Studies

in the History of Statistics and Probability

Vol. 22

Enlarged reprint of papers by Oscar Sheynin

Berlin

2020

Contents

Introduction by the compiler
I. Markov on treatment of observations, 2006
II. Ivory on treatment of pendulum observations, 1994
III. Treatment of observations by Euler, 1972
IV. Fechner as a statistician, 2004
V. Geometric probability and the Bertrand paradox, 2003
VI. Densities in theory of errors, 1995
VII. Buniakovsky on the theory of probability, 1991
VIII. Gauss and the theory of errors, 1979
oscar.sheynin@gmail.com

Introduction by the compiler

Notation

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

L, M, R = Leningrad, Moscow, in Russian

General comments on some items

I

Markov's Work on the Treatment of Observations

Historia Scientiarum, vol. 16, 2006, pp. 80 – 95

1. Introduction

I (1989) have discussed Andrei Andreevich Markov's (1856 – 1922) work in probability. The treatment of observations was included there but not nearly as desirable. Moreover, since then I published some pertinent materials, Markov's manuscript (1903) and the letters written to him by his former student (Koialovich 1893). Here, I review the relevant sources in detail which was not done before and I also refer to my own translations of Markov's papers. My contribution is critical which does not at all weaken our admiration for Markov's great achievements in probability. That I compiled this paper roughly 150 years after Markov's birth was a coincidence.

The first (mimeographed) edition of Markov's course of lectures in probability theory including the treatment of observations appeared in 1883. Since then, he invariably dwelt on the latter subject in no less than four mimeographed editions of his subsequent lectures on probability and in all four editions of his Russian *Calculus of probability* (1900; 1908, translated into German in 1912; 1913; posthumous, 1924) as well as elsewhere (1899; MS 1903/1990).

Opinions concerning his findings in the method of least squares (MLSq), while having been greatly differing, nowadays seem to be unanimous: Markov had not achieved much. My aim is to touch on old debates and to discuss Markov's achievements in this field.

2. The Year 1899

The title of Markov (1899) shows that he was concerned with two topics and the connection between them.

1) Here are Markov's general considerations about the MLSq, see Markov (1899/1951, p. 246), and Item 2 below.

We estimate the merit of each approximate equality by its weight, and […] *for each of the unknowns we determine such an equality whose weight is maximal*. […] *Only this justification of the MLSq is rational* [since] *it does not obscure* [its] *tentative nature. We do not ascribe* […] *the most probable or the most plausible results to the MLSq and only consider it as a* […] *procedure furnishing approximate values of the unknowns along with a tentative estimate of the results obtained*.

And here is his final opinion (1900/1924, p. 323): the MLSq provides approximate results and estimates their worth. *The approximate equalities* can only be the n initial linear equations (possibly having unequal weights), in m unknowns ($m < n$) from which the m normal equations providing *the approximate values of the unknowns* are formed. Consequently, the first lines of the quotation above are incomprehensible. They can only be applied to the case of one unknown. Here, indeed, is Markov's appropriate statement (p. 326):

If observations provide the possibility of forming several approximate equalities a – *X* ≈ 0 *for some unknown a where X is a number, completely ascertained* [?] *by the results of the observations, we shall choose among these equalities that, whose weight is maximal as the best one for determining a*.

The several values of *X* are likely the means of separate series of observations. It can be difficult to assign properly their weights; if this is nevertheless done, and the weights are not too different, why should we reject all series but one? And if the weights are different, why had the practitioner collected unworthy observations?

It is instructive to mention Newcomb (1872): when adjusting the data contained in several astronomical catalogues, he assigned two different weights, depending on the assumed random and systematic errors, to each of them.

Markov's viewpoint concerning the results provided by the MLSq can also be perceived in a letter of 1893 written by his former student and later eminent scientist (Stokalo 1967, p. 415), Boris Mikhailovich Koialovich, 1867 – 1941 (1893) to him:

You say that, once the data are sound, the results will always be good even without the MLSq. I agree absolutely, but from these good results some might be better than the others depending on how we combine the observations.

Now, why is estimation [of precision] tentative, why are the calculated values of the unknowns neither most probable, nor most plausible (as Gauss called them in his early and mature justifications of the MLSq respectively)? Markov never used the second term, although it conformed to the principle of maximum weight which he entirely approved of, see Item 2 below. At the same time he denied the resulting optimality of the MLSq, but then, does it really demand any justification at all?

In any case, Gauss had not unreservedly trusted his own formulas for estimating precision. He kept measuring the angles of his triangulation until becoming convinced that further work was meaningless; for example, at one station he (1903, pp. 278 – 281) measured one of the angles six times, and another one, 78 times. Since his time, a definitive measure of precision in triangulation is only thought to be obtained after all the field work is done, after considering the closings of all the triangles, after measuring baselines and astronomical azimuths at both ends of the chain and additionally allowing for the ensuing discrepancies. Markov could have thought about the unavoidable and undetected systematic errors (whom Gauss hardly ever forgot), but he had not elaborated.

2) Markov upheld Gauss' definitive justification of the MLSq by the principle of maximum weight [minimal variance] and quoted his celebrated pertinent letter to Bessel of 1839, see its English translation in Plackett (1972). Gauss stated that attaining the minimal value of the [variance], an integral measure of precision, was much more important than achieving maximal, but still infinitely low probability.

Markov (1899/1951, p. 247) added that this substantiation

Provides everything needed for the practice, but it does not furnish the probable errors; I believe [he believes] *that this doubtful* *magnitude is not required; if, however, it should be determined* […], *the well-known expression for the* [normal] *probability* […] *should be assumed for each error separately, independently of whether it was derived from one or from several observations.*

Since normality is assumed, the relation between the probable and the mean square errors becomes known, and we ought to conclude that neither one is required! And in 1852, Bienaymé (Heyde & Seneta 1977, pp. 66 – 71) noted that least variance for each estimator separately was less important than a minimal simultaneous confidence interval for all of them.

Later Markov (1916/1951, p. 535) stated, again without explaining himself, that in correlation theory, and even in the theory of errors, he did not

Attach any great importance to the so-called probable errors and only consider[ed] *them as a means for tentatively comparing the merit of different observations.*

Following contemporary astronomers, Markov had not mentioned the mean square error although Gauss (not naming it by that term) had obviously (and reasonably) considered it as the main measure of precision. Then, the probable error can be calculated provided we know the appropriate density, which need not be normal.

3) Markov (1899/1951, p. 248) continued: normality was usually justified either by its alleged conformity with practice (which is *difficult to ascertain*) or by referring to the [central limit] theorem. This consideration assumed observations whose errors were made up of many errors [of the same order and] *independent of one another*, an idea that s*hould be attributed* […] *to the realm of fancy*, and that the number of observations was infinite (which was patently wrong, as Markov reasonably added).

4) In concluding, Markov (1899/1951, pp. 249 – 250) dealt with attempts at substantiating the MLSq by issuing from Chebyshev's findings.

a) Denote the true value of the unknown sought by a and the arithmetic mean of its n observations by $\bar{x}$. Then, as Markov stated, Maievsky's remark (1881, §§ 3l – 32) that

$$\lim P[|\bar{x} - a| < \varepsilon]=1 \text{ as } n \to\infty. \qquad (2.1)$$

[In modem terminology: $\bar{x}$ was a consistent estimator of a]. However, as Markov stated, it did not mean anything since Maievsky had not compared $\bar{x}$ with other possible means, see also below in Item (b).

Markov called (2.1) the *well-known Gauss assumption* [of 1809]. Actually, however, Gauss only considered a finite number of observations, and, in addition, only for deriving the principle of least squares, so he also had to apply the principle of maximum likelihood.

The first such attempt at justifying the arithmetic mean was made by Usov (1867) whom Markov had not mentioned. Note also that (2.1) was a corollary of Chebyshev's more general proposition and that he only formulated it directly in his lectures (Chebyshev 1936), not later than in 1879/1880, when these were written down by Liapunov.

b) Finally, Markov cited some Russian mathematicians who had striven to derive the MLSq by means of a wrong application of the Chebyshev's theorems. He then mentioned the [Bienaymé –] Chebyshev inequality and continued:

They demanded that the probability should exceed a number assigned beforehand, and sought the narrowest boundaries of the error corresponding to a given inequality for the probability. They forgot that an uncountable set of different numbers obeys the same inequality and that the probability can exceed some given number even if the Chebyshev inequality does not reveal that fact.

The Bienaymé – Chebyshev inequality has to do with a finite number of trials or observations, and in this respect the criticized proposal was better than the former. However, here again, in addition to what Markov stated, the principle of maximum likelihood was still required.

The most interesting point in Markov (1899), as far as it concerned the MLSq, was his resolute defence of the principle of maximum weight. True, several authors [Ivory (1825, p. 7), Galloway (1839, pp. 205 and 210) and Ellis (1844)] had preceded him, but Markov's stature undoubtedly made his point of view generally known (and even led to an overestimation of his pertinent merit, see § 5.1). But, to repeat, his defence was hollow.

3. The Year 1903

I bear in mind Markov's review written in 1903 but only published in 1990, also see Sheynin (1990b). In 1902, Prince Boris Galitzin (as he spelled his name in German, the language of his paper, whereas in Russian his name was Golitzin), the future President of the International Seismological Association and Fellow of the Royal Society, published a careless treatment of experiments concerning the solidity of glass tubes.

Markov refuted his findings by scrupulously checking Galitzin's paper. It is hardly expedient to describe Markov's particular conclusions, suffice it to state that he remained within the boundaries of the classical theory of errors and that he even noticed that Galitzin's graph, placed on a plate inserted several pages beyond the paper itself, did not agree with his tabular values.

At the time, Galitzin was nominated for effective membership at the Petersburg Academy of Sciences, and Markov (followed by Liapunov, who criticized some of Galitzin's investigations in mechanics) opposed several academicians, including the astronomer Bredikhin, who continued to defend the nomination. It was during the ensuing debates that Markov declared that he liked the rule stated by Bredikhin according to which the reality of a computed magnitude may be admitted if it at least twice numerically exceeded its probable error; a full quotation is in Sheynin (1989, p, 351). A related rule had been apparently generally applied in natural science: both Mendeleev (1860, p. 46) and Newcomb (1897, p. 165) thought that a discrepancy between two empirically derived magnitudes was essential if it exceeded the sum of the appropriate probable errors.

For the normal distribution the Bredikhin rule, as Markov called it, apparently meant that, since

$$P(|x| < 1.33\sigma) = 0.816,$$

the probability that the calculated magnitude x was situated in the interval [– 1.33σ; 1.33σ], was sufficiently high, the probability of the contrary event was sufficiently low and indicated that x was an *outlier* and did not belong to the vicinity of zero. The essence of the once popular three-sigma rule (but not its conclusion) is similar: if the deviation of an observation from the mean is greater than 3σ, it is significant, and the outlier should be rejected.

4. The Year 1924: The Method of Least Squares

During the academic year 1920/1921 Markov

Intensively worked on the fourth edition (*published posthumously* [in 1924]) of his

<u>*Calculus of Probability*</u> *which is known to differ considerably from its previous edition*.

That was the testimony of Markov Jr (1951, pp. 612 – 613), a mathematician in his own right. He did not, however, add whether his father had completed this work. I shall now describe Chapter 7 of that source called *Method of Least Squares* (pp. 323 – 473). In the previous edition of 1913 the appropriate chapter had not included the investigation of statistical series (see § 4.3) or the tiny issue of linearization (which allows, at least in geodetic applications of the MLSq, to restrict the attention to systems of linear equations, see pp. 469 – 472) and the subject of § 4.7 was then discussed even in lesser detail than in 1924.

4.1. Preliminary Considerations (pp. 323 – 326). Markov (p. 323n) stated that he adhered to his previous viewpoint on the justification of the MLSq as described in 1899. The sequel confirmed this statement. He (p. 323) assumed that the unknown numbers, whose approximate values were provided by observations, existed. In plain words: he assumed the existence of the real (true) values of the measured magnitudes. No one else had formulated this subtle remark.

Fourier's definition (1826, p. 534) according to which the *véritable objet de la recherche* was the limit of the mean of the n pertinent observations as $n \to \infty$ had been forgotten. Later, many authors including Mises independently from him and from each other repeated it (Sheynin 2007).

Following Chebyshev (1879 – 1880/1936, p. 227), Markov (pp. 323 and 373) considered each observation as a particular case of many possible observations. In the first case he specified, although not definitely enough, that one possible observation corresponded to each actually made which was an unnecessary and damaging restriction.

I also recall the statement concerning the choice of maximum weight (§ 3). But the main point is that Markov kept to his previous idea (1911/1981, pp. 149 – 150):

I shall not defend these basic theorems connected to the basic notions of […] *equal probability, of independence of events, and so* *on, since I know that one can argue endlessly on the basic principles even of a precise science such as geometry.*

4.2 The Case of One Unknown (pp. 327—344). **a)** Possible observations (p. 327). Markov assumed that n observations of an unknown constant were made and provided values a_i, i = l, 2, … n, and stated that he will also consider respective possible observations u_i. It is difficult to understand why, contrary to what Chebyshev had assumed (also see § 4.1c), here and on p. 374 only one possible observation corresponded to each (actual) a_i. Furthermore, Koialovich's letter to him dated 1893 implies that much earlier, in 1891, in his mimeographed course of lectures on the theory of probability (unavailable outside Russia), Markov kept to the alternative pattern. This is what Koialovich wrote:

As far as I understand you, you consider each separate observation as a value of a possible result. Thus, a series of results […] *is possible for each measurement, and one of them is realized. I am prepared to understand all this concerning one observation. However, if there are, for example, two observations, then I cannot understand the difference between the series of all the possible results of the first observation* […] *and the similar series for the second measurement.* […]. *The problem will certainly be solved at once if you say that the probabilities of the same error in these two series are different, but you will hardly want to introduce the notion of probability of error in your exposition*.

We would now rather consider a set of observations as a random sample from a general totality. With regard to Koialovich' last phrase, I note (believing, however, that there is only one totality) that Markov (1900/1924, p. 251) had indeed introduced the notion of density, but did not in essence apply it in his chapter on the MLSq; it only appeared there on p. 420 in a discussion of correlation. True, he also applied the normal distribution (Item **d** below).

b) The adjustment proper. Markov assumed that the unknown magnitude a was a linear function of the independent observations a_i, free from constant error [unbiased],

$$a \approx \lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n, \qquad (4.1)$$

$$\lambda_1 + \lambda_2 + \ldots + \lambda_n = 1. \qquad (4.2)$$

Here and below the sign of approximate equality [not $\approx$ but $\neq$ (!)] was Markov's manner of distinguishing between a constant and its estimator.

He (p. 329) noted that other means (for example, the geometric mean) did not ensure [unbiasedness]; and he added elsewhere (p. 447) that the MLSq was only dealing with linear functions because the unknowns were small corrections to approximately known magnitudes, cf. § 4.7. Here and throughout Markov invariably discussed observations of unequal weight but he never remarked that it was possible to abandon this assumption without loss of generality.

And so, the weights p_i were such (p. 327) that

$$\mathrm{E}(a - u_i)^2 = k/p_i \qquad (4.3)$$

where k was the mean square error of unit weight, as it is called in the classical error theory. He abbreviated the Russian term *math. ozhidanie* (mathematical expectation) as m. o. Then (pp. 329 – 330)

$$\mathrm{E}(a - [\lambda u])^2 = [\lambda\lambda/p] = P$$

where I introduced the Gaussian notation of the type $[abc] = \Sigma a_i b_i c_i$ and P is the weight of the equality (4.1).

It is methodically better to say now that,

$$\text{if } \mathrm{var}a_i = \sigma_i^2, \text{ then } \mathrm{var}[a\lambda] = [\lambda\lambda\sigma\sigma] = [\lambda\lambda/p].$$

The condition of maximum weight P together with the restriction (4.2) leads (p. 332) to

$$a \approx a_0 = [pa]/\Sigma p_i = [pa]/P$$

where a_0 naturally satisfies the least-squares condition

$$\Sigma\sqrt{p_i}(a_i - a_0)^2 = \min.$$

c) Determination of k. Markov (pp. 338 – 339) determined the expectation of k, and thus its approximate value

$$k = \frac{\sum p_i(a_i - a_0)^2}{n-1}. \qquad (4.4)$$

The derivation is not difficult but it cannot be generalized to the case of several unknowns which Markov discussed later. Note that Gauss saw fit only to consider the general case.

d) The normal distribution (pp. 341 – 344). Suppose that

$$\Delta = a - a_i \text{ or } \Delta = a - a_0, \qquad (4.5a, 4.5b)$$

then the approximate value of $\mathrm{E}\Delta^2 = h$ is (4.4). Markov did not say why but provided an approximate value; in the sequel, without explanation, he stated that h was exactly equal to the second moment of the appropriate random variable (a term that he never used). Worse, Markov did not state that (4.5a) and (4.5b) differed from each other in that the respective values of h could not have coincided.

Markov supposed that Δ (which one?) was normally distributed with variance h, distributed according to the law $N(0, \sqrt{h})$, and he (pp. 342 – 344) mentioned two possible appropriate justifications: first, Δ "is considered" as a sum of many independent errors; second, normality agreed with practice (cf. his early pronouncement to the contrary in § 2.3). He did not specify that the partial errors should be of one and the same order (indirectly mentioned in the same context by Laplace (1818, p. 536) who remarked that the normal distribution resulted from the use of repeating theodolites which ensured that the effect of the two main errors of angle measurement were of the same order). Nor did Markov suggest any quantitative test for checking that agreement; n ote that on p. 349 he referred to Pearson's paper which introduced the chi-squared test, cf., however, § 4.3b.

4.3. Statistical Series (pp. 345 – 373). Markov inserted an investigation of statistical series (and called this section *Determination of probabilities by observation*).

a) Bernoulli trials: calculation of the constant probability (α) of success (pp. 345 – 349). After *s* trials the number of successes is σ. The result of each trial is a random variable x_i with a binomial distribution (he used neither term) and possible values 1 and 0 (an indicator variable), and

$$\mathrm{E}(x_i-\alpha)^2 = k_1 \approx \alpha(1-\alpha) = \frac{\sigma(s-\sigma)}{s(s-1)}. \qquad (4.6)$$

The last equality follows from (4.4). On the other hand, since $\alpha \approx \sigma/s$,

$$k_2 \approx \frac{\sigma(s-\sigma)}{s^2}. \qquad (4.7)$$

an expression *not completely free from a constant error*.

It might be thought (no such statement is made) that a constant α would have resulted when $k_1 \approx k_2$. However, one of Markov's remarks (p. 345) was questionable (and unnecessary). He stated that the results of the *s* trials meant that σ times $\alpha \approx 1$, and $(s-\sigma)$ times $\alpha \approx 0$. How then should we understand the constancy of α?

Markov next generalized his account to several series of trials. This time the two expressions for *k*, ensuring a possibility of checking to *a certain extent* the assumed independence of trials and constancy of the probability α (here, Markov did make such a statement), were

$$k_l \approx [1/(n-l)]\Sigma s_i([(\sigma_i/s_i] - \overline{p})^2 \text{ and } k_2 \approx \overline{p}(1-\overline{p}). \qquad (4.8;\ 4.9)$$

Here, *n* was the number of series and $\overline{p} = \Sigma\sigma_i/\Sigma s_i$, the statistical probability of success.

b) Comparison of theoretical and statistical probabilities (pp. 349 – 353). Markov considered Weldon's experiment with throws of 12 dice (Pearson 1900). In $N_0 = 185$ throws neither a 5 nor a 6 has appeared; in $N_1 = 1149$ throws one of these results appeared once; …; and in $N_{11} = 4$ throws, the score was 11. The statistical probability of a 5 or a 6 calculated after 26,306 throws was $p_{\text{stat}} = 0.33770$ with $p_{\text{stat}} - p = 0.00436$. Markov applied the central limit theorem (rather than the Pearsonian chi-squared test) to prove that the actual probability of success was higher than 1/3. He confirmed his finding by the Bayes theorem with a transition to the normal distribution and stated that it was indeed highly probable that $\Delta p \geq 0.00436$. Feller (1950, § 2 of Chapter 6) and Hald (1998, p. 201) also discussed the same experiment as treated by Pearson and Fisher (who had remarked that the discrepancy between theory and experiment could have been caused both by variation of the probability and dependence between the trials).

Finally, Markov treated the same experiment as a totality of 12 series assuming $s_i = 12N_i$ and $\sigma = iN_i$. He obtained $k_1/k_2 = 1.0049$ and, since the weight of p_{stat} was $P = 12 \cdot 26{,}306 = 315{,}672$, formula (4.4) led to

$$E\Delta(p_{\text{stat}})^2 = k/P = 0.71 \cdot 10^{-6}.$$

Markov concluded that the probable error of p_{stat} (here and on p. 418 he tacitly assumed the appropriate normal distribution) was 0.00056 so that with probability 1/2 the probability of success deviated less than by that amount from p_{stat}.

c) The coefficient of dispersion (pp. 353 – 373). Markov (p. 353) called

$$L^2 = k_1/k_2$$

the coefficient of dispersion, noted that L is usually applied instead and (pp. 355 – 356) said a few words about its introduction and investigation by Lexis, Bortkiewicz and Dormoy (who studied another variety of that coefficient). It is not amiss to mention that Bortkiewicz (1930) later compared the merits of Lexis and Dormoy and resolutely decided in the former's favour. I (2008) noted that Bortkiewicz, in his alleged discovery of a law of small numbers, had tacitly introduced a different coefficient.

Markov (pp. 356 – 360) then repeated his earlier difficult proof (1916) that the expected value of L^2 was unity, and, this time without repeating the proof, just wrote out his finding made in the same contribution concerning $E(L^2 - 1)^2$. In concluding, Markov calculated EL^2 for two special cases of a variable probability of success.

On the one hand, as I stated just above, Markov's treatment of statistical series seems too difficult for an educational aid; on the other hand, Markov did not mention Chuprov's relevant papers (1916; 1918 – 1919) the first of which he himself had communicated to the *Izvestia* of the Petersburg Academy of Sciences.

4.4 The case of several unknowns (pp. 373 – 397). Markov considered the determination of m unknowns $a_1, a_2, \ldots, a_m$ given n independent observations ($n > m$) $b_1, b_2, \ldots, b_n$ free from constant error and providing the appropriate linear equations. His general considerations were of course known; as in § 4.2a, unusual was his introduction of possible results of observations.

Markov introduced the principle of maximum weight, obtained the normal equations and calculated the weights of the [estimators of the] unknowns as well as the Gauss formula for the mean square error of unit weight, i. e., formula (4.4) with an appropriately changed numerator and denominator equal to $(n - m)$.

4.5 Interpolation (pp. 398 – 403 and 427 – 446). Given, the values of y_i, $i = 1, 2, \ldots, n$ of an unknown linear function $y(x)$ at points x_i, tacitly assumed exactly known. It is required to derive the value of y at any arbitrary point x in accordance with the principle of maximum weight (pp. 398 – 403).

Introducing $\bar{x}$, the arithmetic mean of the observations, and expressing y as

$$y = a_1 + a_2(x - \bar{x}),$$

where the two coefficients were yet unknown, Markov supposed that

$$\Sigma\lambda_i y_i = y, \qquad (4.10a)$$

which ensured lack of constant error. It followed that

$$\Sigma\lambda_i x_i = x. \qquad (4.10b)$$

Then, assuming that

$$\lambda_i = \mu_1 + \mu_2(x_i - \bar{x}), \qquad (4.11)$$

Markov calculated the unknowns λ_i, μ_1 and μ_2 and stated that the λ_i's thus found also ensured maximal weight. Now, constant error was indeed excluded but not in the same way as before: restriction (4.2) was necessary because each observed a_i was approximately equal to the only unknown (a) whereas here the y_i's can considerably differ one from another.

As to conditions (4.11), they are not readily seen to lead to maximal weight. The problem of interpolation can be directly and easily treated by the MLSq without introducing any multipliers such as the λ_i's (e. g., Idelson 1947, § 17) and Markov's treatment of this subject was unnecessarily difficult.

Markov (pp. 427 – 446) also discussed polynomial interpolation. He followed Chebyshev and provided worked-out examples (sometimes criticizing his teacher's calculations) but he never mentioned Weierstrass. In my context, the only important point is that this time Markov chose the direct approach.

4.6. Correlation (pp. 403 – 427). Markov considered linear correlation and applied the MLSq for determining the parameters of the lines of regression. He also discussed the case of random variables with density of their distribution of a quadratic form, with

$$f = a_{11}x_1 + 2\,a_{12}x_1x_2 + a_{22}x_2$$

in case of two variables, and even with densities of the type $g(f)$ where the function g was only restricted by general analytic and stochastic requirements. He had not referred to one of his previous chapters where, on pp. 275 – 287 he studied *connected* variables with density e^{-f}. At the end of that chapter he included a reference to Slutsky's book (1912) on correlation. He certainly did not repeat his earlier harsh words (1916/1951, p. 533) about correlation, which, when indicating the precision of

Various coefficients, enters … the realm of imagination, hypnotism and belief in mathematical formulas that actually have no scientific foundation. Even then, in 1916, this was probably wrong (Hald 1998, § 27.7).

4.7 The Generalized Case of Several Unknowns (pp. 446 – 469). Suppose that, in addition to the *n* equations, there are several more equations connecting the unknowns and such which ought to be strictly complied with; a simplest example is that the sum of the (measured and therefore corrupted by error) angles of a plane triangle should be exactly equal to 180°.

That case (Gauss 1828) had not been in much use although Bessel introduced a method for dealing with it. Second, Markov left out the important and usual method of the adjustment of triangulation, the so-called method of conditional observations.

Markov concluded by working out an adjustment of the angles of a plane triangle in the general case: the angles a_i, $i = 1, 2, 3$, were measured n_i times with the weights of each measurement being p_i. Such generality was hardly needed: a practitioner would have most likely combined n_i and p_i into one single parameter.

5. Discussion

5.1. Justification of the Method of Least Squares. Neyman (1934, p. 595) mistakenly attributed to Markov the definitive Gauss' justification of the MLSq by the principle of maximal weight (coupled with lack of bias). Then, F. N. David & Neyman (1938) repeated and even aggravated the situation by proving *an extension of the Markov theorem* actually due to Gauss. Neyman (1938/1952, p. 228), however, later acknowledged his mistake, *the confusion*, to which he

Unwittingly contributed by attributing to Markov the basic theorem of least squares.

And Plackett (1949, p. 460) concluded that Markov

May perhaps have clarified assumptions … [made when justifying the MLSq] *but proved nothing new.*

It is instructive to note that Kolmogorov (1946) had not mentioned Markov in his relevant paper. True, he did not at all name anyone after Gauss, but he would have possibly made an exception for his countryman Markov had that been expedient.

Meanwhile, as noted by H. A. David (2001, p. 218), Schéffe (1959, p. 14) had put into scientific circulation the term *Gauss – Markov theorem* which Seneta (1997, p. 265) correctly called a misnomer and added that there was *little originality* in Markov's treatment of the MLSq. Nevertheless, I repeat (see end of § 2) that Markov resolutely defended the principle of maximal weight. This was important because gross mistakes die hard! Even Fisher (1925, p. 260) believed that the MLSq was

A special application of the method of maximum likelihood, from which it may be derived.

5.2 Adjoining Topics. The justification of the MLSq as understood above means that unbiased and effective statistics ought to be chosen, and Linnik et al (1951, p. 637) declared that Markov had actually introduced such notions. They could have just as well stated the same with respect to Gauss. Cf. Neyman (1934, p. 593): the importance of

Markov's pertinent work consisted *chiefly in a clear statement of the problem*.

To a certain extent this conclusion negates Neyman's opinion about the justification of the MLSq (see § 5.1).

At least two authors maintained that Markov had *completed the problems* of the MLSq (Besikovich 1924, p. vii), or, in other words, *brought forward* the Gauss method *to a highest logical and mathematical perfection* (Idelson 1947, p. 14). This was of course wrong. Incidentally, Besikovich (1924, p. xiv) just as wrongly attributed to Markov a *new development of correlation*.

5.3. Methodological Issues. There exist conflicting opinions regarding the methodological value of Markov's work. Thus, Bernstein (1945/1964, p. 425): Markov's treatise and memoirs were

Specimens of preciseness and lucidity of exposition

(I strongly object to the lucidity); and Linnik et al (1951, p. 615): Markov's language is distinct and clear, and he thoroughly trims the details. Again, I disagree. A striking example proving the opposite is Markov's failure to discuss the adjustment of direct conditional observations (§ 4.7). Witness Bauschinger (1900 – 1904, p. 794):

Dieser Fall kommt in der Praxis besonders häufig vor und soll daher besprochen werden, obwohl er ein Spezialfall des vorigen [discussed by Markov] *ist*.

[This case appears especially often and ought therefore to be discussed in spite of its being a particular instance of the previous.]

And I do not trust Chuprov (1925/1981, p. 154) who thought that Markov's treatise was

A handbook of the theory of probability for statisticians.

I do not agree; for one thing, Markov did not stress that the Bernoulli LLN had done away with the need to have equally possible cases in statistics, a point that statisticians somehow did not comprehend for many decades. Chuprov also stated that the exposition in Markov's treatise was *transparently clear*, but he reasonably objected to Markov's discussion of correlation in the framework of the MLSq and essentially criticized it.

With regard to the MLSq Markov himself (Ondar 1977/1981, Letter 15 to Chuprov dated 1910) owned that he had *often heard that my* [his] *presentation is* [was] *not sufficiently clear*; recall, indeed, Koialovich' doubts in § 4.2a. And Idelson (1947, p. 101) remarked that the pertinent chapter was ponderously written. Some elaboration is in order.

Markov barely numbered his formulas; instead, he rewrote them. Thus, on pp. 328 – 330 the equality (4.2) appeared five times! A related point is his apparent disregard of demonstrative pronouns, for example (p. 328; similar cases on pp. 379 and 381 – 382):

The choice of coefficients [a displayed line of these coefficients follows] *is at our disposal. We shall subject the coefficients* [the same displayed line is repeated] *to two conditions*.

Markov refused to use the Gauss brackets (§ 4.2b) and he only introduced notation for the arithmetic mean on p. 463. The terms *normal distribution* and even *coefficient of correlation* were lacking in his works; he never said *random error*, let alone *random magnitude* (as it came to be called in Russia since 1885), see Sheynin (1989, p. 350n). Wherever possible, Markov excluded *the completely undefined expressions* <u>*random*</u> *and* <u>*at random*</u> (Ondar 1977/1981, Letter 53 to Chuprov of 1912). True, in his Chapter 5, especially when discussing geometric probability, he allowed himself to describe a uniformly distributed random variable by the second expression. Sometimes Markov used the expression *indefinite* (rather than random) which was simply bad. Much better was the attitude of Vasiliev (1885, pp. 127 – 131) who was one of the first in Russia to pick up the new term, and to add, on p. 133, that *random errors have all the properties of random magnitudes* (and *their own special properties*). His addition was careless.

Markov's references were not specific enough. On p. 427, when discussing polynomial interpolation, he only cited Chebyshev's *Oeuvres*, tt. 1 and 2, and, for example on p. 163 he provided three references without dates. On p. 10 (unconnected with the MLSq) Markov formulated an axiom (which, as it seems, never outlived him, see § 5.5) without duly isolating it from the context and referred to it on p. 24. His literary style was ponderous and sometimes barely understandable (Markov 1907/1951, p. 341), and, from one edition to another, the structure of his treatise became ever more complicated.

My main point is, however, that the chapter on the MLSq was hardly inviting either for mathematicians or geodesists. The former would have been disappointed by an almost lacking discussion of Pearson's work whereas the latter, in addition, had not needed interpolation or investigation of statistical series although they would have wished to see much more about correlation. And the absence of the Gauss brackets, as well as the appearance of the long-ago dated term *practical geometry* (p. 462) would have annoyed them.

5.4. Attitude towards Work of Other Authors. My § 5.3 is partly relevant here. Chuprov (1925/1981, pp. 154 and 155) noted that Markov had left out the works of other authors not belonging to "the stream" of his own contribution; in particular, even Chuprov's "belonging" papers, see § 4.3c. I myself note that he had not referred to several foreign scientists (Bohlmann, Student, Yule, Fisher) or to his compatriot Bernstein. Neither had Markov mentioned Mises or Lindeberg, possibly because he did not properly know their work owing to the situation in Russia. Indeed, actually, none of his references went beyond 1914.

Markov barely mentioned Pearson's chi-squared test. Even worse: his considerations (see § 4.3b) left an impression that it was not needed at all whereas in actual fact that test, unlike the classical stochastic reasoning applied by Markov, was suitable for a small number of trials as well. Again, Markov hardly discussed correlation (§ 4.6). And, perhaps owing to their insufficient substantiation, he passed over in silence the Pearsonian curves. However, he (1924; reprinted Introduction to the previous edition) held that the use of approximate methods in applied mathematics was unavoidable even when an estimation of their error was impossible; and in 1915 he expressly stated that Pearson's "empirical" formulas did not demand theoretical proof (Sheynin 1989, p. 345). I think that Markov followed here (as he did in the most important example of his chains left without natural-scientific applications) his own rigid principle hardly worthy of exact imitation (Ondar 1977/1981, Letter 44 to Chuprov of 1910):

I shall not go a step out of that region where my competence is beyond any doubt.

On the other hand, as compared with 1916 (§ 4.6), he made some progress towards recognizing correlation, and he tacitly overcame his disbelief in normality of errors (compare §§ 2.3 and 4.2d). And at least he became interested in Slutsky's book (Sheynin 1990a/2011, p. 63).

5.5 A Remark concerning the Theory of Probability. I am unable to pass over in silence Markov's general considerations regarding probability theory. He (p. 4) introduced the classical definition of probability of an event, but added on p. 2 that notions were mainly determined not by words, but rather by our attitude towards *them*. Yes, but why not add: that *definition* is not a definition at all? Then, on p. 10, Markov formulated the following axiom (see also § 5.3): If there are several equally possible events, some of them favourable, the others not, with regard to event A, then, after A occurs, the unfavourable events "fall through" whereas the others remain equally possible. I do not see how it can be otherwise.

On pp. 13 – 19 he proved the addition and multiplication theorems in an excessively complicated way. Thus, in the latter proposition for two dependent events Markov referred to his axiom (which was not isolated from the context) in connection with the occurrence of the first event. Another example of what I would call an excessive desire for rigor, persisting in spite of the then shaky foundation of probability theory, is on p. 1. Maintaining that "we" can only answer a particular question in a certain way, he added:

The word <u>we</u> is current in mathematics and does not impart any special subjectivity to the theory of probability.

And here, finally, is Markov's astonishing conclusion (p. 24):

The addition and multiplication theorems along with the axiom mentioned above serve as an unshakeable base for the calculus of probability as a chapter of pure mathematics.

True, on p. 241 Markov formulated what could now be called the extended axiom of addition (and multiplication!), again, as in the case of his first axiom, without duly isolating it from the context.

As noted by Cramér (1976, § 2), the first systematic exposition of the theory of random variables, and of their distributions and characteristic functions was due to Lévy (1925).

6. Conclusion

Markov's main merit in the field under discussion is his staunch (although hollow!) support of the definitive justification of the MLSq (§ 2). Being a graduate (in 1951) of the Moscow Geodetic Institute, I remember that Russian geodesists only recognized this substantiation and that in general Gauss rather than Laplace was our demi-god. Even during one of the darkest periods of Soviet life, when foreign science had been all but denied, when penicillin was declared a Russian invention, Gauss remained supreme, and later, in 1957 – 1958, two volumes of his *Selected Geodetic Works* appeared in Russian translation. The first of them contained the same contributions on the treatment of observations as the celebrated pertinent German edition of 1887. I believe that without Markov the situation in Russia would have been considerably different.

References

A. A. Markov

(1899), The law of large numbers and the method of least squares. In Markov (1951, pp. 231 – 251). **S, G.** 5.

(1900), *Ischislenie Veroiatnostei* (Calculus of Probability). Petersburg. Later editions: 1908, 1913, 1924 (posthumous). German translation: 1912.

(MS, 1903), On the solidity of glass. Published in Sheynin (1990b, pp. 451 – 467). **S, G**. 1. See also **S, G,** 85.

(1907), The extension of the law of large numbers, etc. In Markov (1951, pp. 341 – 361). **S, G,** 5.

(1911), On the basic principles of the calculus of probability and on the law of large numbers. In Ondar (1977/1981, pp. 149 – 153).

(1916), On the coefficient of dispersion. In Markov (1951, pp. 525 – 535). **S, G,** 5.

(l951), *Izbrannye Trudy* (Sel. Works). N. p.

Other Authors

Bauschinger J. (1900 – 1904), Ausgleichungsrechnung. *Enz. math. Wissenschaften,* Bd. 1, Tl. 2, pp. 769 – 798.

Bernstein S. N. (1945), Chebyshev's work in the theory of probability. *Sobranie Sochinenii* (Coll. Works), vol. 4. N. p., 1964, pp. 409 – 433. **S, G,** 5.

Besikovich A. S. (1924), Biography of Markov. In Markov (1924, pp. iii – xiv).

Bortkiewicz L. (1930), Lexis und Dormoy. *Nord. Stat. J.*, vol. 2, pp. 37 – 54.

Chebyshev P. L. (1936), *Teoria Veroiatnostei* (Theory of Probability). Lectures of 1879/1880 as written down by A. M. Liapunov. Moscow – Leningrad. **S, G,** 3. Great number of math. misprints.

Chuprov A. A. (1916), On the expectation of the coefficient of dispersion. *Izvestia Imp. Akad. Nauk*, vol. 10, pp. 1789 – 1798. **S, G,** 35.

--- (1918 – 1919), Zur Theorie der Stabilität statistischer Reihen. *Skand. Aktuarietidskr.*, t. l , pp. 199 – 256; t. 2, pp. 80 – 133.

--- (1925), Review of Markov (1924). In Ondar (1977/1981, pp. 154 – 157).

Cramér H. (1976), Half a century with probability theory. *Annals Prob.*, vol. 4, pp. 509 – 546.

David F. N., Neyman J. (1938), Extension of the Markoff theorem on least squares. *Stat. Res. Memoirs*, vol. 2. London, pp. 105 – 117.

David H. A. (2001), First (?) occurrences of common terms in statistics and probability. In David H. A., Edwards, H. W. F. (2001), *Annotated Readings in the History of Statistics*. New York, pp. 209 – 246.

Ellis R. L. (1844), On the method of least squares. In author's book (1863), *Mathematical and Other Writings*. Cambridge, pp. 12 – 37.

Feller W. (1950), *Introduction to Probability Theory and Its Applications*, vol. 1. New York, 1957.

Fisher R. A. (1925), *Statistical Methods for Research Workers.* London. In author's *Stat. Methods, Experimental Design and Scientific Inference*. Oxford, 1990, separate paging.

Fourier, J. B. J. (1826), Sur les resultats moyennes. *Oeuvr.*, t. 2. Paris, 1890, pp. 525 – 545.

Galloway T. (1839), *Treatise on Probability*. Edinburgh.

Gauss C. F. (1823 and 1828, in Latin), Theorie der den kleinsten Fehlern unterworfenen Combination der Beobachtungen. In author's book (1887), *Abhandlungen zur Methode der kleinste Quadrate*. Hrsg. A. Börsch, P. Simon. Vaduz, 1998, pp. 54 – 91. English translation in G. W. Stewart, *Theory of Combination* … Philadelphia, 1995.

--- (1903), *Werke*. Bd. 9. Göttingen – Leipzig.

Hald A. (1998). *History of Mathematical Statistics from 1750 to 1930*. New York.

Heyde C. C., Seneta E. (1977), *Bienaymé*. New York.

