as it is done in geometry, in algebra and in other mathematical sciences.

5) The foundation of the theory of probability offered by the frequentist theory is not therefore a logical basis in the sense of modern mathematics. Nevertheless, the frequentist theory claims to be a rival of the purely formal foundational schemes. To determine whether this claim is justified, it is necessary, in the first place, to examine the possibility of a complete formalization of the frequentist theory.

6) Within a formalized pattern, the main concept of the frequentist theory, the collective, is represented by a number sequence satisfying two main requirements: the existence of the limits of relative frequencies of the attributes, and the so-called irregularity. The possibility of completely formalizing the frequentist theory under both these demands appears at least dubious. With respect to those notions, which modern mathematics connects with the concept of an irregular sequence,[61] the existence of limits becomes meaningless.

7) These considerations forced Mises' followers to drop this or that element of the two requirements. Most often we encounter a restriction concerning irregularity. Given this relaxation, a complete formalization of the frequentist theory is possible without difficulties.

8) It is quite possible to compare the formalized frequentist theory with the axiomatic theory, which considers the doctrine of probability a part of the metric theory of sets and functions. Such a comparison shows that, with respect to the latter, the former possesses a lower degree of abstraction. Consequently, an inherent feature of the frequentist theory is its incomparably greater complexity, not at all compensated by its being more concrete and therefore closer to actual reality. The axiomatic theory therefore deserves unquestionable preference, whereas the formalization of the frequentist theory cannot be recognized as expedient.

9) Some partisans of the frequentist theory (Tornier) advocate the denial of the more abstract foundations under the pretext that they lead to patterns that cannot on principle be realized in actual reality. This approach contradicts the very spirit of modern mathematical science and ought to be considered as an obscurantist[62] survival of the remote past.

10) The frequentist theory maintains that, among the existing systems of foundation, it alone enables the theory of probability to express statements directly depicting the real course of events. An analysis shows, however, that this claim is based on an illusion and cannot therefore be recognized as justified. In actual fact, the frequentist scheme as a basis for probability theory is no more capable than any other one of attaching to the theory of probability the role of even approximate direct forecasting of the real course of some phenomena. Each system of foundation requires additional assumptions to the same extent, and in this respect the frequentist theory offers no advantage and does not present any exception.

Supplement. The idea of equipossibility: its importance and perspectives.[63]

The logical defects of the idea of equipossibility that prevent it from becoming the formal basis of the theory of probability should not, however, give occasion for underestimating its extremely great methodological importance either for our everyday practice or for the most subtle problems of modern science. In some (although not in a formal-logical) sense this idea, especially as shown by a number of recent investigations,[64] is able to offer more to the doctrine of mass phenomena than any formal foundational scheme. For the sake of comprehensiveness of the picture we have sketched in our survey, let us briefly consider this problem.

The point is not to move backwards from the frequentist interpretation of the idea of probability adopted by the entire modern science, to its old, metaphysical concept. We shall still understand probability as a number that is capable of offering some notion of the relative frequency of the occurrence of an event under certain, absolutely definite conditions. However, if science is something more than a simple, if very economical description of real processes,[65] we obviously should be interested in how to forecast at an earlier point of time (i.e., before the conclusion of the experiment) the frequencies, or what is on principle the same, the probabilities of some events.

When we want to determine the probability of a "6" for a given die, no other means is open to us, as Mises indicated, except throwing it a sufficient number of times and calculating the appropriate relative frequency. Theoretically, this is correct; practically speaking, however, no gambler will act on this method. He knows that the probability sought is 1/6. How does he know? Perhaps, Mises says, the gambler had many times experienced similar dice, for whose overwhelming majority the probability occurred to be 1/6. But the matter is different. The matter is that the gambler's life experience prompts him to this simple conjecture. The six faces of the die have no distinction between themselves of essential importance for the frequency of their occurrence since it is natural to assume that the material of the die is more or less homogeneous. It follows that there is every reason to expect that all the faces will turn up with roughly the same frequency.

Such conjectures, hypotheses based on the idea of symmetry, are encountered at every step, and in each case we forecast the equiprobability of certain events issuing from ideas prior to the given phenomenon. It is self-evident that this assumption may never claim to be more than a working hypothesis demanding an experimental check.

The role of working hypotheses in science is well known. In the sense of a working hypothesis, the idea of equipossibility may be, and actually is, of utmost importance for the theory of mass phenomena. True, this feature by no means expresses itself in the direction in which the classical system of founding the theory of probability wished to apply it. The idea of equiprobability appears not as a formal logical base of the doctrine of mass phenomena, rather it is in single concrete situations our sole method of theoretically forecasting the probabilities of events.

Seemingly, from what was said on the idea of equipossibility in § 1, its domain of application is very narrow. And, indeed, on the face of it, it is difficult to imagine such a pattern as going considerably beyond the province of games of chance. Nevertheless, this is not so. Investigations dating back to Poincaré, and especially extended in recent years, show that in a somewhat generalized form this idea is able to cover a very wide class of phenomena, in mechanics and physics in the first place. It is about these investigations that we wish to say a few words now. Imagine a usual roulette game and suppose that, in the ideal case, friction is eliminated, so that the ball moves round with a constant velocity.[66] The ball's location at a definite moment after it is set in motion is therefore uniquely determined by specifying its initial velocity v. We are now asking ourselves: what theoretical considerations may be put forward in favour of the natural hypothesis that, for a sufficiently high v, all the positions of the ball on the roulette's circumference are equally probable?

The probability of the ball's being in some part of the circumference evidently depends only on the comparison between the probabilities of the various values of the initial velocity. Of course, the assumption that these values are equally probable greatly deviates from reality since the motive force will have its traditional *favourite* values. However, this assumption is not needed. Mathematical analysis shows that, for whichever distribution $\Phi(x)$ of the velocity, if only it is continuous, the probabilities of all the positions of the ball become equal as the period of time increases infinitely (ergodicity). This remarkable result, important as a splendid scientific precedent, gains a simple and clear interpretation in terms of the idea of equipossibility. Indeed, what does the demand of continuity of the function $\Phi(x)$ signify in real terms? Its obvious meaning is that the probabilities of close values of the initial velocity should also be close to one another. This has to do with some differential, or, as it is now usually formulated, local equiprobability. It is indeed remarkable that the relative values of the probabilities of the more or less considerably differing initial velocities are here absolutely irrelevant. It is only essential that no exceptional values of the velocity occur, for any reason, with a frequency much larger than the frequencies of their closest neighbouring values. It is evident that we have every reason to assume that this demand is fulfilled in reality. We thus obtain a scientific theory capable of explaining how and why all the participants in a roulette game have equal chances. We regard this explanation especially satisfactory exactly because it does not require any special assumptions concerning the appropriate distributions. Local equipossibility is an extremely general demand since it takes place for most various distributions.

We have chosen the roulette game as an example for applying the idea of local equipossibility since its mathematical analysis presents no difficulties. Essentially the same method can be made use of in a very large number of problems belonging in the first place to mechanics. In most cases, however, the purely mathematical difficulties still prove insurmountable.

In particular, consider the now classical example of an irregular die. Mises incessantly put it forward as a specimen of problems where the idea of equipossibility is inapplicable; however, from our generalized viewpoint,[67] it can indeed be on principle interpreted by issuing from this idea. Assume that we know exactly the mass distribution inside the die and imagine that, in the ideal case, it is thrown in such a manner that the face that appears is completely determined by the die's initial position. The frequencies of the various faces will then depend on the frequencies of some values of a small number of parameters determining this initial position. And there is every reason to believe that, assuming also a local equiprobability of these parameters, each face of the die will have a definite probability of turning up depending only on the above-mentioned mass distribution.[68]

So, we see that the notion of equipossible cases, after being upgraded to local equiprobability, is able on principle to bear beautiful fruit for our science. This example shows how many treasures that might enrich modern science are buried in classical ideas. It is only necessary to approach them with modern scientific methods that might provide fertile criticism rather than with servility to age-old traditions.

Translator's Note

Until 1990, when Vladimir A. Uspenskiy (Uspensky), Aleksey L. Semenov, and Aleksandr Kh. Shen published an article tacitly corroborating Khinchin's opinion, it had seemed *impossible to embody Mises' intention in a definition of randomess that was satisfactory from any point of view.* The name of their article was *Can an* (*individual*) *sequence of zeros and unities be random*? and it was published in *Uspekhi matematischeskikh nauk* (1990) 45, pp. 105 – 162 (see § 1.3.4). The journal is being translated into English as *Russian Mathematical Surveys*.

Notes

1 See also Gnedenko B. V. and Kolmogorov A. N. (1960), A. Ya. Khinchin (1894 – 1959). *Russ. Math. Surveys* 15, pp. 93 – 106.
2 This assumption is somewhat doubtful as argued below. Probably the paper was written already around 1936.
3 Khinchin's paper, as will be seen, was more philosophical than mathematical and Gnedenko himself decided to publish it in *Voprosy Filosofii.*
4 We were unable to trace any biographical information about the two mathematicians.
5 A special class of stochastic processes the distribution of which is independent (*homogeneous*) with respect to translation of time. This theory was closely connected with Khinchin's work in ergodic theory. Cf. Khinchin, A. Ya. 1934. "Korrelationstheorie der stationären stochastischen Prozesse." *Math. Annalen* 109, pp. 604 – 615.
6 Maistrov (1920 – 1986) is known especially for his book *Probability Theory: A Historical Sketch*. New York: Academic Press, 1974.
7 Not needed.
8 Mises (1919) is the decisive publication, which contains Mises' axioms for the *collectives*. But Mises (1912) has already used the word *Kollektivmasslehre* in the title, which was coined by physicist and psychologist Gustav Theodor Fechner (1801 – 1887), and hints at the frequentist interpretation of probability. One can therefore assume that Khinchin's paper was written in 1939 at the latest. See Mises, R. von 1912. "Über die Grundbegriffe der Kollektivmasslehre." *Jahresbericht der Deutsch. Mathematiker-Vereinigung* 21, pp. 9 – 20, and Mises, R. von 1919. "Grundlagen der Wahrscheinlichkeits-rechnung." *Math. Z.* 5, pp. 52 – 99. See also Heidelberger, M. 1987. "Fechner's indeterminism: from freedom to laws of chance." In *The Probabilistic Revolution*: vol. 1: *Ideas in History*, ed. by L. Krüger, L. J. Daston and M. Heidelberger, pp. 117 – 156. Cambridge: MIT Press.
9 Obviously referring to Mises (1928) in its various editions, which was translated into Russian in 1930 under Khinchin's supervision. See Mises, R. von 1928. *Wahrscheinlichkeit, Statistik und Wahrheit.* Wien: Springer. Mises (1936) is the second German edition with the same publisher.
10 Mises, R. von 1931. *Wahrscheinlichkeitsrechnung und ihre Anwendungen in der Statistik und der theoretischen Physik*. Leipzig und Wien: Franz Deuticke.
11 As far as I know, no analysis of this kind exists at present. The article by the well-known logician Church and the remarks in the book by R. Péter (p. 218) have a narrower aim. Cf. Church, A., On the concept of a random sequence; *Bull. Amer. Math. Soc.* 46 (1940), pp. 130 – 135, and Péter, R., *Recursive Functions* (Russian translation of German original which appeared in Budapest 1951); Moscow 1954. B. G.

Meanwhile (in 2004) several analyses of von Mises' notion of probability have been published. Cf. Martin-Löf, P. 1969. "The literature on von Mises' kollektivs revisited." *Theoria* 35, pp. 12 – 37, and Hochkirchen, Th. 1999. *Die Axiomatisierung der Wahrscheinlichkeitsrechnung und ihre Kontexte. Von Hilberts sechstem Problem zu Kolmogorovs Grundbegriffen*. Göttingen: Vandenhoeck & Ruprecht. See also the unpublished thesis Bernhardt, H. 1984. *Richard von Mises und sein Beitrag zur Grundlegung der Wahrscheinlichkeitsrechnung im 20. Jahrhundert*. Berlin, 226 pp.
12 In October 1937 a *Colloque sur le calcul des probabilités* was held in Geneva. Fréchet's contribution (1938) in that colloquium deals expressly with Mises' theory. Fréchet, M. 1938. "Exposé et discussion de quelques recherches récentes sur les fondements du calcul des probabilités." *Actualités Scientifiques et Industrielle* 735, pp. 23 – 55. S. N. Bernstein, A. N. Kohnogorov, and E. E. Slutskiy published in the proceedings. [? O. S.]. The Russians apparently did not attend that meeting, but were at least aware of it [Such witnesses of the Big Terror were not allowed to go abroad. O. S.] This gives support to the conjecture that Khinchin's paper was written before 1937.
13 Wald (1938) and Ville (1939) were hardly second rate specialists, this again hints at the chronological conjecture above. Cf. Wald, A. 1938. "Die Widerspruchsfreiheit des Kollektivbegriffs." *Actualités Scientifiques et Industrielles* 735, pp. 79 – 99, and Ville, J.. 1939. *Etude critique de la notion de collectif*. Paris: Gauthiers-Villars.
14 Most likely Khinchin is alluding to the second German edition (1936) of Mises (1928).
15 The word *positive* is not unambiguous here, given the fact, that Kolmogorov's *Grundbegriffe* of 1933 existed at that point in [of] time and Khinchin calls the latter the *predominant direction of the modern theory of probability* (below). Khinchin seems to allude rather to the problem of application and to the origin of the notion of probability in nature. As a matter of fact, already in his *Grundbegriffe* of 1933 Kolmogorov would very clearly point to the positive value of von Mises' theory for the connection to application. See Kolmogorov's, 1933. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Springer, p. 3. The English translation of this book is *Foundations of the Theory of Probability*. New York: Chelsea, 1950.
16 Discussion here is about the article "Mises' *doctrine on probability and the principles of physical statistics, Uspekhi fizicheskikh nauk* 9 (1929), No. 2, pp. 141 – 166. B. G.
17 By *unsatisfactory* Khinchin obviously means *too positive*, because recent results in ergodic theory, especially in Khinchin's interpretation, had made outdated von Mises' work on Brownian motion from 1920 to which (Khinchin 1929), as quoted by Gnedenko in the footnote, is referring. See the introduction in the present edition by R. Siegmund-Schültze.
18 Khinchin mentions neither M. Fréchet, A. Wald, nor J. L. Doob, all of whom were involved in discussions with Mises in 1937 and 1941. Thus, one may assume that Khinchin wrote the paper earlier, partly in 1936, and that he is accordingly referring to Mises (1936), the second German edition of von Mises' book (1928).
19 In the manuscript version to which I resort, the end of the sentence is missing. I have added the words in brackets. B. G.
20 See below the *Supplement*.
21 However one has to be aware of von Mises' predecessors in the nineteenth century both in Germany (Fechner, Bruns) and England (Mills, Venn). See Heidelberger (1987). Elsewhere, in 1928, Khinchin even addressed the frequentists as the *English school*. See the introduction by R. Siegmund-Schültze.
22 In fact, Mises' position was more nuanced and ambiguous. He speaks of the theory of probability as a *natural science* in (Mises 1919, 53), and in the English edition of his book of 1928, Mises, R. 1957, *Probability, Statistics, und Truth*. London: George Allan and Unwin, p. 219. However, he means it rather in the restricted sense of the *mathematical* (*tautological*) part of a natural science.
23 Note that Khinchin is approving of that critical part of the Machian philosophy. But below he does not do justice to Mises' reception of Mach. Von Mises was, indeed, impressed by Ernst Mach as a philosopher but critical of the latter's neglect of mathematics. See Mises, R. 1938. "Ernst Mach und die empiristische

Wissenschaftsauffassung," p. 517. In Mises 1963/64. *Selected Papers*, vol. 2, pp. 495 – 523. Providence, Rhode Island: Amer. Math. Soc. (2 volumes = *Selecta* I und II).

24 Khinchin alludes to Friedrich Engels' understanding of mathematics. Engels' *Dialectics of Nature* was first published posthumously in 1927, and therefore its reception was rather new in Khinchin's time. These unfinished notes, written in 1873 – 1883, played in the past a considerable role in much Marxist philosophical discussions on mathematics.

25 In *Materialism and Empirio-Criticism* and elsewhere, Lenin asserts time and again that the idealist is always in fear of dialectics. In paraphrasing Lenin here, the mathematician Khinchin is obviously interested in labelling mathematics as something *dialectical*. Given von Mises' reproach against Mach's lack of mathematics mentioned above, this commentary is particularly inappropriate. Von Mises considered himself first and foremost a mathematician. See also the introduction to this edition by R. Siegmund-Schültze.

26 This remark by Khinchin is particularly careless and at least strongly exaggerated. In (Mises 1936, pp. 123 – 124) Mises calls *nihilists* expressly those who maintain that *there is no need at all for a definition or explanation of the notion of probability*. He adds that there are certain intermediate stages (Zwischenstufen) between the nihilists and his own (uncompromising) view, and names immediately afterwards the *standpoint mathematicians like to withdraw to* which was according to Mises represented by Kolmogorov's *beautiful and very readable booklet*, namely Kolmogorov (1933).

27 This is an English version of the original passage in Mises' German book (Mises 1936, p. 127), which Khinchin translates correctly into Russian.

28 Von Mises (1936, p. 127) in his quotation juxtaposes *Werk* (task) and *Werkzeug* (tool). He is not directly referring to Kolmogorov but to one of his interpreters. We follow the English translation of the third German edition (1951) of Mises' book in (Mises 1957, p. 100), which is otherwise much less detailed than Mises' first reaction to criticism of his theory in the second edition (Mises 1936).

29 Khinchin formulates now the two famous postulates (axioms) on *Kollektivs* (the equivalent of probability distributions or random variables), which Mises had introduced in (Mises 1919, pp. 55 – 57), and defended ever thereafter. Below in § 4, Khinchin gives another definition for a special case of a *collective*.

30 Mises stipulated excluding a *gambling system*, the existence of which would allow for the selection of a subsequence with a changed limit and thus reveal the original sequence as not *random* (irregular). See (Mises 1919, pp. 58).

31 Here the fear of the *idealism-reproach* is palpable which typically is fended off actively by attacking *idealist philosophy*. See also the introduction to this edition by R. Siegmund-Schültze.

32 As stipulated in Engels' *Dialectics of Nature*.

33 Note that the theory of probability built on such a formal axiomatic foundation admits of a frequentist interpretation. B. G.

34 Mises' approach, outwardly tempting, is tantamount to positivism, pure and simple. It restricts the tasks of the researcher to the description and moderate idealization of direct observations and denies the necessity of the next stage, i.e. that of working through to the essence of phenomena. B. G.

35 This is the field of probability (*Wahrscheinlichkeitsfeld*), see *Kolmogorov* 1950, p. 2. In the Russian original one finds *telo*, which is literally *body*.

36 A set of events is called a field, if besides the events "A" and "B" it also includes events "A or B" and "both A and B" and furthermore the impossible and the certain event. B. G.

The sometimes rather careless redaction of Khinchin's paper by Gnedenko is demonstrated by this footnote, which in the original does not only have errors in writing (which may have originated with the printer), but also fails to postulate the complementary event A^c in the field (algebra) of events as well.

