

Russian Papers on the History of Probability and Statistics

Translated by the Author

Berlin 2004

(C) Oscar Sheynin www.sheynin.de

Contents

Introduction

1. Review of Kendall, M.G., Doig, A.G. *Bibliography of Statistical Literature Pre-1940 with Supplements to the Volumes for 1940 – 1949 and 1950 – 1958*. Edinburgh, 1968. *Novye Knigi za Rubezhom*, ser. A, No. 10, 1969,
2. On the work of Adrain in the theory of errors. *Istoriko-Matematicheskie Issledovania* (IMI), vol. 16, 1965, pp. 325 – 336
3. On the history of the iterative methods of solving systems of linear algebraic equations. *Trudy IX Nauchn. Konf. Aspirantov i Mladsh. Nauchn. Sotrudn. Inst. Istorii Estestvoznania i Tekhniki*, Seksia istorii fiz. i mat. nauk. Moscow, 1966, pp. 8 – 12
4. On selection and adjustment of direct observations. *Izvestia Vuzov. Geodezia i Aerofotos'emka* No. 2, 1966, pp. 107 – 112
5. On the history of the adjustment of indirect observations. *Ibidem*, No. 3, 1967, pp. 25 – 32
6. *Some Issues in the History of the Theory of Errors*. Abstract of dissertation. Moscow, 1967. Published as a manuscript. *Inst. Istorii Estestvoznania i Tekhniki*
7. On the work of Bayes in the theory of probability. *Trudy XII Nauchn. Konf. Aspirantov i Mladsh. Nauchn. Sotrudn. Inst. Istorii Estestvoznania i Tekhniki*, Seksia istorii mat. i mekh. nauk. Moscow, 1969, pp. 40 – 57
8. On the history of the De Moivre – Laplace limit theorem. *Istoria i Metodologia Estestven. Nauk*, vol. 9, 1970, pp. 199 – 211
9. On the appearance of the Dirac delta-function in a memoir of Laplace. *IMI*, vol. 20, 1975, pp. 303 – 308
10. History of the theory of probability. Based on Theory of probability before Chebyshev. *IMI*, vol. 25, 1978, pp. 284 – 306, and *History of the Theory of Probability to the Beginning of the 20th Century*. Berlin, 2004
11. Liapunov's letters to Andreev. *IMI*, vol. 31, 1989, pp. 306 – 313 .
12. On the history of the statistical method in natural sciences. *IMI*, vol. 32/33, 1990, pp. 384 – 408
13. Markov's report on a paper by Galitzin. *Ibidem*, pp. 451 – 467
14. Markov's papers in the newspaper *Den*, 1914 – 1915. *IMI*, vol. 34, 1993, pp. 194 – 206
15. Correspondence of Nekrasov and Andreev. *IMI*, vol. 35, 1994, pp. 124 – 147. Coauthor: M.V. Chirikov
16. The notion of randomness from Aristotle to Poincaré. *IMI*, vol. 1 (36), No. 1, 1995, pp. 85 – 105
17. Correspondence between P.A. Nekrasov and A.I. Chuprov. *Ibidem*, pp. 159 – 167
18. Markov and life insurance. *IMI*, vol. 2 (37), 1997, pp. 22 – 33
19. Slutsky: commemorating the 50th anniversary of his death. *IMI*, vol. 3 (38), 1999, pp. 128 – 137
20. History of the theory of errors. *IMI*, vol. 5 (40), 2000, pp. 310 – 332

Introduction

I am presenting translations of my papers originally published in Russian, mainly in *Istoriko-Matematicheskie Issledovania* (IMI). Only a fraction of historians of mathematics read Russian and some are unwilling to study the contributions published beyond the usual set of periodicals so that my present work seems justified. In actual fact, I am putting out most of the items from a microfiche collection of the same title published by Hänsel-Hohenhausen in 1999 as *Deutsche Hochschulschriften* 2621 but hardly examined by more than a dozen readers; the copyright to ordinary publication was, and is mine. Some items below are translations of publications of materials kept at several Russian archives or newspaper articles and among the former is Markov's critical review of a paper devoted to the treatment of observations.

In translating my papers, I corrected a few mistakes and misprints (largely due, in the new series of the IMI, to the impossibility of reading the proofs), left out dated material, and referred not to Russian translations of classical works but to their original editions.

Abbreviations used throughout:

AHES = *Arch. Hist. Ex. Sci.*;

DHS = *Deutsche Hochschulschriften*;

IMI = *Istoriko-Matematicheskie Issledovania*;

L. = Leningrad;

M. = Moscow;

MSb = *Matematich. Sbornik*;

Psb = Petersburg;
(R) = in Russian;
ZhMNP = *Zhurnal Ministerstva Narodn. Prosveshchenia*.

1. **Review of Kendall, M.G., Doig, A.G. (1968),**
Bibliography of Statistical Literature pre-1940 with Supplements
to the Volumes for 1940 – 1949 and 1950 – 1958. Edinburgh.

This is vol. 3 of the entire *Bibliography* covering the period until 1958; the first two volumes appeared in 1962 and 1965. No further volumes are planned since in 1959 the International Statistical Institute began publishing an abstracting journal now called *Statistical Theory and Methods Abstracts*. According to the authors' aims and methodology as described in vol. 1, the *Bibliography* includes almost all the articles from 12 main periodicals and a number of papers from 42 other journals. In addition, the authors made use of the bibliographies appended to many papers and of the abstracting journals (although not of the Soviet *Matematika*). They believe to have covered 95% of the existing articles on statistics and its applications.

Each volume of the *Bibliography* is actually an author index (no subject indices are provided). The literature published in Russian and in several other languages is described in English, French or German. In all, this vol. 3 lists about 10 thousand monographs and articles separated into two time intervals, – before 1900 and from 1900 to 1939 (2,360 and 7,630 items respectively) as well as 148 sources for 1940 – 1949 and about 1,170 for 1950 – 1958. All the books entered here had appeared before 1900. Neither the second part, nor the first two volumes include any books which is in line with the practice of the abovementioned quarterly. This is an essential setback but the *Bibliography* is nevertheless very valuable.

Vol. 3 is also useful for historians of mathematics since it lists classical works (of Laplace, Gauss et al) including writings of such authors for whom probability was a minor subject (Euler), forgotten writings of eminent mathematicians, commentaries and essays, translations of various works into any of the three main languages.

There are some shortcomings. The selected literature, even of the 20th century, was not checked *in visu*; likely because of the general direction of the *Bibliography* there are hardly any references to collected works; of the 14 writings of Euler included in t. 7 of his *Opera omnia*, ser. 1 (1923) and pertaining to probability and statistics, the authors included only seven, and one of these called *Wahrscheinlichkeitsrechnung* either does not exist or wrongly named; the descriptions contain mistakes and inaccuracies (Süssmilch's *Göttliche Ordnung* first appeared in 1741, then in 1761 – 1762 but not in 1788; the second part of Daniel Bernoulli's "Mensura sortis" (1771) is omitted); and cross-references are lacking. Finally, the spelling Ladislaus von Bortkiewicz as given in the second part does not coincide with that in the first part, Vladislav Bortkevich. Having emigrated from Russia to Germany in 1901 and being a nobleman, he changed his name accordingly but that fact is not explained.

In 1962, the authors estimated that about a thousand articles on their subject were being published yearly. This means that already now it would be expedient to issue a bibliography of this literature for 1959 – 1970. Neither abstracting journals, nor their cumulative author indices are a substitute for bibliographies (to be compiled in the first place by scanning such sources). I also believe that a single bibliography for 1900 – 1970 with books being certainly included is also needed.

On the Work of Adrain in the Theory of Errors
Istoriko-Matematicheskie Issledovania (IMI), vol. 16, 1965, pp. 325 – 336

In translating my paper I took into account its somewhat revised version appended to my unpublished thesis of 1967 (Some Issues ..., partly translated in this collection). Adrain's articles are now reprinted (see Bibliography) and I have therefore omitted his original and hardly understandable derivations of the normal law (leaving however their modernized reconstruction [8]). Their latest discussion is due to Hald [10, pp. 368 – 373] and Dutka [8a]. Also note that Adrain's paper [2] apparently appeared in 1809 rather than in 1808 [13, p. 170].

* * *

Robert Adrain is meritorious for his remarkable findings in the theory of errors. He published two derivations of the normal law of error a year before [or at the same time as] Gauss did and applied it to establishing the principles of least squares and arithmetic mean as well as to determining the flattening of the earth's ellipsoid of revolution.

Adrain was born in Ireland and died in New Brunswick. He learned mathematics mainly by himself and began teaching it at an early age. Then, after participating in the Irish national movement and being wounded in the revolt of 1798, he fled to the United States. Adrain resumed there his teaching activities becoming, in 1809, Professor of mathematics at Queen's College (now, Rutgers College) in New Brunswick. From 1813 to 1826 he was Professor at Columbia University, and, from 1827 to 1836, at Pennsylvania (vice-rector from 1828 to 1836).

Adrain delivered lectures in various disciplines. Thus, in 1829 he taught elementary mathematics, geodesy, cartography, mathematical analysis, mechanics and astronomy. He and Nathaniel Bowditch (1773 – 1838)¹ were among the first American mathematicians. In 1812 Adrain was elected to the American Philosophical Society, and, in 1813, to the Academy of Sciences and Arts. He actively contributed to the first American mathematical periodicals. Coolidge [8] provided a general description of Adrain's work, but his account of the latter's findings in the theory of errors was not comprehensive. In the 19th century several geodesists and astronomers discussed these in more detail (e.g., [1; 9; 26] from among those which I do not mention below) but still not sufficiently. At present [in 1965], however, Adrain is forgotten. Neither Struik [23]² nor Strasser [22] nor many other authors of general contributions on history of mathematics cite him and Cajori [6, p. 382] only devoted a few lines to Adrain's discovery of the law of error.

I turn now to Adrain's paper [2]. He issued from a prize question: A traverse with measured sides and bearings did not close ($\sum \Delta x_i \neq 0$, $\sum \Delta y_i \neq 0$). It is required to determine the most probable corrections to the computed increments Δx_i and Δy_i . The paper itself contained two derivations of the normal law of error; the derivation of the principles of least squares (discovered by Gauss in 1795 or 1794 and offered by Legendre in his publication of 1805) and of the arithmetic mean in the one- and three-dimensional cases; the determination of the most probable position of a ship calculated by dead reckoning given its observed latitude; and the solution of the prize question. In concluding, Adrain stated that, owing to lack of space, he had to postpone his derivation of the most probable flattening of the earth's ellipsoid which he again accomplished [in 1818, see below] on the basis of the normal law.

1) His first derivation of the normal law concerned linear measurements a and b whose errors x and y , as he presumed, obeyed two conditions

$$x/a = y/b \tag{1}$$

(the errors were proportional to the lengths of the lines)³, and

$$x + y = E \tag{2}$$

(the total error was constant, but why?). He also tacitly believed that the errors were independent.

Denote the unknown density of errors by φ ; then, according to the principle of maximum likelihood,

$$[\varphi'(x; a) / \varphi(x; a)] dx + [\varphi'(y; b) / \varphi(y; b)] dy = 0 \tag{3}$$

or, allowing for (1),

$$\varphi'(x; a) / \varphi(x; a) = \varphi'(y; b) / \varphi(y; b), \text{ or } \psi(x; a) = \psi(y; b)$$

whose simplest solution is, when taking into account (2),

$$\psi(x; a) = mx/a, \quad \varphi(x) = c \exp(mx^2/2a), \quad c > 0, \quad m < 0,$$

a function with two essential constants. Neither here, nor in his next derivation did Adrain calculate their values; moreover, he had not considered it important.

2) In his second derivation of the normal law Adrain studied the determination of a station B from a given station A by measured distance AB and azimuth of that line. Supposing that the errors x ; y along and perpendicular to AB were equally probable, he also assumed that

$$x^2 + y^2 = \text{Const} \quad (4)$$

so that relation

$$[\varphi'(x)/\varphi(x)] dx + [\varphi'(y)/\varphi(y)] dy = 0$$

which is similar to (3) led to

$$\varphi'(x)/x = \varphi'(y)/y = n \text{ etc.}$$

Adrain also remarked that, in general, circumferences (4) should be replaced by ellipses. He did not however dwell on that point and thus missed the opportunity of introducing the bivariate normal law (or of any such term as *ellipse of concentration*), but at least he considered such a general case elsewhere (see Item 5).

Each of Adrain's demonstrations issued from geodetic measurements; however, he hardly introduced any properties of their errors so that his result was not related to them; it is not even seen that the errors were random.

Merriman [20] and subsequent authors pointed out that the derivation of the normal law by John Herschel [12] was similar to Adrain's second justification, see below.

3) Adrain derived the principles of least squares and arithmetic mean in the same way as Gauss did. Denote the n observations of an unknown constant x by a, b, c, \dots , then the logarithms of the probabilities of the errors $(x - a), (x - b), (x - c), \dots$ will be

$$-(x - a)^2, -(x - b)^2, -(x - c)^2, \dots$$

and the most probable value of x will correspond to

$$(x - a)^2 + (x - b)^2 + (x - c)^2 + \dots = \min$$

so that

$$x = (a + b + c + \dots) / n. \quad (5)$$

Adrain's derivation of the principle of least squares (for one unknown) was questioned: Coolidge [8] stated that he had Legendre's book in his library.⁴ The derivation of (5) for the three-dimensional case (also considered by Legendre) was similar.

4) The correction of dead reckoning was similar to the adjustment of the traverse. The observation of only one astronomical magnitude, the latitude, leads to only one (a latitudinal) discrepancy between the reckoned and the observed positions of the ship. Consequently, this case is indeed similar to adjusting the traverse with respect to only one coordinate.

5) Adjustment of the traverse. Its measured sides are a, b, c, \dots , their corrections, x, y, z, \dots , the measured bearings, A, B, C, \dots the corrections to the positions of the vertices of the polygon in the directions perpendicular to its sides, $X = a \Delta A, Y = b \Delta B, Z = c \Delta C, \dots$, corrections to the increments Δx_i and $\Delta y_i, D_i$ and L_i , and the total corrections

$$D = D_1 + D_2 + \dots, \quad L = L_1 + L_2 + \dots$$

Adrain derived the simple formulas

$$D_1 = x \sin A + X \cos A, \quad D_2 = y \sin B + Y \cos B, \dots$$

$$L_1 = x \cos A - X \sin A, \quad L_2 = y \cos B - Y \sin B, \dots$$

and determined the minimal value of the function

$$x^2/a + y^2/b + \dots + X^2/p^2a + Y^2/p^2b + \dots \quad (6)$$

The parameter p allowed for an unequal precision of measuring the bearings and the sides of the polygon. Assuming that $p = 1$, Adrain got

$$D_1 = ma, D_2 = mb, \dots; L_1 = na, L_2 = nb, \dots;$$

$$m = D/(a + b + \dots), n = L/(a + b + \dots)$$

and additionally considered the academic cases of $p = 0$ and $p = \infty$.

The minimal value of (6) corresponded to adjusting the sides of the polygon corrupted by systematic errors (the weights were proportional to the sides themselves rather than to their roots). It is nevertheless noteworthy that exactly that principle is still applied for adjusting geodetic traverses of lower precision [7, §164].⁵

In his next article [3] Adrain determined the flattening of the earth's ellipsoid $\alpha = (a - b)/b$ with semiaxes a and b ($a > b$)⁶. He issued from Laplace's data [18, §40] on the lengths of the seconds pendulum at various latitudes. The Clairaut formula connecting the latitude φ with the acceleration of gravity r is, in his notation,

$$r = x + y \sin^2\varphi$$

with x and y determining α and Adrain solved a system of such equations under the least-squares condition imposed on the residual free terms (call them v_i) of his (Laplace's) equations getting $\alpha = 1/319$. His trivial application of least squares was interesting in that it was the first one in its field⁷ and, in addition, because of the result obtained. Laplace himself got $\alpha = 1/335.78$ which almost coincided with another figure calculated at the same time (in 1799). The difference between Laplace's and Adrain's results was mainly caused by two mistakes made by the former and revealed by the latter. According to Laplace's (Boscovich's) conditions

$$\sum |v_i| = \min, \sum v_i = 0,$$

Adrain additionally arrived at $\alpha = 1/316.5$ so that the essential difference between him and Laplace was $1/316.5 - 1/319$.

I also mention Adrain's article [4]. Believing that for some practical purposes it was convenient to consider the Earth as a sphere, he calculated its radius under seven different conditions: equal volume or equal surface with the appropriate ellipsoid of revolution; equal masses of bodies restricted by these surfaces (sphere and ellipsoid) given a certain law of the decrease of mass with depth, etc. As a first approximation, Adrain arrived at one and the same result, $r = (2a + b)/3$. He then stated that he had determined the most probable values of a and b by means of meridian arc measurements according to the method published in his first paper. The corresponding mean radius, as he added without providing either the initial data or his own calculations, was 3,959.36 English miles. With $\alpha = 1/319$ this was equivalent to $a = 3,963.50$ English miles, or, assuming that 1 meter equals 39.370113 inches, 6,378.629km. Here are some later determinations [14].

- | | | |
|---------------------|--------------------|------------------|
| 1. Delambre, 1800: | $a = 6,375.653km,$ | $1/\alpha = 334$ |
| 2. Walbeck, 1819: | 76.896 | 302.78 |
| 3. Krasovsky, 1940: | 78.245 | 298.3 |

Herschel [12] derived the normal law by considering a free fall of a ball on a horizontal plane and its deviations from the point above which it was initially situated. He distinctly formulated the symmetry and the decrease of the density sought indicating that those assumptions were a corollary of complete ignorance of the causes of error and the manner of their action.

Thomson & Tait [24, p. 314], without referring to anyone, offered a similar justification. They considered a free fall of a stone. Denoting by $\varphi(x^2) \delta x$ the probability of its fall on interval $[x; x + \delta x]$ they stated that its fall on point $(x; y)$ had probability $\varphi(x^2) \varphi(y^2) \delta x \delta y$. After rotating the system of coordinates they obtained

$$\varphi(x^2)\varphi(y^2) = \varphi(x'^2)\varphi(y'^2)$$

and it followed (as was already well-known at the time) that

$$\varphi(x^2) = A \exp(mx^2)$$

with $m < 0$ since the probability of a large deviation had to be very small.

Again, both Tsinger [25] and Krylov [17, Chapt. 8] applied the same pattern for deriving the normal law by considering shooting at a vertical target. Kemnitz [16] noted that Krylov (and therefore his predecessors as well) had not made essential use of the properties of random errors.⁸

Notes

1. A mathematician and astronomer, Fellow of the Royal Society. He is mostly remembered for his work in navigation and his translation of Laplace's *Mécanique Céleste*.
2. As stated in the title of his book, that author had indeed restricted his study to the Yankee, rather than to the American science.
3. This property is characteristic of systematic rather than random errors, cf. Item 5 below.
4. Later note: It remains unknown, however, when did Adrain get it [10, p. 371]. More [11, p. 626]: Adrain never used the term *least squares*, nor did he refer to Legendre's treatment of meridian arc measurements, cf. below. Finally, Adrain had not then directly stated that his principle of least squares might be applied to the case of several unknowns.
5. Later note: Adrain properly adjusted the directly measured magnitudes rather than the increments Δx_i and Δy_i .
6. Adrain's definition of the flattening was unusual: the generally adopted formula was and is $\alpha = (a - b) / a$. In some cases, as when comparing his results with those of other authors (below), this is of no consequence. After reading [3], Olbers informed Gauss (24.2.1819; [21, p. 711]) that *ein Amerikaner ... schreibt sich ... die Erfindung der Methode der kleinsten Quadrate zu*. Gauss made no comment.
7. The first published application of this kind is however due to Biot [5, Additions, pp. 167 – 169].
8. Later note: It is generally believed that Maxwell, in his celebrated justification of the normal law of velocities of gas molecules which assumed the independence of the three components of the velocities, issued from Herschel's derivation. Kac [15] and Linnik [19] had since revised Maxwell's proof. Independence is still needed but in a weaker form; however, it should in addition persist under any choice of the coordinate system.

References

1. Abbe, C. Historical note on the method of least squares. *Amer. J. Sci. Arts*, vol. 1, No. 6, 1871, pp. 411 – 415.
2. Adrain, R. Research concerning the probabilities of the errors which happen in making observations (1808). Reprinted in Stigler, S.M., Editor, *American Contributions to Mathematical Statistics in the 19th Century*, vol. 1. New York, 1980. No single paging.
3. Adrain, R. Investigation of the figure of the Earth and of the gravity in different latitudes (1818). Reprinted: Ibidem.
4. Adrain, R. Research concerning the mean diameter of the Earth (1818). Reprinted: Ibidem.
5. Biot, J.-B. *Traité élémentaire d'astronomie physique*, t. 3. Paris, 1811.
6. Cajori, F.C. *History of Mathematics*, 2nd ed. New York, 1929.
7. Chebotarev, A.S. *Геодезия* (Geodesy), pt. 1. M., 1955.
8. Coolidge, J.L. Adrain and the beginnings of American mathematics. *Amer. Math. Monthly*, vol. 33, No. 2, 1926, pp. 61 – 76.
- 8a. Dutka, J. Adrain and the method of least squares. *Arch. Hist. Ex. Sci.*, vol. 41, 1990, pp. 171 – 184.
9. Glaisher, J.W.L. On the law of facility of errors of observations and on the method of least squares. *Mem. Roy. Astron. Soc.*, vol. 39, pt. 2, 1872, pp. 75 – 124.
10. Hald, A. *History of Mathematical Statistics from 1750 to 1930*. New York, 1998.
11. Hammer, E. Zur Geschichte der Ausgleichsrechnung. *Z. Vermessungswesen*, Bd. 29, 1900, pp. 613 – 628.
12. Herschel, J. [Review of] Letters on the theory of probabilities ... by Quetelet. *Edinb. Rev., or, Critical J.*, vol. 92, No. 185, 1850, pp. 1 – 57.
13. Hogan, E.R. Adrain: American mathematician. *Hist. Math.*, vol. 4, 1977, pp. 157 – 172.
14. Izotov, A.A. *Форма и размеры земли по современным данным* (Form and Size of Earth according to Modern Data). M., 1950.

15. Kac, M. On a characterization of the normal distribution (1939). [Sel. Papers.] *Probability, Number Theory and Statistical Physics*. Cambridge (Mass.), 1979, pp. 77 – 79.
16. Kemnitz, Yu.V. On a derivation of the law of error. *Trudy Mosk. Inst. Inzhenerov Zemleustroistva*, No. 3, 1959. (R)
17. Krylov, A.N. *Лекции о приближенных вычислениях* (Lectures on Approximate Calculations). М., 1950.
18. Laplace, P.S. *Traité de Mécanique céleste*, t. 2 (1799). *Oeuvr. Compl.*, t. 2. Paris, 1878.
19. Linnik, Yu. V. Comments on the classical derivation of the Maxwellian law. *Doklady Akad. Nauk SSSR*, vol. 85, 1952, pp. 1251 – 1254. (R)
20. Merriman, M. List of writings relating to the method of least squares with historical and critical notes. *Trans. Connecticut Acad. Arts Sci.*, vol. 4, pt. 1 – 2. New Haven, 1877 – 1882, pp. 151 – 232.
21. Schilling, C. W. *Olbers, Sein Leben und seine Werk*, Bd. 2, Abt. 1. Berlin, 1900.
22. Strasser, G. *Ellipsoidische Parameter der Erdfigur, 1801 – 1950*. Deutsche geod. Komm. Bayer. Akad. Wiss., Bd. A19, 1957.
23. Struik, D.J. *Yankee Science in the Making*. New York, 1962.
24. Thomson, W., Tait, P.G. *Treatise on Natural Philosophy*, vol. 1. Oxford, 1867.
25. Tsinger, N. *Курс астрономии* (Course in Astronomy), Theor. part. Psk, 1899.
26. Wright, T.W. *Treatise on the Adjustment of Observations*. New York, 1884.

3. On the History of the Iterative Methods of Solving Systems of Linear Algebraic Equations
Trudy IX Nauchn. Konf. Aspirantov i Mladsh. Nauchn. Sotrudn. Inst. Istorii Estestvoznania iTekhniki,
 Sektsia istorii fiz. i mat. nauk. Moscow, 1966, pp. 8 – 12

Gauss [7; 4; 6] was the first to solve linear systems by the method now called relaxation. He described in detail its application to adjusting triangulation. Apparently to check each step of his calculations, Gauss added up all his initial equations and joined the summary equation to them. When calculating the k -th approximations of x_i , he took into account all the previous calculations made at that (at the k -th) step, and he did not keep to any fixed order of choosing the index i ; he rather selected that unknown for which the ratio of the free term to its quadratic coefficient was maximal in absolute value. He did not leave any indications about the convergence of his approximations.

Jacobi [10] studied the solution of normal equations with small non-diagonal coefficients

$$(00) x + (01) x_1 + (02) x_2 + \dots = (0m)$$

$$(10) x + (11) x_1 + (12) x_2 + \dots = (1m), \text{ etc}$$

where $(ik) = (ki)$. His formulas were

$$x = a + \Delta + \Delta^2 + \dots, \quad x_1 = a_1 + \Delta_1 + \Delta_1^2 + \dots, \quad (00) a = (0m), \quad (11) a_1 = (1m),$$

etc, and, in general,

$$(00) \Delta = -[(01) a_1 + (02) a_2 + \dots], \quad (00) \Delta^{i+1} = -[(01) \Delta_1^i + (02) \Delta_2^i + \dots]$$

$$(11) \Delta_1 = -[(10) a + (12) a_2 + \dots] \text{ etc, } (11) \Delta_1^{i+1} = -[(10) \Delta^i + (12) \Delta_2^i + \dots] \text{ etc.}$$

Jacobi did not prove the convergence of his process, but he additionally considered in detail the case of significant non-diagonal coefficients and [9]¹ derived some relations between the different solutions of the *characteristic system*

$$[(aa) - x] \alpha + (ab) \beta + \dots + (ap) \omega = 0, \dots, \tag{1}$$

$$(pa) \alpha + (pb) \beta + \dots + [(pp) - x] \omega = 0.$$

Finally, Jacobi [10] generalized his iterative method onto non-symmetric systems of equations. His findings are well-known (e.g., [5, §81]) and I shall only sketch them. In order to eliminate, one by one, significant non-diagonal coefficients Jacobi effectively introduced rotation matrices and achieved his generalization by means of orthogonal transformations. For systems (1) with normed unknowns ($\alpha^2 + \beta^2 + \dots + \omega^2 = 1$) he proved a number of properties. Thus, it occurred that $\alpha' \alpha'' + \beta' \beta'' + \dots = 0$ (orthogonality of different solutions); $\alpha' \alpha' + \alpha'' \alpha'' + \dots = 1$ (the norming of the components of the solutions having the same name); and $\alpha' \beta' + \alpha'' \beta'' + \dots = 0$ (orthogonality of different components).

Jacobi also offered working formulas for calculating the unknowns and their weights for a small number of normal equations. Nowadays these are hardly interesting, but in those times Bessel [2] expressed his high opinion about them.

Seidel [16] considered the solution of m observational equations in n unknowns ($m > n$)

$$x - y = \alpha, \quad x - z = \beta, \quad x - u = \gamma, \dots$$

by consecutive approximations beginning with expressions such as

$$x = (1/m) [(y + \alpha) + (z + \beta) + (u + \gamma) + \dots] \text{ etc.}^2$$

In a later memoir [18] Seidel considered the solution of a system of normal equations

$$[aa] x + [ab] y + [ac] z + \dots + [an] = N_1 = 0, \tag{2}$$

$$[ab] x + [bb] y + [bc] z + \dots + [bn] = N_2 = 0, \text{ etc}$$

which of course corresponded to the system of observational equations

$$a_i x + b_i y + c_i z + \dots + n_i = v_i$$

by consecutively minimizing the function

$$Q = [aa] x^2 + \dots + [cc] z^2 + \dots + 2 [ab] xy + \dots + 2 [cn] z + \dots + [nn].$$

Since $Q = [vv]$, Seidel's method coincided with least squares. Seidel noted that the greatest decrease of Q at a certain step was achieved by assuming that

$$\Delta x = -N_1 / [aa] \text{ or } \Delta y = -N_2 / [bb] \text{ or } \dots$$

and several times stressed that iterations should be done not in a fixed order, but in such a way that $|N_i| / [ii] = \max$. He thus effectively applied the method of relaxation.

Seidel noted that approximations might oscillate and even diverge, but connected that possibility only with a violation of his rule of determining each unknown from "its own" equation and thought that this fact could be "easily proved". He also remarked that all the main minors of the determinant of the system (2) were positive so that the convergence of the solution was ensured. He did not explain his reasoning which should have meant "convergence to the rigorous solution".

Then, Seidel studied a number of points essential for calculation: group iterations; treatment of the so-called indirect observations with conditions (see especially [17]); differential change of the weights of the unknowns after adding new equations; treatment of observations whose part was already adjusted.³ He also published a few papers on the theory of errors peculiar to optical instruments [15].

Nekrasov & Mehmke [11; 12] are known to have studied the Seidel iterative method. In particular, they discovered a large number of sufficient conditions for its convergence (restrictions imposed on the coefficients of the initial equations). Nekrasov [13] also stated that he had chosen his subject on the request of the astronomer V.K. Zerassky. Note, however, that Nekrasov's proof [13] that the iterative solution of normal equations converged was faulty: he only proved that $[vv]$ monotonically decreased, not that it reached its minimally possible value. It is remarkable that, upon concluding his "justification", he (p. 192) uttered a curious statement: "In the limit, the Seidel method undoubtedly draws us nearer to the solution sought".⁴

Southwell [19; 14] discovered the method of relaxation anew and applied it to various calculations in engineering. Together with Black [3], he also sketched the possibility of its use for adjusting geodetic networks. Nevertheless, he did not see fit to prove that the method converged. Iterative methods have recently been widely applied, sometimes without working out the normal equations [1].

Notes

1. It was Seidel who performed the numerical computations in [9].

2. Also note Jacobi's memoir [8]. He proved that in order to obtain a solution coinciding with that secured by least squares it was sufficient to combine the observational equations in groups of n , solve each group and assume that x, y, z, \dots were the arithmetic means of the properly weighted partial solutions over all the groups. This method is hardly practical, but its relation with least squares is thus seen. And it had been usually applied in the 18th century for the case $n = 2$ (when determining the parameters of the Earth's ellipsoid of revolution by means of meridian arc measurements) although without assigning any weights to the separate groups.

3. Gauss had considered the two last-mentioned issues.

4. Or, at the very least, correcting a possible grammatical error, "... approaches the solution ..."

References

1. Abramov, A.A., Khublarova, S.L. Solution of systems of linear algebraic equations by iterative methods. *Trudy Zentr. Nauchno-Issled. Inst. Geodezii, Aeros'emki i Kartografii*, No. 135, 1960. (R)
2. Bessel, F.W. Neue Formeln Jacobi's für einen Fall der Anwendung der Methode der kleinsten Quadrate (1840). *Abh.*, Bd. 2. Leipzig, 1876, pp. 401 – 402.
3. Black, A.N., Southwell, R.V. Relaxation methods applied to engineering problems. *Proc. Roy. Soc.*, vol. A164, No. 919, 1938, pp. 447 – 467.
4. Dedekind, R. Gauss in seiner Vorlesung über die Methode der kleinsten Quadrate (1901). *Ges. math. Werke*, Bd. 2. Braunschweig, 1931, pp. 293 – 306.
5. Faddejew, D.K., Faddejewa, V.N. *Numerische Methoden der linearen Algebra* (1960, in Russian). München – Wien, 1970.

6. Forsythe, G.E. Gauss to Gerling on relaxation. *Math. Tables and Other Aids to Computers*, vol. 5, No. 36, 1951, pp. 255 – 258.
7. Gauss, C.F. Brief nach Gerling 26.12.1823. *Werke*, Bd. 9. Leipzig, 1903, pp. 278 – 281.
8. Jacobi, C.G.J. Über die Bildung und die Eigenschaften der Determinanten (1841, in Latin). *Ostwald Klassiker* No. 77. Leipzig, 1896, pp. 3 – 49.
9. Jacobi, C.G.J. Über ein leichtes Verfahren, die in der Theorie der Säcularstörungen vorkommenden Gleichungen numerisch aufzulösen. *J. reine angew. Math.*, Bd. 30, 1846, pp. 51 – 94.
10. Jacobi, C.G.J. Über eine neue Auflösungsart der bei der Methode der kleinsten Quadrate vorkommenden linearen Gleichungen. *Astron. Nachr.*, Bd. 22, 1845, pp. 297 – 306.
11. Mehmke, R. On the Seidel method, etc. *MSb*, vol. 16, 1891, pp. 342 – 345. (R)
12. Mehmke, R., Nekrasov, P.A. Solution of a linear system of equations by consecutive approximations. *Ibidem*, pp. 437 – 459. (R)
13. Nekrasov, P.A. Determining the unknowns by the method of least squares in the case of a very large number of them. *Ibidem*, vol. 12, 1885, pp. 189 – 204. (R)
14. Nikolaeva, M.B. On the Southwell method of relaxation. *Trudy Steklov Math. Inst.*, No. 28, 1949. (R)
15. Seidel, L. Über die Theorie der Fehler mit welchem die durch optische Instrumente gesehenen Bilder behaftet sind. *Abh. natur-wiss. Comm. Bayer. Akad. Wiss.*, Bd. 1, 1857, pp. 227 – 267.
16. Seidel, L. Resultate photometrischer Messungen an 208 der vorzüglichsten Fixsterne. *Abh. math.-phys. Kl. Bayer. Akad. Wiss.*, Bd. 9, Abt. 3, 1861 (1863), pp. 419 – 609.
17. Seidel, L. Über die Berechnung der wahrscheinlichsten Werthe solcher Unbekannten, zwischen welchen Bedingungs-Gleichungen bestehen. *Astron. Nachr.*, Bd. 84, 1874, pp. 193 – 210.
18. Seidel, L. Über ein Verfahren, die Gleichungen durch successive Annäherung aufzulösen. *Abh. math.-phys. Kl. Bayer. Akad. Wiss.*, Bd. 11, Abt. 3, 1874, pp. 81 – 108.
19. Southwell, R.V. Stress calculation in frameworks by the method of systematic relaxations of constraints. *Proc. Roy. Soc.*, vol. A151, No. 872 and vol. 153, No. 878, 1935, pp. 56 – 95 and 41 – 76.

4. On Selection and Adjustment of Direct Observations

Izvestia Vuzov. Geodezia i Aerofotos'emka No. 2, 1966, pp. 107 – 112

Suppose that $x_1 \leq x_2 \leq \dots \leq x_n$ are direct observations of a constant X . It is required to replace the entire variational series by a single number. I am describing the history of the main methods of choosing such a number.

The arithmetic mean. A doctrine of means (and of the arithmetic mean in particular) existed already in the Pythagorean school [23, p. 63]. In antiquity, that mean occurred in most various formulas for calculating areas of figures and volumes of bodies. An Indian commentary of the 16th century stated that the more measurements of the length, width and depth of an excavation were made, the more precise will the determination be of its size and volume [2, p. 97]. In ancient Babyonia, the area of a quadrilateral plot was in two special cases considered to be the product of the halfsums of its opposite sides, viz., when the plot was not quite precisely a rectangle, and when the measurements of the opposite sides were unequal one to another due to the ruggedness of the terrain [32, p. 204].

Thus, the mean should have compensated the inaccuracy of the models, and, possibly, the influence of the systematic errors of measurement. During the epoch of meridian arc measurements the arithmetic mean began to be applied as an universal estimator. Leibniz [21, Book 4, Chapt. 16] testified that it played an important part in developing the [classical] definition of probability and that it has been applied in the sphere of economic relations.

I especially note that even when treating direct observations it was customary to begin by deriving their binary combinations and only then taking the mean of these. Thus, Boscovich [22, p. 150], having four values of latitudinal differences $\Delta\varphi$ between the endpoints of his meridian arc measurement, derived the six halfsums $(\Delta\varphi_1 + \Delta\varphi_2)/2, \dots, (\Delta\varphi_3 + \Delta\varphi_4)/2$, and calculated their mean. The scattering of these combinations had apparently served as an indicator of error.

Simpson [28; 29] devoted a special memoir to stochastically proving that the arithmetic mean was preferable to a single observation. The immediate cause of his work was to refute the statement of *some persons of considerable note* who had thought that *one single observation taken with due care was as much to be relied on as the mean of a great number of them* [28, p. 82].

Such an opinion was possibly occasioned by rapid advances in the technique of observation. Simpson proved (his main result) that for a symmetric triangular distribution the probability of a certain error was essentially less than that of the same error in a single observation. He thus (indirectly) issued from the properties of random errors rather than, as it became fashionable later on, from ignorance of the causes and magnitudes of errors [16, §4129].

The arithmetic mean has been applied together with **rejection of outlying observations**. Galileo [8, Day Third] recommended such rejection and Lambert applied it systematically.¹ Daniel Bernoulli [3], who did not approve of it, mentioned rejection as something usual, as did Euler. Gauss allowed careful rejection of large deviations; however, if, as he stated, such a deviation was caused by an unfortunate concurrence of circumstances, the pertinent observation ought to be retained [9]. Thus, observations might be rejected either if they were indeed corrupted by blunders (Gauss), or if their errors were larger than some magnitude. Struve [31, 1957, §37] and some German authors [10, p. 68; 11, p. 50] sided with Gauss; or, more precisely, they opposed subjective rejection.

Objective stochastic tests began to be applied to rejection in the second half of the 19th century [25; 5, vol. 2, pp. 558 – 566]. Contrary to Gauss' opinion, they did not take into consideration the causes of deviation. Some participants in the ensuing discussions stressed that it was reasonable to sacrifice a few possibly sound observations and to avoid the dangerous influence of large mistakes. That attitude was of course in line with modern statistical notions on errors of two kinds. The application of objective tests naturally demanded the knowledge of the appropriate distributions (the normal law was almost always presumed). Robust tests, i.e., such which hardly depended on deviations of the distributions from their assumed type, remained unknown; the only exception was apparently the criterion of three sigma [13]. A number of statistical tests (e.g., [30]) were offered in the mid-20th century, but the state of the issue, as Rider [26, pp. 21 – 22] formulated it, did not apparently change.²

Posterior weights. Another estimator

$$\hat{e} = \Sigma x_i p(\hat{e} - x_i) / \Sigma p(\hat{e} - x_i) \quad (1)$$

can be used instead of the arithmetic mean. Here $p(\hat{e} - x_i)$ are the posterior weights assigned to equally precise observations x_i in accordance with the distances $(\hat{e} - x_i)$ and \hat{e} had to be calculated by consecutive approximations. The weights might be discrete or continuous functions of their argument, and, from the

18th century onward, mathematicians and astronomers repeatedly proposed estimators (1). Some authors thought that posterior weights can allow for changing conditions of observation over long periods of time.

For symmetric distributions estimators (1) provide a correction to the arithmetic mean due to the deviation of the observations from pairwise symmetry. In addition, at least in the usual case of posterior weights decreasing to the tails of the distribution, their use enables to do away with rejection. In some instances (1) is at the same time the maximum likelihood estimator. Suppose indeed [3] that the density law is

$$\varphi(x) = r^2 - (\hat{e} - x)^2$$

with an unknown parameter \hat{e} . Then, according to the principle of maximum likelihood,

$$\Sigma \{(\hat{e} - x_i) / [r^2 - (\hat{e} - x_i)^2]\} = 0, \hat{e} = \Sigma p_i x_i / \Sigma p_i, p_i = 1 / [r^2 - (\hat{e} - x_i)^2].$$

In this case, the weights increased towards the tails. Daniel Bernoulli had not expressly indicated that fact and it might have remained overlooked.

The median. Possibly the most active partisan of the median was Estienne [7].³ Maintaining that random errors were only characterized by the symmetry of their density, he calculated the probability that, out of n observations, m will be negative, and $(n - m)$ positive. The probability was maximal at $m = n/2$ for an even n and at $m = (n - 1)/2$ for $n = 2m + 1$, hence the median. Estienne then formulated several properties of the median. In particular, he argued that it was closer to the true value of the constant sought than the arithmetic mean [even] if the smaller errors were more probable than the larger ones, and he also stated that the median was the most probable estimator if

$$\varphi(x) = k \exp[-|f(x) - f(a)|] \quad (2)$$

but did not specify $f(x)$. Estienne did not use the decrease of density to prove his first statement (which thus failed); moreover, it was formulated in a deterministic rather than stochastic sense.

Bervi [4] repeated many of Estienne's assertions and he also proved by a simple reasoning that

$$P(x_1 < X < x_n) = 1 - 2^{-(n-1)} \quad (3)$$

where X was the constant sought. Kornfeld [15] argued that the estimation of precision of observations should be restricted to the use of formula (3) but this was an anachronism.

According to modern notions [14], for some distributions the median is nevertheless preferable to the arithmetic mean; and, in particular, in the case of unknown densities. It would therefore be sensible to test the use of the median when treating the observations made by modern rangefinders since the densities of their errors are hardly known. In concluding, I note that Mendeleev [24] suggested to separate the variational series into three groups and choose as the estimator of the constant sought the arithmetic mean of the middlemost third, this being a peculiar combination of the median and the mean.

Notes

1. In the theory of errors, Lambert ([17, §§271 – 306]; and a large part of [18] and [19]) is the main predecessor of Gauss. He was the first to expound systematically many of its main issues, and to offer the very term, theory of errors. He also was the first to estimate methodically, but not successfully, the precision of observations (by the deviation of the arithmetic mean of all the observations from that of all of them excepting the most outlying observation) and even before Daniel Bernoulli he put forward the principle of maximum likelihood. At the same time, when deriving the law of distribution of certain observational errors, he issued not from their real properties, but from an alleged lack of causes for any other law.

2. See the passage in [27, p. 113]. The three-sigma test is due to Jordan [12]; the Charlier test, to Czuber [6, p. 206]; and the chi-squared distribution, to Abbe [1].

3. Estienne published two pertinent notes in the *C.r. Acad. Sci. Paris* (t. 130, 1900, pp. 66 – 69 and 393 – 395) and returned to his subject many years later [7a]. The comparison of the median with the arithmetic mean with respect to their precision began with Laplace (1818, the Second Supplement to his *Théorie analytique des probabilités*) and he [20] was also the first to introduce density of the type of (2).

Later addendum: The abstract of Estienne's paper mentioned in [7a] is complemented by a report on the ensuing discussion (Lévy, Hadamard). Lévy stated that, contrary to Estienne's opinion, the precision of the results increased with the number of observations (provided that the errors were not systematic); that the arithmetic mean was best for the normal distribution but the median might be preferable for other cases; that sometimes the mean square error "ne reste pas finie" which is "un argument sérieux" in Estienne's favor; but that it would then be better to reject the extreme observations "dans une proportion déterminée" and to take

the mean of those retained. Hadamard's remarks were less interesting: Experience proved that precision increased with the number of observations; the increasing precision of astronomical observations revealed that previous results obtained in the classical way by less precise measurements were exact [?]. H.L. Harter (1977, date of preface), *Chronological Annotated Bibliography on Order Statistics*, vol. 1. Wright-Patterson Air Force Base, Ohio, described this material but omitted Hadamard.

References

1. Abbe, E. Über die Gesetzmässigkeit in der Vertheilung der Fehler (1863). *Ges. Abh.*, Bd. 2. Olms, 1989, pp. 55 – 81.
2. *Algebra and Mensuration from the Sanscrit of Brahme Gupta and Bhascara*. Transl. H.T. Colebrooke. London, 1817.
3. Bernoulli, D. The most probable choice between several discrepant observations etc. (1778, in Latin, with companion commentary by Euler). *Biometrika*, vol. 48, 1961, pp. 1 – 18.
4. Bervi, N.V. Determination of the most probable value of a measured object irrespective of the Gauss postulate. [*Trudy*] *Mosk. Obshchestvo Liubitelei Estestvozn., Antropol. i Etnografii*, section phys. sci., vol. 10, No. 1, 1899, pp. 41 – 45. (R)
5. Chauvenet, W. *Manual of Spherical and Practical Astronomy*, vols 1 – 2 (1863). New York, 1960 (reprint of the edition of 1891).
6. Czuber, E. *Theorie der Beobachtungsfehler*. Leipzig, 1891.
7. Estienne, J.E. Etude sur les erreurs d'observations. *Rev. artill.*, t. 36, 1890, pp. 235 – 259.
- 7a. Estienne, J.E. Introduction à une théorie rationelle des erreurs d'observation. *Ibidem*, t. 97, 1926, pp. 421 – 441; t. 98, 1926, pp. 542 – 562 ; t. 100, 1927, pp. 471 – 487. Abstract : *Bull. Soc. Math. France*, sér. 2, t. 55, 1927, pp. 24 – 25.
8. Galilei, G. *Dialogue concerning the Two Chief World Systems* (1632, in Italian). Berkeley – Los Angeles, 1962.
9. Gauss, C.F. Brief nach Olbers 3.5.1827. *Werke*, Bd. 8. Leipzig, 1900, pp. 152 – 153.
10. Gerling, C.L. *Die Ausgleichungsrechnung der practischen Geometrie*. Hamburg – Gotha, 1843.
11. Hagen, G. *Grundzüge der Wahrscheinlichkeitsrechnung* (1837). Berlin, 1867.
12. Jordan, W. Über den Maximalfehler einer Beobachtung. *Z. Vermessungswesen*, Bd. 6, 1877, pp. 35 – 40.
13. Kemnitz, Yu.V. Estimating the precision of equally precise geodetic measurements whose errors have non-Gaussian laws of distribution. *Trudy Mosk. Inst. Inzhenerov Zemleustroistva*, No. 2, 1957. (R)
14. Kolmogorov, A.N. The method of the median in the theory of errors (1931, in Russian). *Sel. Works*, vol. 2. Dordrecht, 1992, pp. 115 – 117.
15. Kornfeld, M. On the theory of errors. *Doklady Akad. Nauk SSSR*, vol. 103, 1955, pp. 213 – 214. (R)
16. de Lalande, J.J., *Astronomie*, t. 3 (1771). New York – London, 1966 (reprint of the edition of 1792).
17. Lambert, J.H. *Photometria*. N.p., 1760.
18. Lambert, J.H. Anmerkungen und Zusätze zur practischen Geometrie. In author's book *Beyträge zum Gebrauch der Mathematik und deren Anwendung*, Tl. 1. Berlin, 1765, pp. 1 – 313.
19. Lambert, J.H. Theorie der Zuverlässigkeit der Beobachtungen und Versuche. *Ibidem*, pp. 424 – 488.
20. Laplace, P.S. Sur la probabilité des causes par les événements (1774). *Oeuvr. Compl.*, t. 8. Paris, 1891, pp. 27 – 65.
21. Leibniz, G.W. *Neue Abhandlungen über den menschlichen Verstand*, Bde 1 – 2 (1765). Frankfurt/Main, 1961.
22. Maire, C., Boscovich, R.J. *Voyage astronomique et géographique etc.* Paris, 1770.
23. Makovelsky, A.O. *Досократики* (Pre-Socratician Philosophers), vol. 1. Kazan, 1914.
24. Mendeleev, D.I. Progress of work on restoring the prototypes of measures of length and weight (1875). *Сочинения* (Works), vol. 22. L. – M., 1950, pp. 175 – 213.
25. Peirce, B. Criterion for the rejection of doubtful observations. *Astron. J.*, vol. 2, 1852, pp. 161 – 163.
26. Rider, P.R. *Criteria for Rejection of Observations*. Wash. Univ. Studies, New Ser., Sci. & Techn. No. 8, 1933.
27. Sheynin, O.B. Mathematical treatment of observations. *AHES*, vol. 11, 1973, pp. 97 – 126.
28. Simpson, T. On the advantage of taking the mean etc. *Phil. Trans. Roy. Soc.*, vol. 49, 1755 (1756), pp. 82 – 93.
29. Simpson, T. Same title. In author's book *Misc. tracts on Some Curious Subjects etc.* London, 1757, pp. 64 – 75.
30. Smirnov, N.V. On estimating the maximal term in a series of observations. *Doklady Akad. Nauk SSSR*, vol. 33, No. 5, 1941. (R)

- 31.** Struve, V. Ya. *Дуга меридиана* (An Arc of the Meridian). (1861). М., 1957.
- 32.** Veiman, A. A. *Шумеро-вавилонская математика и т. д.* (Sumerean-Babylonian Mathematics in Third – First Millenia B C). М., 1961.

5. On the History of the Adjustment of Indirect Observations
Izvestia Vuzov. Geodezia i Aerofotos'emka, No. 3, 1967, pp. 25 – 32

The determination of n unknowns x, y, z, \dots from the equations

$$a_i x + b_i y + c_i z + \dots + s_i = v_i, \quad i = 1, 2, \dots, m > n, \quad (1)$$

where v_i are the unavoidable residual free terms, was being done from the 18th century onwards when treating meridian arc measurements by imposing, directly or tacitly, various conditions on the values of v_i . One of the first to solve a problem of this type was Euler [7] who determined the figure of the Earth from four arcs. Excluding the parameters of the Earth's ellipsoid, he got two equations between the v_i 's (between the corrections to the lengths of one degree of the meridian), and, without applying any definite algorithm, restricted his efforts to rough estimations.

The first classical method of solving systems (1) was the **combination of the equations** (in pairs for the case of two unknowns). All possible combinations of two equations each were formed and the unknowns calculated for each such combination under a tacit assumption that $v_i = 0$. The final values of the unknowns were assumed to be the arithmetic means over all the combinations. Boscovich used this method to determine the parameters of the Earth's ellipsoid [22] but he made use of another method as well (below). In 1827, even after the introduction of least squares, Muncke [25, p. 872] followed suit.

Moreover, Boscovich [26] applied the same method for adjusting direct observations. In general, scientists of the 18th century attempted to treat both direct and indirect observations by a single algorithm, and the relation between the two cases was well understood as witnessed by the coincidence of terminology. Lambert [17, §6] applied the same word *Mittel* to designate both the arithmetic mean and the solution of systems (1); Lalande [15, §2699] used *milieu* in both these cases.

The method of combinations was also used for a qualitative estimation of precision, which, because of unavoidable systematic errors, should have hardly been based on deviations from the arithmetic mean. Tycho Brahe [5, p. 349] apparently pursued this goal when he, for the first time ever (and certainly before Boscovich, see above) applied the method of pairwise combinations for adjusting direct observations (of the distance between Venus and the Sun when the planet was to the east and to the west from the latter, and, as far as possible, with all other conditions being equal). Recalling the method of measuring angles in all combinations, we may ask whether Gauss arrived at it by issuing from the described method.

The method of means

$$\sum v_i = 0 \quad (2)$$

was applied by Tobias Mayer [24] who solved a system of 27 equations in three unknowns by forming three preliminary summary equations according to condition (2). Mayer was compelled, as he himself wrote, to introduce this method so as to avoid the difficult work of deriving and solving all the possible combinations of three equations. More precisely, he thus used a generalization of the method of means (which, in its pure form, allows to determine only one unknown). Condition (2) might be considered as the limiting case of the method of combinations with a single subset identical with the entire system.

At about the same time Euler [8, §115] actually applied the same method. Having obtained two equations

$$x = s_i + b_i y + c_i z + \dots, \quad i = 1, 2,$$

with pairwise roughly equal coefficients, he assumed that x was equal to the half-sum of their right sides.

Laplace [19, p. 121] mistakenly attributed the method of means to Cotes:

Cotes has prescribed that the equations of condition¹ be set out in such a way that the coefficient of the unknown element is positive in each of them and that all these equations be then added to form a final equation ...

Actually, however, Cotes [4] provided no equations and, in essence, his few lines ran as follows:

The point Z [the center of gravity] will be the most probable position of the thing which with the greatest plausibility may be considered its true position.

My statement does not detract from Cotes, who, incidentally, was well thought of by Newton. In his time, there were no quantitative substantiations of any particular method of treating observations. Even Legendre, more than 80 years later, did not justify least squares by anything other than qualitative considerations.

Like Eisenhart [6], I feel that condition (2) was understood in the 18th century as following from the equal probability of errors of each sign, and, as I shall add, as leading to the arithmetic mean in case of direct observations.

Lambert [17] used a condition of the type of (2) for fitting empirical straight lines and curves to points, – to observations ($x_i; y_i$). He divided the observations into two (for curves, into several) intervals with lesser and greater abscissas, determined the center of gravity in each interval and constructed the straight line or curve passing through these. Then, Cauchy [3; 21, Chapt. 14, §5] also used condition (2).

Boscovich [22, p. 501ff] pointed out the inadequacy of the method of combinations and proposed a new one so as to

obtain the mean in such a way that it would not be a simple arithmetic mean, but would conform to the rules of random combinations and calculation of probabilities according to a definite law ...

Specifically, he proposed to adjust the results of meridian arc measurements under three conditions, the first of which demanding that the dependence between the unknowns be of the type of (1). The other two were:

second, that the sum of the positive corrections be equal to the sum of the negative ones; third, that the sum of all the corrections, positive and negative, be minimal among those possible when the two first conditions are satisfied. ... the second condition is required for an equal degree of probability for the deviations of the pendulum² and errors of observation that increase or decrease the length of a degree. The third condition is necessary for a maximal insofar as possible approximation to observations ...

Boscovich' requirement of a *definite law* was legitimate; however, without mastering density functions he was naturally unable to say just how the *rules of random combinations* corresponded with his conditions. Later on Laplace [20, §40] used the Boscovich method and Gauss [9, §186] mistakenly attributed Boscovich' third condition to him.

Gusak [13] considered the history of the **minimax principle**

$$|v_{\max}| = \min \quad (3)$$

in which the minimum takes place for all possible solutions of (1) and traced it to the Chebyshev problem of the best approximation of an analytical function on a given segment by a polynomial of a certain degree.

Euler [7, §§122 – 123], about whom Gusak did not report, was the first to use this principle. Later on Laplace [20, Livre 3] and many other scientists applied it. Lambert [16, §420] knew the minimax principle and admitted that he was unable to devise an appropriate algorithm. Cauchy [2] busied himself with this problem. A.K. Uspensky had recently recommended the minimax principle for geodetic adjustments whereas in mathematical statistics it is applied in the theory of decision-making.

I shall now dwell on the **connections between least squares and the abovementioned principles**. The solution of (1) by least squares might be obtained [14] as

$$x = \frac{\sum \lambda_i \alpha_i}{\sum \lambda_i^2}, y = \frac{\sum \alpha_i \beta_i}{\sum \lambda_i^2}, \dots$$

where $\alpha_i / \lambda_i, \beta_i / \lambda_i, \dots$ are the solutions of all the possible subsystems of n equations isolated from (1). The least-squares solution differs from the one obtained by the method of combinations in that the weights of the partial solutions are there taken into account. In addition, the weights of [the estimators of] the unknowns can be calculated in a similar way by issuing from the appropriate partial weights [12].

Then, both Gauss [9, §186] and Laplace [18, §24] noted that the principle

$$\lim (v_1^{2k} + v_2^{2k} + \dots + v_n^{2k}) = \min \text{ as } k \rightarrow \infty$$

(which, in the case of large but finite values of k may be considered as a generalization of least squares) leads to the minimax principle. Indeed, for any sufficiently large k , the term v_i^{2k} with $v_i^2 = \max (v_j^2)$ will exert the greatest influence so that the minimax condition will be fulfilled.

It is usually thought that **Laplace and Gauss** approached the principle of least squares from considerably differing viewpoints. Tsinger [27, p. 1] asserted that Laplace had made

a rigorous [?] and impartial investigation; it can be seen from his analysis that least squares provide results having more or less significant probability only when the number of observations is large.[...] Gauss, on the basis of extraneous considerations, attempted to attach to this method an unconditional significance ... it will be easy to see the correctness of Laplace's conclusion; but with a limited number of observations we cannot count on a mutual cancellation of errors and ... any combination of observations can ... just as well lead to an increase of error as to its diminution.

Tsinger exaggerated: the arbitrariness of the principle of maximum weight does not yet mean that it is unsuitable; practice had long ago refuted such a conclusion. And Markov [23] unreservedly supported this principle (without ignoring its arbitrariness). It is hardly proper to set off Laplace against Gauss. Their common interest in the treatment of observations enabled these scholars to imagine better the problem that faced them and to approach their goals with clearer understanding of the general situation. And Laplace was no armchair scientist. In particular, he actively participated in the introduction of the metric system of measurements and in the determination of the figure of the Earth which means that he could have hardly restricted his attention to limit theorems. Indeed, he [19, p. 46] pointed out that “it seems natural to use” the method of least squares even when the number of observations was small. And (p. 48)

The optimal procedure is clearly that for which the same error in the results is less probable than it would be under any other procedure.

In the same chapter of [19] Laplace several times returned to his idea about the principle of maximum weight and connected its application with the need for the most rapid decrease of the density function. He understood *weight* as the positive parameter k of a law of the type $\exp(-kx^2)$ and pointed out that the weight of the mean result “increased like the number of observations divided [?] by the number of parameters” (p. 45).

Again in the same source [19, p. 123] Laplace additionally stated that

The slight uncertainty that the observations, when there are not very many of them, leave about the values of the constants [...] causes a slight uncertainty in the probabilities determined by the analysis. But it is almost always enough to know if the probability that the errors in the observed results are contained within narrow limits approaches closely to 1; and when this is not the case, it is enough to know just how many more observations should be taken in order to obtain such a probability that no reasonable doubt remains about the quality of the results.

Much of the above is also contained in [18, Supplements 1 and 2]. Thus, the optimal result corresponds to the maximal weight, the weight is inversely proportional to the sum of the squares of the deviations. At the same time, weight is a parameter of the normal law and its maximum corresponds to the minimal probability of errors or the minimal length of the “confidence interval”. The principle of maximum weight is thus formulated, but actually reduced to “confidence probability” with a “confidence interval” of minimal length which makes it impossible to dispense with an assumption of a definite (of the normal) law. Laplace unquestionably issued from the theorem now called after De Moivre and him.

It might be assumed that these thoughts essentially assisted Gauss, but the latter did not mention them. I describe now how he developed the concept of weight. There is no such notion in his *Theoria motus* [9] where we find only *Genauigkeitsgrad* (§173). ... Gauss actually understood it as the root of the weight. However, he also introduced the *Maass der Genauigkeit* h , a parameter of the distribution

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

but he did not mention the analogy between the *Genauigkeitsgrad* and this *Maass*.

In [10, §3], issuing from the maximal value of the function

$$h^m \exp[-h^2(\alpha^2 + \beta^2 + \gamma^2 + \dots)]$$

(where $\alpha, \beta, \gamma, \dots$ were observational errors, m in number), proportional, as it would be said now, to the likelihood, Gauss derived the “most probable” relation

$$h = \{m/[2(\alpha^2 + \beta^2 + \dots)]\}^{1/2}. \quad (4)$$

Finally Gauss [11, §6] introduced the *mittleren zu befürchtendem Fehler*, the mean error to be feared (*jac-tura*) m^2 , as he called it in §7, and noted (§9) that, for

$$\begin{aligned} \varphi(x) &= (1/h\sqrt{\pi}) \exp(-x^2/h^2), \\ m &= h/\sqrt{2} \end{aligned} \quad (5)$$

and called the magnitude inversely proportional to m^2 the relative weight (*Gewicht*). He could have derived a formula of the type

$$m = [\sum \Delta_i^2 / n]^{1/2} \quad (6)$$

by issuing from (4) and (5) but he did not proceed in such a manner apparently because the result would have depended on the existence of the normal distribution. This fact is extremely important; Gauss had indeed obtained a formula of this type (in §15) but independently of the density. “The rule of least squares was already concealed” there [27, §13].