Idelson N. I. (1947), *Sposob Naimenshikh Kvadratov* etc. (Method of Least Squares). Moscow. **S, G,** 58 (Chapter 1).
Ivory J. (1825), On the method of least squares. *Lond., Edinb. and Dublin Phil. Mag.*, vol. 65, pp. 1 – 10, 81 – 88, 161 –168.
Koialovich B. M. (MS, 1893), Letters to Markov. **S, G,** 5.
Kolmogorov A. N. (1946), On the justification of the method of least squares. *Uspekhi Matematich. Nauk,* vol. 1, No. 1, pp. 57 – 71. Translated in *Sel. Works,* vol. 2. Dordrecht, 1992, pp. 285 – 302.
Laplace P. S. (1818), Deuxième Supplément to *Théor. anal. des prob. Oeuvr. Compl*, t. 7, No. 2. Paris, 1886, pp. 531 – 580.
Lévy P. (1925), *Calcul des probabilités*. Paris.
Linnik Yu. V. et al (1951), A sketch of the work of Markov on the theory of numbers and the theory of probability. In Markov (1951, pp. 614 – 640). **S, G,** 5.
Maievsky N. (1881), *Izlozhenie Sposoba Naimenshikh Kvadratov* (Exposition of the Method of Least Squares). Petersburg.
Markov A. A., Jr. (1951), Biography of A. A. Markov, Sr. In Markov (1951, pp. 599 – 613). **S, G,** 5.
Mendeleev D. I. (1860), On the cohesion of some liquids. *Sochinenia* (Works), vol. 5. Moscow – Leningrad, 1947, pp. 40 – 55**.**
Newcomb S. (1872), *On the Right Ascensions of the Equational Fundamental Stars*. Washington.
--- (1897), A new determination of the precessional motion. *Astron. J.*, vol. 17, pp. 161 – 167.
Neyman J. (1934), On two different aspects of the representative method. *J. Roy. Stat. Soc.*, vol. 97, pp. 558 – 625.
--- (1938/1952), *Lectures and Conferences on Mathematical Statistics and Probability*. Washington.
Ondar Kh. O., Editor (1977, in Russian), *The Correspondence between Markov and Chuprov.* New York, 1981. I (1990/2011, Chapter 8.2) added some letters and (Ibidem, Chapter 8.3) corrected many mistakes made because of Ondar's incompetence. Moreover, he corrupted the archival text by trimming it to his own stupid satisfaction.
Pearson K. (1900), On a criterion that a given system of deviations … can be … supposed to have arisen from random sampling. *Lond., Edinb. and Dublin Phil. Mag.*. vol. 50, pp. 157 – 175.
Plackett R. L. (1949), Historical note on the method of least squares. *Biometrika*, vol. 36, pp. 458 – 460.
--- (1972), Discovery of the method of least squares. Ibidem, vol. 59, p. 239 – 251. Reprint: Kendall M. G., Plackett R. L. (1977), *Studies in History of Statistics and Probability*, vol. 2. London, pp. 279 – 291.
Schéffe H. (1959), *The Analysis of Variance*. New York.
Seneta E. (1997), Markov. In Johnson N. L., Kotz S., Editors (1997), *Leading Personalities in Statistical Sciences*. New York, pp. 263 – 265.
Sheynin O. (1979), Gauss and the theory of errors. *Arch. Hist. Ex. Sci.*, vol. 20, pp. 21 – 72.
--- (1989), Markov's work on probability. Ibidem, vol. 39 pp. 337 – 377; vol. 40, p. 387.
--- (1990a, in Russian), *Chuprov: Life, Work, Correspondence*. Göttingen 1996, 2011.
--- (1990b), Markov's report on a paper by Galitzin. *Istoriko-Matematich. Issledovania*, vol. 32/33, pp. 451 – 467. Translation: Sheynin (2004, pp. 117 – 131).
--- Translator (2004), *Probability and Statistics. Russian Papers*. Berlin. **S, G,** 1.
--- (2007), True value of a measured constant and the theory of errors. *Historia Scientiarum*, vol. 17, pp. 38 – 48.
--- (2008), Bortkiewicz's alleged discovery: the law of small numbers. Ibidem, vol. 18, pp. 36 – 48.
--- (2017), *Theory of Probability. Historical Essay*. Berlin. **S, G,** 10.
Slutsky E. E. (1912), *Teoria Korreliatzii* (Theory of Correlation). Kiev. **S, G,** 38.
Stokalo I. Z., Editor (1967), *Istoria Otechestvennoi Matematiki* (History of National Mathematics), vol. 2. Kiev.

Usov N. A. (1867), A remark concerning the Chebyshev theorem. *Matematich. Zbornik*, vol. 2, pp. 93 – 95. **S, G,** 5.
Vasiliev A. V. (1885), *Teoria Veroiatnostei* (Theory of probability). Kazan. Mimeographed edition.

II

Ivory's Treatment of Pendulum Observations

Historia Mathematica, vol. 21, 1994, pp. 174 – 184

Abstract

James Ivory (1765 – 1842) contributed to the mathematical theory of attraction. I describe his efforts (1826 – 1830) at determining the earth's ellipticity (*e*) through the adjustment of pendulum observations. At the time, several dozens of such observations had already been made in various latitudes, and their adjustment presented difficulties owing to the local anomalies of gravity. The very possibility of deducing a single ellipticity for the earth remained questionable. While achieving his goal, Ivory made methodical mistakes which he gradually corrected. His final result, $0.00333 < e < 0.00338$, favourably compares with the value $e = 0.00335$ of the so-called Krasovsky ellipsoid. Ivory's work was forgotten mainly because new data, especially on meridian arc measurements, became available rather soon after its publication.

1. Introduction

James Ivory, a Fellow of the Royal Society, is best remembered as a contributor to the mathematical theory of attraction [22, vol. 2]. His attempts to adjust pendulum observations seem to have been largely overlooked: Strasser [21] did not mention him at all. Ivory [4] also offered several substantiations of the method of least squares (MLSq).

Later authors, for example Czuber [2, pp. 301 – 304], sharply criticized them. Gauss himself, in a letter to Olbers dated March 15, 1827 [18, pp. 475 – 476], found them unsatisfactory. However, Ivory's paper is extremely interesting since he was one of the first to suggest that the MLSq be based on the principle of maximum weight (of least variance). He did not refer to Gauss [3]; possibly he had not yet read it. I discuss this topic in [20].

In his letter to Olbers, Gauss also criticized Ivory's contribution (obviously, [5]) on the adjustment of pendulum observations. He mentioned, without going into detail, that the *spirit* [Geist] of the MLSq was *utterly alien* [ganz fremd] to the English scientist; that his manner of combining observations was *utterly unworthy* [ganz unwürdig], and that his paper was unmethodical [had wenig logische Ordnung]. In the same letter, however, Gauss remarked that long ago he had appreciated Ivory as an *acute* [scharfsinnigen] mathematician.

Here, I describe a series of Ivory's papers published over a very short period of time on the adjustment of pendulum observations[1]. I begin in § 2 by drawing the connections between pendulum observations and the ellipticity of the earth. I then discuss Ivory's adjustment procedures (§ 3); take notice of how he estimated (or failed to estimate) the precision of observations (§ 4); and study his thoughts on the existence of local gravimetrical anomalies (§ 5). I offer my assessment of Ivory's work in § 6. An additional § 7 is devoted to field computations.

2. Gravity and ellipticity

The main formula connecting the acceleration of gravity g with the length L and semi-period of vibration T of a pendulum is

$$T = \pi\sqrt{L/g}.$$

After measuring g at two stations it becomes possible to determine g_p and g_0, the gravity at the pole and the equator (compare (l) below) and, after that, the earth's ellipticity e by means of the Clairaut theorem

$$e = (a - b)/a = \frac{5F_0}{2g_0} - \frac{g_p - g_0}{g_0},$$

where[2] F_0 is the centrifugal force at the equator and a and b are the semi-axes of the earth's ellipsoid of rotation ($a > b$).

The general practice was to make use of seconds pendulums having $L \approx 1$ *meter*, which led to $T \approx 1$ sec. The observed magnitude at each station was the number of the pendulum's swings per day ($N \approx 86{,}400$). Assuming that the length of the pendulum used (or, more appropriately, the approximate pendulum) did not vary during a given voyage[3] and knowing the length of a pendulum which beats seconds exactly (that is, the exact pendulum) at their base station, astronomers were able to calculate the length of the exact pendulum at any point of observation. I discuss only the exact pendulum.

The equation connecting the lengths of the pendulum at latitude φ (L_φ) with those at the pole (L_p) and at the equator (L_0) is

$$L_\varphi = L_0 + f\sin^2\varphi,\ f = L_\varphi - L_0. \qquad (1)$$

Consequently, the equations of condition are

$$L_0 + f\sin^2\varphi - L_\varphi = v_\varphi,\ \varphi = \varphi_1, \varphi_2, \ldots, \varphi_n, \qquad (2a)$$

or, in general notation,

$$a_ix + b_iy + m_i = v_i,\ i = 1, 2, \ldots, n \qquad (2b)$$

with $a_i = 1$.

Many stations, rather than only two, were needed in order to compensate for random errors and, it was hoped, to diminish systematic influences such as the effects of local attraction or irregularity of the earth's shape. At the time, measurements were reduced to mean sea level; nowadays, reduction is a much more delicate procedure.

Among the first absolute determinations of gravity, involving the measurement of both the number of vibrations of a pendulum per day and its length, were those made by C. M. de la Condamine; by J. C. de

Borda and Jacques Cassini; and by F. W. Bessel. Borda and Cassini, in 1792, determined the length of the seconds pendulum at Paris bearing in mind Condamine's much earlier and then not yet rejected idea of defining the unit of length as that of the seconds pendulum at latitude 45° [16, pp. 198 – 200].

3. Adjusting observations

3.1. Using pairs of observations. Ivory [5, p. 9] began his work by adjusting six pendulum observations made by E. Sabine. He combined the only southern station, Maranham with latitude $\varphi = 2°32'$ with each of the others and justified this by noting that the corresponding variations of the pendulum's length

May be supposed very great in proportion to the errors of observation

and that therefore the dependence between the pairs may be neglected. His unknowns were L_0 and e rather than L_0 and f as in (l), but this difference was hardly essential.

Adding two more observations made by French scientists and building up two more combinations with Maranham accordingly, Ivory [5, p. 10] calculated the mean ellipticity of all seven results, which happened to be *extremely near* 1/300.

In the second part of his article, he continued his calculations in the same manner, combining Maranham with three different groups of northern stations.

A large variation between the pendulum's length at the two stations of a pair was really essential: Ivory [5, p. 94] noted that the error of e, as deduced from a pair, increased as

$$(\sin^2\varphi_1 - k\sin^2\varphi_2),\ k = L_l/L_2 \approx 1 \qquad (3)$$

decreased. He did not, however, determine the weights of the results obtained from a given pair, tacitly assuming that all the pairs were of equal value. Actually this is wrong; the values of e obtained from different pairs have differing weights[4].

3.2. An indirect use of the MLSq. Next, Ivory [5, p. 98] considered 13 observations made by Sabine. He rejected two of them and adjusted the remaining by a strange procedure. He stated that the usual condition of least squares, was not good enough, but it is very difficult to understand the reasoning behind his claim.

He subtracted the first of equations (2a) from each of the subsequent and found the only remaining unknown, f, by least squares! Calculating backwards, he then determined L_0 by assuming $v_i = 0$, although he also noted that it was possible to take

$$\Sigma v_i = 0. \qquad (4)$$

Ivory then made similar calculations issuing from the same observations but joined Sabine's station Maranham to them [5, p. 100]. His results for the two sets of observations were $e = 0.00333$ and 0.00329, respectively.

Elsewhere, Ivory [6, p. 245] stated that condition (4) was much better than the assumption $v_i = 0$ for some i. Now, for $a_i = 1$, see (2b), this condition will coincide with the first normal equation. It follows that Ivory's method of adjustment involving condition (4) did not differ from the MLSq.

Ivory [6, p. 244] had also declared that neither [*av*] nor [*bv*] should vanish, thus reiterating his previous statement above. This, however, was a remarkable misunderstanding since he actually used least squares and, in addition, since the condition of the MLSq applied to residuals v_i's rather than to errors of observation.

Ivory supplemented his declaration by checking Sabine's calculations. The latter, who had adjusted his observations by least squares, arrived at $e = 0.00346$, whereas Ivory, upon leaving out two, and then four stations from the original thirteen and *calculating in the same manner* [6, p. 242], obtained $e = 0.003405$ and 0. 00337, respectively.

Ivory seemed prepared to believe that the discrepancies were occasioned by the deficiency of the MLSq rather than by an *irregular deviation of the earth's surface from the elliptical figure* [6, p. 242]. Does this mean, then, that he forgot his own conclusion that the first two rejected observations were corrupted by *some local anomaly* [5, p. 95]? In [7, p. 246] Ivory admitted that his previous investigation [6] *can be considered as no more than a preliminary inquiry*, and he treated a larger number of observations[5], again indirectly using the MLSq. Finally, he [10, p. 168] stated that both this method and its *modification* (note this insufficient acknowledgment!) were useful only *in bringing out a first approximation*. What he actually did [10, pp. 169 – 170] was to determine a first approximation in a somewhat arbitrary way and to correct it by least squares, concluding that $e = 0.00338$ [10, p. 172]. It seems that he just did not notice that his approach was tantamount to using least squares from the very beginning.

3.3. Combining stations little differing in latitudes. Even if *f* in (1) is not known precisely, the length of the seconds pendulum L_i at one station can be deduced with considerable accuracy from its measured length L_j at another station with approximately the same latitude. Ivory (§ 5.1 below) used this procedure for checking the precision of some observations, whereas Biot [1, p. 16 – 17] went one step further by calculating the mean latitude corresponding to the mean of several observed values of L_i. He did not say anything about the weight of the mean result, nor did he expressly recommend combining stations before adjusting all observations available.

Now, Biot should have given more thought to this possibility, given that he believed that the coefficient of $\sin^2\varphi$ somewhat depended on φ, that is, that the meaning of *f* in (1) should be different [l, p. 18], see also § 4. It followed that stations with approximately the same latitudes should be combined into one having the same weight as that of the initial stations. However, if such stations were not far apart (if, in addition, their longitudes did not differ significantly one from another), then all the relevant observations could be corrupted by a local anomaly to (almost) the same extent and their combination, as described above, would be all the more necessary.

4. Estimating precision

After adjusting 40 pendulum observations, Ivory [10, p. 172] remarked that 35 of them were within the limits of the probable errors. He explained that in each of those 35 cases the difference between the (indirectly) observed and adjusted lengths of the pendulum was less than *what would arise from an error of* 2 *vibrations in a mean solar day*.

I believe that Ivory used the term *probable error* in a loose sense, perhaps not even knowing its exact meaning. Indeed, after considering the discrepancies between the lengths of pendulums observed twice at each of three stations, he concluded that

Such experiments are liable to an error amounting to … from two to three vibrations in a mean solar day,

and, in fact, the error *may be much greater* [10, p. 167]. Elsewhere, he [5, p. 9] mentioned a discrepancy of the same order.

When adjusting observations, Ivory never once determined the mean square difference between observed and final results. Similarly, as noted in § 3.1 above, he did not calculate the precision of L_0 or f.

While discussing a statement made by Biot [l, p. 18] on the variability of f in (l), however, Ivory [14, pp. 413 – 415] attempted to study the precision of f as determined from equations of the type of (1), or, more precisely, from such equations with L_0 equal, in turn, to any of the observations made in the equatorial zone and reduced to latitude $\varphi = 0$.

Instead (or additionally), he should have used appropriate differential formulas. Ivory [14, p. 415] concluded that until L_0 was *decisively ascertained* the coefficient f and the ellipticity of the earth will remain *in some degree indeterminate*.

He thus confirmed his conclusion in [9, p. 353] which constituted a reversal of his earlier optimism [5, p. 9]. Still, upon obtaining five values of e, call them $e_1, e_2, \ldots, e_5$ with $e_1 < e_2 < \ldots < e_5$, Ivory expressed satisfaction with his results.

Suppose that, in my own notation, the mean value of e is $\bar{e}$. Then, as he remarked, the values of e were sufficiently close to, each other since $e_5 - e_1 < 1/20\,\bar{e}$ and both $(\bar{e} - e_1)$ and $(e_5 - \bar{e})$ were less than $1/46\,\bar{e}$.

It is difficult to share Ivory's satisfaction, however: all three differences were apt to increase with the increase in the number of observations, and he failed to calculate the variance of his value (of $\bar{e}$). Moreover, regarding the combination of results obtained by several observers, Ivory, upon adjusting all available observations by least squares, naturally calculated the corrections (the residuals), but did not notice any systematic differences between them [10, p. 171] although they were obvious.

Accordingly, it was desirable to adjust the observations anew, at least tentatively, assigning different weights to groups of observations and, in two cases, introducing unknown general corrections to the groups.

So here are some of his data (observer, number of observations, mean square correction, mean correction):

Biot	6	131	79

Kater	7	94	67
Sabine	10	182	16

The unit of measurement was 10^{-5} inches. In all, Ivory treated 34 observations made by more than 12 observers. The mean square correction of all these observations was 153×10^{-5} inches.

5. Local anomalies

5.1. Rejection of outlying observations. Ivory [5, p. 92] believed that each pair of properly combined observations (as discussed in § 3.1 above), which provided his two unknowns, should be taken into account. At the same time, after comparing with each other those made at several stations near the equator at about the same latitude, he [5, p. 95] rejected two observations. He [5, p. 94; 8, pp. 323 and 326; 9, p. 352; 10, p. 165] also noticed that, in general, observations near the equator were irregular.

On other occasions, after (why after?) adjusting the observations by least squares, Ivory rejected a large proportion of them, 31, 27, and 12% in [6, p. 242], [7, p. 250], and [10, pp. 169 – 170], respectively. He justified this [5, p. 95; 7, p. 250] by local anomalies, see § 3.2.

Elsewhere, because of *great anomalies*, Ivory [11, p. 243] even expressed doubts about the possibility of determining a single figure of the earth. He also stated [14, p. 416] that to ascertain *the exact quantities* of the anomalies and to detect their causes was *the most important and interesting part* of gravimetric investigations[6].

Nevertheless, Ivory (10, pp. 172 – 173] referred to *the splendid speculation about local attraction* and called it *premature*. Finally, he [10, pp. 206 – 207] stated that he

Always thought it necessary to leave out a few of the experiments that were inconsistent with the rest,

for otherwise it would have been impossible to

Deduce … any conclusion respecting the figure of the earth in which much confidence can be placed.

5.2. Special adjustment procedures. Forgetting his doubts, Ivory declared that because of local anomalies special methods of adjusting pendulum observations were needed. Thus, [8, pp. 321 – 322] a adjustment of all observations might lead to a

Mean figure of the earth … considerably different from the true figure belonging to the consistent observations alone[7].

Accordingly, he recommended to subdivide the measurements into *partial combinations*, investigate *the ellipticity of every separate combination*, and examine *whether all the results agree or disagree* [8, p. 322]. In the latter case, he continued, it was necessary to reject anomalous observations.

Note that Ivory’s treatment of separate pairs of observations (as in § 3.1) was an extreme case of dealing with partial combinations. The subdivision of the measurements could have been accomplished in more than one way, and he himself stated elsewhere that arbitrary combinations can lead to *any ellipticity we choose* [9, p. 353].

6. Some comments

For many years, Laplace [19, p. 48 – 49] was uncertain whether the figure of the earth may be represented by an ellipsoid. Consequently, in 1825, just before Ivory began his investigations, Laplace had recommended the use of certain lunar observations not susceptible to local terrestrial anomalies[8] and thus capable of providing more consistent data. This, then, was the main point: at the time, the treatment of meridian arc measurements and pendulum observations was done principally to test the hypothesis just mentioned.

Accordingly, Laplace used the minimax principle on several occasions. He attempted to determine such values of his unknowns, for example, x and y in equations (2b) that led to the minimal value of the maximal $|v_i|$, $i = 1, 2, \ldots, n$, over all possible sets of (x, y). In other words, he checked whether his hypothesis fit the observations. Provided that the answer was positive (or "almost" positive, after rejecting a few observations), the actual adjustment could have been done, for example, by least squares.

Ivory attempted to check that hypothesis and to adjust the observations simultaneously, and he therefore had to proceed by trial and error. As a result of the serious difficulties he experienced, he came to believe, as Laplace did, that observations were corrupted by local anomalies and that the equatorial zone was inadequately covered by the observations. The latter fact, in particular, was impossible to overlook.

In pursuing his research, however, Ivory made mistakes. First, while pairing observations, he combined one and the same station with many other stations. He thus multiplied the southern observation as if adding a group of stations, all of them located at the same latitude and having the same acceleration of gravity. On the other hand, a southern station was badly needed for the pairing.

Ivory's later use of the MLSq allowed him to avoid this method, but even in 1828 he did not acknowledge that the number of southern stations was far less than the number of northern, so that the figure of the earth was not really studied in the equatorial zone.

Second, while actually adjusting observations by the MLSq, Ivory declared that it was unfit in this context. As noted in § 3.2 above, his explanation of this was less than satisfactory. Third, he did not properly estimate the precision of field measurements or of his final results. Fourth, while proposing to adjust observations by separate groups (and adjusting them in pairs), he did not give thought to weighing these groups or pairs. All this means that in treating observations, Ivory was an amateur.

In his various investigations Ivory offered many *final* values of e, all in the interval [0.00329; 0.00340]. He thus reasonably assumed that no single value might be chosen at the time [11, p. 242]. Instead, he stated that $0.00333 < e < 0.00338$. It is instructive to compare this estimate with the same figure for the so-called Krasovsky ellipsoid[9]: $e = 1/298.3 = 0.00335$ [15]. Again, according to Strasser [21, pp. 28 – 32], a better value of e was determined only once, in 1818, by J. C. Bonsdorff: $e = 1/298.5$.

Gauss's criticism (§ 1) of Ivory's papers was just; and, moreover, as noted above, Ivory's later work was also faulty in several respects. Taken together, his efforts could have been fruitful in the practical sense, but later scientists evidently overlooked Ivory's work since it offered no theoretical novelties and since new data, especially on meridian arc measurements, became available rather soon after its appearance.

Ivory, however, had tackled a difficult problem. The adjustment of observations corrupted by considerable systematic errors is extremely vexing even in our time. In the case of pendulum observations, it is even unclear when a large local anomaly should be treated as such, or considered as a distinctive feature of the earth's gravitational field.

7. A note on field computations

I provide the data from Sabine [17]. It shows how he (and, no doubt, other observers as well) determined the number of the daily vibrations of their pendulums. The final magnitudes in this particular case were

$N_{\text{London}} =$ 86,455.6490 (the base station),
$N_{\text{Melville Island}} =$ 86,530.3827.

The differences in the final column were obviously calculated in order to check each of the four results against each other (and against their mean at various stations). Indeed, the figures in each of the two middle columns scattered greatly since they pertained to various pairs of instruments used, but those differences should have been, and actually were, much closer. Such calculations showed astronomers in the early decades of the 19th century whether or not a given combination of instruments was stable.

Indirect relative determination of gravity, Sabine [17, p. 188]
daily vibrations at London (L) and Melville Island (M), also
clock (1 or 2) and pendulum (1 or 2) and difference (L – M)

1, 1	86,392.4513	86,466.4793	74.0280
1, 2	545.0623	620.6646	75.6023
2, l	388.0967	462.5289	74.4322
2, 2	496.9855	571.8580	74.8725

I have corrected an insignificant error in the differences. They testify that only one decimal place was needed. Compare § 4.

Acknowledgments. This paper represents a part of the research programme on the history of the theory of errors undertaken at the Mathematical Institute of the University of Cologne (Professor J. Pfanzagl) with the support of the Axel-Springer Stiftung. The editor, Karen Parshall, corrected my English and revealed quite a few ambiguities which I have addressed. Any residual faults are my own.

Notes

1. For the sake of completeness, I also include two of his related articles [12] and [13] which were devoted to the adjustment of meridian arc measurements.
2. Ivory [5; 6] introduced $2\varepsilon = (a - b)/b = e$. He did not write out the first of these two relations, that is, the definition of ε, but it can easily be determined. Then Ivory forgot his notation and replaced ε by e [5, p. 93]. My own definition of ε is given above. The difference between $(a - b)/b$ and $(a - b)/a$ is of the second order. Ivory

later denoted $(a^2 - b^2)^{1/2}/a$ by e and called $(a - b)/a = \varepsilon$ *the compression or the ellipticity* [12, pp. 343 – 344]. These changes of notation are unfortunate.
3. Several corrections were applied, one of which, for example, allowed for the change of air temperature.
4. I have calculated the weight of e as determined from n observations,

$$p = (L_1\sin^2\varphi_2 - L_2\sin^2\varphi_1)^2 + \dots$$

where the dots stand for similar terms with all subscripts changing cyclically ($1 \to 2$; $2 \to 3$; … ; $(n - 1) \to n$; $n \to 1$).
5. The number of observations rose sharply from 13 to 26 which warranted Ivory's repeated attempts at adjusting them.
6. Compare his earlier pronouncement [9, p. 352] (later largely repeated by Biot [l, p. 14]):

The purpose of a formula … is, not to extinguish discrepancies actually existing in Nature, or supposed so to exist, but to exhibit them as they really are.

Ivory [10, p. 173] correctly suspected that *the great defect of density in the waters of the ocean* corrupted insular observations but dismissed *conjecture and opinion*. He naturally did not know that the geoid (the equipotential surface of gravity coinciding with mean sea level, a term coined by Johann Benedict Listing in 1873) deviates from the earth's ellipsoid over large territories.
7. Biot [l, p. 14] was of the same opinion.
8. Ivory [8, p. 322] mentioned such observations only once, in passing, without indicating their advantage or referring to Laplace.
9. Feodosy Nikolaevich Krasovsky deduced the parameters of this ellipsoid in 1940. His closest associate and one-time student, Aleksandr Aleksandrovich Izotov, published a detailed account [15] of their work and of some further developments. After 1940, geodesy underwent further dramatic development due to the invention of essentially new rangefinders and the use of observations of artificial earth satellites. Consequently, I do not compare Ivory's results with the most recent findings.

References

1. J. B. Biot, Mémoire sur la figure de la terre, *Mém. Acad. Sci. Paris*, t. 8 (1829), pp. 1 – 56.

2. E. Czuber, *Theorie der Beobachtungsfehler.* Leipzig, 1891.

3. C. F. Gauss, Theoria Combinationis …, 1823, translated as Theorie der den kleinsten Fehlern unterworfenen Combination der Beobachtungen, in author's *Abh. zur Methode der kleinsten Quadrate*, Eds. A. Börsch, P. Simon. Berlin, 1887. Latest edition: Vaduz, 1998. English translation: G. W. Stewart, *Theory of combination* … Philadelphia, 1995.

4. J. Ivory, On the method of least squares. *Phil. Mag.*, vol. 65 (1825), pp. l – 10, 81 – 88, 161 – 168, and vol. 68 (1826), pp. 161 – 165.

5. ---, On the ellipticity of the Earth as deduced from experiments made with the pendulum. Ibidem, vol. 68 (1826), pp. 3 – 10, 92 – 101.

6. ---, On the methods proper to be used for deducing a general formula for the length of the seconds pendulum from a number of experiments made at different latitudes. Ibidem, pp. 241 – 245.

7. ---, Disquisition concerning the length of the seconds pendulum and the ellipticity of the Earth. Ibidem, pp. 246 – 251.

8. ---, On the grounds for adopting the ellipticity of the Earth deduced by Captain Sabine from his experiments with the pendulum in his work lately published. Ibidem, pp. 321 – 326.

9. ---, Short abstract of de Freycinet's experiments for determining the length of the pendulum. Ibidem, pp. 350 – 353.

10. ---, On the ellipticity of the Earth as deduced from experiments with the pendulum. Ibidem, new ser., vol. 3 (1828), pp. 165 – 173, 206 – 210.

11. ---, Letter to the Editors relating to the ellipticity of the Earth as deduced from experiments with the pendulum. Ibidem, pp. 241 – 243.

12. ---, On the figure of the Earth as deduced from measurements of different portions of the meridian. Ibidem, pp. 343 – 349.

13. ---, On the figure of the Earth as deduced from measurements of the meridian. Ibidem, pp. 431 – 436.
14. ---, On the figure of the Earth. Ibidem, vol. 7 (1830), pp. 412 – 416.
15. A. A. Izotov, *Forma i razmery Zemli po sovremennym dannym* [Figure and Size of the Earth According to Modern Data]. Moscow, 1953.
16. A. A. Mikhailov, *Kurs gravimetrii i figury Zemli* [A Course in Gravimetry and the Figure of the Earth]. Moscow, 1939.
17. E. Sabine, An account of experiments to determine the acceleration of the pendulum in different latitudes. *Phil. Trans. Roy. Soc. London* (1821), pp. 163 – 190.
18. C. Schilling, *Wilhelm Olbers. Sein Leben und sein Werk*, Bd. 2, Tl. 2. *Briefwechsel zwischen Gauss und Olbers*. Berlin, 1909. Reprint: C. F. Gauss, *Werke, Ergänzungsreihe*, Bd. 4. Hildesheim, 1976.
19. O. Sheynin, Laplace's theory of errors. *Arch. Hist. Ex. Sci*., vol. 17 (1977), pp. 1 – 61.
20. ---, C. F. Gauss and geodetic observations. Ibidem, vol. 46 (1994), pp. 253 – 283.
21. G. Strasser, *Ellipsoidische Parameter der Erdfigur (1800 – 1950)*. München, *Deutsche geod. Kommission Bayer. Akad. Wiss.* Bd. A19, 1957.
22. I. Todhunter, *History of the Mathematical Theories of Attraction and the Figure of the Earth*, vols. 1 – 2. London, 1873; reprinted, New York, 1962; Nabu Press, 2010.

III

On the mathematical treatment of observations by Euler

Arch. hist. ex. sci., vol. 9, 1972, pp. 45 – 56

Abstract

Euler's memoirs on the mathematical treatment of direct observations are described in § 1. His commentaries on memoirs of Lagrange and Daniel Bernoulli are expounded and, in particular, one of Euler's remarks is heuristically related to the method of last squares (MLSq). The treatment of indirect observations is considered in § 2. There also Euler's use of several methods of calculation which preceded the MLSq is studied.

A special § 3 is devoted to a short description of Euler's work in population statistics and to the question why his achievements in the theory of probability were insignificant is raised.

Being an outstanding astronomer[1], Euler mainly treated its general theory but he also repeatedly took up the mathematical treatment of observations and (as also was in the case of map projections) performed numerical calculations in general. When discussing them, I restrict myself to calculations related to the theory of errors and probability.

1. Treatment of direct observations

1.1. Euler's commentary on Lagrange's memoir. He devoted two of his writings[2, 3] to the treatment of such observations, but both only appeared as commentaries on the memoirs of Lagrange[4] and Daniel Bernoulli[5].

The memoir of Lagrange is an important extension of the work of Simpson published in 1756 and 1757, mostly in the general mathematical direction[6–8]. I only consider Subbotin's appraisal[9] of Lagrange's astronomical work in general: Lagrange engaged in astronomy (in the theory of errors)

Not like a natural scientist who desires to penetrate deeper into nature's mysteries, but like a mathematician who seeks new problems and aspires to extend the range of use of his mathematical methods.

In a letter of 10 February 1777 to Euler Lagrange[11] wrote:

Si vos occupations et l'état de votre santé vous ont permis de jetter les yeux sur le peu que j'ai donné … je vous supplie de vouloir bien m'en dire votre avis.

I may suppose that lacking such request Euler would not have written the commentary. It appeared only in 1788, after Laplace had published his first memoirs on probability and it seemed hardly interesting. It added nothing either in ideas or mathematical methods. The following problem which was appropriate in the memoir itself was solved by Euler: errors α, β, γ occur with probabilities proportional to a, b and c respectively. Determine the probability that the error of the arithmetic mean of n observations is λ/n. Answer: it is the coefficient of x^{λ} in the development of $(ax^{\alpha} + bx^{\beta} + cx^{\gamma})$ divided by $(a + b + c)^n$.

1.2. Daniel Bernoulli's memoir. He repeatedly took up the theory of errors[12a]. Long before that memoir appeared, he had compiled a manuscript later described by J. Bernoulli[14, 15]. Daniel had sent this manuscript, *un petit écrit latin*, under the same title to J. Bernoulli in 1769, see Note 5. Its goal was the calculation of the *real value* of an observed constant whose somewhat discrepant observations were x_1, x_2, …, x_n and the frequency of their errors a semi-ellipse or a semi-circumference with a subjectively assigned radius r. On the notion of real value see Sheynin (2007b).

Bernoulli was dissatisfied with the usual arithmetic mean and proposed instead

$$\xi = \Sigma p_i x_i : \Sigma p_i,\ p_i = r^2 - (\xi - x_i)^2, \qquad (1;\ 2)$$

Successive approximations are needed and the first approximation to ξ can be the usual arithmetic mean.

Bernoulli tackled the same problem in his published memoir[16]. He considered the usual mean advisable, as he thought, only in the case of equal probability of all errors. But *such an assertion would be quite absurd* (§ 2) and would mean that *the most skilful shot would have no advantage over a blind man* (§ 5).

Here, however, is K. Pearson's (1978, p. 168) qualitative statement*: Small errors are more frequent and have their due weight in the mean*.

For convenience of calculation Bernoulli left the semi-circumference for an arc of a parabola

$$y = r^2 - (\xi - x)^2,\ y \geq 0, \qquad (3)$$

but he certainly did not know that the variance of the result will therefore change.

Instead of the arithmetic mean Bernoulli proposed a maximal likelihood estimator which coincided with the mode of his curve (3). Calculations proved too difficult and he only considered the case of $n = 3$, and even then only numerically. He somehow did not notice that his estimator could be calculated from (1) with weights being the inverse of weights (2), denote them by (4).

Just as in his earlier manuscript, successive approximations are possible and perhaps two or three would be enough. Now, the weights of observations (4) increase with the increase of their distance from their central group which Bernoulli did not state. This strange fact was confirmed only recently by Lloyd's best linear estimators.

And so, ξ depended more upon the extreme rather than on central observations whereas Bernoulli's contemporaries could have mistakenly concluded from his qualitative initial reasoning that the weights of the extreme observations should be reduced.

In a historical sense, posterior estimators anticipated the best linear estimators of the location parameter. The first to use them was apparently A. G. Pingrè, 1711 – 1796 (1761)[17]. Idelson[18] provided a survey of their use but he only begun with S. Newcomb's memoir of 1886 and continued with several developments of the 20th century.

1.3. Euler's commentary on Daniel Bernoulli's memoir. I do not describe the well known reasoning of Gauss who originally (in 1809) assumed the principle of maximum likelihood as the basis of the theory of errors but afterwards (in 1823) rejected it in favour of the principle of maximal weight (of least variance).

Euler was the first to misunderstand Bernoulli. In § 2 he stated that the proper weights are (2) and in § 4 he ascribed them to Bernoulli. His mistake cannot be explained away as a mathematical blunder made while reducing the maximal likelihood estimator to (1). Indeed. First, Euler (§ 6) altogether denounces the principle of maximal likelihood; second, he recommends the use of estimators of the type of (1) which he regards as corresponding with the *undoubted precepts of the theory of probability* (§ 7). He did not notice the connection of (1) and the maximal likelihood estimator.

Euler's objection to the principle of maximum likelihood is that if *among the observations … there is one that should be almost rejected,* even the maximum value of the likelihood function will become extremely small. Perhaps he thought about the impossibility of a precise calculation of a slurred maximum. In any case, according to (1) and (4) errors of the extreme observations will significantly affect the estimator sought, so that a thorough preliminary discussion of the observations with possible rejection of those most outlying is necessary. On the other hand, Euler argued, the result of an adjustment should barely change whether or not a deviating observation was adopted, which meant that the never mentioned (!) median should be the estimator of the parameter of location. Instead, he recommended the mean with posterior weights which arguably followed from *the undoubted precepts of the theory of probability.*

Indeed, Euler recommended, instead of the arithmetic mean, the estimate (1) with posterior weights (2) and he mistakenly assumed that Bernoulli had actually chosen these same weights. While developing his thoughts, and denoting the n observations by $\Pi + a$, $\Pi + b$, $\Pi + c$, …, where

$$a + b + c + \ldots = 0, \qquad (5)$$

he formed the equation

$$nx^3 - nr^2x + 3Bx - C = 0,$$
$$B = a^2 + b^2 + c^2 + \ldots,\ C = a^3 + b^3 + c^3 + \ldots,$$

from which the estimator $\Pi + x$ should have been calculated with x equal to its root least in absolute value. Condition (5) meant that the estimator sought was the closest possible to the arithmetic mean; Euler himself (§ 9) justified his choice of the root by noting that $x = 0$ as $r \to \infty$, that is, as $n \to \infty$.

Neither Bernoulli, nor Euler offered a definite rule for defining r. Euler noted that it should equal the distance between ξ and the observation *which is to be all but rejected* (§ 3). Certainly, since this is the essence of *r*; it tends to increase with the number of observations but remains bounded.

In a normed way curve (3) becomes

$$y = \frac{3}{4r^3}[r^2 - (\xi - x)^2],\ \xi - r \le x \le \xi + r$$

and the best linear estimator of ξ results[19] in posterior weights in formula (1) sharply rising towards the tails of the observational series whereas Euler's weights (2) decrease towards the tails.

Euler provided examples. In one of them (§13) he assumed that the differences between the meridians of Paris and St. Petersburg was 1°52' instead of $1^{h}52^{min}$.

Euler (§ 11) also remarked that estimator (1) with weights (2) can be obtained from the condition

$$[r^2 - (\xi - a)^2]^2 + [r^2 - (\xi - b)^2]^2 + [r^2 - (\xi - c)^2]^2 + \ldots = \max. \quad (6)$$

The magnitudes in parentheses are the deviations of observations from the estimator sought and their fourth powers are negligible so that condition (6) is equivalent to the requirement

$$(\xi - a)^2 + (\xi - b)^2 + (\xi - c)^2 + \ldots = \min, \quad (7)$$

whence, in accordance with condition (5), follows the arithmetic mean. Condition (7) is heuristically similar to the principle of least squares (which in case of one unknown indeed leads to the arithmetic mean) and condition (6) with weights (2) resembles the Gaussian principle of maximum weight (of least variance). True, if the density of the observational errors is known (which was the unrealistic assumption of both Bernoulli and Euler), then other estimators can be better than the arithmetic mean.

A small deviation from condition (7) does exist and it is easy to see that it is occasioned by inevitable deviations of the observations from the proposed (or tacitly assumed) symmetrical law. Bernoulli himself noted this fact when (see above) numerically adjusting several sets of three observations. So, I say once more that in actual fact Bernoulli proposed the general arithmetical mean.

After hearing about the forthcoming publication of his memoir and its commentary, Bernoulli[20] wrote to Fuss:

Je suis surtout glorieux de ce que M. Euler en a pris occasion de régaler le public d'un autre mémoire sur la même matière. Cependant je suis sûr que ce grand analyste aura envisage la question d'un autre point de vue, que je n'ai fait.

But then, after undoubtedly noticed Euler's misunderstanding of his memoir, Bernoulli kept silent.

Euler was acquainted with Lambert's *Photometria*[21], with its first introduction of the maximum likelihood principle[22] and a reference to Lambert was highly desirable. Apparently, however, Euler was not interested in the appropriate section of the *Photometria*.

2. Treatment of indirect observations

Euler's interest in the problem of the figure of the Earth is generally known and perhaps its latest illustration is his recently published correspondence with P. L. Maupertuis, H. Kühn and N. L. De La Caille[23]. This topic as well as purely astronomical calculations in which he also engaged led to an indirect deduction of some unknowns $x, y, z, \ldots$ from redundant physically independent (linear independence was yet unknown) simultaneous linear algebraic equations

$$a_i x + b_i y + c_i z + ... + l_i = 0,\ i = 1, 2, \ldots, m > n \qquad (8).$$

with directly measured l_i.

These equations are evidently inconsistent and various additional conditions had to be imposed on the residual free terms which were usually designated v_i. It is natural to assume that they, or the initial equations, are mutually independent, unbiased and subject to some law of distribution although unavoidable systematic errors violated unbiasedness.

Euler was the first to use a definite rule, the minimax method, and among the first to apply the method of averages. Both had been among the main methods before the advent of the MLSq. On the other hand, Euler had not considered the adjustment from a general theoretical viewpoint.

Both Johann Albrecht and Christoph Euler directly participated in the treatment of observations in astronomy, meteorology etc.[24] although neither contributed to the appropriate theory. Their father undoubtedly dominantly influenced them, possibly helped them, but there possibly existed some feed-back from sons to father, or at least an additional airing of the adjustment.

2.1. The minimax method. Its condition is

$$|v|_{\max} = \min \qquad (9)$$

where the minimum is sought among all possible sets of adjusted unknowns. Goussac[25] described the history of this method and traced it to the Chebyshev problem of the best approximation of an analytic function on a given interval by a polynomial of a given order but he began his account with Euler[26]. In 1778 Euler modified the map projection of De Lisle so that the maximal distortion of the length of an arc of a parallel became minimal.

However, even in 1749 Euler[27] applied condition (9) for solving equations (8). When treating astronomical observations he constructed a system of 21 equations in six unknowns, studied a few (only a few) solutions and chose that for which the maximal v_i's, positive and negative taken separately were minimal. Those v_i's had the same absolute values, a fact which Euler did not explain either here or in his memoir of 1778 where the same equality of extremal distortions had occurred. It was Laplace[28] who provided the explanation without mentioning Euler.