37 It is remarkable that the notion of "sum" is not specified here, especially with respect to the need to include denumerably infinite sums of events, as usual in Borelian probability fields. Maybe this was another compromise with respect to the readership of a philosophical journal. Another interpretation (which was suggested to me by B. Bru, Paris) is that Khinchin is only interested in showing that the axioms

of set theory can be related to properties of frequencies where infinite sums do not originally occur.
38 In a similar, but more positive, non-discriminatory sense argues Kolmogorov (1933, p. 3), leaving von Mises' theory the *relation to experience*.
39 See Dörge, K. 1930. "Zu der von R. von Mises gegebenen Begründung der Wahrscheinlichkeitsrechnung. Erste Mitteilung: Theorie des Glücksspiels." *Math. Z.* 32, pp. 232 – 258, and Tornier, E. 1936. *Wahrscheinlichkeitsrechnung und allgemeine Integrationstheorie.* Leipzig und Berlin: Teubner.
40 A review by Khinchin of Tornier (1936) ends with the following words:
It is not understandable why a method was chosen, which uses much more mathematical formalism, and still leads to an identification with measure theory, although a rather restricted and less promising one. See Khinchin, 1936. "Review of E. Tornier: *Wahrscheinlichkeitsrechnung und allgemeine Integrationstheorie*." *Zentralbl. Math. und ihre Grenzgebiete* 13, p. 359.
41 In fact, the field of events Φ has to include events "0" and "1" but also E = "either 0 or 1" and M = "both 0 and 1." B. G.
42 The example of *absolutely prime* as opposed to just *prime* numbers which looks a bit strange and superfluous here may have been motivated by the need to emphasize that the selection is absolutely independent of the outcome of the trial and shall not be weakened by the condition that two outcomes have only *relative prime* positions in the sequence.
43 Note that cautious reference to constructability which points to Khinchin's interest in Brouwer's intuitionism. See below.
44 This example is mentioned in (Mises 1936, p. 113), without referring to specific names of mathematicians. Martin-Löf (1969, 27), discussing a slightly different counterexample concerning the irregularity axiom alone, finds that example *very insensitive to von Mises's intentions*.
45 In the English version of von Mises' book on positivism, on page 129, appears the motion as *sequences of free formation*. Mises, R. 1951. *Positivism, a Study in Human Understanding*. Cambridge, Mass, essentially a translation of the German original from 1939 (The Hague). The concept is mentioned in (Mises 1936, p. 112), as *Folgen freien Werdens*. Neither in his original probability book (1928) nor in its third edition of 1957 does Mises refer to Brouwer's *sequences of free formation*, which were apparently invoked by him as a temporary defence strategy against criticism of his collectives in the 1930s.
46 Note that Khinchin in the 1920s praised Brouwer's intuitionism as revolutionary and as *mathematical bolshevism* and that Kolmogorov developed an intuitionistic *calculus of assignments*. Possible connections between the intuitionistic interests of Khinchin's and Kolmogorov's and their adherence to the Moscow school of *descriptive function theory* (N. N. Luzin), which had close relations to E. Borel's semi-intuitionism, would be interesting to investigate. See Khinchin 1926. "The ideas of intuitionism in the struggle for content in modern mathematics." (Russian) *Vestmk kommunisticheskoy akademii* 16, pp. 184 – 192, and Kolmogorov (Kolmogoroff 1932. "Zur Deutung der intuitionistischen Logik." *Math. Z.* 35, pp. 58 – 65.
47 Von Mises' position with respect to Hilbert's definition of existence was much more nuanced than here presented by Khinchin, and it changed over time. See his original position (Mises 1919, pp. 59 – 60), which is still close to Hilbert's notion, but in (Mises 1936, p. 111), under the influence of criticism, von Mises uses arguments from Bouwer's neo-intuitionism.
48 Cf. Kamke, E. 1932. *Einführung in die Wahrscheinlichkeitstheorie*. Leipzig: Hirzel. This is partly commented upon in (Hochkirchen 1999, pp. 209 – 210). Von Mises called Kamke's standpoint to renounce irregularity' a *curious standpoint* (*sonderbarer Standpunkt*) (Mises 1936, p. 116).
49 Cf. Dörge (1930), Tornier (1936), and Copeland, A. H. 1936. "Point set theory applied to the random selection of the digits of an admissible number." *Amer. J. Math.* 58, pp. 181 – 192.
50 It is interesting to note that von Mises welcomed measure-theoretic arguments if they supported his theory. Mises stressed that in Copeland's model (1936) of a special collective the limit of relative frequency remained fixed for *almost all* subsequences after place selection. Cf. (Mises 1936, p. 118) and (Mises 1957, p. 92). Mises was not fully satisfied with Copeland's *admissible numbers* (Martin-Löf

1969, p. 24). His willingness to compromise, however, contradicts somewhat Khinchin's remark.
51 Kinchin himself undoubtedly should also be named here. B. G.
52 Khinchin refers here to (Mises 1928) in its Russian translation of 1930, which he himself organized. This translation indeed omits the word *Truth* in the title. Obviously the publishers were wary of unwanted propaganda for the philosophical content of von Mises' book. Cf. Mises, R. 1930. *Probability and Statistics*. (Russian translation of Mises 1928, *under the redaction of Prof. A. Ya. Khinchin*). Moscow, Leningrad: State Publishers (Gosizdat). See the cover illustration and the introduction by R. Siegmund-Schültze.
53 Mises mentions that example several times and discusses it in more detail in "Diskussion über Wahrscheinlichkeit," *Erkenntnis* 1 (1930/31), pp. 260 – 285. p. 279. He thereby alludes to Felix Klein (1849 – 1925), who had considered such a kind of geometry to be a part of *Approximations-mathematik*, as opposed to *Präzisions-mathematik*. In that *Diskussion* which took place in Prague in 1929, Mises stated that the geometry mentioned was not (!) very successful. He went on to say that his theory of probability introduced infinite sequence as an idealization (i. e. as a kind of *Präzisions-mathematik*) because *for practical purposes* it was useful to have such abstractions. It appears rather strange and careless that Khinchin - Gnedenko would turn of all examples exactly this one against Mises' probability theory.
54 This is quoted from (Tornier 1936, p. iii), a book which received a critical review by Khinchin (1936), as seen above. It has to be mentioned that Tornier was a proponent of the intrusion of Nazi-ideology in mathematical research in the manner of the racist pseudo-doctrine *Deutsche Mathematik* and that Khinchin was certainly aware of this. Cf. also Hochkirchen, Th. 1998."Wahrscheinlichkeitsrechnung im Spannungsfeld von Mass- und Häufigkeitstheorie – Leben und Werk des 'Deutschen' Mathematikers Erhard Tornier (1894 – 1982)." *NTM-Schriftenreihe* (N. S.) 6, pp. 22 – 41.
55 Camille Jordan's (1838 – 1922) notion of measure and integral was replaced around 1900 by the more modern ones by E. Borel (1871 – 1956) and H. Lebesgue (1875 – 1941).
56 Von Mises was, in fact, aware of the problems connected to infinite collectives and the availability of only finite sections in practice. In (Mises 1936, p. 107) he discussed this problem of application for instance with respect to an article by Hempel, C. G. 1935. "Über den Gehalt von Wahrscheinlichkeitsaussagen." *Erkenntnis* 5, pp. 228 – 260.
57 In fact, in the process of changing, a variable can, prior to approaching sufficiently a limit (in case it exists), assume values arbitrarily distant from the limit. Therefore, fiom the fact that the frequency is close to 1/4 after a finite number of trials, it does not follow at all that it converges to 1/4, if the number of trials goes to infinity. That limit can be arbitrary. B. G.

This last footnote by Gnedenko, which uses the fundamental but rather trivial notion of limit in mathematical analysis, is quite obviously intended for non-mathematicians.
58 Khinchin is obviously referring to Mises (1936, p. 106), where he is, in fact, discussing measurements of specific weights and compares them with conclusions in statistics. Gnedenko uses exactly the same example from physics in order to criticize von Mises' theory. Cf. Gnedenko, B. V. 1962. *The Theory Probability*. New York: Chelsea, p. 51.
59 Mises (1936, p. 105), says: *I must defend myself most emphatically against the recurring misapprehension that in our theory infinite sequences are substituted for all finite sequences of observations*.
60 The modern theory of probability does not abandon altogether the classical approach to defining probability which issues fiom the principle of symmetry. On this more in the *Supplement* B. G.
61 This remark, if maybe justified for the time when the article was written in the 1930s, appears strange given the renaissance of von Mises' notion of randomness around 1960. At least some commentary on the part of Gnedenko would have been appropriate here.

62 This very wording *obscurantist* is at the same time an allusion to the Nazi-ideology involved in Tornier's arguments as mentioned above (see Hochkirchen 1998).
63 This *Supplement* is obviously influenced by Khinchin's turn toward statistical mechanics in the second half of the 1930s and in particular by the so-called *theory of objective probability in classical systems* and the *method of arbitrary functions* in the tradition of Henri Poincaré and M. Smoluchowski. In the 1930s, this theory, which represented one branch of ergodic theory, was particularly pursued by the German Eberhard Hopf (1902 – 1983). Sec Hopf, E. 1936. "Über die Bedeutung der willkürlichen Funktionen für die Wahrscheinlichkeitstheorie." *Jahresbericht Deutsche Mathematiker-Vereinigung* 46, II. Abt., pp. 179 – 195. Von Mises reflected rarely on this kind of investigation which aimed at explaining statistical regularities, independence of events and other problems which the von Mises theory largely took as lying outside the proper theory of probability. See Engel (1992) for a modern revival, and Plato (1983) for the history. Engel, E. 1992. *A Road to Randomness in Physical Systems; Lecture Notes in Statistics* 71; Berlin etc.: Springer. Plato, J. von (1983). "The method of arbitrary functions." *Brit. J. for the Phil. of Sci*. 34, pp. 37 – 47. Also Gnedenko was obviously very impressed by this theory which seemed to give support for major tenets of Marxist epistemology, particularly in connection to probability, and enlarged the philosophical explanatory potential of mathematics as a whole. Khinchin gives a similar philosophical interpretation as here in Khinchin, 1952. "The Method of Arbitrary Functions and the Struggle against Idealism in the Theory of Probability" (Russian). In *Filosofskiye voprosy sowremennoy fiziki* (Philosophical problems of modern physics), pp. 522 – 538 (German translation in *Sowjetwissenschaft*. Naturwissenschaftliche Abt. 7 (1954), No. 2, pp. 261 – 273. French translation in *Questions Scientifiques* 5 (1954), pp. 7 – 24).
64 See e.g. Hopf (1936, p. 180) [see Note 63], which interestingly alludes to the results of the Moscow school of descriptive function theory, referring the result that "measurability is not very different from continuity" to N. N. Luzin, the teacher of both Khinchin and Kolmogorov. See also Hopf, E. 1937. *Ergodentheorie*. Berlin: Springer. (*Ergebnisse der Mathematik und ihrer Grenzgebiete* V,2).
65 One may think of Mach's *Denkökonomie* here, because von Mises was his adherent.
66 According to Krengel (1990, p. 476), the following result on the roulette based on the "method of arbitrary functions" is due to Hopf. Cf. Krengel, U. 1990. "Wahrscheinlichkeitstheorie." In *Ein Jahrhundert Mathematik 1890 – 1990. Festschrift zum Jubiläum der DMV.* ed. by G. Fischer et. al., pp. 457 – 489. Braunschweig: Vieweg. The same example is also discussed in Khinchin (1952).
67 This means, based on "local equiprobability" instead of equipossibility.
68 The ideas incidentally mentioned here by Khinchin were more fully developed in his contribution, "The Method of Arbitrary Functions" etc., see Note 63. B. G.

Khinchin was able to send his manuscript to that same periodical, to *Voprosy Filosofii* which was established in 1947. Did he? And was it rejected if he did? We will never know.

Gnedenko had also published an obituary of Khinchin (*Teoriya Veroiatostei i Ee Primenenia*, vol. 5, No. 1, 1960, pp. 3 – 6). He described Khinchin's capital achievements in various branches of mathematics and called him an *excellent lecturer* whose contributions were *literary masterpieces* and a *citizen in the most elevated sense*.

Now, Khinchin's *invasion* of statistical physics in 1943 was unsuccessful (Novikov 2002, p. 334). Then, in 1953 Khinchin published a Russian *Short Course in Mathematical Analysis* intended, as the author's Foreword stated, as the main manual for students of mathematical-mechanical and physical-mathematical faculties of universities. He had not achieved his aim, the course was too short. This was the opinion of an eminent professor (whose name I forgot) expressed about 1957 to us, students of the evening extension of the mathematical-mechanical faculty at Moscow University.

A very special comment on the Russian booklet *Elementary Introduction to the Theory of Probability* by Gnedenko and Khinchin is warranted. Its Moscow publisher reported that from 1946 to 2013 it was issued thirteen times (half a million of copies in all) and translated into fifteen languages. Nevertheless, I say that it is a most detestable trash which testifies to the greed of the publisher and the ignorance of the public.

In 2015 I translated it anew into English (**S, G,** 65) and added a fitting commentary and many Notes. Gnedenko, who died in 1995, was mostly responsible, but he certainly followed Khinchin.

And now my main question: why Khinchin was only a corresponding member of the Academy of Sciences? I have only one explanation. In 1937, at the peak of the Great Terror, Khinchin glorified the Soviet regime (*Front Nauki i Tekhniki*, No. 7, pp. 36 – 46, **S,G**, 7). At first, I thought that he had somehow become guilty of most cruel Soviet laws and regulations and atoned for it. Now, I tend to believe that he indeed acted as a *citizen in the most elevated* (official Soviet) *sense* and thus alienated himself from the scientific community. A tiny episode confirms my decision.

A second edition of the Russian translation of Jakob Bernoulli's *Ars Conjectandi* appeared in 1986 complete with commentaries, one of them my own. A subeditor told me to suppress my most proper reference to Khinchin. He had not elaborated and, regrettably, I did not ask for an explanation. The editor was Yu. V. Prokhorov, a most eminent student of Kolmogorov.

When mentioning various fields of application of probability theory, Khinchin forgot medicine and meteorology. The example of an irregular die first appeared in a manuscript of Newton (Sheynin 2017, p. 49).

Novikov S. P. (2002), The second half of the 20^{th} century and its result etc. *Istoriko-Matematicheskie Issledovania*, vol. 7 (42), pp. 326 – 356. In Russian.

Sheynin O. (2017), *Theory of Probability. Historical Essay*. Berlin. **S, G,** 10.

XIV

Oscar Sheynin

Pirogov as a statistician

Historia Scientiarum, vol. 10, No. 3, 2001, pp. 213 – 225

1. Introduction

Nikolai Ivanovich Pirogov (1810 – 1881) was the founder of modern military surgery and cofounder of surgery in general, an outstanding educationalist, a public figure and, in 1854 – 1855, an active participant of the Crimean war.[1] He published a considerable part of his writings (and almost all of those which are important for my subject) both in Russian and German (in one case, in Russian and French) so that he was undoubtedly well known in Western Europe. The German spelling of his second name was Pirogoff.

I discuss his works from the point of view of statistics, and, whenever possible, I quote him in German or French. Before listing his relevant contributions I note that all of them except [l; 3; 11] are available in (their original or subsequent) Russian, in the author's *Sobranie Sochinenii* (Coll. Works), vols. 1 – 8. Moscow, 1957 – 1962. In particular, all my references to the Russian version of [7] are from vol. 5 (1961) of that source. Here now is a list of his pertinent works.

1. On the application of statistics, physics and pharmacology in surgery during the last three years (1849). *Protokoly i trudy Russk. Khirurgich. Obshch. Pirogova* for 1882 – 1883 (1883), pp. 125 – 134. Publ. by N. Zdekauer. In Russian.
2. *Rapport d 'une voyage médical au Caucase*. St.-Pétersburg, 1849.
3. On the achievements of surgery during the last five years. *Zapiski po chasti vrachebn. nauk Med.–Khirurgich. Akad*., year 7, 1849, pt. 4, sect. 1, pp. 1 – 27. In Russian.
4. *Sevastopolskie Pisma* (Letters from Sevastopol) (1850 – 1855). *Sobr. Soch*. 8, 1961, pp. 313 – 403. In Russian.
5. Betrachtungen über die Schwierigkeiten der chirurgischen Diagnose und über das Glück in der Chirurgie. *Klinische Chirurgie* No. 1. Leipzig, 1854, pp. 22 – 111.
6. Statistischer Bericht über alle meine im Verlauf eines Jahres, Sept. 52 bis Sept. 53, in Hospitälern, Kliniken und in der Privatpraxis vorgenommenen oder beobachteten Operationsfälle, this being the author's booklet *Klinische Chirurgie* No. 3. Leipzig, 1854.
7. *Grundzüge der allgemeinen Kriegschirurgie*. Leipzig 1864.
8. *Bericht über die Besichtigung der Militär-Sanitäts-Anstalten in Deutschland, Lothringen und Elsass im Jahre 1870*. Leipzig, 1871.
9. *Das Kriegs-Sanitäts-Wesen und die Privat-Hilfe auf dem Kriegsschauplätze in Bulgarien und im Rücken der operierenden Armee 1877 – 1878*. Leipzig, 1882.
10. *Tagesbuch eines alten Arztes* (orig. Russian version 1884 – 1885). Stuttgart, 1894.
11. *Sevastopolskie Pisma i Vospominania* (Letters from and reminiscences about Sevastopol). Moscow, 1950. In Russian.

Only a very short note on the statistical aspect of Pirogov's work was published [Belitskaia, 1950], and I myself touched on this subject in the context of a much more general contribution [Sheynin, 1982]. I am partly issuing both from it and from another short note [Sheynin, 1981]. For German readers, of special interest are Pirogov' s relations with von Baer which Raikov [1968] described in his monograph on the latter' s life and work.

The history of applying the statistical method in natural science might be separated into three stages. During the first of these, empirically noticed regularities were put on record; thus, the Hippocrates aphorisms. The second stage is characterized by the availability of statistical data. The material at hand was sometimes so convincing as to lead directly to extremely important discoveries. So it happened that unpurified drinking water was found, in the mid-l9th c., to favour the spread of cholera. Alternatively, elementary stochastic ideas and methods had to be applied (John Graunt). The third stage began in earnest at the turn of the same (19th) century with the introduction of quantitative tests (e. g., for comparing hypotheses).

The statistical method penetrated medicine by several routes. Of course, it had been applied in population statistics (closely linked with this science). Then, by the mid-19th c., two new disciplines, public hygiene and epidemiology, having much in common both with each other and with demography, have emerged, and the field for applying the statistical method considerably widened.

The first branch of medicine proper to yield to the statistical method was surgery; thus, Simpson [1847 – 1848] statistically estimated the influence of anaesthesia on mortality due to amputations of the limb. He then [1869 – 1870] proved that the mortality of surgical patients increased with the size of the hospital (because of the worsening of hygienic conditions) and coined an appropriate term, *hospitalism*. Other physicians expressed similar thoughts [Simon 1887; Virchow 1868 – 1869, p. 21]. The latter even mentioned (but did not offer an exact reference to) the *Patres von Regensburg* who preceded him by about six hundred years. Both these authors as well as Pirogov (§ 2) left Simpson behind, but they did not confirm their statements by statistical data.

Louis [1825] introduced the so-called numerical method into medicine by calculating the frequencies of the symptoms of various diseases so as to facilitate diagnosing. Gavarret [1840], a physician and Poisson's former student at the Ecole Polytechnique, sharply criticized Louis for the lack of any estimation of reliability of the results obtained.