In §38 Gauss generalized his finding onto the case of several unknowns. According to the context, he was concerned with deriving m through the deviations from the adjusted values and he obtained

$$m = [(\lambda_1^2 + \lambda_2^2 + \dots) / (\pi - \rho)]^{1/2} \quad (7)$$

where the meaning of π and ρ is obvious. His working formula was therefore (7) rather than a generalization of (6). Gauss himself, in his *Anzeige* of [11], noted that the Δ_i in (6) and in its generalization were “always” taken to be the most probable deviations but that it was now possible to apply the more precise formula (7) and thus to observe the *Würde der Wissenschaft*. I doubt that the formula (7) should be called after Bessel. The only writing where he could have preceded Gauss is [1] but (7) is lacking there.

After Gauss’ lifetime the **theory of errors** became an engineering discipline with an established sphere of solved problems and its development mostly followed a “technological” direction (its application to the treatment of various geodetic constructions). However, beginning roughly in the 1920s, the theory became a chapter of mathematical statistics although statistical methods (mostly correlation theory and analysis of variance) have been until recently only applied in geodetic literature for special investigations. Without disparaging these at all, it might be said that they did not touch on the essence of the theory of errors.

Quite recently a number of articles on confidence estimation in the theory of errors have appeared; however, neither did this fact essentially change anything since the classical mean square error is also related to such estimation. The basic content of the theory of estimation as applied to the treatment of observations is the attempt to use more fully the information provided by each observation by means of order statistics. According to Gauss, the arithmetic mean of equally precise measurements, independently from the appropriate (but not “bad”) law of distribution, had minimal variance among linear estimators. This point of view is somewhat dated. It is now possible to arrange the observations in ascending (say) order, and to take into account the information furnished by each of them. For example, observations might receive weights depending on their distribution and number. Extreme observations can obtain weights larger than those of the other ones, some weights might even be negative. This very approach, entirely different from the classical notions, increases the precision of treating observations and actually determines the divide between the classical theory and mathematical statistics.

Gauss’ “second method” is now inadequate and methodologically even imperfect. The arithmetic mean is the best estimator only under normality; decrease of variance is provided by taking account of the dependence between terms of the variational series. Interestingly, no decrease is possible here for the case of the normal law.

To what degree is the increase of precision real? Geodesy, characterized by effective and multiform checking of observations, is the proper sphere for verifying this. A final comment: posterior weights had also been introduced time and time again in the classical error theory and they were assigned to equally precise observations depending on their place in the variational series [26]. They were determined almost without taking into account the appropriate distributions but at least they also aimed at improving the classical arithmetic mean.

Notes

1. These differ from equations (1) and correspond to the second main version of adjustment of observations.

2. Pendulum observations provide the possibility of obtaining the flattening of the Earth’s ellipsoid of revolution.

References

1. Bessel, F.W. Untersuchungen über die Bahn des Olbersschen Kometen. *Abh. Preuss. Akad.* [Berlin], math. Kl., 1812 – 1813 (1816), pp. 119 – 160.
2. Cauchy, A.L. Sur le système de valeurs etc. (1831). *Oeuvr. Compl.*, sér. 2, t. 1. Paris, 1905, pp. 358 – 402.
3. Cauchy, A.L. Sur l'évaluation d'inconnues etc. *C.r. Acad. Sci. Paris*, t. 36, 1853, pp. 1114 – 1122.
4. Cotes, R. Aestimatio errorum etc. (1722). *Opera misc.* London, 1768, pp. 10 – 58.
5. Dreyer, J.L.E. *Tycho Brahe*. Edinburgh, 1890.
6. Eisenhart, C. Boscovich and the combination of observations (1961). *Studies in History of Stat. and Probability*, vol. 2. Editors, Sir Maurice Kendall, R.L. Plackett. London, 1977, pp. 88 – 100.
7. Euler, L. Recherches sur des inégalités du mouvement etc. (1749). *Opera omnia*, ser. 2, t. 25. Zürich, 1960, pp. 45 – 157.
8. Euler, L. Eléments de la trigonométrie sphéroïdique etc. (1755). *Ibidem*, t. 27. Zürich, 1954, pp. 309 – 339.
9. Gauss, C.F. *Theoria motus* (1809). German transl.: Aus der Theorie der Bewegung etc. In author's book *Abh. zur Methode der kleinsten Quadrate*. Hrsg. A. Börsch, P. Simon. Berlin, 1887, pp. 92 – 117. Latest edition: Vaduz, 1998.
10. Gauss, C.F. Bestimmung der Genauigkeit der Beobachtungen (1816). *Ibidem*, pp. 129 – 138.
11. Gauss, C.F. *Theoria combinationis* etc. (1823 – 1828). German transl.: Theorie der den kleinsten Fehlern unterworfenen Combination der Beobachtungen. *Ibidem*, pp. 1 – 91.
12. Gleinšwik, P. The generalization of the theorem of Jacobi. *Bull. Géod.*, t. 85, 1967, pp. 269 – 281.
13. Goussac, A.A. La préhistoire et les débuts de la théorie de la représentation approximative des fonctions. *IMI*, vol. 14, 1961, pp. 289 – 348. (R)
14. Jacobi, C.G.J. Über die Bildung und die Eigenschaften der Determinanten (1841, in Latin). *Ostwald Klassiker No. 77*. Leipzig, 1896, pp. 3 – 49.
15. de Lalande, J.J. *Astronomie*, t. 3. Paris, 1771. Reprint of the third edition (1792): New York – London, 1966.
16. Lambert, J.H. Anmerkungen und Zusätze zur practischen Geometrie. In author's *Beyträge zum Gebrauche der Mathematik und deren Anwendung*, Tl. 1. Berlin, 1765, pp. 1 – 313.
17. Lambert, J.H. Theorie der Zuverlässigkeit der Beobachtungen und Versuche. *Ibidem*, pp. 424 – 488.
18. Laplace, P.S. *Théorie analytique des probabilités* (1812). *Oeuvr. Compl.*, t. 7. Paris, 1896.
19. Laplace, P.S. *Philosophical Essay on Probabilities* (1814, in French). Transl. from the edition of 1825: New York, 1994.
20. Laplace, P.S. *Celestial Mechanics*, vol. 2 (1799, in French). Transl. by N. Bowditch. Boston, 1832; reprinted, New York, 1966.
21. Linnik, Yu.V. *Method of Least Squares* etc. (1958, in Russian). Oxford, 1961.
22. Maire, C., Boscovich, R.J. *Voyage astronomique et géographique* etc. Paris, 1770.
23. Markov, A.A. The law of large numbers and the method of least squares (1898, in Russian). Transl.: DHS 2514, 1998, pp. 157 – 168.
24. Mayer, T. Abhandlung über die Umwälzung des Mondes um seine Axe. *Kosm. Nachr. u. Samml.*, 1748 (1750), pp. 52 – 183.
25. Muncke, Erde. *Gehlers phys. Wörterbuch*, Bd. 3. Leipzig, 1827, pp. 825 – 1141.
26. Sheynin, O.B. On selecting and adjusting direct observations (1966). Transl. in this collection.
27. Tsinger, V.Ya. *Способ наименьших квадратов* (Method of Least Squares). A thesis. M., 1862.

6. Some Issues in the History of the Theory of Errors

Abstract of dissertation. Moscow, 1967.

Published as a manuscript. Inst. Istorii Estestvoznania i Tekhniki

The rise of the theory of errors as a discipline belonging to experimental quantitative science is connected with the spectacular successes of instrumental astronomical observations of the 17th century and with the beginning of the epoch of meridian arc measurements (17th – 18th centuries). The advances in, and subsequent new problems of astronomy put onto the agenda various issues concerning the treatment and evaluation of the precision of instrumental observations; and the calculation of arcs of triangulation and the determination of the figure of the Earth demanded, furthermore, the ability to treat redundant systems of linear algebraic equations. The theory of errors thus naturally included a number of problems pertaining to the treatment and estimation of the precision of direct and indirect observations. The treatment of direct observations led to the justification, of qualitative, and then of a quantitative, stochastic nature, of the long prevalent arithmetic mean. The treatment of indirect measurements demanded the development of a number of algorithms which were independent of probabilistic considerations; later, however, Gauss derived the method of least squares by issuing from the principle of maximum likelihood.

In the process of my work, I became acquainted with a large number of sources in mathematics, astronomy and geodesy. A large part of this literature remained little known or completely ignored. This very situation can explain to a certain degree why I succeeded in getting unexpected important results and among them

1. The description of the work of the forgotten American mathematician, Robert Adrain, who, in particular, published two derivations of the normal distribution in the theory of errors a year before Gauss [or at least not later than he].

2. The establishment of the priority of Ernst Abbe in considering the chi-squared distribution.

3. A critique of the derivation of the normal distribution as given by Gauss in 1809.

The literature on the theory of errors and the method of least squares paid great attention to the justification of the normal law. That problem occupies only a secondary place in my work. Indeed, this is perhaps a separate topic in which, first of all, it is necessary to examine the theory of elementary errors and the central limit theorem. And a large number of other derivations turned out to be dead ends and today represent only mathematical exercises.

My work consists of three chapters, an appendix, and an addendum. **Chapt. 1** investigates the early application of the arithmetic mean in approximate calculations, games of chance, in astronomy and theory of probability proper. That it was used in antiquity has been known for a long time. However, the connection of this fact with the stochastic nature of the mean was not noted, – and precisely this fact is stressed throughout almost the entire chapter. Furthermore, a large amount of factual material is collected there.

The science of mean values, including the arithmetic mean, already existed in the Pythagorean school, and the latter had been widely applied in antiquity in approximate calculations of the areas of figures and volumes of bodies. The formula for the calculation of the area of a quadrangle as the product of half the sums of opposite sides was used in ancient Babylonia either for not exact rectangles or when the opposite sides were unequal because of the ruggedness of the terrain (A.A. Vaiman). This means that the arithmetic mean was invoked to compensate both for the lack of strict applicability of the formulas (models) and for the systematic (not random, as nowadays) errors of measurement.

Commercial practice aided the spread of the idea of the arithmetic mean, and claims were even made that in this sense the sphere of economics was primary. According to Leibniz, the principle of equal allowance of equally tenable assumptions was the fundamental hypothesis in the contemporaneous theory of probability. I believe that Leibniz thought about the origin of the first stochastic concepts, probability and expectation, and that his idea provided the possibility of the later subjective understanding of probability.

The history of games of chance contains evidence of the widespread knowledge of the idea of the arithmetic mean which served as an intuitive statistical indicator of the totality of possible outcomes and perhaps of the appropriate expectation. During the epoch of meridian arc measurements the arithmetic mean began to be used as an universal estimator in all stages of their treatment. In the same period and even earlier the first qualitative statements about the benefits of applying it had appeared (Picard, Condamine).

In 1809 Gauss postulated the principle of the arithmetic mean and essentially used it in his derivation of the normal distribution. This attracted attention to his postulate and a number of authors tried to reduce it to a “more obvious” premise. These attempts were, however, of a purely deterministic nature and I do not dwell on them. In the first half of the 18th century Cotes applied an analogy from mechanics (the center of gravity) for justifying the arithmetic mean, and Lambert tried to substantiate it on a stochastic basis.

In **Chapt. 2**, in connection with the history of the treatment of direct observations, I studied the work of Galileo and Lambert, then dwelt on the appropriate memoirs of Simpson and Lagrange. Also there I investi-

gated estimators with posterior weights, the principle of maximum likelihood and the rejection of outlying observations.

Galileo was the first to formulate a number of basic theorems in the theory of errors (Maistrov). Lambert, who laid the foundation for that theory, was the chief predecessor of Gauss in this direction. When substantiating the advantages of the arithmetic mean, Simpson issued from the objective properties of observational errors. His immediate aim was to refute the opinion of “some persons of considerable note” that one careful observation can be relied on as much as on the mean of a great number of them, and in this connection I studied the works of Flamsteed and Bradley. I concluded that the abovementioned opinion, to which these astronomers never subscribed, was the result of the great successes of observational techniques and did not have a lasting influence on experimental science. Lagrange, without mentioning Simpson, reproduced his results and studied several continuous distributions.

Denote the direct observations of an unknown constant x by x_i , $i = 1, 2, \dots, n$, then

$$\hat{e} = \frac{\sum x_i p(\hat{e} - x_i)}{\sum p(\hat{e} - x_i)} \quad (1)$$

will be an estimator of x with posterior weights $p(\hat{e} - x_i)$. Such estimators with discrete or continuous posterior weights had repeatedly been proposed beginning with the second half of the 18th century (Short, Euler, De Morgan, Newcomb, Ogorodnikov). Some authors thought that posterior weights could allow for the change over time of the parameters of the appropriate law of distribution. In my opinion, with an even law (a natural assumption) estimators (1) only provide a correction to the ordinary arithmetic mean for the deviation of the observations from pairwise symmetry. However, these estimators may be considered as an historical analog of some modern statistical estimators.

In 1778 Daniel Bernoulli proposed an estimator with posterior weights increasing towards the tails of the distribution adopted (an arc of a parabola). He sharply criticized the arithmetic mean considering it suitable only for uniform distributions; instead, he proposed the principle of maximum likelihood arriving at estimator (1). The unusual behavior of the posterior weights would have seemed unacceptable; however, Euler, in a companion commentary, mistakenly concluded that the weights decreased to the tails. Incidentally, in such “unusual” cases posterior weights are no alternative for rejecting the outliers.

I have shown the similarity in the use of the principle of maximum likelihood by Adrain (1808 [or 1809]) and Gauss (1809) for deriving the principle of least squares and the arithmetic mean and investigated the justification of maximum likelihood by inverse probability (Laplace, Gauss).

Rejection of outliers was recommended by Galileo and systematically applied by Lambert. Stochastic criteria for rejection were only devised in the second half of the 19th century. Their appearance was inevitable both because of the desire to abandon arbitrary rejection and of the expansion of the domain of applications of probability. On the other hand, the development of such tests was delayed by the fetish made of the normal law according to which any error was possible and perhaps by the opinion of Gauss who allowed rejection only in cases of gross errors.

The first stochastic criteria for rejecting outliers (Pierce, Chauvenet and others) were based on direct calculation, according to the normal law, of the odds for and against dubious observations and gave rise to a drawn-out polemic where opinions in essence leading to the consideration of errors of the two kinds were expressed: better to sacrifice a few possibly reliable observations but get rid of the dangerous influence of large errors. Thus, Gauss notwithstanding, these criteria rejected large errors regardless of their origin. The errors of these tests resulting from small divergences of the distribution of observational errors from normality (non-robustness of criteria) were not investigated.

At the end of the chapter I mention that a quantitative estimation of precision began to be used relatively late. With the exception of Lambert (who, for that matter, did not norm his measure of precision and therefore could not directly compare several series of observations), no-one until Gauss (1823) introduced any such measure. Incidentally, I found a normed estimator of precision in a work of Delambre written sometime during 1818 – 1822.

In **Chapt. 3** I study the history of the mathematical treatment of indirect observations (the solution of redundant systems of linear algebraic equations with the help of some supplementary conditions imposed on the residual free terms v_i). Above all, for the frequently occurring case of two unknowns (in particular, the unknown parameters of the terrestrial ellipsoid of revolution), I investigated the practice of using pairwise combinations of measurements (condition: $v_i = 0$). A similar method was traced up to 1827. In addition, I discovered that the method of pairwise combinations was used while treating direct observations (Boscovich) with the subsequent calculation of the arithmetic mean over all the combinations. I assumed that the combinations were applied here for a qualitative evaluation of the precision of observations, and, in addition, so as to apply a single algorithm for treating both direct and indirect observations.

I also traced the connection of the method of pairwise combinations with the method of means ($\sum v_i = 0$). Tobias Mayer (1750), while solving a system of 27 equations in three unknowns, grouped them in three

summary equations. He justified this procedure (a generalized method of means) by the practical impossibility of forming and solving all the possible combinations of the equations in threes. The condition of the method of means was naturally regarded as resulting from the equal probability of errors of each sign and leading to the arithmetic mean in the case of direct observations.

I also describe the treatment of meridian arc measurements by Boscovich. Not being satisfied with the method of combinations, he proposed the conditions

$$\sum v_i = 0, \sum |v_i| = \min$$

which were also used later on by Laplace.

A.A. Gusak described the history of the minimax method (condition: $\text{Max } |v_i| = \min$ with the minimum being taken among all possible solutions of the system). I supplemented his account by several remarks and indicated, in particular, that Euler had applied elements of this method in 1749 (not in 1778).

Gauss (1809) derived the method of least squares on the basis of the normal distribution of random observational errors. However, their usual properties were only connected with this law through the arithmetic mean. In 1823 Gauss published his second derivation of least squares by issuing from the principle of greatest weight. He thus renounced his previous tacit assumption that the normal law was the only possible law of error. The principle of maximal weight for a finite series of measurements was already known to Laplace who thought that the optimal result corresponded to maximum weight with the weights being inversely proportional to $\sum v_i^2$. But at the same time Laplace defined the weight as the positive parameter k of the law of the type $\exp(-kx^2)$ and he reduced the condition of its maximum to the least probability of errors, or to the shortest length of a “confidence interval”. This point of view did not permit him to renounce the normal law as the universal law of error. However, one should recognize a greater similarity of Laplace’s and Gauss’ ideas than it is usually recognized; it is hardly opportune to contrast these great men of science to each other. Precisely their community of interests in treating observations enabled each of them to formulate better the unsolved problems, and, when attacking them, to rely on the results of each other.

In the **Appendix**, I cite short biographical data on the American mathematician Robert Adrain (1775 – 1843) and investigate in detail his work in the theory of errors. I also trace a number of later derivations of the normal distribution. In 1808 [or 1809] Adrain published an article which contained two derivations of the normal law of random observational errors; a derivation of the principle of least squares (it was supposed, however, that Adrain was acquainted with the work of Legendre) and of the arithmetic mean; a determination of the most probable position of a ship from dead reckoning and an observation of its latitude; and an adjustment of a closed compass traverse.

In 1818 Adrain published two articles devoted to the derivation of the flattening and the size of the Earth. Using the data collected by Laplace, he applied the principle of least squares and obtained 1/319 for the flattening. In his second article, he arrived at a remarkably good estimate of the Earth’s radius. His results were wonderful, but of course his work cannot be compared with the achievements of Gauss either directly or even less in its significance for the later development of the theory of errors since he remained virtually unknown.

In the **Addendum** I attempted to sketch a general outline of the history of the theory of errors, and, in particular, to explain the reason for the existence of two versions of the theory, the *mathematical-statistical*, and the *astronomical-geodetic*.

Having worked without a scientific mentor, I consider it my duty to mention with even greater appreciation the participants and the heads of the seminar on the history of mathematics and mechanics at the Moscow Lomonosov State University. The atmosphere at the seminar in general, as well as the advice received, essentially helped me.

Judith A. Behrens and Walter L. Sadowski had translated this piece about thirty years ago; I have revised their work.

7. On the Work of Bayes in the Theory of Probability

Trudy XII Nauchn. Konf. Aspirantov i Mladsh. Nauchn. Sotrudn. Inst. Istorii Estestvoznania I Tekhniki, Sektsia istorii mat. i mekh. nauk. Moscow, 1969, pp. 40 – 57

Subsequent note

I have examined both parts of the Bayes memoir [2]. Since the first part was reprinted [3] and studied by many authors; and since Wishart, whose work [5] I overlooked, had discussed the second part, I am translating only a few lines from my paper.

* * *

[...] In a posthumous note [1] Bayes proved that the series applied when calculating the sum

$$\log 1 + \log 2 + \dots + \log n$$

by the Stirling formula was divergent. He maintained that only a certain number of the terms of that series might be taken into account and concluded that non-rigorous methods should not be trusted. Actually, Bayes directed all this against De Moivre who had applied the Stirling formula for deriving the “De Moivre – Laplace limit theorem”. [...]

De Moivre proved that

$$P \{ \alpha \leq [(\mu - np) / \sqrt{npq}] \leq \beta \} \sim (1 / \sqrt{2\pi}) \int_{\alpha}^{\beta} \exp(-x^2/2) dx \quad (1)$$

where μ was the number of successes in n independent Bernoulli trials with probability of success p , $q = 1 - p$ and $n \rightarrow \infty$. H could not have known that $\text{var } \mu = npq$.

Neither Bayes nor Price considered the case of $n \rightarrow \infty$, and the latter indicated, in his supplement to pt. 1 of the Bayes memoir, that the De Moivre formula was only applicable in that case, and, furthermore, only when p (or q) was not small. Nevertheless, Timerding [4] modified the Bayes formula from pt. 2 and effectively obtained

$$P \{ -z \leq [(\bar{p} - a) / (\sqrt{pq} / n\sqrt{n})] \leq z \} \sim (\sqrt{2} / \sqrt{\pi}) \int_0^z \exp(-x^2/2) dx.$$

Here, p and q were the numbers of successes and failures in n trials ($p + q = n$), $a = p/n$ and \bar{p} was the unknown probability of success in a single trial. For my part, I note that $E \bar{p} \approx a$ and $\text{var } \bar{p} \approx pq/n^3$.

It is remarkable that Bayes, who certainly did not know anything about variances, was apparently able to perceive that an elementary and formal transformation of the left side of (1) leading to

$$P(\alpha \leq \{[(\mu/n) - p] / \sqrt{pq/n}\} \leq \beta)$$

would not provide the proper answer to his problem.

References

1. Bayes, T. A letter ... to J. Canton. *Phil. Trans. Roy. Soc.*, vol. 53, 1763 (1764), pp. 269 – 271.
2. Bayes, T. An essay towards solving a problem in the doctrine of chances. Communicated and commented upon by R. Price. *Ibidem*, pp. 360 – 418 (pt. 1) and vol. 54, 1764 (1765), pp. 296 – 325 (pt. 2).
3. Bayes, T. Reprint of pt. 1 of his main memoir. *Biometrika*, vol. 45, 1958, pp. 296 – 315. Preceded by a biographical note by G.A. Barnard (pp. 293 – 295) who did not even mention the existence of pt. 2 of the memoir.
4. Bayes, N. *Versuch zur Lösung eines Problems der Wahrscheinlichkeitsrechnung*. Hrsg. H.E. Timerding. *Ostwald Klassiker* No. 169. Leipzig, 1908.
5. Wishart, J. On the approximate quadrature of certain skew curves with an account of the researches of Bayes. *Biometrika*, vol. 19, 1927, pp. 1 – 38 + 442.

8. On the History of the De Moivre – Laplace Limit Theorem¹

Istoria i Metodologia Estestven. Nauk, vol. 9, 1970, pp. 199 – 211

1. Jakob Bernoulli's *Ars Conjectandi* greatly influenced the development of the theory of probability and was always considered a classic. In particular, Laplace [21, p. 118] highly praised it. I touch on its pt. 4 that contains Bernoulli's law of large numbers (a term due to Poisson) and discuss Karl Pearson's extremely negative opinion about it.

The essence of Bernoulli's law is as follows. At first, he considers the binomial $(r + s)^{nt}$ where $t = r + s$, n is a large number and r and s are natural numbers. He proves that, for a sufficiently large nt , the sum of the $2n$ middle terms, even excluding the middlemost one, will become c times greater ($c > 0$) than the sum of its other terms. Bernoulli then makes use of that algebraic fact in his stochastic reasoning.

Let $p = r/(r + s)$ be the probability of success in each of nt (independent) trials. Then, given a sufficiently large number of these, the probability that the number of successes μ is within the boundaries $n(r \pm 1)$ can be made c times greater than the probability of the contrary event with c being fixed beforehand. In other words, Bernoulli proved that

$$\lim P [|(\mu / n) - p| < \varepsilon] = 1 \text{ as } n \rightarrow \infty.$$

Both Markov [22, pp. 44 – 52] and Pearson [26] described the appropriate mathematical steps in detail.

Bernoulli then inverts his problem and maintains, without any special proof, that if some (posterior) probability of success at any trial $p = r/(r + s)$ is obtained after nt trials, then the probability that the true value of p lies within $[p \pm 1/(r + s)]$ can also be made c times higher than the probability of the contrary event. He also provided a somewhat lesser known estimate: for $r = 30$ and $s = 20$ [and, consequently, for $t = 50$ and $1/(r + s) = 0.02$] it occurred that $c = 1,000$ for $nt = 25,500$; $c = 10,000$ for $nt = 31,258$, etc. Thus, when nt increases by 5,758, c increases tenfold. It was hardly noted that this estimate means that

$$nt = 25,500 + 5,758 \lg (c / 1,000) = 8,226 + 5,758 \lg c, \quad (1)$$

or that

$$c = 10^{(nt - 8,226) / 5,758} \quad (1')$$

which is not difficult to write down for base e .

Bernoulli did not aim at estimating the change in c with the change in the boundaries of the number of successes. Expressions (1) and (1') are deterministic relations between nt and c and they show that Bernoulli effectively formulated his law as a local limit theorem.

In 1913, two hundred years after the publication of the *Ars Conjectandi*, its pt. 4 was translated into Russian under Markov's editorship. The same year Markov put out a third, a jubilee, as he called it, edition of his treatise [22] and supplied it with Bernoulli's portrait. And, again in 1913, the Imperial [Petersburg] Academy of Sciences organized a special sitting devoted to Bernoulli's work in probability with Markov, Vasiliev² and Chuprov reading their reports. However, only in 1924, in the posthumous edition of his treatise, Markov [22, 1924, pp. 44 – 52] improved Bernoulli's numerical estimate (above) obtaining 16,655 instead of 25,500. He ensured the main correction (17,324) by specifying Bernoulli's intermediate inequalities. He did not apply the Stirling formula, apparently because Bernoulli naturally had not known it. And Markov's residual correction followed from his abandoning the condition that nt be divisible by $r + s = t$.

Pearson [26] attained even better results by means of the Stirling formula and secured a practically precise coincidence of his estimate with what would have followed from the normal distribution as the limiting law for the binomial. He (pp. 202 and 210) concluded that Bernoulli

adopted a very crude method of inequalities [...] He gets most exaggerated values for the needful number of observations, and for this reason his solution must be said to be from the practical standpoint a failure; it would ruin either an insurance society or its clients, if it were adopted. All Bernoulli achieved was to show that by increasing the number of observations the results would undoubtedly fall within certain limits, but he failed entirely to determine what the adequate number of observations were for such limits. That was entirely De Moivre's discovery.

After all, I think, we must conclude that it is somewhat a perversion of historical facts to

call the method [...] by the name of the man who after twenty years of consideration had not got further than the crude values[...] with their 200 to 300 per cent. excesses. Bernoulli saw the importance of a certain problem; so did Ptolemy, but it would be rather absurd to call Kepler's or Newton's solution of planetary motion by Ptolemy's name! Yet an error of like magnitude seems to be made when De Moivre's method is discussed without reference to its author, under the heading of "Bernoulli's Theorem". The contributions of the Bernoullis to mathematical science are considerable, but they have been in more than one instance greatly exaggerated. The Pars Quarta of the Ars Conjectandi has not the importance which has often been attributed to it.

Pearson's opinion is hardly correct since the practical uselessness of the Bernoulli estimate is not that important (to say nothing about his impossibility of applying the Stirling formula). On the contrary, I stress that the very existence of that estimate and of Bernoulli's law of large numbers was extremely essential. Pearson [25, p. 404] also noted that Bernoulli did not provide a measure of precision determined by $n^{-1/2}$. However, we should not fault Bernoulli for that either. Properly praising De Moivre, whose merits had been attributed to Bernoulli by all French and German authors known to him, Pearson at the same time profaned a great scholar.

2. Niklaus Bernoulli estimated the ratio of the middle part of the binomial series to its other parts and applied his calculations to a stochastic deduction concerning the sex ratio at birth. He communicated his results to Montmort in a letter of 23 January 1713 and the latter included them in his book [24, pp. 388 – 394] published the same year, before or at least independently from the appearance of the *Ars Conjectandi*.

Niklaus issued from Arbuthnot's data³. Denote the sex ratio (boys / girls) by m/f , the yearly number of births by n of which μ are boys, the i th term of the binomial $(m + f)^n$ by u_i , and introduce

$$s_1 = u_{fr+1} / u_{fr-l+1}, \quad s_2 = u_{fr+1} / u_{fr+l+1}, \quad l = o(\sqrt{n}), \quad t = \min(s_1; s_2).$$

Then, as Niklaus approximately calculated,

$$P(|\mu - rm| \leq l) = (t-1)/t,$$

$$t \approx \{1 + [l(m + f) / mfr]\}^{l/2} \approx \exp[l^2(m + f)^2 / 2mfn],$$

so that, denoting $p = m / (m + f)$ and $q = f / (m + f)$,

$$P(|\mu - rm| \leq l) \approx 1 - \exp(-l^2 / 2pqn) \quad (2)$$

where $pqn = \text{var } \mu$.

Thus, Niklaus indirectly arrived at the normal distribution. It is easily seen, however, that his formula is indeed not applicable for large values of l ; for example,

$$P(|\mu - rm| \geq 0) \neq \int_0^\infty \exp(-l^2 / 2pqn) dl$$

since this integral is not equal to unity.⁴

3. A French national, De Moivre (1667 – 1754) [2 ; 23 ; 33, pp. 135 – 136 ; 34 ; 8]⁵ was forced to leave France after the revocation of the Edict of Nantes (1685). His mathematical education (his teacher was Ozanam) occurred to be patently insufficient but he managed to fill in the gaps in his knowledge all by himself and was even elected to the Royal Society (1697). Newton favored and respected him (De Moivre actively participated in editing the Latin version of Newton's *Optics*), and, in his later years, habitually referred those, who asked him questions of a mathematical nature, to De Moivre. When the Royal Society appointed a commission for deciding the priority strife between Newton and Leibniz with regard to the analysis of infinitesimals, De Moivre was elected its member (another member was Arbuthnot).

Todhunter [33, §233] correctly noticed that

In the long list of men ennobled by genius, virtue and misfortune, who have found an asylum in England, it would be difficult to name one who has conferred more honor on his adopted country than De Moivre.

This, however, is not the whole truth. De Moivre's new homeland did not at all secure him a worthy way of life. He was never able to take up a permanent position and had to support himself by private lessons and consultations. In 1735 De Moivre was elected to the Berlin Academy of Sciences, and in 1754, shortly before his death, to the Paris Academy.

Todhunter [33, §336] concluded that "the theory of probability owns more to him than to any other mathematician with the sole exception of Laplace".⁶ However, when listing De Moivre's concrete achievements, he only mentioned his investigations of the duration of play, his theory of recurring series and his "extension of the value of Bernoulli's theorem by the aid of Stirling's Theorem". Considering that this "extension" led De Moivre to the normal law, we should estimate his merits much higher.⁷ His main pertinent writings are

a) *The Doctrine of Chances* [13] greatly expanded from its initial version [11].

b) *Misc. Anal.* [12] with two supplements apparently bound up to the main text at a later date. Pearson [25; 27] ascertained that not all the copies of the book have the first supplement, and only a few of these have the second one dated 1733 [10], reprinted by Archibald [4]. Owing to its importance, I list it separately:

c) *Approximatio ...* [14].⁸ De Moivre included its English translation in the second and the third edition of his *Doctrine* and introduced it [13, 1756, p. 242] in the following way:

*I shall here translate a Paper of mine which was printed November 12, 1733, and communicated to some Friends, but never yet made public ...*⁹

Pearson [25] stressed that the *Approximatio* had contained the normal distribution, but he hardly knew that this fact was already noticed by Eggenberger [16] and that Czuber [9] and Haussner [6, No. 108, pp. 158 – 159] mentioned the latter's discovery.

I shall first dwell on **De Moivre's theological views** which he expounded more fully in the second English version of his *Approximatio* [13, p. 253]. There, illustrating his thoughts by a game of dice, he maintained that

The probability of an assigned Chance, that is, of some particular disposition of the Dice, becomes as proper a subject of Investigation as any other quantity or ratio can be. But Chance, in atheistic writings or discourse, is a sound utterly insignificant: It imports no determination to any mode of Existence [...] nor can any Proposition concerning it be either affirmed or denied ...

Arbuthnot clearly formulated the problem of a determinate versus random origin of the observed predominance of male births over those of females, and concluded that that fact was occasioned by Divine design, – but why was it impossible to formulate similar problems "in atheistic writings" with the same clarity?

Derham (1657 – 1735), another Fellow of the Royal Society and a clergyman, pronounced a similar and vigorous statement [15, p. 313] likely known to De Moivre¹⁰:

Should we be so besotted by the devil, and blinded by our lusts, as to attribute one of the best contrived pieces of workmanship [man] to blind chance, or unguided matter and motion, or any such sottish, wretched, atheistical stuff?

And already in 1738 De Moivre [14, p. 251] quite definitively wrote:

Altho' Chance produces Irregularities, still the Odds will be infinitely great, that in the process of Time, those Irregularities will bear no proportion to the recurrency of that Order which naturally results from ORIGINAL DESIGN.

Pearson [27, p. 552] remarked in this connection:

De Moivre expanded the Newtonian theology and directed statistics into the new channel down which it flowed for nearly a century. The causes which led De Moivre to his Approximatio or Bayes to his theorem were more theological and sociological than purely mathematical, and until one recognizes that the post-Newtonian English mathematicians were more influenced by Newton's theology than by his mathematics, the history of science in the 18th century – in particular, that of the scientists who were members of the Royal Society – must remain obscure.

Above, I indicated that Newton had respected De Moivre. Here now is a phrase from the Dedication of the first edition of the *Doctrine* to Newton, as reprinted in its third edition [13, 1756, p. 329]: He, De Moivre, will think himself “very happy” if he could, by his *Doctrine*,

Excite in others a desire [...] of learning from yours [Newton’s] philosophy how to collect, by a just Calculation, the Evidences of exquisite Wisdom and Design, which appear in the Phenomena of Nature throughout the Universe ...

In other words, how to choose between Design and Chance. The aim of the theory of probability was thus formulated.

I conclude here by quoting De Moivre’s answer to a man “who, apparently intending to pay him a compliment, remarked that mathematicians had no religion”. He replied: “I will prove that I am a Christian by forgiving you the insult you are offering” (Walker [34, p. 363], repeating an earlier author [2, p. 184]).

Book 5 of the *Misc. Anal.* is called “De binomio $a + b$ ad Potestatem permagnam evecto”. There, De Moivre had provided a long passage from Jakob Bernoulli, described Niklaus Bernoulli’s letter to Montmort (above) and solved two problems on expected winnings in games of chance and a few important algebraic problems which he applied later on in his *Approximatio*. While commenting on Niklaus, he [12, p. 98] correctly remarked that Niklaus

Did not investigate the probability that the probability of the number of occurrences or non- occurrences of an event was contained within definite boundaries.

The two abovementioned problems are also in the *Doctrine* (1738 and 1756; NNo. 72 and 73 in the latter). The second, but not the third edition has a Table of Contents where De Moivre characterized them as tending to establish the degree of consent that should be attached to experiments whereas the *Approximatio* was modestly described as the same subject continued further.¹¹

The Corollary to Problem 73 actually states that the statistical probability of an event will be close to its theoretical counterpart, and the closer the more observations are made. Still, De Moivre continued,

Considering the great Power of Chance, Events might at long run fall out in a different proportion from the real Bent ...

and he was therefore adducing a translation of the *Approximatio* so as to solve “the hardest problem that can be proposed on the Subject of Chance ...”

Like the *Ars Conjectandi*, the *Approximatio* consists of an algebraic and a stochastic part. In the first supplement to the *Misc. Anal.* De Moivre derived, independently from Stirling and at the same time as the latter, an approximation for $n!$. It involved a constant B such that

$$\ln B = 1 - 1/12 + 1/360 - \dots, \quad 1/12 = B_1/1 \cdot 2, \quad -1/360 = B_2/3 \cdot 4, \dots$$

and B_1, B_2, \dots were the Bernoulli numbers. It was Stirling, however, who informed De Moivre that $B = \sqrt{2\pi}$. Nevertheless, commentators [22; 25] indicate that the Stirling formula should be called after both him and De Moivre. This is all the more reasonable since De Moivre, in the same supplement (and also in the *Doctrine* [13, 1756, p. 333]), published a table of $\ln n!$ with mantissas given to 14 digits for $n = 10$ (10) 900. When comparing it with a modern table [28, Anhang, Tafel 6, 18-Stellige $\ln n!$] I found out that it is correct up to 11 – 12 digits with a single misprint in the fifth digit of the mantissa of $\ln 380!$.

De Moivre distinctly recognized the importance of \sqrt{n} as a measure of precision and called it “the Modulus by which we are to regulate our estimation” [14, p. 248]. True, its first appearance was caused by an “algebraic” fact: the value $l = \sqrt{n}/2$ was the boundary between two methods of integrating the exponential function.

De Moivre (p. 247) also maintained that

The number n should not be immensely great; for supposing it not to reach beyond the 900th power, nay not even beyond the 100th, the Rule here given will be tolerably accurate, which I had confirmed by Trials.¹²

He did not elaborate, but the mere fact of checking the precision is remarkable. Walker [34, p. 355] maintained that De Moivre had

Made few practical applications of his discoveries, and he never resorted to physical

experimentation or to induction of empirical law from observed phenomena. He did not weigh and measure and count to secure objective verification of his discoveries in the theory of probability. ... he does not set up experimental checks ... he would doubtless have exhibited extreme astonishment at the suggestion that his Approximatio, which he thought merely an exercise in pure mathematics, contained a law which would ...

De Moivre's "Trials" (above) hardly belonged to natural sciences, but they, as well as his Table of $\ln n!$, and his calculations of annuities on lives testify that at least in mathematics he carried his work up to practically useful results. Consider also his Dedication of the *Doctrine* to Newton (above), and Walker's statement will be dismissed. As to the *Approximatio*, it was written to strengthen statistical deductions (see the description of the Corollary to Problem 73 above).

Bearing in mind that De Moivre, in concluding his *Approximatio*, noted that his deductions might be [readily] extended onto the general case of $(a + b)^n$, his finding should be interpreted as proving the local and the integral theorems on the convergence of the binomial distribution to the normal law, but of course he did not know anything about the uniform convergence that takes place there.

Independently from De Moivre, Daniel Bernoulli (1770 – 1771) derived the De Moivre – Laplace limit theorems, and I hope to discuss this topic elsewhere.¹³

Notes

1. The appearance of serious studies [18; 19] as well as of a reprint of Montmort [24] made it possible to leave out some mathematical transformations originally included here. This paper intersects my previous article [29].

2. Aleksandr Vasilievich Vasiliev (1853 – 1929), Professor at Kazan University, a mathematician and historian of mathematics, played an active part in popularizing Lobachevsky's ideas. In 1885 he published a course in probability (Kazan, a mimeographed edition). However, in this branch of mathematics he is primarily remembered as Markov's correspondent. It was in a letter to Vasiliev that Markov expounded his ideas on justifying the method of least squares.

3. John Arbuthnot (1667 – 1735) [1; 5; 32], a physician and mathematician, Fellow of the Royal Society (1704), was well acquainted with Jonathan Swift and Alexander Pope and published a few pamphlets directed against the Whigs. The name of one of his heroes, John Bull (from his *History of John Bull*) is still with us. Arbuthnot also wrote *An essay on the Usefulness of Mathematical Learning* (1700; reprinted in [1]) and *Tables of the Grecian, Roman and Jewish Measures, Weights and Coins* (1707) and he was the main translator of Huygens' *Of the Law of Chance* (1692).

For my subject, however, the most interesting of his writings is his note [3] where he, for the first time ever [17], tested a statistical hypothesis. At present, such a procedure is understood as a test of the realization of some law of distribution, or of some value of a parameter of some definite law. Arbuthnot, however, attempted to test whether a phenomenon under his study (the prevalence of male births over those of females) was random or determinate, and he decided in favor of the latter, – of Divine design. A number of later scholars (Daniel Bernoulli, Michell) including Laplace tested hypotheses in the same sense as Arbuthnot did. Newton apparently respected Arbuthnot. Thus, he discussed Flamsteed's observations with him (letter to Flamsteed of 1711 [7, vol. 2, p. 489]).

4. Later note: I indicated that the factor $\sqrt{2/\pi}$ is lacking in formula (2). Hald [19, p. 17] did not repeat this remark.

5. Maty's memoir proved unavailable. However, an article "Sur la vie et sur les écrits de De Moivre" is contained in the *J. Britannique* (La Haye, t. 18, Sept. – Oct. 1755), a periodical edited by him.

6. Later note: This seems to be too strong.

7. The *Misc. Anal.* is not translated into any modern language, and the works of De Moivre are not collected together in any edition. From Lagrange's letter to Laplace of 30.12.1776 [20] it follows that they both thought of translating De Moivre's *Doctrine* into French. De Moivre began his *Approximatio* by stating that only Jakob and Niklaus Bernoulli had preceded him. And he continued:

tho' they have shewn very great skill, and have the praise which is due to their Industry, yet some things were farther required; for what they have done is not so much an Approximation as the determining very wide limits, within which they demonstrated that the Sum of the terms [of the binomial] was contained.

8. Later note: In the *Approximatio* itself [13, 1756, p. 243] De Moivre also stated: "It is a dozen years or more since I had found what follows ..." These years should be reckoned from 1733 (not 1738) since the Latin version of 1733 mentioned "Duodecim jam sunt anni ..." In other words, De Moivre made his outstanding discovery in 1721 or a bit earlier.

9. In a letter of 1714 to Newton Derham [7, vol. 2, p. 520] asked the former to honor his promise of giving “castigations” for the third impression of his *Physico-Theology*.

10. Just as it was in several instances above, I do not describe these problems anymore. However, I refer readers to my later paper [31, p. 236] in connection with the *Spectator* introduced here by De Moivre and with the role of such outsiders.

11. See [30].

12. “Power” likely referred to “binomial to the power of n ”.

13. Since then published in 1970, in *Biometrika*, vol. 57.

References

1. Aitken, G.A. *Life and Works of John Arbuthnot*. Oxford, 1892.
2. Anonymous, Eloge de De Moivre. *Hist. Acad. Roy. Sci. Paris*, 1754 (1759), pp. 175 – 184.
3. Arbuthnot, J. An argument for Divine Providence taken from the constant regularity observed in the births of both sexes (1712). In *Studies in Hist. Stat. and Probability*, vol. 2. Editors, Sir Maurice Kendall, R.L. Plackett. London, 1977, pp. 30 – 34.
4. Archibald, R.C. A rare pamphlet of Moivre and some of his discoveries. *Isis*, vol. 8, 1926, pp. 671 – 684.
5. Beattie, L.M. *John Arbuthnot*. Harvard Studies in English, vol. 16. Cambridge (Mass.), 1935.
6. Bernoulli, J. *Wahrscheinlichkeitsrechnung* (1713, in Latin). Hrsg. R. Haussner. *Ostwalds Klassiker* No. 107 – 108 (1899). Frankfurt/Main, 1999.
7. Brewster, D. *Memoirs of the Life of Newton*, vols 1 – 2. Edinburgh, 1855.
8. Clarke, A.M. Moivre. *Dict. Nat. Biogr.*, vol. 38. London, 1894.
9. Czuber, E. Die Entwicklung der Wahrscheinlichkeitstheorie. *Jahresber. Deutsche Mathematiker – Vereinigung*, Bd. 7, No. 2, 1899. Separate paging.
10. Daw, R.H., Pearson, E.S. De Moivre’s 1733 derivation of the normal curve: a bibliographic note. *Biometrika*, vol. 59, 1972, pp. 677 – 680.
11. De Moivre, A. De mensura sortis, or the measurement of chance (1712, in Latin). *Intern. Stat. Rev.*, vol. 52, 1984, pp. 229 – 262.
12. De Moivre, A. *Miscellanea analytica de seriebus et quadraturis*. London, 1730.
13. De Moivre, A. *Doctrine of Chances*. London, 1718, 1738, 1756. Reprint of last edition: New York, 1967.
14. De Moivre, A. *Approximatio ...* (1733). Engl. transl.: [13, 1738; 1756, pp. 243 – 254].
15. Derham, W. *Physico-Theology*. Preached in 1711 – 1712. London, 1768 (13th edition).
16. Eggenberger, J. Beiträge zur Darstellung des Bernoullischen Theorems etc. *Mitt. Naturforsch. Ges. Bern* NNo. 1305 – 1334, 1893 (1894), pp. 110 – 182. Also publ. separately: Berlin, 1906.
17. Freudenthal, H. 250 years of mathematical statistics. In *Quantitative Methods in Pharmacology*. Editor H. De Jonge. Amsterdam, 1961, pp. xi – xx.
18. Hald, A. *History of Probability and Statistics and Their Applications before 1750*. New York, 1990.
19. Hald, A. *History of Mathematical Statistics from 1750 to 1930*. New York, 1998.
20. Lagrange, J.L. *Oeuvres*, t. 14. Paris, 1892.
21. Laplace, P.S. *Philosophical Essay on Probabilities* (1814, in French; transl. from edition of 1825). New York, 1994.
22. Markov, A.A. *Исчисление вероятностей* (Calculus of probability). Psk, 1900, 1908, 1913. M., 1924. German transl : Leipzig – Berlin, 1912.
23. Maty, M. *Sur la vie de De Moivre*. La Haye, 1760.
24. Montmort, P.R. *Essay d’analyse sur les jeux de hazard* (1713, second edition). New York, 1980.
25. Pearson, K. Historical note on the origin of the normal curve of errors. *Biometrika*, vol. 16, 1924, pp. 402 – 404.
26. Pearson, K. Bernoulli’s theorem. *Ibidem*, vol. 17, 1925, pp. 201 – 210.
27. Pearson, K. De Moivre. *Nature*, vol. 117, 1926, pp. 551 – 552.
28. Peters, J. *Zehnstellige Logarithmentafeln*, Bd. 1. Berlin, 1922.
29. Sheynin, O.B. On the early history of the law of large numbers. *Biometrika*, vol. 55, 1968, pp. 459 – 467.
30. Sheynin, O.B. Daniel Bernoulli on the normal law. *Ibidem*, vol. 57, 1970, pp. 199 – 202.
31. Sheynin, O.B. Early history of the theory of probability. *AHES*, vol. 17, 1977, pp. 201 – 259.
32. Stephen, L. Arbuthnot. *Dict. Nat. Biogr.*, vol. 2. London, 1885.
33. Todhunter, I. *History of the Mathematical Theory of Probability* (1865). New York, 1949 and 1965.
34. Walker, H.M. De Moivre (1934). In [13, 1756, pp. 351 – 368].

9. On the Appearance of the Dirac Delta-Function in a Memoir of Laplace

IMI, vol. 20, 1975, pp. 303 – 308

In one of his memoirs devoted to the theory of errors Laplace [1] assumed the density of observational errors as

$$f(x) = (1/2a) \ln(a/|x|), \quad a > 0, |x| \leq a. \quad (1)$$

Without repeating his suppositions [2, esp. pp. 294 – 298], I note that this density was patently unsuitable for any practical applications (it was infinite at $x = 0$) and that in case of need Laplace would have undoubtedly modified his law.¹ Just as in his previous memoir [5], Laplace searched for the “best” estimator of the “true value” of an astronomical phenomenon by means of repeated observations² choosing the number e determined by the equation

$$\int_{-a}^e f(x - x_1) f(x - x_2) \dots f(x - x_n) dx = \int_e^a f(x - x_1) f(x - x_2) \dots f(x - x_n) dx \quad (2)$$

where x_1, x_2, \dots, x_n were the observations.

Apparently bearing in mind his memoir [5], Laplace maintained that the rule of the arithmetic mean can be deduced from (1) given that a was unbounded. It was there that he effectively introduced the Dirac delta-function, or, rather, one of its interpretations, a sequence of functions (1) at $a \rightarrow \infty$. Laplace did not prove his statement; instead, he (p. 480) went on to a “much more general theorem” for the density

$$y = \varphi(\alpha x) = \varphi(-\alpha x) = q \text{ if } \alpha x = 0; \text{ and } = 0 \text{ if } \alpha x \neq 0, \alpha \rightarrow 0. \quad (3a; 3b)$$

He actually considered a sequence of functions $\varphi(\alpha x)$; and the condition (3a) should more precisely be written down as

$$\varphi(\alpha x) = q(\alpha), \quad \alpha = \{ \alpha_1, \alpha_2, \dots, \alpha_n, \dots \} \rightarrow 0. \quad (4)$$

Suppose now that $\alpha x = t$, then the relations (3) will become

$$\varphi(t) = q, \quad t = 0, |x| < +\infty; \text{ and } = 0, t \neq 0, |x| = +\infty, \quad (5a; 5b)$$

and, in accord with stochastic demands,

$$\int_{-\infty}^{\infty} \varphi(t) dt = C \quad (= 1, \text{ say}) \quad (6)$$

so that in (5a) $q(\alpha) \rightarrow \infty$. The same demands lead to $q(\alpha) \rightarrow 0$ in (4).

This interpretation of relations (4) – (6), which are lacking in Laplace’s memoir, allows me to state that he introduced the Dirac function. Laplace could have understood the function (3) or (5) as a uniform law with an unbounded interval of possible observational errors. In a sense, such a distribution (as, also, the normal law which possesses the same property) would have been more suitable than the logarithmic function (1) with a finite interval $[-a; a]$ of permissible errors.

I repeat that Laplace considered (3) as a generalization of (1) which can be understood thus: the derivative of (1) is $f'(x) = -1/(2ax)$. For large values of a it decreases very slowly beginning with $x > c^2$. It is possible to assume here that c^2 is a number decreasing with an increasing a , and to discuss only the case of $x > 0$. It can even be thought that, at $x > c^2$, the function (1) is almost constant. Finally, at $x > c^2$ it decreases with an increasing a . On the other hand, the function (3) is constant excepting the case $x = \infty$, and, on the strength of (4), it decreases with an increasing interval of permissible errors (e.g., with a decreasing α).

Laplace proves a theorem concerning (3): e in an equation of the type of (2) written down for the function $\varphi(\alpha x)$ is the arithmetic mean of the observations x_i , $i = 1, 2, \dots, n$. Here is how he reasoned. For any x_i

$$\varphi[\alpha(x - x_i)] = \varphi[\alpha(x - e)] + \alpha \varphi'_e[\alpha(x - e)](x - x_i)$$

so that

$$y = \varphi[\alpha(x - x_1)] \varphi[\alpha(x - x_2)] \dots \varphi[\alpha(x - x_n)] = \varphi^n[\alpha(x - e)] + \alpha \varphi'_e[\alpha(x - e)] \Sigma(e - x_i) = \varphi^n[\alpha(x - e)] \quad (7)$$

and, obviously,

$$\int_{-\infty}^e \varphi^n[\alpha(x - e)] dx = \int_e^{\infty} \varphi^n[\alpha(x - e)] dx.$$

It was not in Laplace's manner to explain exactly how did he understand the derivative of (3)! Furthermore, his notation was not symmetric; he denoted the observational errors by x , $p - x$, $p' - x$, ... where p , p' , ... were the distances of the observations from each other.³ Second, he identified the variable with some constant (which he then varied); third, the expansion in a Taylor series which he made use of was unnecessary difficult. Thus, he wrote

$$\varphi(\alpha x + \alpha p) = \varphi(\alpha x) + \alpha p d\varphi(\alpha x) / d(\alpha x)$$

so that

$$\varphi(\alpha x) \varphi(\alpha x + \alpha p) \varphi(\alpha x + \alpha p') \dots = \varphi^n(\alpha x) + \alpha \varphi^{n-1}(\alpha x) [d\varphi(\alpha x) / d(\alpha x)] (p + p' + \dots).$$

Unlike formula (7) for which the second term in the right side vanished owing to the definition of the arithmetic mean, here the integrals of both terms had to be later taken account of. Fourth, as far as technique was concerned, Laplace's calculations were difficult to follow because of his permanent inattention to detail, but I shall not dwell on this point.

After denoting

$$\int_0^{\infty} \varphi^n(\alpha x) dx = A, \quad (8)$$

Laplace actually assumed that $A > 0$ so that (8) was his analog of formula (6).

Introducing interpretation

$$\varphi(t) = \lim(\lambda / \sqrt{\pi}) \exp(-\lambda^2 t^2) \text{ as } \lambda \rightarrow \infty$$

for the function (5) it is possible to prove Laplace's theorem more rigorously. The product

$$\varphi(t - t_1) \varphi(t - t_2) \dots \varphi(t - t_n) \quad (9)$$

with t_i being the observations will be rewritten as

$$\lim \{ (\lambda / \sqrt{\pi})^n \exp[-\lambda^2 \{(t - t_1)^2 + (t - t_2)^2 + \dots + (t - t_n)^2\}] \} \text{ as } \lambda \rightarrow \infty.$$

The exponent is here

$$-n \lambda^2 [t^2 - (2t/n)(t_1 + t_2 + \dots + t_n) + (t_1^2 + t_2^2 + \dots + t_n^2)/n] = -n \lambda^2 \{ [t - (t_1 + t_2 + \dots + t_n)/n]^2 + (t_1^2 + t_2^2 + \dots + t_n^2)/n -$$

$$(t_1 + t_2 + \dots + t_n)^2 / n^2 \},$$

$$(t_1^2 + t_2^2 + \dots + t_n^2) / n - (t_1 + t_2 + \dots + t_n)^2 / n^2 \equiv \beta^2 \geq 0.$$

The equality ($\beta^2 = 0$) is only attained if $t_1 = t_2 = \dots = t_n$ in which case the theorem is evident so that I assume that $\beta^2 > 0$. The product (9) can now be written down as

$$\lim_{\lambda \rightarrow \infty} \{ (\lambda / \sqrt{\pi})^n \exp(-\lambda^2 \beta^2 n) \exp[-\lambda^2 n (t - t_0)^2] \} =$$

$$\varphi^n(\beta) \varphi^n(t - t_0) \lim_{\lambda \rightarrow \infty} (\sqrt{\pi} / \lambda)^n$$

where t_0 is the arithmetic mean of t_i . Integrating both sides of (9) I have

$$\int_{-\infty}^{\infty} \varphi(t - t_1) \varphi(t - t_2) \dots \varphi(t - t_n) dt = \varphi^n(\beta) \varphi^n(t - t_0) \lim_{\lambda \rightarrow \infty} (\sqrt{\pi} / \lambda)^n$$

but, obviously,

$$\int_{-\infty}^{t_0} \varphi^n(t - t_0) dt = \int_{t_0}^{\infty} \varphi^n(t - t_0) dt$$

so that

$$\int_{-\infty}^{t_0} \varphi(t - t_1) \varphi(t - t_2) \dots \varphi(t - t_n) dt = \int_{t_0}^{\infty} \varphi(t - t_1) \varphi(t - t_2) \dots \varphi(t - t_n) dt.$$

In his later writings Laplace abandoned condition (2) and did not return to the delta-function. For this reason its appearance is connected with Cauchy and Poisson [6, §4 from Chapt. 5], a source indicated to me by F.A. Medvedev

Notes

1. I have already stated this [2] and now I can refer to Laplace himself [3, p. xi; 4, p. 116]. Thus, in the second source he argued that hypotheses should be regarded “only as a means of connecting the phenomena together to discover their laws”; no “reality” ought to be attributed to them; and they should “continually” be corrected by new observations.

2. I restrict my description by what is needed for understanding my subject, and I am following the terminology of the classical theory of errors whose cofounder was Laplace. Statisticians would have discussed the determination of the location parameter e of the density $f(x - e)$.

3. By supposing that the “true value” of the constant sought is at the origin of the coordinates, I have identified observation x_i with error x_i .

Later note. Integrals in equality (1) have no meaning in the language of generalized functions. Still, even an interpretation on a physical level, as attempted above, seems interesting.

References

1. Laplace, P.S. Sur les probabilités (1781). *Oeuvr. Compl.*, t. 9. Paris, 1893, pp. 383 – 485.
2. Sheynin, O.B. Finite random sums. AHES, vol. 9, 1973, pp. 275 – 305.
3. Laplace, P.S. *Traité de Mécanique céleste*, t. 3 (1802 or 1803). *Oeuvr. Compl.*, t. 3. Paris, 1878.

4. Laplace, P.S. *Philosophical Essay on Probabilities* (1814, in French. Transl. from edition of 1825). New York, 1995.
5. Laplace, P.S. Sur la probabilité des causes par les événements (1774). *Oeuvr. Compl.*, t. 8. Paris, 1891, pp. 27 – 65.
6. Pol, B. van der, Bremmer, H. *Operational Calculus*, second edition. Cambridge, 1955.

10. History of the Theory of Probability

Based on Theory of probability before Chebyshev. IMI, vol. 25, 1978, pp. 284 – 306,
and *History of the Theory of Probability to the Beginning of the 20th Century*. Berlin, 2004

This essay is based on my previous writings, about which see **General sources** below. It is usual to separate the development of probability theory into the following periods: [Pascal & Fermat – Jakob Bernoulli]; [the 18th century – Chebyshev]; and [Chebyshev, Markov, Liapunov]. What followed, at least until the mid-20th century (but what is not described here), can be characterized by the introduction of the ideas and methods of the set theory and the theory of functions of a real variable, and of course by the axiomatization of the theory and the birth of mathematical statistics. I stress that there existed a prehistory of probability and that Jakob Bernoulli, De Moivre and Bayes, taken together, created the first version of the theory. The modern stage began with Chebyshev and is characterized by systematic application of the concept of random variable and expectation, and by recognition that the goal of the theory of probability is to determine the probabilities of some events given the probabilities of other events rather than the discovery of the laws of nature (Laplace).

1. The Prehistory

Ancient scholars recognized logical or subjective probability and randomness. For Aristotle, randomness implied lack of aim (a sudden meeting of two people knowing each other) or deviation from law (birth of monsters or even females or girls). Small causes in the first case could have prevented the meeting so that this explanation is close to Poincaré's idea (end of §4.5). The second example (repeated by Thomas Aquinas) is unconvincing insofar as females are concerned since Aristotle himself thought that random events occurred seldom. By probable Aristotle meant something occurring in most cases.

Randomness in the sense of Aristotle's first example is mentioned in the Old Testament, e.g., 1 Kings 22:34, and the Talmud attempted to distinguish between randomness and causality, for example, between "usual" deaths and deaths occasioned by an outbreak of plague. The same attempt is evident in the ancient Indian *Laws of Manu*: the witness in a law-suit, who soon suffered a misfortune, was thought to be divinely punished for perjury. Hippocrates and Galen reasoned in the spirit of qualitative correlation. The latter knew that small causes could essentially influence weak people (unstable equilibrium!). As astrologers, Ptolemy and Al-Biruni believed that the influence of heaven on man was a [correlative] tendency rather than a fatal drive and Tycho Brahe and Kepler were of the same opinion.

Maimonides mentioned expectation on a layman's level. He noted that there existed a more or less fixed expected value of a future possible gain ensuing from a marriage settlement. Life insurance in the form of annuities existed in Europe from the 13th century (although it was prohibited for about a century until 1423) but it had been connected with expectancies of life only in a most generalized way. A similar situation existed in marine insurance.

When selecting a point estimate for the constant sought, ancient astronomers were reasonably choosing almost any number within appropriate bounds. This testifies once again that science in those times was qualitative, but it also agrees with modern notions: such an attitude is justified when dealing with observations whose errors possess a "bad" distribution. Only Kepler (indirectly) stated that the arithmetic mean had become "the letter of the law". This change had to be brought about by the increase in the precision of observations and was likely made easier by the time-honored idea that mean behaviour and moderation possessed optimal properties.