The next after Euler was Lambert[29] who mentioned condition (9) but noted that was unable to apply it. Prior to Laplace and Gauss who introduced distribution laws no one was able to justify the various methods of solving equations (8). The minimax method was an obvious exception: any other method of solution will lead to a greater value of $|v_{max}|$, the gap between theory and observation will be wider, and the correctness of the former might have been mistakenly questioned. Alternatively, that method proved that the quality of the observations was not good enough.

A tendency for explanations did exist but the provided answers were purely qualitative. For example, in 1722 R. Cotes substantiated his method of treating direct (not indirect) observations by referring to their centre of gravity. Then, when treating arc measurements, Boscovich[30] strove for compliance with laws of probability:

Pour prendre ce milieu, tel qu'il ne soit point simplement un milieu arithmétique, mais qu'il soit plié par une certaine loi aux règles des combinaisons fortuites et du calcul des probabilités …

But he was naturally unable to prove that his method achieved the formulated goal. Even Legendre, in 1805, justified the proposed MLSq only by noting that it ensured equilibrium of sorts between the v_i's.

2.2. The method of averages. Also in 1749 Euler[31] used another method of treating observations, the method of averages. Having, e. g., in § 115, two equations of the type

$$x = a_i y + b_i z + \ldots + l_i, \; i = 1, 2,$$

with approximately equal coefficients ($a_1 \approx a_2$ etc.) Euler assumed that

$$x = \frac{a_1 + a_2}{2} y + \frac{b_1 + b_2}{2} z + ... + \frac{l_1 + l_2}{2}$$

Euler did not elaborate but his method is equivalent to assuming that

the sum of v_i is zero. (10)

He could have thought that (10) resulted from the equal probability of the errors of each sign which also led to the arithmetic mean in case of direct observations[32, 33]. Condition (10) could have barely used before, but it is implicitly contained in the proposal of Cotes of 1722 and Mayer[34] had applied it in a much more interesting case.

He solved a system of 27 equations in three unknowns by introducing three intermediate summary equations

$$x\Sigma a_i + y\Sigma b_i + z\Sigma c_i + \Sigma l_i = 0$$

where the summations were for i = 1, 2, …, 9; 10, 11, …, 18 and 19, 20, …, 27 respectively. His was a generalized method of averages. The plausibility of the results depended on the expediency of separating the initial equations and it seems that Mayer had made a reasonable choice; my own separation above was only an example.

In the 19th century, Cauchy[35] introduced a method of solving equations (8) which used condition (10), see a modern description of his method in Linnik[36]. Referring to an unpublished study of L. S. Bartenyeva, he provided a simple proof of the unbiasedness of the estimators in Cauchy's method and calculated their effectiveness.

2.3. Short cuts. A few years after 1749 Euler[37] calculated the flattening and dimensions of the Earth from four arc measurements. Eliminating the unknown parameters of the spheroid from the equations, he got a system of two equations whose unknowns were the corrections to the preliminary lengths of the arcs. Employing no definite rule, Euler found a few solutions and chose that which seemed most reasonable. His calculations were apparently unsuccessful and criticism followed[38, 39].

3. Other applications of the theory of probability

In Euler's time

Social demands … on the theory of probability did not yet overstep calculations pertaining to lotteries or theoretical elaborations of various games [of chance].

He thus explained the insignificance of the heritage bequeathed by Euler in the theory of probability. This is wrong. Apart from population statistics in which Euler himself worked (see below), from statistical studies of inoculation of smallpox; Daniel Bernoulli, 1766, (who had other achievements as well[40 – 42]) should be mentioned first of all; or some statistical studies of the influence of the Moon on the weather (and, specifically, on the air pressure by Lambert which Daniel encouraged), there was serious work in astronomy (Galileo, Flamsteed, Bradley, and, first and foremost, Kepler). I (2017) have discussed this material.

Euler's achievements in population statistics are serious[43 – 46]. He was the essential co-author of Chapter 8 of Süssmilch's *Göttliche Ordnung*, edition of 1765. This chapter is included in Euler's *Opera omnia*[47], where, on p. 533, the commentaries of the editor of the appropriate volume include references to Euler made by Süssmilch.

In one of his memoirs[48] Euler posed and solved problems about the probabilities of the duration of life, values of life annuities, tontines and deduced an approximate problem for the increase in the population in time.

As always, his reasoning is elegant and convincing and he methodically elaborates relevant calculations which are still interesting. True, it does not directly bear on the theory of probability. Thus, Euler did not try to introduce theoretical laws of mortality.

I still ought to mention Lambert once more. He[49, 50] persistently strove to establish a method for delimiting randomness and Design and laid the foundation of the theory of errors.

But why do ideas about probability form such an insignificant part of Euler's work? Why are they barely seen even in his *Letters to a Princess*[51]? Only in Letter 119 Euler discusses various types of certainty. Why had not he theoretically contributed to the adjustment of observations? Contrary to Daniel Bernoulli, Lambert and Laplace, Euler was apparently not specifically interested in the ideas or methods of probability. I even venture to suspect that he was somehow influenced by his known deep religiousness.

Acknowledgements. This is a generalized and extended version of my report at the section of history of mathematics of the International Congress of Mathematicians (Moscow, 1966). Professor Truesdell corrected my English both here and in two earlier papers published in his *Archive*.

References

OC = *Oeuvr. Compl.*
OO = *Opera omnia*

1. Subbotin M. F., 1958, Die astronomische Arbeiten L. Eulers. *Sammelband der zu Ehren des 250. Geburtstages L. Eulers*. Moskau, pp. 268 – 375. In Russian. Title of volume and contributions also in German. German summary on p. 376.

2. Euler L., 1785 (1788), Eclaircisemens sur la mémoire De La Grange … (Eneström 628). OO, ser. 1, t. 7. Leipzig – Berlin, 1923, pp. 425 – 434.

3. ---, 1777, 1778, Observationes in praecedentem dissertationem illustr. Bernoulli.
(E 488). Ibidem, pp. 280 – 290. English translation: *Biometrika*, vol. 48, 1961, pp. 13 – 18.

4. Lagrange J. L., 1776, Sur l'utilité de la méthode de prendre le milieu entre les résultats de plusieurs observations. OC, t. 2. Paris, 1868, pp. 173 – 234.

5. Bernoulli Daniel, 1777, 1778, Dijudicatio maxime probabilis plurium observationum discrepantium atque verisimillima inductio inde formanda. Euler's OO, ser. 1, t. 7,
pp. 262 – 279. English translation: together with translation of [3]; also together with translation of [3] in *Studies in history of statistics and probability*, Editors E. S. Pearson,
M. G. Kendall. London, 1970, pp. 157 – 167 and 167 – 172 respectively. D. B. refers to his unpublished manuscript which is now available in an English translation in *Festschrift for Lucien Le Cam*. Editors, D. Pollard et al. New York, 1997.

6. Encke J. F., 1853, 1853, Über die Anwendung der Wahrscheinlichkeitsrechnung auf Beobachtungen. In one of the volumes of his *Gesammelte math. und astron. Abh.*,
Bde 1 – 3. 1888 – 1889.

7. Todhunter I., 1865, *History of the mathematical theory of probability*. New York, 1949, 1965.

8. Czuber E., 1891, *Theorie der Beobachtungsfehler*. Leipzig.

9. Soubbotin M. F., 1937, Les travaux astronomiques de La Grange. In Russian, title of contribution also in French. *Zbornik statei k 200-letiu so dnia rozhdenia Lagranzha* (Coll. of articles on the bicentenary of La Grange's birth). Moscow – Leningrad, p. 47 – 84. Quotation from p. 48.

10. La Grange J. L., 1892, *Oeuvres complètes*, t. 14. Paris, pp. 40 – 42.

11. Fuss P. N., 1843, Correspondance mathématique et physique de quelques celèbres géomètres du XVIII siècle, tt. 1 – 2. New York – London, 1968 note 20

12. Euler L., see Note 2.

12a. In Note 49 I have described eight of his memoirs devoted to probability theory and among them a hitherto forgotten memoir of 1780 in which he used the normal distribution and formally defined systematic errors.

13. See Note 5.

14. Bernoulli J., 1789, Milieu. *Enc. méthodique. Dict. Enc. des math.*, t. 2. Paris, pp. 404 – 409. Commentators including Todhunter (Note 7), §§ 825 and 598) believe that the author's name was Johann III.

15. Todhunter I., see Note 7.

16. Bernoulli Daniel, see Note 5.

17. Pingrè's work was published in the *Mém. Acad. Sci. Berlin* for 1761. Better known is
Short J., 1763, Second paper concerning the parallax of the sun. *Phil Trans. Roy. Soc. London 1665 – 1800 Abridged*, vol. 12, 1809, pp. 22 – 37.

18. Idelson N. I., 1947, *Metod naimenshikh kvadratov* etc. (Method of least squares etc.). Moscow. **S, G,** 58 (Chapter 1).

19. Sarhan A. E., Greenberg B. G., 1962. Certain symmetric distributions. In *Contributions to order statistics*, edited by them. New York – London, pp. 391 – 397.

20. Fuss P. N., 1843, *Correspondance mathématique et physique de quelques celèbres géomètres du XVIII siècle*, tt. 1 – 2. New York – London, 1968. See t. 2, pp. 674 – 677 (letter dated March 18, 1778).

21. Bopp K., 1924, Euler's und Lambert's Briefwechsel. *Abh. Preuss. Akad. Wiss.*, phys.-math. Kl., No. 2 (the whole issue). See Euler's letter dated 20 May 1760, pp. 15 – 17.

22. Sheynin O., 1966, Origin of the theory of errors. *Nature*, vol. 211, No. 5052, pp. 1003 – 1004.

23. Euler L., 1963, *Pisma k uchenym* (Letters to scholars). Moscow – Leningrad. The letters are published in their original languages with translations into Russian.

24. Stäckel P., 1910, J. A. Euler. *Vierteljahrschr. Naturforsch. Ges. Zürich*, Bd. 55, pp. 63 – 90.

25. Goussac A. A., 1961, La préhistoire et des débuts de la théorie de la représentation approximative des fonctions. *Istoriko-matematich. issledovania*, vol. 24, pp. 239 – 348. In Russian, title also in French.

26. Euler L., 1777 (1778), De projection geographica Delisliana in mappa generali Imperii Russici usitata (E 492). OO, ser. 1, t. 28. Turici, 1955, pp. 288 – 297. Russian translation, 1959.

27. Euler L., 1749. Recherches ssur la question des inégalités du mouvement de Saturne et de Jupiter (E 120). OO, ser. 2, t. 25. Turici, 1960, pp. 45 – 157.

28. Laplace P. S., 1798 – 1825, tt. 1 – 5, *Traité de mécanique céleste*, t. 2, livre 3, § 39. OC., t. 2. Paris, 1878. English translation by N. Bowditch: *Cel. Mech.*, vols. 1 – 4, 1829 – 1839. New York, 1966.

29. Sheynin O. B., see Note 22.

30. Maire [**C.**], **Boscovich** [**R. J.**], 1770. *Voyage astronomique et géographique dans l'Etat de l'Eglise*. Paris. See p. 501.

31. Euler L., see Note 27.

32. Eisenhart C., 1961. Boscovich and the combination of observations. In: *R. J. Boscovich. Studies of his life and work*. Editor, L. L. Whyte. London, pp. 200 – 213. Shorter version: *Actes du symposium intern. Boscovich 1961*. Beograd, 1962, pp. 19 – 25.

33. Sheynin O. B., 1967. On the history of the adjustment of indirect observations. *Izv. Vuzov. Geod. i aerofotos'emka* ser. No. 3, pp. 25 – 32. In Russian. **S, G,** 111.

34. Mayer [**T.**], 1748 (1750). Abhandlung über die Umwälzung des Mondes um seine Axe. *Kosmogr. Nachr. und Samml.*, pp. 52 – 183.

35. Cauchy A. L., 1863. Sur l'évaluation d'inconnues déterminées par un grand nombre d'équations. OC, sér. 1, t. 12. Paris, 1900, pp. 36 – 46.

36. Linnik Ju. W., 1961. *Methode der kleinsten Quadrate in moderner Darstellung.* Berlin. See § 14.5. Russian original: 1958. Unfit for geodesists.

37. Euler L., 1753 (1755). Eléments de la trigonométrie sphéroidique tirès de la méthode des plus grands et plus petits. (E 215). OO, ser. 1, t. 27. Turici, 1954, pp. 309 – 339.

38. Maire [**C.**], **Boscovich** [**R. J.**], See Note 30, § 305.

39. Todhunter I., 1873. *History of the mathematical theories of attraction and the figure of the earth*, vols. 1 – 2. London. [New York, 1962; Nabu Press, 2010.]

40. Gnedenko B. V., 1958. Über die Arbeiten Eulers zur Wahrscheinlichkeitstheorie, zur Theorie der Auswertung von Beobachtungen, zur Demographie und zum Versicherungs- wesen. See *Sammelband* mentioned in Note 1, pp. 184 – 208. In Russian. German summary on p. 209. Quotation from p. 196.

40. Todhunter I., See Note 7, chapter 11.

41. Sheynin O. B., 1970. Daniel Bernoulli on the normal law. *Biometrika*, vol. 57, pp. 199 – 202. Reprinted: **Kendall M. G., Plackett R. L.** 1977, *Studies in history of statistics and probability*, vol. 2. London, pp. 101 – 104.

42. Sheynin O. B., 1972. Daniel Bernoulli's work on probability. Reprinted Ibidem, pp. 105 – 132.
43. Du Pasquier L. G., 1909. Euler's Verdienste um das Versicherungswesen. *Viertel-jahrschr. Naturforsch. Ges. Zürich*, Bd. 54, pp. 217 – 243.
44. Gumbel E.-J., 1916. Eine Darstellung statistischer Reihen durch Euler. *Jahresber. Deutsch. Mathematiker-Vereinigung*, Bd. 25, No. 7 – 9, pp. 251 – 264.
45. Pajevsky V. V., 1935. Les travaux démographiques de Euler. *Recueil des articles et matériaux en commemoration du 150-anniversaire du jour de sa mort*. Moscow – Leningrad, pp. 103 – 110. In Russian. Titles of volume and contributions also in French.
46. Sofonea T., 1957. Euler und seine Schriften über die Versicherung. *Het verzerkeringsarchief*, Bd. 24(1). See Supplement to this journal called *Actuarieel Bijvoegsel*, pp. 87* – 104*.
47. Euler L., 1765. Von der Geschwindigkeit der Vermehrung und der Zeit der Verdoppelung [of the population]. Written by J. P. Süssmilch and him as chapter 8 of that year's edition of Süssmilch's *Göttliche Ordnung*. Euler compiled its mathematical considerations. See Euler, OO, ser. 1, t. 7, pp. 507 – 532.
48. Euler L., 1767. Recherches générales sur la mortalité et la multiplication du genre huimain. (E 334). OO, ser. 1, t. 7, pp. 79 – 100.
49. Sheynin O. B., 1971, Newton and the classical theory of probability. *Arch. hist. ex. sci*., vol. 7, pp. 217 – 243.
50. Sheynin O. B., 1971. Lambert's work on probability. Ibidem, pp. 244 – 256.
51. Euler L., 1768 – 1772. *Lettres à une princesse d'Allemagne*. (E 343, 344, 417). OO, ser. 3, t. 11. Turici, 1960, pp. 7 – 173, 177 – 312; t. 12, pp. 1 – 52, 55 – 265.

Some later contributions

Pearson K. (1978), *History of statistics in the 17th and 18th centuries* etc. Lectures 1921 – 1933. Editor, E. S. Pearson. London.
Sheynin O. B. (1973), Boscovich's work on probability. *Arch. hist. ex. sci.*, vol. 9, pp. 306 – 324.
--- (2007a), Euler's work in probability and statistics. In *Euler reconsidered. Tercentenary essays*. Editor, R. Baker. Heber City, Uta, pp. 281 – 316.
--- (2007b), True value of a measured constant and the theory of errors. *Historia Scientiarum*, vol. 17, pp. 38 – 48.
--- (2017), *Theory of probability. Historical essay*. Berlin. **S, G,** 10.

IV

Fechner as a statistician

Brit. J. Math. Stat. Psychology, vol. 57, 2004, pp. 53 – 72

I describe the work of Gustav Theodor Fechner (l80l – l887) related to probability and statistics and, in particular, to the treatment of observations. From a mathematical point of view his arguments were often insufficient, but his work proved to be highly fruitful, and I present the relevant comments of such scholars as Pearson and von Mises.

As the originator of psychophysics, Fechner opened up a new field for quantification. Subsequent workers rejected some of his findings, while at the same time acknowledged their debt to him.

1. Introduction

Previous commentators have described Fechner as the founder of psychophysics and as a philosopher (Heidelberger, 1987; Jaynes, 1971; Kuntze, 1892; Lasswitz, 1902). Some thought has also been given to his contribution to statistics, especially to his posthumously published work (Fechner, 1897), edited and supplemented by Gottlob Friedrich Lipps (1865 – 1951). Often, however, some of the new material there has been mistakenly attributed to Fechner[1].

It was Fechner s work in physics that obviously led him to quantify his psychophysical studies, and to base them on statistics. In that field of science, Fechner started by translating Biot's (1828 – 1829) treatise. Biot had not discussed the treatment of observations and neither did Fechner say anything relevant; this subject had still to catch the attention of physicists[2]. Later Fechner published several of his own physical contributions which show his great scholarship, and in 1846 Wilhelm Eduard Weber accordingly adapted his model of electric current (Archibald, 1994, p. 1214).

However, even in his *Atomenlehre* Fechner (1864) missed the opportunity to comment on the kinetic theory of gases then being developed by Clausius and Maxwell. Moreover, he repeatedly treated physics on a par with (practical) astronomy by stating that both these branches of natural sciences had to do with symmetric distributions and true values of the magnitudes sought (1874b, pp. 7 and 9; 1897, p. 15); cf. the beginning of § 4. He certainly came to recognize the treatment of observations in physics, but did not go any further.

Fechner's style is troublesome. Very often his sentences occupy eight lines, and sometimes much more, sentences of up to 16 lines are easy to find[3]. On the other hand, he made a nice pronouncement by which he (1877, p. 215)[4] estimated his own work in psychophysics:

The Tower of Babel was not completed because the workers were *unable to explain to each other how should they build it My psychophysical structure will probably survive because the workers cannot see how they might demolish it*.

In this paper I describe Fechner's attempts at constructing a theory for treating mass observations in natural sciences (the introduction of a random variable as an object of study, the choice of estimators, the description of asymmetric observational series, and a measure of dependence between observations) and give an appraisal of his work.

However, I begin with a sketch of the history of the Weber – Fechner law and of Fechner' s experimental work.

2. Psychophysics and the Weber – Fechner law

2.1. Psychophysics. Fechner is acknowledged as the father of experimental psychology (Boring, 1950, Chapter 14; Singer, 1979, pp. 6 – 7). Galton, in a letter of 1875 (Pearson, 1930, p. 464), praised him for having laid, in his *Elemente* (Fechner 1860), *the foundation of a new science* (psychophysics), and he continued:

A mass of work by Arago, Herschel, and various astronomers falls in as a part of the wide generalizations of Fechner, and much criticism and recognition of him will be found in Helmholtz.

Galton apparently thought about experimentation and his opinion is of course noteworthy even though he did not provide any references. Concerning the astronomers, Galton undoubtedly bore in mind Fechner (1859), see § 4.4.1.

Fechner (1860, Bd. 1, p. 8; 1877, p. 213) defined psychophysics as an

Exact doctrine on the functional correspondence or interdependence of body and soul[5].

He distinguished between external psychophysics, which has to do with physics and can be studied by the relations between stimuli and sensations, and internal psychophysics, which is concerned with the not directly observable work of the nervous system (1860, Bd. 1, pp. 11 and 57; 1877, p. 12)[6].

Internal psychophysics is beyond the scope of this paper. Indeed, I doubt whether present-day psychophysicists recognize it, and in any case Fechner himself certainly did not describe it quantitatively. For that matter, his writings related to this subject abound with natural-scientific accounts lacking mathematical support. This is strange since he left some mathematical thoughts, for example, on oscillating stimuli (Fechner, 1860, Bd. 2, Chapter 32), to say nothing about the subjects of my §§ 4 and 6.

Fechner (1859, p. 490) found *the first fundamental law conforming to experience*[7] concerning psychophysics in studies of the sensation of light. His main subject there was the relation between star magnitudes and their luminosities, and on p. 491 he effectively stated that his *Elemente* (1860) will generalize this issue to other sensations.

2.2. The Weber – Fechner law. Fechner (1859, p. 531; 1860, Bd. 1, p 64; 1877, p. 8) attributed to Ernst Heinrich Weber an independent study *in some generality* of the connection between stimuli (x) and sensations (y),

$$y = C\log x. \qquad (1)$$

Consequently, Fechner named it after Weber who had first considered this issue in 1834 and later somewhat enlarged on his thoughts (Boring, 1950, p. 113). In 1851 Weber described the possibility of distinguishing between the weights of two objects, the lengths of two segments, and the pitch of two tones. In the first case, for example, differentiation was generally possible when the weights were in the ratio of 39**:**40, and his final conclusion (Weber, 1905, pp. 117 – 118) was that

The ability to perceive the relation between magnitudes themselves, without either measuring them in smaller units or finding out the absolute difference between them, is an extremely interesting psychological phenomenon[8].

For small values of Δx, the law (1) leads to $\Delta y \approx \Delta x/x$ and Δy becomes perceptible when $|\Delta x|/x$ exceeds some threshold value. The converse transition from Weber to Fechner is, however, methodologically difficult. Furthermore, commentators agree that Fechner had correctly regarded the law (1) as a much more general regularity than did Weber. Thus, Galton (1879, p. 366) simply called it *Fechner's law,* and Spearman (1937, vol. 2, p. 157) stated that *Experimental psychology must be credited with the logarithmic law of Fechner*.

Again, it was Fechner who carried out numerous related experiments (see § 3) and revealed that the issue was much more complicated than it had seemed at first sight.

Concerning this latter point, Fechner (1860, Bd. 1, p. 17 and Bd. 2, pp. 39 – 41) discussed the case of *negative* sensations and paid due attention to the phenomenon of threshold values of x, or more precisely, of such values of x that led to non-zero values of y as well as of such Δx that produce non-zero values of Δy (Fechner 1877, pp. 10 – 11 and 238 – 242); see also § 3.2.

Finally, Fechner (1877, pp. 211 – 212; 1882, p. 419; 1887a) spent much effort on ascertaining the unresolved issue of the limits of applicability of the law, of how different were the sensations of light and sound (Fechner 1860, Bd. 2, p. 267), etc. This fact reflects the continual debates that were going on about the various aspects of the nascent psychophysics. Large sections of some of Fechner's contributions, and especially of his book (1877), were indeed devoted to the discussion of the arguments of other researchers.

3. Experimentation

3.l. General information. It is difficult to imagine how many different experiments Fechner carried out. He listed seven circumstances that might have influenced his study of the sensation of weight (Fechner, 1860, Bd. 1, pp. 80 – 81). In general, he mentioned the need to examine several factors such as fatigue, in particular versus experience (1860, Bd. 1, pp. 80 and 82; 1861b, § 10[9]; 1882, p. 377; 1887b, § 294) and attentiveness (1860, Bd. 1, p. 82; 1861b). When studying eyesight, he attempted to reveal the differences between binocular and monocular vision (Fechner, 1859, p. 458; 1861b, § 9). It is also noteworthy that he claimed that a blind experiment was not more expedient than its counterpart (Fechner, 1860, Bd. 1, p. 119).

Fechner (1860, Bd. 1, p. 85) paid much attention to the correct recording of observations and to the checking of the ensuing calculations. He (Ibidem) stated that the rejection of

Unusual observational values has neither any underlying principle nor boundaries and leads to arbitrariness[10].

Fechner (1860, Bd. 1, pp. 79 – 84; 1887b, pp. 288 – 292) also formulated some general recommendations: an experiment in accordance with a prearranged plan, changes the influencing factors methodically[11], and avoids a formal combination of results obtained by different researchers (1887a; 1887b, p. 218).

Fechner (1860, Bd. 1, pp. 88 – 93 and 112 – 115; Bd. 2, pp. 122 and 134; 1887b, pp. 292 – 295; 1882, p. 359) repeatedly discussed constant errors of observation and their elimination (cf. § 3.2), though he never mentioned the general case of systematic errors with non-zero expectations. It is worth noting his relevant but vague remark that the *mean* values of *irregular chance magnitudes*[12] should remain constant (1860, Bd. 1, p. 77), and his curious statement that observations in physics and astronomy might be not as precise as elsewhere (Ibidem, p. 78). He enthusiastically continued:
the law of large numbers *rules* over randomness *so far as it accumulates*.

3.2. Special methods. Fechner (1860, Bd. 1, pp. 71 – 75; 1882) applied and developed three previously known methods for measuring threshold or near-threshold values of stimuli and discussed them in the context of his experiments on lifting weights.

The method of scarcely perceptible differences. Here, it was required to estimate the least perceptible difference (ΔP) between two weights. Fechner reasonably recommended to approach the unknown threshold both from below (beginning with differences that were too small) and from above.

The method of right and wrong cases. The problem was, in estimating such a ΔP, that, in differing circumstances, the ratio of right and wrong decisions on which of the two weights was heavier, remained constant. This ΔP was evidently larger than its threshold value.

Assisted by A. F. Möbius, best remembered for the Möbius strip but not known for any studies in probability theory, Fechner (1860, Bd. 1, pp. 104 – 107 and 112 – 115) took stochastic considerations into account which allowed him to estimate some constant influences inherent in lifting weights. His reasoning nevertheless left room for some doubts (Stigler, 1986, pp. 246 – 249). However, because of its simplicity, Möbius' interpretation of Fechner's problem deserves mention. In essence, he considered the error in deciding, without measurement, which of the two given segments was longer and he assumed that the error Δs in evaluating the length s of a segment was (up to some value of $|\Delta s|$) normally distributed. Fechner made a similar (and less obvious) assumption concerning his own experiments.

Nowadays, the method of right and wrong cases, regarded in Fechner's stochastic sense, is called the method of paired comparisons, and is a special case of incomplete ranking, in which the observer expresses a preference for one of the two objects (stimuli) under judgement (David, 1988, pp. 11 – 13).

The method of average error. Here, the changeable weight P_2 had to be made equal[14] to a given weight P_1. Each time coincidence was reported, the appropriate value of P_2 was recorded together with $|\Delta P|$ and, eventually, the average $|\Delta P|$, which was probably smaller than its threshold value, was calculated.

4. Theory of errors

Fechner used the error-theoretic term *true value* of the constant sought (§ 4.2), which is my justification for the title of this section. Beginning with Fourier, this theory equates true value to the limit of the arithmetic mean as the number of observations increases indefinitely (Sheynin 1996b, p. 118; 2007). Practically speaking, von Mises (1931,p. 370) assumed the same definition.

Fechner hardly distinguished between a sample estimator and its expected value, and thus I do not stress this difference either.

4.l. The choice of means. Let Δ_k be the deviation of the k-th observation from the mean. Fechner chose a mean by imposing some condition on these Δ_k's. He (Fechner 1874b, p. 4) certainly knew that, with respect to the arithmetic mean,

$$\Sigma\Delta_k^2 = \min,$$

and he (p. 29) went on to determine the mean for which

$$\Sigma\left|\Delta_k\right| = \min^{15}.$$

Fechner (1874b, pp. 40 ff.) took up a more general problem of determining the mean M_n according to the conditions

$$\Sigma\left|\Delta_k\right|^3,\ \Sigma\left|\Delta_k\right|^4,...,\ \Sigma\left|\Delta_k\right|^n = \min,$$

and noted that it involved difficult algebraic work.

Then he (p. 57) remarked that the choice of M_n depended on the appropriate density function $\varphi_n(x)$, and, without theoretical proof or proper empirical justification, assumed (p. 64) that

$$\varphi_n(x) = C_n b\exp(-\,b^{n+1}|x|^{n+1}),\ b^{n+1} = \frac{1}{(n+1)\mathrm{E}\,|\,x\,|^{n+1}}, \qquad (2)$$

where E denotes mean value and C_n is a constant.

Fechner (p. 54) also offered some not quite accurate remarks about Gauss's choice of $\Sigma\Delta_k^2$ as a measure of precision, as well as a comment on Laplace. When estimating precision, the latter, Fechner stated, had made use of $|\Delta_k|$, and he should therefore have recorded these deviations with respect to the median rather than to the arithmetic mean.

Here is a specimen of Laplace's reasoning (Sheynin, 1977, § 5.1). Given, a system of *m* equations in one unknown, *z*,

$$p_i z - s_i = e_i$$

with unknown errors e_i. Adding these equations pre-multiplied by some integers q_i he obtained

$$z = \frac{[sq]}{[pq]} + \frac{[eq]}{[pq]} = \frac{[sq]}{[pq]} + z',$$

where, in general, in Gauss's notation,

$$[ab] = a_1b_1 + a_2b_2 + \ldots + a_mb_m.$$

Non-rigorously proving an appropriate version of the central limit theorem, Laplace derived a normal distribution $\varphi(z')$ for $[eq]$, demanded that

$$\int_{-\infty}^{\infty} |z'|\varphi(z')dz' = \min \tag{3}$$

and calculated the corresponding (optimal) values of the multipliers q_i. A generalization of this problem to two unknowns led Laplace to the MLSq.

Condition (3) was not therefore connected with the median, but the main weakness of the Laplacian approach was the assumption of the requirements for the central limit theorem. Then, Fechner's idea (see also 1874b, p. 53), that the precision of observations be measured by a statistic whose choice depended on the selection of the mean, contradicted the Gaussian and the Laplacian attitude of keeping to one universal estimator, the variance or the absolute expectation, respectively[16].

In addition to the arithmetic mean (A) and the median (C) Fechner (1874b, p. 11) introduced the *most dense* value (D) whose probability was maximal (p. 12)[17]. Beyond the error theory it was more important than A (p. 13); indeed, Fechner applied it when introducing his double-sided Gaussian law (see § 7 below). He also stated that all the observations were equally plausible, so that A should not be singled out (Fechner, 1897, p. 16). He thus had not grasped Gauss's mature justification of least squares according to which the arithmetic mean had maximal weight under general conditions.

4.2. Estimating precision. Failing to appreciate the notion of unbiassedness, Fechner (1860, Bd. 1, p. 125) stated that, for one unknown, the celebrated Gauss formula for the sample variance of an observation in the case of *m* observations, $\mathrm{E}\Delta_k^2/(m-1)$, unlike its predecessor (with *m* rather $(m - 1)$ in the denominator) allowed for the finiteness of observations.

Accordingly, he attempted to correct the statistic

$$\varepsilon = \Sigma|\Delta_k| : m \qquad (4)$$

where the Δ's were calculated from A in a similar way. He indicated the *correct* formula on p. 126 and went on to justify it in Bd. 2, pp. 368 – 372, as follows.

Denote the true value of the magnitude sought by V. Let V – A = a and $x_k - V = \delta_k$. Then the error caused in A by one observation, x_k, will be $|\delta_k|/m$ and

$$a = \frac{\Sigma|\delta_k|}{m^{3/2}}.$$

Fechner then calculated the (expected) number of observations smaller and larger than A and V corresponding to the normal law whose measure of precision was determined by (4). Then he corrected each (expected) x_k accordingly, and found that

$$\varepsilon_1 = \frac{\pi n}{\pi m - 1}\varepsilon \qquad (5)$$

should be taken instead of ε.

In his next contribution Fechner (1861a, p. 57) repeated his desire to correct ε and admitted that his earlier investigation was not good enough; see also Fechner (1877, p. 216), where he repeated this admission. Elsewhere Fechner(1874a, p. 74) noted that, according to the Gaussian approach, each Δ_k should become

$$\Delta_k \sqrt{m/(m-1)}$$

and the correct formula for ε should therefore be

$$\frac{\Sigma|\Delta_k|}{\sqrt{m(m-1)}}. \qquad (6)$$

rather than (5). Finally, Fechner (1897, pp 20 – 21) stated that he had empirically justified expression (6).

Now, (6) coincided with the famous Peters formula which its author substantiated in 1856, albeit only for the normal distribution. In 1875 Helmert considered it anew (Sheynin, 1995, § 5) because Peters had tacitly and wrongly assumed that the Δ's were mutually independent, but the formula persisted.

Fechner (1874a, p. 66) continued his investigation by starting from the formula for the probable error of the arithmetic mean, which has $m(m - 1)$ rather than m in its denominator as in the very beginning of this section, and multiplied the derived expression by 0.675, denote the result by (7). That coefficient meant that Fechner had assumed the normal law.

He then set

$$\Sigma\Delta_k^2 = \pi\frac{\Sigma|\Delta_k|^2}{2m} \quad (8)$$

and called it a generally known relation (see Gauss, 1880, § 5, a relation for the most probable Δ's and the normal law) and arrived at formula (6) with coefficient 0.845, denote the result by (9). For m= 2 this result was unacceptable.

Fechner (pp. 70 – 77) discovered that the mistake was due to the unaccounted for differences between $E(\Sigma|\Delta_k|)^2, (E\Sigma|\Delta_k|)^2$ and $E\Sigma\Delta_k^2$.

Instead of (8) and (9), he derived

$$\Sigma\Delta_k^2 = \pi\frac{m-1}{m(2m+\pi-4)}(\Sigma|\Delta_k|)^2,$$

$$w = 0.675\sqrt{\frac{\pi}{2m+\pi-4}}\frac{\Sigma|\Delta_k|}{m}$$

respectively where the symbols of expectations were lacking. Helmert, in 1876, and then Fisher improved on the second formula (Sheynin, 1995, § 10). Fechner (p. 66n) called his paper a preliminary extract from another contribution which apparently remained unpublished.

4.3. Correcting observational readings. When reading an instrument scale, the position of a point situated on an interval of Width i between two consecutive graduated points has to be estimated, and this estimate is necessarily rounded off. Consequently, Fechner distinguished two sources of error. He touched on this issue (1860, Bd. 1, p. 127) and returned to discuss it somewhat later (Bd. 2, pp. 373 – 376). The mistakes made because of the second cause will not compensate one another because, as Fechner indicated, the errors were not uniformly distributed over any interval.

Assuming a normal law with a measure of precision determined by formula (5)[18] Fechner calculated the correction for different values of the ratio i/ε_1. Elsewhere he noted that the causes of the error in estimating the position of a point are both objective and subjective and recommended, when large mistakes were possible, to abandon the estimation altogether, and record either endpoint of the interval. He then went on to calculate the corrections to $\Sigma|\Delta_k|$ and $\Sigma\Delta^2_k$, again on the strength of the normal distribution, should the estimation be done (Fechner, 1861a, pp. 71, 93 – 105, 108 – 1l3).

He (1897, pp. 10 – 11 and 142 – 143) returned to the same subject once more, but added little of importance. Nor did he cite his previous work (1877, p. 217) where he had decided that his first correction, even as revised (1861a, pp. 93 – 105), was insufficiently general and recommended to abandon it altogether (and to neglect only sufficiently small intervals). It appears that in 1843, Cournot (1984,

§§ 139 – 140) was the first to turn his attention to this subject, but he had not even mentioned the vernier. In any case, Fechner's efforts show that he attempted to make the most of his data.

4.4. Treating observations

The method of least squares. When studying the relation between the magnitudes of the stars (G) and their luminosities (i) (see § 2.1), Fechner (1859, pp. 508 – 509) reprinted John Herschel's data on 60 stars[19] and determined the two constants, k and c, in his own formula,

$$G = -k\log i + c, \; k > 0, \qquad (10)$$

by least squares without, however, providing the calculations[20].

He also checked his formula as follows (pp. 505 – 506). He combined the 60 equations arranged in increasing values of G into six equal groups. Each group provided an arithmetic mean of the appropriate star magnitudes and a geometric mean of the i's, and Fechner calculated the (six values of) G and compared them with their (mean) observed values; he then calculated the i's from the (mean) values of G and compared them with their (mean) observed values. In both cases the coincidence of the observed and the calculated values was *striking*, and the signs of the differences, again in both cases, constituted a reasonable sequence +, +, –, –, +, –. Fechner did not have to assume the normal distribution here (see the next paragraph), but his second calculation was hardly necessary.

Comparing two competing rules Herschel had proposed another relation,

$$(G + \sqrt{2} - 1)^2 i = 1,$$

and Fechner (p. 510) compared the two formulas by means of the residuals ΔG_j, $j = 1, 2, ..., 60$. Both sums, $\Sigma|\Delta G_j|$ and $\Sigma \Delta G^2_j$, were smaller for the Fechner relation (10)[21], though he did not determine the mean square error of k or c.

He then calculated the expressions

$$P = \frac{2m\Sigma\Delta G_j^2}{(\Sigma|\Delta G_j|)^2},$$

again for the two cases, and noted that his relation provided a number closer to π; cf. formula (8). He remarked here that (10) furnished a better approximation because π would have appeared *under a normal distribution of errors*[22] *which presupposes the true observed magnitudes as the starting point of the errors*[23].

Lastly, Fechner noted that

$$Q = \frac{\Sigma|\Delta G_j|}{\Sigma'|\Delta G_j|}$$

with the Σ’ which extended over the terms exceeding the appropriate mean square deviation, was, in his case, closer to √*e*, which meant a better fit[24]. He had not substantiated this reasoning either here, or elsewhere (1860, Bd. 2, p. 360) where he reiterated that *Q* should be equal to √*e*. Then, however, Fechner (p. 371n), without citing these considerations, all but proved them for the normal distribution. I repeat and conclude his calculations. Introducing a constant *c*, we have, for errors *x*,

$$\Sigma|x| = \frac{2cb}{\sqrt{\pi}}\int_0^\infty x\exp(-h^2x^2)dx = \frac{c}{b\sqrt{\pi}},$$

$$\Sigma'|x| = \frac{2cb}{\sqrt{\pi}}\int_0^\alpha x\exp(-h^2x^2)dx,$$

and, for $\alpha = 1/b\sqrt{2}$

$$\Sigma'|x| \ = \frac{c}{b\sqrt{e\pi}} \text{ etc.}$$

Combining equations. In passing, Fechner (1860, Bd. 1, pp. 224 – 225) mentioned the possibility of solving systems of equations in two unknows by arranging them in pairs, solving each pair and calculating the appropriate mean values. Later (1887b, pp. 214ff.), he returned to this method for dealing with equations in two unknowns, *b* and *k*,

$ib + k = t_i$, $i = 1, 2, 3, 4$,

because the *most plausible* MLSq required more calculations and was fraught with mistakes. In addition, as Fechner argued, the method of combinations provided a check.

This method had been used as far back as the 18th century, and C. G. J. Jacobi and Binet independently proved that the least-squares solution was identical to some weighted mean of the partial solutions provided by combining the equations (Sheynin, 1995, pp. 44 – 46). This connection between the two methods apparently contradicts Fechner’s (1887b, p. 217) unsubstantiated remark that, as $i \to \infty$, their results *in principle* coincide. He also stated (p. 218), that for small values of *i* even the MLSq was not good enough[25].

5. The collective and random variables

Fechner (1874b, p. 3) introduced the collective (*Kollektiv-gegenstand*), a very large number of randomly varying objects of the same type[26]. Repeating this formula, he (1897, p. 5) added that the objects were distributed in accord with *general probabilistic laws of chance*[27], which do exist, as *every mathematician knows*[28], and noted that various branches of the natural sciences provided appropriate examples; see also Fechner (1874b, pp. 8 – 9).

Thus, the notion of a random variable had appeared on a natural-scientific level. It was effectively used at least from the 17th century onwards (winnings in lotteries); then came life tables (Graunt, in 1662) and the theory of errors (Simpson, in 1756 and 1757), and in 1829 Poisson introduced this concept formally, although calling it by a purely provisional term (Sheynin 1978, pp. 250 and 290).

Elsewhere I have argued that from Chebyshev's time until about the 1930s mathematicians developed the theory of probability by ever more fully using the power of the concept of the random variable (Sheynin 1998, p. 103).

Fechner (1897, pp. 5 – 6) also invented the term *Kollektivmasslehre* (which became the title of his book), whose *most important* problem was (the study of) frequency distributions of the appropriate objects (p. 4). He believed that an attempt to consider randomness from a philosophical standpoint would bear little fruit, remarking that the random variation of the objects was neither arbitrary nor regular (p. 6)[29].