Pirogov, however, positively appraised the numerical method [1, p. 125; 3, p. 5]. Thus, in the first case, while mentioning syphilis, the stone disease and amputations, he remarked, although without adducing any references, that

Surgeons used the statistical method even before [Louis] *for the determination of symptoms of diseases and indications for certain ways of treatment and operations …*

And, according to Davidov [1854], a mathematician and Professor at Moscow University, Louis actually carried out preliminary treatment of observations (in medicine). In essence, Louis's work may be attributed to the second stage of the statistical method.

The prehistory of the numerical method can be traced back to the last decades of the 18th c. when Black [1788, pp. 65 – 68] presented a catalogue *of all the principal diseases and casualties*. And, even a hundred years earlier, Leibniz, acting in the spirit of *Staatswissenschaft* (statecraft), advocated compilation of *Staatstafeln*, in particular for bringing under control *vielen Krankheiten* [Sheynin 1982, p. 248n].

Neither was the numerical method restricted to medicine. In astronomy, for example, Proctor [1873] compiled charts of 324 thousand stars of the first six magnitudes so as to get along without any theories on the structure of the stellar system. The flaw in such conclusions consisted exactly in that their authors did not pay due attention to theory Actually, *real* statistical studies lead to the advancement of the appropriate science – and, often, to the uselessness of the previously compiled statistical data!

By the mid-19th c. statistical writings began to appear in medical literature (Gavarret, Guy, Farr; Davidov) but as a whole physicians including hygienists and epidemiologists voluntarily or not restricted their attention to the numerical method which had also been applied, for example, for comparing the merits of different ways of treating a given disease. Simpson [1847], who furthered medical statistics (above) and referred to Laplace, Gavarret and Quetelet, in point of fact adhered to the numerical method.

2. The unreliability of statistical data

Pirogov time and time again indicated that medical statistics provided discordant and unreliable data and described several pertinent reasons.

1) In an early contribution he [5, pp. 24 – 26] stated that there were lucky and unlucky physicians and that both lucky and unlucky cases usually occurred in sequences.[2] He hardly repeated this opinion in his later writings but his belief in the existence of such sequences (runs) of [random] events might be corroborated by standard stochastic considerations; alternatively, runs might have been more probable because of hospitalism (also see below).

In his next work Pirogov [6, pp. 4 – 9] mentioned several causes influencing the course of surgical diseases such as climate, the manner of managing the hospital in question;[3] individual features of the patients; and the skill of the surgeon. He also noted that a numerical estimate of the influence of these causes *ist auch zur Zeit noch unmöglich* (p. 8), but I doubt that the situation is much better in this respect even now.

I take up three of these causes.

2) Hospitalism. Pirogov (p, 4) referred to *Hospitalmiasmen*, stated that *Jedes Hospital hat seine Krankheitskonstitution* (p. 5), went into appropriate explanations but did not provide any statistical data. He first discussed this topic even earlier [2, p. 191], then also without adducing data:

Je me suis convaincu par expérience, combien les résultats sont différents entre les opérations faites dans les petits établissements cliniques, et les opérations exécutées dans les grands hôpitaux; et même, combien la différence est grande dans les résultats obtenues par les opérations dans les différents hôpitaux de la même ville, exécutées dans les conditions exactement semblables en apparence.[4]

3) The skill of the surgeon. For separate cases, this was indeed of paramount importance [6, p. 9]. However,

Die Beobachtung der Fälle en masse macht uns schnell demütig, indem wir die beschränkten Grenzen unserer Kunst wahrnehmen.

Accordingly, skill was not revealed *in der ganzen Masse der Fälle*. In any case, Pirogov [6, p. 9] quite consistently thought that unusually optimistic reports should mainly be explained by suppression of unfavourable cases.[6] And, in a letter of 1855, he [11, p. 490] reasonably argued that a proper management of the military medical service was incomparably more important than the skill of the physician.

4) Individuality. Pirogov listed two aspects of individuality which I shall call physical and psychological. The latter [6, pp. 6 – 8] concerned the patient's attitude towards his treatment and his relations with the medical staff. Much more attention Pirogov apparently paid to the former although it is sometimes difficult to say which aspect he was discussing.

Individual features, as he [7, pp. 5 – 6] stated, were important not only in themselves; once they were allowed for, medical statistics will be able to establish the objective danger of diseases and treatments.[7]

About thirty years later Pirogov [l0, pp. 452 – 453] put forward a question which nobody has ever answered either in medicine or elsewhere. Suppose (in my own words) that treatments A and B fail with probabilities P(A)= 0.6 and P(B)= 0.5. Should the physician choose B? Indeed,

Woran soll ich denn erkennen, dass mein Patient gerade zur Zahl der sechzig vom Hundert gehört, welche sterben müssen, und nicht zu den vierzig, welche am Leben bleiben?

He then formulated a similar question concerning B.

Pirogov went on to attach too much importance on individuality by wishing to attribute each man to some definite group.[8] And again [3, p. 5]: individual peculiarities themselves *are subject to statistical inferences*, so that

Only statistical considerations can determine the degree of the influence of the patients' individualities on the course and the treatment of their diseases.

I ought to add that Pirogov [3, p. 6] reasonably believed that the application of statistics in surgery was *in complete agreement* with the latter *because the diseases included in its province depend incomparably less on individual influences or modifications*.

5) Malpractice. This, of course, is a special cause.

Die Statistik nur dann sicher ist, wenn sie keinen antizipierenden Zweck hat und die persönlichen Interessen … dabei nicht im Spiele sind,

he emphasized [7, p. 685]. However, as practised even by the *berühmtesten Hospitalärzte*,

Man die Kranken bei zweideutigen Fällen baldmöglichst nach der Operation aus dem Hospital entfernt, um die Statistik irgend einer Operationsmethode, die ihnen z. B. eigen war, zu verschönern. Auch das entgegengesetzte Verfahren gehört hierher, dass man nämlich die

Aufnahme von Kranken ins Hospital, die sich in verzweifelter Lage befinden, verweigert.[9]

Pirogov did not say anything about Russia; anyway, however, his Russian colleagues considered his honest confession of his own mistakes as unusual.[10]

6) Difficulties peculiar to military surgery. Many times Pirogov (e. g. [7, pp 685 – 686]) pointed out that the inevitable displacements and evacuation of large numbers of wounded (and sick) personnel under war conditions made surgical statistics unreliable. In one instance known to him

Manche frühzeitige Amputationen, die in den ersten 24 Stunden tödlich ablaufen, werden gar nicht notiert, andere wieder zu den Namen solcher Verwundeten, die gar nicht amputiert sind, geschrieben, gleiche oder ähnliche Namen der Kranken verwechselt.

Pirogov went on to discuss other causes leading to faulty statistics and stated (p. 686) that

Sind noch unsicherer die Prinzipien, welche zur Grundlage der statistisch vergleichenden Beobachtungen dienen.

Explaining this proposition, he mentioned several statistical points, such as the need to have a sufficient number of observations and roughly equal numbers of appropriate cases when comparing two methods of treatment; to allow for important circumstances (for example, for mortality from complications or tuberculosis); and, when estimating the danger of some treatment, to take into account *verschiedene lebensgefährliche Zufälle* (p. 694).

He specifically dwelt on the advantages and dangers of late versus early amputations (p. 700). The Russian version of this source[11] is here richer. He argued there [7R, p. 438] that one must understand that the latter were not always performed *early*, whereas the former were really above their reputation: surgeons postponed amputations in the hope of preserving the limb and in some cases they obviously achieved their goal.

Pirogov thus implied that the appropriate statistics was not comprehensive. He returned to the comparison of these amputations [7, pp. 751 – 753], this time directly maintaining that in both cases statistics should be essentially improved (and, in particular, should allow for a number of medical circumstances).

One of Pirogov's correct conclusions [7, p. 699] concerned the use of discordant reports:

Da die Resultate der einzelnen statistischen Berichte zu sehr von einander abweichen, so wird auch die daraus gezogene Mittelzahl zu einer zu willkürlichen und unsicheren Grundlage der Handlungen. Etwas sicherer ist es, wenn man sich nur nach den modernen, für die Verletzungen und Amputationen verschiedener Teile der Extremitäten ausgefertigten Berichten richtet.

His main advice here may be compared with a wrong opinion on a similar matter (mortality from amputations under etherisation and without it during 1794 – 1846 in England).[12]

The (then still practically unknown) Bienaymé – Chebyshev inequality that estimates the probability of the deviations of a random variable from its expectation shows that heterogeneous data can hardly be used in a reasonable way.

3. Attitude towards statistics

Around 1849 Pirogov became convinced in the essential importance of statistics. He indicated that it had been successfully applied in surgery for several decades [1, p. 123] and that [3, p. 4] its application

For determining the diagnostic importance of the symptoms and the merits of operations might be … considered as an essential latest gain of surgery.

Until then, he hardly thought much about statistics. After describing the previous prevalent attitude of military surgeons in favour of earliest possible amputations of injured limbs versus their conservative treatment, Pirogov stated that, both for medical reasons and *durch Einführung einer rationellen Statistik in der Chirurgie* [8, p. 74], this practice was called into question. And he added (p. 75): *folgte ich noch 1847 ohne Zaudern der früheren Doktrin*.

However, also in 1847 Pirogov [6, p. 66], after compiling the appropriate data, first questioned the inevitability of amputations. Later he [7, p. 690] expressed his views quite definitely:

Bei dieser Unbestimmtheit des Quantums von beiderseitigem Risiko schwankt man in der komparativen chirurgischen Statistik fortwährend zwischen zwei Extremen: bald setzt man zu wenig Risiko auf Rechnung der Amputation, bald auf die der konservativen Kur zu viel.[13] *Die alte, nicht statistische Schule überschätzte übrigens den Wert des Lebens. … Die mit der Amputation selbst verbundene Lebensgefahr hielt sie für zu geringfügig, um sie in die Wagschale zu legen. …*

Wir leben offenbar in einer Übergangsperiode. Die geheiligten Grundsätze der alten Schule, deren Ansichten im ersten Dezennium dieses Jahrhunderts vorherrschten, sind durch die Statistik erschüttert, dass muss man ihr lassen, mit neuen Grundsätzen hat sie aber die alten nicht ersetzt, was auch unmöglich ist, so lange die kriegschirurgischen Statistiker nicht nach einem bestimmten und für alle Nationen festgestellten Plane handeln.

Pirogov's common plan was in line with Quetelet's lifelong efforts to standardize population statistics, and he expressed thoughts on compiling statistical data in two more instances. One of his pronouncements was a pipe-dream, pure and simple;[14] his earlier precept [4, p. 382] was at least more definite:

The main point is, record everything, do not rely on your memory; compare the successes of lucky and unlucky physicians, if possible in identical surroundings, and only then estimate the results. Discard old wives' gossip, bureaucratic reports, boastful stories of the rapturous, the quacks and the weak-minded – go on from the operating room to the hospital ward, from the ward to the gangrenous station, and from there to the morgue – this is the only way to discover the truth. …

Elsewhere [7] Pirogov called himself *ein eifriger und aufrichtiger Verehrer der medizinischen Statistik* (p. 5) and (p. 692) formulated its essential aim:

Wollte man auch annehmen, dass das Zufällige selbst durch statistische Untersuchungen in eine bestimmte gesetzliche Form *gestellt wird und dadurch aufhört ein reiner Zufall zu sein; so besteht doch die erste Aufgabe der Statistik darin, das häufiger Vorkommende, das Konstante, herauszukalkulieren und numerisch festzustellen. Das allein kann auch als Grundlage des praktischen Handelns dienen*.

Then, however, he (p. 698) listed many important demands on statistical data hardly attainable under war conditions and concluded that *in der gegenwärtig unverkennbaren Übergangsperiode m der Kriegschirurgie* the surgeon was unable to make considerable use of statistics. It seems that Pirogov had thus set too much store by the future possibilities of the numerical method (§ 1) and on applying statistics to individual cases. On the other hand, he was compelled to make another decision taking into account the actual state of statistics. Here is his most typical utterance:[16]

Even a slightest oversight, inaccuracy or arbitrariness makes [the data] *far less reliable than the figures founded only on a general impression with which one is left after a mere but sensible observation of cases* [7R, p. 20].

In *Ermangelung einer sicheren* [Statistik] *will ich also lieber gar keine, sondern eben nur die Resultate eines solchen Eindruckes in dieser Schrift mitteilen* [7, p. 6].

I myself did not yet [in 1849] *imagine all the blind alleys into which figures sometimes lead* [7R, p. 20].

This mode of action, if understood literally, implied a return to the first stage of the statistical method (§ 1) and was admissible for such a gifted physician as Pirogov. By (following) blind alleys Pirogov possibly meant an unwarranted confidence in the calculated rates of mortality from various injuries and operations (§ 4).[17]

4. The rate of mortality

Pirogov’s discussion of late versus early amputations (§ 3) was of course based on the relevant rates of mortality. So how did he, in general, regard this indicator? In § 2 I quoted him as saying that, in particular, for a large number of observations these rates become stable both for pathological processes and medical treatments if only the initial data were subjected to scientific analysis.[18] Later, however, Pirogov began adding qualifying remarks. Thus [7, pp. 688 – 689]:

Aber m der Frage über vergleichende Resultate verschiedener Behandlungsmethoden in der Kriegspraxis handelt es sich nur um eine mittelgroße Zahl von Beobachtungen, wobei die Schwankungen in den Endresultaten immer bedeutend genug sein werden, um den Einfluss der verschiedensten Verhältnisse auf das Endresultat der Operation oder der Verwundung bemerkbar zu machen, und das wird desto bemerkbarer, je größer der Komplex dieser Verhältnisse sein wird. Gerade aber über diese Verhältnisse erfahren wir nichts aus den statistischen Berichten.[19]

Pirogov apparently implied here that for a really large number of observations the rates of mortality will become stable.[20] This, however, would have been hardly true because the *Verhältnisse* were too diverse. No wonder that, from a practical point of view, Pirogov [7, p. 691] admitted that

Das Mortalitätsverhältnis an und für sich selbst entscheidet gar nicht über den Wert einer Behandlungsmethode, wenn nicht dabei folgende Umstände berücksichtigt werden …

He then listed too many *Umstände* so that his passage could have well ended even before the word *wenn*. Two additional points. First, Pirogov [9, p. 94] concluded that a low rate of mortality *in den Hospitälern während eine Krieges* did not yet testify to *ein günstiges Heilresultat*. G. Blane, at the beginning of the 19th c., and Florence Nightingale, in 1859, expressed the same opinion with respect to hospitals in general [Sheynin 1982, p 265].

Second, before 1879, when the original Russian version of [9] appeared, Pirogov had likely discussed mean rates of mortality, but later in life he [9, p 298] mentioned the minimal rate:

Jede traumatische Verletzung, also auch jede chirurgische Operation ihr wohl mehr oder weniger schwankendes, aber immerhin bestimmtes Sterblichkeitsminimum hat, welches sich trotz aller unserer Bemühungen und trotz aller Erfolge unserer Kunst der Totalsumme der Fälle nicht vermindern lässt.[21]

Geselevich et al [1960, p. 557] rejected Pirogov's belief in the existence of such minimal rates but I do not agree with them. Exactly these rates, corresponding to favourable conditions (and a given period of time), should be more stable than the mean rates reflecting most variable circumstances.

Nevertheless, Pirogov did not resolutely turn his attention from mean to minimal; mortality: even in his last contributions there are quite a few places [8, p. 80; 9, p. 44, 476 and 524] where he discussed mortality with no adjective attached to it.

5. "Der Krieg ist eine traumatische Epidemie"

This was Pirogov's' celebrated statement [9, p. 295] and he understood it in a wide sense. Thus [7, p. 693].

Der Soldat zu Anfang und zu Ende des Kriegs nicht derselbe ist. Die Sterblichkeit nimmt gradatim mit der Zunahme der Erschöpfung in Folge der Kriegsstrapazen zu.

Or, in the same vein [7, p. 5],

Daher erwarte ich den wahren Fortschritt der Medizin vielmehr von dem Aufsuchen solcher Maßregeln, welche den menschlichen Organismus vor Leiden schützen und dem Leiden vorbeugen können.

The second (the Russian) version of this source is here more definite and explains his expression *traumatische Epidemie* [7R, p. 20]: the mortality of the wounded is the result of the *degree of various sufferings and privations rather than of the seriousness of the wounds and operations*. Later Pirogov [9, p. 302] expressed himself in the same way. It is therefore understandable that he devoted a large part of two of his books [8; 9] to the organization of military surgery (the choice of the optimal number of the various types of hospitals and of their location; the distribution of the medical personnel; etc.).

Such problems, whose solution could have been based only on an approximately known or even crudely estimated figures, now belong to operations research. That Pirogov attached great importance to this aspect of medical activities is also evident from his direct pronouncements [9, p. 98][23]:

Für unvorgesehene Fälle eine hinreichende Anzahl von Unterkünften, Hospitalpersonal u. s. w. bereit zu halten, ... und noch mehr alles vorher zu überlegen und richtig auszuführen, wiewohl der Überschlag nur annähernd und auf die Gesetze derWahrscheinlichkeit gestützt sein kann – hierzu gehört Genialität [brilliant intuition] *und Erfahrung*.

I doubt that the laws of probability could have helped: Pirogov himself [9, p. 98] maintained that hospitals should

Immer bereit sein die dreifache Zahl von Kranken und Verwundeten aufzunehmen, als der Etat der Lagerstellen beträgt.

And it is not out of place to mention here that Pirogov's mathematical education was not sufficient [10, p. 144]:

Ich in der Schule wohl der beste Schüler in Geschichte und russischer Literatur, nicht aber in der Mathematik gewesen bin. Dabei glaube ich aber wohl behaupten zu dürfen, dass in mir etwas von mathematischer Ader steckte; sie entwickelte sich aber, glaube ich, nur sehr langsam mit meinem fortschreitenden Alter, und als ich das sogar sehr lebhafte Verlangen fühlte, etwas von Mathematik zu verstehen, da war es schon zu spät dazu.

Summary

Pirogov was mainly concerned with the unreliability of statistical data in surgery, especially under war conditions. He advocated careful recording of cases and a preliminary study of the pertinent circumstances, and he believed that this attitude will eventually ensure the determination of stable mortality rates for various surgical diseases (and treatments). He was too optimistic just as many other scientists of the 19th c. were (physicists then thought that their science was all but completely studied; Darwin and Quetelet believed in a near softening of the relations between nations; Marx had his own dreams …).

For the time being, Pirogov preferred to rely on general impression rather than on discordant figures, and he believed in the regularities inherent in mass random phenomena. This enabled him to compare the expediency of the conservative treatment of injured limbs in contrast to their amputation. Again, the trust in statistical regularities apparently governed his successful activities in organizing military surgery, cf. his maxim which I chose as the title of § 5.

In Pirogov's time, neither probability nor statistics were included in the curriculums for student physicians, and it was undoubtedly his practical experience that led him to become statistically inclined.

Pirogov's main pertinent contributions were published in German (and in some cases also in Russian) and apparently became widely known, but, unlike Florence Nightingale (Note 1) or perhaps Simpson (§ 1), with regard to his statistical ideas he remained a lone figure. Without additional studies it is impossible to assess the influence of Pirogov on the subsequent development of medical statistics.

Acknowledgements. This is an essentially revised version of my earlier Russian paper published in the *Izvestia* of the Petersburg University of Economics and Finance, No. 3 – 4 (1995), pp. 144 – 151. Note 1 is written on the recommendation of Prof. R. Diez and I

am also indebted to the reviewer who suggested that my *Summary* should be largely rewritten.