Galileo formulated the properties of "usual" random errors and Kepler likely knew them as well. When deriving his first law, he adjusted observations by corrupting them by arbitrary corrections, which should have been selected according to these properties. Kepler's system of the world had to leave room for randomness: he explained the eccentricities of the planetary orbits by random corruptions of the predetermined circular paths (Aristotelian deviation from law!).

2.1. Probability and Social Life

Games of chance provided natural problems whose solution led to the development of the theory of probability. In a similar way, jurisprudence propagated stochastic ideas. Leibniz thought of including the nascent theory within a general system of logic (within a statistical decision theory!) and therefore urged to study games of chance. Descartes re-introduced the ancient concept of moral certainty and connected it with ethics, and Arnauld & Nicole (1662) and then Jakob Bernoulli extended its use onto human activities and especially onto jurisprudence. Calculations of the cost of annuities compelled De Witt (1671) to introduce a certain stochastic law of mortality (without justifying it) and to accepting expectation as his criterion. He thus

opened up a new field for probability (already studied by Huygens in his then yet unpublished correspondence of 1669).

Population statistics originated within political arithmetic (Petty, Graunt). Most influential became the latter's mortality table (1662). He established that both sexes were approximately equally numerous and that the sex ratio at birth was 14 boys : 13 girls. He also arrived at important conclusions concerning the causes of death. Halley, in 1694, compiled the second mortality table that served as a point of departure for De Moivre's studies of annuities on lives.

2.2. Pascal and Fermat

In 1654 they introduced the concept of expected winning in a game of chance and applied it as a criterion for solving the already then venerable problem of points. They used combinatorial methods and tacitly applied the addition and multiplication theorems (for chances) whereas one of Pascal's arguments might be interpreted in the language of conditional probabilities. Pascal also used partial difference equations. If a tabular representation of a function is recognized on a par with its analytic expression, then Pascal's arithmetic triangle might be identified with the application of a generating function for the symmetric binomial distribution.

2.3. Huygens

In 1657, following after those French scholars, Huygens independently introduced expectation and effectively applied conditional chances for solving similar problems. He did not use combinatorial analysis and, when investigating games in which the expectation of winning varied from one round to another, he had to make involved calculations. In his correspondence of 1669 (published 1895), he considered problems connected with mortality and life insurance, introduced the notions of mean and probable durations of life, calculated the expectations of order statistics for a discrete empirical distribution, methodologically applied a graph of the function $[1 - F(x)]$ with $F(x)$ being an unspecified distribution function and $0 \leq x \leq 100$.

2.4. Newton

In a manuscript of 1664 – 1666 Newton generalized the concept of chances onto the continuous case and remarked on the possibility of using statistically observed chances. He also applied simple stochastic considerations for checking the duration of the rule of a dynasty or of an individual in ancient kingdoms. Most important was his general philosophical views. De Moivre devoted the first edition of his *Doctrine of chances* (1718) to Newton and stated that his goal was to discover rules for separating chance and Design in nature.

2.5. Arbuthnot

In 1712 Arbuthnot collected the data on baptisms in London during 82 previous years (1629 – 1710) and noted that more boys (m) had invariably been born than girls (f). He declared that the inequality $m > f$ was occasioned by Divine design since its random occurrence had "value of expectation" 2^{-82} . Strictly speaking, his argument was unconvincing (any sequence of ten, say, throws of a coin are equally possible) although practically correct. He had not, however, thought that the births of both sexes obeyed a (not yet studied) binomial distribution but his finding served as a starting point for future important studies.

3. 1. Jakob Bernoulli

The first three parts of Jakob Bernoulli's posthumous *Ars Conjectandi* (1713) contained a reprint of Huygens' treatise of 1657 with essential comments and a solution of interesting problems such as investigation of random sums for the discrete uniform and the binomial distributions; a similar study of the sum of a random number of terms for a discrete distribution; the derivation of the distribution of the first order statistic for the discrete uniform distribution; calculation of the probabilities for sampling without replacement. His analytic methods included combinatorial analysis and the calculation of expected winnings in each round of a finite or infinite game and their subsequent summing.

The last part of the *Ars* contained the "classical" definition of probability of an event, a formulation of the aims of the *ars conjectandi* (derivation of probabilities of events for optimal decision-making in economics and politics), elements of stochastic logic (involving non-additional probabilities whose source is the medieval doctrine of probabilism declaring that the opinion of each theologian was probable) and his law of large numbers. A fragmentary proof of that law was already contained in Bernoulli's diary for 1684 – 1690. In studying the law of large numbers, Bernoulli proved that the statistical probability tended to its theoretical counterpart and estimated the rapidity of that convergence. His law ensured the equivalence of the statistical and theoretical probabilities and established, in the context of the stochastic branch of the theory of knowledge, the first accordance between the deductive and the inductive methods.

3.2. Niklaus Bernoulli

In 1709 Bernoulli's nephew Niklaus published a dissertation on the application of the art of conjecturing to jurisprudence. Drawing on Graunt's mortality table, he calculated the mean duration of life for persons of different ages, and, assuming a continuous uniform law of mortality, he determined the expected life of a last survivor of a group of men. While being important and fostering the public's interest to probability because of these and other findings, Niklaus' dissertation contained borrowed passages from his uncle's yet unpublished book and even his diary.

In Niklaus' correspondence with **Montmort** of 1710 – 1713 published in the latter's book in 1713, he studied the sex ratio at birth and indirectly arrived at the normal distribution as the limiting law for the binomial. He also invented the celebrated Petersburg game involving a random variable possessing an infinite expectation. The game was paradoxical since no-one would have paid any considerable sum in exchange for that expectation. Many scholars have been studying it since then and it became the reason for the introduction of the once famous moral expectation (§4.1). Montmort himself studied many games of chance by applying combinatorial analysis, recurrent formulas and infinite series as well as the formula of inclusion and exclusion. Of special interest was the strategic game *le her* whose investigation only became possible on the basis of the theory of games, but Niklaus at least noted that in that game the gamblers should keep to mixed strategies.

3.3. De Moivre

De Moivre's main finding was the proof, in 1733, of the "De Moivre – Laplace limit theorem" this being the first version of the central limit theorem. He greatly influenced Laplace but became forgotten to the turn of the 19th century. De Moivre was also the most influential student of life insurance in his time. In accordance with Halley's table of mortality, he assumed the uniform law of mortality for all ages beginning with 12 years and a maximal duration of life equal to 86 years. De Moivre as well as Montmort and Niklaus Bernoulli continued to study games of chance; they applied generating functions for solving the problem of points and paid special attention to "ruining" problems first introduced by Pascal. Continuing this line of research, De Moivre, in 1730, created his theory of recurring sequences.

3.4. Bayes

The so-called Bayes theorem for $P(A/B)$ is absent in Bayes' memoir of 1764 – 1765. It contains the formula for the posterior distribution of an event given that it occurred p times and failed to occur q times in $p + q = n$ trials and that its prior distribution was continuous and uniform. It is methodologically important that Bayes' inverse probability is tantamount to conditional probability given that the stipulated condition was fulfilled. Beginning with Fisher, the Bayes formula had been denied for about three decades, mostly because it was based on hardly known prior probabilities (irrespective of what Bayes actually thought).

Bayes studied the case of a large finite n whereas Timmerding, the Editor of the German translation (of 1908) of his memoir, additionally examined the transition to $n \rightarrow \infty$. It occurred that the Bayes formula became a limit theorem improving on the De Moivre proposition. Without introducing any measure of scattering, Bayes actually proved that De Moivre (and Jakob Bernoulli), who had thought that their theorems described both the direct and the inverse cases on a par, were mistaken. The Bayes formula thus accomplished the construction of the first version of the theory of probability whose previous most important propositions were due to Jakob Bernoulli and De Moivre.

4.1. The 18th Century before Laplace

Together with De Moivre, **Daniel Bernoulli** was the main predecessor of Laplace. In order to remove the infinite expectation involved in the Petersburg game, he suggested, in 1738, that the *moral expectation* of a gambler be introduced as $\sum p_i f(x_i) / \sum p_i$ where p_i were the probabilities of gaining (or losing) x_i and $f(x) = C \ln x/a$ with a being the initial capital of the gambler. Bernoulli applied his notion to prove that even a just game with zero expected losses for each gambler was disadvantageous to each and to suggest that moral expectation should be taken into account in hazardous commercial enterprises. His innovation became so popular in the theoretical sense, that Laplace re-named the usual expectation calling it *mathematical*. In 1770, Bernoulli considered an urn problem later examined by Laplace. In essence, it coincided with the celebrated Ehrenfests' model of 1907 usually considered as the beginning of the history of stochastic processes.

Bernoulli stochastically investigated most important problems of population statistics such as sex ratio at birth (independently deriving the local De Moivre – Laplace theorem), mortality (especially mortality from smallpox with inoculation as a preventive measure proving that that dangerous treatment increased the mean duration of life), and duration of marriages. **Lambert** studied the number of children in families, and **Euler** was the joint author of a chapter in **Süssmilch's** book on the rate of increase and the period of doubling of the population. Süssmilch was rather careless in treating his data but he at least paved the way for Quetelet. Lambert followed Leibniz in attempting to create a doctrine of probability as a component of a general teach-

ing of logic. He connected randomness with disorder and heuristically approached the modern notion of normal number.

Several other scholars ought to be mentioned. **Dalembert** formulated patently wrong statements but he also put forward reasonable considerations; taken together, they meant that the theory of probability should be applied cautiously. He also objected to Bernoulli's conclusions concerning inoculation. Thus, he noted that inoculation of children involved moral issues and that an increase in the mean duration of life did not remove the fear of immediate death because of that measure. In 1777 **Buffon** definitively introduced geometric probabilities so as to "put geometry in possession of its rights". Many scholars effectively used them even earlier when introducing densities. In 1767 Michell attempted to determine the probability that the distance between two stars from among all of them uniformly scattered across the sky was not larger than 1° and his problem became classical. **Condorcet** applied the theory of probability to jurisprudence in the tacitly assumed case of independent judgements reached by jurors or judges. His writings were extremely obscure but Laplace and Poisson continued his work.

Simpson, Lambert, Daniel Bernoulli and Euler contributed to the treatment of *direct* observations. In 1756 Simpson, effectively applying generating functions, proved that for the uniform and the triangular discrete distributions the arithmetic mean was stochastically preferable to a separate observation and in 1757 he extended his finding onto the continuous triangular distribution. He, and Lambert, in 1760 – 1765, taken together, originated the theory of errors and that term itself is due to the latter. Lambert was the first to formulate the principle (not yet method) of maximum likelihood for estimating the location parameter of unimodal distributions. In 1778 Bernoulli independently introduced maximum likelihood. His considerations resulted in recommending posterior weighting of observations of equal precision depending on their position in the variational series. For observational errors possessing (as he assumed) an arc of a parabola as density these weights increased to the tails of the distribution, a fact that Bernoulli did not explicitly state. In 1780 he studied the errors of pendulum observations and introduced the normal distribution as the limiting law into the theory of errors and formally distinguished between systematic and random errors. In a companion paper to Bernoulli's memoir of 1778 Euler misunderstood the latter's weighting of observations; for his own part, he proposed a principle heuristically reminiscent of the Gauss condition of maximum weight.

Several methods for adjusting *indirect* observations (for solving n linear equations in m unknowns, $m < n$, under additional conditions imposed on the residual free terms, call them $v_i, i = 1, 2, \dots, n$) were being applied in the 18th century. The **Boscovich** method of 1770 involved additional conditions

$$v_1 + v_2 + \dots + v_n = 0, |v_1| + |v_2| + \dots + |v_n| = \min.$$

In 1809 Gauss remarked that the second condition taken by itself demanded that exactly m residuals should be zeros which meant that he knew an important theorem in linear programming. The first condition alone corresponded to another method of adjustment.

4.2. Laplace

As an astronomer, he was unable to ignore the theory of errors. In his first memoirs devoted to that subject he as though sized it up and did not achieve general results because his formulas were too involved. Laplace then turned to the case of a large number of observations. Applying characteristic functions of random variables and the inversion formula for lattice distributions, he non-rigorously proved the central limit theorem for sums, linear functions, sums of absolute values and of sums of squares of identically distributed random variables with a finite domain of possible values. This approach enabled him to assume normality in the theory of errors. Using his criterion, minimal absolute expectation, Laplace thus arrived at the method of least squares (already known to Gauss). He also compared the median and the arithmetic mean (this time, with variance as his test) and, although without achieving concrete results, introduced the Dirac delta-function.

Laplace's theory of probability included known issues which he essentially developed (the proof of the "De Moivre – Laplace" theorem by the Maclaurin – Euler summation formula; investigation of finite random sums; determination of posterior distributions; examination of an urn problem tantamount to the Ehrenfests' model) and new material (study of a pattern of trials connected into a Markov chain and of trials now called after Poisson; an effective construction of realizations of random functions and their expectations; introduction of a partial differential equation, of the "Dirichlet" formula, discontinuity factors and the Chebyshev – Hermite polynomials into probability). Sometimes he issued from the principle of insufficient reason but stressed that conclusions should be checked by new observations.

Laplace applied the theory of probability to a wide range of problems in astronomy and population statistics and to jurisprudence and "moral" issues. His probability was an applied discipline, its level of abstraction was not high. He did not define a random variable even on a heuristic level, and was thus unable to consider densities or characteristic functions as mathematical objects. After Laplace the theory had to be created anew.

4.3. Poisson

In 1825 he formally introduced the notions of random variable and integral distribution function. In the field of limit theorems he examined the case of a low probability of the occurrence (or non-occurrence) of an event in Bernoulli trials and definitively introduced the “Poisson” trials. He non-rigorously extended the central limit theorem onto sums of non-identically distributed random variables having possible values on a restricted interval. Poisson is especially remembered for his law of large numbers that he proved for the “Poisson” pattern by means of the central limit theorem. An unclear exposition, severe criticism (Bienaymé, Bertrand) and the failure of other scientists to understand the importance of that law led to Poisson’s protracted oblivion.

Poisson invariably determined the statistical significance of discrepancies between empirical magnitudes and thus anticipated the theory of stability of statistical series (Lexis) as well as the intentions of the Continental direction of statistics. By means of the central limit theorem he examined the admissibility of using sample parameters of densities. With respect to the domain of problems studied and the applied analytic methods Poisson continued the work of Laplace. In particular, he devoted much attention to applications of probability to jurisprudence by investigating the stability of the rate of conviction and the probability of judicial errors. In spite of adverse criticisms such applications are useful for perceiving the general picture of administration of justice (in the assumed ideal case of independent decisions made by jurors).

4.4. Gauss

In 1809 Gauss justified the principle of least squares which he had applied since 1794 or 1795. He proved that, among unimodal, symmetric and differentiable distributions of the observational errors there existed only one (the normal) law for which the maximum likelihood estimator of the location parameter coincided with their arithmetic mean. Once more making use of maximum likelihood Gauss came to the principle of least squares.

This elegant reasoning led to the uniqueness of the law of error; and Gauss also reasonably thought that the principle of maximum likelihood was inferior to a method based on a minimal integral measure of error. Assuming variance as such a characteristic, Gauss, in 1823, proved that among linear estimators the least variance was provided by the method (this time, method) of least squares.

The Gaussian theory of errors got rid of the Laplacean assumption of a large number of observations, and the replacement of the latter’s criterion (minimal absolute expectation) by the variance proved successful and fruitful with respect to calculation. Gauss became the creator of the theory of errors which had, until the 1920s, remained the main field of stochastic applications whereas Laplace’s role, practically speaking, was not essential.

Legendre, in 1805, was the first to introduce the principle of least squares and to substantiate it by qualitative considerations. Adrain arrived at the same principle by roughly the same way, and at the same time (in 1809) as Gauss did, but his derivation was not rigorous at all.

4.5. From Gauss to Chebyshev

Among the scientists of this period **Cauchy**, **Bienaymé**, **Cournot**, **Quetelet**, and **Buniakovsky** ought to be mentioned as well as **Bertrand** and **Poincaré**, whose work, although appearing later, in essence also belonged to the pre-Chebyshev stage. **Cauchy** studied the mathematical treatment of observations, especially in connection with his debate with **Bienaymé**, in 1853, on the application of the method of least squares to interpolation of functions. He investigated several approaches to treating observations, proved a theorem in linear programming, introduced stable distributions (although the Cauchy distribution was due to Poisson), used discontinuity factors for studying functions of observational errors. Referring to Poisson and Fourier, he widely applied the Fourier cosine transforms. Also in 1853 he proved the central limit theorem for linear functions of identically distributed errors having possible values on a finite interval and estimated the errors caused by assumptions made. Still, his proof was not sufficiently rigorous.

In 1853 **Bienaymé** proved the celebrated “Bienaymé – Chebyshev” inequality but had not paid due attention to it. In 1852 he introduced cumulants and the multivariate Gram – Charlier series and, in 1845, formulated the properties of criticality of a branching process.

Cournot proposed a common definition of probability for the classical discrete and the continuous cases as the ratio of the “extent” (now we would say “measure”) of the favorable chances to the complete extent of all the chances. He attempted to estimate the dependence between the verdicts reached by jurors or judges and in general paid much attention to statistics. Chuprov called him “the real founder of the modern philosophy of statistics”.

Quetelet directed the attention of the general public to statistics. He introduced the notion of Average man (which dates back to Buffon), introduced mean inclinations to criminality and marriage, declared that the rate of crime was constant, and exerted all his great influence to standardize population statistics on a worldwide

scale. As a popularizer, he made serious mistakes; he did not stress that his inclinations did not concern individuals, that the alleged constancy of crime was only possible under constant social conditions, etc. After his death statisticians denounced his findings and, for good measure, rejected probability as a whole.

Buniakovsky, along with several European scholars (Lacroix, Cournot, De Morgan), attempted to simplify and explain Laplace's and Poisson's writings. He originated a serious study of probability theory in Russia and contributed to population statistics.

In 1888 **Bertrand** published a treatise on probability theory written in an excellent literary style and impregnated with its non-constructive negative and often unjustified, sometimes downright wrong attitude towards his subject. It included some interesting findings (for example, a study of the ballot problem) and is especially remembered in connection with his paradoxical problem on the probability that a "random" chord of a given circle was longer than the side of an equilateral triangle inscribed in the circle. Bertrand proved that the expression "at random" (actually, even "uniformly random") was not definite enough. A sudden pertinent result (De Montessus, in 1903) consisted in proving that the Bertrand problem had an uncountable set of answers. Bertrand had exerted a strong influence on **Poincaré** who apparently did not read (at least did not refer to) other previous authors, such as Chebyshev and Markov and even Laplace and Poisson.

Poincaré's main merit consisted in examining the notion of randomness. True, he offered several pertinent explanations without bothering to provide a single formula, but his main definition that connected randomness with instability of motion or position, as well as his statement that randomness always accompanied necessity, were extremely important. Poincaré also considered vivid examples showing the effect of random causes; thus, by means of his method of arbitrary functions he explained why, in the game of roulette, the probabilities of the two main outcomes coincided.

Poincaré proved that the existence of several answers to Bertrand's problem was caused by its actually describing various situations differing one from another. His finding made it possible to proceed further in the theory of geometric probabilities.

5.1. Chebyshev

He formulated the new and still being recognized aim of the theory of probability in the beginning of his essay of 1845: given, probabilities of some events, to find the probabilities of other events, connected with the given ones. In 1851 **Boole** followed suit (true, he thought about propositions rather than events) and he is meritorious for having been the first to stress, in 1854, that the theory should be axiomatized. Chebyshev's main findings concerned the law of large numbers and the central limit theorem, and among his numerous students were such scholars as Markov and Liapunov.

In 1846 Chebyshev rigorously proved the law of large numbers in the Poisson form, and, as was typical for him, he also estimated the precision of the appropriate pre-limiting relation. Then, in 1867, Chebyshev offered his derivation of the "Bienaymé – Chebyshev" inequality. He highly praised his predecessor and Markov, in several later contributions, kept to the same opinion. Chebyshev's name is however connected with that finding because Bienaymé paid no special attention to that inequality whereas Chebyshev, by issuing from it, developed the method of moments for proving the central limit theorem.

Indeed, in 1887, after achieving a breakthrough by formulating his inequalities involving integrals, Chebyshev was the first to prove that theorem "almost" rigorously and later on Markov corrected his teacher. Markov also preceded the former in publishing the demonstration of those inequalities.

For several decades Chebyshev's work remained largely unnoticed in the West, mostly because the theory of probability had still been then considered as an applied discipline not requiring rigor on a par with pure mathematics.

5.2. Markov and Liapunov

Markov definitively proved the central limit theorem by the method of moments, and he managed to extend essentially that method after Liapunov had demonstrated the same theorem under weaker conditions. Markov's second main contribution was the study of dependent random variables. In 1906 he began publishing findings connected with the applicability of the law of large numbers and the central limit theorem for such variables connected into various versions of "Markov chains". He possibly perceived that these chains could find wide applications, but had not himself worked in that direction. Among his other achievements were the publication, in 1888, of a table of the normal distribution, which, along with a later table of another author, had remained beyond compare for several decades. Markov also successfully studied stability of statistical series, and, by corresponding with Chuprov, drew him into the domain of mathematical statistics.

In 1900 and 1901 **Liapunov** had proved the central limit theorem by the method of characteristic functions and in 1922 Lindeberg highly praised the achievement of his predecessor.

5.3. Statistics and Natural Sciences

During the middle, and especially the second half of the 19th century, several new scientific disciplines, intrinsically connected with statistics, came into being. Among these, the kinetic theory of gases and biometry (born at the turn of that century) should be singled out; and **Helmert** completed the construction of the theory of errors. In 1860 **Maxwell** introduced his famous law of distribution of the velocities of monatomic molecules and greatly fostered the statistical approach to studying nature. He, and even to a larger extent **Boltzmann**, originated kinetic theory and statistical physics.

Darwinism prompted English mathematicians (**Pearson** in the first place) to study biological phenomena statistically and led to the birth of biometry. Darwin's hypothesis of evolution can be described as a discrete random process.

Helmert's findings include tests for revealing systematic errors (1875 and latter, with an independent derivation of the chi-squared distribution first discovered by Abbe in 1863), anticipation of the Student – Fisher theorem on the independence of the arithmetic mean and variance for the normal distribution; introduction of the “Helmert transformation”; and investigation of the precision of the sample variance (correcting a mistake made by Gauss and preceding the same finding achieved by Kolmogorov and others in 1947), – all of it done in 1876.

6. The Continental Direction of Statistics and the Biometric School

Up to the beginning of the 20th century statistical investigations on the Continent were mostly restricted to population statistics whereas in England the main field of statistical applications was biology. The originator of the Continental direction was **Lexis** whose predecessors had been Poisson, Bienaymé, Cournot and Quetelet. In 1879 Lexis proposed a test for revealing changes in probabilities of “success” in different series of Bernoulli trials; he tacitly dismissed as uninteresting the case of dependent observations. His work was really important since previous statisticians tended to restrict their attention to constant probability and independent trials.

Subsequent scientists (**Bortkiewicz, Chuprov, Markov**) studied the Lexis criterion deriving its expectation and variance, extended the investigation to include random variables in general but discovered that the Lexian theory had to be reconstructed. In the process, many findings were made, which, regrettably, were not taken up by the biometricians. Chuprov also exerted great and only partly successful efforts to unite both streams of statistical thought.

Karl Pearson, the originator of biometry and leader of the biometricians, developed the principles of correlation theory and contingency, introduced the system of “Pearsonian curves” for describing empirical distributions and the chi-square test and compiled numerous statistical tables. The methodological difference between the work of the two streams consisted in that, for several decades, biometricians kept to empiricism, to recognizing only statistical indicators and were paying scant attention to the theoretical patterns underlying their studies.

General sources

This essay, which replaces and extends my Russian article on the theory of probability before Chebyshev (IMI, vol. 23, 1978, pp. 284 – 306) is based on my previous writings, and, in particular, on a series of papers in the *Archive for History of Exact Sciences* (from 1971, vol. 7, onward) and on my *History of the Theory of Probability to the Beginning of the 20th Century* privately printed in 50 copies (Berlin, 2004). Below, I reprint the *General Sources* indicated in the bibliography to the just mentioned book.

Bernoulli, J. (1975), *Werke*, Bd. 3. Basel. Includes reprints of several memoirs of other authors and commentaries.

--- (1986), *О законе больших чисел* (On the Law of Large Numbers). Editor, Yu.V. Prokhorov. M. Contains commentaries.

David, H.A., Edwards, A.W.F. (2001), *Annotated Readings in the History of Statistics*. New York.

Farebrother, R.W. (1999), *Fitting Linear Relationships. History of the Calculus of Observations 1750 – 1900*. New York.

Freudenthal, H., Steiner H.-G. (1966), Aus der Geschichte der Wahrscheinlichkeitstheorie und der mathematischen Statistik. In *Grundzüge der Mathematik*, Bd. 4. Göttingen, 1966, pp. 149 – 165. Editors, H. Behnke et al.

Gauss, C.F. (1887), *Abhandlungen zur Methode der kleinsten Quadrate*. Hrsg. A. Börsch, P. Simon. Latest edition: Vaduz, 1998.

Gillispie, C., Holmes, F.L., Editors (1970 – 1990), *Dictionary of Scientific Biography*, vols 1 – 18. Gillispie edited the first 16 vols.

Gnedenko, B.V., Sheynin, O.B. (1978), Theory of probability. A chapter in *Math. of the 19th Century*, vol. 1. Editors, A.N. Kolmogorov, A.P. Youshkevich. Transl.: Basel, 1992 and 2001, pp. 211 – 288.

Hald, A. (1990), *History of Probability and Statistics and Their Application before 1750*. New York.

--- (1998), *History of Mathematical Statistics from 1750 to 1930*. New York.

- (2003), *History of the Law of Large Numbers and Consistency*. Univ. Copenhagen, Dept. applied math. & statistics, Preprint No. 2.
- Heyde, C.C., Seneta, E., Editors (2001), *Statisticians of the Centuries*. New York.
- Johnson, N.L., Kotz, S. Editors (1997), *Leading Personalities in Statistical Sciences*. New York. Collection of biographies partly reprinted from Kotz & Johnson (1982 – 1989).
- Kendall, M.G., Doig A.G. (1962 – 1968), *Bibliography of Statistical Literature*, vols 1 – 3. London.
- Kendall, M.G., Plackett, R.L., Editors (1977), *Studies in the History of Statistics and Probability*, vol. 2. London. Collected reprints.
- Kotz, S., Johnson, N.L., Editors (1982 – 1989), *Enc. of Statistical Sciences*, vols 1 – 9 + Supplement volume. New York.
- Kotz, S., et al, Editors (1997 – 1999), *Update Volumes 1 – 3 to Kotz & Johnson* (1982 – 1989). New York.
- Kruskal, W., Tanur, J.M., Editors (1978), *Intern. Enc. of Statistics*, vols 1 – 2. New York.
- Pearson, E.S., Kendall, M.G., Editors (1970), *Studies in the History of Statistics and Probability* [vol. 1]. London. Collected reprints.
- Pearson, K. (1978), *History of Statistics in the 17th and 18th Centuries against the Changing Background* etc. Lectures 1921 – 1933. Editor E.S. Pearson. London.
- Schneider, I., Editor (1988), *Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfängen bis 1933*. Darmstadt. Collected classical fragments almost exclusively in German.
- Sheynin, O. (1996), *History of the Theory of Errors*. Egelsbach.
- Sheynin, O.B. et al (1972), Theory of probability. Chapter in *История математики с древнейших времен до начала 19го столетия* (History of Math. from Most Ancient Times to Beginning of the 19th Century), vol. 3. Editor, A.P. Youshkevich. Moscow, 1972, pp. 126 – 152.
- Stigler, S.M. (1986), *History of Statistics*. Cambridge, Mass.
- Todhunter, I. (1865), *History of the Mathematical Theory of Probability*. New York, 1949, 1965.

11. Liapunov's Letters to Andreev

IMI, vol. 31, 1989, pp. 306 – 313

1. The life of Aleksandr Mikhailovich Liapunov is well known [1; 2]. Konstantin Alekseevich Andreev (1848 – 1921) graduated from Moscow University. From 1873 to 1898 he taught in Kharkov, then in Moscow. In 1884 – 1899 he was President of the Kharkov Mathematical Society (KhMSoc) and Editor of its *Soobshchenia* [3]. In 1884 he was elected Corresponding Member of the Imperial (Petersburg) Academy of Sciences.

Gordevsky published passages from Andreev's letters to Liapunov [3]¹ but provided no commentaries and neither did he say anything about Liapunov's letters to Andreev which I discovered in the Archive of the Moscow State University.²

Andreev's first letter (17.2.1899; all dates are given here in the old style) is not connected with the newly discovered letters of Liapunov, and I leave it alone. Liapunov's first letter (11.6.1897) is also left out since it was only concerned with his stay in Crimea with one of his brothers and described how useful was it for his brother's health. He did not name his brother, but a paper by Academician Boris Mikhailovich Liapunov [1, p. 11], a philologist, mentioned a trip to the Crimea in May – June 1897 by A.M., his wife, mother-in-law and himself (but did not provide any details).

Almost all of Liapunov's subsequent letters are connected with the appearance of his paper [4] as a response to Nekrasov's criticisms [5]. Another subject was the preparation of a new charter for Russia's universities. Liapunov participated in the work of the appropriate commission at Kharkov University [1, p. 11]. On 29 April 1901 the Ministry of Public Education circulated proposals concerning the new charter [6, pp. 1 – 4], but even before that some universities had begun to discuss the causes of the then occurring students' unrest and to suggest "measures for putting university life in order" (Ibidem, p. 5).^{2a}

2. Pavel Alekseevich Nekrasov (1853 – 1924) was an eminent scholar, a professor at, and for several years rector of Moscow University, but at the turn of the 19th century his scientific work underwent a radical change. He started connecting mathematics with religion and politics and became unimaginably verbose so that his writings of that period are still unstudied. Here is Pavel Youshkevitch's pertinent opinion from his forgotten newspaper article [7]: Nekrasov is

A great lover of philosophy[...] but the philosophy of this honorable scholar is of an absolutely special nature. It is a strangest medley of senseless profundity with tedious verbiage and dried up words.

Nekrasov [5] blamed Liapunov for mistakes and shortcomings allegedly committed in the classical memoir [8]. His considerations were however either unfounded or indefinite, or did not bear any relation to the substance of Liapunov's work. A similar conclusion can be made about Nekrasov's accusations of Chebyshev and Markov which he adduced for good measure. Being blinded by his criticism, Nekrasov even mixed up the notions of limit and asymptotic representation of a function.

3. I reproduce now Liapunov's letters

Liapunov – Andreev, 29.3.1901

Highly respected Konstantin Alekseevich,

I am applying to you with a great request. You have probably already received an offprint of Nekrasov's note [5] where he makes charges against me, Markov and Chebyshev. If you had acquainted yourself with my papers you certainly noticed that Nekrasov completely perverts the truth and does it with impudence beyond any measure. Although I have not the slightest desire to enter into a debate with him, I am compelled to answer him. And I am therefore sending you a manuscript of my *Answer* and am asking you to assist in its publication in the *Matematicheskyy Sbornik*, – if possible, in the same issue where Nekrasov's note is to appear. I would not have troubled you with this request if it were possible to publish my *Answer* in the *Soobshchenia*.³ But our relations with Silberberg have definitely deteriorated and we are compelled to postpone the printing of the *Soobshchenia* at least for the time being.⁴

This is why I have to apply for help to the *Matematicheskyy Sbornik*. I think, however, that I have some right to do so since I am a member of the Moscow Mathematical Society [MMS].⁵ You will perhaps find the tone of my *Answer* rather sharp. But what can I do? I attempted to assuage it as much as possible, but everything has its boundary and I cannot go any further in this direction. It is already sufficient that I do not accuse Nekrasov of deliberately perverting the truth (about which I have no doubt) and that I explain his strange attacks against me by his ignorance of the substance of my first paper.

Understanding that my request is putting you to some inconvenience, I am asking you to excuse me after taking into consideration that my position with respect to this case is almost desperate: I am unable to publish my *Answer* elsewhere. However, your assistance can be restricted only to passing my *Answer* to the President of the Society⁶ and to informing him of my desire. You will do me a great favor by fulfilling my request.

How are things going on at the University? I heard about your Commission, but it is impossible to obtain here any definite information about its purpose. [...]

Andreev – Liapunov, 31.3.1901 [3, pp. 40 – 41]

Andreev will pass Liapunov's manuscript to B.K. Mlodzeevsky, the Secretary of the MMS, and will speak to Bugaev. He will not undertake to judge the debate between Liapunov and Nekrasov. Nekrasov

Reasons perhaps deeply, but not clearly, and he expresses his thoughts still more obscurely. I am only surprised that he is so self-confident. In his situation, with the administrative burden weighing heavily upon him,⁷ it is even impossible, as I imagine, to have enough time for calmly considering deep scientific problems, so that it would have been better not to study them at all.

Andreev is sceptical with regard to the university reform, and he congratulates Liapunov with his being elected Corresponding Member of the Imperial (Petersburg) Academy of Sciences and even with his "future fuller entry into the Academy".

Liapunov – Andreev, 8.4.1901

Highly respected and dear Konstantin Alekseevich,

I thank you for congratulating me with my election to corresponding membership of the Academy, and for your good wishes. As to my fuller entry into the Academy, at which you hint, this is not yet decided, and it is impossible to say how will it be decided. But, since we are discussing this subject, I ought to tell you that I was asked to stand, and gave my consent. But this will only be definitively decided by autumn. At present, it would please me if this business is not spoken about.

V.A. Steklov, who had just arrived from Moscow, visited us today. He told us many interesting things about your university life. It was very pleasant to find out that the report of our faculty committee is finally somewhat on the move and that it is now being used as an initial material by your committee.

I asked Steklov to visit you before leaving Moscow and to take the manuscript of my *Answer* if its publication was not considered possible. But Steklov informed me that you had already passed it to Mlodzeevsky and that he was unable to ascertain whether it will be published. I would therefore ask you, highly respected Konstantin Alekseevich, to inquire of Mlodzeevsky, while meeting him, how was this business decided. If the article will not be published in the *Matematichesky Sbornik*, I would like to receive the manuscript back. In this case I shall publish it in the *Universitetskie Zapiski*⁸ (about whose existence, as I ought to add to my shame, I had completely forgotten when sending you my manuscript).

If, however, it is decided to publish the article in the *Matematichesky Sbornik*, I would ask you to inform those responsible that I certainly desire to read the proofs [...]

Please excuse me for all the troubles I am inflicting on you. I am very grateful for the assistance rendered me in this disagreeable business. [...]

Andreev – Liapunov, 13.4.1901 [3, pp. 41 – 43]

Bugaev and Nekrasov do not want to publish Liapunov's manuscript. Nekrasov, however, agreed to its being put out, but not earlier than in a year, and with his objections added in a separate note. Andreev considers it expedient to extend the manuscript and soften its tone.

Liapunov- Andreev, 21.4.1901

Dear and highly respected Konstantin Alekseevich,

I am grateful to you for sending me the manuscript and for all the troubles encountered when taking it back. From your previous letter I concluded that it was hardly possible to count on its speedy return and therefore began to edit a new version of the *Answer*. And, in accord with your advice, I have essentially extended my article depicting in detail the entire factual aspect of the business without leaving a single objection of Nekrasov unanswered. And I think that because of this very circumstance my new *Answer* will cause

Nekrasov considerably more annoyance. Perhaps he will even regret (tacitly of course) that he was not quick to publish the *Answer* in its old version.

Did you conclude work in the commission? If its results were reported to the University's Council, it would be interesting to know the attitude of that body. [...]

Liapunov – Andreev, no date

Dear and highly respectable Konstantin Alekseevich,

It was extremely pleasant to hear from you. We came to know that the operation essentially benefited you and that at present your health is largely restored which greatly gladdened us. It would be nice to meet you. However, [...] Yesterday I had informed A.N. Krylov over the telephone about your wish to have the book that he published,⁹ and he answered me that it will be sent to you in a few days.

This autumn we moved into a kazenny apartment.¹⁰ It is small but cosy and sufficiently spacious for the two of us. And it is very warm, which is indeed valuable in this severe winter. Its only, but really essential defect is that it is somewhat dark: it is on the ground floor and its windows, opening on the street, are directed towards the north-east. [...]

I am now busying myself far less diligently than before. The occurring international events hold my attention to such an extent that I do not even wish to think about anything else. In addition, scientific pursuits demand a calm mood whereas the events occurring around us are very often so disgusting that they can only strongly irritate and embitter. In such cases scientific pursuits can only serve for distracting the thoughts and cannot be fruitful.¹¹ [...]

Notes

1. These letters are kept at the Archive of the Academy of Sciences of the Soviet Union in Leningrad. Fond 257, Inventory 1, No. 29.

2. Fond 217, Inventory 1, No. 87.

2a. Later note. See Correspondence between P.A. Nekrasov and A.I. Chuprov (translated in this collection), Nekrasov's letter of 17 Febr. 1899, concerning the students' unrest at Moscow University and description of similar events in Kiev directly involving Slutsky [9].

3. Liapunov bears in mind the periodical of the KhMSoc.

4. The *Soobshchenia* were printed in the Kharkov printing office M. Silberberg & Sons. Judging by Liapunov's *Imprimatur* inscriptions, the appearance of the issues of its vol. 7 had been irregular. The first issue even had two such signs, 30.11.1900 and 10.4.1902.

5. Liapunov was member of that Society from 1892 (*Matematichesky Sbornik*, vol. 16, 1891, p. 845).

6. The President of the MMS was N.V. Bugaev.

7. Nekrasov was then warden of the Moscow educational region and vice-president of the MMS.

8. More precisely, in the *Zapiski Khark. Univ.*

9. The sequel proves that the letter was written during World War I. According to the Bibliography of the works of Krylov [10, pp. 39 – 41], the only book that he then “published”, was his Russian translation of Newton's *Principia*. It appeared in 1915 – 1916.

10. In this particular instance, the proper translation seems to be: apartment, belonging to the Academy.

11. In 1916 Liapunov published two papers, both in the *Izvestia* of the Petrograd Academy of sciences. The same year he submitted one more paper, and it appeared in 1917, in the same periodical [11].

References

1. Liapunov, B.M. A brief essay on the life and work of A.M. Liapunov. *Izv. Akad. Nauk SSSR*, Otdel fiz.-mat. nauk, ser. 7, No. 1, 1930, pp. 1 – 24. (R)

2. Zesevich, V.P. *A.M. Liapunov*. M., 1970. (R)

3. Gordevsky, D.Z. *K.A. Andreev*. Kharkov, 1955. (R)

4. Liapunov, A.M. An answer to P.A. Nekrasov. *Zapiski Khark. Univ.*, vol. 3, 1901, pp. 51 – 61. Transl. in DHS 2579, 1998, pp. 53 – 63.

5. Nekrasov, P.A. Concerning a simplest theorem on probabilities of sums and means. *MSb*, vol. 22, No. 2, 1901, pp. 225 – 238. Transl. In DHS 2579, 1998, pp. 43 – 51.

6. *Труды высочайше учрежденной комиссии по преобразованию высших учебных заведений* (Proc. Roy. Comm. on reforming Acad. Institutions). Psb, 1903, No. 1. Publ. as a manuscript.

7. Youshkevich, P. [S.] On one scientific debate (1915). *IMI*, vol. 34, 1993, pp. 207 – 209. (R)

8. Liapunov, A.M. Sur une proposition de la théorie des probabilités. *Bull. [Izvestia] Acad. Imp. Sci. St.-Pétersb.*, t. 13, 1901, pp. 359 – 386. Short version : *C.r. Acad. Sci. Paris*, t. 132, 1901, pp. 126 – 128.

9. Chetverikov, N.S. Life and scientific work of Slutsky (1959). In author's book *Статистические исследования* (Statistical Investigations). M., 1975, pp. 261 – 281. Translation to appear in 2004 in the DHS series.

10. Krylov, A.N. *Собрание трудов* (Coll. Works), vol. 12, pt. 2. М. – Л., 1956.

11. Lukomskaia, A.M. *А.М. Ляпунов. Библиография* (Liapunov. Bibliography). М.-Л., 1953.

12. On the History of the Statistical Method in Natural Sciences

IMI, vol. 32/33, 1990, pp. 384 – 408

This is a translation of the Russian résumé of my English articles [1 – 5] devoted to the application of the statistical method in separate branches of natural sciences mainly during 1750 – 1870. I omit the beginning of my résumé where I discussed the definitions of statistics, statistical method and exploratory data analysis since I treated all this elsewhere [6].

1. Conclusions Not Formulated in My English Papers

1) The statistical method had been developing independently in each branch of natural sciences. From 1830 that process was going on against the background of a fading interest in probability; indeed, neither Darwin nor Boltzmann ever referred to Laplace.

2) There existed a permanent contradiction between statistics and the concrete science where it had been applied. Statistical data showed the direction for developing the latter, but, after some progress was achieved, they usually became useless.

3) By the mid-19th century statistical populations requiring examination were revealed in separate branches of natural sciences.

4) A number of natural-scientific disciplines directly connected with statistics originated in the 19th century; furthermore, for a long time statistics determined the development of some other branches of knowledge.

5) Most important discoveries in some branches of natural sciences were obtained by studying the collected statistical data without applying any (still unknown) statistical tests.¹

6) “Tabular” statistics appeared also in natural sciences.

7) In some branches of natural sciences the statistical method was initially reduced to deriving mean values (or states); then, however, it began to include the study of the pertinent deviations.

8) The Darwinian evolution of species can be represented as a discrete stochastic process.

9) Humboldt’s theoretical findings in natural sciences were based on his studies of the mean (or averaged) states of nature.

10) The first quantitative study of a correlation dependence occurred in 1865 (Seidel), and in 1912 Kapteyn introduced an “astronomical” correlation coefficient.

11) In 1858 Clausius introduced a linear function of an integral law of distribution, and the distribution itself was infinitely divisible.

12) In 1873 Maxwell anticipated Poincaré’s idea that a random event was present when small causes under unstable equilibrium led to essential consequences.

13) The prehistory of exploratory analysis should include the works of Halley (1701), Humboldt (1817) and Galton (1863).

In the sequel, I cite these conclusions by indicating the appropriate numbers in curly brackets.

2. Medicine

2.1. The Numerical Method {6}

The statistical method found its way into medicine along several directions. Population statistics was always closely linked with medical statistics. Epidemiology and public hygiene, that date back to the mid-19th century (§§2.4 and 2.5), were based on statistics {4}; even the inoculation of smallpox (from the 1720s to Jenner) required statistical data and appropriate investigations. Surgery, also from the mid-19th century, needed statistics as an indispensable tool of research (§2.3) {4}. Finally, the numerical method (**Louis**, 1825) of examining the symptoms of various diseases, which amounted to compiling statistical summaries, had been in vogue for a few decades. Pirogov (1849) and Davidov (1854) favorably mentioned it although the latter perceived its restrictions. Even before Louis physicians and other scientists recommended a similar usage of statistical data in medicine (Dalembert, 1759; Black, 1788; Condorcet, 1795), and Pinel (1801) published a statistical study of the treatment of mental patients. Neither he, nor Louis demanded estimation of the plausibility of their conclusions.

2.2. Elements of Mathematical Statistics

In 1835 the Paris Academy of Sciences debated the problems of applying probability to therapeutics but did not decide anything. **Poisson**, however, declared that, from the angle under discussion, medicine did not differ from other sciences. **Gavarret**, his former student at the Ecole Polytechnique, became a physician and published a book (1840) on the principles of medical statistics. Indicating that he had issued from Poisson’s ideas, he recommended the estimation of the plausibility of any conclusions by the De Moivre – Laplace limit theorem in accord with a previously chosen level of significance (a later term) and the checking of initial hypotheses by a method due to his teacher. For many decades, authors invariably included Gavarret’s

formulas in their treatises and at the very least he thus paved the way for a speedier introduction of statistical ideas and methods into medicine. In Russia, Davidov continued this line of development (Ondar, 1971).

2.3. Surgery

The statistical method began to be applied in surgery from 1839 for studying the results of amputations. The introduction of anesthesia (that sometimes led to serious side effects) demanded a comparison of mortality from amputations made with and without it. **Simpson** published such a study in 1847 – 1848, and **Pirogov**, in 1849. The latter began practising anesthesia in military surgery. Simpson used heterogeneous data mistakenly believing that that approach secured higher plausibility, cf. §5.4. The definitive introduction of anesthesia and of the Listerian methods of antiseptics, whose beneficial influence became immediately evident {5}, made previous data useless {2}.

In 1860 the International Statistical Congress adopted **Florence Nightingale's** proposals for an uniform plan of hospital statistics. She indicated that the conditions in surgical hospitals were better described by post-operational complications than by mortality. Nightingale (and Pirogov, Virchow and Simpson) noted that mortality in large hospitals was higher than in smaller medical institutions. According to Simpson (1869 – 1870), it increased monotonically with the number of beds. By then, a steady and practically real change of an indicator became a convincing argument in medical statistics {5}.

In military surgery, statistical data were unreliable and Pirogov advised to trust only “sober” observations. Comparing the conservative treatment of the limbs with amputation, he called his time an “Übergangsperiode”. During the first decades of the 19th century, he continued, “die alte, nicht statistische Schule” did not allow for the danger of operations; now, however, the old “geheiligten Grundsätze [...] sind durch die Statistik [comparison of the appropriate death-rates] erschüttert”, but no new principles have yet emerged, and cannot emerge in the absence of reliable data.

Pirogov did not engage in mathematical statistics, but he widely used the statistical method and came to realise that mass phenomena exhibit stable regularities.² This helped him to solve problems inherent in organizing military surgery {5}.

2.4. Epidemiology

Modern epidemiology attempts to predict the course of epidemics. In the 18th – 19th centuries scientists did not yet aim at such goals; **Farr's** investigation (1866) provided an exception, and even it had to do with veterinary science rather than with medicine. Another exception was **Enko's** study (1889) of measles.

Even in the 1720s inoculation of smallpox led to statistical problems. **Daniel Bernoulli** (1766) tried to examine thoroughly the benefits of that treatment, but neither he, nor **Dalembert**, who criticized him and put forward his own proposals, were able to answer all the necessary questions (Karn, 1931). The Jennerian vaccination fared otherwise. From the very beginning, its results were splendid {5}, but even in this case some concrete technological issues had risen and Simon (1887) argued that reliable national statistics was needed to solve them.

In the 19th century, cholera epidemics repeatedly struck Europe. In 1855 **Snow** compared the data on mortality from cholera in London with the quality of drinking water there. It occurred that the mortality of those who drank purified water was eight times lower than of the others and that fact at once indicated the cause of cholera epidemics {5}. In 1886 – 1887 **Pettenkofer** published a review of writings on cholera. He adduced a large number of statistical tables but was unable to interpret them adequately. Nevertheless, Pettenkofer argued that no cholera epidemics was possible “wenn der Ort [...] keine locale Disposition besitzt” and his opinion is still carrying some weight. The International Statistical Congress repeatedly discussed the statistics of epidemics (especially of cholera). In 1872 it recommended to test the Pettenkofer statement statistically, but this was not implemented.

In 1865, **Seidel**, an astronomer and mathematician, compared the monthly number of typhoid fever cases with the level of subsoil water. Calculating the deviations of these magnitudes from their yearly mean values, he found out that their signs for the same months coincided twice more often than not. This meant that the investigated dependence was significant. In 1866 Seidel extended his research by additionally taking into account precipitation. He thus quantitatively estimated the significance of a correlative relation between two and even three variables, although only in a roundabout way connected with loss of information. Soyka and Virchow made similar studies but on an elementary mathematical level.

2.5. Public Hygiene

Already **Leibniz**, in his manuscripts, formulated recommendations pertaining to public hygiene. And, from its very origin in the mid-19th century, this discipline began to study statistically a great range of problems which were also discussed by the International Statistical Congress and at least some of which nowadays belong to ecology.

In 1842 **Chadwick** described the unsavory sanitary conditions in England. **Pettenkofer** (1873) estimated the financial losses incurred by Munich from such diseases as typhoid fever. The city council adopted his recommendations and mortality from that disease fell there from 0.15 to 0.08%. In 1887, his student, **Eris-mann**, published a treatise on sanitary statistics. He indicated that “even recently” there had existed a negative opinion about that subject; turned the attention of his readers to the methods of collecting data; and explicated Davidov’s ideas about the quantitative tests of the reliability of statistical inferences.³

3. Biology

3.1. Various Problems before Darwin

The attempts to apply statistical methods in biology began not later than in the mid-18th century. Botanists (**Adanson**, 1757) desired to use “natural” methods of classifying plants that preserved the “distances” between species; they had thus searched for an answer to a problem belonging to multivariate statistics.

In 1738 **Réaumur** discovered the law of the “sums of temperatures”; he established that leaves, flowers and fruit appeared on plants of a given species after the sum of the mean daily temperatures had attained certain values. **Aug. De Candolle** (1832) qualitatively compared botanical observations with the results calculated according to that law and recommended to standardize the method of such observations. In 1846 **Quetelet** proposed to replace the sums of the temperatures by the sums of their squares but was unable to estimate quantitatively the advantages of his law as compared with the old one. De Candolle (1832) also published vast statistical data on the consumption of oxygen by plants in darkness, on the content of water and sugar in fruits, etc. {6}.

Not later than in 1833 **Babbage** began collecting statistical data of the life of animals and compiled a statistical questionnaire about mammals {6}. A statistical study of fishing in Russia (**Baer** et al, 1860 – 1875) was of direct practical importance {6} and possibly directed Baer towards theoretical problems in animal ecology (Valt, 1978). In 1882 **Pasteur** tested his vaccine against anthrax on many thousands of animals. The results were brilliant, and he did not have to worry about their mathematical treatment {5}.

Compilation and analysis of statistical data became a most important component of geography of plants, a new discipline created by **Humboldt** at the beginning of the 19th century {4}. In 1858 the International Statistical Congress published a questionnaire partly devoted to it and to zoogeography. Anthropometry, which originated in the second half of the same century {4}, was also directly linked with statistics. Its pioneer was **Quetelet** (1871) but its name was due to Humboldt.

3.2. Various Problems: Darwin

Darwin engaged in various aspects of statistics. His writings include a large number of statistical tables. Thus, he compiled data on the sex ratio at birth for a few species of animals; when studying the occurrence of six-fingered humans, he formulated and asked Stokes to solve a concrete problem on the realization of rare events,⁴ when investigating the advantages of cross-fertilization as compared with spontaneous pollination, he asked Galton to check the significance of his conclusions.⁵ Darwin’s requests for assistance deserve every praise.

When studying the life of earthworms, Darwin examined how did they carry away paper triangles into their burrows. He considered three possible cases of random dragging and reasonably rejected all of them in favor of a non-random process (of some kind of sensible dragging). In 1888 Bertrand proved that the term “at random” was not sufficiently precise (cf. §5.2) and he could have well referred to Darwin.

3.3. Evolution of Species: Darwin

Darwin was the main author of the theory (more precisely: hypothesis) of the evolution of species. He studied the origin of new varieties, subspecies and species, the nature of the variations between individuals of the same generation (“horizontal” variations) and between parents and offspring (“vertical” variations) {3}. Darwin made use of such notions as natural and sexual selection, variations, variability, without precisely defining any of them and it is sometimes difficult to interpret his pronouncements.⁶ And, while repeatedly reasoning on randomness of varieties, Darwin certainly did not understand that notion in a unique way.

According to Darwin, the evolution of a species was caused by the action of small permanent influences or by small differences between the probabilities of two or several events. He did not mention any theorems,⁷ nor did he provide any quantitative estimates {1} (which would have been very difficult). Indeed, how can we determine the probabilities of mating of two individuals having definite qualities, of some vertical variation, etc.?

I represent now the evolution of species according to Darwin in the following way {8}. Introduce an n -dimensional system of coordinates, of the parameters of individuals of a given species,⁸ and the corresponding Euclidean space with the usual definition of the distance between two of its points. At moment t_m of discrete time the i th individual is represented by a point $U_i (u_{i1}; u_{i2}; \dots; u_{in})$; at moment t_{m+1} the corresponding points are the individuals of the next generation, which, due to the vertical variations, will be in

somewhat differing positions. Introduce also a moving point or subspace $V(t)$ corresponding to the optimal conditions for the existence of the species. Then the evolution of species will become a discrete stochastic process in whose course the individuals approach V and the set $\{U\}$ of the individuals of generation t_m constitutes its section. Also required are the appropriate probabilities, but they remain unknown (see above). Moreover, Darwin attached great importance to changes in mind and instinct of animals and even mentioned the natural selection of spontaneous variations of instinct. But how can we insert such notions in a quantitative model?..

The Darwinian theory led to the appearance of new fundamental problems, and, together with the **Mendelian** theory of heredity (unnoticed until the beginning of the 20th century), for many decades determined the development of biology. In addition, Darwin had decisively influenced a group of English scientists who created the Biometric school for a mathematical cum statistical study of biology {4}. Mathematical statistics itself became a discipline in its own right owing to a large extent to the work of that school headed by **Karl Pearson**. The stochastic essence of the evolution theory was evident both for Darwin's partisans and opponents, only Boltzmann (§6.4) thought that it was of a mechanical nature.

3.4. Statements Made by Biologists

In the adduced summary I describe the statements made by several scientists (excepting Darwin) in a generalized and sometimes formalized way. They, the statements, mostly consider the evolution of species that became studied from about the mid-18th century. It may be thought that from then onward variations gradually became one of the main objects of study in biology {3}.

- 1) Adanson (1757), Aug. De Candolle (1813): natural classification of plants; individuals are points in a many-dimensional space.
- 2) Goethe (1790): space forms of plants; same.
- 3) Maupertuis (1745): heredity; vertical variations are random and small.
- 4) Maupertuis (1751), Cournot (1851): randomness; its role in evolution is restricted.
- 5) Lamarck (1809): changes in external conditions indirectly lead to hereditary variations in individuals.
- 6) E. Geoffroy Saint-Hilaire (1822): new species originate due to random mutations.
- 7) I. Geoffroy Saint-Hilaire (1851): restricted evolution of species is due to changes in external conditions.
- 8) Goethe (1831): evolution of species is due to random changes both in the species and external conditions.
- 9) Cournot (1861): same; vertical variations are random, the evolution of species is restricted.

4. Meteorology

4.1. Stages of Development

In 1850 **Buys Ballot** isolated three stages in the newest (from 1801) history of meteorology: the study of mean states (**Humboldt**); of the deviations from these (**Dove**); and the future stage "wo wir versuchen können meteorologische Begebenheiten voraus zu sagen" {7}. The first two periods were obviously of a statistical nature {4} whereas the third one apparently started in the 1870s when observations made in different countries began to be coordinated, and weather maps applied. Indeed, under these conditions the study of the space distribution of meteorological elements became possible and the forecasting of some meteorological phenomena could have started. At the same time, the confidence in general conclusions made by issuing from observations at separate locations (in particular, the belief that the Moon influenced the weather, see §4.2 {2}) was shattered. But already the study of the *deviations* marked the beginning of the investigation of the temporal, if not spatial-temporal distribution of the elements.

4.2. The Influence of the Moon

This was studied already in the beginning of the 18th century. In 1777 **Toaldo** summarized the data pertaining to 1671 – 1772 on the changes of the weather in different places as compared with the phases of the Moon. He concluded that the influence of the Moon had been essential and he could have corroborated his reasoning by the De Moivre – Laplace theorem (already known to De Moivre) but did not do so. Toaldo certainly made use of heterogeneous observations, and the binomial pattern adopted by him was hardly appropriate (see §4.7).

Lamarck thought that the Moon strongly influenced the weather. In 1810 he isolated 25,520 (!) *genres* of that influence depending on the mutual position of the Earth, Moon and Sun and on other circumstances, but he did not even hint at a quantitative theory. Schuebler (1830) remarked that the Moon possibly affected the weather because of "chemischer Verbindungen und Zersetzungen" of particles in the atmosphere rather than owing to the "Gesetzen der Attraktion". He thus sidestepped the insignificance of the lunar tides proved by Laplace and Bouvard.

Only two years had to pass before **Arago** reported that scientists do not anymore believe in the influence of the Moon; he himself was then yet at a loss. But in 1845 he declared that "les influences lunaires et

cométaires sont presque insensibles”. Nevertheless, Muncke (1837), in an authoritative review, arrived at an opposite conclusion. Glaisher (the father), who studied the influence of the Moon on the direction of the wind (1867) and precipitation (1869) in Greenwich, agreed with him, but in 1873 **Koepfen** stated as a self-evident fact that the influence of the Moon was insignificant.

4.3. Observations

Networks of meteorological stations began to appear not later than in the mid-17th century. During the 1730s and 1740s regular observations had been carried out in several Siberian cities and in 1733 **Daniel Bernoulli** compiled a manual for the Siberian stations. In 1780 the *Societas meteorologica Palatina* (in Pfalz, Germany) was founded. It existed for about 20 years and its stations in several European countries had been working according to a common set of rules. The *Societas* was the first to organize field observations (of any kind) on an international scale.

In 1801 **Lamarck** compiled a plan of meteorological observations for France. They were indeed commenced (with his participation) and continued until 1809 or 1810. **Koepfen** (1875) highly praised **Quetelet’s** merit of compiling and systematizing meteorological observations in Belgium “since the early 1840s”, and in 1850 Faraday indicated, in a letter to Quetelet, that his observations of atmospheric electricity were important. Not later than in 1843 **Humboldt** and **Kupffer** proposed a plan of observations in Russia to the Petersburg Academy of Sciences.⁹ Unification of observations was the main subject under discussion at the International Meteorological Congress in 1873.

4.4. Mean States

In 1818 **Humboldt** attempted to find out “les mouvements moyens de l’atmosphère”. Later on, in 1845, he conditioned the investigation of the totality of phenomena in nature by discovering the appropriate mean values (or states) {9}. True, his definition of climate (1831) was not directly connected with mean values, but subsequent authors have formulated this tie ever more explicitly, and Chuprov (1922), evidently expressing an established opinion, identified climate with a system of certain mean values.

In 1817 Humboldt introduced the notion of isotherms and contour lines of equal temperatures for two seasons, winter and summer; plotted the isotherms of 0, 5, 10, and 15° on a world map; estimated the fall of temperature with altitude; and calculated the mean temperature of the seasons for the belts situated between his isotherms. He thus isolated climatology from meteorology {4}. Even in 1811 – 1817 Humboldt turned his attention to the importance of generalizing and eliminating the “causes locales” before plotting the isotherms. That problem, as also the very introduction of the isotherms (“ganz nach Analogie von **Halley’s** isogonischen Curven [1701] geformt”), belong to the prehistory of the exploratory data analysis {13}.¹⁰

4.5. Deviations from the Mean

Dove (1837) came out against the “Herrschaft der Mittel” and maintained that the deviations from the mean state of the atmosphere should be studied. In 1848 he argued that a spatial-temporal study of the air temperature was necessary {3}. He also published vast data on the weather and it seems that he thought of abandoning the monthly isotherms, which he himself introduced, in favor of weekly contour lines or even of reducing the study of the weather to compilation of data {6}.