Denote the observed values in a series (in a collective) by

$$u_1, u_2, \ldots, u_\mu, v_1, v_2, \ldots, v_\nu, u_1 \le u_2 \le \ldots, \le u_\mu \le H \le v_1 \le v_2 \le \ldots, \le v_\nu,$$

where H is some chosen mean. For convenience, write x_k, $k = 1, 2, \ldots, \mu + \nu = m$ instead of u_i and v_j, and Δ_k instead of $(u_i - H)$ or $(v_j - H)$. Fechner (1897, pp. 84 – 85) attempted a (vague) general description of the collective[30] by (u_i), (v_j), by several (at least three) means and their position relative to one another; and by the deviations $|u_i - H|$, $(v_j - H)$, $\Sigma|u_i - H|/\mu$ and $\Sigma(v_j - H)/\nu$. In general, he paid much attention to calculating the two last-mentioned sums without specifying the appropriate density function (1874b, pp. 24 – 37; 1897, pp. 154ff.). And he attempted to discover a unique empirical distribution of the observational values at least for most asymmetric collectives.

By the mid-l9th century the significance of asymmetric distributions began to be recognized (Sheynin, 1984, § 4.3; 1986, § 5.4). In 1845, Auguste Bravais provided appropriate examples from biology, meteorology and even practical astronomy, and in 1846 Quetelet used such distributions to describe atmospheric pressure. Although he then abandoned this approach, his curves describing inclination to crime (in 1869) were again asymmetric. Towards the end of the century, in 1891, Hugo Meyer declared that the theory of errors could not be applied in meteorology because of the asymmetry of meteorological densities but Pearson (1898) used Meyer's data for illustrating his theory of asymmetric curves. Fechner (1874a, p. 9; 1897, p. 16) also insisted that asymmetry was the rule rather than the exception.

It is also worth noting that Fechner (1897, pp. 6 – 7) asserted, again quite reasonably, that the study of a collective should begin with the compilation of an initial list (*Urliste*) of observations, and then of a table of the empirical distribution (*Verteilungstafel*).

6. Asymmetric collectives

Fechner attempted to study the asymmetry of collectives by the. relative positions of A, C, and D, the arithmetic mean, the median, and the most dense (maximum likelihood) estimator. He (1874b, pp. 11 – 13) argued that collectives were generally asymmetric with symmetry being possible only with respect to D (and to A, if A = D), when some simple equalities involving the deviations were fulfilled; for example, the case $\mu = \nu$ and

$$\frac{\Sigma\,|\,u_i - D\,|}{\mu} = \frac{\Sigma(v_j - D)}{\upsilon}$$

corresponded to *absolute symmetry*. He returned to this issue on p. 32 but did not propose a definite measure of asymmetry. One such measure is (Yule & Kendall, 1958, p. 161)

$$\text{skewness of distribution} = \frac{3(\text{mean -- median})}{\text{standartdeviation}}.$$

Skewness vanishes if and only if A = C.

Then Fechner (1897) returned to the issue of asymmetry. He (p. 66) stated that it occurred (this time, with respect to A) when $\mu \neq \nu$. However, he also formulated *special laws* of asymmetry. Among these he mentioned the double-sided Gaussian law, two differing laws governing, respectively, the observational subseries (u_i) and (v_j) which transformed themselves into each other at D (p. 70), and the *laws* describing the relative positions of A, C and D (pp. 71 – 72). He stated without proof that, for small values of |C —D| as compared with $\Sigma|u_i - \text{D}|/\mu$ and $\Sigma(v_j - \text{D})/\nu$,

$$\frac{\text{C - D}}{\text{A - D}} = \frac{\pi}{4},\ \frac{\text{A - C}}{\text{A - D}} = \frac{4 - \pi}{4} = 0.215. \qquad (11a, 11b)$$

Yule and Kendall (1958, p. 117) remarked that the relation

$$\text{D} = \text{A} - 3(\text{A} - \text{C}) \qquad (12)$$

holds *with a surprising closeness for moderately asymmetric distributions*, and it follows that the result in (11b) should be 0.333.

Another of Fechner's *laws*, also apparent from (12), was that the three means fell in the order D, C, A (or A, C, D). He concluded (pp. 81 – 82) that, for ever more asymmetric collectives, the proper distributions were the Gaussian law, the double-sided Gaussian law, and the same double-sided law with all the observations initially replaced by their logarithms. He did not justify the use of the lognormal law, which indeed describes a strongly asymmetric frequency and has since proved its worth in various branches of science.

Fechner (1897) several times (e. g. on p. 204) distinguished between essential and random asymmetry, the latter occasioned by an insufficient size of the collective, but he was not really able to provide an appropriate criterion[31]. It is true that on pp. 205 – 205 he noted that with the increase in *m* the real component of asymmetry became ever more pronounced as compared with its random component. He

(p. 198) also declared, however, that stochastic formulas were useless in this case.

Nevertheless, Fechner (pp. 206 – 209) attempted to make use of them. Count μ and ν with respect to A, write μ – ν = α and suppose that for a given *m* this difference was recorded *n* times. His most interesting formula here was

$$\frac{\Sigma|\alpha|}{n}=\sqrt{\frac{2}{\pi}(m\pm 0.5}. \qquad (13a)$$

Lipps (Fechner, 1897, pp. 212 – 214) improved this reasoning. With the ratio μ/ν as his starting point, he assumed that the appropriate probabilities of the positive and negative deviations, *p* and *q*, were in the same ratio to each other, but he also supposed that μ and ν were recorded with respect to D. Then, instead of (13a), he wrote out the generalized expression

$$\frac{\Sigma|\alpha|}{n}=\sqrt{\frac{2}{\pi}4pqm}. \qquad (I3b)$$

Both Fechner and Lipps thus started from the formula for the variance of the frequency of *successes* in *n* Bernoulli trials with $p = q$ and $p \neq q$, respectively, and relation (8). The underlying pattern was therefore complicated: the observed values were either less than or greater than A (more precisely, D) in accordance with a binomial distribution, but their exact position in each of the two intervals (on either side of D) was apparently governed by the appropriate Gaussian law.

Formula (13a) or (13b) described the essential component of the asymmetry, and both authors obviously believed that, if the asymmetry of a collective was not sufficiently corroborated, a further increase in *m* was necessary.

7. The double-sided Gaussian law and the mode

Lipps (Fechner, 1897, p. 295) wrote out the magnitudes

$$f_i(u_i) = \frac{2\mu^2}{\pi\Sigma|u_i-D|}\exp\left(\frac{-\mu^2(u_i-D)^2}{\pi(\Sigma|u_i-D|)^2}\right),$$

$$g_j(\upsilon_j) = \frac{2\upsilon^2}{\pi\Sigma(\upsilon_j-D)}\exp\left(\frac{-\upsilon^2(\upsilon_j-D)^2}{\pi[\Sigma(\upsilon_j-D)]^2}\right),$$

that is, the *distribution of the numbers* (the expected numbers) of deviations $|u_i$ – D| and (υ_j – D) corresponding to the double-sided Gaussian law with common point D and measures of precision

$$\frac{\mu}{\sqrt{\pi\Sigma|u_i-D|}} \text{ and } \frac{\upsilon}{\sqrt{\pi\Sigma(\upsilon_j-D)}}$$

respectively. For $u_\mu = D = v_1$ we have $f_\mu = g_1$ so that at D the laws coincide, and

$$\frac{\mu^2}{\Sigma|u_i - D|} = \frac{\upsilon^2}{\Sigma(\upsilon_j - D)} \text{ or } \frac{\mu}{\upsilon} = \frac{\Sigma|u_i - D|/\mu}{\Sigma(\upsilon_j - D/\upsilon} \qquad (14)$$

was Fechner's main condition for determining D (p. 296).

In an elementary way Lipps (Fechner, 1897, p. 305) was then able to derive relations (11) for *very weak* asymmetry of the collective.

8. Dependent observations

Many scholars (John Dalton, in 1795; Lamarck, ca. 1805; Quetelet, in 1852; Wladimir Köppen, in 1872) knew that the weather depended on its previous states, and Quetelet was the first to use elements of the theory of runs to study this phenomenon (Sheynin, 1984, § 5). For Fechner, meteorology provided examples of his collectives, and he studied Quetelet's data on daily air temperatures, comparing them with the results of a reputable German lottery (Fechner, 1897, pp. 45 – 47, 365 – 366). Each day, during which the temperature was higher than a certain mean, he entered as a plus, and otherwise as a minus. On the other hand, he arranged the lucky lottery tickets in chronological order of their being drawn. The tickets were numbered so that the signs of the differences between these numbers on consecutive tickets could have been recorded.

Fechner (p. 366)[32] noted that, for a large number of tickets (m), there were about twice as many changes (w) of sign as runs (f) with $f + w = m - 2$. The extreme cases of independence and complete dependence thus corresponded to

$f = m/3$ and $f = m$,

respectively, and Fechner introduced a measure of dependence (*Abhängigkeit*)

$$\text{Abh} = (3f - m)/2m \qquad (15)$$

with Abh = 0 and 1 for the above-mentioned cases. He did not offer any quantitative estimate for the dependence between consecutive air temperatures. However, it is now known (Moore, 1978) that for random *runs up and down*

$$\text{E}f = (2m - 1)/3$$

which agrees with Fechner's estimate above.

9. Discussion

I begin with psychophysics and supplement Fechner's reference (§ 2.1, Note 6) to Newton's *Optics* (Part 1 of Book 3, qu. 14 and 15). Elsewhere, Newton (1934, p. 544; 1950, p. 49) attributed eyesight to Providence. With respect to Euler (same Note), see Helmholtz (1913, pp. 375 – 379). Euler had at least felt the need to connect sensation with stimulation.

One issue that Fechner did not follow up psychophysically was the capability of estimating the position of a point on an interval (cf. § 4.3). Another related subject is connected with the frequency of making computational errors (blunders). This is now being investigated when training astronauts. And in the context of internal psychophysics Fechner could have discussed the personal equation, that is, the difference between the moments of the passage of a star through the crosshairs of an astronomical instrument as recorded by two observers. This phenomenon was discovered by Bessel in 1823; see Sheynin (2000, § 2), where Bessel's mistake is indicated.

Ebbinghaus (1911, p. 67) noted that [for psychophysics] the discovery of the personal equation was a *lucky chance* that led to an *entire class* of investigations. He praised Fechner's study of the sensation of weight but remarked that its results were corrupted by subjective factors (p. 399). He also noted (pp. 604 and 606) that Fechner had overestimated the applicability of the law (1), and claimed (pp. 617 – 618) that his predecessor had wrongly interpreted negative sensations.

Another commentator (Cowles, 1989, p. 29) bluntly declared that Fechner's

Basic assumptions and the conclusions he drew from his experimental investigations have been shown to be faulty,

but

The revolutionary nature of Fechner's methods profoundly influenced experimental psychology.

It is worth mentioning Edgeworth's (1996, p. 563) reference to Fechner's *classical experiments on the accuracy of the senses*. However, Ebbinghaus (1911, pp. 85 – 87), who referred to a previous author, found fault with (or at least shortcomings in) Fechner's version of the method of right and wrong cases.

Elsewhere Ebbinghaus (1908, p. 11)[33] called Fechner a *philosopher full of fantasies* and a *most strict physicist* who had

Put … together psychophysics as a new branch of knowledge.

Freud (1961, p. 541) also talks about *the great* (*grosse*) *Fechner* and (1963a, p. 4) described him as *an insightful* [*tiefblickender*] *researcher*. He (1963b, p. 86, as quoted by Misiak & Sexton 1966, p. 387) further says[34]:

I was always open to the ideas of Fechner and have followed that thinker upon many important points.

During the 19th century, many new scientific disciplines based on, or intrinsically connected to, statistics emerged, for example climatology, epidemiology, biometry and the kinetic theory of gases. Psychophysics, as developed by Fechner, would also have been impossible without statistics, so that he ranks in this context alongside such figures as Humboldt, Pearson, Maxwell and Boltzmann.

Before going on to statistics, I discuss a marginal subject, experimentation. So as to put factorial analysis in perspective, the history of the determinate error theory should be considered. When measuring angles in the field, two factors were dealt with simultaneously: the order in which sightings were taken at adjacent stations, and the position of the sighting telescope relative to the vertical circle of the theodolite (Bessel, 1838, § 15). Then, in a pure physical (not psychophysical) way, J. Ch. Borda and later Gauss had studied the measurement of the difference of two roughly equal weights, and Helmert described their work in 1872. Pukelsheim (1993, p. 427) connected ‚Gauss' ideas with modern concepts in the design of experiments, but neither he nor Helmert cited one of the original sources, Gauss' letters to H. C. Schumacher of 1836 and 1839; see Sheynin (1996b, p. 149). Both Borda and Gauss were apparently able to eliminate largely the influence of two factors at once, but the latter s pattern of experimentation was more expedient[35].

Fechner' s study of the precision of observations contained an innovation (§ 4.2) that properly belonged to the theory of errors. Furthermore, his use of the variance of the number of *successes* in Bernoulli trials (§ 6) should be noted: although this estimator had entered the De Moivre – Laplace limit theorem, its more or less direct application beyond the error theory began with Lexis, and even he did not emphasize this point (Lexis, 1903, § 6). In short, Fechner was undoubtedly one of a very few natural scientists who furthered the theory of errors.

Most interesting among Fechner's other findings were the double-sided Gaussian law and the lognormal distribution (§§ 6 and 7), but they were neither original to him nor general enough, since they could have described only a portion of asymmetric laws. This was pointed out by Ranke and Greiner (1904) and then, much more forcefully, by Pearson (1905). Galton (1879) and McAlister, in a companion paper, had introduced the lognormal distribution, and De Vries, in 1894, had applied the double-sided law. Pearson (p. 196) also alleged that *every one* of Gauss's three assumptions which led him in 1809 to the normal law is *negatived when the double Gaussian curve is used*, so that Fechner's reasoning was illogical; and he correctly stated that Fechner had determined the mode (the common origin of both of the one-sided curves) by a *rough process* much inferior to his (Pearson's) method of moments. His first point seems meaningless: Fechner had abandoned the arithmetic mean and only justified the double-sided law empirically.

Asymmetrical series of observations were known about at least from 1845 (§ 5), and Fechner insisted that asymmetry was the rule in the natural sciences.

Although in metrology the situation was different, it is noteworthy that Dmitri Mendeleev, the eminent chemist and metrologist, called a series of observations harmonious, if, in Fechner's notation, C = A (Sheynin, 1996a, § 6). Mendeleev however preferred another criterion: the coincidence of the mean of the series' middlemost third ($\overline{x}_2$) with the mean of the means ($\overline{x}_1$ and $\overline{x}_3$) of its extreme thirds:

And here is what Pearson (1905, p. 189) had to say about asymmetrical densities:

All the leading statisticians from Poisson[36] *to Quetelet, Galton, Edgeworth and Fechner* [sic!], *with botanists like De Vries, zoologists like Weldon have realized that asymmetry must be in some way described before we can advance in our theory of variation* (in biology).

Fechner said nothing about general applications of his measure (15), nor was it suitable for estimating *negative* dependences. It passed largely unnoticed, and, at least as far as publication is concerned, Galton preceded Fechner (and originated a theory, the correlation theory). Nevertheless, the latter's modest proposal should not be forgotten.

Fechner persistently and fruitfully considered the treatment of natural-scientific observations in a general way by means of his collectives. Bruns (1898, pp. 342 – 343) thought that a collective was an arithmetical counterpart of a density curve. However, contrary to what he also stated, Fechner had not really constructed a doctrine of frequencies (*Häufigkeitslehre*), or any other system at all. Later Bruns (1906, p. 95) declared that Fechner, by making use of *most primitive* tools, had originated an independent chapter of applied mathematics situated alongside [*neben*] the calculus of probability.

In a paper written in 1906, Chuprov (1960, fourth footnote in § 3, p. 116) translated *Kollektivmasslehre* as *doctrine of mass phenomena*, and, perhaps not quite accurately, directly linked Fechner with Galton, Edgeworth and Pearson. Later, in 1909, he (1959, p. 24) stated:

Among those men of natural sciences whose work had paved the way for the revival of theoretical statistics, Fechner should be mentioned in the first place. His anthropometrical [?] *investigations compelled the celebrated psychophysiologist to apply statistical methods*.

At that time (1909), Chuprov was not yet a mathematically minded statistician (Sheynin 1990/2011, p 26 – 27), and he rather overdid his praise of Fechner.

Fechner outlined a theory for treating observational series in natural sciences[37], but his mathematical approach (not just his tools!) was primitive. As a result, almost everything he achieved had to be repeated at a much higher level. Nevertheless, von Mises (1972)[38] argued that the work of Fechner and some later authors was close [*nahe*] to the frequentist theory of probability (p. 26); that, owing to Laplace's authority, Fechner did not dare base the theory of probability on his collectives but instead established an approach close [*neben*] to it (p. 61); and that Fechner did not think at all about securing a basis for a *rational notion of probability* (p. 99).

The reference to Laplace hardly explained the situation: it would have taken a mathematician of von Mises' own calibre to build such a basis! And it was Pearson and his students who had established an approach *close* to probability theory. Von Mises (1972, pp. 204 – 205) expressed his high opinion of the Biometric school, and stated that *sometimes* their investigations[39] *lacked a deeper stochastic justification.* See also Sheynin (1990/2011, § 15.3), where, in particular, Kolmogorov's similar opinion of 1948 is quoted.

The most important point here, however, is that Fechner's

Constructions prompted at least me [Mises 1972, p. 99] *to adopt a new viewpoint*[40].

Fechner had also influenced other, earlier mathematicians, such as Lipps and Heinrich Bruns, and their work is yet to be studied[41].

I conclude with two early reviews of Fechner (1897), by Lipps (1898) and Bertrand (1899). Naturally enough, Lipps praised Fechner highly, but his analysis seems superficial; and Bertrand (1899, p. 5) argued that Fechner had not formulated any conclusions, nor did he solve *the problem*. Bertrand really had a point; recall, however, that Fechner had not finished his work. At the same time, Bertrand credited Fechner with stimulating new ideas and creating psychophysics.

Acknowledgement. Professor Herbert A. David discovered a few shortcomings, oversights and linguistic mistakes in the preliminary version of this paper. The criticisms and suggestions made by the referees prompted me to improve on some points of the account.

Notes

1. See for example, Kruskal (1958, p. 853) on the measure of association, Harter (l977, p. 96) on the distribution of extreme values, and Hald (1998, pp. 378 – 379 and 363 – 364) on the expression for the double-sided Gaussian law and its justifcation, and on the choice between distributions.
2. Paucker (I819), however, provided a lone (and elementary) application of the MLSq in physics.
3. See Fechner (I860, Bd. 1, pp. 65, 73, 303; l877, p. 2I3; 1860, Bd. 1, p. 301): I4, I5 and l6 lines, respectively.
4. Here and below I give the original German passages.
Der babylonische Thurm wurde nicht vollendet, weil die Werkleute sich nicht verständigen konnten, wie sie ihn bauen sollten; mein psychophysisches Bauwerk dürfte bestehen bleiben, weil die Werkleute sich nicht werden verständigen können, wie sie es einreißen sollen.
5. *Eine exacte Lehre von den functionellen oder Abhängigkeitsbeziehungen zwischen Körper und Seele.*
6. Fechner (1860, Bd. 2, p. 284) quoted Book 3 of Newton's *Optics, On the connection between the nervous system and sensation of light*. He (l859, p. 531; 1860, Bd. 1, p. 65 and Bd. 2, pp. 549 – 550) also cited Euler (1926) on the sensation of sound and repeatedly referred to Bouger, Arago and others, and even to Daniel Bernoulli's moral fortune since its mathematical expression (and, in a sense, its substance) coincided with that of the Weber – Fechner law (see below). Neither did Fechner forget his contemporaries, such as Helmholtz, see especially his historical remarks (Fechner 1860, Bd. 2, Chapter 47), or the astronomer. J. F. Encke.
7. *Das erste feste erfahrungsmäßige Fundamentalgesetz.*
8. *Die Auffassung der Verhältnisse ganzer Größen, ohne dass man die Größen durch einen kleineren Maßstab ausgemessen und den absoluten Unterschied beider kennen gelernt hat, ist eine äußerst interessante psychologische Erscheinung.*
9. In spite of its title, Fechner (l86lb) also dealt with hearing.
10. *Ungewöhnliche Beobachtungswerthe … hat* [haben] *weder Princip noch Grenze und führt zu einer Willkür.*
11. In other words, apply factorial experimentation; cf. § 3.2. For the usual case of two possible states, the pattern was, say, $a_1b_1 – a_1b_2 – a_2b_2 – a_2b_1$. Understandably, Fechner never applied randomization.
12. *Durchschnittsgroße, unregelmäßige Zufälligkeiten.*
13. [The law] *beherrscht* [randomness] *sofern sich derselbe häuft.*
14. *Apparently the change was again achieved both from below and above* (Fechner 1860, Bd. l, p. 8l).
15. Boscovich had already applied this condition and Cournot (l984, § 68) called the pertinent mean (the *Centralwert*, as Fechner called it) the *median*.
16. Laplace's estimation of precision was inseparably linked with the normal law (Sheynin, I977, § 11.3). And in one instance, in about 1819, Laplace (1886, p. 585) preferred the variance. Fechner (l897) abandoned the densities (2) as well as the idea

just mentioned. For that matter, such densities could not have described asymmetric collectives; cf. the beginning of § 6. Fechner's recommendations were sound, but their implementation required the knowledge of the appropriate density.
17. I disregard several other means which Fechner (l897, p. l60) also defined but hardly ever applied. He also discussed the extreme values of an observational series, x_1 and x_m (pp. 32 l – 326) and expressed his desire to discover the law governing their change with m (disagreeing with Encke, who had denied the existence of any such laws); and provided an example in which $x_1 + x_m$ remained almost constant when a series was separated into several (n) groups with m increasing from 2 to 360 and $mn = 360$. Fechner had not chosen $(x_1 + x_m)/2$ as a possible mean.

Fechner (pp. 170 – 171) admitted that the determination of D might be difficult and Lipps (Fechner, 1897, p. 182n) declared that the existence of several most dense values meant that the appropriate series belonged to a mixture of *incompatible* collectives . Lipps also wrongly stated, on p. 88, that the error theory regarded D as the true value of the magnitude sought.
18. Recall (§ 4.2) that Fechner later abandoned this formula.
19. Actually 68, but Herschel himself had disregarded eight of them.
20. Note that formulas (1) und (10) essentially coincide, cf. the end of § 2.1. Fechner (1859, p. 522) mentioned Norman Pogson, who in l850 had devised a scale according to which (m_i, stellar magnitudes, i_i, luminosity of stars)

$$m_1 - m_2 = -2.5\lg_{10}(i_1/i_2).$$

I followed a modern explanation since Fechner's description is difficult to understand.
2l. Concerning the sums of the second kind, Fechner had stated elsewhere (l887a, p. 88; l887b, pp. 216 – 217) that the residuals should be calculated with respect to directly observed quantities, a fact scarcely known to other natural scientists of the day.
22. Fechner also introduced an *absolut normale Fehlervertheilung* (distribution of errors) not attainable when the number of errors was finite, and argued that the approximation to π would become better *the more normal* the distribution became. He used the same term, *normale Fehlervertheilung*, many times more (I860, Bd. 1, p. l25 and Bd. 2, p. 369; l86la. p. 75; 1897, pp 69 and 208), and in any case the noun *Vertheilung* had scarcely appeared before.
23. *Bei einer normalen Fehlervertheilung welche die wahre Beobachtungsgröße als Ausgang der Fehler voraussetzt*.
24. In his last contribution Fechner (1897) repeatedly calculated sums of deviations (see § 5), and noted on p. 283 that the Gaussian law had not yet been applied for this purpose.
25. His statement is not quite definite (although much better than his pronouncement quoted at the end of § 3.1). In general, no calculation can improve second-rate observations, whereas it is not really necessary to secure many good redundant measurements.
26. He had used this term without defining it earlier (1874a, p. 67).
27. *Allgemeinen Wahrscheinlichkeitsgesetze des Zufalls*.
28. At the same time, however, he felt that stochastic formulas will not help to distinguish between essential and random asymmetry (§ 6). This attitude contradicted his earlier belief (l860, Bd. l, p. l28) in the great potential of probability theory.
29. At the time, it was Poincaré (Sheynin, 199l, § 9) who offered the best explanation of randomness and of its relation with necessity. Fechner's unwillingness to discuss randomness was probably reasonable, but it is worth noting (Heidelberger 1987, p. 139) that from about l860 he became reluctant to consider philosophical issues; witness his strange pronouncement (1864, p. 86) that *Philosophers argue, but things follow their normal course* (*Philosophen streiten, und die Dinge gehen ihren Gang*). Now, arbitrary variation might have been a hint at a chaotic change.
30. Lipps (Fechner l897, pp. 86 – 87) if not Fechner himself assumed that the empirical distribution recorded for a collective might turn out to be *an irregular heap of values* (*regellose Ansammlung von Werten*). In such cases, he concluded, the

arithmetic mean was the optimal choice for representing the *tabular values*. A median seems better.
31. His simple advice (p. 67) was to check the sign of the difference (μ – ν) as *m* increased. Its constancy would have indicated an essential asymmetry. The letter ν as printed in my computer differs in its form from the same letter provided in my math. programme.
32. Lipps (Fechner, 1897, p. 366n) illustrated this relation with a simple combinatorial example.
33. *Phantasievoller Philosoph*; *höchst exakter Physiker*, *fasst … zusammen als einen neuen Wissenszweig, die Psychophysik*.
34. *Ich war immer für die Ideen G. T. Fechners zugänglich und habe mich auch in wichtigen Punkten an diesen Denker angelehnt*.
35. Biot (l828 – l829, Bd. 1, p. 169) thought that the *precise determination of weight is one of the most important elements of physics* (*die genaue Gewichtsbestimmung eines der wichtigsten Elemente der Physik ist*). He mentioned Borda but did not discuss the elimination of errors. Incidentally, this is evidence for the fact that, as an experimentalist, Biot hardly influenced Fechner, his translator. Note that Biot (1828 – 1829, Bd. 3, p. 473) knew that the sensation of *Galvanismus* was subjective.
36. Pearson was probably referring to the binomial distribution (whose limiting behaviour De Moivre had studied back in l733). With regard to Fechner, he apparently (and reasonably) restricted his attention to his final work (Fechner, l897); the only alternative is that he committed a glaring error of chronology.
37. Cf. von Mises (l964, p. 9)*: The subject of the Kollektivmasslehre is the result of repeated observations* (*Gegenstand der Kollektivmasslehre sind die Ergebnisse wiederholter Beobachtungen*). Later he abandoned Fechner's term (see Note 4l below).
38. As chance would have it, Fechner used the pen name *Dr. Mises* for some of his non-scientific writings.
39. *Eine tiefere wahrscheinlichkeitstheoretische Begründung vermisst*.
40. Fechner's *Ausführungen bildeten, wenigsten für mich, die Anregung zu der neuen Betrachtungsweise*.
41. von Mises (l972, p. 197) praised Bruns for his work on non-Gaussian distributions *within the boundaries of the so-called* [!] *Kollektivmasslehre* (*im Rahmen der sogenannten Kollektivmasslehre*).

References

AHES = *Arch. Hist. Ex. Sci.*
Sächsische Abh., i(j) =*Abh. Kgl. Sächs. Ges. Wiss*., Bd. i of the entire series being Bd. j of its Math.-Phys. Kl.
Sächsische Berichte = *Berichte* of the same Society

G. T. Fechner

(1859). Über ein wichtiges psychophysisches Grundgesetz und dessen Beziehung zur Schätzung der Sterngrössen. *Sächsische Abhandlungen*, 6(4), 455 – 532.
(1860). *Elemente der Psychophysik*, Bde 1 – 2. Leipzig.
(1861a). Über die Correctionen bezüglich der Genauigkeitsbestimmung der Beobachtungen. *Sächsische Berichte*, Bd. 13, 57 – 113.
(1861b). Über einige Verhältnisse des binocularen Sehens. *Sächsische Abh.*, 7(5), 357 – 563.
(1864). *Über die physikalische und philosophische Atomenlehre*. Leipzig: First published in 1855.
(1874a). Über die Bestimmung des wahrscheinlichen Fehlers eines Beobachtungsmittels durch die Summe der einfachen Abweichungen. *Annalen Phys. Chem.*, Jubelband, 66 – 81.
(1874b). Über den Ausgangswerth der kleinsten Abweichungssumme, dessen Bestimmung, Verwendung und Verallgemeinung. *Sächsische Abh.*, 18(11), No. 1, 3 – 76.
(1877). *In Sachen der Psychophysik*. Leipzig.
(1882). *Revision der Hauptpunkte der Psychophysik*. Leipzig.
(1887a). Über die Frage des Weberschen Gesetzes und Periodicitätsgesetzes im Gebiete des Zeitsinnes. *Sächsische Abh*., 22(13), 1 – 108.

(1887b). Über die Methode der richtigen und falschen Fälle in Anwendung auf die Massbestimmungen der Feinheit oder extensiven Empfindlichkeit des Raumsinnes. Ibidem, 109 – 312.
(1897). *Kollektivmasslehre* (G. E. Lipps, Editor). Leipzig:

Other authors

Archibald T. (1994). Mathematical theories of electricity and magnetism to 1900. In I. Grattan-Guinness (Editor). *Companion Enc. Hist. and Phil. Math. Sci*., 1208 – 1219. London.
Bertrand J. (1899). Review of Fechner (1897). *J. des Savants*, No. 1, 5 – 17.
Bessel F. W. (1838). *Gradmessung in Ostpreussen*. Berlin.
Biot J.-B. (1828 – 1829). *Lehrbuch der Experimentalphysik*, Bde 1 – 5. Translated from French by G. T. Fechner. Leipzig.
Boring E. G. (1950). *History of Experimental Psychology*. New York.
Bruns H. (1898). Zur Kollektivmasslehre. *Phil. Studien*, Bd. 14, 339 – 375.
---, (1906). *Wahrscheinlichkeitsrechnung und Kollektivmasslehre*. Leipzig.
Chuprov A. A. (1959). *Ocherki po teorii statistiki* [Essays in the theory of statistics]. Moscow. First published in 1909.
---, (1960). *Voprosy statistiki* [Issues in statistics]. Moscow.
Cournot A. A. (1984). *Exposition de la théorie des chances et des probabilités* (B. Bru, Editor). Paris. First published in 1843. **S, G,** 54.
Cowles M. (1989). *Statistics in Psychology. An Historical Perspective.* Hillsdale, NJ.
David H. A. (1988). *The Method of Paired Comparisons*. London. First published in 1963.
Ebbinghaus H. (1908). *Abriss der Psychologie*. Leipzig.
---, (1911). *Grundzüge der Psychologie*, Bd. 1. Leipzig. First published in 1897.
Edgeworth F. Y. (1996). The element of chance in competitive examinations. *Writings in probability, statistics and economics*, C. R. McCann, Jr. (Editor). Cheltenham, vol. 3, 529 – 564. First published in 1890.
Euler L. (1926). Tentamen novae theoriae musicae. In *Opera Omnia*, ser. 3, t. 1, 197 – 427. Leipzig. First published in 1739. French translation 1865.
Freud S. (1961). Die Traumdenkung. In *Werke*, Bd. 2/3. Frankfurt am Main. First published in 1900.
---, (1963a). *Jenseits des Lustprinzips*. In *Werke*, Bd. 13. Frankfurt am Main: First published in 1920.
---, (1963b). Selbstdarstellung. In *Werke*, Bd. 14, 31 – 96. Frankfurt am Main. First published in 1925.
Galton F. (1879). The geometric mean in vital and social statistics. *Proc. Roy. Soc*., vol. 29, 365 – 367.
Gauss C. F. (1880). Bestimmung der Genauigkeit der Beobachtungen. In *Werke*, Bd. 4, 109 – 117). First published in 1816.
Hald A. (1998). *History of mathematical statistics from 1750 to 1930*. New York.
Harter H. L. (1977). *Chronological annotated bibliography on order statistics*, vol. 1. Wright-Patterson Air Force Base, OH: US Air Force.
Heidelberger M. (1987). Fechner's indeterminism: From freedom to laws of chance. In L. Krüger, L. J. Daston, M. Heidelberger (Editors). *Probabilistic revolution*, vol. 1, 117 – 156. Cambridge, MA.
Helmholtz H. (1913). *Die Lehre von den Tonempfindungen*. Braunschweig. First published in 1863.
Jaynes J. (1971), Fechner. *Dict. Scient. Biogr*., vol. 4, 556 – 559.
Kruskal W. (1958). Ordinal measures of association. *J. Amer. Stat. Assoc*., vol. 53, 814 – 861.
Kuntze J. E. (1892). *Fechner*. Leipzig.
Laplace P. S. (1886). Third supplement to *Théorie anal. prob.* In *Oeuvr. Compl.*, t. 7, No. 2, 581 – 616). Paris. First published in 1819.
Lasswitz K. (1902). *Fechner*. Stuttgart. First published in 1896.
Lexis W. (1903). Über die Theorie der Stabilität statistischer Reihen. In *Abh. Theor. Bevölkerungs- und Moralstatistik*, 170 – 212). Jena. First published in 1879.
Lipps G. F. (1898). Über Fechners Kollektivmasslehre und die Verteilungsgesetze der Kollektivgegenstände. *Phil. Studien*, Bd. 13, 579 – 612.
Misiak H., Sexton V. (1966). *History of psychology*. New York:

Moore P. G. (1978). Runs. In W. H. Kruskal, J. M. Tanur (Editors), *Intern. enc. of statistics*, vol. 1, 655 – 661. New York.
Newton I. (1934). *Mathematical principles of natural philosophy*. Cambridge. First published 1729.
---, (1950). *Theological manuscripts*. Liverpool:
Paucker M. G. (1819). Über die Anwendung der Methode der kleinsten Quadrutsumme auf physikalische Beobachtungen. Mitau (Jelgava). *Programm zur Eröffnung des Lehrkursus auf dem Gymnasium illustre zu Mitau.*
Pearson K. (1898), Cloudiness. *Proc. Roy. Soc*., vol. 62, pp. 287 – 290.
--- (1905). *Das Fehlergesetz und seine Verallgemeinung durch Fechner und Pearson*. A rejoinder. *Biometrika*, vol. 4, 169 – 212.
---, (1930). *Life, letters and labours of Fr. Galton*, vol. 3A. Cambridge.
Pukelsheim F. (1993). *Optimal design of experiments*. New York.
Ranke K. E., **Greiner G.**, (1904). Das Fehlergesetz und seine Verallgemeinung durch Fechner und Pearson. *Arch. Anthropologie*, Bd. 2(30), 295 – 332.
Sheynin O. B. (1977). Laplace's theory of errors. AHES, vol. 17, 1 – 61.
---, (1978). Poisson's work in probability. AHES, vol. 18, 245 – 300.
---, (1984). On the history of the statistical method in meteorology. AHES, vol. 31, 53 – 95.
---, (1986). Quetelet as a statistician. AHES, vol. 36, 281 – 325.
---, (1991). Poincaré's work on probability. AHES, vol. 42, 137 – 171.
---, (1993). On the history of the principle of least squares. AHES, vol. 46, 39 – 54.
---, (1995). Helmert's work in the theory of errors. AHFS, vol. 49, 73 – 104.
---, (1996a) Mendeleev and the mathematical treatment of observations in natural science. *Hist. Math.*, vol. 23, 54 – 67.
---, (1996b). *History of the theory of errors*. Egelbach.
---, (1996c, 2011). *Chuprov*. Göttingen. Originally published in Russian, 1990.
---, (1998). Theory of probability, its definition and its relation to statistics. AHES, vol. 52, 99 – 108.
---, (2000). Bessel: some remarks on his work. *Hist. scientiarum*, vol. 10, 77 – 83.
---, (2007), The true value of a measured constant and the theory of errors. Ibidem, vol. 17, pp. 38 – 48.
---, (2017), *History of probability. Historical essay*. Berlin. **S, G,** 11.
Singer B. (1979). Distribution-free methods for non-parametric problems: a classified and selected bibliography. This *Journal*, vol. 32, 1 – 60.
Spearman C. E. (1937). *Psychology down the ages*, vols. 1 – 2. London.
Stigler S. M. (1986). *History of statistics*. Cambridge, MA.
von Mises R. (1931). *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*, this being his *Vorlesungen aus dem Gebiete der angewandte Mathematik*, Bd. 1. Leipzig.
---, (1964). Über die Grundbegriffe der Kollektivmasslehre. In *Sel. papers*. Birkhoff G., Frank Ph., Goldstein S., Kac M., Prager W., Szegö, G. (Editors), vol. 2, 3 – 14. Providence, RI. First published in 1912.
---, (1972). *Wahrscheinlichkeit, Statistik und Wahrheit.* Wien. First published in 1928. English translation New York, 1981.
Weber E. H. (1905). *Tatsinn und Gemeingefühl*. Leipzig. First published in 1851.
Yale G., Kendall M. G. (1958). *Introduction to theory of statistics*. London. First published in 1937.

V

Geometric probability and the Bertrand paradox

Historia scientiarum, vol. 13, 2003, p. 42 – 53

I describe the early history of geometric probability (§ 1) and consider its later developments which include the appearance of the Bertrand paradox (§ 2). Finally, in § 3, I dwell on the later history of that paradox. Materials, previously unknown, utterly forgotten or left unconnected with my subject are in §§ 1–4, 1–6 (Buffon in 1835), 2–3, 2–5 (Newcomb), 2–8 (Darboux), 3–2 and 3–4 to 3–6 whereas I myself (1971) treated the subject of § 1.1. Mathematicians continued to study geometric probability not knowing the reasoning of Poincaré (§ 2–9). Taken together with Seneta et al (2001), see § 2–4), this paper sufficiently describes the history of geometric probability, but in addition I (§ 4) show that the Bertrand paradox should have been discussed on the base of the theory of information.

1. The early history

1. In a manuscript written sometime between 1664 and 1666 Newton (1967) considered geometric probabilities. Suppose that a circle is divided into two sectors whose areas are as 2**:**√5. Then, as he stated, the chances that a ball, falling vertically on the centre of the circle, *tumbles* in either sector, were in he same ratio. Newton also remarked that the chances of the various castings of an irregular die might be calculated similarly.

2. The English translation of 1692 of Huygens' classical treatise of 1657 by Arbuthnot (ascribed by Todhunter 1865, p. 49) contained a simpler version of the second problem, but its solution was due to Simpson (1740, pp. 67 – 70). For a cuboid with sides proportional to *a*, *b* and *c*, he obtained the probability that the ball will rest on face *a*, *b* (say)

$$P = A°\mathbf{:}90, \ \sin A = 2ab/\sqrt{a^2+b^2+c^2}.$$

The dimensionality of his formula was however wrong. Peres Larigno (1985, p. 101) noted that Simpson made a mistake and offered her own formula without explanation:

$$P = \frac{2}{\pi}\mathrm{arctg}\,ab/c\sqrt{a^2+b^2+c^2}.$$

She also wrote down the (now obvious) expressions concerning the two other faces but had not indicated that the sum of all three probabilities was indeed unity.

3. Daniel Bernoulli (1735) applied geometric probability when he discussed the uniformity of the planetary system. He considered the inclinations of the orbits of the five then known planets relative to the ecliptic as random variables with a continuous uniform distribution. They were small and the probability of a random origin of that circumstance, as he decided, was negligible[1]. Todhunter (1865, p. 223) remarked that it was also possible to consider the arrangement of the orbital poles.

4. Scholars tacitly applied geometric probability when introducing densities. Thus De Moivre (1743, p. 323): *The probability of life's failing in any interval of time AF is measured by the fraction FA/SA* … He assumed a continuous uniform law of mortality for ages exceeding 12.

Simpson (1757) assumed that the probabilities of observational errors were proportional to appropriate rectilinear or curvilinear areas.

In the first part of his memoir (1764 – 1765) Bayes actually repeated Newton's assumption (Item I above) about a vertically falling ball.

Much later Poisson (1837, p. 274) illustrated the continuous uniform distribution in a similar way. Cf. § 2.2[2].

5. John Michell (1767) calculated the probability that two stars out of the n scattered over the sky *by mere chance* (uniformly) were only separated by 1°. Select a point A on a sphere of radius R; draw the radius passing through it; and cut the sphere by such a plane perpendicular to that radius that any point on the circumference thus obtained is situated at distance 1° from A. The surfaced areas of the spherical segment with vertex A and of the entire sphere are, respectively,

$$s = \pi R^2/57.296^2,\ S = 4\pi R^2,$$

and the geometric probability sought will be $s/S = p = 1/13{,}131$.

Michell's main question was, whether the actual scatter of stars was random or designed. The expected number of stars situated not further apart than 1° is $a = pn$. For $n = 6000$ which is roughly the number of stars visible by the naked eye, $a = 0.45$. William Herschel, however, discovered a few hundred visual binaries. Nowadays more than 60 thousand are listed of which a thousand are real double stars. This means that the stars are not uniformly (in Michell's sense) distributed.