Notes

1. His biography can be found in the *Bolshaia Sovetskaia Enziklapedia*, in vol. 19 of its third edition available also in an English translation (*Great Sov. Enc.*, 32 vols. New York – London 1973 – 1983, see vol. 19, pp. 555 – 556).

Florence Nightingale was another participant in the Crimean war (on the other side of the front) which occurred almost in the beginning of their professional careers and profoundly influenced their scientific future. As is shown in the sequel, and as it was natural for a physician, Pirogov only applied statistics in medicine (more precisely, in surgery), whereas Nightingale was a versatile statistician. In addition to important work in hospital statistics, on hygienic conditions in the British army in India, etc., she engaged in various other issues of public interest (e. g., in studying the results of the then recently established compulsory education). She was able to secure help from leading statisticians (Farr); Pirogov however had to remain a lone figure. On the statistical work of Nightingale see Kopf [1916]; Spiegelhalter [1999].

2. *Glück* was also mentioned in the title of [5]; I quote from this source (pp. 25 – 26):

Müssen wir doch gestehen, dass es Leute gibt, die immerfort gute Karte bekommen, und möge man von der Wahrscheinlichkeitslehre sagen was man will, so ist es doch keineswegs ausgemacht, dass, wenn man z. B. eine gewisse, sogar größere Zahl von Kranken in zwei Reihen teilt, sich diese Reihen gleichmäßig gruppieren werden und dass die Zahl der günstigen Fälle auf beiden Seiten dieselbe sein wird.

3. A related topic was the administration of prisons. Much earlier it was convincingly argued that mortality in prisons greatly depended on the general way of their management and that even a slight slackening of the routine could result in an immense decrease of the death-rate among the inmates [Villermé 1829, p. 21].

4. An important feature, Pirogov continued, was

La constitution générale d'un hôpital. Cette constitution générale, comme étant la consequence de l'organisation d'un hôpital, de son installation, de sa situation, et enfin aussi souvent de certaines maladies que l'on traite particuliérement dans tel ou tel hôpital.

5. And again [6, p. 2]:

Der Einfluss des Arztes aber, die verschiedenen Kurmethoden und die mechanische Fertigkeit spielen so sekundäre Rolle, dass sie nur ein in der Masse kaum bemerkbares Schwanken der Zahlenverhältnisse hervorrufen.

Concerning *Kurmethoden* Pirogov explained his idea elsewhere [7, p. 5]: each of these *sein eigenes Quantum des Schadens in sich enthält* etc. Then, however, he actually went back on his opinion, or at least qualified it [7, p. 688].

6. Cf. Darwin's [1887, p. 143] feelings about an unnamed source, later proved to be based on fabricated facts:

A Belgian author … stated that he had interbred rabbits in the closest manner … without the least injurious effects. … I could not avoid feeling doubts, I hardly know why, except that there were no accidents of any kind, and my experience in breeding animals made me think this improbable.

7. I qoute:

Wenn aber in Resultaten der Forschungen über Letalität eines und desselben pathologischen Prozesses oder einer und derselben Operation, bei einer mittelgroßen Zahl von Beobachtungsfällen, bedeutende Schwankungen vorkommen, so hängt dies offenbar davon ab, dass die Lehre von der Individualität, eine der wichtigsten, noch so gut wie gar nicht existiert, und dass wir nicht einmal annäherungsweise im Stande sind, die Verschiedenheit der Verhältnisse und der subjektiven Eigenschaften verschiedener Beobachter zu taxieren. Wenn es mit der Zeit gelingen wird, auch diese Momente in medizinisch statistischen Arbeiten einer wissenschaftlichen Analyse zu unterwerfen und ein großes Beobachtungsmaterial nach bestimmten Gruppen zu ordnen, dann werden und müssen auch die Schwankungen in Bezug auf die Resultate immer geringer werden. Dann wird es uns vielleicht möglich sein, auch das von der Persönlichkeit des Kranken und Arztes [did he think about skill or luck? Or about both?] *sowie das von andern Nebenumständen abstrahierte Verhältnis der Letalität jedes pathologischen Prozesses und der Heilkraft jedes Mittels festzustellen*.

8. Indeed:

Wenn das Studium der menschlichen Individuen so weit gefördert sein wird, dass man jedes Individuum nach verlässlichen Anzeichen zu der einen oder der andern scharf bestimmten Kategorie rechnen kann und andrerseits die Eigenschaften bekannt sein werden, die eine jede Kategorie zum Widerstande wider äußere und organische (*innere*) *Einflüsse befähigen – dann wird auch diese Statistik mit ihren ziffermäßigen Daten eine ganz andere Bedeutung bekamen*.

Elsewhere Pirogov [7, p. 695] mentioned age and *der noch völlig unbekannte Grad der Vulnerabilität der Rassen* and he also stated [9, p. 440] that

Außerordentlich verschiedene Ursachen teilen die ganze Masse der Data in zu unbedeutende, einander sehr unähnliche Gruppen, welche keinen richtigen Schluss über den Wert einer jeden Amputation zulassen.

This apparently meant that separation of cases into groups was a delicate procedure.

9. A similar attitude (pp. 26 – 27) was to refer undesirable patients to other physicians; again [6, p. 9], unsuccessful cases were not always recorded and in some instances (application of the Listerian antiseptic bandage in the field [9, pp. 427 and 456]) arbitrary combination of data had been possible because of the ill-considered forms of the record cards. Elsewhere Pirogov [10, p. 453] added:

Während meines Aufenthalts im Auslande hatte ich mich zu Genüge davon überzeugen können, dass die wissenschaftliche Wahrheit bei weitem nicht der Hauptzweck selbst berühmter Kliniker und Chirurgen ist. Ich hatte mich hinlänglich davon überzeugt, dass in berühmten klinischen Anstalten gar oft Maßregeln nicht zur Enthüllung, sondern zur Verdeckung der wissenschaftlichen Wahrheit getroffen wurden.

10. I quote Raikov (Note 5, see p. 200) who did not regrettably justify his statement concerning the beginning of the 1840's:

Seine Zeitgenossen staunten über die Ehrlichkeit, mit der er selbstgemachte Fehler nie verschwieg, sondern offen in der Spezialliteratur besprach.

11. Denoted here and below by [7R].

12. Simpson [1847 – 1848, p. 102]:

The data I have adduced … have been objected to [by whom?] *on the ground that they are collected from too many hospitals, and too many different sources. But, on the contrary, I believe that our highest statistical authorities will hold that this very circumstance renders them more, instead of less, trustworthy*.

13. Not needed.

14. Here it is [9, p. 529]:

Wenn wir gewillt sind alte, sich immer wiederholende Fragen der Kriegschirurgie mit Hilfe der Statistik zu entscheiden, so ist ... ein besonderes Institut von Spezialisten erforderlich, welche verpflichtet sind auf den Verbandplätzen und in den Hospitälern ... persönlich zugegen zu sein.

15. Here is his final statement.

Nur auf diese Weise [only after fulfilling all of them] *könnte, nach meiner Ansicht, die kriegschirurgische Statistik auf die Stufe der exakten Wissenschaft erhoben und dann auch zur Bestimmung der das Schicksal von Tausenden von Verwundeten entscheidenden Indikationen auf dem Verbandplätze benutzt werden. So lange aber dieser Vorschlag nur eine philanthropische Utopie bleibt, so lange wird auf gleiche Weise auch das Streben, eine wissenschaftliche Indikation für die Erhaltung .und die Amputation des Glieds bei komplizierten Knochenschuss-verletzungen stellen zu wollen nur als Utopie angesehen werden dürfen*.

16. I am now quoting from both (non-identical) versions of the same source [7].

17. Thus, he [2, p 191; 3, pp. 7 – 8] previously remarked that in some cases anaesthesia had increased the rate of mortality (because of subsequent respiratory complications). True, even then he doubted the accuracy of the data which included his own records and he had also noted that amputations without anaesthesia had been partly performed in small clinics, that is, under better hygienic conditions. Surprisingly, he did not add that the new procedure enabled the surgeon to widen essentially his possibilities.

18. He said much the same elsewhere [6, p. 2]:

Jede Krankheit und jede chirurgische Operation in Bezug auf nichtgelingen und tödlichen Ausgang ihr festes und bestimmtes Verhältnis hat. Dies Verhältnis hängt ab von der kontinuierlichen Einwirkung der äußeren Bedingungen auf die verschiedenen Krankheitsformen, von der Natur der Krankheit, von der Individualität der Kranken, so wie von der Art des traumatischen Eingriffe der mit jeder Operation verbunden ist.

Here is another early statement [4, p. 382]: mortality is constant for any epidemic as well as for mass *considerable* operations. And, again [7, p. 5]:

Ich neige zur Ansicht, dass die Letalität jedes pathologischen Prozesses, jeder Verwundung und jeder Operation, die verschiedensten Verhältnisse ungeachtet, doch im Ganzen etwas Konstantes und bestimmtes sein muss.

19. And [6, p. 8]

Nur in wenigen chirurgischen Krankheiten kann das Sterblichkeitsverhältnis, die äußeren Bedingungen mögen sein wie sie wollen, durch eine ziemlich konstante Ziffer ausgedrückt werden …

Later he repeated this opinion with regard to the amputations of the thigh [9, p. 440].

2(). He [9, p. 442] expressly stated that *Widersprüche und Inkonsequenzen* remain *unausbleiblich* if conclusions were based on a small number of observations.

21. In the same source Pirogov [9, p. 525] formulated a test for the advisability of the conservative method of treating injured limbs:

Ob das Minimum der Mortalität bei konservativer Behandlung der Schussfrakturen des Oberschenkels und des Knies plus dem bekannten Mortalitäts-prozent der sekundären Amputationen (als Folgen der konservativen Behandlung) gleichkommen wird dem Minimum der Mortalität nach primären Oberschenkel-amputationen bei denselben Verletzungen.

22. A similar statement is in [9, p. 297].

23. Also [8, pp. 48 – 49]:

Wovon hängen die Erfolge der Behandlung oder die Mortalitätsverminderung in den Armeen ab? Doch gewiss nicht von der Therapie und der Chirurgie an und für sich. ... Für die Massen steht von der Therapie und Chirurgie ohne eine tüchtige Administration auch in Friedenzeiten wenig Nutzen zu erwarten; um so weniger also noch bei solchen Katastrophen, wie ein Krieg.

Bibliography

References to authors other than Pirogov

Belitskaia, E. A., 1950, Problems of military medical statistics in the works of Pirogov. *Voenno-Medizinsk. Zhurnal*, No. 3, pp. 57 – 61. In Russian.

Black, W., 1788, *Comparative View on the Mortality of the Human Species.* London.

Darwin, C., 1887, *Autobiography*. London, 1958.

Davidov, A. Yu., 1854, Application of probability theory to medicine. *Moskovsk. Vrachebn. Zhurnal* No. 1, pp. 54 – 91. In Russian.

Gavarret, J., 1840, *Principes généraux de statistique médicale*. Paris.

Geselevich, A. M., Seneka, S. A., 1960, Commentary to Pirogov [9]. Pirogov's *Sobranie Sochinenii* 7, pp. 533 – 603. In Russian.

Kopf, E. W., 1916, Florence Nightingale as a statistician. *J Amer. Stat. Assoc.* 15, pp. 388 – 404.

Louis, P.-C.-A., 1825, *Recherches anatomico-pathologiques sur la phtisie*. Paris.

Proctor, R. A., 1873, Statement of views respecting the sidereal universe. *Monthly Notices Roy. Astron. Soc*. 33, pp. 539 – 552.

Raikov, B. E., 1968, *Karl Ernst von Baer.* Leipzig. Orig. publ. in Russian in 1961.

Sheynin, O., 1981, Poisson and Statistics. In *S.-D. Poisson et la science de son temps*. Eds, M. Métivier et al. Paris, pp. 177 – 182. (Also [ix].]

Sheynin, O., 1982, On the history of medical statistics. *Arch. Hist. Ex. Sci*. 26, pp. 241 – 286.

Simon, J., 1887, *Public Health Reports*, 1 – 2. London.

Simpson, J. Y., 1847: Value and necessity of the numerical method. *Monthly J. Med. Sci*. 8, No. 17 (83), pp. 313 – 333.

---, 1847 – 1848, Anaesthesia. *Works*, 2. Edinburgh, pp. 1 – 288.

---, 1869 – 1870: Hospitalism". Ibidem, pp. 289 – 405.

Spiegelhalter, D. J., 1999, Surgical and statistical lessons from Nightingale and Codman. *J Roy. Stat. Soc*. A162, pp. 45 – 58.
Villerme, L. R., 1829, Sur la mortalité dans les prisons. *Ann. hyg. publique,* pt. 1, pp. 1 – 100.
Virchow, R., 1868 – 1869: Uber Hospitäler und Lazarette. *Ges. Abh.* 2. Berlin, 1879, pp. 6 – 22.

XV

Oscar Sheynin

Social statistics and probability theory in the 19th Century

Historia Scientiarum, vol. 11, No. 1, 2001, pp. 86 – 111

Based on a large number of sources, this paper traces the relation between statistics (excluding demography) and the theory of probability in the 19th century and links this subject with the development of the statistical method in natural sciences. The battle between the partisans of the Staatswissenschaft and the adherents of political arithmetic continued well into the mid-19th century, and, except for some mean indicators, probability hardly entered statistics before Lexis. The causes of this situation were the contemporaneous absence of many issues introduced into statistics in the 20th century as well as the real or imaginary difficulties of applying probability.

1. Introduction

By social statistics I understand statistics in its usual sense but without population statistics and I depict my subject as it developed before the 20th century, mainly in the 19th century.[1] Medical statistics is traditionally treated as a discipline in its own right[2] and, accordingly, I left it out. Neither did I study the beginnings of the Biometric school except for linking its approach with empiricism.

Among my main sources are Knies (1850), the contributions of Bortkiewicz and Chuprov and a number of my own writings (the overlapping with which is minimal). I examine the period before and after 1850 in §§ 2 and 5 respectively; a separate section (§ 3) is given over to Quetelet and in § 4 I discuss empiricism as it was manifested in statistics during the entire 19th century. Finally, in § 6 I offer a few remarks on the work of Lexis.

My article is a comprehensive account of its subject. For the first time ever, I linked it with the history of the statistical method in natural sciences; the relevant documentation is in my previous papers, see the References. Then, new material is contained in § 2.4 (the essence of the battle between the two approaches to statistics) and in §§ 5.2 and 5.3 which explain the late introduction of probability theory into statistics. Finally, I described many unknown or hardly known statements of Chr. Bernoulli, Butte, Humboldt, of many Russian authors, et al, as well as Chebyshev's opinion on the need to reconstruct the theory of probability (Note 42).

2. The situation before 1850

2.1. Mathematics. The infantile stage of the theory of probability ended in 1713 when Jakob Bernoulli's law of large numbers was published and became the bridge between probability and statistics. Bernoulli also thought of, but did not have time for applying probability to civil, moral and economic issues.

Referring to his unaccomplished goal, Montmort (1713, p. xiii) reported that he chose not to apply probability to the *sujets politiques,* *oeconomiques ou moraux*; he was reluctant to *faire des hypotheses* and lefl this subject *à une autre temps*. Then Montmort (p. xvi) added that *le caprice* guided men much more than *la raison* and that (p. xix) the postponed task would have demanded the disregard of free will, cf. Note 29. Finally, he (p. 322) stated that he had *quelques* [appropriate] *idées & quelques materiaux*, but that he was unable to study this subject with mathematical precision.

The first direct use of probability in statistics was due to Arbuthnot, who, in 1712, examined the sex ratio at birth by employing a quite elementary stochastic analysis. The same problem prompted De Moivre to prove, in 1733, the De Moivre – Laplace limit theorem, as it is now called.[4] Daniel Bernoulli invariably based his statistical studies (mortality from smallpox; duration of marriages; and, again, the sex ratio at birth) on probability.

In statistics, Laplace is mostly remembered for his estimation of the population of France by means of sample data.[5] He also systematically examined the statistical significance of empirical discrepancies and small effects both in natural sciences and population statistics. Then, following Condorcet, he applied stochastic methods to study the testimonies of witnesses and the verdicts of law courts. Only in passing he (1812, p. 523) noted that he had assumed mutual independence of the judges (or jurors), and this shortcoming led to the denial of such applications of probability. That his (and Poisson's later) arguments helped to understand the administration of justice in the ideal case was not recognized; true, Cournot (1843)[6] described this subject without criticizing him.

Laplace (1814, p. 62) not only urged scientists to apply *the method based on observation and calculus* to the *political and moral sciences*; he also noted the [relative] constancy of dead letters and of the profit gained from lotteries (p. 37). Still, *moral* applications of probability were difficult, cf. Montmort's opinion above. Fourier (1821a, pp. iv – v) declared that

L'esprit de dissertation et de conjectures est, en général, oppose aux véritables progrés de la statistique, que est surtout une science d'observation. Somewhat before 1826 he (Quetelet 1869, t. l, p. 103) stated that such progress was only possible if statistics be confined to subjects examined by *théories mathématiques*, but he did not mention the theory of probability.

Gauss is known to have been collecting various statistical data, partly for his own pleasure. He also studied infant mortality, mortality of the members of tontines and occupied himself with life insurance but apparently without applying any serious stochastic considerations.

Poisson resolutely supported the application of probability theory to the *baute statistique*. Thus (Libri Carrucci et al 1834, p. 535):

The most sublime problems of the arithmétique sociale can be resolved only with the help of the theory of probability.

Social arithmetic was a short-lived term, possibly coined by him; it denoted demography, medical statistics and actuarial science (Sheynin 1978, pp. 296 – 297). Poisson et al. (1835, p. 174) also argued that

Statistics carried into effect is, after all, always the functioning *mechanism of the calculus of probability, necessarily concerning infinite masses, an unrestricted number of facts*.

Referring to his correspondence with Poisson, Quetelet (1869, t. l, p. 103) testified that the former had *parfois* derisively expressed himself about those statisticians who were apt to substitute their *fantasies* for the *véritables principes de la science*. Poisson (1837) systematically studied the statistical significance of empirical discrepancies and Cournot (1843, end of Chapt. 8), who referred to Bienaymé, and Lexis (1875, p. 103ff) followed him. His former student, Gavarret, who later took to medicine, advocated such methods as the comparison of two differing treatments and the check of null hypotheses (1840). On his p. XIII he warmly acknowledged Poisson's influence.[8] Poisson may well be called the godfather of the Continental direction of statistics (§ 6; cf. Note 51).

Cournot (1843), as the context of his Chapt. 9 testifies, was all for the connection of probability and statistics. Bienaymé (Guerry 1864) stated that it was *absolument impracticable* to separate the two disciplines. Nevertheless, opposition to such a link became fierce, although it was mostly directed against the application of probability for examining the work of the law courts and the voting procedures. One such antagonist was Poinsot (Poisson 1836a, b).[9] Much later Poincaré (*Le procés Dreyfus* 1900, t. 2, p. 331) voiced a similar opinion.

Quetelet (*Congrés* 1873, p. 139) mentioned a related point. He noted that [apparently in the mid-century] mathematicians had moved away from statistics[10] which resulted in *large mistakes* in calculations. He did not elaborate and he could have said *moved away from probability theory*.