Koepfen (1874) indicated that the “Einführung der arithmetischen Mittel” into meteorology enabled an orientation in the weather, but that the appropriate causal connections should still be cognized. He added that the “allerverschiedensten Zustände” were “zusammen vergraben” in the arithmetic mean which was therefore “Nichts Wirkliches sondern eine abstracte Grösse“ {2}. **Davidov** (1857, not with respect to meteorology) and **Lamont** (1867) indicated that some means were often of an abstract nature. The latter also stated that

Die unregelmässigen atmosphärischen Änderungen nicht als Zufälligkeiten, im Sinne des Probabilitäts-Calculs, sondern als Schwankungen von ungleicher Zeitdauer auffassen müsse.

He apparently thought that, for example, the boundaries of these changes remained unknown, but his opinion was too pessimistic: temporal changes (e.g., in the weather, or in the number of sunspots, see §5.1) are now considered in the context of time (statistical) series. From about 1837 Lamont abandoned temporal changes; instead, he studied the differences of simultaneous observations made at stations situated close to each other, but hardly anyone followed suit.

4.6. Abandoning the Theory of Errors

Quetelet (1846) knew that temporal changes in a meteorological element were often asymmetric. He published the letters sent him in 1845 by **Bravais** who had provided examples of asymmetric densities from astronomy, anthropometry and meteorology. Nevertheless, in 1853 Quetelet declared that only “causes spé-

ciales” and anomalies corrupted the [normal] distribution of meteorological elements. In 1891 **Meyer**, after referring to the asymmetry of the temporal changes, declared that “Die Fehlerrechnung ist in der Meteorologie principiell unzulässig”. It is remarkable that **Pearson** (1898) applied Meyer’s data for testing the applicability of his theory of asymmetric curves to studying antimodal densities.

4.7. The Weather Depends on Its Previous States {5}

Already **Lamarck** (1804) knew that fact. In 1793 **Dalton** studied the influence of auroras on the weather at one station. He compared the number of periods of fair weather occurring after an aurora with the number expected if that phenomenon was not taken into account. The first number considerably exceeded the second one, and Dalton concluded that auroras were advantageous for the weather. He did not however count the actual periods of fair days occurring irrespective of the auroras.

Quetelet (1852) and **Koeppen** (1872) studied the tendency of the foul (or fair) weather to persist by applying the elements of the theory of runs. The first problem belonging to that theory concerned games of chance, but the first application of that theory in natural sciences apparently began in meteorology. When, at the beginning of the 20th century, Markov created the theory of his chains, he considered the interchange of vowels and consonants in the Russian language but did not mention meteorology.

4.8. Lamarck

Lamarck distinguished himself by concrete findings in meteorology in which he had been engaged during almost his entire scientific life. He thought, although not consistently, that that science was the “*théorie de l’atmosphère*” (1802), and, in 1802 – 1810, introduced the term “*météorologie statistique*” and its equivalent, “*statistique atmosphérique*” (the study of the climate and the winds, and of the influence of meteorological elements “*sur les animaux, sur les végétaux et sur le sol même*”). In 1846 **Quetelet** reasonably argued that the statistical part of meteorology should not be included into statistics; indeed, disciplines such as climatology nowadays belong to the appropriate branches of natural sciences.

Lamarck postulated the existence of statistical meteorology, – of a somewhat more extensive part of meteorology than climatology, – and urged that the laws governing the variations in the atmosphere be established. Without disavowing that standpoint, **Humboldt** directed his efforts towards the then solely possible statistical approach to meteorology. In 1800 – 1810 Lamarck published eleven meteorological *Annales* containing hardly successful forecasts of the weather in France as a whole and his theoretical considerations about meteorology.

5. Astronomy

5.1. The system of the World

Beginning with Kepler, astronomers attempted to discover numerical regularities in the solar system, and astronomy to a large extent caused the Laplacean theory of probability. **Laplace** deductively proved a number of astronomical facts after establishing their existence by a stochastic analysis of observations (regrettably left out of his writings). Two exceptions are, a statement (due to **Daniel Bernoulli**) that the coincidence of the directions of rotation of the planets and their satellites was unlikely; and a calculation of the expected mean inclination of the planetary orbits (1776).

In 1869 **Newcomb** compared the theoretical (calculated according to the uniform distribution) and the actual parameters of the orbits of the minor planets, but he was naturally unable to appraise quantitatively his results. For him, these planets constituted elements of a single statistical population {3}. In 1900 he completed a methodologically similar study. **Poincaré** (1896) estimated the total number of minor planets through their known number by applying simple stochastic considerations. Both he and Newcomb confused the notions of mean and probable values of a random variable.

In 1844, **Schwabe**, drawing on his observations of 1826 – 1843, established that the number of sunspots varied periodically {3}. Without providing any analysis, he noted that the period (T) roughly equalled 10 years. In 1859 **Wolf** compiled the observations of the sunspots beginning with the mid-18th century and concluded that $T = 11.1$ years. Then, about 1881, he analysed observations which lasted 120 years by comparing 19 hypotheses concerning the value of the period sought. He calculated the deviations of the mean data for the separate periods from the general mean number of sunspots and applied two tests: the range of the deviations, and the root of the sum of their squares divided by their number, and he concluded that there existed two periods, 10.0 and 11.3. At present, the existence of a precise period is denied, but, anyway, the numbers of sunspots constitute a time series, i.e., an object studied by the theory of probability.

Sabine (1852) noted that the variations of the number of sunspots were positively correlated with those of magnetic declinations {5}. Later, in 1878, this connection was denied, but Sabine’s opinion has eventually been found to be true. During 1872 – 1880 the connection between sunspots on the one hand, and cyclones and atmospheric pressure on the other hand, was established (Meldrum, Lockyer, Blanford) {5}. It is remarkable that no-one had then suggested that a mathematical theory (of correlation) would have been desirable.

5.2. Michell's Problem

Supposing that a few thousand stars were scattered “by mere chance” (uniform randomness) over the sky, Michell (1767) attempted to determine the probability that two stars were close to each other. His calculations were wrong, but his problem became classical. In addition, he used geometric probabilities which came to be generally accepted in 1777 (Buffon). Many astronomers of the 19th century returned to the Michell problem and remarked on the difficulty of its proper solution. **Forbes** (1849 and 1850) noted that an assumption of a prior distribution was doubtful and that the uniform law was “more inconsistent with a total absence of Law or Principle” than the existence of “condensation” and “paucity” of stars. By applying the Poisson law **Newcomb** (1860) calculated the probability that some surface with a diameter of 1° contained s stars out of N scattered “at random”. In 1904 he reasoned (as **Boole** did in 1851) about the difference between a “chance” and a uniform distribution. **F.G.W. Struve**, who essentially contributed to the study of double stars, applied stochastic reasoning for a preliminary proof that a connection of two stars was physical rather than purely optical. In 1827 he calculated the probability that two or three stars were close one to another.

A related problem on the distance between two random points on a sphere dates back to **Laplace** (1812), **Cournot** (1843) and **Newcomb** (1861). Each of these scholars formulated it in his own way. Without referring to anyone, **Bertrand** (1888) solved it just as Laplace and Cournot did. Among other problems, Bertrand used it for proving that the term “randomly” should be defined more precisely (cf. §3.3). He called the Michell problem insufficiently determined.

5.3. William Herschel: the Sidereal System {6; 3}

From about 1784 Herschel began to count the number of stars seen in the field of view of his telescope in various regions of the sky. Supposing that the stars were distributed uniformly and that his telescope penetrated to the boundaries of the finite (!) starry system, he thus attempted to determine the relative distances to those boundaries. He did not apply sampling, but for him the stars were elements of a single population {3}.

Later Herschel abandoned these premises, and, in 1817, he introduced a model of the distribution of the stars according to their distances. He placed the stars of each given magnitude i , $i = 1, 2, \dots, 7$, between two appropriate concentric spheres ($i + 1$ and i) but inside the rings thus obtained the stars were allowed to be randomly distributed, a feature also present in **Struve's** study (§5.5). Herschel calculated the differences between the actual number of stars of each of the seven magnitudes and the number corresponding to his model. The sum of these differences for the first four magnitudes was small and he concluded that his model provided for them a fair approximation. Nevertheless, the individual discrepancies were too large so that his opinion was hardly warranted.

Herschel's criterion (small value of the sum of deviations), which he also applied indirectly when determining the direction of the sun's motion through the proper motion of stars (1805),¹¹ resembled the main condition of the Boscovich method of solving redundant systems of linear equations (1770). When determining the velocity of the sun's motion (1806), Herschel had to choose between the arithmetic mean and the median. Apparently following Laplace's early considerations, he decided in favor of the median.

5.4. Herschel: an Instructive Mistake

In 1817 Herschel indicated that the size of a star, “promiscuously chosen” out of the 14 thousand stars of the seven first magnitudes, “is not likely to differ much from a certain mean size of them all”. Since then, it became known that the sizes of the stars differ enormously so that the notion of their mean size is nonsensical.¹²

Herschel apparently based his reasoning on a heuristic idea which became formalized (and specified in a way unforeseen by him) and is now known as the Bienaymé – Chebyshev inequality. His mistake illustrates the fact that, in order to establish some proposition, the theory of probability, just like any other scientific discipline, has to issue from definite data, and that it is powerless otherwise.

5.5. F.G.W. Struve

In 1847 he statistically studied the linear and angular (reckoned from a certain plane) distances of the stars {3}. He did not indicate that his findings were of a statistical nature; true, at that time empirical distributions in natural sciences were hardly mentioned. He also provided a formula with statistically determined parameters for calculating the maximal distances of stars of a given magnitude. Finally, drawing on statistical data, he argued that the interstellar space absorbed light. His proof was based on essential assumptions, but the existence of that phenomenon was subsequently confirmed.

5.6. The Proper Motion of Stars {3}

The study of these motions for hundreds of stars (for the time being, only in the directions perpendicular to the appropriate lines of sight) began in the 1830s – 1840s. The results obtained (**Argelander**, 1837) enabled to determine the direction of the sun's motion much more reliably. In 1842 **O. Struve** calculated the mean proper motion for the stars of each of the first seven magnitudes, and in 1852 **F.G.W. Struve** analyzed an even more extensive material. The middle of the 19th century can be considered as the beginning of stellar statistics {4}; **Herschel's** merits in that field are unquestionable but his statistical calculations did not lead to any real results.

When studying the sun's motion, astronomers, beginning with Herschel, assumed that the peculiar motions of the stars were random. In 1902 **Newcomb** assumed that the projections of the stellar motions on an arbitrary axis were distributed normally and determined the distribution of the motions themselves and of their projections on an arbitrary plane. Both of these occurred to be connected with the chi-squared distribution.

Already in 1843 **Cournot** declared that "la statistique des astres [...] doit servir un jour de modèle à toutes les autres statistiques", but only 40 years later **Hill & Elkin** (1884) stated that a general statistical study of stellar populations was more important than a precise determination of the parameters of some star.

5.7. Is a Statistical Analysis Needed? {6}

A large number of astronomical catalogs and yearbooks as well as star charts had been published in the 19th century. Their compilation may be attributed to the tabular direction of statistics, which was contrasted with theoretical studies. Thus, **Proctor** (1873) compiled charts of 324 thousand stars and claimed that he did not need any theories on the structure of the stellar system. He did not find followers and the subsequent development of astronomy (§5.8) refuted his opinion.

5.8. A Statistical Description of the Stellar System {3}

Kapteyn (1906 and 1908) vividly depicted the stellar universe describing it by means of the laws of distribution of parallaxes and proper motions of the stars. The (almost) determinate models of the starry heaven (**Herschel**, **F.G.W. Struve**), as well as the mean proper motions of stars of a given magnitude, were since forgotten {7}. **Kapteyn** (1906) also initiated a sample study of the stellar universe and until our time the characteristics of faint stars are being determined in some areas uniformly distributed over the sky and additionally at places of special interest, i.e., by stratified sampling. **Kapteyn** did not mention sample surveys of population that had come into practice at the turn of the 19th century.

5.9. An Astronomical Version of Correlation {10}

Kapteyn (1912) was not satisfied with the Galton definition of the correlation coefficient and offered another one that enabled to estimate the connection between the errors of two functions some of whose measured arguments were common. He had not referred to Gauss although his proposal was in the spirit of the latter's thoughts. **Kapteyn's** innovation could have been applied in geodesy if not astronomy, but it remained unnoticed.

5.10. Karl Pearson

In 1905 – 1910 he published (partly as coauthor) several papers on the application of statistics to stellar astronomy. He was mostly interested in the relations between the parameters of the stars. He applied non-Gaussian frequency curves and noted that statistics was able to indicate the directions for further astronomical research but his attempts remained unsupported partly because he was insufficiently acquainted with the astronomical literature.

6. Physics

6.1. Introduction

Already **Daniel Bernoulli** qualitatively justified the elements of a kinetic theory by statistical considerations. In essence, that discipline originated in the mid-19th century after the statistical method had penetrated physics {4}. True, the kinetic theory had then been developing on a low stochastic level and mechanical discourses occurred side by side with probability-theoretic arguments (**Khinchin**, 1943), which was inevitable because probability theory "was still in an infantine state" {1}.

6.2. Clausius

In 1857 he introduced the notion of mean velocity of molecules and in 1858 he went on to study the distribution of their free paths. Denote the random free path by ξ , then, in modern notation, **Clausius** calculated $E \xi$ and determined $[1 - F(s)]$ where $F(s)$ was the distribution function of the free paths. **Poisson** (1832 and 1837) was the first to introduce distribution functions and **Davidov** (1884 – 1885) followed him, but these functions had not been essentially applied until the 20th century so that **Clausius'** merit is here obvious. Moreover, the function $F(s)$ occurred to be infinitely divisible {11}. **Clausius** had also applied stochastic

considerations for solving concrete problems and it was largely under his influence that **Maxwell** turned his attention to probability. **Gibbs** (1869) contrasted Clausius to Maxwell and **Boltzmann** and stated that Clausius was concerned with mean values of random variables rather than with distributions. This is not altogether true, see above, but the general picture of proceeding from means to distributions is correct {7}. It would have been possible to contrast, in the same way, Daniel Bernoulli to the three later scholars.

6.3. Maxwell

In 1860 he established his celebrated distribution of the velocities of monatomic molecules and effectively applied the chi-squared distribution for three ($n = 3$) degrees of freedom.¹³ His derivation assumed that the components of the velocity were independent. It was repeatedly criticized and Kac (1939) and Linnik (1952) weakened the Maxwellian conditions.

In 1879 Maxwell introduced fictitious physical systems and became able to consider the probability of a system being in a certain phase. Actually both he and **Boltzmann** (§6.4) made use of an infinite general population (indirectly introduced by Laplace). Maxwell left many pronouncements on the statistical method. Physicists, he maintained, were “compelled” to adopt it (1871); it “opened up new views of nature” and is quite sufficient “for all practical purposes” (1873). Without directly mentioning randomness, he (1873) actually connected it with an unstable behavior of physical systems {12}. In 1875 he noted that the motion of a molecule could not be predicted and thus abandoned Laplace’s opposite statement, but he did not add that the statistical populations of molecules should therefore be studied {3}. Prophetically declaring that in future physicists will possibly study “singularities and instabilities”, he illustrated this opinion by the randomness connected with refraction in biaxial crystals (1873).

6.4. Boltzmann

With respect to separate molecules Boltzmann offered two definitions of posterior probability: 1) If m molecules out of M possess property A , then the probability that some molecule possesses it is m/M ; and 2) If, during time period T , some molecule possesses property A for time t , the probability that it possesses A is t/T . Note that the second definition is based on geometric probability. In 1872 Boltzmann declared that the two definitions were equivalent; with respect to gas as a whole this statement is nowadays formulated as the ergodic hypothesis.

When studying the distribution of the kinetic energy of a gas among its separate molecules, Boltzmann (1877, 1878) employed the “classical” definition of probability. Finally, in 1899 he introduced fictitious systems, and, accordingly, the phase probability (cf. §6.3). In 1871 Boltzmann defined the probability of the state of polyatomic gas by the product such as $f d\omega$ where f was some function (varying in time) of the coordinates and velocities of the separate molecules and $d\omega$ was the product of the differentials of these parameters. Boltzmann’s arguments were mostly concerned with the case of invariable f , but he did not reason them out: he understood f as the number of molecules whose parameters satisfied certain restraints, and, at the same time, as the distribution of the states of motion among the molecules. When examining stochastic processes such functions determine the distribution of a system of random variables at the appropriate moment t . Boltzmann’s writings are verbose, sometimes poorly organized and make difficult reading.¹⁴ Khinchin’s negative opinion (§6.1) about the early kinetic theory was perhaps partly caused by that circumstance.

Already in 1871 Boltzmann began to connect the proof of the second law of thermodynamics with stochastic considerations; his fundamental relation between the probability of the state of a system with its entropy is generally known. However, he could have illustrated the stochastic meaning of the second law by the **Daniel Bernoulli – Laplace** urn problem, independently devised, in 1907, by the **Ehrenfests** {1}. In 1886 Boltzmann declared that the 19th century will be called “das Jahrhundert der mechanischen Naturauffassung, das Jahrhundert Darwins” and in 1904 he stated that the evolution theory was of a mechanical nature. Similarly, Boltzmann (1902) argued that it would perhaps become possible to explain electricity and heat by “verbogene mechanische Bewegung” whereas entropy and irreversibility will possibly be reduced

Durch Anwendung der Wahrscheinlichkeitsrechnung auf das Verhalten sehr zahlreicher materieller Punkte.

He also noted that the mechanical picture of the world could not yet be abandoned, owning, however (obviously bearing in mind **Gibbs’** achievements) that it was changing. Thus, although Boltzmann began to re-view somewhat his standpoint, he was not prepared to leave mechanical philosophy behind, possibly because he, unlike Maxwell (§6.3), did not perceive the existence of objective randomness.

Even in 1872 Boltzmann understood that the main difficulty in applying the theory of probability was in expediently defining probability, and he unintentionally rejected the Laplacean explanation of that notion as incomplete knowledge. **Zermelo** (1900) agreed with Boltzmann in this respect, and, also in 1900, **Hilbert**

suggested that the probability theory should be axiomatized (his Problem No. 6), but he did not mention physics in this connection.

Notes

1. Some statistical data compiled in the 19th century led to conclusions now either forgotten or rejected (§§2.4 and 4.2). It would perhaps be opportune, at least in the methodological sense, to treat the available material anew.

2. Thus, he thought that the differences in the physicians' skill led only to a "kaum bemerkbares Schwanken" in the overall results of treating a given surgical disease.

3. A more detailed description of public hygiene should have included a study of mortality in the army and prisons (data on morbidity were then fragmentary). I adduce now Farr's statement (ca. 1857) which testified to the importance of statistics in this field):

Any deaths in a people exceeding 17 in 1,000 annually are unnatural deaths. If the people were shot, drowned, burnt, poisoned by strychnine, their deaths would not be more unnatural than the deaths wrought clandestinely in excess of [...] 17 deaths in 1,000 living.

4. Stokes had indeed solved it, apparently by applying the Poisson distribution.

5. Without providing a rigorous analysis, Galton had confirmed that the former was more advantageous, and Fisher corroborated that opinion.

6. Ruse (1971) studied the first of those concepts with the express purpose of clarifying it.

7. He could have mentioned Laplace's appropriate general statement:

L'action des causes régulières et constantes doit l'emporter à la longue sur celle des causes irrégulières.

8. Males and females can be considered separately.

9. Kupffer thought about that in 1829 whereas Humboldt had even earlier considered it desirable [7, pp. 94 – 95].

10. In 1863 Galton proposed a system of symbols for weather maps and was immediately rewarded for his efforts by ascertaining the existence of anticyclones.

11. Such determination lead to the necessity of correcting the proper motions. This is an example of a mild settlement of the contradiction between astronomy and statistics {2}.

12. Mean proper motion of stars of a given magnitude became another meaningless concept (§5.6).

13. Boltzmann used the same law for $n = 2$ (1877), $n = 3$ (1873) and in the general case (1881).

14. In 1873 Maxwell complained that Boltzmann's "length" was "a stumbling block" for him, and, according to Klein (1973), after about 1870 Maxwell "apparently never read" any of Boltzmann's papers. In turn, Boltzmann, even in 1868, remarked that one of Maxwell's deductions "wegen ihre grossen Kürze schwer verständlich ist".

References

1. – 5. Sheynin, O.B. Five papers on the history of the statistical method in various branches of natural sciences (biology, medicine, astronomy, meteorology, and physics). AHES, vols. 22, 26, 29, 31 and 33, 1980, 1982, 1984 and 1985.

6. Sheynin, O.B. Statistics. In: *Enc. of Stat. Sciences, Update vol. 3*. Editors, S. Kotz et al. New York, 1999, pp. 704 – 711.

7. *Переписка А. Гумбольдта с учеными* (Humboldt's Correspondence with Russian Scientists and State Figures). M., 1962.

14. Markov's Report on a Paper of Galitzin

IMI, vol. 32/33, 1990, pp. 451 – 467

1. Debates among Academicians [13]

In 1902, Prince Boris Borisovich Galitzin (1862 – 1916),¹ a Corresponding Member of the Imperial (Petersburg) Academy of Sciences, established that a center of an earthquake can be determined from data obtained at one seismological station and thus solved an important problem of seismology. Together with his other scientific merits, this fact prompted seven academicians, including the astronomer F.A. Bredikhin, to nominate Galitzin for effective membership at the Academy. The ensuing discussion was published in 1903 [13]; this source shows that Markov, and then Liapunov, negatively spoke about some of Galitzin's investigations. The proposal of the seven academicians was not carried. Liapunov criticized one of Galitzin's works belonging essentially to applied mechanics, whereas Markov destructively reported on the latter's paper [2], and, not being content with the debates, expounded his views in a special article which remained however unpublished until now.

Galitzin [2] treated some theoretical material but his main goal was to publish and adjust the results of some of his experiments. His article [2] did not by any means belong to the central sphere of Galitzin's scientific interests and some other scholar could have well agreed with those nominating him (but perhaps admonished the author for his lame publication). Markov, however, owing to his peculiar straightforward nature, was unable to do so. Indeed, mathematical treatment of observations was (and still is) an important step in the work of natural scientists, and the appearance of a weak paper on that subject in a periodical of the Academy could have additionally annoyed Markov, the more so since this happened just before the nomination of its author.

Markov [13, p. 5] justly declared that

Having no scientific importance, the article [...] [2] hardly offers anything essential for practical purposes because of the large disagreement between the results of the observations reported there.

He added that Galitzin's seismological achievements had not yet stood the test of time.

Neither Galitzin, nor Bredikhin, who supported him, were able to object meaningfully to Markov. They stated that large errors were inevitable when investigating the solidity of substances, but that argument was ill-founded since Galitzin had apparently put forward precise numerical conclusions. One fact additionally characterizes his article. Markov noted that his graph did not agree with his own tabular values. Galitzin argued that the graph was nothing but a sketch; Markov, however, refused to accept that strange statement (and repeated his opinion in the paper here published). Indeed, Galitzin did not say so in his article, and his "sketch" showed the scales adopted for each of the coordinate axis.

2. Markov's Paper

The *Izvestia* of the Imperial Academy of Sciences (ser. 5, vol. 18 for 1903, p. xix) announced that Markov had presented his paper "On the solidity of glass" and that it will indeed be published there. Nevertheless, it did not appear either there, or, as it seems after studying general bibliographic sources, in any other periodical. However, the manuscript of the paper was being kept in Markov's family, and several years ago Andrei Markov, Junior, had given it for some time to Grodzensky who mentioned it in his book [3, p. 65], turned it over to the Archive of the Soviet Academy of Sciences and informed me about its existence.

I say *manuscript*; however, Markov's work was set up in a printing office in a format used by the *Izvestia* (in particular, his name was given in an oblique case). The first page of the *manuscript* carries a typographic stamp "Printing office of the Imp. Academy. 3 [i.e., third] proof. Sent 9.IX.903. Returned ..." with the number 3 and the date written down by hand. It follows that the *manuscript* is a proof (corrupted by a large number of insignificant misprints which I have corrected without any special notice).

It may be supposed that Markov did not return that proof; in other words, that he suppressed his work. A natural reason for this could have been the hardly suitable form of his paper (not an original contribution but a review; and a detailed review at that, difficult to read without having Galitzin's article [2] at hand). Another possibility is that the *Debates* [13] were (perhaps suddenly) to be published, or had already appeared. And, finally, it cannot be ruled out that by the end of 1903 Galitzin had a change of heart and at least essentially agreed with Markov's criticisms.

3. Mathematical Treatment of Observations

Reviewing the Galitzin's article demanded, in particular, accurate and punctilious calculations and Markov definitely had to spend much time on them. He was always interested in treating observations. In that area he

is mostly known as a staunch supporter of the then almost forgotten Gaussian definitive justification of the method of least squares by the principle of maximal weight (minimal variance). He included a chapter devoted to that method in his celebrated treatise [6] and combined it with investigation of statistical series, interpolation and the Lexian theory of dispersion. Such an approach is controversial, especially from the standpoint of methodology, but at least it reflected an attempt (then proved unsuccessful) to include directly the method of least squares into theoretical statistics.

Markov paid due attention to estimation of the plausibility of observations. Here is his hardly known pronouncement [...]²

4. The Theory of Correlation

Markov remarked that the main quantity interesting Galitzin depended not only on many variables, but in addition on unknown variable circumstances. At present, that correct indication would have meant that Galitzin's results should have been adjusted in accordance with correlation theory that by that time (1903) had only taken its first steps. I ought to say that Markov did not recognize it either then, or even later. Here is his pertinent opinion [8, p. 200]: The positive side of the “fashionable” correlation theory

Is not significant enough and consists in a simple usage of the method of least squares in order to discover linear dependences. However, not being satisfied with approximately determining various coefficients, it also indicates their probable errors, and enters here the realm of imagination, hypnotism and mathematical formulas that actually have no sound scientific foundation. [...]

5. Some Subsequent (after 1903) Events

Galitzin naturally continued his scientific work and in 1908 he was indeed elected effective member of the Academy. This time, without recalling his previous misgivings, Markov benevolently mentioned him.³ In 1911 Galitzin was elected President of the International Seismological Association, and, in 1916, Fellow of the Royal Society. He is justly considered as cofounder of modern seismology.

Galitzin's relations with Markov remained (or at least once more became) normal. Indeed, in 1913, in a letter to Chuprov, Markov [12, p. 70] wrote:

*I have not dwelt on the question of the importance of the law of large numbers to physics. It would be appropriate to talk about this with Prince Galitzin.*⁴

Nevertheless, in 1916 Markov [8] again, although this time indirectly, came out against Galitzin: he expressed his negative opinion about correlation theory (§4) while criticizing a paper [16] that appeared in a publication of the Main Physical (now, Geophysical) Observatory, and Galitzin was then both the Director of the Observatory and the Editor of the pertinent periodical. Markov [8] reported at the Academy during Galitzin's lifetime, but it was published after the latter's death.

On the Solidity of Glass

A.A. Markov

Many physicists will perhaps be surprised that a mathematician who did not carry out experiments brings himself to speak about the solidity of glass. I hope, however, that their surprise will blow over when they find out that the matter under discussion is the evaluation of the conclusions made on the basis of experimental data by the method of mean numbers.⁵ The data and the conclusions on them are provided in [2].⁶

Without dwelling on whether it is necessary, for scientific or practical purposes, to search for a new formula expressing the limiting pressure sustained by a glass tube instead of the Neumann formula [11, p. 145]

$$400 (n^2 - 1) / (n^2 + 1);$$

without discussing whether the theoretical part of the article should have been devoted to describing the conclusions made in a certain manual [1] concerning an earlier solved problem [5]⁷; and, finally, without touching upon the issue on the extent of the agreement between the circumstances of the experiments and the assumptions made in the theoretical part, I shall appraise the conclusions reached in this article.

First of all, however, I ought to raise a number of questions about the data there provided. Table 1 gives the values of eight magnitudes

$$V_m, V, R, R', d, n, P_m, T_m$$

three of which (d, n, T_m) are not furnished by direct observations but determined from their connections with R, R', P_m .⁸ In order to estimate the errors of n and T_m under such circumstances it was necessary to estimate beforehand the possible errors of these last-mentioned magnitudes, which is not even hinted at in [2]. From a mathematical point of view, it cannot be denied that, when P_m is given only to two significant digits, the calculation of T_m to three digits is a mistake.

It would have been very simple to estimate the influence of the error of P_m on T_m by means of Table 1, at least for such cases in which, for one and the same n , we find a number of values of T_m corresponding to different values of P_m provided that the values of T_m were calculated quite correctly in accord with the formulas accepted by the author and issuing from the values n and P_m indicated by him. We are obliged, however, to point out the opposite.

Formulas (25) and (26)⁹ show that for a constant n the increment of T_m should be proportional to the increment of P_m , whereas in Table 1 we find for $n = 1.12$

$T_m = 3.22$	at $P_m = 40$	(No. 49) ¹⁰
3.50		42 (NNo. 51 and 54)
3.55		44 (NNo. 48 and 50)

Thus, when P_m is increased by 2, T_m increases at first by 0.28, then by 0.05. In the same table, for $n = 1.15$ we have

$T_m = 3.44$	at $P_m = 52$	(No. 57)
3.76		55 (No. 58)
3.86		58 (No. 61)

Here, when P_m is increased by 3, the increment of T_m is at first 0.32, then only 0.10.

It certainly would not have been difficult to estimate the influence of the errors of R and R' on n in accord with the formula $n = R/R'$; however, it is not worthwhile to dwell on this until the problem about the errors of R and R' is ascertained. The importance of these errors increases when we turn over to the magnitude T_m which is calculated from n and P_m . In Table 1 for $P_m = 42$ we find

$T_m = 3.28$	at $n = 1.13$	(No. 36)
3.50		1.12 (No. 54)
3.90		1.10 (No. 86) ¹¹

This clearly shows that the errors in the first digit of n influence the second digit of T_m .

Under these conditions the problem about the errors of R and R' becomes very important, if, like the author of [2], we shall not restrict our attention to the first approximation, i.e., to determining the mean value of T_m over all the observations, but shall rather try to determine some dependence between T_m and n expressed by empirical formulas. Otherwise all conclusions might be corrupted by constant errors different for the inner and the outer sections of the tube since the influence of these errors on n and T_m varies with R and R' .

Concerning R and R' it is necessary to take into account that they do not persist for different parts of one and the same tube; thus, $R = 4.37$ for No. 48, and 4.50 for No. 51 whereas both these numbers, as shown by the bracket that combines them,¹² denote parts of the same tube. This fact can be caused to some extent by errors of measurement, and, in part, by the form of the tube deviating from a right circular cylinder. However, there are no indications in [2] that the magnitudes R and R' were measured repeatedly, and in various directions. Furthermore, only the arithmetic means of both R and R' are given rather than their values at the end of the tube.

One of the conclusions in [2] is that T_m does not depend on the rapidity of the increase in pressure [...]. This inference, however, cannot be considered as well-founded. The magnitude T_m defined by equality (25) and multiplied by a constant factor¹³ represents a function not only of many variables, but, in addition, of unknown variable circumstances, and these latter, as it may be presumed, play a very important part.

Under such conditions T_m cannot be considered as a function of one variable, be it V_m or n or some other magnitude. At the same time, it is very difficult to decide whether T_m depends on some quantity. In such cases only the method of mean magnitudes can offer some indication which will however be fairly uncertain, especially if the experiments are not methodical but rather of a random nature, and the variables on which the studied quantity can depend are changing all at once.¹⁴

In order to prove that T_m does not depend on the rapidity of the increase in pressure the author of [2] indicates that identical values of T_m were obtained at different rapidities and that larger values of the rapidity were sometimes accompanied by lesser, and, sometimes, quite to the contrary, by greater values of T_m . At

the same time, the author considers it possible to allege that T_m depends on n , although, taking as an example the Thüringen glass, Table 1 shows that for very different values of n there occur values of T_m equal or near to 4.21,¹⁵ and that when n increases the magnitude T_m increases in some cases, and, quite to the contrary, decreases in other instances (and, moreover, in complete disagreement with either Table 21 or Table 23).¹⁶ It was thus possible to conclude that T_m does not depend on n , and, admitting (as is done in [2]) deviations of 50%,¹⁷ to obtain $T_m = 5.4$ or 5.5 . Thus, it was apparently possible to consider that, in the first approximation, in accord with the theoretical assumption,¹⁸ T_m was constant.

Had the author restricted his attention to this first approximation, and omitted the conclusion of the manual [1], he would have compiled a short note perhaps suitable for some technical periodical and I would have left such a note without notice. But the author had not stopped there, he believed it possible to state that T_m depended on n , and even attempted to represent that dependence by a table, or a graph, and I consider it necessary to indicate that we have the same or even more right to state that T_m depends on V_m , the mean rapidity of the increase in pressure.

To this end, I have compiled the following table¹⁹ from which

N	V_m	T_m	N	V_m	T_m	N	V_m	T_m
89	1.0	4.21	21	2.2	8.03	75	2.9	7.04
73	1.1	4.78	71	2.2	7.03	94	3.3	5.50
84	1.2	4.57	91	2.2	5.87	96	7.2	6.29
86	1.3	3.90	95	2.2	3.70	22	7.3	8.03
79	1.4	6.73	93	2.5	5.14	83	7.4	6.28
19	1.4	7.07	81	2.7	4.23	80	8.8	7.79
92	1.5	5.56	99	2.7	6.67	74	15.8	7.83
97	1.5	3.57	82	2.8	6.79			
			85	2.8	6.08			
Sum		40.39			53.54			48.76
Mean		5.05			5.95			6.97

it is possible to conclude that for the Thüringen glass T_m increases with V_m . I have borrowed all the numbers from Table 1 [2] concerning that glass excepting those which either characterize the thick-walled tubes (NNo 12 – 18) or are not accompanied by indications of V_m . I have omitted the thick-walled tubes because the author himself thought it necessary to reject three out of his seven pertinent results.

Going on to compare this table with Tables 20 and 21 [2] which serves its author for conclusions about the dependence of T_m on n , I note first of all that a large part of the numbers in these tables are arithmetic means of three, four or five numbers. Then, in order to ascertain the matter, we shall estimate, in accord with the usual method, the possible errors of the difference (6.97 – 5.05) in my table and of some differences in [2] which serves as the basis for the conclusions there made.

Simple calculations²⁰ provide a magnitude less than 0.21 for the square of the mean error of 5.05, which, in my table, is T_m for the Thüringen glass if $V_m < 1.5$. For 6.97, which is T_m if $V_m \geq 2.9$, the same magnitude is less than 0.14. Therefore, applying the known formula²¹, we find that the mean error of the difference (6.97 – 5.05) = 1.92 is less than $(0.21 + 0.14)^{1/2} < 0.6$ and does not amount to 1/3 of the derived number 1.92.²²

If, however, first considering the Jena glass, we take the difference (6.53 – 5.50) = 1.03 of the two values of T_m in Table 20 corresponding to $n = 1.46$ and 1.33, we shall find by means of just as simple calculations that the mean error of this difference is greater than $\sqrt{0.38} > 0.6$ and is therefore greater than one half of the obtained number 1.03. Yet the author brings himself to say (p. 23) that *Betrachten wir nun näher die Zahlen der Tabelle 20, so sehen wir, dass für Iennäer Glas mit wachsendem n innerhalb der Beobachtungsdata T_m stetig wächst und zwar ist die Abhängigkeit fast eine lineare.*

Finally, for the Thüringen glass (Table 21) we have for the difference (6.84 – 5.16) = 1.68 of the values of T_m for $n = 1.36$ and 1.18 the mean error greater than $(0.81 + 0.13)^{1/2} > 0.95$; and for the difference (6.84 – 5.59) = 1.25 of the values of T_m for $n = 1.36$ and 1.51 the mean error is greater than $(0.81 + 0.25)^{1/2} > 1$. Nevertheless, the author ventures to say that

Für Thüringer Glas wächst am Anfang T_m mit n bis zu einem gewisser Maximum (etwa bei $n = 1.36$) um dann allmählich abzunehmen.

These conclusions can by no means be admitted as well-founded either for the Jena, or still less for the Thüringen glass, and not only because of the considerable magnitude of the mean error which indicates a

large disagreement between the main numbers, but also in virtue of the reasons explained above. So as to eliminate misunderstanding, I remind readers that [2] provides no information for deciding whether T_m depends only on $R/R' = n$ rather than on both R and R' separately.

In concluding, I believe it worth noting a number of facts characterizing [2]. For the Jena glass it offers a formula

$$T_m = A + Bn,$$

and it furnishes even two formulas for the Thüringen glass,

$$T_m = A + Bn + Cn^2,$$

$$T_m = A + B/n + C/n^2 + D/n^3 \quad (1)$$

which are alleged to serve for differing values of n . The coefficients of these formulas are however lacking. Not knowing them, we naturally cannot ascertain to what extent do these formulas agree with Tables 20 and 21 on which they were based; or, how much are Tables 22 and 23 and the curves on the appended figure concordant with the formulas. We can only note that, for the Thüringen glass, the number of measurements recorded in Table 21 is equal to the total number of the coefficients of the empirical formulas, but that in spite of such an abundance of coefficients, some numbers in Table 23 essentially deviate from the corresponding numbers in Table 21: for $n = 1.10$, $T_m = 3.64$ (Table 21) and 3.9 (Table 23); for $n = 1.18$, $T_m = 5.16$ and between 4.5 and 5 respectively.

And, when comparing Tables 21 and 23 with the appended figure, I found out that the line for the Thüringen glass on the figure does not agree with either table: for $n = 1.20$, Table 23 provides $T_m = 5.0$, whereas, according to the figure, T_m is essentially greater and approximately equal to 5.5; for $n = 1.26$, Table 21 furnishes $T_m = 5.44$, but, according to the figure, it is about 6. From among the numbers in Table 23 more than a half are in the interval from $n = 1.55$ to 2.30 for which there are no measurements. Finally, the author's attitude towards the empirical formulas and tables compiled by him can be judged by the calculation of the maximal pressure sustained by a thick-walled tube ($n = \infty$). In this calculation (p. 28), T_m is taken to be 8.03 although such a number is lacking both in Tables 22 and 23 and, at $n = 23$, the latter gives $T_m = 4.4$.²³ We certainly do not know what value of T_m at $n = \infty$ is provided by formula (1), but we cannot forget the author's statement on p. 29 that

Bei wachsendem n wächst am Anfang auch die Festigkeit T_m (Jenaer und Thüringer Glas) um später bei fortgesetztem Wachsen von n mehr oder weniger langsam abzunehmen (Thüringer Glas).

On p. 28 the author assumes, as an example, that $n = 10$. I do not know to what extent are the conclusions from the experiments with tubes, for which the values of n are several times less than 10, applicable to such thick-walled tubes, but I think that in any case it would have been more interesting to choose for an example such a value of n as occurred in the author's experiments. At $T_m = 8.03$ and $n = 2.30$ the formula (26) gives $P_m \approx 469$; in Table 1 (No. 13) we find $P_m = 448$ which is somewhat lesser, but it concerns a tube that was eliminated from consideration, whereas Table 25 provides, for $n = 2.3$, $P_m = 258$.

The final judgement about [2] should naturally be made by physicists.

Notes

1. The Russian spelling of his name was Golitzin.
2. Together with my commentary, it is now included elsewhere [15, p. 351].
3. Archive of the Soviet Academy of Sciences, Fond 1, inventory 1a – 1908, delo 155, pp. 118 – 118 reverse. I am indebted to Dr. Natalie Ermolaeva for this information.
4. Markov discussed his future report [7] concerning Jakob Bernoulli's *Ars Conjectandi*. Nothing is known about his (possible) conversation with Galitzin and in any case he [7] did not say anything about physics.
5. Markov called the method of means "the method of mean numbers" (or, below, "the method of mean magnitudes"). It was Condorcet, who, in a posthumous publication of 1805, introduced the term "theory of means" but he did not indicate its connection with probability theory. Not later than in 1830 the term became established. Later on Quetelet applied it and Humboldt mentioned the "einzig entscheidende Methode, die der Mittelzahlen". In 1857 Davidov published a work entitled *Theory of Mean Magnitudes* (in Russian), and even in 1901 Hilbert, in his celebrated report on the problems of mathematics, maintained that, hand in hand with axiomatization of probability,

eine strenge und befriedigende Entwicklung der Methode der mittleren Werte in der mathematischen Physik, speciell in der kinetischen Gastheorie,

should go on.

The term “Theorie der Fehler” was introduced in 1765 by Lambert, but neither Laplace nor Gauss ever used it. True, Bessel had several times applied it, and in 1845 – 1861 it became current. For several decades both terms had been existing on a par although the theory of means considered any means (e.g., the mean stature of male adults) whereas the theory of errors concentrated on observations of constants. Davidov was the first to stress the common character of treating any observations. See [14, pp. 310 – 312].

6. Markov never mentioned the author of [2] by name.

7. Galitzin referred to [1]. Markov’s remark was far from self-evident. Indeed, Galitzin did not consider the history of the subject, and in general his article was mainly concerned with treating observations rather than with theory, also see §1 of my general commentary above. Incidentally, Markov himself (below) referred in similar circumstances to contemporaneous authors rather than to Gauss. I also note that Mendeleev [9; 10, pp. 481 – 490] investigated the solidity of tubes but did not explain his method of treating the pertinent measurements. He determined the resistance of tubes against rupturing under pressure from within.

8. In Table 1 Galitzin compiled the results of investigating the solidity of glass under pressure from within. He had experimented on 99 tubes of different kinds of glass, mainly of the Jena and Thüringen varieties. Markov mentioned the following magnitudes: V_m , mean velocity of the increase in pressure just before rupture of tube; R and R' , outer and inner radii of a tube; $d = R - R'$; $n = R/R'$; P_m , maximal pressure sustained by a tube; and T_m , measure of solidity of glass.

9. Formula (25) theoretically deduced by Galitzin was

$$T_m = (1/4) (5P_m + 7 \{ [(P_m - 1) / (n^2 - 1)] - 1 \}).$$

His formula (26) derived quite simply from (25) determined P_m through T_m ; however, Galitzin had multiplied the latter magnitude by a constant factor thus allowing for transition from one unit of measurement to another one.

10. Markov naturally kept to Galitzin’s numbering, and, for example, No. 49 meant experiment No. 49.

11. Experiments NNo. 36 and 54 concerned Jena glass, and experiment No. 86 was done with Thüringen glass.

12. In his Table 1 Galitzin used brackets to denote experiments which concerned one and the same tube cut into parts.

13. See Note 9.

14. In those times natural scientists tried their best to exclude the possibility of several arguments (factors) changing all at once in their experiments. The theory of experimental design originated by Fisher in the 1920s – 1930s, being based on a more effective statistical treatment of observations, was able to abandon that restriction.

15. Such for example were experiments NNo. 18, 20, 81 and 89.

16. Galitzin’s article carried 25 tables. I explained Table 1 in Note 8. Each of the Tables 2 – 19 is an extract from Table 1 and is concerned with one single kind of glass, and, furthermore (with a single exception not important for me), with a small interval of values of n . Each of them shows n as well as d and T_m . Tables 20 and 21 are summaries for the Jena and Thüringen glass respectively. Here, for example, is Table 20. The three numbers in the first line show the weighted mean values of Tables 5 and 6 which were concerned with almost

the same value of n ($n = 1.12$ and 1.13); just the same, the left part of the second line is a combination of Tables 7 and 8; and the other lines are connected in the same way with Tables

n	T_m	d	G
1.12	3.67	0.55	12
1.33	5.50	1.04	7
1.36	5.53	0.80	4
1.39	5.96	1.25	4
1.46	6.53	1.36	4

9 – 11. Thus, the last line repeats the mean values previously calculated for Table 11. The last column shows the number of experiments (12 is the total number of them represented in Tables 5 and 6 taken together). Tables 22 and 23 provide the results of Galitzin’s calculations of T_m by his empirical formulas for the two mentioned above kinds of glass respectively, see below. Finally, in Tables 24 and 25 he summarized the magnitudes P_m calculated for given values of n and for values of T_m taken from Tables 22 – 23.

17. Markov bears in mind the relative error of T_m .

18. It is unclear which assumption is mentioned here since one of Galitzin’s formulas connected T_m with n , see Note 9.

19. In this table, N denoted the number of the experiment.

20. Here and below Markov calculated the square of the mean square error of the arithmetic mean. He referred to [4, p. 244] and [17, Chapters 5 – 7].

21. Here and below Markov bore in mind the formula for the mean square error of a sum of several independent terms.

22. Markov could have mentioned here his test of significance, see §3 of my commentary.

23. Number 8.03 is the maximal value of T_m in Table 1. The tube with the thickest wall in Table 23 has $n = 2.35$ and the corresponding value of T_m is 4.4.

References

1. Evnevich, I.A. *Руководство к изучению ...* (Manual on Studying the Laws of Resistance of Building Materials). Psb, 1868.
2. Galitzin, B., Fürst, Über die Festigkeit des Glasses. *Izvestia Imp. Akad. Nauk*, ser. 5, vol. 16, 1902, pp. 1 – 29 + Plate (graph) facing p. 52.
3. Grodzensky, S.Ya. *Марков* (Markov). М., 1987.
4. Khvolson, O.D. *Курс физики* (Course in Physics), vol. 1. Psb, 1897.
5. Lamé, G. *Leçons sur la théorie mathématique de l'élasticité des corps solides*. Paris, 1852, 1866.
6. Markov, A.A. *Исчисление вероятностей* (Calculus of Probability). Psb, 1900, 1908, 1913; М., 1924.
7. --- The bicentennial of the law of large numbers (in Russian, 1914) [12, pp. 158 – 163].
8. --- On the coefficient of dispersion (1916). Transl.: DHS 2514, 1998, pp. 195 – 204).
9. Mendeleev, D.I. [Report on the investigation of the solidity of glass tubes.] (1874). *Сочинения* (Works), vol. 6. М. – Л., 1939, p. 187. (R)
10. --- On the elasticity of gases, pt. 1 (1875). Ibidem, pp. 221 – 589. (R)
11. Neumann, Franz, *Vorlesungen über die Theorie der Elasticität der festen Körper und des Lichtäthers*. Leipzig, 1885.
12. Ondar, Kh. O., Editor, *Correspondence between Markov and Chuprov* (1977, in Russian). New York, 1981.
13. *Прения между академиками* (Debates among Academicians in the Sitzings of the First Section of the Academy of Sciences). Psb, 1903.
14. Sheynin, O.B. Quetelet As a Statistician. AHES, vol. 36, 1986, pp. 281 – 325.
15. --- Markov's work on probability. Ibidem, vol. 39, pp. 337 – 377 and vol. 40, p. 387, 1989.
16. Tikhomirov, E. Correlation method and its applications in meteorology. *Geofisich. Sbornik*, vol. 2, No. 3, 1915, pp. 21 – 48. (R)
17. Tsinger, N.Ya. *Курс астрономии* (Course in Astronomy), pt. theoretical. Psb, 1899.

14. Markov's Letters to the Newspaper *Den*, 1914 – 1915. IMI, vol. 34, 1993, pp. 194 – 206

1. Introduction

1.1. General Explanation

Markov wrote a large number of letters to several newspapers. His sharp statements on burning social issues are indeed interesting, and Grodzensky, who reprinted some of them in his book [1], rendered an essential service to his readers. However, a number of letters, addressed to the newspaper *Rech* and found by Grodzensky in several archives, never appeared in print [1, p. 100]. Apparently, they were considered audacious or impudent. Markov's social activity and certainly his newspaper articles in particular resulted in that the press nicknamed him *Militant Academician* [2, p. 9]. Neyman [3, p. 486] reported that he had a second nickname as well, *Andrew the Furious*. Later Neyman also attributed to Markov a "limerick" about a *Duke Dundook* [4, p. v], but it was Pushkin who had composed a satiric verse (not a limerick) about Prince (not Duke) Dundook.¹

I am reprinting three newspaper letters written by Markov and devoted to the methodology and teaching of mathematics. All of them were published in a leftist newspaper *Den* (Day)². Minkovsky [5] briefly described Letter No. 1, Nekrasov [2, p. 8] referred to Letter No. 3, and I discovered Letter No. 2 making use of Markov's obscure indication in No. 3.

Pavel Alekseevich Nekrasov (1853 – 1924) [6], who will be prominently mentioned below, graduated from Moscow University. He taught there from 1885 as Privat-Dozent, then as Professor and became Rector in 1893. He published important investigations in algebra and probability theory; his dissertation, "The Lagrange series" (1886), is a noteworthy writing. Regrettably, his work is hardly studied. Youshkevich insufficiently described it in his book [7], and he was not even included in the well-known biographical dictionaries compiled by A.N. Bogoliubov or A.I. Borodin & A.S. Bugai. This was certainly caused by Nekrasov's reactionary political views and his activities as Rector of Moscow University, warden of the Moscow educational region, and, from 1905, in Petersburg, high-ranking official at the Ministry of Public Education.

Archival documents [8, p. 378] testify that Nekrasov pursued a tough policy towards the revolutionary-minded students. An anonymous author [9] vaguely corroborated this and expressed his hope that Nekrasov, in his next position as warden, will continue to educate young people "in the spirit of duty to God, Czar, and Fatherland". At the turn of the 19th century Nekrasov's activities underwent a sudden change and it may be supposed that his long-term work as an administrator coupled with failure to understand the spirit of the times was at least partly responsible for this fact. The writings of this eminent scientist became unimaginably verbose and hardly intelligible; he began to connect his mathematics with religion and politics and for this reason alone neither Markov, nor Liapunov perceived there any word of truth. Furthermore, mathematicians concluded that no scientific debate with Nekrasov were possible. Here is Liapunov's opinion [10, p. 62] about his pronouncements on the limit of function:

All of Nekrasov's objections are based on various misunderstandings. Then, some of them are not more than unsubstantiated declarations [...] whereas the other ones either do not at all relate to the subject-matter of the criticized papers, or are distinguished by extreme vagueness.[...] If Nekrasov will see fit to put forward objections of the same kind, I shall consider myself free from answering them.

Similar statements were due to Markov [11, p. 77] and Posse (also see Note 14).

Nekrasov's unpublished letters (1916) to P.A. Florensky testify to his political views. Here, for example, is a passage from his letter of 26 November:

The Moscow school put forward the principles of the language of Christian science and repulses the language in the style of Karl Marx, Markov, Ya.I. Lintsbach. The comparison of the books of Lintsbach, Markov[...] with those of the representatives of the Moscow philosophical-mathematical school clearly shows the crossroads to which the German-Jewish culture and literature are pushing us.

It is hardly amiss to note that Lintsbach [12] did not discuss either Marxism or religion.

So it happened that Nekrasov joined hands with the Black Hundred, as stated in plain terms by Markov Jr. [13, p. 89]. But, to repeat, Nekrasov's scientific achievements should not be forgotten. As an additional point, I note that Zhukovsky [14, p. 639] indicated that Nekrasov had helped him in solving a certain mathematical problem. Seneta [15, §§6 – 7] described Nekrasov's findings in probability and put on record that he had definitely influenced Markov (who sometimes referred to Nekrasov without criticizing him [16]).

1.2. The Teaching of Probability Theory in Schools (Letter No. 1)

In 1914 Nekrasov made an attempt to introduce probability theory into the school curriculum. He distributed a memorandum written by P.S. Florov; then, in 1915, he published an extensive article [17] containing Florov's pertinent program, comments made by many mathematicians during a discussion organized by the Ministry of Public Education and carried out by correspondence, the appropriate materials of the Second All-Russian Congress of Teachers of Mathematics (1914) and his own considerations in favor of the Florov's program.

Many participants in the discussion overstepped its limits by insisting on the inclusion of the fundamentals of mathematical analysis and analytic geometry in the school curriculum, but all of them understood the additional difficulties of any changes caused by war conditions. Markov apparently was not invited to take part in the discussion but he expressed his opinion on the subject in an ad hoc paper [18]. "The leading idea" of the Florov and Nekrasov project, he indicated (p. 33), consisted in "the need to acquaint the school students with their works". In essence, he (p. 26) stated that his viewpoint was close to the opinions of A.V. Vasiliev and B.M. Koialovich (his student, an author of textbooks, including those on probability theory, which had been appearing at least until 1931). Vasiliev in principle approved Nekrasov's project, and he, like Koialovich, was mainly objecting to Florov's program. Vasiliev also remarked that the theory of probability provided apt illustrations for combinatorial analysis and fostered logical reasoning, and that Kraevich, already in 1864, had included a felicitous section on probability in his celebrated collection of problems for school students [19]. Thus, Markov favorably commented on Nekrasov's idea but had not said so directly, nor did he offer his own program.³

Nekrasov appealed to the Vice-President of the Imperial (Petrograd) Academy of Sciences with a request to consider the possibility of implementing his proposal [20, pp. 96 – 97], and, on Markov's initiative, the Academy established an ad hoc commission, which included him, to study the Florov project. The commission agreed with Markov's negative opinion about the project and accused Nekrasov of aiming "to turn pure science into an instrument for bringing religious and political pressure to bear on the rising generation" [21, p. 105 – 106]. The commission did not examine in essence the teaching of probability, but "some" of its members (p. 102) were in principle against including that discipline into the curriculum. In itself, the theory of probability is certainly not a weapon of religion or politics, and it may be thought that the conclusion of the commission was conditioned by Nekrasov's well-known ideological views (§1.1).

Nekrasov continued to advocate his proposal (which was never implemented). Thus, he published a letter received from B.V. Stankevich (1916), the Head of the Physical Institute at Moscow University. There Stankevich [22, pp. 27 – 29] testified that probability "is only taught to senior students, and solely at the Mathematical Department" and that he felt himself "greatly strained in that the [...] listeners are not acquainted even with the elements of probability". He therefore cannot help wishing "the introduction of a short course in the theory of probability in the school curriculum".⁴ It seems that in 1914/1915 probability was not taught even at the Mathematical Department in Moscow, and this very fact makes Stankevich's opinion hardly convincing.⁵

Nekrasov devoted a special booklet [2] to the described episode, and, what is more important, to his relations with Markov. It is still unstudied and I shall only indicate that Nekrasov, first, published six of Markov's letters/postcards to him (1915/1916, pp. 56 – 62) and alleged that they had contained swear-words (which he omitted before publication); and, second, maintained (not, however, for the first time) that Markov had without due acknowledgement made use of some of his findings. Finally, back in 1898 – 1899 Nekrasov asked A.I. Chuprov to assist him in introducing a course in probability theory at the Law Department in Moscow. No practical consequences ever followed, see the translation of their correspondence in this collection.

1.3. Seminarists (Letter No. 2)

For at least many decades a considerable part of the graduates of theological seminaries had been entering universities. Thus, at Petersburg in 1875, 45.7% of the students at the Department of Natural Sciences were former seminarists; at the Department of Mathematical Sciences they amounted to 11.5%; and there were 29.2% of them at the University in general [23, p. 90]. The same source (p. 93) stated that the seminarists were much worse prepared for studying in universities than the recent alumni of gymnasiums, and that, for Russia as a whole, 53.3% of those who entered the universities in 1875 were such alumni whereas former seminarists comprised the "main" part of the rest 46.7%.

I have no such data for the turn of the century, but in any case by the beginning of the 20th century the secular education at the seminaries undoubtedly worsened. In 1911 The Most Holy Synod, "in executing the Imperial will", had worked out new regulations for the theological academic institutions "in the spiritual direction" [24, p. 3]. Before those regulations came into effect, "many members of the scientific staff" had abandoned the theological academies, and "cases became possible when figures absolutely unknown in science, but sufficiently noted in the arena of political struggle within the Church, were appointed to be profes-

sors” [25, p. 209]. And we may assume that a similar process had been going on in the seminaries, so that Markov’s opinion that seminarists, when entering universities, should not be preferred to the “realists”, was quite justified.

1.4. Infinitesimals (Letter No. 3)

An infinitesimal is a variable whose limit is zero. Citing that definition due to Cauchy, Markov added that he did not reckon zero among the values of an infinitesimal, see however Note 19. Nekrasov [26, p. 459] declared that Markov [11, pp. 11 – 12] “was apparently restoring” the Eulerian terminology. Euler, however, is known to have considered an infinitesimal as equal to zero [27, p. 267]. Youshkevich [27] traced the development of the concept of infinitesimal in the 18th century, and, since the difference between actual and potential infinitesimals had not affected Markov’s scientific work, I leave this subject alone. Note, however, that Markov [11] did not treat infinitesimals and that the page numbers (11 – 12) as given by Nekrasov did not belong there. On the other hand, Markov [11, p. 81] quoted Nekrasov who had declared [28, pp. 49 – 50] that a variable (P) should be considered as equivalent to its limit (L), i.e. that the ratio P/L should tend to zero. He then ascribed that trivial statement to the case of a variable ($?$) L and for some reason accused Chebyshev, Liapunov and Markov in that they had believed that condition $(P - L) \rightarrow 0$ was sufficient for such a value of L to be the limit of P . Thus, as Nekrasov continued, quite logically following his own ideas, his opponents had thought that, for example, with $n > 0$ x^n tended to $\sin x$ as $x \rightarrow 0$!

Markov indicated in his Letter No. 3 that Nekrasov had attempted “to direct the teaching in schools to a wrong track”. Posse [29, p. 72] was of the same opinion:

Nekrasov attempts to discredit an entire school of mathematicians whose representatives are all the Petrograd (and not only Petrograd) professors in the eyes of the teachers and students of high schools, and to direct the teaching of mathematics in these schools to a wrong track. I do not consider it possible to pass that attempt over in silence the less so since Nekrasov published it on the pages of an official periodical of the Ministry of Public Education ...

2. Markov’s Letters

2.1. Letter No. 1. Newspaper *Den*, 30 Jan. 1914, p. 4

A question for the Ministry of Public Education

I have recently found out that the Min. of Public Education is occupied with the problem of introducing the teaching of the elements of the theory of probability into high schools. I even have in my hands a printed memorandum on this subject written by P.S. Florov with comments by P.A. Nekrasov, a Member of the Council of the Minister of Public Education. Concerning its substance, I shall only say that it is very controversial and cannot serve as a foundation for establishing the teaching of probability theory in high schools; if required, I shall analyse it in detail at a later date.⁶

At present, I ought to indicate that the Min. of Publ. Educ. did not yet ask the representative of the “theory of probability”⁷ at Petersburg University, i.e., me, to submit his conclusion about that subject. I think that no serious business can be done without the participation of appropriate specialists, provided there is no deliberate intention to do it pell-mell and badly. I ought therefore to ask the Min. of Publ. Educ. whether it is really seriously engaged in the problem of teaching the “theory of probability” in high schools or considers it as an amusement of Professor Nekrasov now idling away his time?⁸

2.2. Letter No. 2. Seminarists and Realists. Newspaper *Den*, 11 Aug. 1915, p. 3

As stated in the newspapers, the seminarists are allowed to enter the physico-mathematical department⁹ without any special examination, whereas the issue about admitting realists¹⁰ to the department is left open. It is difficult to say that such a situation is normal. There is no sharp contrast between the schooling¹¹. Latin, that distinguishes the first from the second, is not necessary for the physical and mathematical education. As to the seminarists, they are getting accustomed by their schooling to a special kind of reasoning. A seminarist must subordinate his mind to the indications of the Holy Fathers and replace it by the texts from the Scripture. The seminary’s wisdom can be very deep as is shown by the fundamental writing [30],¹² but it is far from real science and is only able to establish religious truth for the believers. Such wisdom easily leads to the desire to place “the religious – scientific – political experience under the educational coat of arms” above science. Referring the readers to Nekrasov’s article [32] where they will find not only the senseless verbiage quoted just above, but also specimens of sage statements of special subtlety, I cannot help expressing great doubts about the suitability of the seminarists for the physical and mathematical department. In any case, they should not be preferred to the realists.