6. Buffon (1777, p. 471) forcefully introduced geometric probability. Until then, as he mistakenly indicated, *l'analyse est le seul instrument* applied *dans la science des probabilités*. He intended to *put geometry in possession of its rights in the science of chance* and formulated his celebrated problem (below). He (p. 474) also considered a number of problems in which nothing but the ratio of some *étendue*'s had to be accounted for.

This ratio may be regarded as Buffon's indirect definition of geometric probability, cf. § 2.1. And here is his main problem. A needle of length $2r$ is dropped *randomly* on a set of parallel lines $a > 2r$ apart. Required is the probability that the needle intersects one of the lines. A simple calculation renders[3]

$$p = 4r:\pi a. \qquad (1)$$

Actually, Buffon (pp. 473 – 474) only sought the ratio r**:**a for p = 1/2. On the other hand, he (pp. 471 – 473) studied several other problems of the same type including the mentioned just below. A summary of Buffon's work (undoubtedly compiled by him himself) appeared several decades earlier. Here is its beginning (Anonymous 1735):

This year M. le Clerce de Buffon presented to the Academy solutions of problems regarding the game of franc carreau.

An ecu or a louis is thrown on a set of equal and supposedly regular tiles (carreaux) and it is required to find out the chances (*combien il y à parier*) *that the coin falls only on one tile*.

And after as few lines:

The issue presented here is of a new kind in the sense that it belongs to geometry and to figures that had not at all entered the subject.

Todhunter (1865, p. 203) noted the remark just quoted but later authors do not cite it. I myself (1991, p. 203) mentioned the relevant work of Buniakovsky (in 1837 and 1846) and Markov (in 1900). Possibly the first commentator was Laplace (1812/1886, pp. 365 – 366) who called that problem an example of a new *genre particulier de combinaisons du hazard* that might be applied *pour rectifier les courbes ou carrer leur surfaces*, cf. § 3.7 and Note 10. He noted that the number π can be experimentally (although only approximately) derived by formula (1).

Several such allegedly successful attempts were indeed carried out. Gridgeman (1960) however questioned their results[3]; See also Shneider (1966, §§ 1 – 2 of Chapter 1). There, the precision of such derivations is regarded as low.

2. Nineteenth century

1. Cournot (1843, p. 29) was the first to offer directly a definition of geometric probability, or, rather, to unite the discrete and continuous cases. He stated that the *mathematical probability* was the ratio

De l'étendue des chances favorables à un événement à l'étendue totale des chances.

If modernized, *measure* would have replaced the *étendue*.

2. On pp. 89 – 92 Cournot applied geometric probability for deriving the law of distribution of a function of several random arguments, and in Chapter 6 he explained the notion of density function by geometric considerations, cf. § 1.4. Here is one of his examples.

Given, function $u = |x - y|$ of arguments uniformly distributed on [0, 1]. After considering the areas of the appropriate figures, he concluded that, for $0 \le a \le 1$,

$$P(u \ge a) = (1 - a)^2.$$

Had he calculated the contrary event, he would have been the first to formulate the once-popular encounter formula (Whitworth 1886 and possibly 1867; Laurent (1873, pp. 67 – 69): two people are to meet at a definite spot during a specified time interval. They arrive independently and occur *at random*. The first to arrive waits for a certain time, then leaves. Required is the probability of the encounter.

3. Boltzmann (1868, p. 50), tacitly applied geometric probability. Boltzmann defined probability that the velocity of a molecule was contained in an infinitesimal interval as the ratio of time during which that happened to the total time of observation. I leave aside an earlier definition of probability in physics as well as the problem of ergodic hypotheses.

4. Darwin (1881, pp. 52 – 55), see also Sheynin (1980, pp. 349 – 350), strewed paper triangles over some ground. Earthworms dragged them away but Darwin recovered most of them. He found out that the worms had not seized *indifferently by chance any part* of the triangles, considered a few versions of *indifference* and rejected all of them. Earthworms are necessary for the vegetable kingdom and, what is noteworthy, Darwin's experiment forestalled the Bertrand problem.

5. Geometric probability became topical. Seneta et al (2001) studied the relevant work of Sylvester, Crofton, Barbier and Bertrand. On p. 506 they quoted Crofton's biographer J. Larmor who had concluded, in 1915, that

The subject of most of his (Crofton's) *original memoirs was that beautiful combination of geometry with the integral calculus to which has been given* (perhaps by himself) *the name of Local Probability*.

Crofton, for example, had solved Sylvester's remarkable problem that required the probability that four points taken *at random* within a finite convex domain form a convex quadrilateral. See Czuber (1908, pp. 99 – 102) for a few particular cases of that problem.

6. Several commentators questioned Michell's understanding of chance as well as the later (not described in § 1.5) calculations and examined the probability of the distance between two random points situated on a sphere (Sheynin 1984, pp. 158 – 160). It turned out that the notion of a random arrangement of the two points was indefinite and that therefore several remarkable configurations of stars should be studied (Cournot 1843, pp. 175 – 177; Betrand 1888, pp. 6 – 7 and 170 – 171). Thus (Bertrand p. 7):

Les probabilités relatives à la distribution des étoiles, en les supposant semées au hasard sur la sphère céleste, sont impossibles à assigner si la question n'est pas précisée davantage.

Newcomb also was a commentator and I adduce his related statement (1862, p. 21) about another astronomical problem:

The probability that a (certain) *point will … fall in any portion of the* (appropriate) *circumference will be equal to the length of that portion divided by the entire circumference*.

While discussing yet another astronomical problem, Poisson (1837, p. 306) assumed that the probability for a point randomly situated on a hemisphere to belong to its infinitely narrow circular belt was proportional to the belt's *étendue*.

7. Czuber devoted a book (1884) to problems in geometric probabilities which shows that those probabilities became widely studied.

8. The outcome of the developments in the 19th century (§ 2–5) was the creation of integral geometry at the junction of geometry and measure theory, and, more to the point, of combinatorial integral geometry. Another upshot of the same developments was Bertrand's discovery that the expression *at random* (actually, *uniform randomness* was meant) was not sufficiently specific (§ 2–6). I describe his celebrated problem[4], note Crofton's earlier opinion (§ 3–4) and Darwin's conclusion (§ 2–4).

Bertrand (1888, pp. 4 – 5) suggested to determine the probability that a randomly drawn chord of a given circle is longer than the side of an inscribed equilateral triangle. He considered three cases, or rather three different problems, cf. § 2–9.

a) One of its endpoints is fixed; $p = 1/3$.

b) Its direction is fixed; $p = 1/2$.

c) Its centre is located in any point of the circle with the same probability; $p = 1/4$.

A curious relevant statement was due to Darboux (1912, p. 50):

Par des raisonnements qui peuvent paraître également plausibles, il trouve pour la probabilité cherchée deux valeurs différentes, tantôt 1/2, tantôt 1/3. Cette question l'a préoccupé; il en avait trouvé la solution, mais il la laisse à chercher à son lecteur.

Darboux was referring to Bertrand, and in failing to mention the third solution he apparently followed Poincaré (§ 2–9).

9. Poincaré (1896, p. 97; 1912, p. 121) noted that the probability that a point (x, y) was situated within some figure S can be defined as

$$\int\int \varphi(x, y)dxdy \text{ over } S$$

with φ specified in accord with the particular problem. On his next pages he passed to the Bertrand problem (or rather to its first two cases) and tacitly assumed that $\varphi \equiv 1$. The chord can be fixed with respect to the centre of the circle O and the polar axis passing through, and beginning in O, by two parameters , ω and α, the polar angles of A, an endpoint of the chord, and of P, its centre; or by two other parameters, θ and ρ, the polar coordinates of P. Now,

$$\int\int d\omega d\alpha \neq \int\int d\rho d\theta$$

with the integrals taken over the given circle. This inequality, as Poincaré stated (happily forgetting Bertrand's third case), explained the paradoxical nature of the problem.

Actually he began his deliberations by discussing the probability of a point x belonging to some segment $[a, b]$. He decided that that probability was equal to

$$\int_a^b \varphi(x)dx.$$

However, Poincaré had not normed his formula. The function φ had to be appropriately chosen.

3. The subsequent history

First, I mention other natural solutions of the Bertrand problem and it is for this reason that I somewhat violate the chronology of presentation.

1. Czuber (1908, pp. 107 – 108) provided three more such solutions:

d) One endpoint of the chord is fixed, and the chord passes through any given point of the circle[5]; $p = 1/3 + \sqrt{3}:2\pi \approx 0.609$.

e) Both endpoints of the chord are chosen randomly. This case is identical with a).

f) A much more difficult case. Two points of the chord inside the circle are chosen randomly; $p = 1/3 + 3\sqrt{3}:2\pi \approx 0.746$.

2. Much more interesting was De Montessus' discovery (1903) that the problem had an uncountable set of answers. Suppose that Ox is the x-axis and mark points D and C on its positive half,– its intersections with concentric circumferences with common centre in point O and radii $OD = 1/2$ and $OC = 1$. Arbitrary

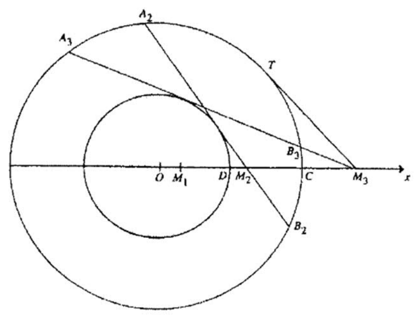

Fig. 1. De Montessus (1903).
A point moves along the axis from D to infinity, and, correspondingly, the probability sought in the Bertrand problem is seen to have an uncountable set of values. $OD = 1/2$, $OC = 1$.

points $M_2(x)$ and $M_3(x)$ are situated on the same semiaxis, between the two circles and beyond the larger of them respectively. Tangents A_2B_2 and A_3B_3 to the smaller circumference pass through M_2 and M_3 respectively, and M_3T is the tangent to the larger circumference with point of contact T. Finally, $M_1(x)$ is an arbitrary point on the same semiaxis inside the smaller circle.

For points M_2 and M_3 the probability sought is, respectively,

$$P_2 = \text{angle } A_2M_2O/\pi = [2\arcsin(1/2x)]/\pi, \qquad (2)$$
$$P_3 = \text{angle } A_3M_3O/\text{angle } TM_3O = [\arcsin(1/2x)]/[\arcsin(1/x)], \qquad (3)$$

with $1/2 \le x \le 1$ and $x \ge 1$ respectively.

When moving from point O in the positive direction (say), the probability P_2 decreases from 1 at point D to 1/3, and, from point C to infinity, probability P_3 increases from 1/3 to 1/2. It is rather difficult to prove that P_3 increases monotonically (and De Montessus had not done it), but already for $x = 1.01$ and 1.1 it is 0.36 and 0.41 respectively and it reaches value $(1/2 – 1/1{,}600)$ at $x = 10$.

Note that the coincidence of points M_2 or M_3 with D leads to Bertrand's first solution and the movement $M_3 \to \infty$ provides his second case. His third solution concerned a point rather than a straight line and was thus different.

De Montessus calculated the general mean probability of the studied event. However, it was hardly proper to include in the calculation, as he did, points such as M_1 for which the stipulated condition was certainly satisfied. More important, while calculating the mean probability for the continuous case, De Montessus first determined a finite sum, and, when adding together the appropriate fractions, he added separately their numerators and their denominators. Nevertheless, his mean probability ($P = 1/2$) was correct and could have been established by noting that the studied interval beyond the circle was infinite.

3. In 1909 Borel published a book whose edition of 1950 I saw. It contained two chapters on geometric probabilities, and here are some of their considerations.

a) Points *b, d, c* and *a* belong to one and the same straight line and $b < d < c < a$ (p. 120). Then

$$P(d < x < c) = [f(d) - f(c)]:[f(b) - f(a)]$$

where the monotonic function *f* should be somehow chosen. Borel did not cite Poincaré but likely followed him.

b) Borel (p. 132) solved the encounter problem (§ 2–2).

c) Borel (p. 137) considered the distance between two random points on a sphere (§ 2–5). He chose the solution bearing in mind practical points: a point on a sphere can only be fixed to a certain degree of precision which depends on its position relative to the sphere's equator.

d) The Bertrand problem should be specified. Most of its natural solutions lead to $p = 1/2$ (pp. 148 – 149).

4. Schmidt (1926) issued from Poincaré. Suppose that a straight line belonging to domain A is also situated in domain B. Then (pp. 35 – 36) the probability of that event is the ratio of two integrals of φ, of a continuous and differentiable function, over those domains.

He also stated that his conclusion was valid for other geometric objects as well[6]. Schmidt stipulated that the probability should persist under translation and rotation of the coordinate system[7] and proved that this condition was only satisfied for

$$\int\int d\rho d\theta \;\text{ or }\; \int\int[\partial(\rho,\theta)/\partial(\xi,\eta)]d\xi d\eta$$

where θ and ρ were the same as in § 2–9.

For the Bertrand problem the probability sought thus becomes equal to 1/2. Finally, Schmidt remarked that Crofton (1868) had asserted, although without justification, that θ and ρ should be preferred to any other system of coordinates. Schmidt had not provided an exact reference and anyway Crofton's memoir does not contain any such statements[8].

It is not amiss, however, to indicate that Crofton (p 181) mentioned *The new Theory of Local or geometrical probability* and concluded on the next page, before Bertrand but somewhat vaguely, that *at random* was not always definite.

5. Without citing anyone except Czuber, Bower (1934) proved that the Bertrand problem had infinity of solutions. He derived formula (3), then specified it and obtained, in particular, all the six solutions (§§ 2–8 and 3–1). His approach was essentially similar to what Poincaré (§ 2–9) had sketched.

However, Bower had not explained the situation or his mathematical reasoning well enough, he only stated that the *equilikely* element (the differential in the appropriate integral) can be *weighted* in different ways. Suppose (his p. 508) that two points on a unit circumference with centre O are randomly chosen. Let the midpoint of the chord this defined be N and denote $ON = x$. Then the probability that a point on ON belongs to interval $[x, x + dx]$ will be not $dx/2$ but rather

$$[\pi(x + dx)^2 - \pi x^2]/\pi \approx 2xdx.$$

The *equilikely* element was the element of the appropriate area.

6. I attempted to locate all the pertinent literature up to ca 1940. Without claiming success, I mention one more paper, Petrini (1937). He only referred to Bertrand and to Borel, but did not provide the exact source. He offered his own definition of geometric probability. Thus, when points in a circle are considered, it should be partitioned into n elements égaux $\delta\omega$ with $n \to \infty$. For Bertrand's case c), see § 2–8, two concentric circles with radii r and $2r$ provide the immediate solution. Here Petrini gets

$$P = \lim(p\delta\omega/q\delta\omega) = p/q = 1/4,\ n \to \infty. \qquad (4)$$

Here, p and q are, respectively, the numbers of the elements in the two circles. For the two other cases which he also considers he assigns appropriate weights to the elements but does not obtain the same answer (4) that he believes to be the only correct solution.

7. Here, I conclude my account since later contributions do not yet belong to history. Nevertheless, I mention Kendall & Moran (1963). These authors certainly pay attention, as Crofton (1868) did, to the fact that geometric probability can essentially simplify the calculation of integrals[9] and I indicate that they worthily discuss both the Buffon and Sylvester problems.

Finally, I note that geometric probability and integral geometry (§ 2–8) are the new tools of the newly created discipline, stochastic geometry (Ambartzumian 1999).

4. Conclusion

Geometric probability appeared in a published work at the end of the 17th century. It became tacitly used in the mid-18th century and was expressly introduced by Buffon. In the 19th century mathematicians actively studied geometric probability which led to the origin of integral geometry. The Bertrand paradox required a more attentive attitude towards geometric probability and Poincaré showed the proper approach to stochastic problems involving geometric objects. For a few decades, however, his considerations remained forgotten. De Montessus proved that the Bertrand problem had an uncountable set of solutions but was unable to conclude definitely his discovery. Then several decades ago, geometric probability became a tool of the new discipline, stochastic geometry. Astronomers have been tacitly applying geometric probability from 1735 and Boltzmann and Darwin used it in the same way in physics and biology respectively.

Most commentators finally decided that the proper probability sought was 1/2 which could have been immediately agreed upon since according to the theory of information that probability 1/2 follows from ignorance. We may recall the renown Latin proverb *Ex nihil nihilo fut* (Nothing follows from ignorance) which Ellis (1850, p. 57) cited on another occasion.

Notes

1. I do not distinguish strict from non-strict inequalities. Bernoulli's reasoning was typical: when based on the same argument, design had been inferred even in the Talmud (Sheynin 1998, pp. 191 – 192). Such conclusions are justified when only the observed regularity makes sense, see the celebrated D'Alembert – Laplace problem (Todhunter 1865, p. 273).

2. Later Czuber (1908, p. 10) remarked that there existed *a second point of view* according to which positive knowledge was required for stating that a probability was equal to 1/2. Poisson (1837, p. 47) determined the subjective probability of extracting a white ball from an urn containing white and black balls. He arrived at 1/2 by assuming that all possible outcomes were equally probable. See § 6.

3. For n as the number of throws and $a = 1$, Gridgeman's formula for the variance of the experimental value of π was

$$\sigma^2 = \frac{\pi^2[(\pi/4r) - 1]}{n}.$$

The value $r = 1/2$, its maximally possible value, was therefore also its best value. Laplace (1812), in the first edition of that classic, had arrived at the same result, but then, without explanation, changed over *from truth to error* (Todhunter 1865, pp. 590 – 591). In the later editions he stated that the optimal value of r was π**:**8 (= 0.39), and Todhunter attempted to reconstruct his reasoning.

4. Bru and Jongmans (2001, p. 187) remarked that Bertrand had (initially)

Constructed his problem in his handwritten lecture notes for the École Polytechnique … as a transformation of the famous Buffon needle problem.

5. Or, which is the same: a point of the chord situated inside the circle and the direction of the chord are chosen randomly (Barth and Haller 1996, p. 391).

6. Schmidt wrote his paper somewhat carelessly. Thus B obviously belongs to A, and for a straight line rather than for a segment both domains should be unbounded which is fraught with difficulties.

7. Invariance under reflection is now also included.

8. Note also Prokhorov's opinion (1988): from the geometrical viewpoint, the most natural assumption in the Bertrand problem is that θ and ρ are independent and uniformly distributed on intervals $0 \le \theta \le 2\pi$, $0 \le \rho \le 1$.

9. Recall Laplace's relevant remark (§ 1–6) which was regrettably too concise. Note also the expression *squaring the surfaces of curves* which is difficult to understand.

References

Ambartzumian R. V. (1999), Stochastic geometry. In *Veroiatnost i matematich. statistika. Enziklopdia*. (Prob. and math. stat. Enc.), p. 682. Editor Yu. V. Prokhorov. Moscow.

Anonymous (1735), Géomètrie. *Hist. Acad. Roy. Sci. avec Mém. de math. et de phys*., pp. 43 – 45 of the *Histoire*.

Barth Fr., Haller R. (1996), *Stochastik. Leitungskurs*. München. Fifth edition.

Bayes T. (1764 – 1765), Essay towards solving a problem in the doctrine of chances. Second part under another title. Reprint of pt. 1: *Biometrika*; vol. 45, 1958, pp. 296 – 315 and E. S. Pearson, M. G. Kendall (1970), *Studies in the history of statistics and probability*. London, pp. 131 – 153.

Bernoulli Daniel (1735), Recherches physiques et astronomiques … *Werke*, Bd. 3. Basel, 1987, pp. 303 – 326.

Bertrand J. (1888, 1907), *Calcul des probabilités.* Paris. Reprint of first edition: New 1970, 1972. Second edition barely differs from the first one.

Boltzman L. (1868), Studien über das Gleichgewicht der lebenden Kraft. *Wiss. Abh*., Bd. 1, pp. 49 – 96. Leipzig, 1909.

Borel E. (1909, 1950), *Eléments de la théorie des probabilités*. Paris.

Bower O. K. (1934), Note concerning two problems in geometrical probabilities. *Amer. Math. Monthly*, vol. 41, pp. 506 – 510.

Bru B., Jongmans Fr. (2001), Joseph Bertrand. In Heyde C. C., Seneta E. *Statisticians of the centuries*, pp. 185 – 189. New York.

Buffon G. L. L.de (1777), *Essai d'arithmétique morale*. *Oeuvr. Phil*. Paris, 1954, pp. 456 – 488. English translation in Internet.

Cournot A. A. (1843, 1984), *Exposition de la théorie des chances et des probabilités*. Editor B. Bru. **S, G,** 54.

Crofton M. W. (1868), On the theory of local probability applied to straight lines drawn at random in a plane. *Phil Trans. Roy. Soc*., vol. 158, pp. 181 – 199.

Czuber E. (1884), *Geometrische Wahrscheinlichkeit und Mittelwerte*. Leipzig.

--- (1903, 1908), *Wahrscheinlichkeitsrechnung und ihre Anwendung auf Fehlerausgleichung, Statistik und Lebensversicherung*, Bd. 1. New York, 1968.

Darboux G. (1912), Eloge historique de Bertrand. Lu 1901. *In author's Eloges académiques*, pp. 1 – 60. First publ. in J. Bertrand's *Eloges académiques*, nouv. sér. Paris, 1902, pp. xii – li.

Darwin Ch. (1881), *Formation of vegetable mould*. London, 1945.

De Moivre A. (1725), *Treatise of annuities on lives*. Edition of 1743 incorporated in author's *Doctrine* … (1756, 1967, pp. 261 – 328).

De Montessus R. (1903), Un paradoxe du calcul des probabilités. *Nouv. annales math.*, sér. 4, t. 3, pp. 21 – 31.

Ellis R. L. (1850), Remarks on the alleged proof of the method of least squares. Reprinted in author's *Math. and other writings*. Cambridge – London, 1863.

Gridgeman N. T. (1960), Geometric probability and the number π. *Scripta math.*, vol. 25, pp. 183 – 195.

Kendall M. G., Moran P. A. P. (1963), *Geometrical probabilities*. London.

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

Laurent P. H. (1873), *Traité du calcul des probabilités*. Paris.

Michell J. (1767), Inquiry into the probable parallax and magnitude of the fixed stars. *Phil. Trans. Roy. Soc. Abridged*, vol. 12, 1809, pp. 423 – 438.

Newcomb S. (1862), Determination of the law of distribution of the nodes and perihelia of the small planets. *Astron. Nachr*., Bd. 58, pp. 210 – 220.

Newton I. (ca. 1664 – 1666), MS without title. In author's *Math. papers*, vol. 1, pp. 58 – 61. Editor D. T. Whiteside. Cambridge, 1967.

Peres Larigno M. T. (1985), On the history of the concept of geometric probability. *Voprosy Istorii Estestvoznania i Tekhniki*, No, 4, pp. 100 – 103. In Russian.

Petrini H. (1937), Le paradoxe de Bertrand. *Arkiv for matematik, astronomi och fysik*, t. 25. This source consists of four issues each having pts. A and B with separate paging for each item. Petrini's paper is in No 3, A16.

Poincaré H. (1896, 1912), *Calcul des probabilités*. Sceaux, 1987.

Poisson S.-D. (1837, 2003), *Recherches sur la probabilité des jugements* etc. Paris. **S, G,** 53.

Prokhorov Yu. V. (1988), The Bertrand paradox. *Enc. Math.*, vol. 1, pp. 370 – 371. Dordrecht.

Schmidt O. (1926), On the Bertrand paradox. *Matematich. Zbornik*, vol. 33, pp. 33 – 40. In Russian.

Seneta E., Parshall Karen H., Jongmans Fr. (2001), 19th century developments in geometric probability etc. *Arch. Hist. Ex. Sci*., vol. 55, pp. 501 – 524.

Sheynin O. (1971), Newton and the classical theory of probability. Ibidem, vol. 7, pp. 217 – 243.

---, (1980), On the history of the statistical method in biology. Ibidem, vol. 22, pp. 323 – 371.

---, (1984), On the history of the statistical method in astronomy. Ibidem, vol. 29, pp. 151 – 199.

---, (1991), On Buniakovsky's work in the theory of probability. Ibidem, vol. 43, pp. 199 – 223.

---, (1994), Bertrand's work on probability. Ibidem, vol. 48, pp. 155 – 199.

---, (1998), Statistical thinking in the Bible and the Talmud. *Annals of Science*, vol. 55, pp. 185 – 198.

Shreider Yu. A., Editor (1966), *The Monte-Carlo method. The method of statistical trials*. Transl. D. M. Parkyn. Oxford. Initially published in Russian (1962).

Simpson T. (1740), *Nature and laws of chance*. London.

--- (1757), On the advantage of taking the mean etc. In author's book *Misc. Tracts on some curious … subjects* … London, pp. 64 – 75. This is an extended version of a memoir of 1756 of the same name.

Todhunter I. (1865), *History of math. theory of probability*. New York, 1949, 1965.

Whitworth W. A. (1959) *Choice and chance*. New York. Reprint of edition of 1901. Initially published 1867.

VI

Density curves in the theory of errors

Arch. Hist. Ex. Sci., vol. 49, 1995, pp. 163 – 196

To the memory of Churchill Eisenhart

1. Introduction

1.1. Aim and scope of this paper. My question is: What kind of densities were introduced into the theory of errors[1], especially as laws of error? The purposes of the stochastic theory of errors are not restricted to determining appropriate densities (they usually even remain unknown): we still ought to estimate the true value(s) sought (see § 1.2) and the plausibility of our conclusions. The study of the history of these problems is beyond the scope of my paper. Some findings concern De Morgan (1864) and Lüroth (1875).

From the 1750-s to the turn of the 19th century the theory of errors is known to have been a most important branch of probability. Thus, Poincaré (posth. publ. 1921, p. 343) indicated that [in probability] *La théorie des erreurs était naturellement mon* [his] *principal but*. Much later Lévy (1925, p. VII) stated that without the theory of errors his main contribution on stable laws *n'aurait pas de raison d'être*. [Actually, the theory of errors has no connection with stable laws.]

In § 2 I discuss earlier developments mostly belonging to the 18th century. In § 3 I deal with the history of the normal law and in § 4 I dwell on its generalisations and especially upon mixtures of normal frequencies. I devote § 5 to Lévy's (vain) attempts to introduce stable laws into the theory of errors whereas § 6 is given over to distributions which do not belong to laws of error. Finally, in § 7 I describe the stages in the history of treating observations.

In § 3, I leave aside the limiting case, i. e., the central limit theorem (CLT)[2] so that the normal law is understood there as the density of observational errors in their own right. Neither do I review its numerous replacements which Knobloch (1985; 1990) non-technically studied but I dwell on the opposition to the universality of the normal law. The authors whom I discussed in § 4 hardly (but sometimes tacitly) believed in that universality.

1.2. Some explanations. I use the notation introduced by Gauss:

$[ab] = a_1b_1 + a_2b_2 + \ldots + a_nb_n,$
$[bc1] = [bc] – [ab][ac]/[aa],$
$[cd2] = [cd1] – [bc1][bd1]/bb1]$ etc.

Two terms should be discussed. *First term*. When have mathematicians and natural scientists realised that a random error was a random variable? The latter concept was effectively used in probability theory from its infancy, for example, when discussing possible gains (losses!) in the Genoise lottery and Graunt's mortality table. [Beyond the theory randomness was actually mentioned in the Old Testament and the Talmud (Sheynin 2017, pp. 16 – 17)].

However, only Poisson (1837, pp. 140 – 141) introduced a special, though obviously a provisional term, *chose A* for that notion. The present term, or rather its Russian equivalent, random magnitude, appeared at the end of the 19th century (Sheynin 1989, p. 350, note 17) and one of the first to introduce it (in a barely noticed source) was Vasiliev (1885, pp. 127 – 131). On p. 133 he also stated that *random errors have all the properties of random magnitudes*. In Russia, for a long time both that present term and *random magnitude* had been in use but finally *magnitude* regrettably won.

Simpson (1756, 1757) effectively introduced random errors but the theory of errors somehow avoided that term.

That errors are unavoidable and finite was always self-evident, but neither Ptolemy nor Al-Biruni said anything about the other properties of ordinary errors which are easily formulated in terms of frequency curve: it should be unimodal and even and therefore decrease at both sides of the mode. It was Galileo who indirectly and without using any precise terms formulated these statements[3], see his *Dialogo* on the two chief systems of the world (Hald 1990, pp. 149 – 150).

Second term: true (real) value of a measured constant. Fourier (1826, p. 534) defined it as the limit of the arithmetic mean as the number of observations tends to infinity. He had predessesors. Lambert (1765b, § 3): if

Gleich große Abweichungen auf beiden Seiten gleich möglich sind, … das Mittel … dem wahren desto näher kommen müsse je mehr der Versuch ist wiederholt worden.

I (1994, pp. 257 – 258); 2007) discussed this subject at length (and mentioned Laplace and von Mises). Modern authors introduced that term independently from each other and neither recalled Fourier.

Cameron (1982, p. 546) stated that true value is unknowable and leads to confusion. Instead, he advocated *correct value*, the one given by the appropriate standard. This is obviously insufficient so that correct values do not exist for example in positional astronomy or geodesy.

2. The first laws of error

The origin of the theory of probability was occasioned by the study of games of chance. Population statistics came next. Combination of observations, properly speaking, lagged behind and became prominent in the second half of the 18th century with the introduction of laws of error[4].

2.1. Boscovich. He considered a discrete uniform distribution of errors and calculated the probability of the error of their sum in several specific cases. He had not published his calculations and the date of his manuscript remains unknown (Sheynin 1973a, pp. 279 – 280; 1973b, pp. 317 – 318). Stigler (1984) discovered that Simpson and Boscovich had met in 1760 and concluded (p. 619) that *their contact would seem to suggest that Boscovich' manuscript was posterior to 1760 and may have been suggested by Simpson*.

Anyway, Boscovich (1758, § 481, also see § 479), Sheynin (1973b, p. 321) reasoned about particles of matter, *moving together with practically the same velocity* and rather obscurely stated that the sum (not the mean!) of *n irregular inequalities* between the velocities tended to zero as n tended to infinity. That the velocities were essentially different was not yet known. His mistake was possibly evoked by his earlier calculations mentioned above. Kepler (Sheynin 1973c, p. 120) made the same mistake earlier and even in the 20th century Helmert (1905, p. 604) warned his readers against it.

Farebrother (1990) considered in detail the encounter of Simpson and Boscovich.

2.3. Lambert. He (1760, § 303), see Sheynin (1971, pp. 250 – 251) published an extremely important contribution: he described for the first time ever the principle of maximal likelihood for an unspecified unimodal frequency curve. Elsewhere he (1765a, §§ 429 – 430), see Sheynin (1971, p. 253) introduced a definite frequency, a semi-circumference justifying its choice by speculative reasoning and using it for studying (again, speculatively) the properties of observational errors.

2.4. Daniel Bernoulli. He (1778), also see Sheynin (1972, pp. 47 – 48), recommended the estimation of the constant sought by a weighted arithmetic mean; or, rather, without citing Lambert he introduced the principle of maximal likelihood which led to him to posterior weighting. Accordingly, he stood in need of a certain law of error and mentioned a semi-ellipse and a semi-circumference but finally chose an arc of a parabolic curve.

Back in 1769, Daniel had sent a preliminary version of his work to J. (obviously, Johann III) Bernoulli who described it (after Daniel's death) in 1789, see Todhunter 1865, § 825; Sheynin (1972, p. 47). There, Daniel only thought about the semi-ellipse and semi-circumference but not about maximal likelihood[6].

2.5. Lagrange. He (1776)[7] greatly enlarged on the work of Simpson (§ 2.2) but did not refer to him[8]. I (1973a, pp. 282 – 286) have dwelt on a portion of Lagrange's contribution and noted that he, unlike his predecessor, had studied several continuous distributions of which only the triangular, and to a certain extent the uniform law were practically important. Eisenhart (1983, p. 535) remarked that Lagrange (§ 18) was obviously the first who had introduced the term *la courbe de la facilité des erreurs*, cf. § 2.6.

2.6. Laplace. He (1774), see also Sheynin (1977, pp. 4 – 5) offered a speculative derivation of the density

$$\varphi(x) = {m/2}\exp[-m|x - \alpha|],\ m > 0. \qquad (1)$$

I replaced Laplace's $|x|$ by $|x - \alpha|$[9]. Using this law, Laplace chose various estimators for the true value of the constant sought. He did not say (but hardly failed to notice) that the median of that distribution corresponded to the maximal probability of the occurrence of a series of observations $x_1, x_2, \ldots, x_n$. Indeed, if

$$|x_1 - \alpha| + |x_2 - \alpha| + + \ldots + |x_n - \alpha| = \min,$$

then α is the median.

Elsewhere Laplace (1781), see also Sheynin (1973a, pp. 293 – 294), introduced three different continuous laws of error including the known uniform[10] and triangular distributions. The last one was

$$y = \frac{1}{2a} \ln a/|x|, \; 0 < |x| \le a \qquad (2)$$

or, I would replace $|x|$ by $|x - \alpha|$.

Laplace derived this distribution upon solving an interesting problem involving discrete random variables (not errors of observation) and passing to the continuous case (Sheynin 1973a, pp. 294 – 298). He recommended this curve as a law of error even though its use was apparently difficult for small values of $|x|$ or $|x - \alpha|$.

On p. 396 Laplace mentioned *quantités variables* thus making the fist step in the long way to the introduction of a term for *random variable*, see § 1.2. Actually following Lagrange (§ 2.5), he (p. 396ff) also used two special expressions, *loi de possibilité* and *loi de facilité*. He showed no preference for either but later in life he gradually replaced both by *loi de probabilité*[11].

2.7. Gauss. It was he (1809, § 175; 1821, p. 193; 1823, § 4) who, properly speaking, defined density and thus ultimately subordinated the stochastic theory of errors to probability:

1. *Die Wahrscheinlichkeit, welche irgend einem Fehler* Δ *beizulegen ist, wird daher eine Funktion von* Δ *ausgedrückt, welche wir mit* $\varphi(\Delta)$ *bezeichnen wollen*.

2. *Die Funktion, die Wahrscheinlichkeit der Fehler dargestellt …*

3. *Bezeichnet man mit* $\varphi(x)$ *die relative Häufigkeit des Totalfehlers x bei einer bestimmten Gattung von Beobachtungen, so wird wegen der Stetigkeit der Fehler die Wahrscheinlichkeit eines zwischen den unendlich nahen Grenzen x und x + dx liegenden Fehlers* = $\varphi(x)dx$ *zu setzen sein*.

2.8. The opposition. The introduction of densities met with opposition. As late as 1888 Bertrand (p. 212) argued that the probability of an error does not remain constant[12] and the errors themselves are not independent; and that (p. 222), because of systematic influences and blunders, *les résultats* [of observation] *échappent à toute théorie*. Bertrand criticized everything possible without proposing anything instead but at least he clearly stated the difficulty involved in the application of mathematical considerations.

Not without reason, Bessel (1838) stated to the contrary, in the very first sentence of his memoir, that the probability of observational errors may be assumed to depend on their magnitude.

Bertrand (1888, p. 181) also disapprovingly remarked that, even if observations were distributed according to a certain law, the errors of their functions will obey other laws. True, but not crucial or new, see § 3.5, Item 1.

Sampson (1913) was perhaps one of the last astronomers to doubt the validity of densities. The error of an angle, he (p. 164) asserted, may be a function of the angle itself,

But it may also be a function of the temperature, the hour of the day … and so forth, and all of these are equally ignored.

Exactly! Sampson (p. 173) recognized, however, that

There is no objection to regard the actual observations as a mere selection, taken at random or on any system not deliberately unrepresentative, from an infinite sequence which it is open to take, and so as generally yielding results that are not representative of the whole.

He meant that a series of observations was a random sample from an imagined general population. And even by the end of the 19th century Markov (Sheynin 2017, p. 234) meant the same. Even Laplace, in the very beginning of § 23 of his *Théor. anal. prob*., mentioned *numerous not yet made observations*.

3. The normal law

I repeat (Sheynin 1984a, p. 183, Note 47) that the term *normal law* (or *curve*) began appearing in 1873 (Peirce, noticed by Kruskal in 1978) and that its definitive introduction was due to Pearson (1894).

3.1. The first occurrences. Nikolaus Bernoulli, who considered the sex ratio at birth, was the first to derive indirectly the normal density in the limiting case, see his letter of 1713 to Montmort (1713, pp. 388 – 394; also Sheynin (1968, only in its reprint of 1970, p. 232; 1970a, pp. 201 – 202).

In 1733, while pursuing the same goal, De Moivre proved what is now called the De Moivre – Laplace limit theorems and thus arrived at the normal law[13]. In 1770 – 1771 Daniel Bernoulli followed suit but had not derived the necessary result (Sheynin 1970b).

3.2. Gauss. The normal density in its own right first appeared in his *Theoria motus* (1809) and he had been applying the principle of least squares from 1794 or 1795[14]. But when did he derive the normal distribution? Certainly earlier than in 1809, since the published text of the *Theoria motus* is Gauss'own translation from its German original, but not before 1797 when Gauss (1821, p. 193)

Diese Aiufgabe [adjustment of indirect observations] *nach den Grundsätzen der Wahrscheinlichkeitsrechnung zuerst untersuchte* [and] *fand sobald*

that it was impossible to derive the most probable values of the unknowns without determining the law of error.

3.3. Adrain. He (1808, actually 1809) publicly justified the normal law at about the same time as Gauss, but his mathematics was unacceptable and his paper remained unknown for about sixty years. C. Abbe (1871) seems to be the first who took notice of Adrain whereas its latest description is apparently Dutka (1990).

I myself (1965) had not quite satisfactorily discussed Adrain`s paper, but I paid attention to the geodetic applications of the normal law as considered by him and described his two later papers of 1818 which concerned the determination of the size and the figure of the Earth and I cited his first publication. The later papers became more or less known[16], although Strasser (1957) had not mentioned them. Quite a few scholars including Maxwell (§ 3.4, Item 7) actually followed one of Adrain's (unsuccessful) proof of the normal law but had not repeated his mistakes.

3.4. The success. Gauss rejected his first justification of the MLSq because of two circumstances. He preferred to substantiate it independently from the law of error (1823, § 17) and to estimate the plausibility of the obtained results by an integral rather than a differential measure of precision as he stated in his correspondence (Sheynin 1979, pp. 40 – 41 and 46). Gauss (1809, § 177) was also dissatisfied with the underlying postulate of the arithmetic mean, see also Sheynin (1994, p. 277). That mean and therefore the normal law only emerged as the sole law of error and moreover under ideal conditions.

For many decades Gauss' mature ideas were largely ignored. First, the mathematics in the *Theoria motus* was sufficiently simple as opposed to the hardly understandable exposition of 1823. Second, the exponential function was handy. Third, the normal law more or less successfully described the scatter of observational errors, no doubt because of the CLT which, moreover, made the Gaussian law respectable. Fourth, the normal law became entrenched in natural science. Fifth and last, it was stable (for the time being, not in Lévy's sense, see § 5). I illustrate some of these points.

1. Laplace (1818, p. 536) recognized the normal law in its own right:

Cette supposition [of its existence], *la plus naturelle et la plus simple de toutes, résulte de l'emploi du cercle répétiteur dans la mesure des angles des triangles*[18].

2. De Morgan (1864, p. 409) made a curious although not altogether correct remark:

The peculiar pliability of the normal law is so dexterously used that we hardly know how much any result is independent of it.

3. Crofton (1870, p. 176) stated that

This law of error seems in our day to have been adopted by general consent … as expressing the law of frequency of single errors of observation.

4. Repeating G. Lippman's oral jocular remark, Poincaré (1912, p. 171) noted that

Les expérimentateurs s'imaginent que c'est un théorème de mathématiques, et les mathématiciens que c'est un fait expérimental[19].

5. Quetelet (1846) studied the distribution of the chest measurements of several thousand Scottish soldiers and stated that it was approximately normal. Late in life, in 1873, he (Sheynin 1986, p. 313) maintained that the normal law was *une de plus générales de la nature animée*. At the same time Quetelet did not at all deny the existence of other densities (§ 3.5, Item 3) which (together with many other instances) goes to show that he was a happy-go-lucky author.

6. Quetelet (1846, pp. 12 – 24) published a few letters of 1856 from Bravais, a natural scientist known for two papers of 1838 and 1845. Bravais had much to say about asymmetric densities (§ 3.5, Item 4) but he concluded (Ibidem, p. 422) that

Pour exemple, dans l'astronomie et la géodésie de precision, chaque résultat est toujours déduit d'un assez grand nombre de mesures, et rentrant dans la classe des moyennes, son erreur probable [?] *doit suivre* [the normal law].

Only a feeling of that law is discerned here. He (p. 418) hesitatingly (*si je ne me trompe*) based his argument on the law of large numbers.