2.2. Political arithmetic. It came into being in the mid 17th century. Neither Petty, who coined this term, nor his friend and colleague Graunt (1662) ever defined it, but they strove to base their studies on numerical data and shrunk from applying qualitative characteristics. Graunt was able to use fragmentary statistical data for estimating the population of London and England and to compile the first mortality table. Already the title of his book implied that Graunt attempted to study the influence of various factors on mortality; and he also made several conclusions about London (a *Head too big for the Body*, p. 320), the electoral system, etc. In the 1680s, Leibniz wrote several manuscripts (first published in 1866) partly devoted to political arithmetic and recommended to compile *Staats Tafeln*, with or without using numbers (Sheynin l977b, § 2.4.4).

Süssmilch is best remembered for originating demography, and, especially, for compiling a life table which continued in use well into the 19th century. He attempted to prove divine providence by revealing the stability of vital statistical ratios (e.g., of deaths to births). And, like Graunt, Süssmilch discussed the pertinent causes and offered conclusions. Thus, he (1758) thought of examining the dependence of mortality on climate and geographical position and he knew that poverty and ignorance were conducive to the spread of epidemics.

Statistics (in the modern sense) thus took over from political arithmetic. However, since Achenwall's disciples continued to include the former in the Staatswissenschaft (§ 2.3), the latter term persisted for a long time in statistical writings.

2.3. Staatswissenschaft. This was a branch of science founded by Conring (1606 – 1681), and Achenwall (1719 – 1772) first presented it systematically. According to both scholars, the aim of the Staatswissenschaft or statistics (!) was to describe the climate, geographical position, political structure and economics of a given state and to estimate its population.[11] It is to Schlözer (1804, p. 86), Achenwall's most eminent student and a member of his Göttingen school of statistics, that this discipline owes its celebrated early definition:

Geschichte ist eine fortlaufende Statistik und Statistik eine stillstehende Geschichte.[12]

He thus opposed political arithmeticians by leaving aside the study of causes and consideration of probable developments.

2.4. The competition. The fierce competition between political arithmetic and Staatswissenschaft continued for more than fifty years with the points of dissent being both the subject and the method of research.

1) The subject. Political arithmeticians persistently concentrated on population statistics, whereas, during the time from Achenwall to Bluntschli (1867, p. 152), the subject of the Staatswissenschaft narrowed. The latter did not anymore include climate or political structure in its province, nor did he equate Staatswissenschaft with statistics.[13] Another innovation consisted in that the previous stress on political problems was even earlier replaced by focussing Staatswissenschaft on both politics and economics (on both government and society). Knies (1852, p. 654) connected this second change with Schlözer' s statement.

Several Russian statisticians sided with Schlözer. Herrmann (1809, p. 47) argued that statistics *did not judge* and Poroshin. (1838, p. 101) declared that the statistician was *not obliged to treat causes or consequences*. In France, Delambre (1819, p. LXVII) argued that statistics *exclut presque toujours les discussions et les conjectures*.[14] His next phrase was also significant:

L' arithmétique politique … doit aussi étre distinguée de la statistique.

And the London (later, Royal) Statistical Society declared (Anonymous 1839, p. 1) that statistics does not discuss causes, nor reason upon probable effects[15] so that all conclusions *shall admit of mathematical demonstrations* (p. 3). The second proposition was also too restrictive. It is now known that statistics can reject a hypothesis (stochastically rather than in the implied sense of the Society's belief), but that it can hardly *demonstrate* a positive conclusion.

The Society then stated that statistics *are connected with* or *enter* into several sciences; and declared that statistics *enter* more or less into every branch of science (Ibid., pp. 2 and 3). Thus, they did not recognize statistics as a separate discipline (witness also the appeared plural form!), nor did they mention mathematics at all. At the same time, they did not agree that statistics *consist merely of columns of figures*, cf. § 4.2. The proponents of the other viewpoint were also numerous. Even in the 18th century Gatterer (1775, p. 15) declared that, just like history ought

Nicht nur das Pourquoi, selbst auch das Pourquoi von dem Pourquoi zu erforschen, so wird es auch bey der Statistik ... den gegenwärtigen Zustand eines Staat aus dem vergangenen Zustände begreiflich zu machen.

The author likely borrowed his French expression from Leibniz (Krauske 1892, p. 680 with no reference provided):

Das Warum des Warum hätte sie [Sophie Charlotte, Queen of Prussia] *am liebsten ergründet, wie Leibniz meint*.

Obodovsky (1839, p. 113) and Knies (1852, p 661) argued that the Schlözer definition of statistics (§ 2.3) was inadequate because it excluded the study of causes. Vernadsky (1852, p. 223) quoted the full title of Dufau (1840) which strengthened his conclusion (pp. 224 – 225) that the [main] goal of statistics was to discover the appropriate laws.

Knies (1850, p. 68) quoted the same title and cited from Dufau (p. 144):

La statistique a pour objet de conduire … à la découverte des lois d'aprés lesquelles se développent les faits sociaux.

Also see Niemann's opinion in Note 26.

2) The method. Are numbers sufficient for describing a state? This issue is related to the application of mathematics to other fields of knowledge. A statistician (say) should work together with a mathematician rather than reject the 1atter's science. Numbers met with resolute opposition. In 1806, the *Göttingische gelehrte Anzeigen* (John 1883, p. 670; author's name not given) denounced political arithmeticians who had *attempted to make others believe that one could ascertain the importance of a state* by numbers. John also quoted another source from 1811, again without naming the author, who had stated that measuring *everything by numbers* was *supremely ridiculous*.

A strange utterance was due to Mone (1824), see Knies (1850, p.„80):

Une description simple va au fond des choses, tandis que les nombres et les täbleux [§ 4.2] *s'arrétent a la surface*.

Then, Roslavsky (1839, pp. 181 – 182) thought that the *number element* was an innovation [?] to be admitted *with greatest caution*.[16]

On the other hand, not only political arithmeticians, some statisticians[17] and the partisans of the numerical method (§ 4.1) recognized the primary importance of numbers. In 1838, Humboldt (Knies 1850, p. 145) declared that

Im politischen Haushalt, wie bei Erforschung von Natur-erscheinungen sind die Zahlen immer das Entscheidende; sie sind die letzten unerbittlichen Richter.

Provided that the numbers are not objected to, there still remains the issue of mean statistical indicators. It was Quetelet (§ 3.2) who introduced both them, and, actually, probabilities into statistics even beyond demography. Vernadsky:(1852, p. 228) declared that statistical laws were *only revealed in mean values*, and Yanson (1871, p. 264) more correctly argued that

The entire aim of statistics consists in constructing mean values and studying deviations from them.[18]

In 1850, bearing in mind natural sciences, Humboldt (Sheynin 1984b, p. 68, note 36) stated that *die einzig entscheidende Methode* was that of the *Mittelzahlen*. However, even if a statistician accepts the traditions of political arithmetic, he (Knies 1852, p. 660)

Verwirft aber entweder die Anwendung der Wahrscheinlichkeits-rechung [actually: of means], *weil ihr Ergebnis kein adäquater Ausdruck der Wirklichkeit ist, oder will zum Mindesten die Operation des Statistikers keinenfalls auf sie beschränkt wissen*.

Also see Note 17.

And so, the main conclusions concerning the first half of the 19th century are

a) The battle between political arithmeticians and the followers of Achenwall and Schlözer was not yet decided.

b) The theory of probability hardly entered statistics.

A special remark: Knies (1850) reasonably inferred that only political arithmetic represented statistics proper, and he even called this discipline *mathematische Statistik* (p. 163). Vernadsky (1852), who referred to Knies on another occasion, repeated this term on his p. 237.

3. Quetelet

3.1. His work. Quetelet followed the traditions of political arithmeticians. He (1846, p. 351) recommended to study the changes brought about by the construction of telegraph lines and railways and even to work out ideas on the future of the *population* (1829, p. i; 1843, p. 27).[19] Then he (1845, p. 207) distinguished between constant, variable (in particular, periodic) and accidental causes[20] and stated that statisticians should eliminate the last-mentioned ones and study the constant and variable causes. Here is his example. A small number of the newborn is not registered and the ensuing mistake in the sex ratio at birth is random. If, however, some newborn sons are concealed in order to save them from military service, the corruption becomes variable, depending on the chances of war (1846, p. 193).

Another illustration concerned meteorology. Confirming a contemporaneous and correct belief in that fair or foul weather tended to last (1852, p. 57), Quetelet examined this fact by applying elements of what is now called the theory of runs.

Preceding Quetelet by a few years, Cournot (1843, §§ 104 and 117) isolated causes acting in the same way on a whole series of trials, and in the second case he referred to Bienaymé, see Heyde & Seneta (1977, p. 43). Apparently no one mentioned astronomy where systematic errors (and therefore systematically acting causes) were undoubtedly known to Tycho if not to Ptolemy. Quetelet also quantitatively studied the significance of causes and Yule highly praised his reasoning (Sheynin 1986, § 5.6).

3.2. Mean inclinations. Quetelet introduced the [mean] inclinations to crime and to marriage. Subsequent statisticians vigorously attacked the former; Rümelin (1867, p. 25) declared himself free of any such inclination, but, at the same time, he agreed that the conclusions derived from an appropriate mortality table were valid for him also.

He was thus inconsistent. Moreover, unlike Quetelet, e. g. (1848, pp. 82 and 93), he apparently had not noticed that mean statistical indicators did not necessarily apply to any given person.[21]

3.3. The theory of probability. Fourier[22] rather than Laplace (as it is usually stated) influenced Quetelet. Even when popularizing the theory of probability, Quetelet never mentioned the latter's main stochastic tool, the central limit theorem; in general, he made many a nice pronouncements, e. g. (1869, t. l, p. 134), on the theory but used it rather seldom and unprofessionally. Thus, when introducing his Average man, he did not mention the Poisson law of large numbers although he cited it on another occasion (1846, p. 216).[23]

True, Quetelet (1869, t. l, p. 112) argued that, because of large mistakes in the data, *mathematical corrections* were useless. Indeed, the situation was likely dangerous, but only mathematics can detect such errors, at least to some extent. More precisely, it is the modern exploratory analysis, an important chapter of theoretical statistics, that critically appraises the data. This analysis is largely informal, and, what is noteworthy, Quetelet himself repeatedly advocated such studies, see for example his *Lettres* (1846), and revealed gross errors and deliberate distortions in official figures. His reference to mathematical corrections was thus unclear but it possibly reflected a widespread feeling.

Quetelet's *simplest* procedure here (1846, pp. 199 and 308 – 311) was to separate the data into groups and to study these one by one. Closely linked with the examination of data is the estimation of their plausibility whose importance in those times was apparently not yet duly recognized (cf. § 4.2.5); for one thing, the variance was then hardly used beyond the theory of errors as a measure of unreliability. Indeed, in 1869 the International Congress of Statistics (*Congrès* 1870, p. 534) recommended that statistical investigations be accompanied by information on the number (n) of observations (x_i,) made and on the differences ($x_i - x_j$), $i \neq j$, and [$x_i - (1/n)\sum x_i$]. The variance was obviously still lacking!

One more point was there included: It was *desired* to indicate both the *means* and the number of *oscillations* so as to calculate the *mean deviation* of the results of a series from its mean. This advice was formulated awkwardly, but it seems that at least a subdivision of observations into series was thought about.

All these recommendations followed upon the statement of the preceding Congress (*Congrès* 1868, p. 6) that *statistical questions* should have their *scientific base* in mathematics and should be studied *in direct connection with the theory of probability*.

Quetelet who died m 1874, is known to have greatly influenced the work of these congresses which regularly took place from 1853 to 1876.[24]

3.4. Conclusions. Pearson (1924, p. 420) highly praised Quetelet's achievements *in organizing official statistics in Belgium and in unifying international statistics* and called him *the great Belgian statistician* (Ibid., p. 12). Indeed, Quetelet was a highly respected scholar who originated moral statistics,[25] shaped statistics into an important scientific discipline and remained its main representative for a few decades. As the material above (aims of statistics; initial examination of data; study of causes) testifies, his work undoubtedly belonged to political arithmetic or statistics proper.

At the same time, he was careless when teaching his conclusions. Thus, he wrongly claimed that the [relative] crime figures were stable. Moreover, Quetelet did not adequately qualify his statements and his attractive style diverted the attention of his readers from studying his contributions more attentively. Indeed, how else can we explain that subsequent statisticians mistakenly assumed (§ 3.2) that his mean inclination to crime was directly applicable to any individual? Consequently, suchlike indicators (and probabilities in general) associated with *Queteletism*, along with the theory of probability in its entirety, became regarded with suspicion as an abstract theory.

4. Empiricism

4.1. The Numerical method. This is usually attributed to the French physician Louis. He and his followers strove to base themselves on statistical data rather than on *doubtful, incomplete or wrong information* [faits] (Louis 1825, p. XVII), but they hardly ever applied stochastic considerations or studied causes. This method, whose advocates had actually appeared in medicine by the end of the 18th century (Black, in 1788), remained in vogue until ca. 1850.

Furthermore, it was actually used in other branches of natural sciences as well, for example in biology (Aug. De Candolle, in 1832). The spread of cholera was explained (Snow, in 1855) just by subdividing the data m two groups.

At the same time, however, this attitude was not sufficient. Thus, astronomers were unable to abandon theories on the structure of the stellar system although Proctor, in 1872, had plotted 324 thousand stars on his charts with the express goal of doing away with all such theories.

Another aspect of the numerical method was the desire to amass observations. Pettenkofer, in 1886 – 1887, who hardly recommended it openly, published a monstrous survey of writings on cholera epidemics adducing a large number of graphs and tables, but he was unable to analyse his data.

At least in natural sciences the abundance of materials led to the wrong opinion that mass heterogeneous data were better than a small number of reliable observations. In 1847 – 1848, Simpson collected information on mortality from amputations pertaining to many hospitals for the period 1749 – 1846 and declared that this fact ensured trustworthy conclusions for the future. More definitely, William Herschel, in 1817, argued that the size of any star of the first seven magnitudes *promiscuously chosen* out of the 14 thousand of them will not be *likely* to differ *much* from a *certain mean size of them all*. He naturally did not know that with respect to their size stars enormously differed one from another, cf. Note 21.

4.2. Empiricism. In actual fact, the numerical method originated in 1741 with Anchersen, when *statisticians* have begun to describe states in the tabular form (as foreseen by Leibniz, see § 2.2) and thus facilitated the use of numbers. Achenwall, however, opposed tabular statistics which was in essence a connecting link between the

Staatswissenschaft and political arithmetic. It continued to be victimized (Knies 1850, p. 25; also Knies 1852, p. 658):

Man unterschied zwischen höherer [Achenwallian] *und gemeiner Statistik … bis auf unsere Tag*.

At least by implication, the latter kind of statistics was not only its tabular form, but political arithmetic as well. Recall (§ 2.1) that Poisson apparently identified political arithmetic with the *haute statistique*!

From 1827 onward the French Ministère de la Justice had been publishing its yearly *Compte général …*,which led to the origin of criminal statistics. In general, those who collected and applied statistics of some kind rapidly increased in number so that some confusion had to occur[26] and later statisticians (Mayr 1874, p. 26) had to deny the purely empirical approach:

Die amtlichen Statistiker seien zu etwas Höherem als zu bloßen Tabellenknechten berufen.

Nevertheless, Chuprov (1903, p. 42; 1905, p. 422) accused practitioners (perhaps even some of those who kept to political arithmetic):

a) *Such statisticians who observe without thinking about the <u>why</u> or the <u>how</u>, who make most involved computations without understanding where all their multiplications and divisions might and will lead them, are extremely numerous. And statistics has to thank them for its ill fame*.

b) *Allgemein anerkannte Prinzipien, an denen die Richtigkeit der Schlüsse und die Zweckmäßigkeit der angewendeten Methoden geprüft werden könnten, gibt es gegenwärtig in der Statistik gar keine*.

Sei es nun auch bloß, um die Individualwillkür einzuschränken und den Ergebnissen der statistischen Forschung das wichtigste Abzeichen der Wissenschaftlichkeit, die Allgemeingültigkeit, zu verschaffen, muss demnach das gegenwärtige empirische Verfahren der Statistiker rationalisiert werden.

4.3. Biometry. To make the next step: Anderson and Chuprov properly accused the Biometric[27] school of empiricism (Sheynin 2011, § 15.3). Thus, Chuprov (1918 – 1919, pp. 132 – 133) stated that *English scientists* had avoided the concepts of probability and expectation (restricting their attention to the appropriate empirical indicators)[28] and concluded:

Nicht <u>Lexis gegen Pearson</u>, sondern <u>Pearson durch Lexis geläutert, Lexis durch Pearson bereichert</u> sollte gegenwärtig die Parole deren lauten, die, von der geistlosen Empirie der nachqueteletischen Statistik unbefriedigt, sich nach einer rationellen Theorie der Statistik sehnen.

5. Statistics in the second half of the 19th century

5.1. The general situation. Fourier (§ 2.1) and Cauchy (Ibid., Note 9) declared that statistics should be based on mathematics and Poisson (§ 2.1) resolutely stated that the theory of probability ought to be this foundation. Cournot (1843, § 105) argued that statistics should be applied to phenomena both of the *ordre physique et naturel* and of the *ordre social et politique*. And in § 113 he added, without any qualification, that the theory of probability was not *indifférente* for statisticians. His §§ 111 – 114 meant that stochastic calculations should be accompanied by statistical reasoning.

Quetelet is known to have considered his Average man as a specimen of both physical and moral qualities; he thus believed that, at least to a certain extent, stochastic reasoning was applicable to sociology.[29] And here are a few related pronouncements. Chr. Bernoulli (ca. 1842, p. 17):

Das numerische Verhältnis der jährlichen Trauungen, der unehelichen Geburten oder der Findelkinder variirt fast überall weniger als das der Sterblichkeit ...

Guy (1885, p. 85) repeated his earlier opinion of 1869:

Subject to many exceptions, ... events brought about by physical causes are subject to greater fluctuations ... than events in the production of which the will bears a part.

Cournot (1843, § 118) even thought that the figures of criminal statistics were more stable than those determined by the *concours des forces aveugles de la nature* because the causes of different crimes were virtually independent one from another. His conclusion hardly followed from this independence (which in itself was dubious).

Bortkiewicz (1894 – 1896, p. 356) stated that

Diejenigen statistischen Größen, die sich in die Schemata der Wahrscheinlichkeitsrechnung am ehesten fügen lassen, zugleich solche sind, die meistens kein großes materiell-statistisches Interesse in Anspruch nehmen können.[30]

He was thus sceptical, but on p. 360 he added that a *Richtung der theoretisch statistischen Forschung* recognized that the *Massenerscheinungen der menschlichen Gesellschaft* can only be understood *aus den Prinzipien der Wahrscheinlichkeitsrechnung.*[31]

Bortkiewicz linked his first passage with the concluding pages of Lexis (1879) whose exposition was there rather indefinite; he likely thought about the case of constant probabilities (see § 5.3.2). Here, I note that Bortkiewicz (1904, p. 241) and Chuprov (1905, p. 424), without providing the exact reference, remarked that, according to Lexis, the *menschliche Handeln* fell *ganz außerhalb des Rahmen der Naturgesetzlichkeit*.

Chuprov (pp. 473 – 474n) also stated that *previous* theoreticians of statistics were inclined to contrast nature with man and society. The Editor of the Russian edition of Chuprov's collected papers, his student N. S. Chetverikov (Chuprov 1960, p. 85), mentioned here Chuprov's later statements where his teacher argued that this point of view was definitely obsolete. He did not cite anyone (see above).