2.3. Letter No. 3. Newspaper *Den*, 28 Oct. 1915, p. 3

Dear Mr. Editor, – Allow me to raise two questions for lawyers through your respected newspaper. What measures can be taken, in the hope of succeeding, against the misuse of the [freedom of the] press if done by an official periodical? And, is the Editor of such a journal as responsible a figure as the editors of other organs of the press, or is all his responsibility restricted by unquestionably executing the will of the authorities?

In addition, since the *Журнал Министерства ... [J. Min. Publ. Educ.]* corrupted the facts, I am asking you to publish the following letter which I had sent to Mr. Radlov but which he refused to make public.¹³ I ought to indicate that the business concerns an issue of a social importance since Nekrasov, a Member of the Council of the Minister of Public Education, profiting by his influence, attempts to direct the teaching of mathematics in high school on a wrong track.

Dear Sir, Ernest Lvovich [Radlov], – Since the periodical which you are editing, carried two polemic articles by Nekrasov directly concerning me [32; 33], I am asking you to find place in your journal for the following explanation.

Professor Posse [29] had indicated some peculiar features of Nekrasov's style.¹⁴ Without calling in question the existence of those features, Nekrasov [33] maintains that he caught his style from me.¹⁵ This statement is not true since such a style is absolutely foreign for me. Just as I disagree with him on mathematical topics, so also I differ from him in style. In particular, I think it reasonable to pass over in silence all that which does not concern the business at hand and do not allow any confusion of mathematics with religion and politics.

Then, Nekrasov repeatedly referred to scientific societies. He said that I had taken the principles of the theory of limits to the court of scientific societies and that our debate was done away by them.¹⁶ Later he declared that his definition had passed the crucible of the judgement of a scientific society so that the reader can be sure of its truth.¹⁷ These references do not accurately reflect reality. Never had I taken the principles of the theory of limits to the court of scientific societies. Regarding these principles I ought to say that I did not introduce there any innovations. In my lectures on the diff. calc. I defined infinitesimals similar to how already Cauchy [34, p. 26] did it

(On dit qu'une quantité variable devient infiniment petite lorsque sa valeur numérique décroît indéfiniment de manière à converger vers la limite zéro)

and how they are defined in many foreign and Russian textbooks (for example, in the abovementioned book of Posse¹⁸), only I add the following words:¹⁹

*It is important to indicate that we do not reckon its limit, zero, to the totality of the values of an infinitely small number.*²⁰

These words do not unite me with Nekrasov, but in this particular case they are decisive.

I shall not repeat or explain what I said before about Nekrasov's discoveries,²¹ and I shall end my letter by indicating the resolutions of scientific societies. Nekrasov's paper [26] begins thus:

Editorial note. Nekrasov submitted this paper as an answer to Markov's article [11] and it is published in accordance with the decision of the Moscow Mathematical Society: to place in the Matematicheskyy Sbornik one article written by each of the two authors and devoted to the issue under consideration.

And, the proceedings of the sittings of the Kharkov Mathematical Society on 19 Jan. 1914 [38] carry the following:

*The letter of Markov concerned with his debate with Nekrasov was heard out. The Society, having heard and discussed it, being guided by the principle that both sides ought to be put as far as possible in identical positions, resolved to open the pages of the Soobshchenia for Markov's answer, if only he wishes to avail himself of that opportunity.*²² 9 Oct. 1915

So that the readers of the newspaper *Den* can form some opinion about Nekrasov's style, I consider it not amiss to refer to the end of one of his phrases already mentioned by me [in Letter No. 2]:

*Then only the religious – scientific – political experience under the educational coat of arms, similar to the one contained in the Arithmetic published (about 1700) by order of Peter the Great, is left.*²³

24 Oct. 1915²⁴

Notes

1. It seems possible that Markov recited the satirical verse “Prince Dundook” (rather than composed it) and directed it against Prince Galitsin, see the appropriate piece in this collection. The verse stated that the imaginary prince Dundook became Academician only because having been endowed by an ass.

2. After the February revolution (1917), some Mensheviks (Russian social-democrats opposed to the Leninist Bolsheviks) were included in its editorial staff. The newspaper was suppressed in 1918.

3. He also opposed the idea of including mathematical analysis and analytic geometry in the curriculum, but did not elaborate.

4. Senkevich also reported that V. Ya. Tsinger, in the 1880s, “had used to say” that the kinetic theory of gases was “anarchic”, whereas N. Ya. Sonin, in the 1890s, “although he did not approve of the theory, had regarded it already with some leniency”.

5. See Note 19 to Nekrasov’s correspondence with A. I. Chuprov (in this collection).

6. Later on Markov had indeed published a paper [18] devoted to the teaching of probability theory in school (§1.2).

7. The term *theory of probability* occurs several times in this Letter, twice in quotation marks. The Editor had apparently inserted them believing that the term was not sufficiently known.

8. A few lines above Markov (correctly) indicated Nekrasov’s position, so that his phrase is hardly understandable.

9. The singular form is strange.

10. Alumni of non-classical schools, cf. the German *Realschule*.

11. Obviously, between the seminaries and the non-classical schools.

12. This book contains an extensive supplement of a natural-scientific and mathematical nature and several hundred appropriate references, also see [31].

13. On this point see Radlov’s letter to Markov of 8 Oct. 1915 (Archive, Russian Academy of Sciences, Fond 173, inventory 1, 59 No. 3; fragment translated in DHS 2579, 1998, p. 100).

14. Posse [29, p. 71] indicated that Nekrasov

Likes to strike his opponent with apparently very serious, but actually very obscure phrases [...] and [...] when quoting the words of his opponent, he sometimes changes them and attributes to them something that they nowhere and never said.

15. Here are Nekrasov’s words [33, p. 97]:

I was challenged to the debate by the argumentation of my opponent, whom I answered by duty and catching his polemic style.

16. Here is the appropriate passage [32, p. 12]:

Here [18, pp. 27 – 28] Markov provided as proof the principles of the theory of limits which he [11, p. 81] had already taken to the court of scientific societies and gotten there my adequate response [26, p. 459].

In these articles nothing, however, was said about a “court of scientific societies”. Also see §1.4.

17. I quote Nekrasov [33, p. 101]:

My definition [26, p. 459] does not please Posse, but it passed the crucible of the judgements of a scientific society so that the reader can be sure of its truth ...

He never explained what did he mean by the truth of a definition.

18. Markov apparently bears in mind the definition in Posse’s book [35], but he had not mentioned any such book before.

19. Markov referred to his textbook of 1898 which I did not see; however, the same definition is contained in its later edition [36, p. 45]: “A variable tending to zero is called an infinitesimal. It is important to note ...” The report of the Commission established by the Academy of Sciences (1916) contained an explanation to the effect that Markov’s qualification remark was called forth by considerations of convenience [21, p. 101]. Markov was a member of that commission.

20. Hardly a proper term.

21. In 1910 Markov [37] declared: “I never confirmed and cannot confirm any of Nekrasov’s discoveries ...”

22. Markov’s answer never appeared.

23. Nekrasov referred to Magnitsky [39]. The coat of arm (let it even be “educational”) depicted there is described by Gnedenko [40, p. 57]. Nekrasov maintained that a course in mathematics restricted to deduction (i.e., without probability theory, as it followed from the context) was only a “religious – scientific – political experience”. It is hardly possible to understand him, but in any case induction is connected with statistics rather than with probability.

24. The newspaper *Den* provided a brief summary of the debate between Markov and Nekrasov by publishing a paper by Pavel Youshkevich [41] three weeks after Letter No. 3 had appeared, cf. [42, p. 307].

Acknowledgements. S.S. Demidov acquainted me with P.A. Nekrasov’s letters to Florensky which are kept by the family of the latter. M.V. Chirikov suggested some editorial changes in my manuscript.

References

1. Grodzensky, S. Ya. *Марков* (Markov). М., 1987.
2. Nekrasov, P.A. *Средняя школа ...* (The High School, Mathematics and Scient. Training of Teachers). Psk, 1916.
3. Neyman, J. Review of the original Russian edition of *Correspondence between Markov and Chuprov*. М., 1977. *Hist. Math.*, vol. 5, 1978, pp. 485 – 486.
4. Neyman, J. Introduction to the English transl. of same (New York, 1981). On pp. v – viii.
5. Minkovsky, V.L. Markov’s pedagogic ideas and activity. *Matematika v Shkole*, No. 5, 1952, pp. 10 – 16.
6. Anonymous, P.A. Nekrasov. *Энци. Словарь Брокгауз и Ефрон* (Brockhaus & Efron Enc. Dict.), vol. 20A, 1897, p. 861.
7. Youshkevich, A.P. *История математики ...* (History of Math. in Russia before 1917). М., 1968.
8. *История Московского университета* (History of Moscow Univ.), vol. 1. М., 1955.
9. Anonymous, The new warden of the Moscow educational region. Newspaper *Mosk. Vedomosti*, 13 (25) March, pp. 2 – 3, and 15 (27) March, p. 2, 1898. (R)
10. Liapunov, A.M. An answer to P.A. Nekrasov. *Zapiski Khark. Univ.*, vol. 3, 1901, pp. 51 – 63. Transl.: DHS 2579, 1998, pp. 53 – 63.
11. Markov, A.A. A rebuke to P.A. Nekrasov. MSB, vol. 28, 1912, pp. 215 – 227. Transl.: DHS 2579, 1998, pp. 77 – 82.
12. Lintsbach, Ya. *Принципы философского языка* (Principles of the Philosophical Language). Petrograd, 1916.
13. Markov, A.A. Jr, Biography of A.A. Markov. In Markov, A.A. (Sr), *Избранные труды* (Sel. Works). N.p., 1951, pp. 599 – 613. Transl.: DHS 2696, 2000, pp. 81 – 94.
14. Zhukovsky, N.E. On the form of ships (1890). *Собрание сочинений* (Coll. Works), vol. 2. М. – L., 1949, pp. 627 – 639. (R)
15. Seneta, E. The central limit theorem and linear least squares in pre-revolutionary Russia. *Math. Scientist*, vol. 9, 1984, pp. 37 – 77.
16. Markov, A.A. On probability a posteriori. *Soobshchenia Khark. Matematich. Obshchestvo*, ser. 2, vol. 14, No. 3, 1914, pp. 105 – 112. (R)
17. Nekrasov, P.A. Theory of probability and mathematics in the high school. ZhMNP, 1915, 4th paging, No. 2, pp. 65 – 127; No. 3, pp. 1 – 43; No. 4, pp. 94 – 125. (R)
18. Markov, A.A. On the draft of teaching the theory of probability in the high school drawn up by P.S. Florov and P.A. Nekrasov. Ibidem, No. 5, pp. 26 – 34. (R)
19. Kraevich, K.D. *Собрание алгебраических задач* (Collection of Algebraic Problems). Psk, 1864, 1867, 1874, 1882.
20. Nekrasov, P.A. Letter to Vice-President of the Imperial Academy of Sciences as quoted and described by the Permanent Secretary of the Academy in his letter of 5 Nov. 1915 to Markov. Archive, Russian Acad. Sci., Fond 173, inventory 1, 57 No. 1. Transl.: DHS 2579, 1998, pp. 96 – 99.
21. Report of the Commission on discussing some issues concerning the teaching of mathematics in high school. *Izvestia Akad. Nauk*, vol. 10, No. 2, 1916, pp. 66 – 80. Transl.: DHS 2696, 2000, pp. 95 – 108.
22. Nekrasov, P.A. *Принцип эквивалентности...* (Principle of Equivalence of Magnitudes in the Theory of Limits and in Successive Approximate Calculus). Petrograd, 1916.
23. Anonymous, On the number of graduates of the gymnasiums and of those who entered the universities in 1875. ZhMNP, No. 2, 1876, sect. Sovremenn. Letopis, pp. 88 – 93. (R)
24. *Журналы особой комиссии ...* (Minutes of Special Commission of the Synod for Working Out Draft Regulations for High Theological Schools). Psk, 1911.
25. Nikolsky, N. Theological academic institutions. *Энци. Словарь Гранат* (Enz. Slovar Granat), vol. 19. N.d., pp. 202 – 209. (R)

26. Nekrasov, P.A. The general main method of generating functions. MSb, vol. 28, 1912, pp. 351 – 460. (R)
27. Youshkevich, A.P. Differential and integral calculus. In *История математики ...* (Hist. Math. From the Most Ancient Times to the Beginning of the 19th Century), vol. 3. Editor, A.P. Youshkevich. M., 1972, pp. 241 – 368.
28. Nekrasov, P.A. Concerning a simplest theorem on probabilities of sums and means. MSb, vol. 22, 1901, pp. 225 – 238. Transl.: DHS 2579, 1998, pp. 43 – 51.
29. Posse, A.K. A few words on Nekrasov's article [32]. ZhMNP, No. 9, 1915, sect. Sovremenn. Letopis, pp. 71 – 76. (R)
30. Florensky, P.A. *Столп и утверждение истины* (Pillar and Consolidation of Truth). M., 1914 and 1990. (R)
31. Demidov, S.S. et al, On the correspondence of N.N. Lusin and P.A. Florensky. IMI, vol. 31, 1989, pp. 116 – 191. (R)
32. Nekrasov, P.A. On Markov's paper [18]. ZhMNP, No. 7, 1915, sect. Sovremen. Letopis, pp. 1 – 17. (R)
33. Nekrasov, P.A. A reply to Posse's objections. Ibidem, No. 10, pp. 97 – 104. (R)
34. Cauchy, A.L., *Course d'analyse de l'Ecole Royale Polytechnique*, 1^{re} pt. *Analyse algébrique*. Paris, 1821.
35. Posse, K.A. *Курс дифференциального ...* (Course in Diff. and Integr. Calculus). Psk, 1903.
36. Markov, A.A. *Дифференциальное ...* (Diff. Calculus). Psk, 1901. Lithogr. Edition.
37. Markov, A.A. Correcting an inaccuracy. *Izvestia Akad. Nauk*, ser. 6, vol. 4, No. 5, 1910, p. 346. (R)
38. *Soobshchenia Khark. Matematich. Obshchestvo*, ser. 2, vol. 14, No. 6, 1915, second paging, p. 8. (R)
39. Magnitsky, L.F. *Арифметика* (Arithmetic). M., 1703.
40. Gnedenko, B.V. *Очерки по истории ...* (Essays Hist. Math. in Russia). M. – L., 1946.
41. Youshkevich, P.[S.] On a scientific debate (1915). Appended to this paper, pp. 207 – 209 of the Russian original. (R)
42. Sheynin, O.B. Liapunov's letters to Andreev (1989). Transl. in this collection.

15. The Correspondence of Nekrasov and Andreev

IMI, vol. 35, 1994, pp. 124 – 147. Coauthor: M.V. Chirikov

1. General information

1.1. Introduction

Pavel Alekseevich Nekrasov (1853 – 1924) [1, §5; 2], Professor and Rector of Moscow University, then a prominent official at the Ministry of Public Education, was a distinguished mathematician and a religious person. A Platonist according to his philosophical views, he kept to reactionary political convictions. Konstantin Alekseevich Andreev (1848 – 1921) was a Corresponding Member of the Imperial (Petersburg) Academy of Sciences, a geometer and Professor at Kharkov and Moscow. The correspondence of Nekrasov and Andreev¹ was devoted to many subjects: the teaching of probability theory in high school; the encounters of both of them (but mostly of Nekrasov) with Markov;² the foundations of mathematical analysis; the central limit theorem (CLT). The main student of the contemporaneous history of that theorem is Seneta [4, §§6 and 7].³ For us, it suffices to say that in 1898 Nekrasov [5], having applied the methods of the theory of functions of a complex variable, sketched the proofs of the local and integral forms of the CLT for large deviations for sums of lattice random variables. His work made difficult reading and nobody appreciated it; the fate of his later writings proved to be just as dismal (cf. §1.3).

In his letters to Andreev Nekrasov repeatedly asked him to ascertain the possibilities of discussing some problems, and of reporting at the MMS, and we emphasize that the former, although being an “elder” there (see Letter 11), never occupied any official position at the Society. On the other hand, it follows from the concluding salutations in the letters of both Nekrasov and Andreev that there existed ties between their families.

1.2. The Teaching of Probability Theory in Schools

Nekrasov’s attempts [6] to introduce the theory of probability into high school are well known [2, §2].⁴ Like many other reformers, he had not thought about the difficulties of management which would have arisen had his proposals been implemented. Furthermore, for some reason he based them on P.S. Florov’s program compiled by that mathematician on a low theoretical level. A number of scientists beginning with Markov [8; 9] had therefore come out against Nekrasov’s attempt and killed it.

Not feeling himself defeated, Nekrasov continued to explain his theoretically correct viewpoint in private letters. His main step, however, was an appeal to the Vice-President of the Academy of Sciences, the philologist P.V. Nikitin.⁵ As a result, on Markov’s initiative, the Academy established a commission which sharply denounced the Florov – Nekrasov proposal [10], but, at the same time, missed the opportunity to reform the Russian school program. Still, Nekrasov stood his ground [3]. In particular, he (pp. 44 – 45) listed the commissions and congresses of the teachers of mathematics, which, over the years, had to do with the school mathematical curriculum and stressed that in 1914 the commercial schools had included elements of probability theory into their program “in spite of the brakes created by Markov and his colleagues”. For the sake of comprehensiveness we list Nekrasov’s writings at least partly devoted to the teaching of probability in schools [11 – 17; 6; 18; 3] and we quote his generalizing declaration [3, p. 51]:

*At bottom, my official activities in defining the various types of schools and mathematical programs [...] are reduced to an ideological struggle that aims at completely upholding the classical values of the mathematical education in all types of the general school.*⁶

1.3. Nekrasov’s Writings. Some Conclusions about Them

From about 1900 Nekrasov’s mathematical writings became unimaginably verbose, sometimes obscure and confusing, with mathematics being inseparably connected with ethical, political and religious considerations. Markov [19; 20] expressed himself against this manner of exposition and Youshkevich [21] offered a number of Nekrasov’s phrase-mongering to illustrate his intolerable style. Much earlier Borkiewicz, in a forgotten paper [22], accused Nekrasov of oily words (p. 215), reactionary longings (p. 216) and of attempting to justify “the principles of strong rule and autocracy” by the theory of probability (p. 219).

Nekrasov [13], however, expressly declared that a consolutistic [!] basic education should be of a scientific – religious – national – state nature”. It is still possible to understand this statement but not his own writings compiled according to the same principle. Then, coming out for the introduction of logic into schools of all types, and considering [11, p. v] that school mathematics should be based on logic, he (p. iii) included elements of probability theory and the Jakob Bernoulli theorem into it. Mathematics, as he declared in addition on p. 9, accumulated

Psychological discipline as well as political and social arithmetic or the mathematical law

of the political and social development of forces depending on mental and physiological principles!

This monstrous phrase apparently had to do with the works of Quetelet. Elsewhere, in connection with the statistical method, Nekrasov [7, p. 29] mentioned “problems of labor, and public wealth, of credit, insurance of life and capacity to work”. Not restricting his efforts by upholding his own views, Nekrasov had been accusing Markov of pan-physism⁷ and of following Nietzsche only because his opponent did not lump together mathematics with ethics, philosophy etc. [18]:

The mathematical language [must] [...] embrace supreme ethics, [be] together with conscience (with theology) [...] However, the mathematical language of such pan-physicists as Markov is of another kind, it is Nietzschean and does not recognize supreme ethics (theology).

All the above excepting the support of Florov’s unfit program (§1.2) is yet compatible with subjective scientific honesty. However, mathematical mistakes and unwarranted statements indicated by Markov [24; 20], Liapunov [25] and Posse [26], also see [2, §4], impede even this conclusion.⁸ The abovementioned peculiar features of Nekrasov’s style prevented the recognition of his works, and, to the contrary, favored his being considered only as a muddleheaded reactionary, We are unable to comment on a statement [27, p. 225] that he suffered from a mental illness, but it is impossible to deny Andreev’s opinion formulated by him in a letter to Liapunov in 1901 [28, p. 40]: Nekrasov

Reasons perhaps deeply, but not clearly, and he expresses his thoughts still more obscurely. I am only surprised that he is so self-confident. In his situation, with the administrative burden weighing heavily upon him, it is even impossible, as I imagine, to have enough time for calmly considering deep scientific problems, so that it would have been better not to study them at all.⁹

Agreeing with Andreev, we believe that all of Nekrasov’s philosophical and mathematical statements should be regarded as doubtful. At the same time, we provide illustrations of his deep thoughts which enable us to consider him as some mathematical Nostradamus. Thus,

1) His ideas about the dominance of logic (above) sounds really modern since it is possible to assume that he also bore in mind mathematical logic.

2) In connection with the mathematical study of indeterminacies Nekrasov [7, p. 23] mentioned almost all the main problems of the then not yet existing theory of catastrophes (and used the term *catastrophe*).

A special point concerns three Nekrasov’s mistakes of constructing/spelling and his own coined words.¹⁰ I attempted to preserve the former in translation so that the reader will see *consoliticism* (§1.3), *equivalentness* (Letter 9) and *illiteracism* (Note 38). A mistake of another kind is his use of *pamphlet* for an article.

1.4. Markov’s Polemic Style

Nekrasov repeatedly complained in his writings about the sharp tone of Markov’s polemic statements and even about the rudeness of his private letters [3, pp. 56 – 62]. Andreev believed that scientific debates with Markov were simply impossible. Here is a passage from his letter of 13 April 1901 to Liapunov [28, p. 42]:¹¹

I have experienced on myself all the annoyance of debating with a man who does not like to restrict his sharp expressions at somebody else’s expense. Markov all but scolded me.

Then, Slutsky (letter of 22 Nov. 1912 to Chuprov [68, p. 44]) tactfully remarked that Markov possessed an “unusual” manner of writing private letters, whereas Chuprov, in a letter to an English statistician Isserlis written late in 1925 or early in 1926 (Ibidem, p. 55), indicated that

Markov’s temper was no better than Pearson’s; he could not tolerate even slightest contradictions either.

However, Markov’s very critical letter of 29 April 1913 to N.A. Morozov,¹² a former political prisoner, was polite and ended in a way unusual for him: “Please be assured of my perfect esteem and devotion”.

Acknowledgement. S.S. Demidov acquainted us with Nekrasov’s letters to Florensky which are being kept by the latter’s family.

Notes to §1

1. Archive of Moscow State University, Fond 217, inventory 1, No. 45 (Andreev's letters) and No. 92 (letters written by Nekrasov).

2. This second topic is also described in [2], which, however, was based on Markov's newspaper letters. Nekrasov himself [3, p. 52] attributed the beginning of his sharp scientific debates with Markov (not yet regarding probability) to the beginning of the 1890s. Seneta [4, p. 70] noted that in those days (in 1892) the relations between the two mathematicians were yet normal: Nekrasov even read out one of Markov's reports to the Moscow Mathematical Society (MMS).

3. Seneta also provided sufficient information about Nekrasov's life. In lesser detail he described Nekrasov's efforts to introduce probability theory into high school.

4. Very interesting are Nekrasov's more general thoughts [7, pp. 30 – 31] on the teaching of mathematics in school. They allow us to perceive his notion of "classical values" of the appropriate course. He recommended to include the theory of probability, elements of analytic geometry and analysis as well as the "consecutive approximate analysis" into the school curriculum. He related the last-mentioned subject to induction (understood in a wider sense) and he mentioned in this connection Laplace, Poincaré and other scholars (p. 19). He attached much importance to the establishment of mathematical classrooms and the educational use of cinema (pp. 30 – 31).

5. Nekrasov [3, pp. 55 and 58] subsequently published two letters to Nikitin written on 29 Sept. 1915 and 5 April 1916.

6. Cf. Nekrasov's statement from his letter to P.A. Florensky of 2 Nov. 1916:

For the sake of the future of our fatherland, it is necessary to raise the standard of mathematical education in the school but protect it from the Markov & Co's frame of mind by those precepts, emblems and exercises which are included in our native tongue, in Magnitsky's arithmetic, in Bugaev's arithmology, in the theory of probability of Buniakovsky, Chebyshev, Mendeleev and me.

The term *arithmology* introduced by Nekrasov's teacher, Bugaev, meant *theory of numbers* but later became a synonym of a doctrine of discrete functions and even a *Weltanschauung* based on discreteness. It is now dated. Mendeleev did not have either any probability theory or even systematized indications on treating observations. True, Nekrasov [17, p. 4] declared that "the maps and the principles of nomography" in the great scholar's book *К познанию России* (On Coming To Know Russia) "are adapted to the Chebyshev theorem" but this obscure remark did not explain anything at all. Magnitsky was the author of the first Russian treatise on arithmetic.

7. See Letter 12 and the appropriate commentary. Creation of new words by amateurs was then in vogue.

8. It is hardly known that Nekrasov [23, p. 11; 4, pp. 45 – 46] committed an elementary logical mistake when proving the convergence of an iterative process.

9. This statement was possibly prompted by Liapunov's lost commentary on a letter of 16 March 1901 from Nekrasov to him [29, p. 84]. Nekrasov advised Liapunov not to hurry with publications on probability theory and maintained that the latter's theorems "like those of Chebyshev" were corrupted by mistakes. It is quite possible that Liapunov could have imparted his thoughts about that letter to Andreev. Seneta [4, p. 63] indicated Nekrasov's unjustified declaration [30, §168] about Liapunov's memoir [31].

10. See Note 7 above and Note 50 about the latter. The last invented word was *juridism* instead of "jurisdiction of sorts".

11. Andreev apparently had in mind Markov's note [32]. Similar statements are in his Letter 3 (below), and, indirectly, in his Letter 7 (see Note 34). It is sufficient to indicate here that an editorial note attached to [33] explained that Markov had "declared that he considered it [...] impossible" to replace some phrases in his letter [32] "by insertions without a sharp tone and not containing references to some personal relations". Consequently, [32] was published with cuts.

12. Archive of the Russian Academy of Sciences, Fond 543, inventory 4, No. 1130. Markov denied Morozov's paper devoted to the application of the statistical method to linguistics and even publicly expressed his opinion [34].

2. The Correspondence between Nekrasov and Andreev

We adduce now these letters (with insignificant abridgements). In a number of cases we had to specify the references.

1. Nekrasov – Andreev, 14 Oct. 1915

Nekrasov had sent Andreev his "pamphlets" including an offprint of [35].

"As in the past,¹³ so now also K.A. Posse [26] appears as Markov's advocate when the latter, in attempting to discredit his opponents, gets entangled in his own nets. This time Posse came out because Markov was

painfully flogged in my article [17] [...] This encounter has a double lining. One of these is the natural struggle between schools differing in their principles;¹⁴ but the other one, less visible, represents a tacit aspiration of a group of Petrograd mathematicians for subordinating [other] schools to their practical influence. Thus, for example, the reviews written by the members of the scientific committee, Posse and Koialovich, and academician Markov, had killed, for all purposes, the talented works of P.S. Florov on analysis and probability theory. You probably know Florov who was a student at Kharkov University; he possesses a gift of explaining issues of higher mathematics in an elementary way.¹⁵

“Leaving aside the second lining since it touches the academic-managerial system, I wish to seek your advice about the first one that concerns the principles of mathematics of *main* importance for science and education. I am deeply convinced that the comparison of the Brashman – Davidov – Bredikhin – Imshenetsky – Bugaev – Tsinger school¹⁶ with that of Posse – Markov reveals the greater value of the former’s principles. At the same time, however, the latter is militant, and, by means of Markov & Co.’s 16-inches’ debate [cannonade], it is attempting to overthrow the best principles so as to replace them by their own ones. I alone have to withstand the charge of an entire bloc.

“Perhaps the Mathematical Society can objectively (without any personal debate) compile definite decisions on the points of disagreement on principle as they are revealed in a number of my encounters with Markov & Co. Issues to be decided could be formulated about classifying the concepts; on relations of analysis and arithmetic with mechanics and probability theory; and about the preference of some methods to other ones (e.g., about comparing the Bienaymé – Chebyshev – Markov method with the method of Cauchy – Chebyshev – Nekrasov – Pearson¹⁷ [...]”

2. Nekrasov – Andreev, 15 Oct. 1915

“... I have prepared a brief report which I will entitle thus: *The concepts limit and asymptotic equivalent of a function in the calculus of probabilities of sums and means when integer m* [apparently: the argument of the function] *increases unboundedly.*¹⁸ There will be no personal debates in this talk, but, nevertheless, the main concepts will be discussed so exhaustively as to overturn completely the Markov & Co.’s declaration [24; 8; 26] that I, rather than they (Posse and Markov), am abusing the concepts *limit* and *infinitesimal*.”

The end of the letter is lost.

3. Andreev – Nekrasov, 24 Oct. 1915

Andreev received the manuscript of Nekrasov’s report. It had not indeed “contained any polemical sharp words” but it can lead to new discussion. The MMS resolved to transfer the decision about publishing the manuscript in the *Matematicheskyy Sbornik* to its Bureau.

“Naturally, I have absolutely abstained from any personal testimonial about the debate. Assuming that you are interested in my opinion about the differences between you, and Posse & Markov, I venture to formulate it. [...] I believe that it is completely out of question to decide who is right in your debate. Its essence is not to determine the correctness or otherwise of some judgements based on rigorously established assumptions, but to establish these very assumptions. Even if this does not belong to metaphysics, it at least lies in the province of intuition in the broadest sense of this word. Here, along with the mind, [...] appear, with a certain degree of being in the right, [...] tastes, inclinations, habits, acquired outlooks, sometimes even random points of views, [...] about which [...] *non est disputandum* [...]

“My life experience showed me that the champions of either direction sin, each of them in their own way. Some, while attempting to build firmly on a not yet prepared and still quaking soil, are compelled [...] to make use of diffuse and tempting explanations, in verbose formulations, etc. That is your sin.

“Others, being unable to justify the [mistaken] proposition that progress in science is conditioned by barring the expansion of the mental outlook, resort to sophistry and cannot resist the temptation of stinging their opponents not by scientific, but by journalistic weapons. That is Posse’s sin.¹⁹ [...] Not being a sinner in either of these senses, [Markov] [...] to this day remains an old and hardened sinner in provoking debate. I had understood this long ago, and I believe that the only way to save myself from the trouble of swallowing the provocateur’s bait is a refusal to respond to any of his attacks [...]”²⁰

4. Nekrasov – Andreev, 25 Oct. 1915

Having received no answer to his Letters 1 and 2, and understanding that his requests were difficult to fulfil, Nekrasov is prepared to abandon them, but he asks Andreev to return him the “suggested report” to the MMS entitled *Criticism of the connection and difference between the concepts limit and equivalent of a function of an unboundedly increasing number N* . Nekrasov then supplements the text of his report by the following considerations.

“...the concept *equivalent of a function*, that I am widely using in the calculus of discrete functions $\varphi(N)$ of a discrete and very large number N , has also been applied for a long time in another section of mathematics, namely, in the analysis of infinitesimals having to do with continuous magnitudes. Imshenetsky made use

of that concept in 1873, in his *Supplements* to [39]²¹ and thus essentially added to the Lagrange's and Todhunter's concept known as the *theory of analytic functions*, and, at the same time, he paved the way for bringing together the theories of continuous analytic and of discrete (arithmological) functions.

"Boussinesq [40] does not apply the term *equivalent*, but, like Imshenetsky, he (pp. 64 – 66) established principles equivalent [tantamount] to this concept and indicates their importance for simplification.²² These simplifications, as Imshenetsky put it, lead to *imperfect equations*, and, for continuous functions, they lead, in the limiting case, to *perfect equations* between differential coefficients. [...]"

5. Nekrasov – Andreev, 30 Oct. 1915

Nekrasov thanks Andreev for fulfilling his requests formulated in Letters 1 and 2.²³ He agrees beforehand with any future decision made by the MMS about his debate with his *main opponent*, Markov, but he asks that his work mentioned in Letter 4 be additionally considered. That letter will "reveal more perfectly the criminal [?] sense of what was said by me [Nekrasov] and Markov [41, p. 459; 30, vol. 22, p. 326; 24, pp. 223 – 224].

"The fundamental issues of the calculus of equivalents and limits are closely linked with the main tendency of mathematicians to *simplify* formulas and to admit, for the sake of simplification and saving of time, [...] a rightful dose of subjectivism(Boussinesq) and of *active intuition* (of experience, of the *ars conjectandi* [art of conjecture] in the spirit of Jakob Bernoulli) so as to approach the *objective truth*,²⁴ – to admit it to an extent which will not lead to overstepping the *right* to make slight errors. Mathematics of approximate and asymptotic calculations has an exact *juridism* with rules, customs, instructions of/in computation which must be categorically observed. The question is, who of us, Markov or I, oversteps this extent in the differential calculus of probability ΔP ?²⁵ [...]"

"My prosecutor in the person of Markov had not, however, calmed down and continues to charge me definitely with introducing fundamental mistakes into the theory of limits, the doctrine of infinities and of infinitely low probabilities".

Nekrasov listed a number of papers published by Markov, Posse and by himself [24; 9; 26; 17; 35] and continued:

"... from 1898 onward, while rendering proper homage to Chebyshev, I am, however, publicly maintaining that his *theorem* on probabilities [37] is rather a *postulate*²⁶ demanding *critical* attitude, and that the *fundamental* faults in the calculus of probability are to be found *not in my work, but in Markov's writings*. Indeed, he, even after the publication of my memoir [5], *persisted in claiming that Chebyshev's postulate is a theorem* [46; 48 – 50; 51, 1900, pp. 88 – 89]. Later, evidently under my influence, Markov [51, 1913, pp. 88 – 97] changed the theorem; however, *he introduced a lacunary (a molar)*²⁷ *reckoning when measuring a varianta*²⁸

$$X = \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_m.$$

*But he passed over in silence both that lacunarity and the fact that filling it in will demand the recognition of all the power of the fundamental base of my and Pearson's critical attitude towards the known differential calculus of the probability of the value of X.*²⁹ [...]"

"I am quite sharing your opinion that it would have been best to refuse to respond to Markov's provocative attacks so as not to swallow his bait. And I had indeed ignored them until the attacks were perpetrated in some newspaper (the provocation in the newspaper *Den* [20] became Markov's usual business)".

However, Nekrasov cannot keep silent when Markov publishes such attacks in the periodical of the Ministry of Public Education [9] or of the Academy of Sciences.³⁰ The letter in the newspaper [20] "impertinently slanders [Nekrasov] as though I attempt to direct the teaching of mathematics in schools on a wrong track"³¹ [...] Actually, my project, and even not my personally, but the collective project compiled by the professors and teachers of the Moscow educational region and the Petersburg Ministerial Commission in 1899 – 1900 [was] later [discussed] at the All-Russian congresses of teachers of mathematics.³² [...]"

"I would like to ask you to consider personally and carefully the essence of the debate and to take a firm stand with confidence whether I am wrong, or is Markov wrong.[...]"

6. Nekrasov – Andreev, 15 Nov. 1915

"The debate is going on not about some depths of metaphysics, but only about the calculus of actual infinitesimal probabilities. The abuse of mathematics (*petition principii* [begging the question]) is not mine, but on Markov's side since he scolds [denies] my right to apply the known simplifying principle of replacing actual infinitesimals by their equivalents [...]"³³

7. Andreev – Nekrasov, 17 Nov. 1915

The MSS's Bureau indicated that Nekrasov's suggested report concerns the subject of the debate which the Society had earlier (*Matematich. Sb.*, vol. 28, 1912, p. 351) resolved to restrict by publishing one paper of each of the debaters. The MMS will apparently approve the Bureau's opinion.

"You could have plainly understood [Letter 3] that I consider any investigation of this debate [...] as at least a useless business. [...] In essence, all that you wish is that at least one such person will be found who explicates your own ideas more clearly [...] I may assure you [...] that you are severely mistaken if you think that my firm stand taken with confidence can lead to the resolution of the debate. Not confidence is here needed but clarity. However, an explanation of somebody else's ideas is a thankless and very risky business. [...] Only once in my life did I allow myself to explicate somebody else's ideas [54], the ideas of the late Imshenetsky [55], but I have since felt sorry for myself because I saw that I was occupied with a needless business which did not benefit anyone or anything."³⁴

"[...] I certainly cannot approve of the academician's [Markov's] appeal to the public opinion formulated as a newspaper feuilleton [20]. However, once he brought himself to come out on that arena, he had the right to use all the means usually applied there, and, in particular, to parade in an ungainly fashion the weak points and the blunders of his opponent. Markov [20] makes use of this weapon very skillfully and deftly [...]"

Andreev then comments on Nekrasov's unfortunate, to say the least, statement quoted by Markov:

"[...] It is utterly unthinkable that you were seriously convinced that the publication of an article in the transactions of a scientific society may serve as some crucible [...] or that it indicates the article's approval by the society. [...] All the previous lines were evoked by a feeling of my sincere liking and goodwill towards you."

8. Nekrasov – Andreev, 5 Dec. 1915³⁵

Upon considering Nekrasov's letter to its Vice-President, the Academy of Sciences "resolved to constitute a commission" for looking into the teaching of the theory of probability in the school. The mathematical section of the Pedagogical Museum of the military schools discussed the reports of A.N. Krylov, S.A. Bogomolov, Ya.V. Uspensky and Nekrasov himself.³⁶

"At the [...] Museum I delivered a talk on the more elementary part of my report *Criticism* etc [see Letters 2 and 4] and the educational and simplifying significance of the principle of equivalence of magnitudes was ascertained".

9. Nekrasov – Andreev, 13 Dec. 1915

"[...] my memoir touched on the central issue of the fundamentals of mathematics, namely, on the equivalence of functions. It shows the normal way of induction from the simpler to the more complex".³⁷

Nekrasov then mentioned Imshenetsky and Boussinesq (see Letter 4), referred to Barbèra [57] and continued:

"My own works on the calculus of functions [of very large numbers], on approximate and asymptotic laws of equivalentness [!] of functions are completed along the entire line and in all rigor so that I am firm in my conviction against all the insinuations of Markov and Posse [...]"³⁸

Nekrasov mentioned the *Zhurnal Ministerstva Narodnogo Prosveshchenia*, cited Markov [24, pp. 223 – 224] and informed Andreev that, in connection with his report at the Pedagogical Museum (see Letter 8) and with the appearance of a Russian translation of Newton's *Principia* [58], he would like to supplement his suggested report at the MMS, and, in addition, to refer there to Barbèra [57].

10. Nekrasov – Andreev, 12 Jan. 1916

Nekrasov once more (see Letter 9) makes known his desire to improve his report indicating the same reasons as before but this time without mentioning Barbèra [57].

11. Nekrasov – Andreev, 5 Febr. 1916

The Commission of the Academy of Sciences "took advantage of the school issue only for settling the score with me and for compiling a new pamphlet against me [10] [...]"

Nekrasov asks permission to answer the Commission via the *Matematich. Sb.* "And I am once more asking you as an elder among the representatives of pure mathematics and pedagogy³⁹ to support me with all resoluteness [...]" The report of the Commission [10, p. 79] includes the "main distortion of the basis of my scientific and philosophical concepts."⁴⁰ [...] I never confuse philosophy [...] with pure mathematics⁴¹ [...] The booklet [59] contains ideas identical in spirit with mine."⁴²

12. Nekrasov – Andreev, 7 March 1916

The MSS had not allowed Nekrasov to publish his answer to the Academic Commission in their periodical but he thanks Andreev for his troubles.

“You have correctly indicated that the struggle is going on on three fronts. (The first one: the fundamentals of the analysis of infinitesimals and of their approximate asymptotic calculus; the second one: fundamentals of the theory of probability; and the third front: the bringing of mathematics and probability theory into proper relation with issues in physics, religion and politics.⁴³) You think that for me the struggle on the third front is hopeless. However, it is here that the Commission has distorted my works most of all.⁴⁴ [...] Markov [51, 1908] treats problems in religion and politics (ignorantly, by means rejected not only by Buniakovsky [60, p. 326],⁴⁵ but also by Boole, Jevons, Bertrand and Karl Pearson⁴⁶), and this is allowed and even commendable[...].”

If, however, Nekrasov [43] discusses the same issues, it becomes

“An inadmissible abuse of mathematics. [...] This is a purely Prussian objectivity of reasoning.⁴⁷ [...] In spite of the Commission’s statements I distinguish two main directions struggling with each other. The motto of the school belonging to one of these is mathematical humanism; the motto of the school of the other direction (trend) is physico-mathematical realism.⁴⁸ If these trends can be united into a single one, it might be done only under the first motto, only when [additionally] stating that *the ideal of science is the lamp of the truth* (V.Ya. Tsinger). Pearson, a mathematician and a humanist, considers the separation of science and philosophy as *obscurantism* [61, p. 55 of the Russian translation]. [...] The founders of the MMS were of the same opinion [62; 63, 64]”.⁴⁹

Nekrasov then criticizes the Russian translation of Newton’s *Principia* [58]⁵⁰ and declares that the theory of probability is the basis of “a wide mathematical induction in the sphere of disputable but vital issues (Poincaré, Pearson, N.A. Umov)[...]”⁵¹ Regrettably, in 1872 the school in Russia took the road to ruthless pseudo-classicism and formalism and reshaped the minds of contemporaries in a different fashion;⁵² physico-mathematics was substituted for mathematics, humanism became thought of as being opposite of mathematics rather than of the extremes of materialism and heartless formalism”.⁵³

The mathematical societies in Moscow and Kharkov came into being, as Nekrasov indicated, for struggling against such obscurantism.

“Had the Mosc. Mathematical Society wished to defend the humanitarian branch of mathematics with its spiritual culture, it would have very, very strongly supported my just claim to correct the distortions committed by Markov & Co.”

Nekrasov begins thinking about leaving the Society. He asks Andreev to inform the MMS about his intention “as about a tentative decision” and to ask them whether they did not become “only physico-mathematical rather than widely mathematical [...]” The formal cause for his leaving will be the Society’s refusal to enable him to defend himself from the Academic Commission. The end of the letter is lost.

13. Nekrasov – Andreev, 13 March 1916

After receiving Andreev’s (lost) reply to Letter 12, Nekrasov asks him “to do nothing” concerning his intended leaving of the MMS.

Notes to §2

13. Nekrasov apparently thought about letter [36], see Note 34.

14. The text below makes it clear that Nekrasov contrasts Moscow scientists with those in Petrograd (Petersburg). The lumping together of such mathematicians as Davidov and Bugaev and the astronomer Bredikhin (below) allows us to question whether such a school existed at all, cf. Notes 16 and 17.

15. Posse and B.M. Koialovich were members of the scientific committee of the Ministry of Public Education. Concerning the latter see *Новый энци. словарь* (New Enc. Dict.), vol. 23. Petrograd [, 1915], p. 46. In 1883 – 1888 Florov published not less than eight papers on mathematical analysis (mostly, on differential equations) in the *Soobshchenia Math. Soc. Kharkov Univ.* In 1912 – 1915 he also wrote three superficial articles on the Jakob Bernoulli theorem; on the Buffon needle; and on insurance of life, all of them in *Vestnik Opytn. Fiziki i Elementarn. Mat.*, and he reported on annuities at the Second All-Russian Congress of Teachers of Mathematics (1913 – 1914). Nekrasov [17, p. 4] also publicly argued that Florov “possesses a wise gift for explicating great truths in the simplest form ...”

We are unaware either of any reviews of Florov’s papers on analysis or of Posse’s comments on his work in probability. Koialovich [6, No. 3, pp. 18 – 19] regarded Florov’s program as scientifically unsatisfactory. We also note that he called mathematical statistics a shaky and poorly substantiated theory.

Later addendum: In 1910, Nekrasov asked the Minister of Public Education to be appointed unpaid member of the Ministry’s Scientific Committee. His request was refused because of the resistance of those mathematicians who already were members of this Committee. They were afraid that Nekrasov’s appointment “can lead to very undesirable conflicts [...] concerning the existing mathematical curricula”, see letter of N.Ya. Sonin to the Minister of 8 May 1910, Ross. Gos. Istorich. Arkhiv, Fond 740, inventory 43, No. 24, p. 2. The entire letter is translated in DHS 2579, 1998, pp. 101 – 102. Dr. A.L. Dmitriev (Petersburg), from

whom I received a copy of this letter, informed me that it is (mistakenly!) kept among materials concerning politically suspect academics of Tsarist Russia.

16. Nekrasov undoubtedly meant V.Ya. Tsinger, also see his Letter 12. All the scientists whom he named excepting Imshenetsky were founders of the MMS. Also see Note 17.

17. It is difficult to agree with the existence of some single Cauchy – ... – Pearson methods. True, Nekrasov applied methods of the theory of functions of a complex variable in probability (§1.1) which explains his mentioning Cauchy (but not Chebyshev or Pearson). In essence, Nekrasov repeated his then already published statement [17, pp. 10 – 11] about two directions in mathematics (where he had indeed contrasted “the ideas of Cauchy, and, on the other hand, of Bienaymé”).

The Academic Commission [10, pp. 67 – 73] established in order to discuss the teaching of mathematics in school, necessarily overstepped its terms of reference and sharply denounced both this statement and Nekrasov’s wrong understanding of the principles of mathematics (and his attacks on Chebyshev’s memoir [37] based on his mistakes). Also see Letter 5.

18. It can be thought that in Letter 4 Nekrasov described the same suggested report but entitled it differently. Later he published its “more elementary part”, cf. Letter 8. In his works that appeared from 1885 onward Nekrasov understood the term “asymptotic equivalence” of functions $f(x)$ and $g(x)$ with a continuous or discrete argument x in several senses, namely (in modern notation) $f(x) \sim g(x)$; $f(x) = O[g(x)]$; and $f(x) = o[g(x)]$. This ambiguity allowed him to formulate some prophetic statements about the possibility of considering probability as an actual infinitesimal in the spirit of non-standard analysis [38, p. 110]. See Letter 5.

19. It seems that Andreev had correctly noticed some features of the debate between Nekrasov and his opponents but had not wished to consider carefully its mathematical essence. Posse’s paper [26] was nevertheless a scientific writing devoted, in particular, to the theory of limits. Finally, Markov, who introduced a new and extremely important object, dependent random variables, into probability, was not interested either in the methods of the theory of functions of a complex variable, or in axiomatizing probability. And for a long time he was regarding the first steps of mathematical statistics with excessive suspicion.

20. Cf. §1.4 and Note 34.

21. Imshenetsky supplemented his translation of Todhunter [39] by considering the application of analysis to three-dimensional geometry and by a chapter on infinitesimals (definition; order of magnitude; equivalence) and on differentials. Todhunter’s book was an educational treatise and Imshenetsky’s chapter was naturally of a methodological rather than scientific nature. Imshenetsky (p. 450) had indeed, see below, introduced *imperfect equations* of the type $\varphi(\alpha; \beta; 0) = 0$ which replaced equations $\varphi(\alpha; \beta; \gamma) = 0$ when all the variables were infinitesimals with γ being of a higher order of magnitude than α and β whose orders coincided. The attribution of the theory of analytical functions to Lagrange and Todhunter remains on Nekrasov’s conscience.

22. On the indicated pages Boussinesq established only one principle by stating that an infinitesimal may be replaced by any other one if their ratio tended to unity. True, he additionally formulated an evident corollary.

23. The extant part of the Letter 2 contains no requests.

24. On the indicated pages Boussinesq [40, pp. 64 – 66] had not mentioned subjectivism. Nekrasov’s reference to Jakob Bernoulli is hardly convincing. We are inclined to understand subjectivism in approximate calculations as a replacement of a given function $f(x)$ by a simpler function $\varphi(x)$ that asymptotically estimates $f(x)$; $\varphi(x)$ is chosen subjectively so as to facilitate calculations. We shall also quote Ashby [42]:

The theory of systems should be based on methods of simplification, and, in essence, it represents a science of simplification. [...] I think that the science of simplification was initiated by mathematicians who study homomorphisms.

25. Nekrasov [43, 1912, p. xiv] also excused, although indirectly, “a relatively infinitesimal error against formal logic”. Markov [9, p. 28] declared that that statement “has nothing to do with common sense”. See Note 29 about the differential calculus of probability.

26. According to the context, Nekrasov meant the CLT. In any case, in 1901 he [30, vol. 22] groundlessly accused Chebyshev (and, for good measure, Markov and Liapunov as well) in that they had mistakenly understood the foundations of mathematical analysis [2, §4]. In 1898 Nekrasov [5] had not at all mentioned the Chebyshev memoir [37], and later [44, p. 24] explained that omission by his aspiration for brevity and by the fact that he had applied a more perfect method than the one used by Chebyshev. In 1900 Nekrasov publicly declared that Chebyshev’s proof of the CLT was of little value, see my paper on his work in probability (AHES, vol. 57, 2003, pp. 337 – 353 (p. 348)).

In 1898 Nekrasov was still able to write concisely, but his arguments were hardly convincing. In 1915 he devoted an article [45] making difficult reading to the memoir [37], also see Note 27, and in 1916 he [7, p. 26] repeated that Chebyshev's statement (the CLT) "is not a theorem in the strict sense but a postulate correct until finite magnitudes of probability are discussed, but having numerous exceptions otherwise". Nekrasov connected the cases in which the "postulate" failed with an actual infinitesimal probability and referred to his "rigorous proof" and to Pearson's "investigations" but had not mentioned any sources. His reference to Pearson is unconvincing, and it was the limit of the sum of variances of the studied random variables divided by their number that should not have vanished [47, p. 240; 1, §7.2]. Note that Nekrasov provided his own definition of a postulate [3, p. 54]; quite consistently, he called it a rule spoiled by exceptions.

27. Nekrasov, here [45, p. 318] and elsewhere [17, p. 12], identified a lacuna with a mosaic pattern of landownership and for some reason called the normal distribution lacunary. Earlier he [43, 1912, pp. 141 and 473] understood lacunarity as a defect (lacunarity of a law) or brevity (lacunarity of a table). Also in [45, p. 321] he explained that he had constructed the term *molar* from the Latin *molares* (a stone block).

28. A variable taking discrete values. The Russian term *varianta* is hardly used anymore although several decades ago Fichtenholz [52, §22] had applied it and referred to H.C.R. Méray (1835 – 1911) without mentioning any source. The German text of Fichtenholz used the German term *Variante*.

29. It is hardly possible to comment duly on this statement before studying in detail Nekrasov's merits in proving the CLT (which Seneta, see §1.1, described somewhat cursorily). Note, however, that Nekrasov [17, p. 3] mentioned the "Nekrasov – Pearson differential form" which had something to do with the theory of the Pearson curves. Neither that source, nor [43, 1912, pp. 519 – 520], to which he had there referred, gives any clue to the understanding what exactly had Nekrasov contributed to this form and what was the essence of Pearson's critical attitude towards the "differential calculus of probability".

30. Nekrasov's reference to the Academy's *Zapiski* is patently wrong. He could have mentioned its *Izvestia* where Markov [53] negatively, although not altogether directly, commented on all of his works in general. Nekrasov had not (and could not have) answered such a general comment.

31. Markov [20] discussed only the fundamentals of mathematical analysis and his statement about the wrong track remained unjustified.

32. Nekrasov's reference to professors and teachers and to a commission was a fabrication, pure and simple. The discussion of his reports [12; 15; 16] at the congresses took place only partly; and, in addition, it had not at all amounted to their general approval.

33. Nekrasov [17, pp. 13 and 16] also publicly accused Markov of begging the question (again without explanation). Not later than in 1935 the concept of convergence in probability, that had been applied long before that, was fully understood. In probability theory based on classical analysis debates about actual infinitesimals became since then pointless.

34. Andreev had published Imshenetsky's posthumous manuscript [55] whereas Markov [31] sharply criticized it. Then, the former [54], while recognizing the paper's incompleteness, reasonably argued that its appearance was nevertheless useful. Markov [34] however declared that his viewpoint had triumphed since the incompleteness of [55] was not denied.

Markov's opinion was apparently too formal. Here, incidentally, is Bezikovich's testimony [56, p. XIV] about Markov: His last article was the only one

Which lacks a complete solution of the formulated problem [...] He brought himself to publish it only having been afraid that he will be unable to complete it.

Finally, we disagree with Andreev in that his paper [54] "had not benefited anyone or anything". Imshenetsky's article [55] continued his previous work of 1887 – 1888 which also provoked debate where Markov, Nekrasov, and Posse et al had participated [36].

35. See §1.2.

36. The Pedagogical Museum published all these reports in 1916 as separate booklets. Neither of them concerned probability theory, but Nekrasov's contribution [7] was an exception. Also see Note 18.

37. The replacement of an infinitesimal by an equivalent magnitude, as Nekrasov himself indicated (Letter 5), is made for the sake of simplification. His idea becomes clear when recalling his pronouncement about induction, see Note 4.

38. Nekrasov apparently meant the papers by Markov [9] and Posse [26].

39. A list of members of the MMS indicating the year of their entry had been published regularly in the *Matematicheskyy Sbornik*. Already in 1913 (vol. 29, No. 1 of the periodical) from among about 90 members from Russia itself not more than five had joined the Society before Andreev (before 1873) did.

40. Nekrasov perceived the "main distortion" in that he allegedly attempted to prove mathematically the omnipotence of God whereas he stated that without faith mathematics was insufficient for that purpose. Nev-

ertheless, the Commission, in the place indicated, only referred to the bygone (and generally known) attempts to prove God's omnipotence by applying various arbitrary rules for summing diverging number series.

41. In a letter of 13 Dec. 1916 to P.A. Florensky Nekrasov argued that he conciliated mathematics with religion and politics "logically, correctly and rightfully". Both his statements were wrong, see for example the Introduction to his treatise [43, 1912].

42. Khvolson [59, p. 11] held that "we must, and we can almost only believe" in physical laws. Thus, for small values of a the law of universal gravitation cannot be empirically isolated from the family of formulas including (in usual notation) r^{2+a} . Then, he stated that hypotheses which indeed included "the veritable essence of physics as a science" (p. 13) were only based on faith (p. 12). Finally (p. 15), Khvolson declared that a number of issues including the problem of free lies beyond the province of knowledge.

43. Nekrasov publicly repeated these definitions of the "fronts" [3, p. 21]. "Mathematics and probability theory" apparently meant "mathematics including ..."

44. Cf. Note 40.

45. In 1916 Nekrasov [3, p. 54] mentioned "the main mathematical doctrine [...] of Buniakovsky and others about the trustworthiness of ancient legends ..." Actually, Buniakovsky [60, p. 326] simply stated that "the spiritual world includes such facts which do not obey physical laws". In 1908 Markov [51, 1924, p. 320] resolutely objected to this, and added later [9, p. 33] that the chapter of probability theory dealing with appraisal of testimonies was its "weakest". We shall say a few words about Markov's unsuccessful petition, in 1912, for excommunication from the Russian Orthodox Church (the Most Holy Synod resolved that he "fell away" from the Church). In his petition, Markov referred to Tolstoy who was excommunicated in 1901. A few days before his death in 1910, the Synod resolved that Tolstoy is to remain excommunicated.

46. Neither Buniakovsky, nor the other scientists (Boole, Jevons, Bertrand) could have rejected an objection only formulated in 1908. And, anyway, we think that denying atheism is as impossible as denying religious faith. Note also that Nekrasov, in the same letter, mentioned Pearson twice more. Had not he realized that Pearson, together with like-minded associates, had created the Biometric school in order to study mathematically the issues of the atheistic theory (more precisely, hypothesis) of the evolution of species?

47. Recall that Nekrasov wrote this in 1916. A much more ugly statement is to be found in his letter to Florensky of 11 Nov. 1915 (also while the war with Germany was going on):

I quite sympathize with your attempt to teach the mathematical encyclopedia at the Theological Academy. At your hands, it will differ from an encyclopedia of Markov & Co. inspired from Berlin.

The letter bears the date 1905, but Nekrasov was obviously mistaken: he also referred to a source published in 1914.

48. It is impossible to understand why humanism cannot be realistic. Nekrasov, however, apparently thought about Christian principles rather than humanism.

49. Here are the relevant statements. Pearson (translated from Russian): "To distinguish between the fields of philosophy and science means promoting obscurantism". Tsinger [63, p. 39]: "Mathematical sciences ... are very closely related to philosophy". The speeches of Davidov [64] and Bugaev [62] lack such pronouncements, but, judging by their contexts, these scientists would have hardly objected to Pearson's opinion.

50. A similar criticism is contained in Nekrasov's letter of 7 Dec. 1916 to Florensky:

Krylov [...] is elected to full membership at the Academy in spite of his scientific illiteracism [!]. He translated Newton's book ignorantly [...] and supplied his biased translation with notes in the spirit of panphysisism, i.e., in the same spirit in which academician Markov distorted the principles of the classical work of academician Buniakovsky [cf. Note 45] by his pseudo-interpretation. [...] Those who collated the authentic Newton's text with its translation also mention many philological mistakes in the latter which had distorted beyond recognition the ideas of the great scholar who believed in God and His prophets.

Three authors (N.N. Luzin, p. 54; T.P. Kravets, pp. 322 – 323; and T.I. Rainov, p. 343) of the collected articles [65] expressed an opposite opinion. True, we had not found there any detailed discussion of Krylov's work but it is quite possible that Nekrasov had exaggerated. And Krylov [66] expressly stated that he had not kept to Newton's mathematical terminology.

Nekrasov had not explained the terms *panphysisism*; *physico-mathematical realism*; *physico-mathematics* (see the same Letter 12). However, since he connected panphysisism not only with Krylov, but also with Markov, who never studied physical problems, it might be assumed that he understood that term as the ex-

planation of nature by mathematical means without turning to God. Indeed, in another letter to Florensky (1 Aug. 1916) Nekrasov argued that “the Moscow school directs the training of teachers in the spirit of panphism with an anti-Christian tinge ...” One question suggests itself: Was Laplace an panphysist? Apparently, yes.

51. Not the theory of probability but mathematical statistics is (partly) based on induction. See however Note 4. Umov was a physicist hardly connected with probability.

52. A strict bureaucratic surveillance of schools was implemented in 1866; in 1872 a “Statute concerning city schools” was adopted so as to weaken the influence of social institutions on education. A “Statute concerning primary public schools was then introduced in 1874 for guarding the school against “pernicious and ruinous influences” [67, pp. 759 – 760]. These facts do not corroborate Nekrasov’s statement which apparently reflected the weakening of the influence of faith on natural sciences.

53. We think (cf. Note 48) that, according to Nekrasov, humanism meant Christian or mystic principles. This corroborates our opinion (§1.1) that he kept to the Platonic tradition.

References

Additional abbreviation: Kazan Izv. = *Izvestia Fiz.-Mat. Obshchestvo Kazan Univ.*, 2nd ser.