7. In 1860, Maxwell derived his celebrated normal distribution of molecular velocities appropriated to a gas in equilibrium. In astronomy, in the mid-19th century, the stellar motions were thought to be also normally distributed[20].

8. It was Bessel (1838, § 7) who proved that the normal law was stable (to repeat; for the time being, not in Lévy's sense, see § 5). Since then, several authors confirmed his obviously forgotten or even unnoticed result. However Czuber (1890) forestalled him. Then he (1903, p. 23) referred to Pizzetti and Lindelöf and proved that

Wenn die unabhängigen Beobachtungsfehler[21] *… einzeln das* [normal] *Gesetz befolgen, so interliegt eine homogene lineare Funktion … derselben einem Gesetz der gleichen Form*.

He called this proposition *ein Hauptsatz* of the theory of errors.

Sampson (1913, p. 170) repeated the proof of the same proposition and cited two predecessors, Father Willaert and d'Ocagne. He attached some importance to *the reproduction of form* but remarked that it was difficult to say *what is to be understood by the same form*.

Indeed, the Cauchy distribution reproduces itself, a fact proved by Poisson (who should replace Cauchy as the author of that distribution), but Sampson was able to confirm this only for the simplest case of

$$\frac{a}{\pi(a^2+x^2)},$$

that is, for $a = 1$. He made a mistake when studying the general case (understandably, without using characteristic functions).

3.5. The opposition. I discuss the actual denial of the universality of the normal law irrespective of the consistent or hesitating nature of the appropriate statements and, just as in § 3.4, I have to go beyond the theory of errors. Before Newcomb (Item 5) the opposition was weak.

1. Bessel (1818, p. 279) was apparently the first to hint at the disagreement between the normal law and the (Bradley's) astronomical observations:

Die Übereinstimmung … ist überall so gut, wie man dies überhaupt … erwartet kann. Aber die schwersten Fehler, die die gewohnten Grenzen weit überschreiten, sind … ein wenig häufiger, als die Differenz zu unbedeutend, als dass sie nicht auch einer noch nicht ausreichend großen Anzahl von Beobachtungen zugeschreiben werden könnte.

However, the *wenig häufiger* should have been accompanied by *wenig seltener* somewhere else, and the number of observations (300, 300 and 470) was large enough. In any case, Bessel presented the data in a generalised form and it is difficult to analyse them thoroughly.

Elsewhere he (1838, § 11) presented the same data and added a summary of some of his own observations. This time he asserted without any reservations that the theory was extremely close to empirical evidence. He somewhat illustrated his conclusion by simple numerical considerations, but, what is extremely important, in the same contribution he proved the CLT (understandably, non-rigorously). **He saved his proof**. In 2016 I (**S, G,** 72) described his second self of an impudent hack-worker.

In the same contribution Bessel (§ 2) studied two instrumental errors. In either case the error itself had a non-normal density and its random influence was also non-normal. Worse: it occurred that one of the resulting densities was antimodal, cf. Item 4. Strictly speaking, Bessel did not refute the possible normality of the total error of observation, neither had practitioners heed his indirect warning the less so since Bessel had devoted his memoir to proving the CLT.

2. Bienaymé (1853, p. 313) remarked that

Elle [the *exponentielle*] *n'est qu'un moyen d'approximation très-commode, mais qui pourrait être remplacé par d'autres formules. … Même dans les questions où elle offre le plus de facilité comme approximation, elle donne fréquemment des résultats dont la fausseté est manifeste dès qu'on veut l'employer à des raisonnements un peu complexes au lieu de la tenir ce qu'elle est réelement.*

This is not sufficiently definite but I am inclined to support Heyde & Seneta (1977, p. 88) who indicated that Bienaymé had [at least] stated that the normal density was not *the* law of error.

3. Quetelet (Sheynin 1986, § 5.4) knew that the curve of inclination to crime and to marriage were exceedingly asymmetric. He (1846, p. 168) also noted that

Les exemples où la moyenne ne tombe pas à égale distance des deux valeurs limites, et où la courbe de possibilité perd de sa symétrie, sont assez frequents.

He mentioned the asymmetric deviations of atmospheric pressure at given moments from its mean daily value, cf. Item 4 below[22].

Quetelet's attitude towards the existence of non-normal laws was muddy, witness both his recognition of universal normality (§ 3.4, Item 5) and his statements of 1848 and 1853 on the *loi des causes accidentelles* (Sheynin, Ibidem). Indeed, although he did not specify this mysterious law, he considered it [no less] universal and indicated that its curve can be asymmetric.

4. Quetelet published a few letters from Bravais, see also §3.4, Item 6. Bravais reasonably stated that many errors, instrumental and observational alike, and that at least some physical phenomena as well, do not obey the normal law. Without referring to Bessel (Item 1) he repeated his example of an error having an antimodal density and he also described the same meteorological example as Quetelet (Item 3).

5. Newcomb (1886, p. 343) stated that the cases in which the errors follow the normal law were *quite exceptional* with large errors being much more numerous than they should be[23]. He (p. 345) added that in certain *classes of important observations* the proportion of large errors was so great that *no separation into normal and abnormal observations* [was] *possible* and mentioned his earlier contribution (1882) which I did not see and which apparently had not yet contained any new advice.

Newcomb himself (1886, pp. 359 – 362) quoted an important passage from its p. 382:

That any general collection of observations of transits of Mercury must be a mixture of observations with different probable errors was made evident to the writer by his observations of … 1878.

6. Pearson (1900, p. 353) harshly noted the contemporaneous treatment of astronomical and geodetic observations and target shooting. He mentioned *current textbooks of the theory of errors* and stated that the normal law was usually derived analytically [he obviously meant the CLT] and that the authors *give as a rule some meagre data of how it fits actual observation. … Perhaps the greatest defaulter in this respect is the late Sit George Biddell Airy*.

Then he censured Merriman and (p. 355) maintained that [e]*ven today there are those who regard* [the normal law] as a sort of fetish[24].

4. The normal law modified

I treat the attempts to improve on the normal law. At least some of those who modified the normal density thought that they were introducing a new universal law of error (§ 4.3) but their belief was hardly fulfilled. Any particular distribution (Eisenhart 1983, p. 565)

Is just a model, a simplification close enough *for fruitful application*.

Only one author (§ 4.6) used a frequency curve of the Gram – Charlier Type A, and even he abandoned it later. This curve represents a near-normal distribution appropriate under certain general conditions. It corresponds to the CLT whose discussion is beyond my aim.

4.1. Cournot. He (1843) was the first to discuss the case of a series of observations of unequal precision. He (§ 81) began by studying an appropriate urn problem, then went on to consider astronomical (or geodetic?) observations[35]. Suppose that $n_1, n_2, \ldots$ observations have densities $f_1(x), f_2(x), \ldots$ The density for the entire series will be (§ 81)

$$f(x) = [n_1 f_1(x) + n_2 f_2(x) + \ldots] : (n_1 + n_2 + \ldots). \qquad (1)$$

He did not specify the distinction between the densities but likely thought that they differed only in the values of their parameters of precision. Nor did he say anything about their type but in § 130 he provided a figure of a density which described the scatter of observational errors and resembled a normal curve. Elsewhere (§ 135, Note) Cournot stated that

La forme de la function qui exprime la loi de probabilité, quand il s'agit d'observations aussi précises que celles des astronomes, doit peu s'écarter de celle qui Gauss lui avait primitivement [in 1809] *assignee*.

It is opportune to stress that $f(x)$ corresponds to a mixture of observational samples, i. e., of values of certain random variables rather than to their sum. The normal law is stable (§ 3.4, Item 8) and the sum of normal variables with differing parameters is again normal (with certain parameters) but a mixture of normal laws, as it became called, is not (§ 4.5).

4.2. De Morgan. He (1864) gave thought to generalising the normal law[27]. At first he (p. 418) decided that it will be proper to determine the *average* of every even moment A_{2k} of the unknown density. These moments, he stated, should be finite and tend to vanish with an increasing k. Consequently, the Cauchy distribution which he mentioned but did not name will not do and neither will the normal law. Then, however, De Morgan acquitted the normal distribution by noting that only A_2 and A_4 really mattered. Without further ado he (p. 410) assumed that the law of error was

$$y = \sqrt{c/\pi}(p + qx^2 + rx^4 + ...)\exp(-cx^2). \qquad (2)$$

He introduced the moments $A_0 = 1, A_2, A_4, \ldots$ and wrote out the ensuing equations in $c, p, q, \ldots$ If (De Morgan's actual case) only two of the additional unknowns were needed, A_0 and A_2 should be used to determine them in terms of c. Substituting the estimates thus obtained in the next equality corresponding to A_4, he got a quadratic equation in c, noted that this parameter took real values when

$$3A_2^2 - A_4 \geq 0 \qquad (3)$$

and obtained the appropriate real c.

Relation (3) meant that the excess of the density (2) was not positive:

$$\varepsilon = (A_4 - 3A_2^2):3 \leq 0.$$

Later empirical evidence, e. g., the frequenter occurrence of larger errors, than those prescribed by the normal distribution (§ 3.5, Item 5), suggested that the law of error should have a positive excess[26]. De Morgan (p. 420) however

Presume[d] *that in any law we shall have to represent, large errors are more infrequent*
[less frequent] than according to the general theory. His law (2) was not directly applied; no one even mentioned it although De Morgan was the first to generalise the normal distribution.

He noted that $p > 0$ and $q < 0$ and that for large values of $|x|$ which, as he (p. 421), concluded, was unimportant:

The numerical effect [is] *too small to require attention*. He had *not yet seen a problem* in which an interpretation of a negative probability *was worth looking for*.

Worse and even terrible is to come and I refuse to call De Morgan a mathematician. Extremely strange as it is, he was an eminent logician.

First, he somehow thought that, if the probability of a certain event was 2.5, *it must happen twice with an even chance of happening a third time*.

Then, in 1842, in a letter to John Herschel, De Morgan (Sophia De Morgan 1882, p. 147) stated that *undoubtedly*

$$\sin\infty = \cos\infty = 0, \tan\infty = \cot\text{an}\infty = \mp\sqrt{-1}.$$

Herschel's answer is lost; De Morgan could have destroyed it.

4.3. Newcomb. During 1873 – 1887 several authors (Peirce; Stone, in *Monthly Notices Roy. Astron. Soc*., 1873 – 1874; Glaisher, Ibidem, 1874; Edgeworth, *Phil. Mag*., 1883 and 1887; Newcomb) stated that observations of a given series can obey normal laws with differing measures of precision[28]. They had not mentioned Cournot (§ 4.1). Harter (1977) described their ideas and efforts and I only discuss the work of Newcomb (1886, p. 351) who set out to modify

The usually accepted law in order that it may be applicable to all cases whatever (!).

He (352) rejected a possible analytic approach to the problem under discussion since the *management* of the new density *might … prove inconvenient*. Instead, he adopted *a very probable hypothesis*, viz., that the law of error was *a mixture of observations* [obeying normal laws] with different measures of precision h_i occurring with probabilities p_i:

$$\varphi(x) = \frac{1}{\sqrt{\pi}}[p_1h_1\exp(-h_1^2x^2) + p_2h_2\exp(-h_2^2x^2) + ... + p_nh_n\exp(-h_n^2x^2)].$$

The parameter h of the normal law became a discrete random variable but the new parameters as well as he number n had to be subjectively assigned.

Newcomb next stated that, given observations $x_1, x_2, …, x_m$, the parameter of location α should be determined from the condition

$$\int_{-\infty}^{\infty}(x-\alpha)^2\varphi(x-x_1)\varphi(x-x_1)...\varphi(x-x_1)dx = \min$$

After introducing simplifying assumptions[29], his thus approximated method meant (Hulme & Symms 1939, p. 644) that the value of α corresponded to

$$\varphi(\alpha - x_1)\varphi(\alpha - x_2) \dots \varphi(\alpha - x_m) = \max.$$

It is evident that Newcomb introduced a new loss function and that, in his own words, he followed Gauss.

Pearson (1894) who did not mention Newcomb investigated a related problem, the dissection of abnormal densities into normal curves. He (74) proved that

A curve which breaks up into two normal components can break up in one way, and one way only.

The dissection, however, required the solution of an algebraic equation of the ninth order.

4.4. Lehmann-Filhés. He (1887) modified Newcomb's proposal by assuming that the measure of precision was a continuous random variable with a normal distribution of its own. I doubt that his proposal had any practical use, and in any case no one used it in practice. Ogorodnikov (§ 3.6) complicated matters still more (and did not cite his predecessor).

4.5. Eddington. He (1933, p. 277) quite simply proved that the excess of the Newcomb distribution (§ 4.3) was positive. In particular, this meant that it was not normal[31]. Idelson (1947, p. 307) called Eddington's theorem and a similar proposition by Ogorodnikov (§ 3.6) *one of the most important results of the contemporary theory of errors*.

4.6. Ogorodnikov. He (1928; 1929a) assumed that the parameter of precision had a not necessarily normal density. No wonder that his proposal was never applied either. Ogorodnikov (1928, p. 16) remarked that the

Frequency curves of the stellar motion have a very pronounced positive excess

and explained it by the fact that stars of different spectral types have different average velocities, cf. § 3.4, Item 7. He also noted that no observational series with a negative excess had yet been collected.

In another paper he (1929b) complicated matters even more. He did not cite Pearson (1894).

5. Stable laws (Levy)

Mathematicians began to study stable laws in the 1920s, cf. § 1.1. Such studies proved their worth in economics, and natural scientists have applied them in various branched of science and technology (Zolotarev 1984, pp. 5, 38 – 39, 52 and 55; Barbut 1991, p. 36 – 43).

Lévy was cofounder of their theory and the sole author who argued that they were necessary for establishing the theory of errors anew[32], see his unmethodically compiled contributions (1924; 1925) which I am now trying to systematise[33]. Previous commentators hardly paid attention to his statements about the treatment of observations.

5.1. Random errors. Lévy (1924, p. 51; 1925, p. 278) noted that their mean values were zero, that (1924, p. 50) they were *independantes et très petites* (which was wrong), or, at least (1925, p. 278) that they appeared *comme la somme* of such errors. He (p. 279) also remarked that

Certains savants, en Russie notamment, se sont préoccupes d'étudier le cas où … les lois des probabilités auxqelles obeissent les erreurs partielles ne sont qu'à peu prés indépendantes.

Lévy dismissed this case by stating that it can be reduced to some primary independent variables (errors). He obviously knew nothing about the pioneer work of Markov.

Lévy (1925, pp. 70 – 71, 278) stated that a random error was normally distributed, but he (p. 73) also argued that *elle n'obéira qu'à peu près à la loi de Gauss*. And on p. 279 he concluded

En définitive, l'erreur accidentelle obéit à la loi de Gauss d'autant plus exactement que le conditions [of the CLT] *sont plus exactement vérifiées*.

Late in life Lévy (1970, p. 71) noted however that in 1919 he had only

Un vague souvenir du fait que les erreurs accidentelles obeissent à loi de Gauss.

Strange indeed! But he was mainly concerned with non-normal laws (and even with peculiar stable laws having index $\alpha < 1$, see

§ 5.5), and it likely follows that he was mostly discussing observations corrupted by systematic influences.

5.2. Precision of observations. It can only be comprehensively described by means of the appropriate law of error (1924, pp. 78 – 79; 1925, p. 75 – 76). Correct, but difficult to apply. Those who *prétendent fonder la théorie des erreurs* on the concept of precision were wrong since precision is not a notion première (1925, p. 74).

Accordingly, Lévy (1924, p. 77; 1925, pp. 80 and 284 – 285) disapprovingly mentioned Bienaymé (1853) who had denied the practical importance of the Cauchy distribution since by definition, as he argued, [sound] observations cannot obey it**[34]**, and upheld the estimation of precision by the variance.

In essence, however, Lévy directed his attack against Laplace and Gauss (1823). He mentioned them elsewhere (1924, p. 77) not forgetting Bienaymé either. The *fausseté* of the work of both cofounders of the error theory, as he declared, *aurait dû apparaître lorsqu'en 1853 Cauchy attira l'attention* [to stable laws and in particular to the Cauchy distribution]**[35]**. But where is the connection to Laplace and Gauss?

Lévy somehow conditioned the possibility of plausibly estimating the precision of observations by the existence of a stable law of error (§ 5.5).

5.3. The mean square error. Lévy considered true errors ω rather than deviations from the arithmetic mean and, although he did not say so, his mean square error was of course $\sqrt{[\omega\omega]/n}$. This statistic, as he (1925, p. 75) stated, repeating his earlier pronouncement (1924, p. 52), corresponded to the *idée la plus simple* and its use was *assez naturel*. Again (p. 74), *il semble qu'en effet on ne puisse pas choisir un meilleur paramètre*. But how to determine true errors?

The mean square error provided (1925, p. 27), *faute de mieux* [obviously, when the law of error remained unknown] … *une certaine idée de l'ordre de grandeur de l'erreur* [of its absolute error].

But then he (p. 61) added that other estimators of the type $\sum|\omega_i|^p/n$ with *p ayant une valeur positive quelconque* were *tout aussi bien*. He (p. 78) returned to this point by stating that under the Gaussian law the mean square error was not better than any *paramètre défini d'une autre manière*, but that for near-normal densities it was for some reason important to use exactly this estimator (pp. 78 and 282). He (1925, p. 77) also expressed himself against the introduction of the variance without its justification:

*Une théorie déduite d'axiomes introduits arbitrairement ne saurait avoir aucune valeur***[37]**.

And so, the sample variance is a convenient estimator of precision; nevertheless, other measures of precision can also be used, and in any case (cf. § 5.5) comprehensive estimation is impossible without knowledge of the pertinent law of error. The greatest trouble is that for stable laws with $\alpha < 2$ the variance just does not exist.

Lévy did not refer here to Gauss or Bienaymé. The latter (1853, p. 313) believed that the variance was not

Un élément arbitraire de l'approximation, ni, comme le croyait M. Gauss[38], *une mesure arbitraire de la précision, à laquelle on pourrait substituer toute autre moyenne de puissances de dégre pair. Tout au contraire*, [the variance] *renferme la condition fondamentale*.

Indeed, the variance of a sum of random terms is equal to the sum of the variances of the terms (Gauss 1823, § 18), but Bienaymé (Heyde & Seneta 1977, p. 89) proved that no similar property persisted for these competing estimators of precision. Gauss did not mention this fact.

5.4. A new concept of precision. Lévy (1924, p. 73) proposed to estimate the precision of a random error ω [better: of an observation corrupted by error ω] with $E\omega = 0$ by a *paramètre* that indicated

L'ordre de grandeur de l'erreur à laquelle on doit s'attendre (*en valeur absolue*).

On p. 75 he noted that that parameter was defined to within an arbitrary multiplier and he (p. 78) justified his approach:

Considérer la notion de paramètre de précision come intuitive, c'est admettre qu'on peut définir par un seul nombre [without knowledge of the pertinent density, cf. § 5.5] *les avantages d'une méthode de mesure*.

For the normal law the parameter of the sample mean is $\sqrt{n}$ times less than that of x_i (Lévy 1925, p. 280) whereas its *module de précision* [its weight] $h = 1/\alpha^2$ *est alors n fois plus grand* (Ibidem)[39].

Lévy's innovation makes more sense in the context of stable laws.

5.5. Stable laws. Lévy offered a definition of stable laws in terms of their characteristic functions, but I am much more interested on its corollary (Lévy 1924, p. 69; 1925, p. 258): Given independent and identically distributed errors $\omega_1, \omega_2, \ldots, \omega_n$ and positive numbers $a_1, a_2, \ldots, a_n$ and that there exists such a positive number A that

$$\mathrm{A}^2 = a_1^2 + a_2^2 + \ldots + a_n^2$$

with $[a\omega]/\mathrm{A}$ having the same distribution as the errors ω_i have, then this distribution is stable. Two conditions are additionally imposed by the initial definition of stability (Lévy 1924, p. 70; 1925, p. 255). The first of these is that $0 < \alpha \le 2$ and the second requires that the variance of a stable law is finite if $\alpha = 2$ and infinite otherwise.

If $a_i = 1/n$ and $\mathrm{A} = n^{-(\alpha-1)/\alpha}$ then $[a\omega]/\mathrm{A} = \omega_{\mathrm{mean}}/\mathrm{A}$ has the same distribution as (any) ω_i. For example, if $\alpha = 2$ then ω_{mean} is distributed according to the same law as $\omega_i/\sqrt{n}$ whereas $\alpha = 1$ leads to ω_{mean} having the same density as ω_i. These two cases correspond to the normal law ad the Cauchy distribution respectively.

The importance of stability consists in that (Lévy 1925, p. 78 and 282)

Les moyennes … calculées avec différents systèmes de coefficients ne donneront lieu à des erreurs du même type, et, par suite, ne seront fascilement comparables au point de vue de la précision [which obviously is estimated by the parameters] *que si ce type est stable*.

Suppose that indeed the law of error is stable. What then? If $\alpha = 2$, the MLSq holds (1925, p. 79)[40]:

La loi de Gauss est bien la seule pour laquelle cette méthode s'applique. If $1 < \alpha < 2$ the weight of observation i should be proportional to $a_i^{-(\alpha-1)/\alpha}$ where a_i is the pertinent parameter of precision (1924, pp. 75 – 76; 1925, p. 283). In this case (p. 77; p. 285) the observations, as compared with the MLSq, should be adjusted *avec quelques modifications* (1924). Suppose that a_i = Const and, consequently, that the mean ω has the same distribution as $\omega_i n^{-(\alpha-1)/\alpha}$ (for example, as $\omega_i/n^{1/3}$ for $\alpha = 3/2$. But what next? Introduce posterior weights decreasing towards the tails (Lévy 1924, p. 77). Indeed, since the variance is infinite, large errors are more dangerous than in the previous case ($\alpha = 2$) and their influence should be diminished. However, Lévy did not use his calculations described just above. And posterior weights are subjective and only provide a correction for the asymmetry of the actual distribution.

Two other cases are sill left, $\alpha = 1$ and $0 < \alpha < 1$. If $\alpha = 1$, choose arbitrary weights and calculate the generalised arithmetic mean whose precision, however, will not be higher than that of a single observation. But why bother? Why not choose any single observation and reject all the others?

In the other case, the mean is worse than a single observation. Therefore (Lévy 1924, p. 76; also 1925, pp. 79 and 284):

On peut aussi écarter, dans une proportion déterminée, les plus grands et les plus petits nombres trouvés, et prendre la moyenne des nombres conserves.

And (Lévy 1925, p. 286), *le procéde le plus simple,* is to retain only a half of the observations or even only 1/3.

Thus (Lévy), the law of error should be stable, otherwise the adjustment is fraught with danger. But[41] is it really stable? And how to distinguish between stable laws with $\alpha < 1$ and $\alpha > 1$? Nevertheless, Lévy's advice to trim (suspected) observations can be followed[42] (Elashoff & Elashoff 1978, p. 233):

There is more to gain than to lose by discarding some extreme observations when long tails are possible[43].

The authors adduce references to contributions which offer formal rules of trimming.

6. Densities other than laws of error

In § 6.1 I discuss a function which Laplace effectively introduced in connection with a law of error although never thought of applying it himself. The two other subsections deal with densities related to the estimation of the precision of observations.

6.1. The Dirac delta-function. Laplace (1781) suggested the function

$$\varphi(x) = \frac{1}{2a}\ln\frac{a}{|x|} \text{ or } \frac{1}{2a}\ln\frac{a}{|x-\alpha|} \qquad (1)$$

as a law of error (cf. § 2.6), then applied several methods for choosing estimators for the true value of the constant sought. Thus, he selected such a value for α that, given observations $x_1, x_2, \ldots, x_n$, the integrals of $\varphi(x_1 - x)\varphi(x_2 - x) \dots \varphi(x_n - x)$ over $(-\infty, \alpha]$ and $[\alpha, \infty)$ are equal, call this equality (2). Thus, α was a median of the distribution

$$\xi(x) = c\varphi(x_1 - x)\varphi(x_2 - x) \dots \varphi(x_n - x).$$

Laplace also proved that this choice was tantamount to setting

$$\int_{-\infty}^{\infty} |x - \alpha|\, \xi(x)dx = \min.$$

He (p. 480) then set out to justify the choice of the arithmetic mean as the estimator of α and maintained that for density (1) equation (2) indeed led to α being that mean of $x_1, x_2, \dots, x_n$.

Actually, however, Laplace considered a sequence of functions

$$y = f(\beta x) = f(-\beta x) = q \text{ if } \beta x = 0 \text{ and } = 0 \text{ otherwise.} \qquad (3)$$

Here, $\beta \to$ o. Denote $\beta x = t$, then

$$f(t) = q \text{ if } t = 0\ (|x| < +\infty) \text{ and } f(t) = 0 \text{ if } t \neq 0\ (|x| = +\infty) \qquad (4)$$

and of course $f(t)$ over he entire real axis equals some C.

Laplace had not described the relation between sequences (3) or (4) and his function (1). For large values of $|x|$ the logarithmic function decreases slowly and, moreover, the larger is α, the less pronounced is this decrease so that those sequences generalize function (1). Nevertheless, barely anything can be said here about the intermediate values of $|x|$.

Be that as it may, Laplace went on to prove that for $f(\beta x)$ or rather for $f(\beta x - \alpha)$ equation (1) led to $\alpha = x_{\text{mean}}$. I (1975) repeated his proof by applying the interpretation

$$f(t) = \lim(\lambda/\sqrt{\pi})\exp(-\lambda^2 t^2),\ \lambda \to \infty$$

of the Dirac-function[44].

6.2.The chi-squared distribution. 1. Assuming that the errors of geodetic observations were normally distributed (cf. § 3.4, Item 1); Laplace (1818) estimated the measure of precision h of their frequency law in terms of the closings ε_i of the triangles of triangulation. Since h related to an observed angle whereas ε_i corresponded to the sum of three angles,

$$2h[\varepsilon\varepsilon]/3 = [\varepsilon\varepsilon]/3\sigma^2$$

and Laplace in effect proved that this fraction had density χ^2_{n+2} (Sheynin 1977, pp. 40 – 41). Lancaster (1966, p. 120) noted that Laplace had derived this distribution of precision under a Bayesian hypothesis.

2. Unlike Laplace or two later authors (Items 3 and 4) Bienaymé (1852) did not study magnitudes of the type of $[\varepsilon\varepsilon]$. He attempted to determine what is now called joint efficient estimators and in this sense he indirectly arrived at the chi-squared distribution (Lancaster 1966, § 4; Heyde & Seneta 1977, p. 69).

3. E. Abbe (1863) derived that distribution while studying means for revealing systematic influences (Sheynin 1966; Kendall 1971).

4. Helmert (1876) studied densities of sums of ε_i^m of errors ε_i distributed uniformly and normally. On pp. 202 – 205 he established the chi-squared distribution whose formula he first published in 1875 without proof.

5. Pearson (1900) definitively introduced that distribution and applied it for checking whether certain observations had indeed normal densities (§ 3.5, Item 6).

6. A number of frequencies are in a sense connected with the chi-square including the distribution of the mean square error m of n observations obeying the normal law N(0, σ) (Eddington 1933, p. 280):

$$f(x) = x^{n-1}\exp(-\,nx^2/2\sigma^2). \qquad (5)$$

The ε_i involved here are the true errors rather than the deviations from the arithmetic mean. Eddington had not proved his formula, and neither did he adduce the necessary numerical coefficient. Perhaps he was mainly interested in calculating the expectation of the mean square error[45].

Student (1908, pp. 5 – 6) derived distribution (5) much earlier than Eddington and provided the appropriate numerical coefficient and his formula corresponded to residuals rather than to true errors.

6.3. The Student distribution. He derived the distribution

$$y_k = A_{k-2}(1 + x^2)^{-k/2}$$

for k observations with independent and normally distributed errors:

$$A_k = \frac{k(k-2)...4\cdot 2}{\pi(k-1)...3\cdot 1} \text{ or } \frac{k(k-2)...3\cdot 1}{2\pi(k-1)(k-3)...4\cdot 2}$$

for even and odd values of k respectively. He considered the case of one unknown (direct observations) and his number of degrees of freedom was $(k - 1)$.

Lüroth (1875) studied the case of k observations with n unknowns ($n < k$), see Pfanzagl & Sheynin (1996). Here, I only mention that he had tacitly applied the yet not formulated Student – Fisher theorem on the independence of the arithmetic mean and the measure of precision.

7. Appendix: three stages in treating observations

The first stage: Scholars enjoyed full power over their observations. Only a small part of the data might have been used with the rest of them remaining unknown to the scientific community. Ptolemy embodied that attitude whereas Tycho in astronomy and Graunt in population statistics apparently heralded the coming of the new period.

The second stage. All the observations were made, or should have been made generally known but they were treated either subjectively or, at best, without stochastic or statistical interpretation. Thus, Boscovich solved redundant systems of linear algebraic equations by imposing two natural conditions on their residual free terms but was only qualitatively able to justify his method nor did he describe the properties of the solutions thus obtained. Much the same can be stated about Legendre's introduction of the principle of least squares.

The third stage. The treatment of observations is accompanied by statements on the stochastic and/or statistical properties of the obtained solutions.

Compare these three stages with the corresponding stages of the statistical method (Sheynin 1982, pp. 242 – 243).

Acknowledgement. Several authors including myself have studied the material of § 3 and especially of § 2. I tried to avoid repetition but had to refer to many of my own previous papers. I profited from Eisenhart's survey (1983).

This article represents a part of a research programme performed at the Mathematical Institute of the University of Cologne (Professor J. Pfanzagl) with the support of the Axel-Springer Stiftung.

Notes

1. By *introducing* I mean recommending or using or both.

2. *This is a common name for a number of limit theorems … stating conditions under which sums or other functions of a large number of independent or weakly dependent random variables have a … distribution close to the normal distribution* (Prokhorov 1988, p. 83). I use this term in a narrower sense to denote convergence to the normal law itself. The history of the CLT is a separate worthy subject.

3. It is difficult to believe that Kepler had not known as much.

4. A few words on earlier developments are in § 1.2.

5. Cotes defined the *most probable place* … as the weighted mean of the appropriate observations. However, he had not adduced any stochastic notions or explained what exactly did he mean by most probable. His rule was published posthumously in 1722. See Gowing (1983).

6. Lagrange (1776, § 40) introduced another curve of the second degree. Curves of algebraic functions continued to appear in the 19th century as approximations to the normal density. Thus, **a)** Jordan (1877) without rejecting the Gaussian law represented it by an even trinomial which allowed him to introduce his celebrated three-sigma rule for rejecting outliers. **b)** Bertrand (1888, p. 267) attempted to prove that Gauss' second justification of the method of least squares (MLSq) did not abandon the normal law. Since the law of error is even, for small values of $|x|$ it can de described by the function $(a + bx^2)$. He did not consider large errors.

7. The exact date of the publication of Lagrange's memoir is unknown but his correspondence (Sheynin 1972, p. 46) ensures that it is 1775 or, much more likely, 1776.

8. Pearson (1978, p. 587 – 612) minutely reviewed his memoir and, in connection with his attitude to De Moivre, called him a *most disreputable character* and (p. 184) *an unblushing liar and a thorough knave at heart*. See Sheynin (1973a, p. 279) for some curious quotations from De Moivre and Simpson.

9. When determining the free path of a molecule of a given substance, Clausius, in 1858, effectively arrived at distribution (1), an infinitely divisible law (Sheynin 1985, p. 358).

10. More precisely, he considered $y = ax^2 + bx + c$ and later assumed that $a = b = 0$.

11. In Chapter 4 of his *Théor. anal. prob.* (1812/1886) he used *loi de probabilité* (p. 338), *loi des erreurs* (pp. 338, 344 and 345), *courbe des probabilités* (pp. 324, 338 and 345) and *loi de facilité* (p. 309), but on p. 335 he denoted by $\varphi[x/((n + n_1)]$

l'ordonnée corespondante à l'erreur x. Only *loi de probablité* occurs in in his *Essai* (1814/1886). Finally, in the supplements to the *Théor. anal. prob*., *loi de facilité* is found only once (1816, p. 514) whereas, in these supplements (1818 and somewhat later, pp. 531 – 612), *loi de probabilité*, or *loi des erreurs* occurs more than 20 times.

12. *L'assimilation des erreurs fortuites à des tirages au sort dans une urne composée de manière à donner à chaque erreur la probabilité qui lui convient est une fiction, non une réalité.*

13. Todhunter (1865, §§ 335, 336 and 995) had not described De Moivre's achievements well enough. De Morgan (1864), Eggenberger (1894) and Czuber (1899) were likely the first to pay attention to it. Pearson (1924), without mentioning his predecessors, highly praised De Moivre's finding, Then he (1925, p. 201) noted that *In all the French and German works … with which I am* [he was] *acquainted*, De Moivre's results were ascribed to Jacob Bernoulli. De Morgan (1864) correctly stated De Moivre's result but had not cited his pamphlet of 1733.

14. Several of his friends and colleagues including Bessel have testified to this fact. I (2017, pp. 139, 140 and 158) have later described the entire situation including Stigler's impertinent attack on the grand Gauss. Here, I only leave my previously collected unfavourable references to von Zach from Gauss' correspondence (which have no connection with the issue at hand). See letters from Gauss to Olbers of 26.5.1807; to Schumacher of 4.3.1821 and from Schumacher to Gauss of 27.2.1824 (Gauss, *Werke, Ergänzungsreihe*).

15. Adrain's paper was actually published in 1809 (Hogan 1987).

16. Witness Olbers' letter to Gauss of 24.2.1819 (Gauss, *Werke, Ergänzungs-reihe*):

Auch ein Amerikaner schreibt sich … die Erfindung der [MLSq] *zu. Er scheint weder von Ihnen, noch von Legendre's, noch La Place's Arbeiten über diesen Gegenstand etwas zu wissen, sondern beruft sich auf seine, doch erst 1808 herausgekommene Algebra.*

Gauss did not comment.

17. Nevertheless, numerous attempts were made to improve on Gauss by basing the choice of the arithmetic mean on deterministic axioms (Sheynin 1994, pp. 273 – 274). In addition, Bertrand (Note 6) tried to save theoretically the normal law whereas Pizzetti, in 1892 (Czuber 1899, pp. 156 – 157) proposed the function $c\exp[-k^2(x-a)^4]$ as the law of error approximately corresponding to the Gauss postulate.

18. When measuring an angle with a repeating theodolite, the observer can lay it out on the limb several (n) times in succession but read off only the first sighting of the left direction and the last sighting of the right direction and divide the multiple angle thus obtained by n. The influence of the much larger error of reading becomes n times less. Though subject to other considerations, this method equalizes the influence of both errors. However, the other errors did not change and the preconditions for the CLT can still be wanting.

19. Eisenhart (1983, p. 531) found the same remark in Poincaré's *Thermodynamique*.

20. This opinion was eventually refuted: in 1896 Newcomb discovered that the centennial proper motions in declination considerably disagreed with the normal law (Sheynin 1984a, pp. 181 – 183). See § 4.6 for a relevant remark by Ogorodnikov.

21. A restriction mentioned only by Bessel.

22. Another meteorologist (Meyer 1891, p. 32) even declared that, since the relevant densities were asymmetric (which was then known long ago, and not only by Quetelet), *Die Fehlerrechnung ist in der Meteorologie unzulässig*. Pearson (1898) made use of Meyer's material for illustrating his theory of asymmetric curves.

23. Obviously, only one such deviation from normality was impossible. Elsewhere Newcomb (p. 359) mentioned another anomaly, viz., a similar preponderance of small errors but effectively dismissed it (p. 360). Eddington (1933, p. 277) indicated that these phenomena (he did not reject the second one) were accompanied by a (corresponding) *defect of intermediate errors*. Cf. Item 1 in § 3.5.

24. Cf. Pearson's earlier remark (1894, p. 72):

In the case of certain biological, sociological and economic measurements there is a well-marked deviation from the normal shape.

Mosteller (1978, p. 219) provided another example. C. Peirce (1873) had analysed sets of observations and inferred that they conformed to the Gaussian law of error, but in 1929 Wilson & Margaret Hilferty refuted his conclusion.

25. *Les diverses mesures sont prises par divers observateurs avec des instruments différents ou dans des circonstances dissembables* (Cournot, § 132). Bru (p. 153) connected Cournot's problem with Poisson (1837, pp. 291 – 292) who had discussed the treatment of observations made *avec des instruments différents* without expressly distinguishing (as Cournot did, see below) between the corresponding densities.

26. This statement is not appropriate for statistical series in general. It was Pearson (1894, p. 93) who introduced the term *excess* and defined it as

$$\varepsilon = (A_4 - 3A_2^2):3A_2^2.$$

There also he put into scientific circulation the terms *standard deviation* (p. 75) and *normal curve* (p. 72, but see § 3).

27. De Morgan (p. 427) also remarked that Oresme (ca. 1323 – 1382) had *a clear idea of fluxional velocity.* I did not find any similar statement in Clagett's account of Oresme (*Dict. Scient. Biogr.*, vol. 10, pp. 223 – 230).

28. A few words about Edgeworth's memoir (1883). First, he uses a special term, *probability curve*, for the normal densities whereas densities in general are called *facility-curves*. Second, for him, any estimator of error is *evil* (p. 361), a term later adopted in a restricted sense by Newcomb (1886), and he (p. 363) discusses the combination of observations as a problem of maximizing *utility*. Third, Edgeworth (Ibidem) somehow believes that, given densities $f_1(x)$ and $f_2(x)$, it is possible that for *every value of x* integral of $f_1(x)$ taken from 0 to x is larger than a similar integral of $f_2(x)$.

29. Here is what he himself (Harter 1977, p. 127) stated in 1912 about the rejection of outliers:

So little is gained by aiming at complete rigour of method, that almost any modification which will prevent the incongruity of changing the weight [of observations] *per saltum from 1 to 0 at a certain point will do.*

30. Lehmann-Filhés (p. 123) also treated the Bradley observations (§ 3.5, Item 1).

31. Eddington (p. 271) did not justify his strange statement that

The purpose of the theory [of combination of observations] *is to assist scientific investigation and not to answer mathematical conundrums.*

In a note attached to the same page he indicated that his §§ 7 – 10 *concern*[ed] *those minutiae, beloved of the mathematical theorist, which cannot be wholly omitted.* Eddington's paper has only eight sections! And a saying comes to mind: *Nothing is more practical than a good theory*!

In the *Discussion* appended to this article I met, on p. 283, one of my previous heroes, N. R. Campbell. In 1928 he (Sheynin 1994, p. 277) fiercely attacked the theory of errors hardly knowing anything about Gauss' mature thoughts. By 1933, he did not become either more knowledgeable or less combative:

The theory of errors is the last surviving stronghold of those who would reject plain fact and common sense in favour of remote deductions from unverifiable guesses, having no merit other than mathematical tractability.

32. See however Item 8 in § 3.4.

33. Cf. Lévy's own opinion (1970, p. 79) about the first part of his book (1925):

Je la trouve inutilement longue, mais je reste convainçu que les idées qui sont développées sont exactes.

34. *J'ai propose d'appeler* [it] *loi de Cauchy* (Lévy 1970, p. 78). Bienaymé (1853, p. 323) correctly indicated that Poisson (1824) was the first to study it, and that Poisson himself had chosen to disregard it. Cf. Helmert (1876, p. 207n):

Dieser Ausnahmefall entspricht … gar keine thatsächlichen Fehlergesetze.

35. Cauchy himself did not say anything of the sort.

36. On the same page he remarked, apparently with regard to the normal law, that

Il est necessaire de prendre [the variance] *pour qu'on puisse les calculer par les formules habituelles en function les moyennes analogues relatives aux erreurs partielles, et justifier la méthode des moindres carrés.*

37. In general, Lévy was dissatisfied with the axiomatic method, or at least with how it was introduced into mathematics:

a) *La géomètrie d'Euclide et celle d'Hilbert ne valent rien si l'on ne soumet d'abord à la critique du bon sens les axioms qui sont à leur base* (1924, p. 79).

He accused *les disciples d'Hilber*t who, in his opinion, forgot to discuss their axioms and whose work was therefore often *sans aucune valeur au point de vue de la justification des résultats qu'ils croient établir.*

b) He quoted Poincaré's *Dernières pensées*:

On peut les [the axioms] *regarder comme des décrets arbitraries qui ne sont que les définitions déguisées des notions fondamentales*.

This idea, he (1925, p. 12n) maintained, was *complètement perdu de vue par certains disciples d'Hilbert.*

Much later Lévy (1949, p. 55) apparently came to recognize the axiomatic theory of probability: *Kolmogoroff a donné à l'axiomatique du calcul de probabilités une forme qui semble definitive*.

Even in 1854 Boole had anticipated the need to base the theory of probability on axioms. On the other hand, although Poincaré did not say anything on this point, his general attitude (above) is disappointing, and the same is true with regard to Markov (Al. Ad. Youshkevich 1974, p. 125) who had underestimated the role of the axiomatic method in mathematics.