Lexis (1877, p. 5) stated that statistics was mostly based on probability and that (1903, p. 241) the

Schema der Wahrscheinlichkeitsrechnung was *auch die höchste wissenschaftliche Form in welche die Statistik ihren Stoff fassen kann*.

And (p. 230) the only [but most essential] aim of applying probability was to obtain' *ein verständliches Schema für die Verteilung der Fälle und … einen Maßstab für die Stabilität der statistischen Verhältniszahlen* [§ 6] *zu bieten*.

Lexis (pp. 242 – 243) also reasonably thought that the causes of social phenomena should be ascertained by the use of both stochastic and deterministic methods.

5.2. The subjects not yet/hardly studied by statisticians. Several important subjects were not yet, or hardly studied in the 19th century which partly explains the situation in statistics.

1) Correlation. Galton introduced the concepts of regression and correlation in the last decades of that century but Continental statisticians did not then become interested. True, in 1912, in Russia, Slutsky published an important treatise explaining the correlation theory to his fellow-countrymen. Kolmogorov (1948) thought that this contribution was still important, but then, in 1912, Markov opposed it (Sheynin 2011, § 7.4.1). Markov (1916, pp. 200 – 202) again criticized the correlation theory, and Linnik, in a modern commentary to Markov, agreed in that correlation was then not yet sufficiently developed.

It is noteworthy that Seidel and Kapteyn pyblished findings related to correlation theory. I mentioned the former in § 1 (Note 2); and the latter, in 1912, being dissatisfied with Galton's work, introduced his own correlation coefficient for applications in astronomy.

Kaufman (1922, p. 152) stated that the

So-called correlation method … does not essentially add anything to the results of an elementary analysis.

It seems however that at least a comparison of two cases, of two appropriate correlation coefficients, could have been interesting for understanding where the connection between factors was tighter. And the sign of the coefficient is also important.

2) Sampling. Its prehistory goes back to the 12th century when the checking of the quality of coinage began in England (Stigler 1977). Simplest sampling estimations of harvest were made in Russia from the 17th century onward (Ptukha 1961). Laplace (§ 2.1) applied sampling for estimating the population of France, and Buniakovsky, in 1846 (Sheynin 1991b, p. 211), in essence repeated his study. Fries (1842, p. 148), Knies (1850, pp. 152 – 153) and Kries (1886, p. 241) utterly rejected Laplace's investigation whereas Cournot (1843) passed it over in silence. Knies also remarked that sampling became widespread in the *nächstvergangenen Zeit*. Fourier (1821b, p. XXXVIII) mentioned sampling without any misgivings, but later authors, just as those referred to above, were critical. Moreau de Jonnès (l847 pp. 53 – 54) cited Vauban, a military engineer and a marshal of France, who, at the beginning of the 18th century, had examined the agricultural production of France by sampling and argued that this procedure *semble étrange aujourd'hui*. Quetelet (1846, p. 293) thought that sampling should be avoided and John (1896, p. 39) mentioned it disapprovingly.

Bortkiewicz (1904, p. 252) and later on Chuprov, see below, stated that Mayr was opposed to sampling. This pronouncement seems plausible, but I am unable to confirm it. Bortkiewicz (1901, p. 825) also argued that the compilation of data in the 19th century brought about the decline of sampling (of *Konjekturalberechnung*, as he called it)**[32]** but he also mentioned the pioneer work of Kiaer. In 1904 and even in 1893 Kapteyn began to consider the stellar universe as a stochastic entity and in 1906 he initiated its international sampling study.

It seems that neither this example, nor the progress in medical statistics, which had to make use of inductive methods, influenced statisticians engaged in sociological studies.

Chuprov (1906, p. 706) stressed the importance of sampling for obtaining materials concerning the time period between adjacent censuses, but he warned his readers against arbitrary assumptions, His father (Chuprov, A. I., 1894) upheld sampling as a means for supplementing previous full-scale statistical investigations.[33] In 1910 A. A. Chuprov delivered a popular report on sampling (Sheynin 1997a). He indicated (p. 662) that Mayr and several other like-minded statisticians had opposed this procedure. Although he did not say so directly, the context of his report clearly shows that Chuprov reasonably believed in a great future for sampling.

Czuber (1921, p. 13) still associated sampling with *Konjektural-rechnungen* and apparently regarded it as an obsolete method.

3) The study of public opinion. This is inseparably linked with sampling. Small (1916), who reviewed sociology in the USA during 1865 – 1915, did not even mention such investigations (and hardly discussed statistics).

4) Statistical control of quality began in the 1920s (Shewhart 1931). True, in 1848 Ostrogradsky suggested sampling for checking the quality of goods supplied in batches but his recommendation was hardly noticed (Sheynin 1991b, pp. 206 – 207).

5) Estimation of precision. Until the 1920s statisticians had not studied it. Even Cournot (1843) actually followed Fourier rather than Gauss and did not therefore directly describe the use (or the benefits) of the variance (for the case of repeated observations). Kries (1886, p. 180) remarked that the probable error of weighing a chemical substance with a sufficiently precise balance was of no significance and that, in general, the calculation of *aller Wahrscheinlichkeiten* might be *sogar irreführend.*

Mayr (1914, p. 46) did not recognize *verfeinerte* investigations of plausibility which were still unable, as he declared, to distinguish between errors and deviations from the assumed model.

And here is _Bortkiewicz (1894 – 1896):

Die Präcision einer statistischen Größe ist stets als etwas Accessorisches anzusehen (p. 353); the *statistische Sinne* hardly ever errs so that the calculation of precision in each case is a *Luxus*; and the value of the probability that characterizes our conclusions is not important if it is not sufficiently close to unity (p. 354). Only the last phrase seems to be quite correct.

The *statistical feeling* should apparently be supplemented by considering the results of similar previous investigations whose precision had since became known; by subdividing the data into groups and comparing these one with another; and, generally, by critically examining the data, cf. § 3.3. And I quote Chr. Bernoulli's sound opinion (ca. 1842, p. 6):

Eben so gewiss aber werden wir eine vorhandene Gesetzmäßigkeit in diesem Sinne [of large numbers] *anerkennen müssen, so oft wir nach solchen Beobachtungen Ergebnisse sich beständig und in demselben numerischen Verhältnis wiederholen sehen, und zwar wie sehr auch die Ereignisse vom Zufall oder dem menschlichen Willen abzuhängen scheinen, oder wie sehr wir über die eigentliche Ursache im Dunkeln sein mögen*.

I still have to refer to Kaufinan (1913, p. 105) who described a modern-looking procedure without however providing any examples or explaining it in any detail. He recommended to construct an *empirische Abweichungskurve* and to compare it with the appropriate *theoretisch berechneten Kurven*.

6) Econometrics originated only in the 1930s and united statistics with the appropriate branches of economics and mathematics. The disputes about the relation between statistics and political economy (economics) became meaningless.

5.3. Difficulties (real and imaginary) of applying the theory of probability. There existed real and imaginary difficulties in applying the theory of probability. When discussing the restrictions governing the theory, statisticians invariably stated that the trials of a given series should be [mutually] independent; the probabilities of *success* in these trials should be constant. The very concept of probability, based on the existence of equally possible cases, was hardly fit for applications; the theory was too abstract; and its law of large numbers was of little use.

I take up these points one by one and I show that in the first two cases some thought was given to broadening the scope of the application of probability.

l) I list three items. a) The problem of the extinction of families (that is, the prehistory of simple branching processes) goes back to Bienaymé (Heyde & Seneta 1977, § 5.9). b) Kries (1886, pp. 242 – 243) remarked that the expectations of life were not independent for man and wife. c) Directly applicable to my subject is Cournot's exposition (1843, §§ 104 and 117): he knew that consecutive trials often depended one on another so that much more observations were then needed than in the case of independence, and he mentioned *l'économie sociale* and natural sciences. In § 206ff Cournot even attempted to examine the likely dependence between the decisions made by the jurors in courts of law.

2) Cournot also devoted a (short) chapter to the *variability of chances*. He did not consider there any statistical examples (nor had he referred to Poisson), but this was a hardly sufficient cause for the statisticians to ignore his study (as they obviously did).

Cournot (§ 116) recommended to partition series of trials into groups so as to reveal whether the appropriate probability was constant. He referred to Bienaymé whereas Bru, the Editor of Cournot's latest edition, additionally cited Fourier.

John (1896, p. 39) and then Czuber (1899, p. 231) noted that a number of conditions should remain permanent for a relative statistical indicator to be (some function of) a [constant] probability and that this circumstance (indicated by Bienaymé and Cournot, as

Czuber (1921, p. 35 – 36) added later) was not taken into account by previous workers.[35] He did not say that, irrespective of probabilities, a change in the relevant conditions depreciated the collected data.

Knapp (1872, p. 117) argued that

Man braucht mehr als nur die Urnen des Laplace mit bunten Kugeln zu füllen um eine theoretische Statistik herauszuschütteln. Die Anwendung auf Bevölkerungsstatistik liegt im Argen: denn es fehlt hier alle Ähnlichkeit der Bedingungen.

Neither he, nor Guerry (1864, p. XXXIV), who preceded him by making a (weaker) statement similar to the first half of the passage just above, mentioned constant probabilities, but they hardly thought of anything else (for example, of statistical control of quality, § 5.2.4, connected with variable probabilities). And Mayr (1914, p. 45)[36] declared that stochastic interpretation of statistical studies was only possible for repeated observations of one and the same object[37] or when they were similar to urn problems concerning constant probabilities.

I ought to add that Lexis naturally (§ 6) did not restrict his attention to the case of constant probabilities:

Schema einer konstanten Wahrscheinlichkeit … nur in den seltensten Fällen auf die menschlichen Massenerscheinungen passt, as Bortkiewicz (1904, p. 247) argued, and added that for this reason Lexis *kein besonderes Gewicht auf dieselben legt*. See the related material concerning Lexis in § 5.1.

3) Equally possible cases. Cournot (1843, § l8) heuristically introduced geometric probability,[38] applied it for defining the density curve (§§ 65 – 66) and described the latter's use in statistics (§ 125). Statisticians however hardly took notice although the issue of equally possible cases was thus softened. Kries (1886, p. 6) justified equally possible cases by the *Princip des mangelnden Grundes*, as he named it.[38a] He (p. 64) then attempted to avoid such cases by noting that in a game of roulette *kleinen Variirungen der Bewegungs-Modi* of the rotating ball were sufficient for a considerable effect.[39] Therefore, as he concluded on an intuitive level, the occurrences of *red* and *black* were equally probable, and, more generally, the appearance of the uniform distribution in most various cases was proved. Poincaré (1896) independently explained the last-mentioned conclusion by introducing the important notion of arbitrary functions, which allow the modern theory of probability to keep to the concept of equipossibility, and he additionally considered such phenomena as the uniform distribution of the ecliptic longitudes of the minor planets (Sheynin 1991a, § 8).

Lexis (1886, p. 436) favourably described and enlarged on Kries but he mistakenly believed that the uniform distribution was necessary for justifying the theory of probability. He (p. 437) also inferred that, because of the *equally possible cases* (which haunted him even much later (1913, p. 2091), the theory of probability was a subjectively based discipline (see below).[40] For some reason he did not repeat here his earlier statement (1877, p. 17) to the effect that equally possible cases might be presumed when a statistical probability tended to its theoretical counterpart, when this later was justified by experience.

Still earlier, apparently in 1874, he (1903, pp. 241 – 242) indirectly argued that the existence of such cases was necessary for the *Schema der Wahrscheinlichkeitsrechnung*.

Cournot (1843, § 86) was perhaps the first who strongly argued that the *Bernoulli principle* allowed the practitioners to apply empirical frequencies instead of probabilities. Nevertheless, mathematicians naturally remained unhappy and even .in the mid-20th century Khinchin [xiii, p. 104] noted that *each author … invariably spoke about equally possible cases attempting however to leave this annoying subject as soon as possible*.

Perhaps the situation in the natural sciences was more difficult than in statistics. Thus, Langevin (1913, p. 3) thought that, for the kinetic theory of gases, *the main difficulty* consisted in providing an *appropriate* (*correcte*) *and clear definition of probability*.

The axiomatic theory of probability naturally does not help the practitioner here so that the theoretically imperfect frequentist theory *found* [and continued to find] *a large number of partisans among both mathematicians and, especially, representatives of natural and applied sciences and in particular among physicists* (Khinchin, p. 101).

These considerations do not exonerate Soviet statisticians, who, for a few decades, had been denying the theory of probability and justifying their attitude by the lack of equally possible cases in the national economy. They were doing their best to protect Marxist dogmas against the pernicious influence of the contemporaneous science (Sheynin 1998, §§ 3.5 and 5).

Already Poisson (1837, pp. 30 – 31) distinguished between objective and subjective probabilities. Cournot (1843, §§ 44 and 46) believed that the term probability usually implied a subjective sense and thought (§§ 86 and 240.4) that there existed a *distinction fondamentale* between the two versions of probability. He also called subjective probabilities *philosophical* (§ 233).

Referring to Kries, Bortkiewicz (1894 – 1896, p. 661) argued that *Jene Unterscheidung zwischen objektiven und subjektiven Wahrscheinlichkeiten anerkanntermaßen nicht stichhaltig ist, weil jede gegebene Wahrscheinlichkeit einen bestimmten Wissens- oder Unwissenheitszustand voraussetzt und in diesem Sinn notwendig subjektive ist*.

Chuprov resolutely disagreed: *A difference, and not a small one does nevertheless exist*. This was a marginal note written (in German or Russian ?) on his copy of Bortkiewicz's paper and now adduced to the Russian translation of the latter (Chetverikov 1968, p. 74).[41]

Subjective probabilities are still with us, but they do not underpin the theory of probability.

4) The history of mathematics proves that the more abstract it becomes, the more fruitful are its applications. However, some statisticians had been complaining that the theory of probability was too abstract. Such was the opinion of Block (1886, p. 134) who also argued that it should not be employed *too often*.

Consider now Knapp's arguments (1872). First (p. 115), he doubted that *dieser schwierige Calcul nützlich gemacht werden kenne* beyond games of chance or insurance. He then declared that *even* and *independence* were *mangelhaft umschreibenden Begriffe*.[42] It is true that, once a mathematical theory becomes axiomatized, its notions should be somehow interpreted before being practically applied. However, probability had not then approached axiomatization and 1 doubt that Knapp 's objections were really important.

5) Jakob Bemoulli's law of large numbers established the relation between probability and frequency, between deduction and induction. He (1713, Chapt. 4 of pt. 4) attempted to find out whether [the right side of our present formula of his law is unity or some positive proper fraction]. The second would have meant, as he himself effectively argued, that induction was incapable of precise conclusions.[43]

And still, at least one statistician (Haushofer 1872, pp. 107—108) declared that statistics

Steht in gar keiner inneren Beziehung with mathematics; *die Mathematik beruht eben auf der Deduktion, die Statistik auf der Induktion*.

Not only did Haushofer disregard Bernoulli's thoughts; he failed to notice that theoretical statistics was only partly deductive (mathematics in general was not here relevant).

Many statisticians, even without mentioning philosophical concepts, denied the applicability of the law of large numbers. Knapp (1872, pp. 116 – 117),[44] for example, alleged that statisticians always made only one observation (as when counting the population of a city) so that for statistics the law *ist von geringerer Bedeutung*! Quetelet's celebrated study (1846), the treatment of the chest measurements of about six thousand soldiers by means of one single distribution, apparently did not concern him at all. Hardly less known is the example of throwing several coins of the same coinage: the relative frequency of heads was here quite stable.[45] Lexis (1879, §§ 6 and 7) applied the same idea for studying series of numbers oscillating around some mean value.

It was in this context that Knapp added that *die Urnen des Laplace* were insufficient for creating theoretical statistics (§ 5.3.2) but he did not explain what else was needed.

Maciejewski (1911) formulated curious thoughts about the law of large numbers.[46] He (p. 95) noted that statisticians understood the law discordantly, and, on p. 96, he introdued *la loi des grands nombres des statisticiens* which stated that

The oscillations of statistical numbers diminishes as the number of observations increases.

He (pp. 94 – 98) also declared that the (real) law did not lead to any remarkable results and even impeded the development of statistics and that the principles of insurance were established before Jakob Bernoulli's discovery.

It is natural to recall here Bortkiewicz's remark (1904, p. 251):

Auch der grimmigste Feind of probability theory *operiert … mit Vorstellungen die gerade diesem Erscheinungsgebiete entstammen*, but *in unmethodischer Weise*!

I did not see any references to Maciejewski's book and came across it by chance.

6. Lexis: stability of statistical series

In §§ 2.1, 5.1 and 5.3.3 1 mentioned Lexis in connection with the examination of expected empirical discrepancies; with ascertaining the appropriate causes and the relations between probability theory and statistics; and with subjective probabilities, respectively. I shall now touch on his study of the stability of statistical series, of discovering possible variations in the probability underlying several consecutive statistical series (1877 and, especially, 1879).[47] Many commentators described these contributions; suffice it to mention Chuprov (below) and Bauer (1955). I also dealt with this subject including the related work of such scholars as Bortkiewicz, Markov and Chuprov (Sheynin 2011, § 14).

Lexis's non-parametric test of stability was based on calculating the ratio (Q) of two expressions for the appropriate variance[48] and he argued that the case $Q < 1$ signified a change of the probability. He also stated that $Q > 1$ characterized interdependence of the series, but he left this case aside. It is understandable that, unlike his followers, Lexis himself did not calculate either the mean value or the variance of his criterion, this problem proved to be very dificult. And he even did not formulate it although he could have done so, because Gauss, when introducing his measure of precision of observations, specifically chose an unbiased statistic (the sample variance, in 1823) and calculated its own variance.

Lexis greatly differed from his contemporaries who never thought of such quantitative studies. Instead of abandoning the case of variable probabilities, he attempted to examine it. Indeed, as noted by Bortkiewicz (1904, p. 248), he (1903, p. 98) had stated
that

Das Interessante in den moral-statistischen Zahlen … ist überhaupt nicht ihre Stabilität, sondern ihre Veränderlichkeit.[49]

Formally speaking, Lexis's study was unsuccessful: Chuprov (1918 – 1919) proved that the use of the criterion Q (or Q^2) was hardly justified.[50] Nevertheless, Lexis originated the Continental direction of statistics,[51] which, as hoped by Chuprov (see end of § 4.3), had merged with the Biometric school thus leading to the creation of a unified mathematical statistics.

Chuprov repeatedly praised Lexis although some of his pronouncements were wrong, as he himself proved (1918 – 1919). 1 reproduce three passages from his Russian contributions and I only mention that another relevant statement appeared in a German source (1905, p. 450).[52]

a) *The coming together* [of probability theory and statistics] *was sketched out by Cournot; and Lexis … accomplished it in a number of writings rich in deep original ideas* (1897, p. 55).

b): *His works, where originality competes with clearness of exposition, opened up the newest epoch in the development of metal statistics* (Ibidem, p. 59).

c) *For a long time, his … investigations … constituted the only source of vivid theoretical thought in our science* (1909/1959, p. 63).

Kries's opinion (1886, p. 287) is also important (although his mentioning the theory of probability instead of theoretical statistics seems wrong):

Darf die Arbeit von Lexis gerade mit Bezug auf die Principien als eine der wichtigsten aus der ganzen Literatur der Wahrscheinlichkeits-Rechnung bezeichnet werden.