1. Sheynin, O.B. (1989), Markov’s work on probability. AHES, vol. 39, pp. 337 – 377.
2. Sheynin, O.B. (1991), Markov’s letters in the newspaper “Den”. Translated in this collection.
3. Nekrasov, P.A. (1916), *Средняя школа и т.д.* (High School, Mathematics and Scientific training of Teachers). Psb.
4. Seneta, E. (1984), The central limit theorem and linear least squares in pre-revolutionary Russia. *Math. scientist*, vol. 9, pp. 37 – 77.
5. Nekrasov, P.A. (1898), The general properties of mass independent phenomena. MSb, vol. 20, pp. 431 – 442. Transl.: DHS 2579, 1998, pp. 21 – 28.
6. Nekrasov, P.A. (1915), Theory of probability and mathematics in high school. ZhMNP, 4th paging, No. 2, pp. 65 – 127; No. 3, pp. 1 – 43; No. 4, pp. 94 – 125. (R)
7. Nekrasov, P.A. (1916), *Принцип эквивалентности* (Principle of Equivalence of Magnitudes in the Theory of Limits and in Consecutive Approximate Calculus). Petrograd.
8. Markov, A.A. A question for the Ministry of Public Education. In [2].
9. Markov, A.A. (1915), On the Florov and Nekrasov project. ZhMNP, 4th paging, No. 5, pp. 26 – 34. (R)
10. Markov, A.A. et al (1916), Report of the Commission for discussing some issues concerning the teaching of mathematics in high school. *Izvestia Akad. Nauk*, vol. 10, No. 2, pp. 66 – 80. Translated: DHS 2696, 2000, pp. 95 – 108.
11. Nekrasov, P.A. (1906), *Основы общественных и естественных наук и т.д.* (Principles of Social and Natural Sciences in High School). Psb.
12. Nekrasov, P.S. (1912), On sections of mathematics necessary for economic sciences. *Matematich. Obrazovanie*, No. 2, pp. 79 – 81. (R)
13. Nekrasov, P.A. (1912), The lyceum system for connecting the education in high school and in universities as a measure for regulating our schools. Newspaper *St. Peterburgsk. Vedomosti*, 17 (30) Oct., p. 2. (R)
14. Nekrasov, P.A. (1913), The intermediate lyceum stage between high school and university. ZhMNP, 4th paging, No. 11, pp. 31 – 48. (R)
15. Nekrasov, P.A. (1915), On the educational features of the two directions of the mathematical course in high school. *Доклады второго всероссийского съезда преподавателей математики* (Reports 2nd All-Russian Congr. Teachers Math.). M., pp. 83 – 93. (R)
16. Nekrasov, P.A. (1915), The second (bachelor) stage in the future high school. *Ibidem*, pp. 175 – 181. (R)
17. Nekrasov, P.A. (1915), On Markov’s paper [9]. ZhMNP, 4th paging, No. 7, pp. 1 – 17. (R)
18. Markov, A.A. (1916), Letter to the editorial office. Newspaper *Novoe Vremia*, 7 (20) Dec., pp. 7 – 8. (R)
19. Markov, A.A. (1915), Seminarists and realists. In [2].
20. Markov, A.A. (1915), Letter to the editorial office. In [2].
21. Youshkevich, P. [S.] (1915), On a scientific polemic. IMI, vol. 34, 1991, pp. 207 – 209. (R)
22. Bortkevich, V.I. (von Bortkiewicz, L.) (1903), The theory of probability and the struggle against sedition. *Osvobozhdenie* (Stuttgart), book 1, pp. 212 – 219. Signed “B”. Bortkevich claimed his authorship in 1910 (ZhMNP, 2nd paging, No. 2, p. 353). Apparently printed only in some copies of the periodical. (R)
23. Sheynin, O.B. (1966), On the history of the iterative methods of solving systems of linear algebraic equations. Translated in this collection.
24. Markov, A.A. (1912), A rebuke to Nekrasov. MSb, vol. 28, pp. 215 – 227. Transl.: DHS 2579, 1998, pp. 77 – 82.

25. Liapunov, A.M. (1901), An answer to Nekrasov. *Zapiski Khark. Univ.*, vol. 3, 1901, pp. 51 – 63. Transl.: DHS 2579, 1998, pp. 53 – 63.
26. Posse, A.K. (1915), A few words about Nekrasov's article. *ZhMNP*, 3rd paging, No. 9, pp. 71 – 76. (R)
27. Mikhailov, G.K., Stepanov S.Ya. (1985), On the history of the problem of the rotation of a solid about a fixed point. *IMI*, vol. 28, pp. 223 – 246. (R)
28. Gordevsky, D.Z. (1955), *К.А. Андреев (Andreev)*. Kharkov.
29. Tsykalo, A.L. (1988), *А.М. Ляпунов (Liapunov)*. М.
30. Nekrasov, P.A. (1900 – 1902), New fundamentals of the doctrine of probabilities of sums and means. *MSb*, vol. 21, pp. 579 – 763; vol. 22, pp. 1 – 142, 323 – 498; vol. 23, pp. 41 – 462. (R)
31. Liapunov, A.M. (1901), Nouvelle forme du théorème sur la limite de probabilité. *Mém. [Zapiski] Akad. Nauk*, vol. 12, pp. 1 – 24.
32. Markov, A.A. (1894), Extract from letter to Andreev. *Soobshchenia Kharkov Mat. Obshchestvo*, vol. 4, No. 4, pp. 146 – 149. (R)
33. Markov, A.A. (1894), On Andreev's commentary. *Ibidem*, pp. 175 – 176. (R)
34. Markov, A.A. (1916), On an application of the statistical method. *Izvestia Akad. Nauk*, vol. 10, pp. 239 – 242. (R)
35. Nekrasov, P.A. (1915), Answer to Posse's objections. *ZhMNP*, 4th paging, No. 10, pp. 97 – 104. (R)
36. Korkin, A.N., Bobylev, D.K., Posse, A.K. (1893), Extract from letter to Moscow Math. Soc. *MSb*, vol. 17, pp. 386 – 391. (R)
37. Chebyshev, P.L. (1887), Sur deux théorèmes relatifs aux probabilités. In Russian. Translation : *Acta Math.*, t. 14, 1891, pp. 305 – 315.
38. Uspensky, V.A. (1987), *Что такое нестандартный анализ (What Is Non-Standard Analysis)*. М.
39. Todhunter, I. (1871), *On the Differential Calculus and the Elements of the Integral Calculus*. 5th edition. London – New York. Translated into Russian by V.G. Imshenetsky. Psb, 1873.
40. Boussinesq, J. (1887), *Cours d'analyse infinitésimale*, t. 1, No. 1. Paris.
41. Nekrasov, P.A. (1912), The general main method of generating functions as applied to the calculus of probability and to the laws of mass phenomena. *MSb*, vol. 28, pp. 351 – 460. (R)
42. Ashby, W. Ross (1964), [Some remarks]. In *Views on General Systems Theory*. Editor, M.D. Mesarovich. New York. Quotation in text translated from the Russian translation of this source (M., 1966, p. 177).
43. Nekrasov, P.A. (1888 – 1912). *Теория вероятностей (Theory of Probability)*. М., 1896; Psb, 1912. Lithogr. Editions: 1888 and 1894.
44. Nekrasov, P.A. (1899), On Markov's article [48] and on my report [5]. *Kazan Izv.*, vol. 9, No. 1, pp. 18 - 26. Transl.: DHS 2579, 1998, pp. 29 – 33.
45. Nekrasov, P.A. (1915), Interpretation of Chebyshev's second theorem and its versions. *MSb*, vol. 29, pp. 315 – 343. (R)
46. Markov, A.A. (1898), Sur les racines de l'équation etc. *Izvestia Akad. Nauk*, vol. 9, pp. 435 – 446.
47. Markov, A.A. (1951), *Избранные труды (Sel. Works)*. N.p.
48. Markov, A.A. (1899), The law of large numbers and the method of least squares [47, pp. 233 – 251]. Transl.: DHS 2514, 1998, pp. 157 – 168.
49. Markov, A.A. (1899), Application of continuous fractions to calculating probabilities. *Kazan Izv.*, vol. 9, No. 2, pp. 29 – 34. (R)
50. Markov, A.A. (1899), Answer to [44]. *Kazan Izv.*, vol. 9, No. 1, pp. 41 - 43. Transl.: DHS 2579, 1998, pp. 35 – 36.
51. Markov, A.A. (1900 – 1924), *Исчисление вероятностей (Calculus of probability)*. Psb, 1900, 1908, 1913. М., 1924. German transl. of 2nd edition: 1912.
52. Fichtenholz, G.M. (1971), *Differential- und Integral-Rechnung*, Bd. 1. 6. Aufl. (1959, in Russian). Berlin.
53. Markov, A.A. (1910), Correcting an inaccuracy. *Izvestia Akad. Nauk*, vol. 4, p. 346. (R)
54. Andreev, K.A. (1894), A commentary on Imshenetsky's article [55]; an answer to Markov's criticism [32]. *Soobshchenia Khark. Matematich. Obshchestvo*, vol. 4, No. 4, pp. 150 – 160. (R)
55. Imshenetsky, V.G. (1894), Comparison of Bugaev's method of discovering rational fractional solutions of differential equations with other methods. *Ibidem*, No. 1 – 2, pp. 60 – 80. (R)
56. Bezikovitch, A.S. (1924). Biographical essay [on Markov]. [51, 1924, pp. iii – xiv]. (R)
57. Barbèra, L. (1876), *Teorica del calcolo delle funzioni*. Bologna. We have not seen it.
58. Newton, I. (1687), [*Mathematical Principles of Natural Philosophy*]. Russian transl. from the Latin original by A.N. Krylov (1915 – 1916) in his *Собрание трудов (Coll. Works)*, vol. 7. М. – Л., 1936.

59. Khvolson, O.D. (1916), *Знание и вера в физике* (Knowledge and Faith in Physics). Petrograd.
60. Buniakovsky, V.Ya. (1846), *Основания математической теории вероятностей* (Principles Math. Theory of Probability). Psk.
61. Pearson, K. (1892), *Grammar of Science*. London. Russian transl.: Psk, 1911.
62. Bugaev, N.V. (1868), Mathematics as a scientific and pedagogical tool. MSb, vol. 3, 2nd paging, pp. 183 – 216. Translation of fragment to appear in 2004 in the DHS series.
63. Tsinger, V.Ya. (1874), Exact sciences and positivism. In *Отчет и речи и т.д.* (Report and Speeches, Grand Meeting Mosc. Univ. 1874). M., pp. 38 – 98 of second paging. (R)
64. Davidov, A.Yu. (1857), Theory of mean values with application to compiling mortality tables. Ibidem, first paging. (R)
65. *Сборник статей и т.д.* (Coll. Articles Honoring 300 Years from Newton's Birth). M. – L., 1943.
66. Krylov, A.N. (1916), *Учение о пределах и т.д.* (Doctrine of Limits As Explicated by Newton). Petrograd.
67. Falbork, G., Charnolusky, V. (1897), Primary public education. *Энци. словарь Брокгауз и Ефрон* (Brockhaus & Efron Enc. Dict.), vol. 40, pp. 728 – 770. (R)
68. Sheynin, O.B. (1996), *Chuprova* (1990, in Russian). Göttingen.

15. The Notion of Randomness from Aristotle to Poincaré

IMI, vol. 1 (36), No. 1, 1995, pp. 85 – 105

Note

Complying with a request made by a French colleague, I left him the manuscript of this paper for publication in a reputed periodical. Then, suddenly, it appeared elsewhere (*Mathématiques, informatique et sciences humaines* No. 114, 29^e année, 1991, pp. 41 – 55). I consider all this as a piracy and am reprinting my manuscript here. Its Russian version has since appeared, see Contents.

*

*

*

1. Introduction

Aristotle and even earlier scientists and philosophers attempted to define, or at least to throw light upon randomness, and in jurisprudence, about two thousand years ago, it was indirectly recognized in an ancient Indian book of instructions [1, §108] which determined the behavior of man both at home and in social life.

In §2 I sketch the attempts to direct the concept of randomness into the realm of mathematical science; in §§3 – 10 I dwell on various interpretations of randomness that were being pronounced in natural sciences and philosophy; my §11 is devoted to the interrelation between necessity and randomness; and, finally, in §12, I formulate my conclusions. Since **Aristotle**, **Darwin** and **Maxwell** described (used; indicated) various aspects of randomness, I repeatedly mention each of these great scholars. The history of randomness is especially interesting since the new approach to its understanding that had recently took shape in physics and mechanics has affected the fundamentals of these sciences [2].

I have examined the work of Poincaré, who paid much attention to randomness, elsewhere [3]. Here, I only mention that he directly linked chance to instability of motion [of the solution of differential equations] and introduced the fruitful method of arbitrary functions [4, pp. 88 – 89]. Then, there is a case for studying the attitude of ancient scientists preceding Aristotle towards randomness. However, my own experience [5, §§2.1 and 2.3] is that this topic is extremely difficult since their thoughts may be interpreted in different ways. Finally, I restrict my paper with the fields of mathematics and natural sciences.

There is no general literature on my subject; one author [6] had discussed randomness from a different point of view, some other [7 – 13] busied themselves with its particular issues; I mention contributions [7 – 9] in the sequel. I myself touched the same topic in many articles published in the *Archive for History of Exact Sciences* and my excuse for doing so, and for returning to the same subject in an ad hoc paper, is that it is patently impossible to compile a contribution such as this one all at once.

2. Mathematics and the Concept of Randomness

Lambert [14, pp. 238 – 239; 15, p. 246; 5, pp. 136 – 137] made an endeavour to formalize randomness. His interest in this problem may be explained by the fact that he was the first follower of Leibniz in attempting to create a doctrine of probability belonging to a general science of logic. Lambert's efforts, founded on an intuitive notion of normal numbers, was ahead of its time. True, **Cournot** [16, pp. 57 – 58] and **Chuprov** [17, p. 188] had noted Lambert's efforts, but no-one became interested in their accounts.

Poisson [18, pp. 140 – 141] hesitatingly offered a definition of a random magnitude as a variable that assumed different values with corresponding probabilities. His definition (independently re-introduced at the end of the 19th century [19, §15.4]) went unnoticed. Poisson [18, p. 80] also attempted to state the nature of chance. Randomness, he argued, was an *ensemble* of causes that produced an event without altering its (the event's) chances of happening or failing. His idea seems unsuccessful, but at least Poisson thus maintained that random events possessing stable probabilities of two possible outcomes do occur.

While attempting to construct the theory of probability anew, **von Mises** [20a, p. 62] introduced his celebrated concept of *Kollektiv* (of an infinite random sequence) and demanded that the order in which its elements followed each other be random (“mit zufallsartiger Zuordnung ...”). Later he [20b, 1939, p. 32] equated chance with “complete ‘lawlessness’”, cf. §6, and (Ibidem, p. 133) noted its “fundamental importance” for the theory of probability. For an evaluation of his efforts see [4]. I only remark here that mathematicians became interested in defining the *collective* and attempts of such kind are now continued in the modern theory of algorithms. Three approaches are now recognized [22, pp. 199 – 214]. The *frequency* viewpoint had originated with Mises (even with **Lambert**) and in 1963 **Kolmogorov** modified it. It demands that the various elements of a random sequence and of its *legitimate* subsequences appear with stable frequencies. According to the approach founded on *complexity* (Kolmogorov 1963), the entropy of the initial part of a random sequence should be sufficiently large. The main idea of the *quantitative* standpoint (**Martin-Löf**, 1966) is that a random sequence may only have a small number of regularities, and, therefore, that it should pass certain tests. It is easy to see that these approaches are not independent. In 1963 Kolmogorov additionally outlined the concept of a finite random sequence; according to his opinion, a finite sequence is the *more random* the more complex is the law that describes it. Quite recently there appeared another Russian paper

[23] on the same subject with no reference being provided to the previous one. As in the case of ref. [22], Uspensky was its coauthor, but this time, the second, or, rather, the first coauthor was Kolmogorov himself.

3. Randomness Does Not Exist

Such was the standpoint of the most eminent thinkers and scholars who believed that semblance of randomness resulted from ignorance of the relevant causes. Sambursky [9, pp. 40 – 41] described the utterances of ancient Greek authors on that subject, and Kendall [8, p. 11] studied similar ideas due to **St. Augustine, Thomas Aquinas, Spinoza** and **Dalembert**. In turn, I discuss the thoughts of several scientists without dwelling on the writings of **Bentley** [24, pp. 316 – 318], who somewhat verbosely explicated **Newton's** point of view, or **Lamarck** [25, pp. 74 and 97; 26, p. 329]. Here are the statements of **Kepler** [27], **Laplace** [28, p. 145] and **Darwin** [29, p. 128], in that order.

1. Chance is an idol, an abuse of God Almighty.
2. Chance is only ignorance of the connections between phenomena.
3. That chance occasions variations between individuals is wrong, but this expression “serves to acknowledge [...] our ignorance” of the relevant causes.

Kepler, however, was unable to deny that the eccentricities of the planetary orbits were random (§5). **Newton** left two pronouncements [30, Query 31; 31, p. 49] which testify that he attached certain importance to chance and to which I return in §§7 and 8:

Blind chance could never make all the planets move one and the same way in orbs concentrick, some inconsiderable irregularities excepted, which may have risen from the mutual actions of comets and planets upon one another, and which will be apt to increase, till this system wants a [divine] reformation. Such a wonderful uniformity in the planetary system must be allowed the effect of choice. And so must the uniformity in the bodies of animals.

Did blind chance know that there was light and what was its refraction, and fit the eyes of all the creatures after the most curious manner to make use of it?

A similar utterance is in [32, p. 544], and another one, formulated in about 1715, in [33, p. 127].

Lamarck [34, p. 450] thought that variations between individuals came into existence because of random causes and the Darwinian theory hinged in its entirety on the action of these same causes. It is extremely strange that, in spite of his own statistical explanation of the second law of thermodynamics, **Boltzmann** failed to recognize either the latter fact or the importance of randomness in nature [35, §4.3]. Finally, I return to Laplace in §5.

4. A Possibility

Randomness is a possibility. This definition goes back to **Aristotle** [36, 1064b – 1065a], who, moreover, apparently believed that a chance event had a logical or subjective probability lower than $1/2$. Similarly, **Thomas Aquinas** [37, vol. 19, p. 297] supposed that random events “proceed from their causes in the minority of cases ...”

The followers of the Indian teaching of Syadvada, that existed as early as in the 6th century B C, studied the concepts of the possible, the indeterminate, etc. Mahalanobis [38] maintained that this doctrine was interesting for the history of statistics. He had not mentioned randomness, but I believe that the Syadvada indirectly recognized it as a possibility.

Darwin [39, vol. 1, p. 449], drawing on stochastic calculations made at his request by **Stokes**, decided that a particular deformity in man was passed from parent to child and did not occur by chance [was not merely possible]. **William Herschel** [40, p. 577] and **Struve** [41, Note 72] left room for randomness of this kind. In their models of the stellar system they only restricted the distances of the stars without indicating their actual position. **Maxwell** [42, p. 274] remarked that neither the form and dimensions of the planetary orbits, nor the size of the Earth were determined by any law of nature [that the relevant magnitudes might have been different]. He had not mentioned randomness and his remark had to do with yet another interpretation of chance (§6).

Hegel [43, p. 383], in addition to understanding randomness as a possibility, formulated the converse proposition:

Das Zufällige ist ein Wirkliches, das zugleich nur als möglich bestimmt [...] was möglich ist, ist selbst ein Zufälliges.

It is easy to illustrate this proposition. If a random variable X takes values x_i with probabilities p_i , $i = 1, 2, \dots, n$, then any possible x_i is random in a sense that it occurs with probability p_i . Note that Aristotle had not connected any definite probabilities with the possible values of x_i , $i = 1, 2$.

5. Deviation from Laws of Nature

Randomness occurs when the purpose of nature is not attained, when hindering causes corrupt the operations of nature. This explanation is due to **Aristotle** [44, 199b] who thought that nature's accidental mistakes brought about the appearance of monsters and that the birth of female animals was the first departure from the *type*, and, at the same time, a *natural necessity* [45, 767b]. His statements were the first to confront necessity and randomness. Indeed, the occurrence of monsters accompanies the necessary acts of regular births whereas the birth of a female is, according to Aristotle, both necessary and random. From a modern point of view the second example is wrong; and it hardly corresponds to his own belief (§4) that the probability of a chance event is lower than $1/2$. Referring to the Philosopher, **Thomas Aquinas** [37, vol. 19, p. 489] pointed out that the birth of a girl was a random event.

Kepler [46, p. 244; 47, p. 932] suggested that only "zufällig" perturbations had forced the planets to deviate from circular motion. True, he also stated that the eccentricities regulated the planets' motions [48, p. 317], but he was naturally unable to say why the eccentricity of a given planet had a particular value rather than any other one. **Kant** [49, p. 337] repeated Kepler's pronouncement on the elliptic paths of their planets. **Lamarck** [26, p. 133] maintained that there existed deviations from the divine lay-out of the tree of animal life and explained them by the action of a "cause accidentelle et par conséquent variable".

The pronouncements described above pertained to determinate laws of nature. However, many natural scientists, while making similar statements, actually thought about mean states. **Adanson** [50, p. 48] regarded intraspecific variations as digressions from the divine order and believed them necessary "pour l'équilibre des choses". **Lamarck** [51, p. 76] argued that "plusieurs causes", some of them "variables, inconstantes et irrégulières dans leur action", corrupted [determined!] the [mean] state of the atmosphere. **Humboldt** [52, p. 68] conditioned the study of all natural phenomena by discovering the appropriate mean values (mean states). As early as in 1817 he isolated climatology from meteorology [53]. His point of view was not, however, quite consistent in that he had not linked his definition of climate [54, p. 404] with mean states, but at least later scholars improved on him [55, p. 296].

De Moivre [56, p. 253] declared that the value of the parameter of the binomial distribution of male and female births was of divine origin. Quite logically, he regarded as random only the deviations of the number of male (say) births from the corresponding number determined by the binomial law. Random, in modern notation, for De Moivre was not X itself, but rather $(X - EX)$. He (Ibidem, p. 251) also argued that

In process of Time, Irregularities [produced by chance] will bear no proportion to the recurrency of that Order which naturally results from Original Design.

And De Moivre [57, p. 329] effectively declared that the aim of the theory of probability was to isolate chance from divine design [from purpose], and thus came close to another understanding of randomness (§6). Being greatly influenced by **Newton**, to whom he devoted the first edition of his book [57], De Moivre had not nevertheless repeated the former's inference on the need for divine reformation (§3).

Similarly, for **Laplace** the theory of probability belonged to natural sciences rather than to mathematics, and its goal was not to study mathematical objects (for example, densities), but the discovery of the laws of nature. He therefore stood in need of analyzing observations, of eliminating randomness from them, of separating chance from law.

6. Lack of Purpose

Randomness is lack of divine law or goal; it occurs when independent chains of events intersect each other. Again, randomness is lack of purpose, and perhaps, "uniformity" (§7) as well. It was in this sense that chance was understood in ancient India, about two thousand years ago, although not in natural sciences, but in civil life [1, §108]: If, shortly after giving evidence at a trial, a misfortune befell the witness or his family, it was believed that God punished him [that the evil had not happened without purpose, i.e., not by chance].

6.1. Lack of Law Or Goal

According to **Aristotle**, an unexpected meeting of two people [36, 1025a] or a discovery of a buried treasure [44, 196b] are chance events. Each of them could have been (but was not) aimed at. **Junkersfeld** [7, p. 22], who considered numerous examples contained in the great scientist's work, inferred that he would not have thought that coming across a stranger or finding a rusty nail were random.

The ancient Indian *Yadrichchha* or *Chance* theory contained a similar interesting illustration of randomness [58, p. 458]:

The crow had no idea that its perch would cause the palm-branch to break, and the palm-branch had no idea that it would be broken by the crow's perch; but it all happened by pure chance.

These examples show that the interpretation of chance as an intersection of chains of events was known even in antiquity. In this connection **Cournot** [16, p. 56] had quoted **Boethius** and Bru [59, p. 306] noticed that Cioffari [60, pp. 77 – 84] had discussed or reproduced appropriate passages from several ancient scholars. **Hobbes** [61, p. 259] maintained that a traveller “meets with a shower” by chance since “the journey caused not the rain, nor the rain the journey”. Much the same was the opinion of many modern scientists [5, p. 133] and of course in any reasoning of this kind the interpretation mentioned above simply suggests itself. **Darwin** [62, p. 395] argued that he had used the word chance only in relation to purpose [to lack of purpose] in the origination of species. He continued: “the mind refuses to look at [the universe] as the outcome of chance, – that is, without design or purpose”.

The **Dalembert – Laplace** problem merits special attention. The word *Constantinople* is composed of separate letters; is it possible that the choice and arrangement of the letters were random? Dalembert [63, pp. 245 – 255], who questioned the fundamentals of the theory of probability, maintained that all arrangements of the letters were equally probable only from the mathematical point of view but not in reality. Laplace [28, p. 152; 54, p. XV] came to a different conclusion: Since the word had a certain meaning [answered a particular purpose], the composition was not likely at all to have been accidental [aimless].

This reasoning helps to understand properly a number of earlier pronouncements. **Aristotle** [65, p. 289b] believed that it was impossible for the stars to move independently one from another [to move at random] and yet to remain fixed, – they possessed common motion. A similar idea can be traced in the theory of errors. A large deviation of an observation from the appropriate arithmetic mean had rather been assigned to a special reason (though not to a goal, or a law, but to a blunder) than attributed to an unlikely combination of admissible and mutually independent [accidental] errors. Note, however, that observational errors hardly belong to natural sciences.

Kepler [66, p. 337] thought that a possible (a chance, see §4) appearance of a new star in a definite place and on a particular date was so unlikely that it had to be occasioned on purpose. By implication, he believed that each place (and date) was equally probable. Thus, Kepler understood randomness not only as lack of purpose, but as something [aimlessly] possible (§4), and, at the same time, as uniform (§7).

6.2. Intersection of Chains of Events

Randomness is an intersection of such chains. This interpretation is due to **La Placette** [67, last page of Preface] who devoted his book to proving that games of chance were not contrary to Christian ethics. He contended that “le Hasard renferme [...] un concours de deux, ou de plusieurs événements contingents”. Each event had its own cause, the author continued, but we did not know why they coincided. La Placette had not explained randomness; his definition amounted to saying that the cause of any chance event was unknown (cf. §3).

Cournot [59, §40; 16, p. 52] took up La Placette's idea and in one instance [16, p. 57] referred to him. Cournot [59] initially mentioned chains of determinate events thus improving on his predecessor:

Les événements amenés par la combinaison ou la rencontre de phénomènes qui appartiennent à des séries indépendantes, dans l'ordre de la causalité, sont ce qu'on nomme des événements fortuits ...

In his later work Cournot [16] regrettably omitted the phrase “dans l'ordre de la causalité”. He [59, §§41 – 48] apparently thought of using his definition of randomness to present the theory of probability as a science of chance events. He could not have succeeded; what was really needed was a systematic use of the notions of random variable (cf. §2) and of its expectation and variance.

7. Uniformity

Randomness is something uniformly possible, it can occur in one out of several equally possible ways.

7.1. Uniform Randomness

In §6.1 I stated that **Kepler** had equated chance with *uniform randomness*. This attitude was characteristic of natural scientists for about two centuries. **Arbutnot** [68], in attempting to explain the prevalence of boys among the newly-born, contrasted uniform randomness and design without thinking of other possible laws of randomness. The same kind of comparison is implied in both of **Newton's** pronouncements (§3).

Jakob and **Niklaus Bernoulli** and **De Moivre** introduced the binomial distribution into the theory of probability; in spite of that, the former understanding of randomness persisted. **Boyle** [69, p. 43], indicating that a

chance composition of a long sensible text was impossible, declared that the world could not have been created randomly. The first part of his statement is also contained in the *Logique de Port-Royal* [70, Chapt. 16]. **Kant** [49, p. 230] and **Voltaire** [71, p. 316] maintained that a uniformly random origin of organic life was even less possible than a similar origin of the system of the world. **Daniel Bernoulli** [72] and **Laplace** [73], likely following **Newton**, calculated the probability that the regularities observed in the Solar system were due to randomness but they only contrasted blind chance and a determinate cause.

Maupertuis [74, pp. 120 – 121] indicated that the seminal liquid “de chaque individu” most often contained parties similar to those of its parents. He also mentioned rare cases when a child resembled one of his remote ancestors (p. 109) as well as mutations (a later term) (p. 121). It could be thought that Maupertuis recognized randomness with a multinomial distribution, but he was not consistent. While discussing the origin of eyes and ears in animals, he [75, p. 146] restricted his attention to comparing “un attraction uniforme & aveugle” and “quelque principe d’intelligence” (and came out in favor of design).

In the 19th century, many scientists, imagining that randomness was only uniform, refused to recognize the evolution of species. While illustrating that idea, both the astronomer **John Herschel** [76, p. 63] and the biologist **Baer** [77, p. 6] mentioned the philosopher depicted in the *Gulliver’s Travels* [but borrowed by Swift from Raymond Lully, 13th – 14th centuries]. Hoping to get to know all the truths, that good-for-nothing inventor was putting on record each sensible chain of words that happened to appear among their uniformly random arrangements.

Also in the 19th century, **Boole** [78, p. 256] argued that the distribution of stars was random, if, owing to the ignorance of the relevant law, “it would appear to us as likely that a star should occupy one spot of the sky as another” (cf. §3). And he continued: “Let us term any other principle of distribution an indicative one”. Even in 1904 **Newcomb** [79, p. 13] called the uniform distribution of stars “purely accidental”. Recalling the definition of a finite random sequence as outlined by Kolmogorov (§2), and bearing in mind that the number of stars of the first few magnitudes is finite, I note, however, that Boole’s and Newcomb’s inferences were quite modern.

The following examples that have to do with finite populations of stars or atoms are similar. Nevertheless, in these instances natural scientists reasonably believed that uniform randomness represented a statistical law of nature. Thus, **Forbes** [80, p. 49] contended that

An equable spacing of stars [...] [was] far more inconsistent with a total absence of Law or Principle, than the existence of [regions of condensation and paucity] of stars.

He [80, 1850, p. 420] also asked which distributions might be called random [as not representing any law, cf. §6.1].

In 1906 **Kapteyn** [81, p. 400] declared that

The peculiar motions of the stars are directed at random, that is, they show no preference for any particular direction.

Struve [82, pp. 132 – 133] pronounced a similar weaker statement even in 1842. **Boltzmann** [83, p. 237; 84, p. 321] held that gas molecules move with equal probability in whichever direction, but did not mention randomness.

Sometimes chance might have been connected with the state of chaos, i.e., with the absence of any law of distribution. Since this possibility was hardly discussed before the 19th century, I believe that either no-one considered it, or, in any case, that it gradually gave way, perhaps unjustly, to uniform randomness. In those times, apparently only **De Moivre** [56, pp. 251 – 252] mentioned chaos, but even he dismissed it out of hand. “Absurdity follows”, he declared, while considering one or another value of the parameter of the binomial distribution, if a certain event happened not

According to any law but in a manner altogether desultory and uncertain; for then the Events would converge to no fixt Ratio at all.

And, when introducing his definition of probability as the limit of statistical frequency, **von Mises** [19, p. 60] effectively excluded chaos.

Against the background of the abovementioned examples, it is interesting to name two philosophers of the 18th century who expressly indicated that non-uniform randomness was indeed possible. **Hume** [85, vol. 1, p. 425], while discussing chance events, illustrated his ideas by considering an imaginary die having “four sides marked with a certain number of spots, and only two with another”. He had not however referred to any law of nature. **D’Holbach** [86, pt. 2, pp. 138 – 139] maintained that the *molecules* of various bodies greatly dif-

ferred one from another and combined with each other in diverse ways. He compared them with “dice pipées [...] d’une infinite de facons différentes” [with irregular dice].

7.2. Specifying Particular Problems

During the 19th century, it gradually became clear that the concept of uniform randomness in general was not sufficiently intelligible. The problem of determining the distance between two random points (A and B) on a sphere is highly relevant since **Laplace** [87, p. 261] and **Cournot** [59, §148] understood it in different senses. Laplace believed that B was, with equal probability, any point of the great circle AB whereas Cournot’s solution implied that equally probable were all possible situations of B on the sphere. Similarly, **Daniel Bernoulli** had calculated the probability that the planes of the planetary orbits were close to each other due to uniform randomness (cf. §6.1), but Todhunter [88, §396] remarked that it would have been more natural to consider uniform randomness with respect to the closeness of the poles of the orbits.

Darwin [89, pp. 52 – 55] attempted to ascertain whether earth worms carrying small objects into their burrows seize “indifferently by chance” any part of their find. He considered four versions of such randomness with regard to the manner of capturing paper triangles strewn about on the ground. After calculating the appropriate frequencies, Darwin decided that the worms carried the triangles non-randomly, i.e., to a certain extent sensibly. Considering non-randomness on a par with reason, he therefore recognized chance as lack of purpose; in §6.1 I have mentioned him exactly in this connection.

Bertrand [90, pp. 6 – 7] took up the problem of calculating the distance between *random* points on a sphere. Without mentioning Laplace or Cournot, he repeated their solutions and concluded that both were correct. In addition, Bertrand maintained that not only small distances but other geometric features as well might be used to characterize an unlikely scatter of the stars over the sky. He hardly knew about Darwin’s experiment, but he provided a few more examples including his celebrated problem on the probability of the length of a *random* chord of a given circle. He thus proved that uniform randomness was not definite enough and justly insisted that in particular instances that concept be specified.

8. Instability of Motion

Randomness is instability of motion, or of initial conditions; it involves slight causes leading to considerable consequences. **Galen** [91, p. 202], without mentioning randomness, asserted that “in old men even the slightest causes produce the greatest change”. According to **Newton** (§3), the accumulation of irregularities in the planetary system may be interpreted as an action of slight causes giving rise to considerable effects (true, only gradually). Many examples from §6.1 can be also considered in this connection.

Maxwell [92, p. 366] prophetically argued that physicists will study “irregularities and instabilities” and thus move away from mere determinacy. Illustrating his idea, he mentioned unstable refraction of rays within biaxial crystals (p. 364). Maxwell thus connected randomness with instability but had not said so directly. He expressed similar thoughts elsewhere [93, pp. 295 – 296]:

There is a very general and very important problem in Dynamics [...] It is this: “Having found a particular solution of the equations of motion of any material system, to determine whether a slight disturbance of the motion indicated by the solution would cause a small periodic variation, or a total derangement of the motion ...”

Von Kries [94, p. 58], while discussing the game of roulette, noted that

Eine kleine Variirung der Bewegung hinreichend, um an Stelle des Erfolges Schwarz der Erfolg Weiss herbeizuführen ...

His remark was not, however, convincing: the slight variation of the motion could have resulted, first and foremost, in changing the number of revolutions travelled by the ball.

Pirogov [95, p. 518] called an event random if its dependence on the relevant causes was complicated and “mit Hülfe von nur analytischen Functionen gar nicht ausgedrückt werden kann”. His utterance may be considered as another hint at the connection between chance and instability. As to complicated causes, see §9.

As stated in §1, I am not discussing the work of Poincaré, but at least I emphasize that he was the first to say expressly that randomness is instability of motion.

9. Complicated Causes

Randomness occurs when complicated causes are involved. In a heuristic sense **Leibniz** [96, p. 288] anticipated this explanation by declaring that the “zufällige Dingen” were those “deren vollkommener Beweis jeden endlichen Verstand überschreitet”. While formulating his celebrated law of the velocities of gas mole-

cules, **Maxwell** [97] reasonably supposed that the distribution sought sets in “after a great number of collisions among a great number of equal particles”. He had not mentioned randomness. Elsewhere he [98, p. 436] remarked that the motion of heat was “perfectly irregular” and that the velocity of a given molecule could not be predicted. Once more, he did not mention randomness, and he said nothing about complicated causes. I have adduced his second pronouncement since it supplements his previous idea. Note that he actually rejected Laplace’s famous declaration [64, p. VI] on the possibility of calculating the future states of the universe.

10. Slight Causes Leading to Small Effects

Randomness occurs when slight causes lead to small effects. Laplace [99, p. 504] qualitatively explained the existence of trifling irregularities in the system of the world by the action of countless [small] differences between temperatures and between densities in the diverse parts of the planets. He had not mentioned randomness. **Kepler** and **Kant** (§5) referred in similar cases to deviation from purpose.

11. Necessity and Randomness

In discovering laws and regularities of nature and in studying its mean states, scientists determined necessity. Besides that, they often revealed, or even attempted to isolate, the unavoidable accompanying phenomena of the second order, i.e. randomness. And it was exactly in this manner that many natural scientists imagined the relation between necessity and chance. Recall in this connection **Aristotle’s** opinion (§5) on the appearance of monsters, **Kepler’s** reasoning on the eccentricities of the planetary orbits (§5), **Newton’s** thoughts (§3, also see below) on the planetary system, **Lamarck’s** utterance (§5) on the tree of animal life, **De Moivre’s** reasoning (§5) on the sex ratio at birth as well as the isolation of climatology from meteorology achieved by **Humboldt** (§5) and **William Herschel’s** and **Struve’s** models of the stellar system (§4).

Lamarck’s pronouncement [26, p. 169] merits special attention. He apparently believed that necessity and chance were the two main *moyens* of nature. Without proving anything or providing any example, he declared that these “moyens puissans et généraux” were universal attraction and a repulsive molecular action “qui [...] varie sans cesse ...” He also argued that the

Equilibre entre ces deux forces opposées [...] naissent [...] les causes de tous les faits que nous observons, et particulièrement de ceux qui concernent l’existence des corps vivans.

Lamarck likely supposed that the molecular action was random since elsewhere (see §5) he maintained that by definition accidental causes were variable.

Without dwelling on the statistics of marriages, suicides, crime, etc that reveals laws in apparently free (random) behavior of man, I note that **Kant** [100, p. 508] compared the chance birth of a man with the stability of the birth-rate:

Der Zufall im Einzelnen nichts desto weniger einer Regel im Ganzen unterworfen ist ...

Only **Hegel**, after offering his definition of randomness (§4), formulated a proposition on the unity [on the interdependence] between necessity and chance. Exactly this unity, he [43, p. 389] declared, “ist die absolute Wirklichkeit zu nennen”. **Engels** [101, p. 213] approvingly called this thesis utterly unheard of and urged scientists to study both necessity and chance. It was **Poincaré** [102, p. 1], however, who provided the most important statement:

Dans chaque domaine, les lois précises ne décidaient pas de tout, elles traçaient seulement les limites entre lesquelles il était permis au hasard de se mouvoir. Dans cette conception, le mot hasard avait un sens précis, objectif ...

A few words about the theory of probability. At the end of §5 I mentioned **De Moivre** and **Laplace** in connection with the aim of that scientific discipline. They entrusted the theory with delimiting randomness from necessity. In our days, the same goal is being achieved by mathematical statistics created since then. **Pearson** [103] remarked that the development of the theory of probability was much indebted to **Newton**; I shall show that he thought about the great scientist’s idea on the relation between necessity and chance. Here are his words:

Newton’s idea of an omnipresent activating deity, who maintains mean statistical values, formed the foundation of statistical development through Derham, Süßmilch, Niewentyt, Price to Quetelet and Florence Nightingale. [...] A. De Moivre expanded the Newtonian theology and directed statistics into the new channel

down which it flowed for nearly a century. The causes which led De Moivre to his Approximatio [56] [where the normal approximation to the binomial distribution was first discovered] or Bayes to his theorem were more theological and sociological than purely mathematical, and until one recognizes that the post-Newtonian English mathematicians were more influenced by Newton's theology than by his mathematics, the history of science in the 18th century, – in particular that of the scientists who were members of the Royal Society – must remain obscure.

Since Newton never mentioned the maintaining of mean values, I believe that Pearson actually thought about divine reformation, necessary, according to Newton (§3), for neutralizing the propagation of chance corruptions in the Solar system, for preserving the mean states. Thus, Pearson suggested that Newton's theologically formulated idea concerning the relation between necessity and chance had served as a basis for the development of the theory of probability. Pearson's general statement about the science in the 18th century may be specified. First, he apparently bore in mind **Laplace** (end of §5 and above); second, restricting my attention to the theory of probability, I note that Pearson [104, §§10.1 – 10.2] put forward plausible arguments in favor of the thesis on Newton's influence on **Bayes** (and **Price**, who communicated and inserted comments in the Bayes memoir).

12. Conclusions

The denial of randomness (§3) was only formal and nowadays seems to be deservedly forgotten. Possibility (§4) found its way into laws and empirical regularities, but it was **Hegel** who declared that randomness was a possibility, and, moreover, that the possible was random. Chance as deviation from laws of nature (§5) is recognized as a perturbation (a noise) and natural scientists admitted that it indeed was corrupting the laws. As far as the deviations obey the preconditions of the central limit theorem, this randomness is normal. Randomness as lack of law or purpose (§6) may be interpreted as an intersection of independent chains of events. The definitions of §§4 and 6, while reflecting different heuristic features of randomness, essentially coincide. Randomness is a random variable having a uniform distribution (§7), i.e., it is a special case of the possible (§4). Therefore, this *uniform* randomness characterizes lack of determinate law or purpose (§6); at the same time, in some instances it signifies the existence of a special statistical law of nature.

Randomness is occasioned by instability (§8) and/or complicated causes (§9). It can also occur in the context of slight causes leading to slight effects (§10). This case partly includes deviations from the laws of nature (§5), as in meteorology and astronomy. The joint action of a large number of such causes can lead to random variables with a normal distribution (above).

It is scarcely possible to comprehend randomness without studying its interconnection with necessity. Hegel stated that these concepts were united. However, even after Hegel scientists had been recognizing randomness only as a phenomenon of the second order accompanying the main event, necessity (§11). Until the mid-19th century necessity (divine design) had been contrasted only with blind chance (uniform randomness, §7).

The explanations and definitions of chance (§§4 – 10) are heuristically connected with the modern interpretations of randomness (§2). Thus, §9 is closely linked with the complexity approach and to a lesser degree a similar link seems also to apply to §8; §7 illustrates a particular instance of the frequentist approach and the rest of these sections at least do not contradict the quantitative approach. Finally, I note that §§4 – 6 and 10 are linked with §11.

Acknowledgements. B. Bru, B.V. Chirikov and V.V. Nalimov sent me reprints/copies of their contributions or other materials. M.V. Chirikov, a relative of B.V., pointed out a number of mistakes and ambiguities in a preliminary text of this paper. A. Kozhevnikov and E. Knobloch gave me editorial advice, and A.P. Youshkevich counselled me to explain the modern point of view on randomness (§2).

References

1. *Laws of Manu*. Oxford, 1886.
2. Chirikov, B.V. The nature of the statistical laws of classical mechanics. In *Методологические и философские проблемы физики* (Methodological and Philosophical Issues of Physics). Novosibirsk, 1982, pp. 181 – 196.
3. Sheynin, O.B. Poincaré's work on probability. AHES, vol. 42, 1991, pp. 137 – 171.
4. Khinchin, A.Ya. On the Mises frequentist theory; posth. publ. by B.V. Gnedenko. *Voprosy Filosofii*, 1961, No. 1, pp. 91 – 102 and No. 2, pp. 77 – 89. English transl. to appear in *Science in Context*, 2004.
5. Sheynin, O.B. On the prehistory of the theory of probability. AHES, vol. 12, 1974, pp. 97 – 141.
6. Nalimov, V.V., *Язык вероятностных представлений* (Language of Stochastic Concepts). М., 1976.
7. Junkersfeld, J. *The Aristotelian – Thomistic Concept of Chance*. Notre Dame, Indiana, 1945.
8. Kendall, M.G. The beginnings of a probability calculus. *Biometrika*, vol. 43, 1956, pp. 1 – 14.
9. Sambursky, S. On the possible and probable in ancient Greece. *Osiris*, vol. 12, 1956, pp. 35 – 48.

10. David, F.N. *Games, Gods and Gambling*. London, 1962.
11. Byrne, E.F. *Probability and Opinion*. The Hague, 1968.
12. Rabinovitch, N.L. *Probability and Statistical Inference in Ancient and Medieval Jewish Literature*. Toronto, 1973.
13. Gnedenko, B.V. *Из истории науки о случайном* (From the History of the Science of the Random). M., 1981.
14. Sheynin, O.B. Newton and the classical theory of probability. AHES, vol. 7, 1971, pp. 217 – 243.
15. Sheynin, O.B. Lambert's work in probability. Ibidem, pp. 244 – 256.
16. Cournot, A.A. *Essai sur les fondements de nos connaissances*, t. 1. Paris, 1851.
17. Chuprov, A.A. *Очерки по теории статистики* (Essays on the Theory of Statistics) (1909). M., 1959.
18. Poisson, S.-D. *Recherches sur la probabilité des jugements*. Paris, 1837.
19. Sheynin, O.B. *Chuprov : Life, Work, Correspondence* (1990, in Russian). Göttingen, 1996.
- 20a. Mises, R. von, *Grundlagen der Wahrscheinlichkeitsrechnung* (1919). *Sel. Papers*, vol. 2. Providence, RI, 1964, pp. 57 – 105.
- 20b. Mises, R. von, *Wahrscheinlichkeit, Statistik und Wahrheit* (1928, 1936, 1951). English transl.: 1939, 1964.
21. Church, A. On the concept of a random sequence. *Bull. Amer. Math. Soc.*, vol. 46, 1940, pp. 130 – 135.
22. Uspensky, V.A., Semenov, A.L. *Теория алгоритмов ...* (Theory of Algorithms: Main Discoveries and Applications). M., 1987.
23. Kolmogorov, A.N., Uspensky, V.A. Algorithms and randomness. *Teoria Veroiatnostei i Ee Prime-neniya*, vol. 32, 1987, pp. 425 – 455. (R)
24. Bentley, R. Sermons (1693). In Newton, I. *Papers and Letters on Natural Philosophy*. Cambridge, 1958.
25. Lamarck, J.B. Aperçu analytique des connaissances humaines (MS, 1810 – 1814). In Vachon, M. et al, *Inédits de Lamarck*. Paris, 1972, pp. 69 – 141.
26. Lamarck, J.B. *Histoire naturelle des animaux sans vertèbres*, t. 1. Paris, 1815.
27. Kepler, J. De stella nova (1606). French transl. of a passage: Servien, P. *Science et hasard*. Paris, 1952, p. 132.
28. Laplace, P.S. Recherches sur l'intégration des équations différentielles (1776). *Oeuvr. Compl.*, t. 8. Paris, 1891, pp. 69 – 197.
29. Darwin, C. *Origin of Species* (1859). London – New York, 1958.
30. Newton, I. *Optics* (1704). *Opera*, vol. 4. London, 1782, pp. 1 – 264.
31. Newton, I. *Theological Manuscripts*. Liverpool, 1950.
32. Newton, I. *Mathematical Principles of Natural Philosophy*. Cambridge, 1934. Revised reissue of the edition of 1729.
33. Manuel, F.E. *A Portrait of Newton*. Cambridge (Mass.), 1968.
34. Lamarck, J.B. Espèce. *Nouv. Dict. Hist. Natur.*, t. 10, 1817, pp. 441 – 451.
35. Sheynin, O.B. On the history of the statistical method in physics. AHES, vol. 33, 1985, pp. 351 – 382.
36. Aristotle, *Metaphysica*. *Works*, vol. 8. Oxford.
37. Aquinas, Thomas, *Summa theologiae*. English transl.: *Great Books of the Western World*, vols 19 – 20. Chicago, 1952.
38. Mahalanobis, P.C. The foundations of statistics. *Sankhya, Ind. J. Stat.*, vol. 18, 1957, pp. 183 – 194.
39. Darwin, C. *The Variation of Animals and Plants under Domestication*, vols 1 – 2. London, 1885.
40. Herschel, W. Astronomical observations and experiments etc. (1817). *Scient. Papers*, vol. 2. London, 1912, pp. 575 – 591.
41. Struve, F.G.W. *Etudes d'astronomie stellaire*. Psb, 1847.
42. Maxwell, J.C. Discourse on molecules (MS, 1873). In Campbell, L., Garnett, W. *Life of Maxwell* (1882). London, 1884, pp. 272 – 274.
43. Hegel, G.W.F. *Wissenschaft der Logik*, Tl. 1 (1812). Hamburg, 1978.
44. Aristotle, *Physica*. *Works*, vol. 2. Oxford.
45. Aristotle, *De generatione animalium*. *Works*, vol. 5. Oxford.
46. Kepler, J. *Neue Astronomie* (1609, in Latin). München – Berlin, 1929.
47. Kepler, J. *Epitome of Copernican Astronomy*, book 4 (1620, in Latin). In *Great Books of the Western World*, vol. 16. Chicago, 1952, pp. 845 – 1004 (books 4 and 5).
48. Kepler, J. *Welt-Harmonik* (1619, in Latin). München – Berlin, 1939.
49. Kant, I. *Allgemeine Naturgeschichte und Theorie des Himmels* (1755). *Werke*, Bd. 1. Berlin, 1910, pp. 215 – 368.
50. Adanson, M. Examen de la question si les espèces changent parmi les plantes. *Hist. Acad. Roy. Paris avec Mém. math. et phys.*, 1769 (1772), pp. 31 – 48 of the *Mémoires*.

51. Lamarck, J.B. *Annuaire météorologique* [t. 1]. Paris, pour l'an 8 (1800).
52. Sheynin, O.B. On the history of the statistical method in meteorology. *AHES*, vol. 31, 1984, pp. 53 – 95.
53. Humboldt, A. Des lignes isothermes. *Mém. Phys. Chim. Soc. d'Arcueil*, t. 3, 1817, pp. 462 – 602.
54. Humboldt, A. *Fragmens de géologie et de climatologie asiatiques*, t. 2. Paris, 1831.
55. Koerber, H.G. Über Humboldts Arbeiten zur Meteorologie und Klimatologie. In Humboldt, *Gedenkschrift*. Berlin, 1959, pp. 289 – 335.
56. De Moivre, A. A method of approximating the sum of the terms of the binomial etc. (1733, in Latin). [57, 1756, pp. 243 – 254].
57. De Moivre, A. *Doctrine of Chances* (1718, 1738, 1756). New York, 1967 (reprint of the edition of 1756).
58. Belvalkar, S.K., Ranade, R.D. *History of Indian Philosophy*, vol. 2. Poona, 1927.
59. Cournot, A.A. *Exposition de la théorie des chances et des probabilités* (1843). Paris, 1984. Editor, B. Bru.
60. Cioffari, V. *Fortune and Fate from Democritus to St. Thomas Aquinas*. New York, 1935.
61. Hobbes, T. Of liberty and necessity (1646). *Engl. Works*, vol. 4. London, 1840, pp. 229 – 278.
62. Darwin, C. *More Letters*, vol. 1. London, 1903.
63. Dalember, Le Rond J. Doutes et questions sur le calcul des probabilités. In author's *Mélanges de littérature, d'hist. et de philos.*, t. 5. Amsterdam, 1768, pp. 239 – 264.
64. Laplace, P.S. *Essai philosophique sur les probabilités* (1814). *Oeuvr. Compl.*, t. 7, No. 2. Paris, 1886. Separate paging.
65. Aristotle, De caelo. *Works*, vol. 2. Oxford.
66. Kepler, J. A thorough discussion of an extraordinary new star (1604, in German). *Vistas in Astronomy*, vol. 20, 1977, pp. 333 – 339.
67. La Placette, J. *Traité des jeux de hasard*. La Haye, 1714.
68. Arbuthnot, J. An argument for divine Providence etc. (1712). In *Studies in the History of Statistics and Probability*, vol. 2. Editors, Sir Maurice Kendall, R.L. Plackett. London, 1977, pp. 30 – 34.
69. Boyle, R. Some considerations touching the usefulness of experimental natural philosophy (1663 – 1671). *Works*, vol. 2. London, 1772, pp. 36 – 49.
70. Arnauld, A., Nicole, P. *Logique de Port-Royal* (1662). Paris, 1877.
71. Voltaire, Homélies. Première homélie (1767). *Oeuvr. Compl.*, t. 26. Paris, 1879, pp. 315 – 354.
72. Bernoulli, D. Recherches physiques et astronomiques etc.
73. Laplace, P.S. Sur l'inclinaison moyenne des comètes (1776). *Oeuvr. Compl.*, t. 8. Paris, 1891, pp. 279 – 321.
74. Maupertuis, P.L.M. Venus physique (1745). *Oeuvr.*, t. 2. Lyon, 1756, pp. 1 – 133.
75. Maupertuis, P.L.M. Système de la nature (1751). *Ibidem*, pp. 135 – 184.
76. Herschel, J.F. Sun (lecture, 1861). In author's *Familiar Lectures on Scient. Subjects*. London – New York, 1866, pp. 47 – 90.
77. Baer, K. *Zum Streit über den Darwinismus*. Dorpat (Tartu), 1873.
78. Boole, G. On the theory of probabilities (1851). In author's *Studies in Logic and Probability*. London, 1952, pp. 247 – 259.
79. Newcomb, S. *On the Position of the Galactic*. Carnegie Instn of Washington, No. 10, 1904.
80. Forbes, J.D. On the alleged evidence for a physical connection between stars. *London, Edinb. and Dublin Phil. Mag.*, vol. 35, 1849, pp. 132 – 133; vol. 37, 1850, pp. 401 – 427.
81. Kapteyn, J.C. Statistical methods in stellar astronomy. [Repts] *Intern. Congr. Arts and Sci. St. Louis – Boston 1904*. N.p., vol. 4, 1906, pp. 396 – 425.
82. Struve, F.G.W. Review of O. Struve, *Bestimmung der Constante der Präcession*. Psb, 1843. *Bull. Scient. Acad. Imp. Sci. Psb.*, t. 10, No. 9 (225), 1842, pp. 129 – 139.
83. Boltzmann, L. Über das Wärmegleichgewicht (1871). *Wiss. Abh.*, Bd. 1. Leipzig, 1909, pp. 237 – 258.
84. Boltzmann, L. Weitere Studien über das Wärmegleichgewicht (1872). *Ibidem*, pp. 316 – 402.
85. Hume, D. *Treatise on Human Nature*, vols 1 – 2 (1739). London, 1874.
86. D'Holbach, P.H.T. *Système de la nature*, pt. 2. Paris, 1781.
87. Laplace, P.S. *Théorie analytique des probabilités* (1812). *Oeuvr. Compl.*, t. 7. Paris, 1886.
88. Todhunter, I. *History of the Mathematical Theory of Probability* (1865). New York, 1965.
89. Darwin, C. *Formation of Vegetable Mould* (1881). London, 1945.
90. Bertrand, J. *Calcul des probabilités* (1888). New York, 1972.
91. Galen, C. *De sanitata tuenda*. English transl. : *Hygiene*. Springfield, Ill., 1951.

92. Maxwell, J.C. Does the progress of physical science tend to give any advantage to the opinion of necessity etc. (Read 1873). In [42, pp. 357 – 366].
93. Maxwell, J.C. On the stability of the motion of Saturn's rings (1859). *Scient. Papers*, vol. 1 (1890). Paris, 1927, pp. 288 – 376.
94. von Kries, J. *Die Principien der Wahrscheinlichkeitsrechnung* (1886). Tübingen, 1927.
95. Pirogov, N.N. Über das Gesetz Boltzmanns. *Repertorium Phys.*, Bd. 27, 1891, pp. 515 – 546.
96. Leibniz, G.W. Allgemeine Untersuchungen über die Analyse der Begriffe und wahren Sätze (MS 1686). In author's *Fragmente zur Logik*. Berlin, 1960, pp. 241 – 303.
97. Maxwell, J.C. Illustrations of the dynamical theory of gases (1860). *Scient. Papers*, vol. 1 (1890). Paris, 1927, pp. 377 – 410.
98. Maxwell, J.C. On the dynamical evidence of the molecular constitution of bodies (1875). *Scient. Papers*, vol. 2 (1890). Paris, 1927, pp. 418 – 438.
99. Laplace, P.S. *Exposition du système du monde. Oeuvr. Compl.*, t. 6. Paris, 1884. Reprint of edition of 1835.
100. Kant, I. *Kritik der reinen Vernunft* (1781). *Werke*, Bd. 3. Berlin, 1911.
101. Engels, F. *Dialektik der Natur* (written 1873 – 1882, publ. 1925). Berlin, 1971.
102. Poincaré, H. *Calcul des probabilités* (1896). Paris, 1912, reprinted 1923. The Introduction to this edition is a reprint of an article of 1907.
103. Pearson, K. A. De Moivre. *Nature*, vol. 117, 1926, pp. 551 – 552.
104. Pearson, K. *History of Statistics in the 17th and 18th centuries etc.* Lectures 1921 – 1933. Editor E.S. Pearson. London, 1978.

17. Correspondence between P.A. Nekrasov and A.I. Chuprov

IMI, vol. 1 (36), No. 1, 1995, pp. 157 – 167

1. Introduction

I have written about Pavel Alekseevich Nekrasov (1853 – 1924) [1, §1.5; 2; 3, §1]. Now, I mention other sources [4; 5; 6] throwing light on his biography. The author of [6] explains Nekrasov's Weltanschauung and his later work by his aspiration for permeating social life by arithmology (in its wider sense). Be that as it may, I keep to my previous opinion [3, §1.3], and, in particular, I am still believing that, from about 1900, Nekrasov's mathematical writings became unimaginably verbose, intrinsically connected with ethics, religion and politics, and therefore obscure. In addition, the term *arithmology*, even in its narrow sense, is no longer in use, and Polovinkin [6] should have explained his reasoning as well as the title of his article. See Note 6 to Correspondence of Nekrasov and Andreev (translated in this collection).

Here, I only repeat that in 1893 Nekrasov became Rector of Moscow University; in 1898, warden of the Moscow educational region; and, in 1905, a prominent official at the Ministry of Public Education. Aleksandr Ivanovich Chuprov (1842 – 1908), Corresponding Member of the Imperial (Petersburg) Academy of Sciences, was a statistician, the father of zemstvo statistics, an economist and writer on current topics [7; 8]. For a long time, until the autumn of 1899, he taught at the Law faculty in Moscow. More widely known is his son Aleksandr.

Two letters from Nekrasov to Chuprov (1898 and 1899) are kept at the Central State Historical Archive (Fond 2244, inventory 1, No. 2124). The first is devoted to the teaching of the theory of probability at the Law faculty of Moscow University (see §2) whereas the second one characterizes the general situation at the University and I think that it should be also adduced. Here it is.

Nekrasov – Chuprov, 17 Febr. 1899

Dear Sir, Aleksandr Ivanovich, – Today, your lecture, as I heard, had not taken place because of the pressure of a group of students who want to impede the course of studies. Since there exists another group of students seeking after the contrary, I am most zealously [!] asking you not to give in during your forthcoming lecture tomorrow, and, if possible, to carry it out. In this way you will undoubtedly contribute to putting an end to the students' unrest. [...]¹

2. The Letter of 1898

Nekrasov is known to have been advocating the inclusion of the theory of probability into the school curriculum. It occurred that he also thought of teaching this discipline to student-lawyers. His appeal to Chuprov, who was extremely influential in his field, was hardly official: the latter was not the Dean of the Law faculty.

At the turn of the 19th century, the possibility of using probability in statistics was already proved, – in England, for biological research, and, on the Continent, for the theory of stability of statistical series (Lexis, Bortkiewicz). Furthermore, already Quetelet applied elements of probability for studying moral statistics (the statistics of marriages, suicides and crime), which could have undoubtedly been useful for lawyers. Nevertheless, Nekrasov's program (below) hardly mentioned that branch of statistics.

In 1896 Nekrasov accepted the "candidate composition" [9] written by Chuprov's son, then graduating from the Physical and Mathematical faculty of the University. There, the future scientist attempted, in particular, to study the interrelations of the statistical method with philosophy and logic, and Nekrasov could have well included the last-mentioned item in §7 of his program, "The statistical method as one of the methods of cognition". Finally, §8 of Nekrasov's program testifies to his interest in the application of the theory of probability to economics. Later he paid much attention to that issue [10; 11, 1912, Chapt. 5 of pt. 2], and, during 1918 – 1919, he read a special course *On the branches of mathematics necessary for the economic sciences* [12, p. 423] at Moscow University (for a single listener, A.A. Koniushin).² Here, now, is Nekrasov's letter.