38. On p. 319 he mentioned *l'erreur commise* par Gauss in 1823. Gauss (1809, § 186; 1823, § 6) however twice remarked on his preference for the variance. Thus, in 1823:

Bei der unendlichen Mannigfaltigkeit derartiger Funktionen scheint die einfachste vor den übrigen den Vorzug zu verdienen, und diese ist unstreitig das Quadrat.

39. The expression $1/\alpha^2$ is repeated on p. 281. The reader himself ought to interpret it.

40. He obviously meant the arithmetic mean rather than this method, cf. below. Later Lévy (1929, p. 30) came to regard the MLSq (again, to regard that mean) more favourably: it may also be used if, beginning with some *n*, it *conduit à prendre une valeur d'autant plus exacte que n est plus grand.* He thus apparently allowed the use of that mean in case of stable laws having $\alpha > 1$, but it is difficult and perhaps even impossible to check whether these two conditions are fulfilled.

41. For that matter (Zolotarev 1984, pp. 30 – 31) sufficiently simple expressions for the densities of stable laws are known only for those with $\alpha = 2$, 1 and 1/2.

42. In economics, suchlike trimming existed long ago (Gergonne 1821, p. 189; Cournot 1843, § 122). Bru (Cournot 1843/1884, p. 325) who mentions Gergonne, describes contemporaneous attitude towards trimming.

43. Later Lévy himself (1929, p. 31) voiced a similar opinion without mentioning stability.

44. I have inserted the Russian original (Sheynin 1975) of the text of § 6.1 in the text under preparation of Gnedenko & Sheynin (1978). Gnedenko did not comment but Kolmogorov as co-editor without explanation called it rubbish. Gnedenko slavishly struck off my piece although he, if not Kolmogorov should have noticed that anyway Laplace's innovation was historically very interesting. Later I understood that one of the formulas there (I do not remember which one) made no sense in the language of generalised functions. Much, much worse about Gnedenko is at **S, G** 65.

45. While studying stellar motions and their projections on an arbitrary plane, Newcomb (1902) without justification provided two other distributions connected with the chi-square (Sheynin 1984a, pp. 182 – 183). In 1860, Maxwell obtained that distribution for three degrees of freedom whereas Boltzmann, in 1881, derived it in the general case (Sheynin 1985, pp. 360 and 372 – 373).

References

OC = *Oeuvr. Compl.*

Abbe C. (1871), Historical note on the MLSq. *Amer. J. Sci. Arts*, vol. 1, 411 – 415. Reprint: Stigler (1980, vol. 1).

Abbe E. (1863), Über die Gesetzmäßigkeit in der Verteilung der Fehler bei Beobachtungsreihen. *Ges. Abh.,* Bd. 2. Hildesheim, 1989, 55 – 81.

Adrain R. (1808 [actually 1809]), Research concerning the probabilities of the errors which happen in making observations. Reprint: Stigler (1980, vol. 1).

Barbut M. (1991), Note sur les moyennes de variables aléatoires. In *Moyenne, milieu, centre*. Editors: Jacqueline Feldman et al. Paris, 31 – 43.

Bernoulli Daniel (1778), The most probable choice between several discrepant observations etc. *Biometrika*, 1961. Reprint: *Studies* (1970, 157 – 167). Initially in Latin.

Bertrand J. (1888), *Calcul des probabilités.* Paris. Reprints: 1907, and 1970, 1972, New York.

Bessel F. W. (1818), *Fundamenta astronomiae* etc. German translation of pertinent piece: Schneider (1988, 277 – 279).

--- (1838), Untersuchung über die Wahrscheinlichkeit der Beobachtungsfehler. Reprint: Bessel (1876, Bd. 2, 372 – 391).

--- (1876), *Abhandlungen*, Bde. 1 – 3. Leipzig.

Bienaymé I. J. (1852), Sur la probabilité des erreurs d'après la méthode des moindres carrés. *J. math. pures appl.*, sér. 1, t. 17, 33 – 78. Also *Mém. pres. Acad. Sci. Inst. France*, sér. 2, t. 15, 1858, 615 – 663.

--- (1853), Considérations à l'appui de la découverte de Laplace etc. *C. r. Acad. sci. Paris*, t. 37, 309 – 324. Reprint (1867), *J. math. pures appl.*, sér. 2, t. 12, 158 – 176.

Boscovich R. (1758, Latin), *Theory of natural philosophy*. Latin – English edition. Chicago – London, 1922.

Cameron J. M. (1982), Error analysis. In Kotz & Johnson (1982 – 1988, vol. 2, 545 – 581).

Cournot A. A. (1843), *Exposition de la théorie des chances et des probabilités.* Paris. Reprint: Paris, 1984. Editor B. Bru. **S, G,** 54.

Crofton M. W. (1870), On the proof of the law of error of observations. *Phil. Trans. Roy. Soc.*, vol. 160, 175 – 187.

Czuber E. (1890), Zur Theorie der Beobachtungsfehler. *Monatsh. Math.-Phys.*, Bd. 2, 459 – 464.

--- (1899), *Die Entwicklung der Wahrscheinlichkeitstheorie und ihrer Anwendungen. Jahresber. Deutsch. Math. Verein.*, Bd. 7, No. 2, 1 – 279.

---, (1903), Über einen Satz der Fehlertheorie etc. Ibidem, Bd. 12, 23 – 30.

De Morgan A. (1864), On the theory of errors of observation. *Trans. Cambr. Phil. Soc.*, vol. 10, 409 – 427.

De Morgan Sophia Elizabeth (1882), *Memoir of Augustus De Morgan.* London.

Dutka J. (1990), Adrain and the MLSq. *Arch. Hist. Ex. Sci.*, vol. 41, 171 – 184.

Eddington A. S. (1933), Notes on the MLSq. *Proc. Phys. Soc.*, vol. 45, 271 – 287.

Edgeworth F. Y. (1883), The MLSq. *Lond., Edinb. Dubl. Phil Mag. & J. Sci*, ser. 5, vol. 16, 360 – 375. Reprint in author's *Writings in probability, statistics and economics*, vols. 1 – 3. Editor, C. R. MacCann, Jr. Cheltenham, 1996. See vol. 2, 1 – 16.

--- (1887), On discordant observations. Ibidem, vol. 23, 364 – 375. Reprint: Ibidem, vol. 1, 256 – 267.

Eggenberger J. (1894), Beiträge zur Darstellung des Bernoullischen Theorems etc. *Mitt. Naturforsch. Ges. Bern*, No. 1305 – 1334, 1893, 110 – 182. Separate edition: Berlin, 1906.

Eisenhart Ch. (1983), Law of error, pts 1 – 3. In Kotz & Johnson (1982 – 1988), vol. 4, 530 – 566).

Elashoff Janet D., Elashoff R. M. (1978), Effects of errors in statistical assumptions. In Kruskal & Tanur (1978, vol. 1, 229 – 250).

Farebrother R. W. (1990), Further details of contacts between Boscovich and Simpson in June 1760, *Biometrika*, vol. 77, 397 – 400.

Fourier J. B. J. (1826), Sur les résultat moyens etc., OC, t. 2. Paris, 1890, 525 – 545.

Gauss C. F. (1809), *Theoria motus*, extract. German translation: Gauss (1887, 92 – 117).

--- (1811), Disquisitio … German translation: Gauss (1887, 118 – 128).

--- (1821), Selbstanzeige of Gauss (1823), German translation: Gauss (1887, 190 – 195).

--- (1823), Theoria combinationis. German translation: Gauss (1887, 1 – 53). English translation by G. W. Stewart. Philadelphia, 1995.
--- (1887), *Abhandlungen zur Methode der kleinsten Quadrate*. Hrsg. A. Börsch, P. Simon. Vaduz, 1998.
--- (1975 – 1987), *Werke, Ergänzungsreihe*, Bde. 1 – 5. Hildesheim. Correspondence of Gauss. With Bessel: Bd. 1, 1975; with Gerling: Bd. 3, 1975; with Olbers: Bd. 4, 1976; with Schumacher: Bd. 5, 1975.
Gergonne J. D. (1821), Dissertation sur la recherché du milieu le plus probable etc. *Annales math. pures appl*., t. 12, No. 6, 181 – 204. Published anonymously. Stigler (1976) established author.
Gnedenko B. V., Sheynin O. (1978, Russian), Theory of probability. Chapter in *Math. of the 19th century*, vol. 1. Editors, A. N. Kolmogorov, A. P. Youshkevich. Basel, 1992, 2001, pp. 211 – 288.
Gowing R. (1983), *Roger Cotes – natural philosopher*. Cambridge.
Hald A. (1990), *History of probability and statistics … before 1750*. New York.
Harter H. L. (1977, date of introduction), *Chronological annotated bibliography on order statistics*, vol. 1. Published by the U. S. Air Force etc.
Helmert F. R. (1875), Über die Berechnung des wahrscheinlichen Fehlers etc. *Z. Math. Phys*., Bd. 20, 300 – 303.
--- (1876), Über die Wahrscheinlichkeit der Potenzsummen der Beobachtungsfehler etc. Ibidem, Bd. 21, 192 – 218.
--- (1905), Über die Genauigkeit der Kriterien des Zufalls etc. *Sitz.-Ber. Kgl. Preuss. Akad. Wiss. Berlin*, Phys.-math. Kl., Hlbbd. 1, 594 – 612.
Heyde C. C., Seneta E. (1977), *I. J. Bienaymé*. New York.
Hogan E. R. (1987), R. Adrain: American mathematician *Hist. Math*., vol. 4, 157 – 172.
Hulme H. R., Symms L. S. T. (1939), The law of error and the combination of observations. *Monthly Notices Roy. Astron. Soc.*, vol. 99, 642 – 649.
Idelson N. I. (1947), *Sposob naimenshikh kvadratov* (MLSq). Moscow.
Jordan W. (1877), Über den Maximalfehler einer Beobachtung. *Z. Vermessungswesen*, Bd. 6, 35 – 40.
Kendall M. G. (1971), The work of E. Abbe. *Biometrika*, vol. 58. Reprint: *Studies* (1977, 331 – 335).
Knobloch E. (1985), Zur Grundlagenproblematik der Fehlertheorie. In *Festschrift für Helmuth Gericke*. Hrsg. M. Folkerts et al. Stuttgart, 561 – 590.
--- (1990), The hypothetical nature of mathematical proofs etc. *Proc. Intern. Symp. Math. & Theor. Phys., 1989*. Symp. Gaussiana, ser. A. Toronto, 86 – 123.
Kotz S., Johnson N. L., Editors (1982 – 1988), *Enc. of statistical sciences*, vols. 1 – 9. New York.
Kruskal W. (1978). Formulas, numbers, words etc. Reprint (1981): *New directions for methodology of social and behavioural science* etc., 93 – 102. Editor, D. Fiske. San Francisco.
Kruskal W., J. M. Tanur, Editors (1978), *Intern. Enc. of Statistics*, vols. 1 – 2. New York – London.
Lagrange J. L. (1776), Sur l'utilité de la méthode de prendre le milieu etc. OC, t. 2. Paris, 1868, 173 – 234.
Lambert J. H. (1760), *Photometria*. Augsburg. Incomplete German translation: *Ostwald Klassiker*, No. 31 – 33, 1892.
--- (1765a), Anmerkungen und Zusätze zu practischen Geometrie. In author's *Beyträge zum Gebrauche der Math.* etc., Teil 1. Berlin, pp. 1 – 313.
--- (1765b), Theorie der Zuverlässigkeit der Beobachtungen und Versuche. Ibidem, 424 – 488.
Lancaster H. O. (1966), Forerunners of the Pearson chi-square. *Austr. J. Stat.*, vol. 8, 117 – 126.
Laplace P. S. (1774), Sur la probabilité des causes par les événements. OC, t. 8. Paris, 1891, 37 – 65.
--- (1781), Sur les probabilités. OC, t. 9. Paris, 1893, 383 – 485.
--- (1812), *Théorie analytique des probabilités*. OC, t. 7. Paris, 1886.
--- (1814), *Essai philosophique* etc. Ibidem, separate paging. English translation: New York, 1995.

--- (1816, 1818, and somewhat later), Supplements to Laplace (1812). OC, t. 7, 497 – 530; 531 – 580, 581 – 616.
Lehmann-Filhés R. (1887), Über abnorme Fehlerverteilung etc. *Astron. Nachr.*, Bd. 117, 121 – 132.
Lévy P. (1924), La loi de Gauss et les lois exceptionnelles. *Bull. Soc. math. France*, No. 52, 49 – 85.
--- (1925), *Calcul des probabilités*. Paris.
--- (1929), Sur quelques travaux relatifs à la théorie des erreurs. *Bull. sci. math.*, sér. 2, t. 53, 11 – 32.
--- (1940), Les fondements du calcul des probabilités. *Dialectica*, t. 3, 55 – 64.
--- (1970), *Quelques aspects de la pensé d'un mathématicien*. Paris.
Luroth J. (1875), Vergleichung von zwei Werthen des wahrscheinlichen Fehlers. *Astron. Nachr.*, Bd. 87, 209 – 220. See also Pfanzagl, Sheynin (1996).
Meyer Hugo (1891), *Anleitung zur Bearbeitung meteorologischer Beobachtungen* etc. Berlin.
Montmort P. R. (1708, 1713), *Essay d'analyse sur les jeux de hazard. Paris.* New York 1980.
Mosteller F. (1978), Nonsampling errors. In Kruskal, Tanur (1978, 208 – 229).
Newcomb S. (1882), Discussions and results of observations etc. *Astron. Papers Amer. Ephemeris*, vol. 1, 363 – 487.
--- (1886), Generalized theory of the combinations of observations etc. *Amer. J. Math.*, vol. 8, 343 – 366. Reprint: Stigler (1980, vol. 2).
--- (1902), Statistical relations among the parallaxes and the proper motions of the stars. *Astron. J.*, vol. 22, No. 21, 165 – 169.
Ogorodnikoff K. (1928), A method for combining observations etc. *Astron. Zh.* (Moscow), vol. 5, No. 1, 1 – 21.
--- (1929a), On the occurrence of discordant observations etc. *Monthly Notices Roy. Astron. Soc.*, vol. 88, 523 – 532.
--- (1929b), A general method of treating observations. *Astron. Zh.* (Moscow), vol. 6, 226 – 244.
Pearson K. (1894), On the dissection of asymmetrical frequency curves. *Phil. Trans. Roy. Soc.*, vol. A185, pt. 1, 71 – 110.
--- (1898), Cloudiness. *Proc. Roy. Soc*., vol. 62, 787 – 790.
--- (1900), On a criterion etc. *Phil. Mag*., vol. 50. *Early Stat. Papers.* Cambridge, 1956, 339 – 357.
--- (1924), Historical note on the origin of the normal curve of errors. *Biometrika*, vol. 16, 402 – 404.
--- (1925), James Bernoulli's theorem. Ibidem, vol. 17, 201 – 210.
--- (1978, posthumous publ.), *History of statistics in the 17th and 18th centuries*. Lectures 1921 – 1933. London. Editor E. S. Pearson.
Peirce S. (1873), On the theory of errors of observations. *Rept. Coast Survey U. S.*, 1870, 220 – 224.
Pfanzagl J., Sheynin O. (1996), Forerunner of the *t*-distribution. *Biometrika*, vol. 83, 891 – 898.
Poincaré H. (1896, 1912), *Calcul des probabilités*. Paris. Reprint: Seaux, 1987.
--- (1921, posthumous publ.), *Résumé analytique* [of his own works]. In *Math. heritage of H. Poincaré. Proc. Symp. Indiana Univ. 1980*. Editor F. E. Browder. Providence, RI, 1983, 257 – 357.
Poisson S. D. (1824), Sur la probabilité des résultats moyens des observations. *Conn. des temps* pour 1827, 273 – 302.
--- (1837, 2003), *Recherches sur la probabilité des jugements* etc. Paris. **S, G,** 53.
Prokhorov Yu. V. (1988), Central limit theorem. *Enc. of Math.*, vol. 2, 83 – 87. Dordrecht. This source was initially published in Russian, in 1977 – 1985.
Quetelet A. (1846), *Lettres sur la théorie des probabilités*. Bruxelles.
Sampson R. A. (1913), On the law of distribution of errors. *Proc. Fifth Intern. Congr. Mathematicians*, vol. 2. Cambridge, pp. 163 – 173.
Schneider I., Hrsg. (1988), *Entwicklung der Wahrscheinlichkeitsrechnung* etc. Darmstadt. Collection of fragments almost exclusively in German with some comments.
Seidel L. (1863), Über eine Anwendung der Wahrscheinlichkeitsrechnung etc. *Sitz.-Ber. Bayer. Akad. Wiss*., Bd. 2 for 1863, 320 – 350.

Sheynin O. B. (1965), On the work of R. Adrain in the theory of errors. *Istoriko-matematich. issledovania*, vol. 26, 325 – 336. In Russian.
--- (1966), Origin of the theory of errors. *Nature*, vol. 211, 1003 – 1004.
--- (1968), On the early history of the law of large numbers. *Biometrika*, vol. 55. Reprint: *Studies* (1970, 231 – 239).
--- (1970a), On the history of the De Moivre – Laplace limit theorems. *Istoria i metodologia estestvennykh nauk*, vol. 9, 199 – 211. **S, G,** 111.
--- (1970b), Daniel Bernoulli on the normal law. *Biometrika*, vol. 57. *Studies* (1977, pp. 101 – 104).
--- (1971), Lambert's work in probability. *Arch. hist. ex. sci.*, vol. 7, 244 – 256.
--- (1972), On the mathematical treatment of observations by Euler. In this collection.
--- (1973a), Finite random sums. *Arch. hist. ex. sci.*, vol. 9, 275 – 305.
--- (1973b), Boscovich' work on probability. Ibidem, 306 – 324.
--- (1973c), Mathematical treatment of astronomical observations. Ibidem, vol. 11, 97 – 126.
--- (1975), On the appearance of the Dirac function in Laplace's work. *Istoriko-matematich. issledovania*, vol. 20, 303 – 308. In Russian.
--- (1977), Laplace's theory of errors. *Arch. hist. ex. sci.*, vol. 17, 1 – 61.
--- (1979), Gauss and the theory of errors. Ibidem, vol. 20, 21 – 72.
--- (1982, 1984a, 1984b, 1985), On the history of the statistical method in medicine; in astronomy; in meteorology; in physics. (Four papers.) Ibidem, vol. 26, 241 – 286; vol. 29, 151 – 199; vol. 31, 53 – 95; vol. 33, 351 – 382.
--- (1986), Quetelet as a statistician. Ibidem, vol. 36, pp. 281 – 325.
--- (1989) Markov's work on probability. Ibidem, vol. 39, 337 – 377.
--- (1994), Gauss and geodetic observations. Ibidem, vol. 46, 253 – 283.
--- (2007), True value of a measured constant and thje teory of errors. *Historia Scientiarum*, vol. 17, 38 – 48.
--- (2017), *Theory of probability. Historical essay*. Berlin. **S, G,** 10.
Shoesmith E. (1985), Simpson and the arithmetic mean. *Hist. math.*, vol. 12, 352 – 355.
Simpson T. (1756), An attempt to show the advantage arising by taking the mean etc. *Phil. Trans. Roy. Soc.* for 1755, vol. 49, pt. 1, 82 – 93.
--- (1757), Same title. In author's *Misc. tracts on some curious … subjects* etc. London, 64 – 75. Partly reprinted: Schneider (1988, 221 – 227).
Stigler S. M. (1973), S. Newcomb, P. Daniell and the history of robust estimation, 1885 – 1920. *J. Amer. Stat. Assoc.*, vol. 68. Reprint: Studies (1977, 410 – 417).
--- (1974), Cauchy and the witch of Agnesi etc. *Biometrika*, vol. 61, 375 – 380.
--- (1976), The anonymous Professor Gergonne. *Hist. Math.*, vol. 3, 71 – 74.
---, Editor (1980), *American contributions to mathematical statistics in the 19th century*, vols. 1 – 2. New York. Collection of reprints with separate paging.
--- (1984), Boscovich, Simpson and a 1760 manuscript etc. *Biometrika*, vol. 71, 615 – 620.
Strasser G. (1957), *Ellipsoidische Parameter der Erdfigur, 1800 – 1950. Deutsche geod. Komm. Bayer. Akad. Wiss. München*, Bd. A19.
Student (1908), The probable error of a mean. *Biometrika*, vol. 6, 1 – 25.
Studies (1970, 1977), *Studies in the history of statistics and probability*, vols. 1 – 2. Editors, E. S. Pearson, M. G. Kendall (vol. 1); Sir Maurice Kendall, R. L. Plackett (vol. 2). London.
Todhunter I. (1865), *History of the mathematical theory of probability*. New York, 1949, 1965.
Vasiliev A. V. (1885), *Teoria veroyatnostei* (Theory of prob.). Kazan.
Von Mises R. (1931), *Wahrscheinlichkeitsrechnung und ihre Anwendungen in der Statistik und theoretischen Physik*. Leipzig – Wien.
Youshkevich Al. Ad. (1974), Markov. *Dict. Scient. Biogr.*, vol. 9, 124 – 130.
Zolotarev V. M. (1984), *Ustoichivye zakony i ikh primenenia* (Stable laws and their application). Moscow.

VII

On V. YA. Buniakovsky's work in the theory of probability

Arch. hist. ex. sci., vol. 43, 1991, pp. 199 – 223

I. Introduction

Viktor Yakovlevich Buniakovsky (1804 – 1889) was a mathematician, a member of the Imperial Academy of Sciences (Petersburg) and its vice-president from 1864 to 1889. I describe all of his writings devoted to the theory of probability. However, I do not study in detail his investigations in population statistics, and leave aside his more special contributions on annuities and pensions. Buniaskovsky's other works deal with mathematical analysis and the theory of numbers.

Other publications on my subject do exist [41; 45; 56; 91; 42] but they are too brief.

1.1. From Laplace to Buniakovsky. The Laplacian period in the development of the theory of probability lasted for a long time. Laplace kept to an insufficiently high level of mathematical abstraction and did not prove his results rigorously and for these very reasons he was able to achieve outstanding results in natural science. But probability as a truly mathematical discipline had to be created all over again [78, pp. 179 – 183].

Several mathematicians including Buniakovsky primarily restricted their attention to simplifying Laplace's expositions. Without bearing in mind Poisson, who essentially contributed to this discipline, **I** name Lacroix, Cournot, and De Morgan. Lacroix' book [49] ran into four editions and was translated into German. It was very useful but its mathematical level was not high. The same is true for Cournot's interesting writing [31], also translated into German[1] and recently reprinted in Paris.

De Morgan [34] expounded the theory of probability and its applications, although, unlike Buniakovsky, he did not dwell on population statistics. He offered a clear concept of Laplace's mathematical methods and carried out many transformations in more detail than the Master. In 1838 he published a popular booklet on the same subject.

Chebyshev introduced rigour in the development of the theory of probability. Apart from the proof of one limit theorem, his fundamental contributions began to appear in 1867, when Buniakovsky had almost abandoned the theory. Thus, not without reason, the Russian economist and philosopher Struve [88, p. 1318], who took notice of Buniakovsky's note on linguistics (§ 3.5), called him *a Russian representative of the French mathematical school.*

1.2. Buniakovsky's Works. My list of references includes all of Buniakovsky's known writings on the theory and application of probability. Some of his contributions on annuities and pensions were never published and a few auditory reports on the work of pension funds of which Buniakovsky was a coauthor [22, p. 13]

remain unknown (and possibly unpublished or even lost). In addition, *many* [of his] *notes are scattered in various journals and newspapers* [89, p. 7]. Buniakovsky himself [22, p. 16] indicated that in such sources he had inserted abbreviated versions of his reviews of works submitted to the Academy of Sciences.
Almost complete lists of his known works [22; 59] exist, many of them are included in the Royal Society's *Catalogue of Scientific Papers*.

[1] Chuprov [30, p. 30] was the first to acknowledge the achievements of this author in probability and statistics.

After Buniakovsky's death, the Archive of the Academy received fragments of the unpublished part of his *Lexicon* [3]. He gave permission for them to be shown only to those who would continue his work on this explanatory dictionary. In turning over these materials to the Academy, Vladimir Buniakovsky, a son of the deceased mathematician, made known his father's scientific testament which forbade not only the publication, but even any examination of his manuscripts for which no explicit permission was given[2]. His last will explains the absence, in the Archive, of any other of his previously unpublished papers.

Buniakovsky's versatile scientific activities included his important, even if barely noticeable collaboration in compiling explanatory dictionaries of the Russian language (also see above). This work is partly known [73]; in addition, I point to the Pluchart dictionary (17 volumes, 1835-1841) and to the dictionary compiled by the Academy of Sciences and published in 1847 [35]. Buniakovsky edited the mathematical terminology of both works and, in the second, he also supplemented the mathematical word-list [22, p. 13]. The term *probability* (veroyatnost) if not its mathematical definition, was included in vol. **1** of the dictionary of 1847. The *calculus* (ischislenie) of *probabilities (l. c.)* was described as a science dealing with the *laws of probability.* The non-mathematical word *possible* (statochni) is in vol. **4** of the same source. I mention this fact since Buniakovsky [3, p. 182] usually translated the French (and English) *chance* as *statochnost* (possibility). The Introduction to vol. **1** (p. xv) *with special gratitude* acknowledged the contribution made by several scholars including Buniakovsky. And I note that he edited the part pertaining to the exact sciences in yet another dictionary [38].

2. The Principles of the mathematical theory of probability

Buniakovsky's book [6] is his main contribution to the theory of probability. Here (p. ii) he stated that, while following Laplace, he had sought to simplify its exposition. He also expressed a justified hope that he succeeded in making easier the study of the *Théorie analytique* [52], a classic which *is intelligible* [only] *to very few readers*[3].

2.1. The terminology. There is a view [42, p. 213] that Buniakovsky [6] developed the Russian terminology of the theory of probability. However, such terms as *random variable* (or *random quantity*) or *limit theorem* did not then exist in any language. Thus, he had to introduce only two expressions, *viz., matematicheskoye ozhidanie* (expectation) and *zakon bolshikh chisel* (law of large numbers). The term *veroyatnost* (probability) appeared in 1789, in the Russian translation of Buffon and,

[2] *Zapiski Akad. Nauk*, vol. 63, 1890, p. 205

[3] In a report on his book, then not yet completed, Buniakovsky [5] explained his aim as follows:
J'ai eu d'abord pour but de remplir une des lacunes de notre littérature mathématique. Mon second but, beaucoup plus difficile à atteindre, est de rendre plus abordables les théories delicates, dont le calcul des hasards offre tant d'exemple.

for example, much later, in 1821, in a booklet written by Pavlovsky [68]. True, even in 1836 the author of a popular article [48, pp. 29 and 32] recommended rejection of this word in favour of one or another of two artificial Slavophile constructions. Nevertheless, I doubt that anyone ever took his advice seriously.

The expression *sposob naimenshikh kvadratov* (method of least squares) was used in Russia from 1836 [83]. A. N. Kolmogorov, in 1946, and V. V. Petrov, in 1954, replaced its first word with the synonymous *metod* although the *Great Soviet Encyclopedia* did not then (in 1954) follow suit. Finally, Buniakovsky's term *nraystvennoye* [now: *moralnoye] ozhidanie* (moral expectation, § 2.3) is known today only to historians of probability.

2.2. Probability and the Theory of Probability. In the Introduction [6, p. 3], Buniakovsky indicated that some events were more likely than others and called probability the measure of likelihood. Without naming anyone, Ostrogradsky [65, p. 238] sharply criticized this idea. Gnedenko [40, p. 114] quoted a long passage from Ostrogradsky's article and, as it seems, considered his objection immaterial. I note that although Buniakovsky introduced the notion of probability in an intricate manner, he did not mention likelihood at all in his main text (p. 5).

According to the essence of the matter, Buniakovsky [6, p. i] ascribed the analytical theory of probability to applied mathematics. In addition, following Laplace [78, p. 176], he maintained *(l. c.)* that

The analysis of probabilities considers and quantitatively estimates even such phenomena ... which, due to our ignorance, are not subject to any suppositions.

However, in such cases Laplace always considered estimates of the first approximation suitable only until the data be specified. Moreover, Buniakovsky did not corroborate his statement by any examples and he in effect went back on his word by declaring (p. 364) that

The extensive purposes of the theory of probability which embrace *almost the entire range of man's mental occupations are essentially restricted ... by the lack of data ...*

Later Buniakovsky [13, p. 24] repeated himself almost verbatim. Ellis [37, p. 57] was the first to declare expressly that *mere ignorance is no ground for any inference whatever.* He added: *Ex nihilo nihil.*

2.3. The Moral Expectation. If factor x in the expression for the expectation of a continuous random variable is replaced by $\ln x$ the new quantity will be its moral expectation[4].

Daniel Bernoulli [25; 77, pp. 108-114] made use of moral expectation if not the term itself (which had been introduced earlier) to study the Petersburg paradox, an imaginary game of chance whose investigation by means of mathematical expectation patently contradicted common sense. He also noted that an equal distribution of a given cargo on two ships increases the moral expectation of the freight owner's capital as compared with the transportation of the cargo on a single ship[5].

[4] A similar change is made for discrete random variables.

Buniakovsky [6, pp. 103-122] described Bernoulli's reasoning and proved the validity of his remark. Moreover, his proof can be generalized in respect to several ships [77, pp. 112-113]. It is difficult to say why he did not repeat or at least mention Laplace's substantiation of this more general theorem [52, pp. 444445].

Furthermore, in 1880, Buniakovsky [21] found the most advisable division of freight in two arbitrary parts for the case of unequal probabilities of their loss. He also made use of moral expectation to illustrate one general rule of the application of mathematics to statistics Studying the movement of population, he [12, p. 154] noted that the productive population should be considered separately from the children. His remark was hardly original; however, in concluding his statement, he mentioned the moral expectation (recall that it values an increment of capital depending on the amount of the capital itself) and stated:

Anyone who does not examine the meaning of the numbers with which he performs particular calculations is not a mathematician. Buniakovsky passed over in silence Ostrogradsky's attempt to generalize the concept of *moral satisfaction* [67, pp. 293-294; 78, pp. 170-171].

2.4. Geometric Probabilities. Following Buffon and Laplace, Buniakovsky [6, pp. 137-143] considered two versions of the celebrated problem concerning the *Buffon needle:* a needle falls from above on a number of equally spaced parallel lines (on a grid of congruent rectangles); it is required to determine the probability that the needle intersects a line (a side of a rectangle).

Several writers have generalized this problem by studying the possibility that the needle intersects at least one line (one side), and by replacing the needle by a cylinder of finite width.

Referring to his memoir of 1837 [2], Buniakovsky solved one more problem of this kind. This time he considered the fall of the needle on a system of congruent equilateral triangles and determined the probability that the needle should intersect at least one side of the system. His geometric reasoning, which preceded the compilation of the appropriate integrals was extremely involved and his final result (p. 143), as Markov [57, p. 186] maintained in 1900, was wrong *due to an unfortunate choice of the order of integration which* [in addition] *greatly complicated his calculations.*

I did not find any mistakes in the calculations themselves and a more detailed study of his problem is only of special interest. Also note that Markov has considered a more general case: his congruent triangles were scalene. I hasten to add that the geometric aspect of Markov's own discussion was so uninviting that hardly anyone ever checked it.

[5]This proposition is of course a mathematical version of the popular saying: *Don't put all your eggs in one basket.*

In the memoir [2] which I mentioned earlier, Buniakovsky had in addition studied several similar problems. He indicated that the values of special transcendental functions, which appear in their solution, can be approximately determined by means of the Monte-Carlo method, as it is now called[6]. Laplace [52, p. 366] made a similar remark only in regard to π.

Peres Larigno [69][7] published a brief review of the applications of geometric probabilities. Her work contains errors and she does not indicate that this notion in effect first appeared in a methodological problem studied by Newton but not published by him [78, p. 152].

2.5. Numerical Probabilities. Buniakovsky [6, pp. 132-137] solved a problem unusual for his time by calculating the probability that the equation

$$x^2 + px + q = 0$$

with coefficients p and q having random integral values $\pm 1, +2,, \dots, \pm m$ has real roots. The calculations came to counting the number of integral solutions of the inequality

$$p^2 - 4q \geq 0, \ q > 0.$$

Buniakovsky noted that, as $m \to \infty$, the probability sought tends to 1. He referred to his memoir of 1836 [1] where he had solved both this problem and a similar one for a complete quadratic equation.

Holgate [43] described similar stochastic investigations made by Waring (in 1782) and Sylvester (in 1864 - 1865). Neither Sylvester nor Holgate mentioned Buniakovsky's elementary considerations and the latter did not refer to Waring.

Stochastic problems pertaining to sets of real numbers[8] are of course more interesting. An appropriate case in point is provided by Oresme's statement. In the 14th century, he maintained that two ratios [two numbers] randomly chosen were probably incommensurable [64, pp. 40 and 247; 76, p. 131].

2.6. Random Walks. *Given the position of two squares* [A and B] *on ... a chessboard, it is required to determine the probability that a castle standing on one of these squares* [on square A] *reaches the other one in x moves.* This problem [6, pp. 143-147; the passage is on p. 143] should be specified. The castle is to move over the board at random, but in accordance with the rules of the game. Therefore, each move sends the castle from its given position to any of the 14 squares which are within its reach, and its arrival on each of these squares is equally possible. Moreover, the appearance of the castle on square B in less than x moves is not taken into account and, finally, the situation of B relative to A is random so that these two squares may even coincide.

Thus Buniakovsky considered a problem concerning generalized random walks. In spite of its elementary nature (the castle has only three different states, see below), this fact

[6]Its accuracy in such cases is, however, low.

[7]I am not sure of the spelling of this name.

[8]Example [44, p. 8]: A number belonging to segment [0, 1] is chosen at random; what is the probability that it will be algebraic?

deserves to be placed on record. Indeed, it is possible to identify a number of games of chance with a one-dimensional random walk of a particle. However, random walks in their proper sense were hardly considered before Buniakovsky. In any case, Dutka [36] in describing the first contributions in this field began at 1865. Buniakovsky divided the squares of the chessboard in three groups. His first group consisted of square A itself; his second group was made up of the 14 squares lying within reach of the very first move (from A*);* and the third group included the rest of the 49 squares. Accordingly, he compiled and solved a system of three difference equations and suddenly discovered that the mean value of the probability sought is 1/64 and does not depend on x. Buniakovsky's problem was, however, elementary. The castle can only be in two states,– it can reach B either in one move, or in two moves; the case A ≡ B belongs to the latter, but might be isolated for the sake of expediency. He had not interpreted his final result, but properly indicated that it was also possible to solve the problem in an elementary way, by direct calculation. Note that the first n moves ($n \ge 1$), if unsuccessful, do not change anything, and this circumstance apparently explains the situation.

2.7. Statistical Control of Quality. Buniakovsky appended to his treatise [6] a study of military losses (pp. 455 - 469). In 1850, he published it as a separate memoir [9]. Let n soldiers be selected at random from all the men in a detachment, N, and suppose that by a certain moment of an engagement i of these n men are put out of action. What will be the *probable* total number of casualties, Buuniakovsy asked (p. 456)[9].

Denote by x the probability that a certain soldier of the detachment is put out of action. Then the probability of the recorded fact will be

$$P = \frac{n!}{i!(n-i)!} x^i (1-x)^{n-i}.$$

Now, probability x has $(N - n + 1)$ equally probable values i/N, $(i + 1)/N$, ..., $(i + N - n)/N$; the corresponding hypotheses lead to probabilities P, P_1, P_2,…, P_{N-n+1}.

[9]Both here and below Buniakovsky actually thought of the mean number of casualties. He used the same term (*nombre probable*) in his memoir [9,

pp. 234 and 236]. *Cf* the title of his contribution [19]. In a forthcoming publication, I will show that Betrand made the same mistake whereas Poincaré, in accordance with his own definition, used the terms probable, and mean value on a par.

The *probable* (p. 457) number[10] of casualties will be $k = iN/n$[11] and the main question thus reduces itself to determining the probability p that this number will belong to a certain interval $[k - \omega, k + \omega]$. Basing his argument on the so-called Bayes formula, Buniakovsky assumed that

$$p = (P_\alpha + P_{\alpha+1} + \ldots + P_\beta)/(P_1 + P_2 + \ldots + P_{N-n+1}),$$

$$\alpha = k - \omega - i + 1,\ \beta = k + \omega - i + 1.$$

He then made use of the Maclaurin – Euler summation formula for calculating each of the two sums and estimated the value of the appearing incomplete B function; also see § 2.12.

Buniakovsky (pp. 468 - 469), used these calculations to conclude that it was possible to compile a table suitable *for many cases occurring in social life.* As an example, he cited the estimation of *a very large number of articles and supplies* only a fraction of which is actually examined. He had not yet formulated this idea in the French version of his work [9][12] read on February 20, 1846 where, on p. 257, he only mentioned *plusieurs autres usages.*

By 1846 statistical control of quality was still unknown. Gnedenko [41, p. 365], however, noticed that Simpson [85, Problem 6] had considered a highly relevant problem:

There is a given Number of each of several Sorts of Things as (a) of the first Sort, (b) of the second, etc. put promiscuously together; out of which a given Number (m) is to be taken, as it happens; To find the probability that there shall come out precisely a given Number of each Sort ...

About a hundred years later, without mentioning Simpson, Öttinger [62, p. 231] considered an equivalent problem. Then, in 1848, Ostrogradsky [66] formulated a problem directly concerning statistical control. He maintained (p. 322) that

Il est étonnant que la question propre a l'opérer n'ait pas été convenablement traitée; car les solutions que nous en avons sont peu exacter et peu conformes aux principes de l'analyse des hasards.

It is possible that Ostrogradsky was by then acquainted with Buniakovsky's treatise. Indeed, he read his memoir [66] on October 23, 1846, whereas, on October 1, the censors had already authorized the appearance of vol. 44 of the periodical *Sovremennik*, where, on pp. 196 - 204, an anonymous author had published a (non-mathematical) review of Buniakovsky's work.

[10] More precisely, Buniakovsky advised the recording of sample losses in each arm of the engaged force, that is, the use of stratified sampling, as it is now called.

[11] Buniakovsky noted that *iN*/*n* was not necessarily an integer and accordingly he recommended to correct it.

[12]The date of its publication is 1850. However, in 1846 [6, p. 455, note] Buniakovsky stated that it had already appeared. The only plausible explanation is that actually the memoir was then only scheduled for publication in the appropriate volume of the *Memoirs*

Finally, concerning Ostrogradsky's adverse opinion of earlier contributions it is difficult to determine exactly what he meant. And no one has yet checked his own main formula [66, p. 342].

2.8. The Law of Large Numbers. Buniakovsky [6] reasonably attached much importance to Bernoulli's law of large numbers (LLN)[13]. Then he referred to Laplace and obtained, practically speaking, the same result, Buniakovsky derived the De Moivre – Laplace integral limit theorem (with a correction term) calling it *the Bernoulli theorem.*

The Poisson form of the LLN did not earn recognition all at once [80, pp. 273 - 274]; Buniakovsky (p. 35), however, was one of the first to mention it.

2.9. Mathematical Treatment of Observations. Buniakovsky devoted more than sixty pages of his treatise [6] to the mathematical treatment of observations. At first he studied the distribution of the arithmetical mean and, in general, of a linear function of errors of observations. Following Laplace and supposing that the errors were distributed over a finite interval either uniformly or according to an arbitrary even law, he proved the relevant limit theorems [51, § 6; 52, chapt. 4; 79, pp. 18-21, 25-27 and 30-32].

In addition, he briefly described the MLSq. Appropriately referring to Gauss, Buniakovsky regrettably did not throw light on the Master's fundamental achievements. Furthermore, just as had Laplace, he did not use Gauss's elegant notation such as

[13]Omitted

$$[ab] = a_1b_1 + a_2b_2 + \ldots + a_nb_n. \qquad (1)$$

Because of both these circumstances, Buniakovsky's exposition was old-fashioned which is all the more regrettable since even Shiyanov [83] did more justice to Gauss[14].

In 1859 Buniakovsky [10] designed a mechanical device equipped with verniers for calculating sums of squares (and, therefore, scalar products (1) as well) to four significant digits. The instrument was intended for evolving the normal equations occurring in the adjustment of observations, or alternatively, if its accuracy was not sufficient for the purpose, for a rough check of the calculations. I cannot say whether this device was ever actually used.

2.10. Testimonies, elections, verdicts. Buniakovsky allotted another sixty pages of his treatise [6] (cf. § 2.9) to the treatment of the results of elections; to the study of testimonies and legends and of decisions passed by tribunals.