Much later Anderson (1932, p. 531) testified that

Unsere (jungere) Generation der Statistiker kann sich kaum jenen Sumpf vorstellen, in welchen die statistische Theorie nach dem Zusammenbruch des Queteletschen Systems hineingeraten war und der Ausweg, aus welchem damals nur bei Lexis und Bortkiewicz gefunden werden konnte.

And, again:

Nur unter den antimathematischen Statistikern Deutschlands fand er [Bortkiewicz] *keinen tieferen Anklang. Doch scheint jetzt hier eine Neubelebung der mathematischen Statistik zu sein*.

Two years passed and Anderson (1934, p. 539) added:

Germanic countries, despite the work of Lexis and his eminent pupil Bortkiewicz, still continue under the influence of von Mayr's empirical school … Only in very recent years has a faint start been made toward the acceptance of the English theories, and, in economic statistics, of American methods.

And he also stated, on the same page: by 1870 – 1890 *statistics came to designate the science of mass* [stochastic] *phenomena in social life*. Cf. the title of Lexis (1877)!

References

Abbreviations

JNÖS =*Jahrbücher für Nationalökonomie und Statistik*
(R) = in Russian

Anderson, O. (1932), L. von Bortkiewicz. In author's *Ausgew. Schriften*, Bd. 2. Hrsg, H. Kellerer, et al. Tübingen, 1963, pp. 530 – 538.
--- (1934), Statistics: Statistical Method. Ibidem, pp. 539 – 544.
Anonymous (1839), Introduction. *J. Stat. Society London*, vol. 1, pp. 1 – 5.
Bauer, R. K. (1955), Die Lexissche Dispersionstheorie in ihren Beziehungen zur modernen statistischen Methodenlehre insbesondere zur Streuungsanalyse. *Mitteilungsbl. f. math. Stat. u. ihre Anwendungsgebiete*, Bd. 7, pp. 25 – 45.
Bernoulli, Chr. (1841), *Handbuch der Populationistik oder der Volkes- und Menschenkunde*. Ulm.
--- (ca. 1842), Einige Worte über anthropologische Statistik. In *Einladungsschrift zur Promotionsfeier des Pädagogiums und Eröffnung des Jahreskurses von 1842*. Basel. Separate paging.
Bernoulli, J. (1713), *Ars Conjectandi*. German transl. Leipzig, 1899. Reprinted: Verlag Harri Deutsch, 1999.
Block, M. (1886), *Traité théorique et pratique de statistique*. Paris. First edition, 1878.
Bluntschli, J. C. (1867), Staatswissenschaft. ln *Deutsches Staatswörterbuch*, Bd. 10. Stuttgart – Leipzig, pp. 152 – 154.
Bortkiewicz, L. von (1894 – 1896), Kritische Betrachtungen zur theoretischen Statistik. JNÖS, 3. Folge, Bd. 8 (63), pp. 641 – 680; Bd. 10 (65), pp. 321 – 360; Bd. 11 (66), pp. 701 – 705.
--- (1901), Anwendungen der Wahrscheinlichkeitsrechnung auf Statistik. *Enz. math. Wissenschaften*, Bd. 1, Tl. 2. Leipzig, pp. 822 – 851.
--- (1904), Die Theorie der Bevölkerungs- und Moralstatistik nach Lexis. JNÖS, 3. Folge, Bd. 27 (82), pp. 230 – 254.

--- (1910), The aims of scientific statistics. *Zurnal Ministerstva Narodnogo Prosveshchenia*, No. 2, pp. 346 – 372 of the second paging. (R)
--- (1915), W. Lexis zum Gedächtnis. *Z. f. Versicherungswissenschaft*, Bd. 15, pp. 117 – 123.
--- (1922), Knapp als Statistiker. *Wirtschaftsdienst*, März (Sonderheft), pp. 10 – 12.
--- (1930), Lexis und Dormoy. *Nordic Stat. J.*, vol. 2, pp. 37 – 54.
Butte, W. (1808), *Die Statistik als Wissenschaft*. Landshut.
Cauchy, A. L. (1821), *Cours d'analyse de l'Ecole Polytechnique*. In author's *Oeuvr. compl.*, sér. 2, t. 3. Paris, 1897.
--- (1845), Sur les secours que les sciences du calcul peuvent fournir aux sciences physiques ou même aux sciences morales. *Oeuvr. compl.*, sér. 1, t. 9. Paris, 1896, pp. 240 – 252.
Chebyshev,'P. L. (1936), *Teoriya veroyatnostey* (The Theory of Probability). Lectures read in 1879 – 1880. Moscow. **S, G,** 3.
Chetverikov, N. S. (1968), Remarks on Lexis (1879). In *O teorii dispersii* [On the Theory of Dispersion]. Coll. papers of different authors, compiled and translated by Chetverikov. Moscow, pp. 39 – 55. (R)
Chuprov, A. A. (1897), Moral Statistics. **S, G,** 35.
--- (1903), Statistics and the statistical method, their vital importance and scientific aims. In Chuprov (1960, pp. 6 – 42). (R)
--- (1905), Die Aufgabe der Theorie der Statistik. *Jahrbuch f. Gesetzgebung, Verwaltung und Volkswirtschaft im Deutschen Reich,* Bd. 29, pp. 421 – 480.
--- (1906), Statistik als Wissenschaft. *Arch. f. Sozialwissenschaft u. Sozialpolitik*, Bd. 5, pp. 647 – 711.
--- (1909), *Ocherki po teorii statistiki* [Essays on the Theory of Statistics]. Moscow. Reprinted in 1910 and 1959.(R) I quote from the last edition.
--- (1918 – 1919), Zur Theorie der Stabilität statistischer Reihen. *Skandinavisk aktuarietidskrft*, t. 1, pp. 199 – 256; t. 2, pp. 80 – 133.
--- (1960), *Voprosy statistiki* [Issues in Statistics]. Collected reprints and translations of articles. Moscow. Compiled and translated by B. I. Karpenko and N. S. Chetverikov. (R)
--- (1999), On the mean square error of the coefficient of dispersion. Manuscript (1916 or early 1917) (R). **S, G,** 35.
Chuprov, A. 1. (Read 1894), On nomographic descriptions of separate settlements. In author's *Rechi i statii* [Speeches and Papers], vol. 1. Moscow, 1909, pp. 228 – 233. (R)
Congrès international de statistique. *Comptes rendus de la ... session* (1868, 1870, 1873). The sessions were held in 1867, 1869 and 1872 respectively.
Cournot, A. A. (1843), *Exposition de la théorie des chances et des probabilités*. Reprinted: Paris, 1984. **S, G,** 54.
Czuber, E. (1899), Die Entwicklung der Wahrscheinlichkeitstheorie und ihrer Anwendungen. *Jahresbericht deutschen Mathematiker-Vereinigung*, Bd. 7, No. 2. Separate paging.
--- (1921), *Wahrscheinlichkeitsrechnung*, Bd. 2. Aufl. 3. Leipzig – Berlin.
D'Amador, R. (1837), *Sur le calcul des probabilités appliqué à la médecine.* Paris.
Delambre, J. B. J. (1819), Analyse des travaux de 1'Académie pendant 1817, pt. mathématique. *Mém. de l'Acad. des sci. de l'Inst. de France*, t. 2, pp. I – LXXII of the *Hist. de 1'Acad.*
De Moivre, A. (1756), *Doctrine of Chances*. London. Previous editions 1718, 1738. Reprint of last edition: New York, 1967.
Druzinin, N. K., editor (1963), *Khrestomatia po istorii russkoi statistiki* [Sourcebook of History of Russian Statistics]. Moscow. Extracts of the sources. (R)
Dufau, P. A. (1840), *Traité de statistique ou théorie de 1'étude des lois, d'après lesquelles se développent des faits sociaux*. Paris.
Enko, P. D. (1889), On the course of epidemics of some infectious diseases. Partial transl. by K. Dietz in *Intern. J. of Epidemiology*, vol. 18, 1989, pp. 749 – 755.
Fallati, J. (1843), *Einleitung in die Wissenschaft der Statistik*. Tübingen.
Fouricr, J. B. J. (1821a), Introduction to *Recherches* (1821).
--- (1821b), Notions générales sur la population. In *Recherches* (1821, pp. IX – LXXIII).
--- (1823), Sur la population de la ville de Paris In *Recherches* (1823, pp. XIII – XXVII).

Fries, J. F. (1842), *Versuch einer Kritik der Principien der Wahrscheinlichkeits-trechnung*. In author' s *Sämtliche Schriften*, Bd. 14. Darmstadt, 1974, pp. 11 – 254.
Gatterer, J. C. (1775), *Ideal einer allgemeinen Weltstatistik*. Göttingen.
Gavarret, J. (1840), *Principes généraux de statistique médicale.* Paris.
Graunt, J. (1662), *Natural and Political Observations Made upon the Bills of Mortality*. Reprinted: Baltimore, 1939.
Guerry, A.-M. (1864), *Statistique morale de I'Angleterre comparée avec la statistique morale de la France*. Paris. Preceded by a two-page *Extrait du rapport sur le Concours pour le prix de statistique fondation,* **Montyon. Dupin, Mathieu, Boussingault, Passy, Bienaymé, rapporteur**.
Guy, W. A. (1885), Statistical development with special reference to statistics as a science. *Jubilee volume, Stat. Soc. of London*, pp. 72 – 86.
Halbwachs, M. (1913), *La théorie de l'homme moyen.* Paris.
Haushofcr, D. M. (1872), *Lehr- und Handbuch der Statistik*. Wien.
Herrmann, K. F. (1809), General theory of statistics. In Druzinin (1963, pp. 45 – 61). (R)
Heyde, C. C., Seneta, E. (1977), *Bienaymé.* New York.
John, V. (1883), The term statistics. *J. Roy. Stat. Soc.*, vol. 46, pp. 656 – 679. Translated from German (Bern, 1883).
--- (1896), Statistik und Probabilität. *Allg. stat. Archiv*, 4. Jg., pp. 1 – 46.
Johnson, N. L., Kotz, S., Editors (1997), *Leading Personalities in Statistical Sciences*. New York.
Kaufman, A. A. (1913), *Theorie und Methoden des Statistik.* Tübingen. Russian editions: 1912, 1916, 1922, 1928.
Knapp, G. F. (1872), Quetelet als Theoretiker. JNÖS, Bd. 18, pp. 89 – 124.
Knies, C. G. A. (1850), *Die Statistik als selbstständige Wissenschaft*. Kassel.
--- (1852), Die Statistik auf ihrer jetzigen Entwickelungsstufe. *Gegenwart*, Bd. 7, pp. 651 – 688. Appeared anonymously, attributed to Knies by E. Blenk in *Biographisches Jahrbuch und Deutscher Nekrolog*, Bd. 3. Berlin, 1900, p. 112.
Kolmogorov, A. N. (1948), Slutsky. An Obituary (R) *Math. Scientist*, vol. 27, 2002, pp. 67 – 74.
Krauske, O. (1892), Sophie Charlotte. *Allg. deutsche Biogr.*, Bd. 34, pp. 676 – 684. Reprinted: Berlin, 1971.
Kries, J. von (1886), *Die Principien der Wahrscheinlichkeitsrechnung*. Freiburg i. B. Second edition: 1927.
Lambert, J. H. (1772), Anmerkungen über die Sterblichkeit, Todtenlisten, Geburthen und Ehen. In author's *Beyträge zum Gebrauche der Mathematik und deren Anwendung*, Tl. 3. Berlin, pp. 476 – 569.
Lang, J. Scott (1992), Statistical methods. In *Enc.of Sociology*, vol. 4. Eds. Edg. F. and. Mareé L. Borgatta. New York, pp. 2083 – 2090.
Langevin, P. (1913), La physique du discontinu (1913). In collected papers of different authors *Les progrès de la physique moléculaire*. Paris, 1914, pp. 1 – 46.
Laplace, P. S. (1781), Sur les probabilités. *Oeuvr. Compl.*, t. 9. Paris, 1893, pp. 383 – 485.
--- (1812), *Théorie analytique des probabilités*. *Oeuvr. Cornpl.*, t. 7, No. 2. Paris, 1886.
--- (1814), *Essai philosophique sur les probabilités*. I refer to its English translation (New York, 1995).
Lazarsfeld, P. F. (1961), Notes on the history of quantification in sociology. *Isis*, vol. 52, pp. 277 – 333.
Le procès (1900), *Le procès Dreyfus devant le Conceil de Guerre de Rennes 1899*, tt. 1 – 3. Paris.
Lexis, W. (1875), *Einleitung in die Theorie der Bevölkerungsstatistik*. Strassburg.
--- (1877), *Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft*. Freiburg i. B.
--- (1879), Über die Theorie der Stabilität statistischer Reihen. In Lexis (1903, pp. 170 – 212). There, however, a source published in 1898 is mentioned.
--- (1886), Über die Wahrscheinlichkeitsrechnung und deren Anwendung auf die Statistik. JNÖS, 2. Folge, Bd. 13 (47), pp. 433 – 450.
--- (1903), *Abhandlungen zur Theorie der Bevölkerungs- und Moralstatistik.* Jena. In this collection, I refer to his work of 1879 and to

a) Über die Ursachen der geringen Veränderlichkeit statistischer Verhältniszahlen, pp. 84 – 100. A source published in 1901 is mentioned there.
b) Naturgesetzlichkeit und statistische Wahrscheinlichkeit, pp. 213 – 232. A source published in 1899 is mentioned there.
c) Naturwissenschaft und Sozialwissenschaft (Rede, 1874), pp. 233 – 251.
--- (1913), Review of Kaufman (1913). *Jahrbuch für Gesetzgebung, Verwaltung und Volkwirtschaft in Deutschen Reiche*, Bd. 37, pp. 2089 – 2092.
Libri-Carrucci, G. B. I. T., rapporteur (1834), Au nom d'une Commission. Report on Bienaymé's manuscript. Members of the Commission: S. F. Lacroix, S. D. Poisson. *Procès verbaux des séances, Acad. des sci*. [Paris], t. 10, pp. 533 – 535.
Littrow, J. J. (1842), Wahrscheinlichkeitsrechnung. In *Gehler's Physikalisches Wörterbuch*, Abt. 2, Bd. 10. Leipzig, pp. 1181 – 1251.
Louis, P. Ch. A. (1825), *Recherches anatomico-pathologiques sur la phtisie*. Paris.
Lueder, A. F. (1812), *Kritik der Statistik und Politik.* Göttingen.
Maciejewski, C. (1911), *Nouvaux fondements de la théorie de la statistique*. Paris.
Markov, A. A. (1916), On the coefficient of dispersion. In author's *Izbrannye Trudy* (Sel. Works). No place, 1951, pp. 523 – 535. **S, G,** 5.
Mayr, G. von (1874), *Gutachten über die Anwendung der graphischen und geographischen Methode in der Statistik*. München.
--- (1906), *Begriff und Gliederung der Staatswissenschaften.* Tübingen. Other editions: 1901, 1910, 1921.
--- (1914), *Theoretische Statistik*. Tübingen. First edition, 1895.
Mona, F. J. (1824), *Theorie der Statistik*. Heidelberg. French translation, 1834.
Montmort, P. R. (1713), *Essay d'analyse sur les jeux de hasard*. Paris. First edition, 1708. Reprint of second edition: New York, 1980.
Moreau de Jonnès, A. (1847), *Eléments de statistique*. Paris.
Niemann, A. (1807), *Abriss der Statistik und der Statenkunde*. Altona.
Obodovsky, A. G. (1839), *The Theory of Statistics*. In Druzinin (1963, pp. 105—137). (R) **S, G,** 88.
Ondar, Kh. O. (1971), On the works of A. Yu. Davidov in probability theory. (R) **S, G,** 5.
Pearson, K. (1924), *Life, Letters and Labours of Galton*, vol. 2. Cambridge.
Pfanzagl, J., Sheynin, O. (1997), Süssmilch. In Johnson et al (1997, pp. 73 – 75). Also in second edition of *Enc. of Stat. Sciences*. Hobokan, NJ, 2006, vol. 13, pp. 8489 – 8491, somehow appeared anonymously.
Pirogoff, N. I. (1854), *Statistischer Bericht über alle meine … Operationsfälle*, this being the author's *Klinische Chirurgie* No. 3. Leipzig.
Poincaré, H. (1896), *Calcul des probabilités*. Second edition 1912, reprinted 1923 and 1987.
Poisson, S. D. (1836a), Sur le loi des grands nombres. *C. r. Acad. des Sci.* Paris, t. 2, pp. 377 – 382.
--- (1836b), Sur le calcul des probabilités. Ibidem, pp. 395 – 400.
--- (1837), *Recherches sur la probabilité des jugements*. Paris, 2003. **S, G,** 53.
Poisson, S. D., Dulong, P. L., Larrey, F. H., Double, F. J. (1835), [Report on a manuscript by J.] Civiale, Recherches de statistique sur l'affection calculeuse. *C. r. Acad. Sci. Paris*, t. 1, pp. 167 – 177.
Poroshin, V. S. (1838), Critical examination on the foundations of statistics. In Druzinin (1963, pp. 92 – 104).
Ptoukha, M. V. (1961), Sample investigations of the 17th – 18th centuries in agriculture in Russia. *Uchenye Zapiski po Statistike*, vol. 6, pp. 94 – 100. (R)
Quetelet, A. (1829), Recherches statistiques sur le Royaume des Pays-Bas. *Mém. Acad. Roy. des Sci., des Lettres et des Beaux Arts de Belgique,* t. 5, pp. I – VI+ 1 – 55.
--- (1843), Sur le recensement de la population de Bruxelles. *Bull. de la Commission Centrale de Statistique* [*de Belgique*], t. 1, pp. 27 – 164.
--- (1845), Sur l'appréciation des documents statistiques. Ibidem, t. 2, pp. 205 – 286.
--- (1846), *Lettres sur la théorie des probabilités*. Bruxelles.
--- (1848), *Du système social*. Paris.
--- (1852), Sur quelques propriétés curieuses qui présentent les résultats d'une série d'observations. *Bull. l'Acad. des Sci., des Lettres et des Beaux Arts de Belgique*, t. 19, pt. 2, pp. 303 – 317.
--- (1869), *Physique sociale*, tt. 1 – 2. Bruxelles a. o.