Nekrasov – Chuprov, 27 Jan. 1898

Highly respected Aleksandr Ivanovich, – I am sending you a copy of my memorandum about which I told you during our rendezvous and which I, as a person teaching the theory of probability, intend to submit to the Law faculty.³ Other mathematicians will also probably sign it. The extent of teaching is determined by the appended program; for the time being I am raising the issue only in its essence. I am convinced that a proper and skillful teaching of probability will heighten the lawyers' level of education and I hope that you will regard this matter with due sympathy and exert your influence at the Faculty in order to establish this teaching under the most favorable conditions that are especially necessary for an absolutely new undertaking. [...]

[Supplement 1.] A Rough Program for Teaching Probability Theory with Applications to Phenomena of Public Life to Students of the Law Faculty⁴

1. *Random phenomena and their probabilities.* Examples of direct calculations of probabilities. The case of an infinite number of chances. Moral certitude.
2. *The main theorems.* The addition theorem. Contrary events. A group of all possible incompatible events.⁵ A compound event and the notion of conditional probability. The multiplication theorem. Independent events and the multiplication theorem for them. Hypotheses. The theorem on total probabilities.
3. *Probabilities of compound events in numerous trials.* The case of constant probability in all trials. The case in which the probabilities of the events change from one trial to another.
4. *The law of large numbers.* An elementary derivation of the theory of Jakob Bernoulli and Poisson. Definition of the expectation of a random variable.⁶ The Chebyshev form of the law of large numbers. The Poisson and the Bernoulli theorems as particular cases of the Chebyshev proposition. The boundaries of the action of the law of large numbers. Examples provided by Ettingen and Bertrand.⁷
5. *On probabilities a posteriori.* The Baye [!] theorem and its corollaries. Examples. The subjective aspect of the notion of probability. The change of the posterior probability depending on the accumulation of data. The application of the Baye theorem to the derivation of the main theorem on testimonies.⁸
6. *An elementary theory of the method of least squares.* The principle of the arithmetic mean.⁹ The measure of precision. Combination of observations having different measures of precision. The weights of the results. The method of least squares in cases of one and many unknowns.
7. *Application of probability to statistics.* Statistical data and the statistical method. Its field. The need for a special critical appraisal of statistical data. Phenomena in public life and the will as one of its causes.¹⁰ Moral statistics. A classification of mass observations and phenomena. Regularities in phenomena of public life. An empirical determination of probabilities as one of the problems of statistics. Application to determining the probabilities of duration of life. The statistical method as one of the methods of cognition.
8. *The influence of chances on estimating monetary undertakings.* The importance of the law of large numbers in the Bernoulli and Chebyshev forms for determining the value of sums and undertakings exposed to randomness. On fair money games. On insurance of property and life. On buying annuities.
9. *Application of the theory of probability to legal proceedings.* A caution regarding the conditions for the application of probability theory to verdicts and testimonies.¹¹ The difficulties in accomplishing these conditions in full. The change of the probabilities of phenomena after new testimonies and verdicts become known. Criminal statistics.¹²

[Supplement 2.] To the Law Faculty

During the last half-century, the theory of probability together with its applications made more than a small progress for which it is considerably indebted to Russian scientists, suffice it to mention Bunyakovsky, Davidov and especially Chebyshev.¹³ The advances in probability were not however reflected upon the level of educating the students of the Law faculty since the teaching of that discipline is not assigned a proper place.¹⁴

It could hardly be doubted that the subject-matter of a science cannot be isolated from its main methods without causing damage to the teaching. Such an abnormal dissociation always led to stagnation hindering the correct interpretation of the appropriate phenomena, and moreover, precluding the expedient use of methodology. Regrettably, such a dissociation, harmful for the success of education, exists between the sciences of jurisprudence, which are in charge of the phenomena in social and political life, and the theory of probability, which provides mathematical methods for their systematic investigation.

Professor [Yu.E.] Yanson, in his *Теория статистики* (Theory of Statistics). Pskov, 1891, p. 490, characterized the abnormality of the situation in the following words:

*Regrettably, statisticians are not sufficiently acquainted with the theory of probability whereas the mathematicians, who applied mathematical calculations to analyzing numerical data on social phenomena, considered them as abstract magnitudes, did not take into account the special properties of these phenomena and arrived therefore at conclusions bordering on nonsense.*¹⁵

The need to get rid of this dissociation by properly teaching the theory of probability and its applications to phenomena in social life can be justified by many considerations. Thus, the doctrine of probabilities provides a precise formulation and a complete interpretation of the so-called law of large numbers, i.e., of mass phenomena to which social and state phenomena also belong. At the same time, this science offers methods for discovering the most cautious assumptions about future random phenomena, for example

those concerning economic and financial life. Lastly, the application of probability theory to testimonies and verdicts cannot remain uninteresting for an educated lawyer.

Since social phenomena cannot at present be investigated without any knowledge of the theory, its conclusions are even now being partly reported at the Law faculty. Regrettably, the information provided is scanty, extremely fragmentary and not always precise. Furthermore, it is offered on trust, without any substantiation which is necessary not only for cogency, but also for ensuring a distinct understanding of the boundaries for applying the reported methods. So as to justify such an abnormal situation, it was reasoned that the mathematical analysis of probabilities demanded the use of higher mathematics, the acquaintance with which could not have been expected from student-lawyers. At present, however, this consideration had lost its meaning owing to the works of Chebyshev and the attempts of later Russian mathematicians.¹⁶ Chebyshev's outstanding merit consists not only in that he provided a more general expression of the law of large numbers, but also in that he extraordinarily simplified the proof of this most important proposition of the theory.¹⁷ Nowadays, all the essential parts of this doctrine and its applications can be taught in an elementary way when issuing from the mathematical knowledge determined by the gymnasium program. This discipline can thus be made intelligible to student-lawyers. It is self-evident that the lectures in probability adapted for them must differ from those read to future mathematicians, but they can retain precise scientific character and be rich in their content.

The reader in probability must naturally take care that its teaching be of an adequate scientific level and sufficiently disseminated in the University. This consideration prompts me to raise before the Faculty the issue in principle about the teaching of the theory of probability with its applications to student-lawyers. This problem interested me for a long time; at present, I have compiled quite a definite plan for its solution and presented it in the subjoined rough program that can be made use of for teaching the theory of probability with its applications to student-lawyers. Two hours a week lasting for an academic year would be quite sufficient for a conscientious mastering of this course. When studying the theory according to this program, student-lawyers will encounter difficulties; these, however, will be caused not by the complexity of mathematical analysis, that will not go beyond the gymnasium curriculum, but by the intricacy of the ideas and notions that form the subject-matter of the science of random phenomena.¹⁸ He who successfully overcomes these difficulties will more distinctly understand the laws of social phenomena. In raising the issue of teaching the theory of probability at the Law faculty, I consider it necessary to state that favorable conditions ensuring adequate success be provided for this. If this issue will be satisfactorily solved, the teaching will not present any difficulties for the personnel at the disposal of the University.¹⁹

Notes

1. Next year, 9 February 1900, Nekrasov will write to F.E. Kosh (1843 – 1915), a philologist and orientalist: "Up to now, there is absolute order at Moscow University, but the nearest future is full of uncertainty" (Archive, Russian Acad. Sci., Fond 558, inventory 4, No. 235). Also see the appropriate passage in [2, §1].

2. In November 1989 Konius told me that Nekrasov's lectures had included an examination of the work of Walras, the founder of the mathematical school in economics. Judging by its title, Nekrasov's course had much in common with his report [13].

3. I emphasize that Nekrasov, still the Rector of the University [14], did not pull rank.

4. The titles of several sections of this program coincided with those of the appropriate chapters of Nekrasov's treatise [11, 1896 and/or 1912].

5. The term "a group of ... events/phenomena" occurs also in Nekrasov's treatise [11, 1896, p. 13; 1912, p. 221] without any special emphasis on *group*.

6. Nekrasov was one of the first to apply this term (in Russian, *random magnitude* [1, p. 350, Note 17]). However, the absence of the notion of density implies that he restricted his attention here to discrete variables. Incidentally, he was thus unable to mention the normal distribution which considerably worsened his program.

7. I am unaware of the former and I doubt that either was essential.

8. This term is not in common use. Later Nekrasov [15, p. 14] called the formula of the type

$$P = pA / (pA + qB)$$

"the main equation of the probabilities of testimonies"; p and A were the probabilities of the truthfulness of the witness of the event in question and of the subsequent narrator, $q = 1 - p$ and $B = 1 - A$. Already Condorcet [16, p. 400] introduced this formula.

9. Otherwise: the Gauss postulate (1809) according to which the arithmetic mean of observations coincided with the mode of the unimodal curve of distribution of their errors.

10. It seems that Nekrasov had grossly exaggerated the importance of (free) will. The regularities in public life, which he obviously had in mind, were caused not by the action of will, but by the specific nature of mass phenomena. My statement excludes public outbursts, or social will. Nekrasov largely devoted his writing [17] to free will.

11. Nekrasov apparently thought about the (non-existent) independence of the judgements passed by the jurors.

12. Vlasov [18], see Note 15 below, had not discussed the subject-matter of Nekrasov's §§6, 8, or 9, but he included elements of the theory of stability of statistical series lacking in §7.

13. Nekrasov obviously overestimated the influence of the two first named mathematicians.

14. Here and below Nekrasov directly or implicitly stated that some elements of probability were nevertheless reported at the faculty. According to official documents, the theory of probability was not taught there in 1902 – 1903 or 1912 – 1916 [19], and only in 1907 – 1908 A.K. Vlasov delivered a course of lectures in that theory [18]. At that time, and also in 1912 – 1916, statistics was also taught at the Law faculty. I have no information relating to 1899 – 1907 or 1908 – 1912. Note that in 1908 Vlasov edited a Russian translation of Laplace's *Essai philosophique* and that in 1911, after about twenty years there, he had to leave the University [20] because of the worsening of its social and political atmosphere [21, pp. 375 – 377].

15. Quetelet [22, p. 633] pronounced a similar statement, but he only mentioned “des prétendus savants”. Is it true, however, that mathematicians rather than the statisticians themselves arrived at non-sensical results? Buniakovsky [23, p. 154], who was both a mathematician and a statistician, remarked, although not in a statistical context, that

Anyone who does not examine the meaning of the numbers with which he performs particular calculations, is not a mathematician.

16. Along with Nekrasov's example below, it can be indicated that Chebyshev, in his Master dissertation [24], had indeed explicated the theory of probability by elementary means but his description was ponderous. Vlasov (Note 15) also managed without higher mathematics in his textbook.

17. In describing Chebyshev's merits Nekrasov possibly did not want to go beyond the boundaries of his program. Nevertheless, Nekrasov denied the importance of the Chebyshev's proof of the CLT, see Note 26 to Correspondence of Nekrasov and Andreev (translated in this collection).

18. At the time, this definition of probability theory, although formulated indirectly, was indeed fortunate.

19. I do not know whom Nekrasov had borne in mind. He himself left the University two months afterwards [14]. Moreover, during 1902 – 1904, 1912 – 1913, 1914 – 1915 and 1916 – 1917 no-one had been teaching the theory of probability even at the Physical and Mathematical faculty [25]! I have no information about 1904 – 1912, but during 1913 – 1914 and 1915 – 1916 the theory was indeed taught there by L.K. Lakhtin (Ibidem). Incidentally, all this testifies against the Nekrasov – Florov proposal that probability theory be introduced into the school curriculum [26]. Some participants of the then ensuing debate (Ibidem, No. 3) had indeed expressed doubts about the availability of qualified school teachers.

References

1. Sheynin, O.B. Markov's work on the theory of probability. AHES, vol. 39, 1989, pp. 337 – 377.
2. Sheynin, O.B. Markov's publications in the newspaper *Den*, 1914 – 1915 (1993). Translated in this collection.
3. Chirikov, M.V., Sheynin, O.B. Correspondence between Nekrasov and Andreev (1994). Translated in this collection.
4. Anonymous, Nekrasov. *Новый энц. словарь Брокгауза и Ефрона* (New Brockhaus & Efron Enc. Dict.), vol. 28, 1916, p. 272.
5. Sluginov, S.P. Nekrasov. *Trudy Matematich. Seminaria Permskogo Gos. Univ.*, No. 1, 1927, pp. 37 – 38 (R)
6. Polovinkin, S.M. The psycho-arithmo-mechanician. Philosophical features of Nekrasov's portrait. *Voprosy Istorii Estestvoznania i Tekhniki* No. 2, 1994, pp. 109 – 113. (R)
7. Kablukov, N. [A.] A.I. Chuprov, a biographical essay. In Chuprov, A.I. *Речи и статьи* (Speeches and Papers), vol. 1. M., 1909, xiii – xlviii.
8. Koni, A.F. From [my] recollections of Chuprov. Ibidem, vol. 3, ix – xli.
9. Chuprov, A.A. *Математические основания теории статистики* (Math. Principles of Theory of Statistics). Thesis. M., 1896. Unpublished. Gorky Library at Moscow Univ., Section rare books & MSS. The Fond of A.I. & A.A. Chuprov, karton 9, item 1. (R)
10. Nekrasov, P.A. Mathematical statistics, commercial law and financial turnover. *Izvestia Russk. Geografich. Obshchestvo*, vol. 45, 1909, pp. 333 – 398, 565 – 612 and 811 – 896. (R)

11. Nekrasov, P.A. *Теория вероятностей* (Theory of Probability). М., 1896; Psb, 1912.
12. Komlev, S.L. On the conjuncture statistics of the 1920s. Conversation with A.A. Konius. *Ekonomika i Matematich. Metody*, vol. 25, 1989, pp. 423 – 434. (R)
13. Nekrasov, P.A. On the sections of mathematics necessary for the economic sciences. *Matematich. Obrasovanie*, No. 2, 1912, pp. 79 – 81. (R)
14. Anonymous, The new warden of the Moscow educational region. Newspaper *Moskovskie Vedomosti*, 1898, March 13 (25), pp. 2 – 3 and 15 (27), p. 2. (R)
15. Nekrasov, P.A. *Средняя школа и т.д.* (High School, Mathematics and Scient. Training of Teachers). Psb, 1916.
16. Todhunter, I. *History of the Mathematical Theory of Probability* (1865). New York, 1965.
17. Nekrasov, P.A. *Философия и логика и т.д.* (Philosophy and Logic of Science of Mass Manifestation of Human Activities). М., 1902.
18. Vlasov, A.K. *Теория вероятностей* (Theory of Probability). М., 1909 and 1916.
19. *Обозрение преподавания на юридическом факультете Имп. Моск. унив. на ...* (Review of Teaching at Law Faculty, Imp. Mosc. Univ. for ... Academic Year). Appeared yearly. N.p., no dates, no title-pages.
20. Glagolev, N.A. A.K. Vlasov (1868 – 1922). MSb, vol. 32, 1925, pp. 273 – 275. (R)
21. *История Московского университета* (History of Moscow Univ.), vol. 1. М., 1955.
22. Quetelet, A. Unité de l'espèce humaine. *Bull. Acad. Roy. Sci., Lettres et Beaux-Arts Belg.*, sér. 2, t. 34, 1872, pp. 623 – 635.
23. Buniakovsky, V.Ya. Essay on the laws of mortality in Russia and on the distribution of the Orthodox believers by ages. *Zapiski Imp. Akad. Nauk Psb*, vol. 8, 1866, Suppl. 6. Separate paging. (R)
24. Chebyshev, P.L. *Опыт элементарного анализа теории вероятностей* (Essay on Elementary Analysis of Theory of Probability). (1845). *Полн. Собр. Соч.* (Complete Works), vol. 5. М. – L., 1961, pp. 26 – 87.
25. *Обозрение преподавания на физико-математическом факультете [Имп. Моск. Унив.] на ...* (Review of teaching at the Phys. and Math. Faculty [of Imp. Moscow Univ.] for ... Academic Year). Appeared yearly. N.p., no dates, no title-pages.
26. Nekrasov, P.A. The theory of probability and mathematics in the high school. ZhMNP, 1915, 4th paging, No. 2, pp. 65 – 127, No. 3, pp. 1 – 43, No. 4, pp. 94 – 125. (R)

1. Introduction

In 1906 Markov published two polemic articles (letters) on insurance of children in the newspaper *Наша жизнь* (Nasha Zhisn, Our Life). Later he (§4) referred to them without mentioning his authorship; no wonder that they were not included in the Bibliography of his works [1]. Neither did Grodzensky [2] cite them although it was he from whom I first came to know about their existence. I comment on these articles in §4 and reprint them in §5, and in §3 I describe Markov's activities in insurance. My §2 discusses the history of insurance and its material is largely known. There, I draw, among other sources, on my previous paper [3]; Kohli & van der Waerden [4] and Hald [5] devoted much attention to this subject.

I define life insurance as any agreement ensuring payments of definite sums either to the heir(s) of the insured should he/she die within a stipulated period of time (a lump sum), or to the insured himself (regular sums, and, especially, an annuity). According to modern ideas, but not in line with the practice of insurance during the 17th and 18th centuries, the price of such agreements must be determined by means of mortality tables depending on the age and the sex of the insured. My definition does not cover all the existing forms of life insurance (§4), but it is sufficient for a general understanding of the matter. I also note that various kinds of mutual insurance of several persons have also been widely used. Thus, upon paying a necessary sum, a married couple could have enjoyed a fixed annuity until one of them dies, with the surviving spouse continuing to draw it to the end of his/her life.

In England, societies offering mutual insurance had already been in existence in the 17th century. At the turn of the next century that country had several thousand of them. Their members drew insurance in cases of illnesses or death of their wives, and wives received it upon the death of their husbands. It seems [6] that most such societies existed on voluntary dues, but that in any case there had been no connection between the premiums and the ages of their members.

An operation connected with risk is called fair if the expected winning (ξ) is zero ($E\xi = 0$). For an insured, insurance is never fair: since insurance societies cannot exist without profit, his/her expectation is always negative. Nevertheless, insurance might be advantageous for the insured, if, for example, his family will get a lot of money should he die prematurely. And, indeed, such scientists as Laplace [7, p. 454] ardently approved of the institution of life insurance. In 1898, more than 7mln people were insured the world over, about 0.1mln of them in Russia [8, p. 747] which goes to show the scale of the activities of the main insurance enterprises roughly at the time that directly concerns us.

2. From the History of Life Insurance

Population statistics had been the most important branch of political arithmetic that emerged in the mid-17th century and at least until the beginning of the 19th century the former remained significant mainly because of the developing insurance business demanding reliable data on mortality and studies of its laws. These statistical data, insofar as they were being collected by insurance societies, had been kept secret, but the theoretical principles were not concealed. Their development both directly and implicitly heightened the interest in probability and to some extent fostered its advancement.

In 1669, in a letter to his brother Lodewijk devoted to various problems in mortality and published in 1895 [9], Christiaan Huygens calculated the expectations of the order statistics for an empirical distribution, introduced the concepts of mean and probable durations of life and constructed and made methodological use of a graph of a continuous function $y = 1 - F(x)$ where $F(x)$ in my notation was an integral distribution function of mortality. It was in this correspondence that the theory of probability went beyond the province of games of chance (as it also did in 1671 at the hands of De Witt).

In 1709 Niklaus Bernoulli [10] considered a number of problems connected with insurance. In one of these he (pp. 296 – 297, also see [11, pp. 195 – 196]) determined the expectation of the maximal element of a sample from a continuous uniform distribution. Issuing from statistical data published by Halley in 1694, De Moivre [12] proposed to describe mortality, beginning with age 12, by a uniform distribution. There also he introduced the expectation of a random variable thus distributed (Problem 20 from pt. 1) and calculated probabilities of the type $P(\xi \geq x) = 1 - F(x)$ for the same distribution (Chapt. 8 of pt. 2, my own notation).

Laplace [7, Chapt. 9] solved several problems on life insurance in the same way as those pertaining to the treatment of observations, but this time he also discussed the so-called Poisson generalization of the Bernoulli trials. Gauss did not shun life insurance either; he had to solve practical problems while managing the pension fund at Göttingen University [13, pp. 61 – 64]. In Russia, Zernov [14] published a treatise in which he paid special attention to life insurance and Buniakovsky, in 1846, devoted a chapter of his celebrated work to the same subject.

3. Markov's Work in Retirement Funds

Retirement funds began appearing in Russia in the second half of the 19th century. After retirement, their members had been drawing lifelong pensions depending on the duration of their work and their final or mean salary. It was indeed possible to estimate the duration of life of the pensioners by applying mortality tables (although, strictly speaking, statistical inferences suitable for the general population will not do for its special groups), but it was extremely difficult to predict the yearly number of the retiring or their salaries whereas the evaluation of the number of additional members of a given fund admitted for state reasons (see below) was absolutely impossible. Consider also that the widows and children of dying members were also provided with life annuities or longterm pensions, and it becomes evident that any retirement fund could have went broke, and especially so during its first years of existence when experience was still lacking and unreliable guesswork was necessary.

Ostrogradsky [15] participated in the work of the first Russian retirement fund. A few decades later Markov began to busy himself with similar activities; already in 1884 he [16] published detailed calculations for the retirement fund at the Ministry of Justice. In 1890 he became member of its governing board [17, vol. 2, p. 36]. He actively participated in its sittings, offered his advice about concrete issues and checked book-keeping accounts. Thus, he compiled a note [18] (not mentioned in his Bibliography [1]) on the financial conditions necessary for ensuring the payment of pensions. Vol. 1 of the same source [17] contains many references to Markov, and on pp. 90 and 100 it cites pp. 10 and 6 respectively of a certain note, possibly [16], since [18] is only two pages long.

In 1893, 1894 and 1902 Markov received letters of thanks from the Ministry of Justice [2, p. 59]. Incidentally, its retirement fund was considered the "best established" from among the six funds of the "civil departments", and this fact was naively attributed to Markov's "precise mathematical calculations" [19]. It would have been more correct to mention his prudence and foresight, perhaps his intuition and ability to detect the slightest circumstances.

Markov also occupied himself with similar work at the War Ministry [20]. In 1900 Academician Sonin [21], on behalf of the Physical and Mathematical Department of the Academy of Sciences, acquainted himself with the work of the Ministry's retirement fund and expressed his opinion about it in the following way: "The sole reason for the crisis that it experiences now" was the unforeseeable increase in the number of its members occurring through instructions from above. He recommended to liquidate the fund and to transfer its liabilities to the state.

After hearing this out, Markov (Ibidem) declared that he did not agree with Sonin "on any point". The Department resolved that, since the problem posed by the War Ministry [before the Academy] was rather of a practical than purely scientific nature, it should only inform the Ministry that "the members of the Academy are always ready to render assistance" to it.

Also in 1900 the same Ministry established a "Special [standing] Mathematical Conference" for determining the financial state of its fund [22, p. 10]. Its members included academicians Markov, Sonin and I.I. Yanzhul, other eminent scientists (I.I. Pomerantsev, N.Ya. Tsinger) and actuaries (B.F. Maleshevsky). Regrettably, nothing is known either about the work of this Conference or of Markov's even more active participation in practical life insurance after his retirement in 1906 [23, p. 604].

Again in 1900, Markov devoted to life insurance a short chapter of his textbook [24]. There, not aiming at new results, he acquainted his readers with the main stochastic problems of the contemporaneous insurance business. I indicate, finally, that the Markov Fond (Fond 173, inventory 1) at the Archive of the Russian Academy of Sciences includes three letters directly pertaining to my subject.

1) An undated Markov's letter to Maleshevsky (Delo 60, No. 15). Markov disapprovingly mentioned the "just appeared" book of Savich [25] and noted that it "compelled me [him] to turn attention once more [!] to the theory of inability to work". He also discussed one of Maleshevsky's formulas and expressed his opinion about the mortality of the disabled.

2) D.A. Grave's letter to Markov of 21 April 1916 (Delo 5, No. 5). Grave indirectly agreed with Markov in that the granting of some kind of pensions was undesirable.

3) Another letter from Grave to Markov of 13 Dec. 1916 (Delo 5, No. 7). Grave mentioned a "surprising discordance" between the calculations made by Markov and Maleshevsky. These apparently concerned the work of the pension fund in the city of Chernigov.

4. Markov's Newspaper Publications

In §5 I reprint two polemic newspaper articles published by Markov and devoted to the insurance of children. It may be thought that this kind of insurance was more or less widely practised in Russia from at least the mid-19th century. In any case, Kraevich [26] included an appropriate example in the first three editions of his collection of mathematical exercises for school students. Here it is. Upon the birth of a boy, his father deposits 1,000 rubles with an insurance society. In exchange, the son draws x rubles after his 20th birthday,

but the money is lost if he dies before that date. Assuming that the insurance is fair, and that the interest rate is 5%, determine x by means of the appended mortality table (whose origin is not explained).

Both this simplest pattern, and the other one criticized by Markov (below), and, as it may be supposed, any other scheme for insuring children, suffers from one and the same essential defect: they necessarily remain unfavorable for the insured (see §1; fair insurance is only possible in textbooks), and they do not really insure him. Here is a relevant passage whose author mentions, among other types of insurance, the insurance of children against death [27, p. 243]:

Cases that, under the guise of insuring life, conceal deals in paying out some moneys upon the occurrence of a stipulated event not inflicting [pecuniary] loss on the insured, – deals which do not restore actual damage, – should not be attributed to insurance. They abuse the idea of insurance.

Markov's criticism was justified; regrettably, however, he did not take the occasion to explain to his readers that there exist other forms of life insurance (and of insuring property) advantageous for the insured.¹

5. Markov's Letters

Letter No. 1. Newspaper *Nasha Zhisn*, 7 April 1906, p. 1

The "Benefits" of Insurance through the Savings Offices

In order to ascertain the benefits of insuring profits and capitals through the state savings offices [...] it is necessary to consider the tariffs. Judging by the number of these (5 – 10), the insurance of juveniles plays a large part in the new direction of business of the savings offices. Who will benefit from this insurance, excepting "those engaged in this operation"?² To answer this question it is necessary to dwell on the tariffs of insurance from which we extract two lucid examples.

1) According to tariff 6, a downpayment of 1,200 rubles is necessary for a six-year-old child to draw 2,000r after reaching the age of twenty; and, should the child die prematurely, only $1,200 - 60 = 1,140r$ are returned back (5% is retained to cover the expenses). On the other hand, if the same sum, 1,200r, be kept at a bank with an interest rate of 4% (this is the rate underlying the tariffs) for each full hundred rubles,³ then, consecutively,

$$1,200 + 4\% = 1,248 \text{ at seven years; [...] } 1,964 + 76 = 2,040 \text{ at twenty years.}^4$$

My table shows that this insurance is in all cases disadvantageous for the family. If the child survives until age 20, the loss will be expressed by a small sum of 40r; otherwise, it can amount to several hundred rubles since the family loses the interest on the downpayment.

2) According to tariff 8, a yearly grant of 600r during five consecutive years will be paid out to a six-year-old child after his reaching age 18 for a downpayment of 1,789r; and, should the child die before that age, $1,789 - 89 = 1,700r$ are returned back.

$$1,789 + 68 = 1,857 \text{ (age, seven years); [...] } 2,729 + 108 = 2,837 \text{ (age, eighteen years).}$$

So, when paying out the 600r for five years, we obtain consecutively

$$2,837 - 600 = 2,237; [...] 637 - 600 + 24 = 61.$$

It is seen that this operation also inflicts a loss for the family. In the favorable case this loss is expressed by a small sum of 61r; otherwise, it can amount to a thousand rubles.

As indicated above, I have chosen lucid examples, but similar results are obtained in the other cases as well with the only difference being that, for the most favorable instances, the small loss can become a small profit. However, the possibility of large losses for the family because of a premature death of its child persists.

Letter No. 2. Newspaper *Nasha Zhisn*, 2 May 1906, p. 1

The "Benefits" of Insurance through the Savings Offices

The explanations provided by the Directorate of the savings offices [28] do not explain anything; on the contrary, they obscure the essence of the problem that I have raised.⁵ They are composed in such a way as though the whole matter consists in the high cost of insurance through savings offices, which, in my example, was expressed by a small sum of 40r out of 2,000. Dwelling only on this small loss, the Directorate maintains that it is of no consequence owing to the security of savings through the insurance and is compen-

sated by profit sharing. And, assuming an interest rate of 5% rather than 4%, the Directorate promises the insured a payout of 2,180 – 2,280 instead of the 2,000.

Thus, the Directorate completely overlooks those cases in which the insured child dies prematurely and the family's loss due to the insurance is expressed already by hundreds rather than tens of rubles. Only by forgetting these instances is it possible to bring oneself to state that the savings are here secured. Meanwhile, in my first Note I had paid attention to these cases; and, for determining the loss incurred by the insurance to the family, I had adduced, in addition to the figure 2,040 on which the Directorate rests its eyes, a number of other ones. The Directorate apparently chose only the most favorable case; and it vainly tries to prove that in this instance the family's loss can be replaced by some profit. Indeed, I had mentioned this possibility in my Note: suffice it to change the age of the insured and the duration of the contract in such a way that the probability of losing the stipulated insurance heightens.

As to the method by which the Directorate attempts to replace the loss by a profit, it cannot be called proper not only because, instead of providing a detailed calculation, it only indicates an indefinite magnitude between 2,180 and 2,280, but, mainly, since it admits that its estimation was based on changing the interest rate. The Directorate obviously forgot that after 14 years and assuming a 5% yearly interest rate, a capital of 1,200r, when saved without any insurance being involved, fetches not 2,040, but $1,200 \cdot 1,05^{14} = 2,374r$.⁶ Nevertheless, I am quite prepared to agree with the Directorate that my calculations were based on a too low rate of interest, witness for example the latest pleasing loan.⁷ But an increase in the rate increases the family's loss incurred by the insurance. The Directorate's statement about an exaggeration in my reckoning is therefore absolutely wrong. Thus, my indication that some insurance procedures offered by the savings offices always lead to losses remains unshaken. Neither can it be shaken until the mortality table taken as the basis for computing the tariffs of insurance remains unaltered and the expenses (5%) of carrying out the insurance are not lowered.

Indicating profit sharing, the Directorate says that five years after the insurance operations begins profit will be shared among the insured; but it forgets to mention that a considerable part (25%) of the profit will go to the employees of the savings offices.

Defending its future operations of insuring juveniles, the Directorate refers to [private] insurance societies where such operations are carried out according to higher tariffs, but, regrettably, it does not corroborate this statement by comparative excerpts. My remarks undoubtedly concern these societies as well, but it is also obvious that insurance societies aim at getting rich, and this distinct goal can serve as a warning to those insuring. On the other hand, the fact that some operations are being carried out, is no proof that they should indeed be done. For example, a lot of people gather to play the roulette in Monaco – so should not we therefore arrange that game, or something similar, at the savings offices? For anyone who read my first Note it should be clear that all the conclusions there contained only concern the insurance of juveniles. The tariffs of insurance as carried out by the savings offices cover, however, not one single form of insurance, as it could be understood from the words of the Directorate, but several forms, so that, according to the number of the tariffs involved, the insurance of juveniles occupies a rather considerable place among the new operations at the savings offices. And I have provided examples concerning two different tariffs.

I have not touched on other kinds of insurance so that the Directorate apparently vainly defends them; and the more so since its arguments reduce to a statement that for 70r it is possible, given some conditions, to draw 1,000r. Is the Directorate really so naïve as to attach serious meaning to this proposition? Having 70r and playing the roulette game it is possible to win even more than 1,000r.

Thus, I have spoken only about some forms of insurance whereas the Directorate itself raised the question about the high, or the low cost of all kinds of its insurance but has not provided any proof of the latter; it did not even adduce comparative passages from its own tariffs and those of insurance societies. For my part, I remark that if, contrary to expectation, insurance through savings offices will prove to be cheap for the insured, it will be expensive for the state, provided of course that the business will be widespread since the expenses will then not be low.⁸

Notes

1. Elsewhere Markov [24], only in the edition of 1908, on p. 97, when referring to his letters and to the Explanation [28], contrasted various forms of insurance:

There exist also [!] such insurance operations which do not protect against any risks, and in all cases inflict some greater or lesser damage on the insured. Such operations may be justified [...] only by a rather doubtful consideration that they compel people to save money.

Note, however, that parents (when juvenile insurance is concerned) become directly interested in the pecuniary sense in that their insured children remain alive.

2. Markov bears in mind the employees of the offices, see Letter 2.
3. Interest was paid on the sum rounded down to the nearest hundred.
4. Markov had written down all the intermediate results (here omitted). The same will be true in two other cases below. A rough check of his final figure is provided by calculating $1,200 \cdot 1.04^{14} = 2,078$.
5. Markov refers to his Letter 1. The Directorate maintained that the insurance of juveniles ensures but little profit: it “comes close [...] to simple saving”. Then, private insurance societies offer even worse conditions for the insured; the psychological aspect of being protected from chance by insurance is important; if, after some time, the savings offices show a profit higher than 4%, the surplus will be given over to those insured.
6. More correctly, 2,376. It is obvious that the restriction concerning the interest (Note 3) did not apply to the savings offices themselves.
7. Markov possibly referred to the “Short-Term Treasury Bonds” issued on 9 December 1905 and yielding a 5.5% rate of interest [29, p. 67].
8. It seems that the only explanation here is that “cheap” means “almost fair”.

References

1. Alekseeva, V.P. Bibliography of the works of Markov. In Markov, A.A. *Избранные труды* (Sel. Works). N.p., 1951, pp. 679 – 714.
2. Grodzensky, S.Ya. *A.A. Марков* (Markov). M., 1987.
3. Sheynin, O.B. Early history of the theory of probability. AHES, vol. 17, 1977, pp. 201 – 259.
4. Kohli, K., van der Waerden, B.L. Bewertung von Leibrenten. In Bernoulli, J. *Werke*, Bd. 3. Basel, 1975, pp. 515 – 539.
5. Hald, A. *History of Probability and Statistics and Their Applications before 1750*. New York, 1990.
6. Wells, A.F. Friendly society. *Enc. Brit.*, vol. 9, 1965, pp. 935 – 938.
7. Laplace, P.S. *Théorie analytique des probabilités* (1812). *Oeuvr. Compl.*, t. 7. Paris, 1886.
8. Press, A. Insurance. *Энци. Словарь Брокгауза и Ефрона* (Brockhaus & Efron Enc. Dict.), vol. 62, 1901, pp. 736 – 782.
9. Huygens, C. Correspondance (1669). *Oeuvr. Compl.*, t. 6. La Haye, 1895, pp. 483 and 526 – 538.
10. Bernoulli, N. Dissertatio inauguralis mathematico-juridica de usu artis conjectandi in jure (1709). In Bernoulli, J. (see [4]), pp. 287 – 326.
11. Todhunter, I. *History of the Mathematical Theory of Probability* (1865). New York, 1965.
12. De Moivre, A. *Treatise on Annuities on Lives* (1725). In author’s *Doctrine of Chances* (1756); re-printed New York, 1967, pp. 261 – 328.
13. Sheynin, O.B. Gauss and the theory of errors. AHES, vol. 20, 1979, pp. 21 – 72.
14. Zernov, N. *Теория вероятностей и т.д.* (Theory of Probability with Applications Mainly to Mortality and Insurance). M., 1843.
15. Ostrogradsky, M.V. Note on the retirement fund (1858). *Полное собрание трудов* (Complete Works), vol. 3. Kiev, 1961, pp. 297 – 300.
16. Markov, A.A. *Записка о расчетах и т.д.* (Note on Calculating the Probable Turnovers of the Retirement Fund of the Judicial Department). Psk, 1884, 13pp.
17. *Труды по обзору и т.д.* (Review of Activities of the Retirement Fund of the Ministry of Justice for Its First Five Years), vols 1 – 2. Psk, 1890 – 1891.
18. Markov, A.A. Note on calculating the capitals needed for the operations of the fund [16, vol. 2, pp. 131 – 132]. (R)
19. Lykoshin, A.S. Retirement funds. *Энци. Словарь Брокгауза и Ефрона* (Brockhaus & Efron Enc. Dict.), vol. 40A, 1904, pp. 726 – 729.
20. Markov, A.A. *Записка по вопросам и т.д.* (Note on Points Considered by the 4th Control Commission of the Retirement Fund at the Department of Military Land-Forces). [1899].
21. *Протоколы Имп. Академии Наук* (Proc. Imp. Acad. Sci.), Record No. 1, 19 Jan. 1900. Published as a manuscript.
22. Военное ведомство. *Отчет о денежных оборотах и т.д.* (Report on the Turnover of Retirement Fund, Department of Military Land-Forces for 1899). Psk, 1900.
23. Markov, A.A. Jr, Biography of A.A. Markov, Sr. In Markov, A.A., Sr (see [1]), pp. 599 – 613.
24. Markov, A.A. *Исчисление вероятностей* (Calculus of Probability). Psk, 1900, 1908, 1913; M., 1924. German transl.: Leipzig – Berlin, 1912.
25. Savich, S.E. *Элементарная теория страхования* (Elementary Theory of Insurance). Psk, 1900.
26. Kraevich, K.D. *Собрание алгебраических задач* (Coll. Algebraic problems). Psk, 1864, 1867, 1874, 1882.
27. Nikolsky, P.A. *Основные вопросы страхования* (Main Issues in Insurance). Kazan, 1895.

28. On insurance through the savings offices. Explanations provided by the Directorate of the savings offices. Newspaper *Наша жизнь* (Nasha Zhizn, Our Life), 18 April 1906, p. 2. (R)

29. *Ежегодник Министерства финансов* (Yearbook of the Ministry of Finance). Issue 36 for 1906/1907. Psk, 1907.

19. Slutsky: Commemorating the 50th Anniversary of His Death

IMI, vol. 3 (38), 1999, pp. 128 – 137

Note. The original Russian text lacks §3.3.

1. Introduction

Many authors [7; 27; 1; 4; 5, 32; 8; 20]¹ described the life and work of Evgeni Evgenievich Slutsky (1880 – 1948), an outstanding mathematician, statistician, and economist, and his most important writings are available in a one-volume edition [26]. I am therefore restricting my main goal to publishing or describing a few archival letters either written by, or having to do with him (§3). In addition, I say a few words about Slutsky's life (below) and throw light on the events which apparently compelled him to abandon economics (§2).²

In 1920 Slutsky became Professor at Kiev Commercial Institute. However, he had not mastered the Ukrainian language which was then made compulsory for academic institutions, and in 1926 he had to move to Moscow and to start working there at the Central Statistical Directorate [4, p. 268], and, at the same time, at the Conjuncture Institute under the Finance Ministry [15, p. 8].

Already then Slutsky busied himself in real earnest with applying his statistical research to geophysics. Being forced to abandon his activities in economics (§2), he [4, p. 270], for a few years,

*Went over to working in institutes connected with geophysics and meteorology where he [...] hoped to find application for his discoveries in the field of pseudo-periodic waves.*³

He had not found suitable conditions for theoretical research (Ibidem), and in 1934 he moved to the Moscow State University, then (in 1939) going over to the Steklov Mathematical Institute. The University conferred on him the degree of Doctor of Physical and Mathematical Sciences *honoris causa* [4, p. 271].

Slutsky was an original and deep researcher. He is mostly known as a cofounder of the purely mathematical theory of probability and the theory of random processes, and remembered for his application of stochastic ideas and methods in economics and geophysics (especially in studying solar activity) and as a compiler of important mathematical tables which constituted “a masterpiece of the art of calculation” [27, p. 417].

Slutsky's contribution to the theory of consumer's demand is very valuable [1, p. 210]. For a very long time before his death he (Ibidem, pp. 213 – 214) remained

Almost inaccessible to economists and statisticians outside Russia [...] His assistance, or at least personal contacts with him would have been invaluable.

2. Withdrawal from Economics

In 1927, N.D. Kondratiev, the Director of the Conjuncture Institute, published a critical article concerning the first Five-Year-Plan. Soon he was elbowed out of science, arrested (1931) and then (1939!) shot [10]. N.S. Chetverikov, Kondratiev's assistant, served four years in prison, and, in 1937 or 1938, was subjected to new “repressive measures” [3]. Slutsky apparently had not suffered,⁴ but the general situation in statistics became unbearable. Later Chetverikov [4, p. 270] warily remarked that in 1930

The Conjuncture Institute ceased to exist and the Central Statistical Directorate underwent radical change.

I myself add that, also in 1930, the leading statistical journal, *Vestnik Statistiki*, was closed down and only reappeared in 1948;⁵ during that period only a meager number of statistical papers had been published in *Planovoe Khoziastvo*.

Under the changed social conditions, Maria Smit (more correctly, Falkner-Smit), a statistician of the new wave, became especially useful in spite of her crass ignorance (and in 1939 she was even elected Corresponding Member of the Soviet Academy of Sciences). Pearson, she [30, p. 228] wrote,

Does not want to subdue the real world by a single curve [of distribution] as ferociously as it was attempted by Gaus [Gauss] [...] His system [of curves] nevertheless only rests on a mathematical foundation, and the real world cannot be studied on this basis at all.

She [28, p. 168] also declared that Marxist statisticians should help the state security service in exposing the “saboteurs”. Iastremsky (Ibidem, p. 153) effectively agreed and mentioned D.F. Egorov (who died soon afterwards in his exile in Kazan):

I had recently an occasion to hear out [...] the speech of Prof. Egorov, the then not yet

*exposed saboteur.*⁶ *He came out with a program of sorts saying so ardently, even with a cry in his voice, What are you harping here on sabotage? [...] There are no saboteurs worse than you yourselves, comrades, since you standardize reasoning by popularizing Marxism.*

Also see [22] and [23].

3. Archival Sources

Before adducing the promised letters I list similar and already published archival materials concerning Slutsky.⁷

1) In three of his letters to Chuprov, Markov, in 1912 [13, pp. 53 – 58] criticized Slutsky's book [26]. In the same source (p. 143) the Editor, in his review of the Markov – Chuprov correspondence, quoted a passage from a letter written by Slutsky to Markov. I translated and published this letter in full [22, pp. 45 – 46].

2) I myself [22, pp. 43 – 50] made known a few other archival or hardly known materials:

a) Chuprov's review of Slutsky [26] published in 1912 in a newspaper.

b) Slutsky's scientific character written by Chuprov in 1916.

c) Passages from the correspondence of these scholars with each other.

3) Seneta [21] published English translations of two of Slutsky's letters to his wife concerning the author's appraisal of the comparative contribution of Borel and Cantelli to the discovery of the strong law of large numbers.⁸

3.1. D.A. Grave – A.A. Markov, 4 Nov. 1912, Kiev

Archive of the Soviet Academy of Sciences, Fond 173, Delo 5, No. 1

Highly respected Andrei Andreevich, – I got to know E.E. Slutsky under the following circumstances. I was invited to a sitting of the Society of Economists at K. Comm. [Kiev Commercial] Inst. to attend a report on applying the Pearson theory to statistics. The report was delivered by Slutsky, a young man who had recently graduated from the [Kiev] University with a gold medal awarded for a work on political economy, but, because of some reasons, was not left at the University [to prepare himself for professorship].

I inquired directly of Slutsky's professor of political economy the reasons for this, and his answer surprised me by the justification unusual for a mathematical ear. According to his words, Slutsky is quite a talented and serious scientist, but the professor had not ventured to nominate him for being left at the University because of his distinct sympathy with social-democratic theories. And when I was unable to refrain from stating that at the mathematical faculty the author is not usually asked about his political views, the professor advised me to leave Slutsky at the mathematical faculty. I was naturally obliged to say that I have absolutely no desire to intervene in the business of the law faculty and that I am therefore asking him to leave the mathematical faculty alone. After this encounter Slutsky became my student and protégé. Although I am not at all acquainted with his works and had not understood the mathematical part of his report.

The lawyers, professors at the K. Comm. Inst., who did not understand Slutsky's book [26] but desired to acquaint themselves with the Pearson theory, have asked me to explicate it properly in my course in insurance mathematics [6]. I do not know how to find a way out of this difficult situation: it is simply repulsive to read all this ...

[The sequel has no bearing either on Slutsky or probability and/or statistics. As also below, I myself inserted or specified the bibliographic information provided. For Grave, it was "repulsive" to read Pearson; cf. the now published letter of Slutsky to Markov (below).]

3.2. The Extant Part of the Unsigned and Unaddressed Letter (obviously, from Slutsky to Markov; no date)

Same Archive, Fond 173, delo 18, No. 5

are not independent in magnitude from the sum of the already accumulated deviations or that the probabilities of equal deviations are not constant, we shall indeed arrive at the formula

$$(1/y) dy/dx = x/F(x).$$

In an infinite number of cases (naturally, not always!) F can be expanded into a Taylor series, and the first few (e.g., three) terms will ensure a sufficient approximation. These qualification remarks should have certainly been made.

Only experience can show how often do empirical polygons of distribution, which could with a sufficient approximation be interpolated by a Pearson curve, appear in practice. Much material is already collected for answering this question in the positive. In many cases the Gauss curve will not do since asymmetric polygons are often encountered in practice. Interpolation by parabolic curves

$$y = a_0 + a_1x + a_2x^2 + \dots$$

is unsuitable since these curves do not give an adequate picture at the edges of the figure: it is impossible to ensure their asymptotic approximation to the X axis; in addition, they lead to many superfluous inflexions. The Pearson curves constitute the type that occurred to be practically the most suitable.

Since the Gauss curve in very many cases is well suited for representing statistical facts, especially in anthropology [anthropometry], it seems desirable also for the asymmetric Pearson curves not only to indicate that they are corroborated by practice, but in addition to provide a theoretical derivation that would put this curve [these curves] in the same line as the Gauss curve on the basis of the theory of probability (hypergeometric series).

The derivation on pp. 16 – 17 only serves to make the striking practical suitability of these curves less incomprehensible by means of the hypothesis on the action of infinitely many causes combining semi-randomly one with another.

2) The method of moments. Here, I allow myself to remark that neither Pearson, nor Lakhtin [9] say that they proved that the method of moments brings

$$\int (y - Y)^2 dx$$

to its minimal value. They only prove that the method ensures an approximation. It would have been interesting to investigate this problem and to indicate precisely when is the method of moments applicable, and when it is not. Lakhtin does it, but is he not mistaken?

I think that, *quand meme*, approximate formulas should not be objected to. Indeed, you yourself [11, p. iv] admit that such formulas might be used in probability theory even “without estimating their error” since “the aims of applied mathematics” demand this. You also state that approximate formulas should in addition be created for ensuring the calculations [12, p. 77].⁹ At the same time, the method of moments is very convenient; and, since it is proved to provide an approximation for a large number of types of functions, its critical investigation is desirable. In many cases it is simply indispensable since the method of least squares sometimes leads to intolerable or even unrealizable calculations. If desired, I shall next time illustrate this proposition.

3) The theory of correlation. Here, I shall allow myself for the time being ...¹⁰

3.3. Slutsky's Letters to Karl Pearson

I [22, pp. 46 – 47] published Slutsky's letter of 31 March 1913 to Chuprov. It occurred that Slutsky sent Pearson two manuscripts for publication in the *Biometrika*. Pearson had, however, returned both of them, and Slutsky, considering that he was treated improperly, asked Chuprov's advice. Chuprov recommended that Slutsky submit his work to the Royal Statistical Society, and one of these manuscripts was indeed published by it [25]; the other one, on a modification of the difference method, had not appeared anywhere.

Now, I am able to make known three letters from Slutsky to Pearson;¹¹ Pearson's letters are lost. Slutsky invariably gave his address as the Volodkevich Commercial “Schoole” in Kiev. Volodkevich was the name of his (future?) wife, and I am sure that since 1917 Slutsky never mentioned this private enterprise.

3.3.1. Slutsky – Pearson, 23 April 1912

University College London, Library, Pearson Papers 856/4

Dear Sir, – I am sending for your approval a paper concerning a correction to be made in the theory of contingency. If you find no fallacy in chief results, will not the paper be of some interest to the readers of the *Biometrika*? [!] Should you find any fault making idle the whole of my reasoning, I hope you will not refuse to communicate me your kindly criticism. It is a pleasure to acknowledge beforehand my great debt to you for the slightest of hints on the fallacies possibly made in my work. I am,

Yours faithfully E. Slutsky

P.S. The summary of the results is to be found at the end of the paper.

3.3.2. Slutsky – Pearson, 6 May 1912

Kept at the same place, 856/7

Dear Sir, – I had the pleasure to receive your honored letter on the 3^d May and I must excuse myself for answering so late – the reason is that I wanted much time for translating my letter in English. I thank you very much for your long and very interesting letter and for the proof which I am sorry not to have got yet, probably because it must be censured before I get it. Being you really very thankfull for your suggestiv and

very valuable criticism and agreeing with you in many points, I fear nevertheless that I shall not be able to agree with you about their bearing concerning my main thesis. I think I can keep my ancient opinion about the best method of determining the probability we have in view, though after your letter I feel compelled to change its foundation. I take the liberty to begin with some general considerations and then I shall continue with the question in which we disagree.

1. There is not a single method for the determination of the probability that a given system of frequencies has arisen from random sampling.

A) The theoretical frequencies being known à priori, we can determine the probability of the given system of errors:

$$e_1 = m_1 - \mu_1, e_2 = m_2 - \mu_2, \dots P = Q(\chi_{\mu}^2; n') -$$

in the notation of my paper – where n' is the number of groups,

$$\chi_{\mu}^2 = \Sigma [\mu R_{ii} e_i^2 / \mu R \sigma_{\mu_i}^2] + 2 \Sigma [\mu R_{ij} e_i e_j / \mu R \sigma_{\mu_i} \sigma_{\mu_j}],$$

$$R = \begin{array}{cccc} | & 1 & r_{\mu_1\mu_2} & r_{\mu_1\mu_3} & \dots | \\ | & r_{\mu_2\mu_1} & 1 & r_{\mu_2\mu_3} & \dots | \\ | & \dots & \dots & \dots & \dots | \end{array} \quad (1)$$

Now it is to be remarked that the method, even when applied to the same material, gives us very different results, the value of n' being arbitrary. As you have shown [14, p. 160], by infinitesimal grouping $P = 1$ for any value of χ^2 will appear. There is thus a number of groups n'_m which brings the value of P to the minimum, and I think you will agree that this minimal value of P is that really significant for the probability in question. “Really significant” means but this: we cannot assume a value greater than this P_{minimum} to the probability that the given system of frequencies has arisen by random sampling from the supposed theoretical population.

B) Let

$$\theta_1 = f_1(m_1; m_2; \dots; m_n), \theta_2 = f_2(m_1; m_2; \dots), \theta_q = f_q(m_1; m_2; \dots)$$

be functions of empirical frequencies such that

$$f_1(\mu_1; \mu_2; \dots) = 0, f_2(\mu_1; \mu_2; \dots) = 0, \dots, f_q(\mu_1; \mu_2; \dots) = 0$$

and let $\sigma_{\theta_1}, \sigma_{\theta_2}, \dots, r_{\theta_i\theta_j}, \dots$ be their standard deviations and correlations. Then the probability of our frequency distribution being a random sample of the theoretical population $(\mu_1; \mu_2; \dots; \mu_n)$ can be judged

α) From the probability of the deviation of any θ_i from its zero value. In this case

$$P = \sqrt{2/\pi} \int_{\theta_i}^{\infty} \exp[-(1/2) \theta_i^2 / \sigma_{\theta_i}^2] d\theta_i.$$

β) From the probability of the set of deviations from their zero values of a correlated system of functions $\theta_1; \theta_2; \dots; \theta_q$

$$P = Q(\chi^2_{\theta_1; \theta_2; \dots; \theta_q}; q + 1)$$

where q is the number of independent values $(\theta_1; \theta_2; \dots; \theta_q)$,

$$\chi^2_{\theta_1; \theta_2; \dots; \theta_q} = \Sigma [\theta R_{ii} \theta_i^2 / \theta R \sigma_{\theta_i}^2] + 2 \Sigma [\theta R_{ij} \theta_i \theta_j / \theta R \sigma_{\theta_i} \sigma_{\theta_j}], \quad (2)$$

and R is the same as (1) but with θ_i replacing μ_i .

The question of the relations between the results obtained by different methods seems to me to be a very difficult one. I think, however, that the following propositions hardly can meet objections.

Proposition 1. From all the values $\chi_1, \chi_2, \dots, \chi_s$ that is really significant which gives the least value for P . For ex. [15, p. 280 & 283 – 284]: In the case (1 – 3) – Motion of bright Line – the probability of the fre-

quency distribution being a random sample from the general population distributed normally equals $1/23$ if judged from the value of the criterion χ^2 and it is $< 1/1000$ if the probable error of the skewness will be taken into account.

Proposition 2. Should we take a great number of random samples from the general population and evaluate all values

$$\chi^2 \text{ with indices } \mu, \theta_1, \theta_2, \dots, \theta_q, \theta_i\theta_j, \theta_i\theta_j\theta_k, \dots, \theta_i\theta_j\dots\theta_s, \dots$$

for each random sample, the distribution of each χ^2 must be that indicated by the theory within the errors of random sampling.

Proposition 3. Let us have χ_1^2 (for n_1 independent values $\theta_i; \theta_j; \dots; \theta_k$) and χ_2^2 (for n_2 independent values θ with other indices) and let n_1 not be equal to n_2 . Then it is impossible that for all random samples $\chi_1^2 = \chi_2^2 = \chi^2$ say. Indeed, the theoretical distribution of χ_1^2 as given by $Q(\chi^2; n_1 + 1)$ differs from the theoretical distribution of χ_2^2 as given by $Q(\chi^2; n_2 + 1)$ whereas χ_1 being identical with χ_2 their distributions must and will be also identical.

2. I come now to consideration of the point of our divergence and I confess that "if I write

$$e_p = f_p - N'(f_p + 2f_p) / (N' + N'')$$

I vary the constitution of the general population for each pair of samples I take, whereas it must really be a constant, as we take all pairs of samples".

For consequence χ^2 proposed by me as the criterion of divergency cannot be regarded as *your* criterion for goodness of fit as worked out in your paper [14, pp. 160 – 163]. In the notation of this letter it is not χ^2_μ . But nevertheless it is significant. Let us have a contingency table [Table 1] and let us look upon the values like

$$m_{ij} - N_i' N_j'' / N = \varepsilon_{ij}$$

as on the functions of the group frequencies, *varying from sample to sample*, and becoming all zeros for the general population. Then my criterion of divergency χ^2_ε [Slutsky wrote out the right side of (2) with ε replacing θ]; the corresponding value of

$$P = Q[(\chi^2_\varepsilon; (s-1)(t-1) + 1)]$$

measures the probability "that a given system of deviations from the probable ($\varepsilon_{ij} = 0$) in the case of a correlated system of variables (ε_{ij}) is such that it can be reasonably supposed to have arisen from random sampling". It is quite analogous with my $\chi^2_{\theta_1; \theta_2; \dots; \theta_q}$ and it is easily to be subsumed under your general theory in [14, p. 157 – 160].

Let us suppose there is no correlation in the general population and let a great number of random samples be taken from it. Then the distribution of values of χ^2_ε will be that given by $Q[(\chi^2_\varepsilon; (s-1)(t-1) + 1)]$.

I have shown in my paper that my criterion of divergency (χ^2_ε) for a fourfold table is identical as to its numerical value with your square continugency χ^2_μ . If so both theories cannot be valid as it is shown in the proposition 3 above.

I am not able now to see any error in my reasoning and it seems me the divergence in our views resolves as follows: We do not know the theoretical frequencies and we use "the best available values", i.e. $N_i' N_j'' / N$ as it occurs in many other cases.

(A) I think that they are not the best, and it seems to me you will agree that we should obtain far better values if we have had a theory of skew surfaces. Then fitting such a surface to the system of values like $N_i' N_j'' / N$ and integrating its volume for the base elements of the subgroups we have had indeed the *best* available values.

(B) Yet supposed the values like $N_i' N_j'' / N$ be "the best available", there is still no ground that they are sufficiently *good*, for we can safely use the theoretical values deduced from the sample itself instead of the unknown quantities relating to the general population *only* if their probable errors are sufficiently small. That is the case with the standard deviation, when used to determine the probable error of the mean. In determining the goodness of fit we bring into the comparison the empirical frequencies with the theoretical ones deduced from the sample itself. But in using the method of moments for fitting the curves we reduce largely the probable errors of the theoretical group frequencies so that they become small as compared with the empirical frequencies.

For Ex. the frequency in Gaussian distribution, the base element being h , is $\mu_x \approx yh$ whence $\sigma_\mu / \mu = \sigma_y / y$. But in this case

$$\delta y/y (x^2 / \sigma^3) \delta \sigma, \text{ so that } \sigma_y/y = (x^2 / \sigma^2) \sqrt{2N}.$$

For the empirical frequency m_x we have

$$\sigma_m = \sqrt{m[1 - (m/N)]}, \quad \sigma_\mu/\mu = \sqrt{(1/m) - (1/N)} \approx 1/\sqrt{m}.$$

Let $x = (1/2)\sigma$, $h = (1/8)\sigma$, $N = 450$, $m = \mu$. Then $\sigma_\mu/\mu = \sigma_y/y = 0.008$ and $\sigma_m/m = 0.224$ exceeding by 28 times the preceding value of procentual error of theoretical frequency. Let us take now a fourfold table [Table 2] and suppose the values a, b, c, d be proportional to the values in the general population. Let $a' = (a + b)(a + c)/N$. Then

$$\sigma_a = [a(1 - (a/N))]^{1/2},$$

$$\sigma_{a'} = (1/N) [(a + c)^2 \sigma_{a+b}^2 + (a + b)^2 \sigma_{a+c}^2 + 2(a + b)(a + c) \sigma_{a+b} \sigma_{a+c} r_{a+b, a+c}]^{1/2}$$

where

$$\sigma_{a+b}^2 = (a + b)[1 - (a + b)/N], \quad \sigma_{a+b} \sigma_{a+c} r_{a+b, a+c} = a - (a + b)(a + c)/N.$$

For $a = b = c = d = 12$, $\sigma_a = 3$, $\sigma_{a+b}^2 = \sigma_{a+c}^2 = 12$, $r_{a+b, a+c} = 0$, $\sigma_{a'} = 2.45$.

For $a = b = c = d = 1200$, $\sigma_a = 30$, $\sigma_{a'} = 24.5$.

Thus, taking for the theoretical frequency $(a + b)(a + c)/N$ as determined by *any* random sample and dealing with *every possible* random sample we shall have our errors measured from the point the position of which is subject to errors of random sampling almost so great as the values we are measuring thereof. In consequence we shall obtain the values of χ^2 on the average largely reduced as compared with the case we knew the *à priori* frequencies in the general population. In my paper are given the values of χ^2_ϵ evaluated for random samples obtained by the experiment. The values of ϵ which correspond to the ϵ in the notation of this letter were measured from the theoretical frequencies deduced from the data. If we measure them from the frequencies known in my case *à priori*: $a = b = c = d = 12$, we obtain, as a matter of fact, much greater values (given in the table here apart). If we use the same grouping as before we obtain [Table 3].

This seems to me to confirm my views that your theory is to be applied in the cases where we know the *à priori* frequencies but that in the cases we do not know them your χ^2_μ must be replaced by my χ^2_ϵ which is numerically identical with it, so that the whole difference in the results touches only the value of n' being in the case we use χ^2_ϵ , $(s - 1)(t - 1) + 1$.

It seems to me I have found now more stronger grounds for the proposed modification in the theory and I will be immensely grateful to you if you let me know your views on the matter. Again thanking you for your courtesy I am Yours very faithfully E. Slutsky

Table 1

m_{ij}	...	N_1'
N_2'		
N_1''	N_2''	N

Table 2

a	b	$a + b$
c	d	$c + d$
$a + c$	$b + d$	N

Table 3

Values of χ^2_μ	Theory	Statistics	e^2/μ (on <i>à priori</i> grounds)
			μ m
0 - 0.25	1.54	2	0.14
0.25 - 0.50	2.51	1.5	0.41
0.50 - 1	5.88	3.5	0.96
1 - 2	11.44	13	0.21
2 - 3	9.04	5.5	1.39
3 - ∞	19.58	24.5	1.24

whence $P = 0.50$, $\chi^2 = 4.35$

Table 4

The values of χ^2_{μ} (criterion for goodness of fit) – for the experiment described in my paper – if we use the à priori probabilities: $a = 12$, $b = 12$, $c = 12$, $d = 12$.

N	χ^2_{μ}										
1	1 1/3	10	5 1/6								
2	1 2/3	11	4 1/6								
3	1 1/6	12	3 1/2								
4	2 1/6	13	1 1/2								
5	4 1/6	14	4 1/6								
6	3 5/6	15	3 1/2								
7	1 1/6	16	1 5/6								
8	3 1/6	17	1								
9	1/2	18	1 5/6								

19	1 1/3	27	1 2/3	35	1/6	43	2 1/2
20	4 1/2	28	1 1/3	36	1 2/3	44	2 1/6
21	4 1/2	29	2 1/2	37	5 5/6	45	3 2/3
22	1/2	30	1	38	4 1/6	46	3 1/6
23	5/6	31	6 1/2	39	3	47	5 1/6
24	9 1/2	32	3 1/6	40	3 1/2	48	4 1/2
25	1/2	33	0	41	5 5/6	49	3 2/3
26	2 1/6	34	3 1/6	42	1 1/6	50	5 5/6

3.3.3. Slutsky – Pearson, 18 May 1912

Kept at the same place, 856/4

Dear Sir, – I take the liberty to write you again, before I have your answer on my previous letter. I am printing now a treatise (or a text-book) on the theory of correlation and I would be very grateful to you if you let me know whether the probable error of the partial correlation coefficient can be reduced to the same form as the probable error of the total one, as m^r Yule says.¹²

I have also brought fast [replace this German word by the proper English *almost* – O.S.] to the end a paper on a *General test for Goodness of Fit of the Regression Curves*. To keep your valuable time I do not send it to you and I take the liberty only to communicate you an idea of it you will easily appreciate. It is very simple but I am not able to refer to any previous mention of it.

In the notation of your memoir on Skew correlation [16] the criterion will be simply

$$\chi^2 = S(Y - y_x)^2 / (\sigma_{n_x}^2 / n_x)$$

n' = number of arrays + 1 for there is no correlation between the means of the x -arrays and the probability of a deviation is

$$C \exp [- (1/2) (Y - y_x)^2 / [(\sigma_{n_x}^2 / n_x)] \delta (Y - y_x)].$$

Quite analogous will be a criterion which can be applied in the physical sciences to test the probability that a given system of measurements can reasonably be supposed to correspond to the proposed functional rela-

tionship. If you will agree with this I can send you a more elaborate – but still a short paper – with the illustrations taken from your memoir on skew correlation [16].

I excuse myself, dear sir, for my very imperfect English and for the trouble I give you and remain very faithfully yours E. Slutsky

3.3.4. Slutsky's Letter to Aleksandr Nikolaevich Shchukarev, a specialist in physical chemistry (1928)

Archive of the Moscow State Univ., Fond 276, Inventory 1, No. 114

Slutsky made known his opinion about Shchukarev's unnamed paper, perhaps answering the latter's request. This paper [19], which I located without much difficulty, was written extremely carelessly. In essence, Shchukarev vainly attempted to derive the Maxwellian law without introducing any stochastic ideas and it is therefore sufficient to say only a few words about Slutsky's reply.

Slutsky indicated that Shchukarev had not nevertheless managed without stochastic considerations; admitted (perhaps too modestly) that he "hardly understands" physics but "somewhat catches" the logical structure of "suchlike theories"; and offered concrete remarks (unnecessarily since the paper was beyond repair).

Notes

1. Short anonymous and hardly differing articles on Slutsky are included in the 2nd and 3rd editions of the *Bolshaia Sovetskaia Enziklopedia*; the 3rd edition is available in an English translation (entitled: *Great Sov. Enc.*). My references do not at all exhaust the literature on him. Sarymsakov [18] praised his work in geophysics, and the authors of several sections of [31] described his mathematical achievements. Romanovsky [17] indicated that Slutsky was chairman of a commission on applying statistical methods in industry (as a young man he studied for a few years at the machine-building department of the Munich polytechnical school [4, p. 262]). It seems, however, that because of the negative attitude of the Soviet establishment towards statistics in general (§2) that commission was unable to be of essential use.

2. For a background to this section see [23].

3. Slutsky had been applying these discoveries mostly to economics, and his transition to other branches of knowledge was painful: disallowing a report that appeared in 1932 but was delivered by Slutsky in 1928, he had not published anything during 1930 – 1932. I also note that an English translation of his paper of 1927 was published in 1937. It found important application in investigating time series in economics [1, pp. 209 – 210].

4. In 1990 the eminent mathematician Konüs told me that at the time he had also worked at the Conjunction Institute. He was left alone; as Könus explained the attitude of those responsible for the decision-making, they had decided: "He is only a mathematician, not responsible for anything..."

5. In 1929 a paper by the mathematician and statistician N.V. Smirnov appeared in the *Vestnik* and Slutsky even before his move to Moscow had published four articles there.

6. Smit [29, p. 4] clumsily declared that "the crowds of arrested saboteurs are full of statisticians". Anderson, a student of Chuprov, testified [2, p. 294]:

Könnte ich [...] eine ganze Reihe von in Rußland früher sehr geschätzten Statistikern und viel verschprechenden jüngeren Schülern [...] Tschuprows aufzählen, deren Namen nach 1930 aus der sowjet-russischen wissenschaftlichen Literatur plötzlich ganz verschwanden.

7. I also stress that Chetverikov [4] mentions and quotes Slutsky's biography written by his wife, Yu.N. Volodkevich (p. 265), as well as another biography written by Slutsky himself (pp. 267 and 271). In turn, Gnedenko [5, p. 6] quotes Slutsky's autobiography compiled in 1938. They do not provide any information about these sources. Recall (§1) that in 1939 Slutsky started working at the Steklov Mathematical Institute.

8. Also see [4, p. 269]. In 1970 Chetverikov had given me (Russian) typed texts of these letters which I turned over to Seneta (their copies are regrettably lost). Seneta acknowledged my help in obtaining "important materials" but had not elaborated. He was concerned that I could have had problems with the Soviet authorities.

9. Slutsky obviously referred not to the paper itself as put out in the *Matematich. Sbornik*, but to its previously published offprint. Indeed, he mentioned the year 1911 and p. 4 neither of which agree with the periodical. The appropriate page numbers in the translation (see References) are 77 and 78.

10. Slutsky discusses the Pearson curves. At the time (and even in 1928, in his letter to Shchukarev, see §3.3.4, which I only describe but do not quote) he sometimes wrote "theory of probability" instead of the correct Russian "... of probabilities".

Slutsky derived the equation (see beginning of letter) in his book [26, p. 17]. Also there (pp. 15 – 17 rather than 16 – 17) he obtained the normal distribution as the limiting law for the binomial distribution. Assume the unknown law (Y) as, for example, a polynomial of the n -th degree, then, in principle, its $(n + 1)$ parameters can be determined given the appropriate moments. If the class to which Y belongs is not restricted, its unique determination is impossible even if “all” the moments are given. Slutsky’s question apparently touched on this *problem of moments*.

11. For some reason the pressmarks of two of the letters are identical.

12. Slutsky’s reference is *J. Roy. Stat. Soc.*, 1907, pp. 6 and 47. In both these cases Yule was a participant in discussing the contributions of other authors. The paper that Slutsky mentions just below is apparently [25].

References

1. Allen, R.G.D. The work of Eugen Slutsky. *Econometrica*, vol. 18, 1950, pp. 209 - 216.
2. Anderson, O. Mathematik für marxistisch- leninistische Volkswirte. *Jahrb. f. Nat.-Ökon. u. Statistik*, Bd. 171, 1959, pp. 293 – 299.
3. Anonymous, Anniversaries and memorable dates. *Voprosy Statistiki*, No. 11, 1995, p. 77. (R)
4. Chetverikov, N.S. Life and scientific work of Slutsky (1959). In author’s book *Статистические исследования* (Stat. Investigations). M., 1975, pp. 261 – 281. Translation to appear in 2004 in the DHS series.
5. Gnedenko, B.V. Slutsky. [26, pp. 5 – 11]. (R)
6. Grave, D.A. Insurance mathematics. *Izvestia Kievsk. Kommerch. Inst.*, book 16, 1912, pp. i – iv + 1 – 88 of second paging. (R)
7. Kolmogorov, A.N. Slutsky. Obituary (1948, in Russian). Transl.: *Math. Scientist*, vol. 27, 2002, pp. 67 – 74.
8. Könus, A.A. Slutsky. *Intern. Stat. Enc.*, vol. 2. Editors, W.H. Kruskal, Judith M. Tanur. New York – London, 1978, pp. 1000 – 1001.
9. Lakhtin, L.K. On the Pearson method etc. MSb, vol. 24, 1904, pp. 481 – 500. (R)
10. Makasheva, N. N.D. Kondratiev. Brief biogr. essay. *Mirovaia Ekonomika i Mezhdunarodn. Otnosheniia*, No. 9, 1988, pp. 59 – 61. (R)
11. Markov, A.A. *Исчисление вероятностей* (Calculus of Probabilities), 2nd edition. Psb, 1908.
12. Markov, A.A. A rebuke to P.A. Nekrasov. MSb, vol. 28, 1912, pp. 215 – 227. Transl.: DHS 2579, 1998, pp. 77 – 82.
13. Ondar, Kh.O., Editor, *Correspondence between Markov and Chuprov* (1977, in Russian). New York, 1981.
14. Pearson, K. On a criterion that a given system of deviations from the probable etc. *Phil. Mag.*, vol. 50, 1900, pp. 157 – 175.
15. Pearson, K. On the mathematical theory of errors of judgement etc. *Phil. Trans. Roy. Soc.*, vol. A198, 1902, pp. 235 – 299.
16. Pearson, K. *On the general theory of skew correlation etc.* Drapers’ Co. Res. Mem., Biometric Ser., 2, 1905.
17. Romanovsky, V.I. On the application of math. statistics and the theory of probability in the industries of the Soviet Union. *J. Amer. Stat. Assoc.*, vol. 30, 1935, pp. 709 – 710.
18. Sarymsakov, T.A. Statistical methods and problems in geophysics. *Второе всесоюзное совещание по математич. статистике* (Second All-Union Conf. on Math. Statistics). Tashkent, 1948, pp. 221 – 239. (R)
19. Schükarev, A.N. Ein Versuch der Ableitung des Maxwellischen Verteilungsgesetzes. *Phys. Z.*, Bd. 29, No. 6, 1928, pp. 181 – 182.
20. Seneta, E. Slutsky. *Enc. Stat. Sci.*, vol. 8. Editors, S. Kotz, N.L. Johnson. New York, 1988, pp. 512 – 515.
21. Seneta, E. On the history of the strong law of large numbers etc. *Hist. Math.*, vol. 19, 1992, pp. 24 – 39.
22. Sheynin, O. *Chuprov. Life, Work, Correspondence.* (1990, in Russian). Göttingen, 1996.
23. Sheynin, O. Statistics in the Soviet epoch. *Jahrb. f. Nat.-Ökon. u. Statistik*, Bd. 217, 1998, pp. 529 – 549.
24. Slutsky, E.E. *Теория корреляции* (Theory of Correlation). Kiev, 1912.
25. Slutsky, E.E. On the criterion of goodness of fit of the regression lines etc. *J. Roy. Stat. Soc.*, vol. 77, 1914, pp. 78 – 84.
26. Slutsky, E.E. *Избранные труды* (Sel. Works). M., 1960.
27. Smirnov, N.V. Slutsky. *Izvestia Akad. Nauk SSSR*, ser. matematich., vol. 12, 1948, pp. 417 – 420. (R)

28. Smit, M. Planned sabotage and the statistical theory. *Planovoe Khoziastvo*, No. 10, 1930, pp. 139 – 168. (R) Incorporates several reports including that of B.S. Iastremsky.
29. Smit, M. *Теория и практика советской статистики* (Theory and Practice of Soviet Statistics), 2nd edition. M. – L., 1931.
30. Smit, M. Against the idealistic and mechanistic theories in the theory of Soviet statistics. *Planovoe Khoziastvo*, No. 7, 1934, pp. 217 – 231. (R)
31. Stokalo, I.Z., Editor, *История отечественной математики* (History of National Mathematics), vol. 4, pt. 2. Kiev, 1970.
32. Youshkevich, A.A. Slutsky. *Dict. Scient. Biogr.*, vol. 12. Editor, C.C. Gillispie. New York, 1975, p. 461.

20. History of the Theory of Errors

IMI, vol. 5 (40), 2000, pp. 310 – 332

1. Introduction

During the last 40 years I published many writings partly or completely devoted to the history of the theory of errors [1 – 35] and what follows is their résumé. The last three of these contributions [33 – 35] appeared later than this paper and I do not refer to them below. My main writing on the present subject is [30].

I understand the method of least squares (MLSq) both as a certain condition for solving redundant systems of linear equations (§2) and as the appropriate formulas for estimating the precision of the observations and of the computed values of the (estimates) of the unknowns. MLSq was developed by Gauss in 1823, but, keeping to traditional terminology, I also apply this term with regard to his contribution of 1809. Then, the term *normal distribution* which I naturally use had appeared in 1873 (C. Peirce [36]) and was definitively introduced into the nascent mathematical statistics by Pearson in 1894. Finally, I omit the three last words in the expression *density law of observational errors*.

2. Aims and Branches [23]

The theory of errors aims at establishing expedient patterns and methods of observation and at determining their most plausible results. Thus, when intersecting station C from given stations A and B in the field it is required to find the optimal form of the triangle ABC (ensuring the least possible influence of the errors of measurement upon the coordinates of C) and to choose appropriate methods of observation and calculation (so that the systematic errors be eliminated from the appropriate means as much as possible and the random errors of the observation and of the final results be sufficiently small).

Formally speaking, the theory of errors estimates k unknown constants x, y, z, \dots and determines the precision of their determination given “physically” independent observations s_1, s_2, \dots, s_n ($n > k$) and the coefficients of the equations

$$a_i x + b_i y + c_i z + \dots + s_i = 0, \quad i = 1, 2, \dots, n. \quad (1)$$

The approximate values of x, y, z, \dots are known (for example, from a subsystem of (1)) so that the linearity of the equations is warranted. Such an estimation is called *adjustment of observations*, – of *direct* or *indirect* observations for the cases $k = 1$ and $1 < k < n$ respectively. Systems (1) are inconsistent and any vector (x_0, y_0, z_0, \dots) leading to reasonably small residual free terms, call them v_i , is therefore considered as their solution. Thus, the principle of least squares is defined by the condition

$$\sum v_i^2 = \min \quad (2)$$

with the minimum being sought among all such vectors. In the case of direct observations the estimator sought coincides here with the arithmetic mean of the observations s_i .

Neither the search for expedient patterns of observation (e.g., for the optimal form of the triangle, see above), nor the application of methods of revealing and excluding systematic errors demand stochastic considerations, and the appropriate, the *determinate*, branch of the theory of errors (§10) is akin to the exploratory data analysis [37] and design of experiments, the two new disciplines that might at least partly be attributed to applied statistics.

Random errors are random variables and the *stochastic* theory of errors is based on the theory of probability whose development in the 19th century, and even to the 1920s, was mainly determined by the requirement of adjusting observations. Thus, while describing his own treatise of 1912 (first edition 1896) on probability theory, **Poincaré** [38, p. 343] stated that “La théorie des erreurs était naturellement mon [his] principal but” and **Lévy** [39, p. vii] indicated that without the error theory his book, – his main contribution on stable laws, – “n’aurait pas de raison d’être”.

1.2. The Stages

a) The stochastic branch. During the first stage, perhaps up to **Tycho Brahe**, astronomers were dealing with their observations as they saw fit. To him we are indebted for the first ideas and methods concerning preliminary treatment of observations, and to **Galileo**, for discussing the elements of the theory of errors. At the second stage, observations ceased to be the “private property” of astronomers, but their treatment had not yet been corroborated by quantitative considerations. The third stage, that somewhat overlaps the second one, began with **Simpson** and **Lambert**. Simpson proved that, for two distributions, the arithmetic mean was stochastically preferable to a single observation and Lambert introduced the principle of maximum likeli-

hood, adjusted indirect observations and estimated (imperfectly) their precision, and founded the basis of the theory of errors. Finally, **Euler** anticipated the principle of least squares. During the fourth stage, **Laplace** and **Gauss** completed the development of the classical error theory. Gauss was especially meritorious: unlike Laplace, he studied the treatment of a finite number of observations without applying limit theorems and presented his findings in a form suitable for application.

b) The determinate branch. Before the 18th century, only general methods for minimizing the influence of errors were known. During that century, the effect of observational errors upon the final results began to be determined by differential formulas and narrow formal definitions of random and systematic errors were provided. At the beginning of the 19th century **Laplace** examined the precision of simple patterns of geodetic networks. **Gauss** and **Bessel** originated a new stage in experimental science by thoroughly studying all possible instrumental errors and defects of observational methods and by recommending appropriate measures. Finally, **Helmert** began to investigate geodetic networks in much more detail and formulated the problem of attaining a given (or the highest possible) precision under minimal (or given) expenses of time and money.

1.3. Fields of Application

Apart from metrology (whose boundaries with physics and chemistry are fuzzy) both branches of the theory of errors are applied in practical astronomy and geodesy. The determination and specification of the parameters of the Earth's ellipsoid of revolution by meridian arc measurements, and, since the 20th century, by means of geodetic networks remains a fundamental problem of the two last-mentioned disciplines. Furthermore, new most important practical problems have also appeared. In the 19th century the flattening of the Earth's ellipsoid began to be also determined by pendulum observations and nowadays gravimetry, that is certainly unable to manage without the error theory, studies in addition the exterior gravitational field of the Earth. An absolutely new field of application for the theory is the treatment of geodetic observations made by means of artificial satellites.

2. Ancient Astronomy: Its Features [12; 22]

a) Ancient astronomers attempted to determine reasonable boundaries for the possible values of the observed magnitudes. Suppose that observations s_1, s_2, \dots, s_n of the observed constant s are arranged in an ascending order, then s_1 and s_n can determine these boundaries; however, with an increasing n the interval between these observations tends to

increase. It should therefore be checked against some indirect or theoretical considerations, and astronomers had indeed allowed for previous observations.

b) The second feature of ancient astronomy was the understanding of the need to observe regularly. Thus, **Hipparchus** regularly observed the length of the tropical year.

c) Astronomers attempted to observe under optimal conditions, when unavoidable errors least corrupted the final results. Neugebauer [40, p. 101] remarked that observations in ancient times "were more qualitative than quantitative". This is not true literally, but in general ancient science had indeed been qualitative (cf. next point).

d) Commentators agree in that **Ptolemy** arbitrarily selected and treated observations, and many believe that he appropriated Hipparchus' observations. Yes, he possibly did, but he apparently acted in the spirit of his times without considering it disgraceful. And I think that his arbitrariness in treating observations may be explained. Neither the arithmetic mean, nor any other estimator of the constants sought was then generally accepted; even **Al-Biruni**, 900 years later, did not keep to any standard method when adjusting his metrological measurements. So, had not Ptolemy, had not his contemporaries believed that any number within the interval $[s_1; s_n]$ can equally (although also allowing for indirect considerations and information) serve as an estimate of the unknown constant? From a modern point of view, when having observations of low precision, it is indeed possible to rely on only one, and to disregard all the rest of them. This can be justified by a reference to the Cauchy distribution whose variance does not exist at all; for observations obeying this law, their arithmetic mean is not better than any one of them.

Two authors [41; 42, p. 6] to some extent corroborate my assumption. The former remarked that, as a cartographer, Ptolemy was mainly concerned with semblance of truth (I would say: with general correctness) rather than with mathematical consistency. The latter stated that even in the Middle Ages maps "may well have been made from general knowledge of the countryside without any sort of measurement".

3. From Al-Biruni to Kepler [12]

3.1. Al-Biruni [20]

His writings contain many pronouncements concerning the treatment of observations. Thus, when determining the longitudinal difference between two stations by simultaneous observations of a lunar eclipse, he obviously attempted to exclude systematic errors from his final result. Elsewhere he qualitatively reasoned on

the combined effect of observational and calculational errors. As an observer and calculator he surpassed his predecessors and was a worthy forerunner of Galileo and Kepler.

3.2. Galileo

He was the first to discuss the elements of the theory of errors. Buniakovsky mentioned this fact much too concisely and had not provided the exact reference. Maistrov [43, Chapt. 1] described Galileo's ideas in detail but superficially; and Hald [44, §10.3] furnished a comprehensive account. Galileo attempted to determine the diurnal parallax of the New Star of 1572, or, more precisely, to decide whether it was situated "below" the Moon, between the Moon and the stars, or among the stars. He collected observations of many astronomers and was able to arrive at the correct answer. In the process, he formulated the properties of the "usual" random errors and indicated that the parallax ought to be determined under the condition that the sum of the absolute corrections to observations be minimal, cf. §4.3.

3.3. Tycho Brahe

He was the first astronomer of the new time to carry out long series of observations of the planets and stars. Although direct proof is lacking, it is hardly doubtful that he took reasonable measures to exclude random and systematic errors from his observations. And he essentially developed the technique of observations and raised their plausibility. It was on the basis of his observations that **Kepler** was able to construct a new (the new) system of the world.

3.4. Kepler [13]

While attempting to fit in Tycho's observations with the Ptolemaic system of the world, Kepler noted that the discrepancies between them and the theory reached 8' which was inadmissible. He apparently tried out several versions of adjusting the observations, and it might be thought that he applied the principle of minimax demanding that the maximal absolute correction (more precisely, corruption) of the observations or their functions be minimal. The minimax method is now used not in the theory of errors but in statistical decision making.

But how had Kepler calculated these corruptions? He [45, p. 334] assigned them arbitrarily but "within the limits of observational precision". And he should have somehow considered the properties of "usual" random errors, – somehow selected a larger number of corruptions small in absolute value and about the same number of positive and negative corruptions. Kepler apparently approached this problem in the spirit of modern statistical simulation.

After collecting four values of the right ascension of Mars, s_1, \dots, s_4 , say, Kepler [45, p. 200] assumed as his final result some number s calling it *medium ex aequo et bono*. He obviously selected the generalized arithmetic mean with weights $p_1 = p_2 = 1, p_3 = 2$ and $p_4 = 0$. More interesting, his Latin expression meant *in fairness and justice* and implied (as it did in one of Cicero's writings) *rather than according to the letter of the law*. Thus, already in Kepler's time the (ordinary) arithmetic mean became an universal estimator. The first to state, although not clearly enough, that the arithmetic mean was an optimal estimator was **Cotes** and **Laplace** testified that "tous les calculateurs" were keeping to [kept following] his "rule".

In one of his letters of 1627 Kepler remarked that the total weight of a large number of coins of the same coinage [almost] did not depend on the inaccuracies in the weights of the individual coins. This proposition (involving, however, the mean weight) can be justified by a reference to the law of large numbers.

4. The Eighteenth Century [4; 9]

4.1. Tobias Mayer [21; 24]

In 1750, he was the first to adjust indirect observations in a formalized way. Having obtained 27 equations in three unknowns, he separated them into three equal groups, added together all the equations in each group and solved the thus obtained system. If, for example, the first group consisted of the nine first equations, then, as he tacitly assumed (notation as in §2),

$$v_1 + v_2 + \dots + v_9 = 0. \quad (3)$$

Mayer was mostly interested in only one unknown, call it x , and, accordingly, all the equations in his first group were those having the largest positive coefficients a_i whereas all the a_i 's in the second group had largest negative values. We may assume that he thus attempted to secure the largest possible weight of x .

Mayer also indicated that it was too difficult to solve all the possible subsystems of three equations each (so as to calculate the pertinent means over all of them). For the case of two unknowns such a procedure had indeed been done (**Boscovich**), and this suggests that scholars had wished to approach the adjustment of both direct and indirect observations in the same way. **C.G.J. Jacobi** and **Binet**, independently, proved that calculation taking into account all of the *appropriately weighted* subsystems was tantamount to the least-squares

solution of the initial system; no weights were, however, introduced. Note that condition (3) is exactly fulfilled when applying the MLSq if at least one coefficient (a_i , or b_i , or ...) is identical in all the equations.

4.2. Lambert [2; 6]

Lambert classified observational errors in accord with their origin; attempted to prove that the extreme observations ought to be rejected and that the arithmetic mean was the best estimator of the constant sought; like **Mayer**, estimated the precision of observations (although, again like his predecessor, unfortunately); introduced a continuous unimodal density curve (without specifying it) and formulated the principle of maximal likelihood.

He also adjusted indirect observations actually applying condition (3) and suggested the term *Theorie der Fehler* that became generally accepted in the mid-19th century; neither Laplace nor Gauss ever used it, but Bessel applied it without mentioning anyone. Lambert was the first to discuss the issues of the theory of errors in detail.

4.3. Boscovich [11; 15; 21]

Not being satisfied with the existing methods of treating indirect observations (§4.1), Boscovich applied conditions

$$v_1 + v_2 + \dots + v_n = 0, \quad |v_1| + |v_2| + \dots + |v_n| = \min \quad (4a, b)$$

and provided pertinent qualitative arguments. The restriction (4a) was not essential (by eliminating one unknown it was possible to ensure its fulfilment); the second one, (4b), determined the median in the case of direct observations, and, in the general case, led to exactly k zero values of v_i , as **Gauss** remarked in 1809 (without taking into account condition (4a)). Note that he thus knew an important theorem in linear programming.

Boscovich applied his method for determining the (two) parameters of the Earth's ellipsoid of revolution by meridian arc measurements. Eliminating one of them (see above), he proposed a geometric trick for determining the second one whereas **Laplace**, in 1818, used an algebraic method. Both scholars remarked that one of the v_i 's had vanished.

In an undated manuscript Boscovich considered the most simple cases of summing discrete observational errors; he thus possibly (and to a small extent) forestalled Simpson (§4.4).

4.4. Simpson [10; 12]

In 1756 – 1757 Simpson proved that for the discrete uniform law, and for both the discrete and continuous triangular distribution, the arithmetic mean was stochastically preferable to a single observation. He actually considered random errors as random variables whose formal (and hesitating, for that matter) introduction was only due to Poisson. Thus, Simpson applied densities in the error theory; furthermore, he effectively used generating functions. For passing on to the continuous law he replaced the discrete values of errors

$$-v, -(v-1), \dots, -1, 0, 1, \dots, (v-1), v$$

by $-kv, -k(v-1), \dots$ with $k \rightarrow 0, v \rightarrow \infty$ and $kv = \text{Const.}$

4.5. Lagrange [10]

Without mentioning Simpson, Lagrange, in 1776, considered, with the same aim, a number of discrete and continuous distributions, some of them (for example, the cosine law) bearing no relation to the error theory. He applied the term *courbe de la facilité des erreurs* replaced by Laplace by *courbe des erreurs* (or, *des probabilités*). Instead of a limiting transition from discrete to continuous (§4.4), Lagrange was able to apply a (modified) method of generating functions by introducing “generating integrals” [46, p. 170].

4.6. Daniel Bernoulli [8; 24]

In 1778 Daniel Bernoulli objected to the arithmetic mean and proposed instead the maximum likelihood estimator (\hat{e}) for density

$$y = r^2 - (\hat{e}^2 - x^2), \quad y \geq 0$$

where r was the greatest possible error. He arrived at the generalized arithmetic mean

$$\hat{e} = (p_1 x_1 + p_2 x_2 + \dots + p_n x_n) / (p_1 + p_2 + \dots + p_n) \quad (5)$$

of n observations x_i with

$$p_i = [r^2 - (\hat{e} - x_i)^2]^{-1}. \quad (6)$$

Already for $n = 3$ Bernoulli had to solve an algebraic equation of the fifth degree. The posterior weights p_i (6) increased with x_i moving away from the central group of observations. He had not mentioned this fact, whereas, in accord with the context, the opposite should have been expected. The unusual behavior of the weights can now be justified for some distributions.

In 1780 Bernoulli, while examining the precision of pendulum observations, isolated random (*momentaneorum*, obeying the normal distribution) and systematic (*chronicarum*, constant) errors. He introduced the normal distribution in 1770 – 1771 in his study of the sex ratio at birth and it was not at all difficult for him to apply it in the new context. If, in a simplified pattern, the probabilities of the birth of both sexes are identical, the probability that out of $2N$ newly born there will be about m boys is

$$P(m = N \pm \mu) \approx c \exp(-\mu^2/N).$$

Later, in 1780, Bernoulli wrote out an equality of the same type for $2N = 86,400$ oscillations of the pendulum per day with $(N + \mu)$ of them being too slow, having period $(1 + \alpha)$ sec, and $(N - \mu)$ fast oscillations with period $(1 - \alpha)$ sec. Both earlier and in 1780 Bernoulli specifically considered the case of $P(0 \leq \mu \leq r) = 1/2$. In the second case he obtained $r = 100$ and determined the probable error (Bessel's term of 1816) of the sum of daily oscillations.

4.7. Euler [7; 21]

In a commentary of 1778 to Daniel Bernoulli's memoir of the same year Euler objected to the principle of maximum likelihood. Mistakenly thinking that Bernoulli determined the weights as

$$p_i = r^2 - (\hat{e} - x_i)^2,$$

he reduced the expression (5) to a cubic equation

$$n\hat{e}^3 - \hat{e}(nr^2 - 3B) - C = 0 \quad (7)$$

where \hat{e} was not anymore the maximum likelihood estimator,

$$B = a^2 + b^2 + c^2 + \dots, \quad C = a^3 + b^3 + c^3 + \dots,$$

$$a + b + c + \dots = 0, \quad x_1 = \Pi + a, \quad x_2 = \Pi + b, \quad x_3 = \Pi + c, \dots$$

Finally, Euler noted that his equation (7) can be derived from

$$[r^2 - (\hat{e} - a)^2]^2 + [r^2 - (\hat{e} - b)^2]^2 + [r^2 - (\hat{e} - c)^2]^2 + \dots = \max.$$

If the fourth powers of the errors are disregarded, Euler's condition will be tantamount to the principle of least squares!

5. Laplace [4; 9; 10; 15; 16]

Laplace's contributions to the theory of errors can easily be separated into two groups. Until 1810, he examined various criteria for deriving estimators of the true values of the constants sought (cf. §9) by issuing from more or less arbitrary densities. He invariably arrived at involved equations and was compelled to restrict his studies to the case of three observations. Then, after non-rigorously proving several versions of the central limit theorem (CLT; a term due to Polya, 1920), he became able to consider a large number of observations, and began to apply one and the same criterion (see below) for selecting the estimators. Finally, however, in 1818 Laplace abandoned it and chose, even before Gauss did, the condition of least variance (though only in connection with the normal distribution).

5.1. Early Memoirs [15]

In 1774 Laplace, assuming an arbitrary analytic condition, selected the density

$$\psi(x) = (m/2) \exp[-m|x|]$$

or (as I add), more precisely,

$$\psi(x) = (m/2) \exp[-m|x - e|]$$

where e was the unknown (location) parameter. Introducing the product

$$f(x) = \psi(x; x_1) \psi(x; x_2) \psi(x; x_3) \quad (8)$$

where x_i were the observations, he determined an estimator of e by applying two different conditions, viz., by assuming that it was the median with regard to the curve (8) and that

$$\int_{-\infty}^{\infty} |x - e| f(x) dx = \min.$$

It occurred, however, that in both cases the results were identical.

In 1781 Laplace again issued from a curve of the type of (8), but in addition to the two previous conditions he considered two more criteria, namely, that the median was determined with respect to $x f(x)$, and that $f(x) = \max$ (principle of maximum likelihood). When examining a general stochastic problem, Laplace also non-rigorously derived the density

$$y = (1/2a) \ln[a/|x|], |x| \leq a, \quad (9)$$

or, as I myself add,

$$y = (1/2a) \ln[a/|x - e|], |x - e| \leq a.$$

When attempting to estimate e , he introduced, on a “physical” level, the Dirac delta-function. Regrettably, however, in the language of distributions his reasoning becomes meaningless.

I believe that Laplace understood well enough that his derivations were non-rigorous (for example, the function (9) did not exist at $x = 0$) but regarded them as preliminary and provisional attempts.

5.2. The Memoir of 1810

Here Laplace considered n observations with errors obeying an arbitrary distribution possessing variance. Applying an analog of a characteristic function and the inversion formula, he derived a version of the CLT. His calculations were non-rigorous and extremely careless and apparently only his intuition enabled him to arrive at the correct result.

In a supplement Laplace justified the principle of least squares without having recourse to the Gaussian postulate of the arithmetic mean (§7.1). Nevertheless, he had to admit another assumption, and an artificial one at that.

5.3. The Memoir of 1811

Here, Laplace applied indeterminate multipliers (q_i) for solving equations in one unknown

$$a_i x + s_i = \varepsilon_i, \quad i = 1, 2, \dots, n$$

where ε_i were observational errors rather than residual free terms. Multiplying each i -th equation by q_i and summing the equations, he obtained

$$x = [sq]/[aq] + [\varepsilon q]/[aq] \equiv [sq]/[aq] + g$$

where, in Gauss' notation (regrettably not adopted either by Laplace or other French scholars including Poisson),

$$[ab] = a_1 b_1 + a_2 b_2 + \dots + a_n b_n.$$

Assuming an even distribution with a variance and one and the same order for his multipliers, Laplace non-rigorously proved that the error g was normally distributed and determined the q_i 's by demanding that the error possessed least absolute expectation. He got

$$q_i = \mu a_i, x = [as] / [aa]$$

which corresponded to the principle of least squares (for direct observations). Finally, Laplace generalized his account onto the case of several unknowns.

5.4. *Théorie analytique des probabilités*

In this monograph of 1812 Laplace again determined the distributions of several functions of (independent) observational errors ε_i , – namely $\sum \varepsilon_i$, $\sum |\varepsilon_i|$, $[\varepsilon_i \varepsilon_i]$ and $[q \varepsilon]$, this last being a linear function with coefficients of one and the same order, – identically distributed on a finite interval. He obtained the pertinent normal laws and repeated his previous (1811) justification of the principle of least squares.

A few words about the first two supplements to the *Théorie* (1816 and 1818). In 1816, the most interesting was Laplace’s oblique indication that the sample variance and the arithmetic mean for (independent) normally distributed errors were independent. He thus knew the so-called Student – Fisher theorem. In 1818 Laplace estimated the precision of triangulation by the “closings” of the triangles and derived the density of the measure of precision of the normal law (thus regarding that measure as a random variable). Then, when considering the Boscovich method of adjusting observations (§4.3), he compared the expediency of the arithmetic mean and the median depending on the pertinent law of distribution and in essence assumed the condition of least variance as his test.

6. The Principle of Least Squares

6.1. Legendre [21]

Justifying his proposal by qualitative arguments, Legendre, in 1805, adjusted indirect observations by the principle of least squares and indicated that the greatest absolute correction should be minimized. Note, however (Gauss, 1809), that this minimax condition follows from the generalized principle of least squares

$$\lim (v_1^{2p} + v_2^{2p} + \dots + v_n^{2p}) = \min, p \rightarrow \infty.$$

6.2. Adrain [1; 17]

In 1809 Adrain provided two very non-rigorous proofs that observational errors follow the normal law. As a corollary, he easily arrived at the principle of least squares. John Herschel, and then Maxwell, in 1860, essentially repeated one of his proofs. Adrain had published his contribution in a hardly known American periodical, that, moreover, almost immediately ceased to exist. European mathematicians only came to know about his work in the 1870s.

7. Gauss [4; 12; 16; 17; 21; 24; 32]

7.1. The First Attempt

Gauss had been applying the principle of least squares from 1794 or 1795 onward and published its first justification in 1809. Postulating the coincidence of the arithmetic mean and the maximum likelihood estimator for a unimodal, symmetric and differentiable density $\psi(x - x_0)$, he proved that this density was normal and that the principle of least squares followed at once. Thus, according to Gauss, the least-squares estimators were “most probable”. Nevertheless, he was (or at least became) unsatisfied with these findings. Indeed, his postulate was not evident; random errors could have possessed other densities as well; and the principle of maximum likelihood was less reliable than an integral criterion of the type

$$\int_{-\infty}^{\infty} \psi(x - x_0) f(x) dx = \min. \quad (10)$$

7.2. Mature Thoughts

In 1823, having selected condition (10) with $f(x) = x^2$ as a criterion, Gauss proved that the corresponding estimators were determined by the MLSq (and called them “most plausible”). He also derived the formula for estimating the sample variance by the residual free terms v_i of the initial equations (§2)

$$\sigma^2 = [vv] / (n - k) \quad (11)$$

and determined (with a mistake [47; 48]) the boundaries of the variance of this estimator. The solution of this last problem demanded the knowledge of the density; otherwise, Gauss’ findings were distribution free and completed the construction of the classical error theory. True, many sections of his contribution of 1823 (and, to a lesser extent, of 1809) made difficult reading.

7.3. The Fate of the First Approach [17; 24]

The first substantiation of the MLSq proved extremely tenacious: it was elegant and much simpler than the second one, and the normal distribution more or less conformed to reality and became rooted in physics (Maxwell, in 1860). In actual fact, astronomers and geodesists, and other natural scientists of the mid-19th century had hardly been acquainted with Gauss' mature thoughts and some were only gradually grasping the essence of the MLSq (Ivory [26], a "scharfsinnigen" mathematician, as Gauss called him, in a series of his papers in 1826 – 1830, although he it was who at once came out in favor of the second justification). Chebyshev (lectures of 1880) had not described the second justification and, moreover, unsuccessfully dwelt on formula (11): he replaced the magnitudes v_i by observational errors. True, Markov's active defence of Gauss' mature thoughts and an appropriate description of this issue in his treatise bore fruit: in Russia, the first justification left the forefront.

7.4. Appendix: Legendre and Gauss [21; 24]

When publishing his first substantiation of the MLSq, Gauss used the expression "our principle" and in 1820, after his letter of 1809 to Gauss had remained unanswered, Legendre reasonably and openly remarked that a discovery was due to him who first published it. Here is the pertinent opinion of two commentators [51, p. 309; 52, p. 18]:

Gauss cared a great deal for priority. [...] But to him this meant being first to discover, not first to publish; and he was satisfied to establish his dates by private records, correspondence, cryptic remarks in publications. [...] Whether he intended it so or not, in this way he maintained the advantage of secrecy without losing his priority in the eyes of later generations.

Was einem normalen Autor verboten ist, einem Gauss wohl gestattet werden muß, zumindest müssen wir seine Gründe respektieren.

I am dwelling on this issue only because Stigler [53], who described the work of Legendre, Gauss and other scholars on a modern level (but omitted the ancient history as well as Kepler, Lambert, Daniel Bernoulli's last memoir and Helmholtz), unfoundedly questioned Gauss' priority (even in the above sense) and his imparting his discovery to colleagues and, for good measure, declared that Euler had not understood the essence of statistics. I [21, §7; 32] resolutely refuted these allegations which Stigler, moreover, expressed in an inadmissible manner, and I only adduce one pertinent example (see his pp. 57 and 146): Legendre "immediately realized the method's potential" but "there is no indication that [Gauss] saw its great potential before he learned of Legendre's work". In 1999 Stigler [54] repeated suchlike astonishing declarations slightly less impudently and, without recalling his previous opinion, highly praised Euler.

8. Further Work [28]

8.1. Newcomb

Arguments, denying the universality of the normal law, had been multiplying and gradually undermined the first justification. **Cournot** [61] was the first to suggest that the density might be represented as

$$\psi(x) = [n_1 \psi_1(x) + n_2 \psi_2(x) + \dots] / (n_1 + n_2 + \dots)$$

where n_i were the numbers of observations in one and the same series possessing densities $\psi_i(x)$. He had not indicated the type of these partial densities nor did he explain the difference between them. All of them were possibly normal laws with differing variances.

Newcomb's investigations of 1882 and 1886 [62] became generally known. He remarked that in a long series of observations errors were seldom normally distributed and argued that their density was a mixture of n such laws with differing measures of precision h_i appearing with probabilities p_i . Newcomb also developed a method of adjusting observations possessing such a density but it occurred that the simplifications that he introduced reduced his integral criterion to the principle of maximum likelihood. And, what was worse, his method remained subjective because the parameters of the mixture, h_i , p_i and n , could not have been determined quantitatively. It should also be noted that the normal law has special importance in the error theory: its existence is necessary and sufficient for the least-squares estimators to be jointly effective. Already **Bienaimé** [64] indicated that such estimators were desirable.

Newcomb actually replaced the constant measure of precision of the normal law by a discrete random variable. In 1887 **Lehmann-Filhès** proposed, instead, to regard h as a continuous random variable with its own normal distribution. Finally, **Ogorodnikov** [62] argued that the errors were distributed with density

$$\psi(x) = (1/\sqrt{\pi}) \int_0^{\infty} h f(h) \exp[-h^2 x^2] dh$$

where $f(h)$ was not given beforehand.

Eddington [63] in an extremely simple way proved that Newcomb's distribution was not normal. Thus, although the normal law itself is stable (which was known to Gauss, Laplace and Bessel), a mixture of such laws is not.

8.2. Helmert [27; 56 - 59]

Helmert continued the work of Gauss. While examining normally, and sometimes uniformly distributed errors, he provided formulas for revealing systematic influences and estimating the precision of observations by means of different functions of residual free terms v_i (§2). One of his problems led him to the chi-squared distribution (already known to Abbe). He also studied the plausibility of the mean square error, again for the normal law, and introduced the *Helmert transformation*, as it is now called in statistics. Finally, he reasonably thought that biased estimators were admissible. The Gauss classical formula (11), in connection with which I mentioned Helmert in §7.2, provided an unbiased estimator of the sample variance; nevertheless, it was the (biased) mean square error that has always been practically referred to in geodesy.

8.3. Lévy [28]

Lévy [39; 65] thought that the estimation of precision was only possible if observational errors possessed a stable distribution. Otherwise, as he argued, the mean square error only provided a general notion about precision. He also recommended how to treat observations in case of stable laws with various values of parameter α ($1 < \alpha < 2$; $\alpha = 1$; $0 < \alpha < 1$). His considerations were reasonable but useless since these values remain unknown; even the realization of a stable law with $\alpha = 2$ (the normal law) is questionable whereas the worst cases ($0 < \alpha \leq 1$) should not take place at all. To recall, stable laws have variance only if $\alpha = 2$, and $\alpha = 1$ defines the Cauchy distribution.

9. The Stochastic Branch of the Error Theory and Mathematical Statistics [28]

From the 18th century, the theory of errors is based on its own theorems and principles. These theorems, however, now belong to the theory of probability, and the principles (of maximum likelihood and least variance) were appropriated by statistics. Nowadays, the stochastic error theory is the application of mathematical statistics to the treatment of observations; unlike statistics, it (as also the determinate error theory) ought to study systematic errors, which, together with some "physical" dependence between observations of one and the same series, lead to the insufficiency of a formal estimation of precision. Gauss, for example, observed the angles of his triangulation until becoming convinced that (in contradiction with his own formulas) further work will be meaningless.

The theory of errors cannot do without the notion of *true value* of the constants sought whereas statisticians consider estimation of parameters of the appropriate distributions. According to Fourier [66], a true value is the limit of the appropriate arithmetic mean as the number of observations increases unboundedly (§2). This definition, that seems to be akin to the Mises formula of probability of an event, has remained either unnoticed or forgotten, but modern metrologists introduced it independently [67, pp. 30 – 31] (and included the residual systematic error as a component of the true value). Here is an example from that source:

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 absorbed upon its surface under standard conditions.

Owing to systematic errors, rejection of outlying observations in accord with subtle tests is not widely applied in the theory of errors. In addition, practitioners cannot be sure that the errors of their observations obey one and the same (still less, one and the same normal) law.

The theory of errors differs from statistics also in that it studies above all independent magnitudes. Thus, as **Gauss** stated in 1823, the angles of a triangle are dependent only insofar as some components of their errors (almost) coincide. Dependence certainly appears after adjustment, when the sum of the angles becomes equal to a stipulated magnitude (180° plus the spheroidal excess of the triangle).

In 1912, being disappointed with the nascent theory of correlation, **Kapteyn** [68] suggested to estimate the dependence between two functions, whose measured arguments partly coincided, by a peculiar correlation coefficient. His paper had been forgotten (or unnoticed in the first place) perhaps because he had not referred to Gauss.

10. The Determinate Error Theory [5; 12]

I sketch now the history of the determinate branch of the error theory and remain within traditional boundaries. It is not amiss, however, to note that the design of experiments and preliminary data exploration, if understood in a wide sense, could have also been included here. I bear in mind, above all, the study of the free fall of bodies (**Galileo**, **Huygens**), the development of pendulum clocks (**Huygens**) and optical investigations (**Newton**).

10.1. Ancient Astronomy and the 18th Century

Ancient astronomers knew that errors corrupted their observations, attempted to minimize their action (§ 2) and could not have failed to recognize the difference between random and systematic influences. All this is all the more true with regard to such scholars as **Tycho Brahe** and **Kepler**, but nevertheless the determinate branch of the error theory properly originated only in the 18th century.

In 1722 **Cotes** offered differential formulas for estimating the errors of calculated elements of plane and spherical triangles given the errors of their measured sides and angles. Without mentioning him, **Lambert**, in 1765, published similar formulas for other geodetic figures and even invented a pertinent but forgotten term, *Theorie der Folgen*. Although his definitions were narrow, **Daniel Bernoulli** (§4.6) was the first to separate formally the errors into the two categories. **Mayer** and **Boscovich** derived differential formulas connecting instrumental errors with the ensuing observational errors [69] and thus provided the possibility of compiling more expedient observational programs. Mayer also invented the repeating theodolite which enabled to lessen essentially the influence of the error of reading and to equate its order with that of the error of pointing. The total error was made less, and its density became closer to the normal law.

10.2. Laplace [15]

Laplace repeatedly described his ideas about observational programs, about the influence of errors on final results, etc. In 1818 – 1819 he examined the precision of a meridian arc determined by triangulation as well as of the so-called trigonometric levelling and thus made a further step in the investigation of geodetic networks. However, the reader acquainted with the work of Gauss will hardly be satisfied with Laplace's unwieldy mathematical tools connected with the CLT and calculation of integrals.

In 1792 Laplace applied the minimax method for checking whether meridian arc measurements contradicted the hypothesis of an ellipsoidal Earth. Kepler (§3.4), as it seems, applied elements of the same method (and for a similar aim) and, much later, Chebyshev [70] used it in the theory of mechanisms when studying the precision of transforming circular motion into rectilinear.

10.3. Gauss and Bessel [17]

Hipparchus, **Tycho Brahe** and **Bradley** are rightfully remembered as outstanding experimentalists, but the modern period in experimental science began with **Gauss** and **Bessel**. They demanded thorough examination of instruments, compensation of instrumental errors by expedient observational programs and corrections for residual instrumental errors. Gauss also invented the heliotrope (the solar mirror), which allowed to observe triangulation angles more precisely, and essentially improved the existing method of measuring the difference between two approximately equal weights [71, p. 427].

Bessel [72] examined the corruption of the length of a measuring bar due to its weight and the ensuing bending. By means of appropriate differential equations he determined the position of the bar's supporting point securing its least bending. Bessel [73] also discovered the existence of the personal equation. As recorded by a practitioner, the moment of the passage of a star through the cross-hairs of an astronomical instrument strongly depends on his psychological habits; accordingly, two astronomers will hardly ever record, even approximately, the same moment. One of Bessel's experiments was however faulty: its errors were of the same order as that of the suspected difference.

10.4. Helmert [27]

Following Laplace, but applying the Gauss approach to the error theory, he [55] studied the configurations of various geodetic figures ensuring maximal weight (least variance) of their calculated elements. Helmert also broke fresh ground by stating his aims as "einen nothwendigen Genauigkeitgrad [of geodetic systems] mit möglichst wenig Zeit und Geld zu erreichen" (p. 1), or as achieving more precise results "bei gleicher Mühe" (p. 60). Not all of his equations were linear (or even algebraic) and he was therefore unable to develop the theory of linear programming. Nevertheless, one of his conclusions complied with this theory: the second aim, as he remarked, demanded that some observations be lacking altogether. Note, however, that this is impossible since all observations should be carried out (even with a lesser weight) for checking the entire work.

References

O.B. Sheynin

1. On the work of Adrain in the theory of errors (1966). Translated in this collection.
2. Origin of the theory of errors. *Nature*, vol. 211, 1966, pp. 1003 – 1004.
3. On selection and adjustment of direct observations (1966). Translated in this collection.
4. On the history of the adjustment of indirect observations (1967). Translated in this collection.
5. Newton and the classical theory of probability. AHES, vol. 7, 1971, pp. 217 – 243.
6. Lambert's work in probability. *Ibidem*, pp. 244 – 256.
7. On the mathematical treatment of observations by Euler. AHES, vol. 9, 1972, pp. 45 – 56.
8. Daniel Bernoulli's work on probability. In *Studies in the history of statistics and probability*, vol. 2. Editors, Sir Maurice Kendall & R.L. Plackett. London, 1977, pp. 105 – 132.
9. Theory of probability. In *История математики с древнейших времен до начала XIX века* (History of Mathematics from the Most Ancient Times to the Beginning of the 19th Century), vol. 3. Editor, A.P. Youshkevich. Moscow, 1972, pp. 126 – 152. Coauthor, L.E. Maistrov.
10. Finite random sums. AHES, vol. 9, 1973, pp. 275 – 305.
11. Boscovich's work on probability. *Ibidem*, pp. 306 – 324.
12. Mathematical treatment of observations. AHES, vol. 11, 1973, pp. 97 – 126.
13. Kepler as a statistician. *Bull. Intern. Stat. Inst.*, vol. 46, 1975, pp. 341 – 354.
14. On the appearance of the Dirac delta-function in a memoir of Laplace (1975). Translated in this collection.
15. Laplace's theory of errors. AHES, vol. 17, 1977, pp. 1 – 61.
16. Theory of probability (1978, in Russian). In *Mathematics of the 19th Century*, vol. 1. Editors, A.N. Kolmogorov, A.P. Youshkevich. Basel, 1992, 2001, pp. 211 – 288. Coauthor, B.V. Gnedenko.
17. Gauss and the theory of errors. AHES, vol. 20, 1979, pp. 21 – 72.
18. On the history of the statistical method in astronomy. AHES, vol. 29, 1984, pp. 151 – 199.
19. Markov's work on probability. AHES, vol. 39, 1989, pp. 337 – 377.
20. Al-Biruni and the mathematical treatment of observations. *Arabic Sciences and Philosophy*, vol. 2, 1992, pp. 299 – 306.
21. On the history of the principle of least squares. AHES, vol. 46, 1993, pp. 39 – 54.
22. The treatment of observations in ancient astronomy. *Ibidem*, pp. 153 – 192.
23. Theory of errors. *Companion Enc. Hist. Phil. Math. Sciences*. Editor, I. Grattan-Guinness. London – New York, vol. 2, 1994, pp. 1315 – 1324.
24. Gauss and geodetic observations. AHES, vol. 46, 1994, pp. 253 – 283.
25. Chebyshev's lectures on the theory of probability. *Ibidem*, pp. 321 – 340.
26. Ivory's treatment of pendulum observations. *Hist. Math.*, vol. 21, 1994, pp. 174 – 184.
27. Helmert's work in the theory of errors. AHES, vol. 49, 1995, pp. 73 – 104.
28. Density curves in the theory of errors. *Ibidem*, pp. 163 – 196.
29. Selection and treatment of observations by Mendeleev. *Hist. Math.*, vol. 23, 1996, pp. 54 – 67.
30. *The History of the Theory of Errors*. Egelsbach, 1996.
31. Theory of probability: definition and relation to statistics. AHES, vol. 52, 1998, pp. 99 – 108.
32. The discovery of the principle of least squares. *Historia Scientiarum*, vol. 8, 1999, pp. 249 – 264.
33. Bessel: some remarks on his work. *Ibidem*, vol. 10, 2000, pp. 77 – 83.
34. Gauss, Bessel, and the adjustment of triangulation. *Ibidem*, vol. 11, 2001, pp. 168 – 175.
35. Newcomb as a statistician. *Ibidem*, vol. 12, 2002, pp. 142 – 167.

Other Authors

36. Kruskal, W.H. Formulas, numbers, words: statistics in prose. In *New Directions for Methodology of Social and Behavioral Sciences*, No. 9, 1981, pp. 93 – 102.
37. Andrews, D.F. Data analysis, exploratory. *Intern. Enc. Statistics*, vol. 1. Editors, W.H. Kruskal, J.M. Tanur. New York – London, 1978, pp. 97 – 107.
38. Poincaré, H. Résumé analytique [of own works]. In *Math. Heritage of Poincaré*. Proc. Symp. Indiana Univ. Providence, 1980, pp. 257 – 357.
39. Lévy, P. *Calcul des probabilités*. Paris, 1925.
40. Neugebauer, O. Mathematical methods in ancient astronomy. In author's *Astronomy and History. Sel. Essays*. New York, 1983, pp. 99 – 127.
41. Berggren, J.L. Ptolemy's map of earth and the heavens: a new interpretation. AHES, vol. 43, 1991, pp. 133 – 144.
42. Price, D.J. Medieval land surveying and topographical maps. *Geogr. J.*, vol. 121, pt. 1, 1955, pp. 1 – 10.

43. Maistrov, L.E. *Probability Theory: Historical Sketch* (1967, in Russian). New York – London, 1974.
44. Hald, A. *History of Probability and Statistics and Their Applications before 1750*. New York, 1990.
45. Kepler, J. *New Astronomy* (1609, in Latin). Cambridge, 1992.
46. Freudenthal, H., Steiner H.-G. Aus der Geschichte der Wahrscheinlichkeitstheorie und der mathematischen Statistik. In *Grundzüge der Mathematik*, Bd. 4. Hrsg, H. Behnke et al. Göttingen, 1966, pp. 149 – 195.
47. Helmert, F.R. Zur Ableitung der Formel von Gauss für den mittleren Beobachtungsfehler etc. *Sitz.-Ber. Kgl. Preuss. Akad. Wiss. Berlin*, 1904, Hlbbd 1, pp. 950 – 964.
48. Kolmogorov, A.N. et al, A formula of Gauss in the method of least squares (1947, in Russian). *Sel. Works*, vol. 2. Dordrecht, 1992, pp. 303 – 308.
49. Fisher, R.A. *Statistical Methods for Research Workers* (1925). Reprint of the 14th edition of 1973 in author's book *Statistical Methods, Experimental Design and Scientific Inference*. Oxford, 1990, pp. 1 – 362.
50. Markov, A.A. The law of large numbers and the method of least squares (1899). *Избранные труды* (Sel. Works). N.p., 1951, pp. 233 – 251. Transl.: DHS 2514, 1998, pp. 157 – 168.
51. May, K.O. Gauss. *Dict. Scient. Biogr.*, vol. 5, 1972, pp. 298 – 315.
52. Biermann, K.-R. Über die Beziehungen zwischen Gauss und Bessel. *Mitt. Gauss – Ges. Göttingen*, Bd. 3, 1966, pp. 7 – 20.
53. Stigler, S.M. *History of Statistics*. Cambridge (Mass.), 1986.
54. Stigler, S.M. *Statistics on the Table*. Cambridge (Mass.) – London, 1999.
55. Helmert, F.R. Studien über rationelle Vermessungen im Gebiete der höhern Geodäsie. *Z. Math. Phys.*, Bd. 13, 1868, pp. 73 – 120, 163 – 186.
56. Helmert, F.R. Über die Formeln für den Durchschnittsfehler. *Astron. Nachr.*, Bd. 85, 1875, pp. 353 – 366.
57. Helmert, F.R. Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichsten Beobachtungsfehlers. *Astron. Nachr.*, Bd. 88, 1876, pp. 113 – 132.
58. Helmert, F.R. Über die Wahrscheinlichkeit der Potenzsummen der Beobachtungsfehler. *Z. Math. Phys.*, Bd. 21, 1876, pp. 192 – 218.
59. Helmert, F.R. Über die Genauigkeit der Kriterien des Zufalls bei Beobachtungsreihen. *Sitz.-Ber. Kgl. Preuss. Akad. Wiss. Berlin*, Hlbbd 1, 1905, pp. 594 – 612.
60. Cournot, A.A. *Exposition de la théorie des chances et des probabilités* (1843). Paris, 1984.
61. Newcomb, S. A generalized theory of the combination of observations. *Amer. J. Math.*, vol. 8, 1886, pp. 343 – 366.
62. Ogorodnikoff, K.F. A method for combining observations by applying the method of least squares etc. *Astron. Zh.* (Moscow), vol. 5, No. 1, 1928, pp. 1 – 21.
63. Eddington, A.S. Notes on the method of least squares. *Proc. Phys. Soc.*, vol. 45, 1933, pp. 271 – 287.
64. Bienaymé, I.J. Sur la probabilité des erreurs d'après la méthode des moindres carrés. *J. Math. Pures et Appl.*, sér. 1, t. 17, 1852, pp. 33 – 78.
65. Lévy, P. La loi de Gauss et les lois exceptionnelles. *Bull. Soc. Math. France*, No. 52, 1924, pp. 49 – 85.
66. Fourier, J.B.J. Sur les résultats moyens (1826). *Oeuvr.*, t. 2. Paris, 1890, pp. 525 – 545.
67. Eisenhart, C. Realistic evaluation of the precision and accuracy of instrument calibration systems. In *Precision Measurement and Calibration*. Nat. Bureau Standards Sp. Publ. 300, vol. 1. Editor H.H. Ku. Washington, 1969, pp. 21 – 47.
68. Kapteyn, J.C. Definition of the correlation coefficient. *Monthly Notices Roy. Astron. Soc.*, vol. 72, 1912, pp. 518 – 525.
69. Proverbio, E. Boscovich's determination of instrumental errors in observation. *AHES*, vol. 38, 1988, pp. 135 – 152.
70. Gusak, A.A. Prehistory and beginning of the development of the theory of approximation of functions. *IMI*, vol. 14, 1961, pp. 289 – 348. (R)
71. Pukelsheim, F. *Optimal Design of Experiments*. New York, 1993.
72. Bessel, F.W. Einfluss der Schwere auf die Figur eines, auf zwei Punkten von gleicher Höhe auflegenden Stabes (1839). *Abhandlungen*, Bd. 3. Leipzig, 1876, pp. 275 – 282.
73. Bessel, F.W. Persönliche Gleichung bei Durchgangsbeobachtungen (1823). *Ibidem*, pp. 300 – 304.