Recherches (1821 – 1829), *Recherches statistique sur la ville de Paris et de département de la Seine*, tt. 1 – 4. Paris. I refer to the second edition of the first two volumes (1833 and 1834). The *Recherches* appeared anonymously but *La grande encyclopédie* (n. d., t. 17, pp. 908 – 909) stated that its editor was Fourier. This attribution was never challenged.
Romanovsky, V. I. (1923), Review of Chuprov (1918 – 1919). *Vestnik Statistiki*, No. 1 – 3, pp. 255 – 260. **S, G,** 1.
Roslavsky, A. P. (1839), Introduction to statistics. In Druzinin (1963, pp. 169 – 187). (R)
Rümelin, F. (1867), Über den Begriff eines socialen Gesetzes. In author's *Reden und Aufsätze.* Freiburg i. B. – Tübingen, 1875, pp. 1 – 31.
Schlözer, A. L. (1804), *Theorie der Statistik*. Göttingen. S, G, 86.
Seneta, E. (1985), Sketch of the history of survey sampling in Russia. *J. Roy. Stat. Soc.*, vol. A 148, pp. 118 – 125.
--- (1994), Liebenneister's hypergeometric tails. *Hist. Math*., vol. 21, pp. 453 – 462.
Shewhart, W. A. (1931), *Economic Conrol of Quality of Manufactured Product*. New York.
Sheynin, O. B. (1971), On the history of some statistical laws of distribution. *Biometrika*, vol. 58, p. 234 – 236.
--- (1977a), Daniel Bernoulli's work on probability. In *Studies in the History of Statistics and Probability*, vol. 2. Eds, Sir Maurice Kendall, R. L. Plackett. London, pp. 105 – 132.
--- (1977b), Early history of the theory of probability". *Arch. Hist. Ex. Sci.*, vol. 17, pp. 201 – 259.
--- (1978), Poisson's work in probability. Ibidem, vol. 18, pp. 245 – 300.
--- (1982), On the history of medical statistics. Ibidem, vol. 26, pp. 241 – 286.
--- (1984a), On the history of the statistical method in astronomy. Ibidem, vol. 29, pp. 151 – 199.
--- (1984b), On the history of the statistical method in meteorology. Ibidem, vol. 31, pp. 53 – 95.
--- (1986), Quetelet as a statistician. Ibidem, vol. 36, pp. 281 – 325.
--- (1991a), Poincaré's work on probability". Ibidem, vol. 42, pp. 137 – 171.
--- (1991b), Buniakovsky's work in the theory of probability. Ibidem, vol. 43, pp. 199 – 223.
--- (1996, 2011), *Chuprov*. Göttingen.
--- (1997a), Chuprov's early paper on sampling. JNÖS, Bd. 216, pp. 658 – 671.
--- (1997b), Achenwall. In Johnson et al (1997, pp. 5 – 6). Also in second edition of *Enc. of Stat. Sciences*. Hobokan, NJ, 2006, vol. 1, pp. 26 – 27.
--- (1998), Statistics in the Soviet period. JNÖS, Bd. 217, pp. 529 – 549.
--- (1999), Statistics, Definitions of. In *Enc. of Stat. Sciences, Update Volume*, Bd. 3. Ed., S. Kotz. New York, pp. 704 – 711. New version: Statistics: history and principle. In electronic edition: Wiley, *Stat. Ref. Conf.* 2016, pp. 1 – 13.
--- (2017), *Theory of probability. Historical Essay*. Berlin. **S, G,** 10.
Sigwart, Chr. (1878), *Logik*, Bd. 2. Tübingen. 2. Aufl., 1893.
Small, A. W. (1916), Fifty years of sociology in the USA. *Amer. J. Sociology*, vol. 21, pp. 721 – 864.
Stigler, S. M. (1977), Eight centuries of sampling inspection. *J. Amer. Stat. Assoc*., vol. 72, pp. 493 – 500.
Süssmilch, J. P. (1758), Gedancken von dem epidemischen Krankheiten etc. In J. Wilke, Hrsg, *Die königliche Rezidenz Berlin und die Mark Brandenburg im 18. Jahrhundert*. Berlin, 1994, pp.69 – 116.
Tikhomandritsky, M. A. (1898), *Kurs teorii veroiatnostei* [Course in Theory of Probability]. Kharkov. (R)
Vernadsky, I. V. (1852), The aims of statistics. In Druzinin (1963, pp. 221 – 238). (R)
Westergaard, H. (1890), *Grundzüge der Theorie der Statistik*. Jena.
Woolhouse, W. S. B. (1873), On the philosophy of statistics. *J. of the Inst. of Actuaries*, vol. 17, pp. 37 – 56.
Yanson, Yu. E. (1871), Directions in the scientific treatment of moral statistics. In Druzinin (1963, pp. 262 – 280). (R)
--- (1887), Theory of statistics. lbidem, pp. 280 – 293. (R)
Yastrcmsky, B. S. (1964), *Izbrannye Trudy* [Sel. Works]. Moscow. (R)

Young, Th. (ca. 1819), Remarks on the probabilities of error in physical observations. In author's *Misc. Works*, vol. 1. London, 1855, pp. 8 – 28. Reprinted: New York – London, 1972.
Zuravsky, D. P. (1846), On the sources and application of statistical information. In Druzinin (1963, pp. 199 – 219). Reprint of entire book: Moscow, 1946. (R)

Notes

1. Lazarsfeld (1961) studied the period before the 19th century, the work of Quetelet and some other issues, but from another angle.
2. I list several pertinent facts. In 1865 – 1866 Seidel (Sheynin 1982, pp. 277 – 278) quantitatively estimated the correlation of typhoid fever with some meteorological factors; in 1877, in the context of clinical trials, Liebermeister (Seneta 1994) proposed a test for homogeneity of two binomial populations; and Enko (1889) constructed the first epidemiological model. [It is a sad fact that statistics of population neglected even epidemics.]
3. Even in 1709 Niklaus Bernoulli applied the *art of conjecturing* to jurisprudence. He hardly promoted this science, but at least he made an important step in the right direction.
4. De Moivre (1756, p. 348) remarked that censuses, *repeated at proper intervals* with the population distributed *into the proper Classes*, might result in *useful conclusions* and in discovering *the general state of the Nation*.
5 He had also managed to estimate the precision of his calculations (of sampling) by general mathematical and stochastic means. Also see § 5.2.2. Bearing in mind statistics, Laplace (1781, p. 383) appropriately mentioned *une nouvelle brauche de la théorie des probabilités*. It was Lagrange (letter to Laplace of 13.1.1775 in t. 14 of his *Oeuvres*) who first said *nouvelle branche* with respect to one of Laplace's problems. [Pearson criticized Laplace's calculation of the population of France (Sheynin 2017, pp. 103 – 104).]
6. 1 refer to him time and time again, and I quote Chuprov's relevant opinion (1909, p. 30). Cournot was

One of the most original and profound thinkers of the 19th century … who rates higher and higher in the eyes of posterity.

Chuprov likely bore in mind Cournot's contribution to economics as well.

[Poisson's studies of criminal statistics (1837) were mostly aimed at discovering the optimal majority of the juror's voices for accusing a defendant, and the issue of independence of votes was hardly important.]
7. The end of this sentence was the first direct proposition linking statistics with mass (but not infinite!) phenomena. Littrow (1842, p. 1205) followed suit and additionally argued that

Für ganze große Völkerschaften, so lehrt die Erfahrung, verschwindet die Wirkung des freien Willens beinahe gänzlich …,

cf. Note 29.
8. Gavarret opposed D'Amador (1837) who had contended that the foundation of probability theory was doubtful (p. 114), its applications *ou inutile ou illusoire* (p. 15), etc. In Russia, Davidov (Ondar 1971) followed Poisson m applying the theory of probability to therapeutics.
9. He was against the *moral* applications of mathematics. Even earlier Cauchy (1821, p. v) argued that mathematics should not be applied beyond natural sciences.
10. Quetelet obviously disregarded population statistics.
11. Knies (1850, p. 62) added:

Obwohl Achenwall selbst bereits von Folge spricht, … so sehen wir doch darauf weit weniger Gewicht gelegt …
12. Strictly speaking, this statement served Schlözer only as an illustration. And here is a hardly known contemporaneous definition (Butte 1808, p. xi) which almost coincides with the modern formula of mathematical statistics (Kolmogorov & Prokhorov, see Sheynin 1999, p. 707): the theory of statistics is

Die Wissenschaft der Kunst statistische Data zu erkennen und zu würdigen, solche zu sammeln und zu ordnen …

[Essentially important is a comparison of statistical data belonging to differing moments or regions (which was known to Leibniz!). Therefore, Schlözer's saying was lame.]

13. Mayr (1906, pp. 52 and 58; 1910, p. 76; 1921, pp. 100 – 101) still included statistics into the Staatswissenschaft understood *im engeren übertragenen Sinn* (in the last instance, he omitted the *engeren*). Much earlier, Obodovsky (1839, p. 137) stated that *Nobody doubts anymore that statistics is a science*, cf. also the very title of Knies (1850).
14. Delambre (p. LXX) also remarked that *Les descriptions minéralogiques appartiennent sans doute à la statistique*. This statement indeed corresponded to qualitative statistics. And in 1838 J. E. Portlock (Fallati 1843, p. 3) declared that the not entire numerical descriptions of animals, plants and minerals were their respective statistics! [Schlözer did not deny studies of causes. His viewpoint was embivalent.]
15. Woolhouse (1873, p. 37) noted that the *absurd restrictions* imposed by the Society *have been necessarily disregarded*.
16. Later on similar pronouncements were made with respect to the theory of probability (§ 5).
17. Moreau de Jonnès (1847) began his book by stating that statistics was *la science des faits sociaux exprimés par de termes numériques*. At the same time, his subject index lacked such terms as political arithmetic, mathematics, probability theory.
18. Cf. Fourier (1823, p. XX):

It is not enough to collect a large number of observed values and to take the mean value; it is also necessary to examine whether these values approach, or lie wide apart [s'écartent beaucoup] *from the mean result.*

19. Cf. Cauchy (1845, p. 242): Statistics offered the means for judging doctrines and institutions. He did not say anything about its relations with probability, but he recommended to apply it *with full rigour*. His viewpoint apparently changed as compared with 1821, see Note 9.
20. This subdivision was only heuristic since accidental causes are variable. Quetelet enlarged upon Laplace (1814, p. 37) who recognized irregular and constant causes.
21. In usual notation, the magnitude $|\xi - E\xi|$ obeys the Bienaymé – Chebyshev inequality which certainly does not mean that some value x_i of ξ coincides with $E\xi$, and Rümelin was wrong. Chr. Bernoulli (1841, p. 389) and Fries (1842, p. 23) correctly stated that mean indicators cannot be directly applied to individual cases and logicians (Sigwart 1878, p. 537) agreed with this. Nevertheless, some statisticians repeated Rümelin's mistake; Bortkiewicz (1904, pp. 250 – 251) accused *die Neueren* of keeping to the wrong conclusion.
22. I bear in mind the four volumes of tables (*Recherches* 1821 – 1829).
23. Even in the 20th century Halbwachs (1913, p. 172) declared that the Bernoulli law of large numbers cannot be applied to social phenomena because society consists of groups with any individual belonging to some group.
24. In 1885, the International Statistical Institute was established in their stead.
25. Since the time of Quetelet the field of moral statistics greatly widened; even Zuravsky (1846, p. 216) understood it as comprising the issues of crime, bankruptcies, prostitution, religious dissent, vagrancy, alcoholism, the state of prisons, philanthropy, official decorations. [In moral statistics, Süssmilch was Quetelet's predecessor.]
26. When discussing political arithmetic, Niemann (1807) declared that (John 1883, p. 672)

Its employment in determining the social conditions of men ... in considering these [pertinent] *facts for different periods in order to compare them and then see how far and in what particular progress or decline had taken place; in studying the influences exercised by physical and political causes, has done more for the improvement of the political condition of states than the mere piling up of figures which are frequently so little to be relied upon.*

Lueder (1812, p. 9) argued that *Legionenweise erschienen statistische Angaben in Zahlen und statistische Tabellen voll Zahlen*. Cournot (1843, § 103) stated that

In our days ... statistics blossomed out somewhat exuberantly and we even have to guard ourselves against its premature and improper applications which can discredit it for some time ...

In § 105 he added that statistics should have *sa théorie, ses régles, ses principes*. In turn, Quetelet (1846) denounced the reduction of statistics to compilation of tables (p. 432) as practised by some statisticians (p. 278).

27. Chr. Bemoulli (1841, p. 389) used the term *Biometrie* as representing mass observations concerning *einer ganzen Klasse oder Gattung* of men.
28. Cf. Kolmogorov (1948, p. 77):
Notions about the logical structure of the theory of probability, which underlies all the methods of mathematical statistics, remained [in the Pearsonian school] *on the level of the 18th century*.
29. The ensuing discussions on the part of the free will in man's behaviour (also see § 3.2, Montmort's and Littrow's opinion in § 2.1 and Note 7) did not involve mathematics (Chuprov 1897). I adduce Schiller's statement (Wallenstein, 1800; Wallenstein's Tod, Aufzug 2, Auftritt 3, also quoted in 1867 by Drobisch):
Des Menschen Thaten und Gedanken ... sind notwendig wie des Baumes Frucht, Sie kann der Zufall gaukelnd nicht verwandeln.
Compare now the issue of free will with the ability of a physician [xiv, Note 5].
30. It is therefore hardly surprising that Yanson (1887, p. 291 – 293) did not set high store by mathematics. Thus, mathematicians, who analysed data on social phenomena,
Disregarded their special properties and sometimes arrived at conclusions bordering on nonsense.
Stochastic calculations presuppose constancy of the phenomena studied and statisticians should not be *carried away* by mathematical *deductions*; *general elementary knowledge of probability theory is in most cases sufficient* for appraising statistical conclusions.
The first two statements only mean that statisticians and mathematicians should work together (of. § 2.4.2), and the last declaration proved wrong.
31. Angelo Messedaglia, 1866, as quoted by John (1896, p. 18), was much more resolute: *Die Statistik ist, schon an und für sich nichts anderes als eine Wahrscheinlichkeitsrechnung*. The 1atter also remarked that another Italian statistician, Luigi Perozzo, was of the same *Grundanschauung*.
On the other hand, continuing the tradition of the Staatswissenschafl, Block (1886, p. 132) discussed the relations between statistics and history, geography and political economy without saying anything about mathematics.
32. Apparently after 1850, see Knies's testimony above.
33. Aleksandr Ivanovich Chuprov (1842 – 1908) was an economist and a *non-mathematical* statistician and the creator of the zemstvo, of the local agricultural statistics in Russia. In general, see Seneta (1985) for the early history of sampling in Russia.
34. Cf. Young (ca. 1819, pp. 8 – 9); It is vain
To substitute arithmetic for common sense ...; at least as much good sense is required in applying our mathematics to objects of a moral nature as would be sufficient to enable us to judge of all their relations without any mathematics at all.
In demography, the non-formal attitude was justified by the presence of large systematic mistakes. Alteady Lambert (1772, § 108), when studying the size of families, arbitrarily increased the numbers of children by l/2 thus allowing for stillbirths and infant mortality. Geodesists, beginning with Gauss, adhered to a mixed approach: they definitively estimated precision by taking into account all the appropriate discrepancies (the *closings* of their triangles etc.) rather than by examining the repeated observations at single stations.
Common sense failed Westergaard (1890, pp. 90 – 91). As it is borne out by his figures, he decided that the formal error in the sample number of births was equal to $[0.0lqn]^{1/2}$ where n was the sample population and q%, the birth-rate.
35. Lexis (1879, §§ 3 and 5) even indirectly argued that not every such indicator was a probability and that (§ 13) such cases were indeed interesting. The law of large numbers (as well as the De Moivre – Laplace limit theorem) assumes that the probability of a random event exists and is known. For calculating an unknown (but existing) probability $\overline{p}$) statisticians could have rather applied the Bayesian approach or the (hardly known) Bayes limit theorem

$$P[-z \le \frac{\overline{p}-a}{\sqrt{pq/n^3}} \le z] \sim \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^z \exp(-x^2/2)dx$$

where p is the number of the occurrences of the studied event in n trials, $q = n - p$ and $a = p/n$ (Sheynin 1971).
36. While examining the applied aims of mathematical statistics, Mayr (pp. 43 – 46) largely restricted them. During the period from 1895, the year when the first edition of his work had appeared, his narrow viewpoint did not change, as he himself stated (1914, p. 153); see pp. 26 – 29 of this first edition. It is instructive to compare Mayr's account with a modern description of the application of stochastic processes (!) in sociology (Lang 1992).
37. Here, he disregarded Knapp's opinion, see § 5.3.5.
38. Envisioned by Newton in a manuscript written about 1665, hinted at by Jakob Bernoulli (1713, chapter 4 of pt.4) and applied by Buffon and Laplace.
38a. Chuprov (Ziele und Wege …, *Nordisk Statistisk Tidskrift,* Bd. 3, 1924, p. 435), however, remarked:

Der Wahrscheinlichkeitsbegriff der klassischen Wahrscheinlichkeitsrechnung, der sich auf das so genannte Prinzip des mangelnden Grundes stützt, kann Niemand mehr, als Grundlage der statistischen Anwendungen der Wahrscheinlichkeits-rechnung, befriedigen.

39. This is true not only from the gambler's point of view: the effect can well change the number of the revolutions of the ball.
40. He said much the same earlier (1877, p. 14).
41. In 1925, in a Russian contribution, Chuprov again reasonably maintained that the issue of the subjective versus objective nature of probability *acquires* [in applications] *superior importance* (Sheynin 2011, p. 121).
42. Independence of events is defined through their probabilities. As to *event*, I quote Chebyshev (1936, p. 111):

The word event means, in general, everything whose probability is being determined ... the word probability thus serves to denote some magnitude subject to measurement.

This great scientist thus made a tiny step towards an axiomatic theory of probability. Moreover, Tikhomandritsky (1898, p. iv) put on record that Chebyshev, in 1887, had argued that *the entire theory of probability should now be reconstructed*.
43. Cf. Cournot (1843, § 115): The Bernoulli *principle* is the only *base solide* for all the applications of probability theory. [Part 4 of J. B.'s classic is inadequately known. Thus, he thought that statistical probability of an event should be applied even when its theoretical probability did not exist.]
44. Knapp's negative attitude towards probability theory is also seen in his (correct) statement (p. 101) that each crime has its own cause and in his denial (p. 114) of any difference between essential and random causes.

Here is what Bortkiewicz (1904, pp. 252—253) had to say about Knapp (the *anderer*) and Mayr: Knapp was

Ein anderer und wohl der schärfste und konsequenteste Gegner der wahrscheinlichkeits-theoretischen Auffassungsweise

and (1910, p. 358) *a most convinced enemy* of applying probability.

Bortkiewicz was hardly able to speak his mind during Knapp's jubilee (observing his 40 years at Strasbourg) where he (1915, p. 119) even credited the latter with the *Neubegründung einer Theorie der Statistik*. It is true, however, that Knapp was meritorious for studying mortality (Bortkiewicz 1922).
45. This was one of Poisson's illustrations of his version of the law of large numbers.
46. The author was a physician in Petersburg.
47. Dormoy should also be mentioned as cofounder of the same theory. However (Bortkiewicz 1930, p. 53),

Ging Dormoy das Verständnis für Anwendungen der Wahrscheinlichkeits-rechnung auf Erfahrungstatsachen in so starkem Masse ab, dass es der historischen Gerechtigkeit nicht entspricht, ihn, so oft von Dispersionstheorie die Rede ist, in eine Reihe mit Lexis zu stellen.

48. A notion which was then not yet applied beyond the error theory, cf. the beginning of § 5.2.5.
49. Beginning with 1913, Yastremsky treated the Lexian theory as a theory of mutability of statistical series. Romanovsky (1923, p. 160) briefly described the essence of his method which remains hardly known in spite of the appearance of

Yastremsky's *Selected works* (1964). It is instructive to recall that Buys-Ballot, in 1850, argued that modern meteorology was then in its second stage characterized by the [statistical, at least in part] study of the deviations from the mean states of the atmosphere.

50. Chuprov's related manuscript devoted to the derivation of the asymptotic density of Q^2 (Sheynin 2011, p. 101) is now published in an English translation (Chuprov 1999).

51. Poisson, Bienaymé and Cournot might be called its predecessors; Bortkiewicz and Chuprov were among his followers (the former was also a student of Lexis).

52. In 1906 Chuprov also edited the translation of two of Lexis's papers on the statistical measurement of mortality.