Suppose that out of s witnesses whose testimonies have the same probability of truth p exceeding 1/2, r maintain that a certain fact did occur whereas the rest $s - r = q$ ($q < r$) declare the opposite. Then [6, p. 311] the probability that the first group of witnesses tells the truth is

$$P = \frac{p^{r-q}}{p^{r-q} + (1-p)^{r-q}}.$$

This coincides with the probability of a unanimous statement made by $r - q$ people. Both this formula and conclusion can be found in Laplace's classic [52, p. 466]. Thus, Buniakovsky continued, the case of $s = 212$ and $r = 112$ (and $q = 100$) is equivalent to having $s = r = 12$ (and $q = 0$).

Not really convincingly, Buniakovsky corroborated his conclusions by proving that if the first case did take place, the probability of the second would be very low. As a basis he took an integral formula in which the probability of the witnesses' telling the truth was considered variable and, furthermore, taking all values in the interval [0, 1]. Both these circumstances contradicted his own earlier premise of a constant p exceeding 1/2. Also cf. Laplace's formula below.

Buniakovsky borrowed his numerical example from Laplace [53, p. XCVIII; 52, pp. 523-524] who used it to illustrate decisions arrived at by a jury consisting of 12 (or 212) members, provided that for each juror the probability of making a mistake was variable with the interval of possible values of [0, 1/2] which followed from a formal application of the Bayesian approach. And so, the probability of a wrong verdict was equal to

$$\int_0^{1/2} x^p (1-x)^q dx \div \int_0^1 x^p (1-x)^q dx.$$

In addition, Buniakovsky took into account the [prior] probability of the fact being testified to, as though considering the testimony of a new witness. Allowing for the possibilities that the witnesses are mistaken and deceived or are mistaken and tell the truth, *etc.* (four cases in all), and, following Laplace [52, chapt. 11; 78, p. 171], he

[14] Still, Syanov used only some of Gauss's notation; worse, he did not describe Gauss` second substantiation of the MLSq.

determined the probability that the fact had actually happened. The event that he studied was the drawing of ticket i out of an urn containing n tickets numbered from 1 through n $(1 \le i \le n)$.

Buniakovsky paid special attention to the case of an unlikely event. Suppose (p. 314) that two eye-witnesses maintain that letters selected from an alphabet of 36 letters made up the word *Moskva* [Moscow]. Assuming that the witnesses were equally trustworthy and that $p_1 = p_2 = 9/10$, that the letters were drawn at random, and, finally, that the total number of reasonable six-letter Russian words was 50,000, Buniakovsky determined the probability that the witnesses' account was true. Here are his calculations: According to the formula above, P = 81/82 and the probability proper of composing an intelligent word is 50,000:(36•35•34•33•32•31) = 1/28,048.. Lastly, from that formula generalized to include the case of unequal trustworthiness of two witnesses,

$$P = \frac{p_1 p_2}{p_1 p_2 + (1 - p_1)(1 - p_2)}.$$

Buniakovsky took p_1 = 81/82 and p_2 = 1/28,048 and got P = 1/347, *i.e.*, the probability of obtaining any reasonable word rather than a definite word[15].

Upon solving Buniakovsky's problem in its exact sense, that is, determining the probability of composing a definite word, Markov [57, p. 320] added:

This example ... sufficiently illustrates that many arbitrary assumptions are needed in order to solve problems ... that are in essence ... of a very indeterminate nature. … [If] *we admit that witnesses can be mistaken and deny the independence of their testimonies, the uncertainty will deepen*[16].

Formally speaking, Markov was right. Nevertheless, during the few latest decades, the theory of probability and mathematical statistics had to solve perhaps even less definite problems. Thus, although the treatment of the outcome of elections which Buniakovsky had considered (following Borda and Condorect), hardly interest present-day mathematicians, who have begun to study rank correlations (for example, in dealing with expert appraisals).

Incidentally, Laplace's problem concerning elections (e.g., in his *Théorie analytique* [52, § 15]), or, for that matter, appraisals, which Buniakovsky (pp. 341 - 345) solved by using simpler mathematical tools, can also be attributed to rank correlation. Here it is as formulated by Buniakovsky: A certain fact could have resulted only from one of the causes $C_1, C_2, \ldots, C_i$, whose unknown probabilities are $p_1, p_2, \quad, p_i$, whose sum is unity.

[15]Laplace (see, for example, his *Essai* [53, p. XV]) believed that the word *Constantinople* could hardly have been composed of separate letters by chance; much more likely, he indicated, was that the letters had been arranged by someone on purpose. Buniakovsky (p. 315), however, stated that he excluded the action of outside agents.

[16]I have quoted Markov's thoughts concerning Buniakovsky's view of miraculous events [81, p. 340] and I can now add that elsewhere Buniakovsky [13, p. 4] made a more definite, pronouncement of the same kind:

By excluding truths cognized by revelation we shall ... find out for sure that ... almost all the rest of our knowledge is based solely on probabilities.

(Markov refused to believe in revelation.)

Each voter (expert) arranges these probabilities in decreasing order and it is necessary to determine the mean value of the (subjective) probability of each cause.

Buniakovsky, as he himself indicated, described the application of probability to jurisprudence according to Poisson [72; 80, § 6]. Note that he (p. 359) repeated one of his doubtful statements [72, p. 333; 80, p. 287]: Supposing that the probability of a just decision is the same for each juror, Buniakovsky maintained that the probability of a correct majority verdict depended on the difference between the votes rather than on the total number of jurors. Indeed, if the majority is not fixed, the difference for ($2n$ – 1) jurors can be equal to 1, 3, ..., ($2n$ – 3) and Poisson (and Buniakovsky's) statement becomes wrong.

Many scientists opposed the application of probability to jurisprudence (or even, like Cauchy in 1821 and Poinsot in 1836 [75, p. 296], rejected its use beyond natural science). More definitely, Cournot [31, § 214] had noted that prejudices in law courts were of a social nature. In 1906, Poincaré [70, p. 92] maintained that people act like the *moutons de Panurge* but he was ignorant of Poisson's main goal: the minimization of the miscarriage of justice by determining the optimal majority votes of the jurors[17].

2.11. The history of the theory of probability. Buniakovsky included in his treatise a good essay on the history of probability although it is not difficult to indicate several mistakes there[18]; again, he did not describe De Moivre's achievements clearly enough. Finally, this time considering Buniakovsky's book [61 in general, **I** note that the reader will not grasp exactly what Jacob Bernoulli or Gauss accomplished (§§ 2.8 and 2.9).

Still, Buniakovsky was one of the first to publish a study of this kind. He had few predecessors. Montucla [60] devoted about 45 pages to the history of probability, but his account was of a popular character. Moreover, he overlooked Laplace's early memoirs. Laplace himself described the same subject in a section of his *Essai* [53] but his exposition was hardly successful: he rarely referred to definite sources and, furthermore, the complete lack of formulas (throughout the *Essai* in general) impeded reading. Lastly, the contribution of Lubbok & Drinkwater[55], though not without shortcomings, was really useful. Buniakovsky did not mention it. Upon studying popular encyclopedic articles written by Buniakovsky, Prudnikov [73, p. 235] declared that, in general, he displayed *interest in the history of mathematics*. Buniakovsky's encyclopedic dictionary [3] bears witness to the same conclusion. In

[17]In one case Laplace [52, p. 523] indicated that the probability of a just decision made by each juror is

Très près de l'unité ... a moins que des passions ou des préjugés communs n'egarent tous les juges.

Also see my comment [78, p. 172] on the criticisms levelled against applications of probability in jurisprudence.

[18]Buniakovsky [6, p. 368] believed that the *Lettre à un ami sur les parties du jeu de paume* (1713) was a work of an unknown mathematician; in actual fact, its author was Jacob Bernoulli. Then (p. 369), he gave the year of the first publication of De Moivre's *Doctrine of chances* as 1716 instead of the correct date, 1718.

a later contribution, Prudnikov [74, pp. 8, 26 - 31 and 40] himself described all of Buniakovsky's studies on the history of mathematics. Among later Russian mathematicians who interested themselves in the history of their science, I can mention Markov [81, § 3] and, of course, Bobynin.

2.12. Population statistics. Buniakovsky [6, pp. 173-213] discussed the main problems of population statistics. He described various methods of compiling mortality tables; studied the increase of population, resulting, in particular, from the weakening of, or deliverance from a certain cause of mortality; calculated the expected and probable durations of marriages (and associations). True, Laplace, and, to some extent, Euler, described most of these topics although not the compilation of mortality tables.
In addition, Buniakovsky solved two special problems; see also Laplace [52, chapt. 6; 78, pp. 157-161].

The first problem. Suppose that, in a given nation, during a certain period p boys and q girls were born. Then

$$P(\frac{1}{2} \le x \le 1) = \int_{1/2}^{1} x^p (1-x)^q \, dx \div \int_0^1 x^p (1-x)^q \, dx$$

where x is the probability that a newborn baby will be a boy.

The second problem. Denote the population of a small part of a country by m, the number of yearly births in this part by n and, by N, the total number of early births in the country. It is required to estimate the entire population of the country, M, roughly equal to mN/n. Suppose, as did Laplace, that (m, n) and (M, N) are samples from a single universe, and that the samples contain n and N white balls out of m and M balls, respectively. Then, given the sample (m, n), the probability x that a white ball is to appear will be

$$x^n(1-x)^{m-n}dx \int_0^1 x^n (1-x)^{m-n} \, dx$$

and the probability that N white balls will be contained in the second sample will be

$$\frac{M!}{N!(M-N)!} x^N (1-x)^{M-N}.$$

The intermediate formula above (Laplace [52, p. 393]) is appropriate for continuous probabilities as well. Finally, the probability that the total number of balls in the second sample is M, will be

$$P = \frac{M!}{N!(M-N)!} \int_0^1 x^{n+N} (1-x)^{m-n+M-N} \div \int_0^1 x^n (1-x)^{m-n} \, dx.$$

Again, Laplace (l. c.) derived this formula and applied it to solve a number of problems. As to Buniakovsky, he estimated the integrals in that formula and, supposing that $M = mN/n + t$, and using the Maclaurin – Euler summation formula, determined the probability of an inequality such as $|t| < a$.

Thus the chief difficulty in both these problems was to estimate the appropriate integrals and, especially, to represent the values of the incomplete B-function by the integral of the exponential function of a negative square. Like Laplace, Buniakovsky was content with a fairly low accuracy in his calculations. Later mathematicians [86, pp. 43 - 46], who strove for much greater precision, had to overcome considerable obstacles [78, p. 161; 42, p. 192].

Pearson [78, p. 160] criticized Laplace’s inferences concerning the samples (m, n) and (M, N). In particular, he noted that the existence of a single universe was questionable. Moreover [26], the very concept of parent population is not logically rigorous. Nevertheless, Laplace was the first to estimate the plausibility of sampling.

3. Later works

I have mentioned some of Buniakovsky’s contributions [13]; [12] and [21]; [9]; [13]; and [10] in §§ 2.2, 2.3, 2.7, 2.8 and 2.9 respectively, published after 1846.

3.1. Population statistics. Buniakovsky returned (cf. § 2.12) to this subject in a number of later works. In 1866 he compiled mortality tables of Russia's Orthodox believers and tables of their distribution by age [12]. He subsequently corrected the second type of these tables, making allowances for new statistical data [17]. Moreover, he estimated the number of Russian conscripts ten years in advance[19][19]. Many writers maintain that Buniakovsky was a government expert in demography. I cannot corroborate this statement. It is likely that he executed relevant assignments but he hardly doubled as a government official.

Making use of later data, Bortkewitch (Bortkiewicz) [28] sharply criticized Buniakovsky, declaring (p. 1056) that his tables

Do not provide even a roughly accurate picture of mortality in recent times.

He did not change his mind in a later contribution [29], and it is indeed possible that he was right (below). However, he said nothing whether Buniakovsky had any real possibility of achieving better results. He himself repeatedly stressed the inaccuracy and inadequacy of his data [12, pp. 4, 10, 162; 13, pp. 27 and 52; 15, pp. 10 and 20; 17, p. IV] and, consequently, did not regard his conclusions as sufficiently sound [12, pp. 159 and 162; 13, p. 52; 15, p. 20; 17, p. IX; 19, p. 18]. In the last instance, he remarked that, should his figures prove wrong,

It would be necessary to infer that the [relevant] *mistakes lie in the data* on births and mortality.

Bortkiewicz did not mention Davidov [32, pp. 51-52] who called Buniakovsky’s *Essay* [12] *an excellent work,* and *the most*

[19] Nobody ever verified this forecast.

comprehensive and detailed [contribution] *both in regard to the precision of the methods used and with respect to the thorough treatment of its subject.*

Nevertheless, even considering Buniakovsky's data insufficient, Davidov reproached him for making a few blunders and concluded that Buniakovsky had underestimated the death-rate in Russia.

I shall now quote from a later source [61, pp. 54 – 55]:

A new period in the study of mortality in Russia started with ... the demographic investigations ... made by Buniakovsky ... and, especially, ... with this Essay **[**12**].** *Buniakovsky's contributions on population statistics represent an outstanding and remarkable phenomenon not only in our extremely poor demographic literature but in the rich realm of foreign writings as well, and particularly of his time ... Due to the clearness, depth, and nicety of his analysis, Buniakovsky's works fully retain their importance for the present day. … Because of the inaccuracies in the main data, the lack ... of many necessary materials, and as a result of the shortcomings in the method itself employed in composing the tables,* **[**Buniakovsky's mortality tables] *although a great step forward ... do not portray a sufficiently correct picture of Russian mortality.*

Brief information on Buniakovsky's method of compiling mortality tables is contained in the *Comptes rendus* of the International Statistical Congress [47]. Recent publications on the subject are Shusherin's paper [84] and two articles in the *Demographic dictionary* [33], *viz.*, *Buniakovsky's method of compiling mortality tables* and *Mortality tables.*

3.2. Stochastic summing. An autoabacus. In 1867 Buniakovsky [14] solved a few problems on the stochastic summing of a large number of terms. Suppose that *n* numbers are selected at random from integers 1, 2, …, *m* (the numbers in the sample can coincide). It is necessary to determine the prob~~ability~~

$$P(|s-\frac{(m+1)n}{2}|\le l).$$

Here *s* is the sum of the sample numbers, $(m+1)n/2 = a$ is the mean value of *s* and *l* is much less than *a* (I have somewhat changed the notation). The solution of this problem, the main one in Buniakovsky's memoir, can be represented as

$$P(-\alpha\le s-a\le\alpha)\approx\frac{\sqrt{2}}{\sigma\sqrt{\pi}}\int_0^{\alpha}\exp(-z^2/2\sigma^2 dz,\ \sigma=\frac{n(m^2-1)}{12}. \quad (1)$$

Of course, σ is the variance o the sum of integral random variables uniformly distribute on the interval [1, *m*].

Laplace [50, § 3; 42, pp. 194-195] and, later, Buniakovsky himself (§ 2.9) derived a formula equivalent to expression (1) in the context of the theory of errors.

The rest of the pages of the memoir [14] were primarily concerned with applying formula (1) to summing the values of functions (for example, of square or cube roots taken at consecutive natural values of their arguments) or the results of observations (of atmospheric pressures recorded for six months at a given point and at the same time of the day). The author did not indicate the aim of stochastic summing. However, it was evident, at least in the case of the observations, that in calculating the mean value of the atmospheric pressure it was necessary to find the sum of the pressures.

Buniakovsky stressed that his method could be applied only to sum the variable parts of tabulated data, *i.e.*, to sum such numbers all of whose possible values had equal prior probabilities. The summing of the radicals (above) can be then explained by noting that almost all of the relevant irrational numbers, as is now commonly accepted (though not proved), are normal.

I know nothing concerning the practical use of the formula (1) for stochastic summing.

It is appropriate to mention Buniakovsky's autoabacus. His first communication describing this instrument at a meeting of the physical and mathematical department of the Petersburg Academy of Sciences dates back to 1867. Here is the report on this meeting[20]:

Buniakovsky made a report on an instrument he had invented. ... Its purpose is to eliminate the shortcoming inherent in usual abacuses concerning the transfer of units from a lower rank to a higher one. In Mr. Buniakovsky's instrument these units by means ... of a simple mechanism arrange in proper order all by themselves. ... Mr. Buniakovsky produced a specimen of the instrument constructed ... by the mechanic of the Academy ...

No mention was made either here or in the subsequent memoir [20] of the use of the autoabacus for stochastic summing. Nevertheless, the memoir was mainly devoted to the use of this instrument for direct calculations of monthly and yearly means of the values of meteorological elements[21].

Bool [27, pp. 53-62] highly estimated the autoabacus but did not report any comments made by those who used it; moreover, it is difficult to say if the tool was indeed ever applied. Bool, also indicated (p. 62) that the idea of the autoabacus was not new: it had been used in the so-called *Kummer's calculator.* On pp. 14 - 19 he described this *simple but excellent instrument,* which had appeared in Russia in 1847, but, once more, he did not comment on it.

Prudnikov [74, p. 81] suggested that Buniakovsky [20] had led Chebyshev to design an adding machine.

[20]*Zapiski* [Petersb.] *Acad. Nauk*, vol. 11, pt. 1, p. 72. Also see newspaper *St. Petersb. Vedomosti*, 8 March 1867, p. 3.

[21]For example, instead of dividing a sum of the values of an element by 30, it was possible to assign weight 1/30 to each of them.

3.3. The theory of random arrangements. One of Buniakovsky's problems [16] can be attributed to the theory of random arrangements. A certain number of copies of a booklet are faulty in one or another respect. For example, each lacks with equal probability one of the pages; or, even, the missing pages are distributed non-uniformly among the copies; or, extra pages are bound in. It is necessary to determine the probability that a certain number of faultless booklets can be composed from the given ones.

Buniakovsky solved this problem, making use of appropriate generating functions, without encountering any special difficulties. It seems, however, that no one ever applied his study or enlarged on it.

3.4. The partition of numbers. In 1875 Buniakovsky [18] solved a natural but difficult problem: *n* balls numbered from 1 through *n* are placed in an urn out of which *a* balls are then drawn all at once. What is the probability that the sum of the numbers drawn is equal to *s?*

This problem is due to Laurent [54, p. 76], who referred to similar studies made by Euler [39, chapt. 16]. In considering the partition of numbers, Euler allowed the case of identical terms. However, he noted also that the coefficient of $x^n z^m$ in the expanded form of the product $(1 + x^{\alpha}z)(1 + x^{\beta}z)(1 + x^{\gamma}z) \ldots$ indicated the number of ways in which a given number *n* can be represented as the sum of *m* different items, α, β, γ … He did not, however, calculate this coefficient.

Buniakovsky made it clear that the difficulty confronted in his problem lay exactly in determining such a coefficient, or, more precisely, of the coefficient of $t^{\alpha}x^s$ in the development of the product $(1 + tx)(1 + tx^2) \ldots (1 + tx^n)$. He solved this problem for small values of α by means of an extremely complicated equation in finite partial differences, and offered a formula for passing from α to (α + 1).

Poisson [72; 80, § 7.3] considered the same problem in a more general setting although did not intend to solve it. Later, in 1867, Öttinger [63, pp. 335-337] used a simple trick that can be traced to Euler to represent the product $(1+ xz)(1 + x^2z) \ldots (1 + x^mz)$ as

$$1 + v_1z + v_2z^2 + \ldots + v^mz^m$$

and derived an expression for its coefficients:

$$v_r = \frac{(x - x^{m-1})(x^2 - x^{m-1})\ldots(x^r - x^{m-1})}{(1-x)(1-x^2)\ldots(1-x^r)}, \; r = 1, 2, \ldots, m.$$

He found his way out of Poisson's problem, but did not solve it completely.

3.5. Linguistics. In his popular articles Buniakovsky expounded the principles of compiling and using mortality tables and described the benefits secured by various forms of life insurance (especially by participating in pension funds). In addition, basing his arguments on the notion of moral expectation (§ 2.3), he warned of the danger of games of chance and lotteries [4; 11] and, on the contrary, recommended the division of commercial transactions involving risk [4]. In one instance Buniakovsky [7] elucidated the elements of the mathematical treatment of observations. He explained how to estimate the population of a country by sampling and went on to discuss the application of probability to linguistics. The *analysis of probabilities*, he maintained (p. 48), can be used

In grammatical and etymologic studies of a particular language [and] *also in comparative philology*.

I adduce further passages:

My statement is based ... on a critical discussion of the subject, on some of my previous attempts and on analytic formulas which I derived to determine the probabilities of various constructions of words. It is necessary to ascertain *numerical data on the total number of words, ... on the distribution of these words by parts of speech, number of letters, by first letters, endings, etc. Information concerning general rules, exceptions of various kinds, words ... adopted from other languages and so on is also needed. ... Upon obtaining such statistical data for two or several languages it will be possible to compare them in various respects. Thus the conclusions arrived at will achieve a status of authority which philologists in the present state of the* [of their] *science are not always able to demonstrate. ... On another occasion I shall perhaps publish my theoretical studies. ... Mathematicians must certainly enter into relations with experts in this subject.*

At that time, statistical investigations in linguistics were just beginning to appear [46]. Regrettably, Buniakovsky never published (perhaps did not even complete) this research, now certainly lost (§ 1.2).

3.6. The dread of cholera. During Buniakovsky's lifetime, Russia suffered several cholera epidemics which led to panic, and in 1830-1832, to *cholera riots*. It is not surprising, then, that in 1848, when one of the epidemics had started, Buniakovsky published an article [8] in a metropolitan newspaper[22] and called upon the population to remain calm and orderly. In illustrating his ideas by simple numerical calculations, he naturally had to avoid algebraic notation and derivations.

In the absence of cholera, Buniakovsky explained, the *danger* [the statistical probability $p_1(t)$] of dying for a person of age t can be determined by a mortality table; the *danger* [the probability p_2] of dying of cholera, which he supposed to be the same for each inhabitant of the capital, can be found from the data on the previous epidemic (Moscow, 1830).

According to Buniakovsky's calculations, p_2 was considerably less than p_i for any age t so that the father (he never mentioned the mother!) of several children of ages $t \leq 5$ years should not fear that one of them would die. Adducing additional arguments, he indicated that

[22]Not more than a few copies of the newspaper are in existence. I have therefore reprinted Buniakovsky's article [8] in the Russian version of this contribution [82]. I do not publish its translation since (below) it is not really scientific.

The probability of dying of cholera during an epidemic monotonically decreases day after day. Suppose, said Buniakovsky, illustrating this idea, that n persons out of N are to die. Then if $n/3$ (say) have died, the probability of dying will be $(2n/3)/(N - n/3)$

which is less than n/N, the same probability on the first day of the epidemic. And, during an epidemic, other diseases abate.
I conclude with a few adverse comments.

The probability of being infected with cholera evidently depended on the way of life of a given person and, therefore, on his/her sex, age, and social status. Just the same, the probabilities of dying of cholera also differed from person to person. ##

Buniakovsky paid no attention to these facts except for making a superficial remark to the effect that prudent people *had already taken some preventive measures.*

A comparison of p_1 and p_2 was not really instructive. It would have been much more natural to compare p_1 with $(p_1 + p_2)$.

The father of several children should have feared that *at least one of them* will die. Therefore, in the case of k children, p_i, $i = 1, 2$, should have been replaced by $(1 — q_i^k)$, $q_i = 1 – p_i$.

Moreover, the father should have feared cholera as such since, upon recovering, his child would be more prone to die of other diseases. Incidentally, Buniakovsky's statement concerning the lesser danger of these other diseases seems highly questionable.

Similarly, I do not believe that the probability of dying of cholera decreases during an epidemic. Indeed, the existence of a definite number n given beforehand seems doubtful. If the population, upon believing that the epidemic has practically ended, ceases to take precautionary measures, the disease can break out anew.

Buniakovsky supposed that n will be much less than 12,000. Actually, however, in 1848 26,836 people were afflicted and 14,430 of these (53.8%) died; furthermore, in 1849, after a winter lull, the epidemic continued with the corresponding figures being 6,384 and 3,156 (49.4%) [24, p. 3].

4. Conclusions

In 1847, an anonymous reviewer [23] praised Buniakovsky's treatise [6]. Not quite appropriately, the reviewer started by declaring that Ostrogradsky was a genius. He (p. 40) proceeded to maintain that Buniakovsky's merits *though secondary were still merits, worthy of attention.* And, further (p. 44):

In the Russian language, Buniakovsky's book is new in subject matter, complete in its contents and scientifically up-to-date. What else shall we demand of an author [who does not claim originality]?

Nevertheless, the treatise [6] did contain some findings (§§ 2.4-2.6) and new ideas (§ 2.7). Buniakovsky's later works were also of value. Even leaving aside population statistics (§ 3.1), recall the stochastic summing (§ 3.2) and the partition of numbers (§ 3.4).

Buniakovsky's contemporaries did not follow up on his concrete achievements but his contributions for a few decades exerted exceptional influence on the teaching of the theory of probability in Russia. Prudnikov [74] described this aspect in detail and I only adduce a passage from a lesser known work [90, p. 36]:

This thorough and clearly written source [6], *one of the best in European mathematical literature on the theory of probability, considerably helped to disseminate interest in this discipline among*

Russian mathematicians and to raise the importance of its teaching in Russian universities to a higher level as compared with the academic institutions of other nations[23]. *At Moscow university for example, the teaching of probability was initiated in 1850 by A. Yu. Davidov who had been specially invited for this purpose.*

To place so high a value on a treatise published 75 years earlier was of course an exaggeration. However, Markov [58, p. 162], in spite of his criticism (§ 2.10), considered Buniakovsky's writing *a beautiful work* and Steklov [87, p. 177] believed that *for his time* Buniakovsky had compiled *a complete and outstanding treatise*.

But the full story should be told! In 1846, 1867, and, again, in 1887 Chebyshev published his remarkable studies in the theory of probability. Buniakovsky, however, just did not pay attention to them (cf. § 1.1).

Acknowledgements. This paper is a slightly revised version of my Russian preprint [82]. M. V. Churikov and Al. Ad. Youshkevich read the MS of this preprint. They pointed out a few mistakes and ambiguities and offered methodological advice. Michael Davidov checked the English text of my paper.

Addendum to § 3.6. In 1889, P. D. Enko, a Russian physician, published a paper (On the course of epidemics of some infectious diseases) offering the first epidemic model in medicine. Extracts from his contribution translated into English by K. Dietz have recently appeared in the *Intern. J. of Epidemiology* (vol. 18, No. 4, 1989, pp. 749-755). They have prompted me to add that Buniakovsky regrettably did not make the necessary steps to originate mathematical epidemiology.

[23]Drawing on a contribution published by Mansion in 1903, I have described the unsatisfactory situation existing in those times in France and Germany [80, p. 273, note 30]. However, Mansion also maintained that in Belgium the status of probability was much higher. He attributed this fact to the lasting influence of Quetelet.

References

AHES = *Arch. hist. ex. sci.*
IMI = *Istoriko-mathematicheskie issledovania*
L, M, Pg, Psb = Leningrad, Moscow, Petrograd, Petersburg
Petersb. Bull. = *Bull. phys.-math.* (= *Izvestia*) *Acad. sci. St. Petersb.*
Petersb. Mem., t. m/n = *Mem. Acad. sci. St.-Petersb.*, 6me ser., Sci. math., phys., natur., t. m **(**= Sci. math. et phys., t. n)
Petersb. Zap., vol. m, No. n = *Zapiski Acad. Sci. St.-Petersb.*, vol. m**,** Suppl. n. Each Supplement had its own paging.
R = in Russian

V. Ya. Buniakovsky

1R. Determination of the probability that a randomly chosen quadratic equation with integral coefficients has real roots. *Petersb. Mem.*,t. **3/1,** No. 4, 1836, 341 – 351.
2R. On the application of the analysis of probabilities to determining the approximate values of transcendental numbers. *Ibid.*, 457 – 467 and No. 5, 1837, 517 – 526.
3R. *Leksikon chistoi i prikladnoi matematiki* (*Lexicon of pure and appl. math.*), *vol.* **1.** Psb, 1839.

4R. Thoughts about some unfounded notions in social life. *Mayak,* pt. 3, 1840, 80 – 94.
5. Sur la publication, en Russe, d'une *Théorie analytique des probabilités. Bull. Scient. Acad. Sci. St.-Petersb.,* t. **10,** 1842, p. 95.
6R. *Osnovania matematiheskoi teorii veroiatnosstei* (Principles math. theory prob.). Psb, 1846.
7R. On the possibility of introducing definite measures of confidence in the results of some observational sciences and, in the first place, of statistics. *Sovremennik, vol.* **3,** 1847, pp. 36-49 of section Science and Arts.
8R. A few words concerning the dread of cholera. *St. Petersburg Vedomosti* (newspaper), 24 June 1848, No. 140, p. 560.
9. Sur une application curieuse de l'analyse des probabilites. *Petersb. Mem.,* t. **6/4,** No. 3-4, 1850, 233-258.
10. Sur un instrument destiné a faciliter l'application numerique de la méthode des moindres carrés. *Petersb. Bull.,* t. **17,** No. 19 (403), 1859, 289-298.
11R. Games of chance. *Enc. Dict.* [38], vol. **2,** 1861, 119-121. Coauthors I. Ye. Andrevsky, G. I. Kananov.
12R. Essay on the laws of mortality in Russia and on the distribution of the Orthodox believers by ages. *Petersb. Zap.,* vol. **8,** No. 6, 1866.
13R. Tables of mortality and of population for Russia. In *Mesyatseslov* (Calendar) for 1867. Psb, 1866, Suppl., 3-53.
14R. On the approximate summation of numerical tables. *Petersb. Zap.,* vol. **12,** No. 4, 1867.
15R. A few remarks on the laws of the movement of the population in Russia. *Russk. vestnik,* vol. **73,** 1868, 5-20.
16R. On a special kind of combinations occurring in problems connected with defects. *Petersb. Zap.,* vol. **20,** No. 2, 1871.
17R. Anthropobiological studies *etc. Ibid.,* vol. **23,** No. 5, 1874.
18R. On a problem concerning the partition of numbers. *Ibid.,* vol. **25,** No. 1, 1875.
19R. On the probable number of men in the contingents of the Russian army in 1883, 1884 and 1885. *Ibid.,* No. 7.
20R. On the autoabacus and its new use. *Ibid.,* vol. **27,** No. 4, 1876.
21R. On maximal quantities in problems concerning moral benefit. *Ibid.,* vol. 36, No. 1, 1880.
22. *Liste des travaux mathématiques.* Psb, 1883 (manuscript).

Other Authors

23R. Anonymous, Review of treatise [6]. *Finsk. vestnik,* vol. 16, No. 4, 1847, pp. 39-44 of section Bibl. chronicle. Ed., F. K. Derschau.
24R. Archangelsky G. I., *Kholera v Peterburge v prezhnie gody* (Cholera in Petersburg in former times). Psb, 1892.
25. Bernoulli D., Exposition of a new theory on the measurement of risk (1738, in Latin). *Econometrica,* vol. 22, No. 1, 1954, 23-36.
26R. Bolshev L. N., Sampling. *Great Sov. Enc.,* vol. 5, 1971, columns 1532-1534. There exists an English transl. of the entire edition of the *Enc.*
27R. Bool V. G., *Pribory i mashiny …* (Instruments and tools for mech. performance of arithm. operations). M., 1896.
28R. Bortkevich V. I., On Russian mortality. *Vrach, vol.* 10, No. 48, 1889, 1053 - 1056.
29. Bortkiewicz, L., Das Problem der russischen Sterblichkeit. *Allg. stat. Arch.,* Bd. 5, 1898, 175-190 and 381-382.
30R. Chuprov A. A., *Ocherki po teorii statistiki* (Essays on the theory of statistics) (1909). M., 1959.
31. Cournot A. A., *Exposition de la théorie des chances et des probabilités* (1843). Paris, 1984. Ed., B. BRU. **S, G,** 54.
32R. Davidov A. Yu., On mortality in Russia. *lzv. Obshch. liubitelei estestv., anthropol. i ethnogr.,* vol. 49, No. 1, 1886, 46-66.
33R. *Demographich. enz. slovar* (Demographic enc. dict.). M., 1985.
34. De Morgan A., Theory of probabilities. In: *Enc. metropolitana.* Pure sciences, vol. 2. London, 1845, 393-490.

35R. *Slovar tzerkovno-slav. i russk. yazyka* (Dict. of the Church Slavonic and Russ. Languages), vols **1-4.** Psb., 1847.
36. Dutka J., On the problem of random flights. AHES, vol. 32, No. 3-4, 1985, 351-375.
37. Ellis R. L., Remarks on an alleged proof of the method of least squares (1850). In *Math. and other writings of R. L. Ellis.* Cambridge, 1863, 53-61.
38R. *Enz. slovar sostavlennyi russk. uchenymi i literatorami* (Enc. dict. compiled by Russ. scientists and men of letters). Psb, 1861-1863.
39. Euler L., *Introduction to the analysis of infinities,* vol. **1** (1748, in Latin). New York, 1988.
40R. Gnedenko B. V., On the works of M. V. Ostrogradsky in the theory of probability. IMI, No. 4, 1951, 99-123.
41R. ---, A short essay on the history of the theory of probability. In author's *Kurs teorii veroiatnostei* (Course in theory prob.). M., 1954, pp. 360-388. The author suppressed the Essay from the later editions of his *Course*.
42R. Gnedenko B. V., Sheynin O. B., Theory of probability. In *Matematika XIX v.* (Math.of the 19th c.) [vol. 1]. Eds. A. N. Kolmogorov, A. P. Youshkevich. M., 1978, 184-240. Basel, 1992, 2001, pp. 211 – 288.
43. Holgate P., Waring and Sylvester on random algebraic equations. *Biometrika,* vol. 73, No. 1, 1986, 228-231.
44R. Khinchine A. Ya., Metric problems in the theory of irrational numbers. *Uspekhi math. nauk,* No. 1, 1936, 7-32.
45R. Kiro S. N., The scientific and educational activities of M. V. Ostrogradsky and V. Ya. Buniakovsky. In *Istoria otechestv. matematiki* (Hist. of nat. math.), vol. 2. Kiev, 1967, 52-103.
46. Knauer K., Grundfragen einer mathematischen Stilistik. *Forschungen u. Fortschritte,* Bd. 29, No. 5, 1955, 140-149.
47. Koumanine A., Résumé de la méthode de Buniakovsky appliquée a la construction des tables de mortalité et de population. C. *r. Congr. intern. stat.* La Haye, pt. 2, 1870, 138-145.
48R. Kozlovsky, On hope. *Sovremennik,* vol. 3, 1836, 23-47. **S, G,** 78.
49. Lacroix S.-F., *Traité élémentaire du calcul des probabilités.* Paris, 1816, 1822, 1833, 1864.
50. Laplace P. S., Sur les approximations des formules *etc.*, 1809 (1810). *Oeuvr. compl.*, t. 12, Paris, 1898, 301-345.
51. ---, Sur 1es intégrales définies *etc.* 1810 (1811). *Ibid.*, 357-412.
52.---, *Théorie analytique des probabilités* (1812). *Ibid.,* t. 7, No. 1-2. Paris, 1886.
53. ---, *Essai philosophique sur les probabilités* (1814). *Ibid.*, No. 1 (separate paging). English translation: New York, 1995.
54. Laurent H., *Traité du calcul des probabilités.* Paris, 1873.
55. Lubbok W., Drinkwater Bethune J. E., *A treatise on probability* (1830). London, 1844 this being an Appendix to Jones D., *On the value of annuities,* vol. 2.
56. Maistrov L. E., *Probability theory.* A historical sketch (1967, in Russian). New York – London, 1974.
57R. Markov A. A., *Ischislenie veroiatnostei* (Calculus of prob.) (1900). M., 1924. German transl. of the Russian ed. of 1908: Leipzig-Berlin, 1912.
58. ---, The bicentennial of the law of large numbers (1914, in Russian). In *The correspondence between A. A. Markov and A. A. Chuprov* (1977, in Russian). Ed., Kh. O. Ondar. New York, 1981, 158-163.
59R. *Materialy dlia biogr. slovaria …* (Materials for biogr. dict. of the full members of the Acad. Sci.), vol. 1. Pg., 1915 (1917).
60. Montucla J. E., *Histoire des mathématiques*, t. 3. Paris, an 10 (1802).
61R. Novoselsky S. A., *Smertnost i prodolzhitelnost zhizni v Rossii* (Mortality and longevity in Russia). Pg., 1916.
62. Öttinger L., Untersuchungen über die Wahrscheinlichkeitsrechnung. *J. reine angew. Math.,* Bd. 26, 1843, 217-267, 311-332.
63. ---, Über einige Probleme der Wahrscheinlichkeitsrechnung. Ibidem, Bd. 67, No. 4, 1867, 327 – 359.

64. Oresme N., *De proportionibus proportionum* and *Ad pauca respicientes.* Ed., E. Grant. Madison, 1966. Latin and English.
65R. Ostrogradsky M. V., On insurance (1847). *Polnoe sobr. trudov* (Complete works), vol. 3. Kiev, 1961, 238-244.
66. ---, Sur une question des probabilités. *Petersb. Bull., t.* **6,** No. 21-22, 1848, 321-346.
67R. ---, *Pedagogich. nasledie* (Educational heritage). M., 1961. Eds. I. B. Pogrebyssky, A. P. Youshkevich.
68R. Pavlovsky A. F., *O veroiatnosti* (On probability). Kharkov, 1821.
69R. Peres Larigno M. T., On the history of the notion of geometric probability. *Voprosi istorii estestv. techniki*, No. 4, 1985, 100-103.
70. Poincaré H., *Science et méthode* (1906). Paris, 1914.
71. Poisson S.-D., Sur l'avantage du banquier au jeu de trente-et-quarante. *Ann. math. pures et appl.,* t. 16, 1825-1826, 173-208.
72. ---, *Recherches sur la probabilite des jugements etc.* Paris, 1837, 2003. **S, G,** 53.
73R. Prudnikov V. Ye., On the articles of P. L. Chebyshev, M. V. Ostrogradsky,V. Ya. Buniakovsky and I. I. Somov from the *Enc. dict.* [38]. IMI vol. 6, 1953, 223-237.74R. ---, *V. Ya. Buniakovsky kak ucheny i pedagog* (V. Ya. Buniakovsky: scientist and educator). M., 1954.
75. Sheynin O. B., Finite random sums etc. AHES, *vol.* **9,** No. 4/5, 1973, 275-305.
76. ---, On the prehistory of the theory of probability. Ibid., vol. 12, No. 2, 1974, 97-141.
77. ---, D. Bernoulli's work on probability. In: *Studies in the history of statistics and probability,* vol. 2. Eds. M. G. Kendall, R. L. Plackett. London, 1977, 105-132.
78. ---, P. S. Laplace's work on probability. AHES, vol. 16, No. 2, 1976, 137-187.
79.---, Laplace's theory of errors. *Ibid., vol.* **17,** No. 1, 1977, 1-61.
80. ---, S.-D. Poisson's work in probability. *Ibid., vol.* **18,** No. 3, 1978, 245-300.
81.---, A. A. Markov's work on probability. *Ibid., vol.* **39,** No. 4, 1989, 337 - 377.
82.---, *On V. Ya. Buniakovsky's work in the theory of probability*. Inst. hist. nat. sci. & technology, Preprint No. 17. M., 1988. IMI, vol. 4 (39), 1999, 57 – 81.
83R. Shiyanov A. N., On the method of least squares. In: Bolotov A. P., *Geodesy,* pt. 1. Psb, 1836, 321-360.
84R. Shusherin P. P., Academician Buniakovsky as a demographer. *Uch. zap. Mosk. Ekon.-stat. Inst.,* No. 6, 1955, 44-54.
85. Simpson T., *The nature and laws of chance*. London, 1740.
86. Soper H. E., *Numerical evaluation of the incomplete B function.* Cambridge, 1921 (Tracts for computers, No. 7).
87R. Steklov V. A., A. A. Markov. *Izv. Ross. Akad. Nauk,* ser. 6, vol. **16,** 1922 (1924), 169-184. **S, G,** 85.
88R. Struve P. B., Who was the first to indicate that statistics can be applied to philological studies? *Izv. Ross. Akad. Nauk,* ser. 6, vol. **12,** No. 13, 1918, 1317-1318.
89R. Tripolsky P. I., *V. Ya. Buniakovsky*. Poltava, 1905.
90R. Vasiliev A. V., *Mathematics,* No. 1. Pg., 1921.
91R. Youshkevich A. P., *Istoria matematiki v Rossii …* (Hist. math. in Russia up to 1917). M., 1968.

VIII

Gauss and the theory of errors

Arch. hist. ex. sci., vol. 20, 1979, pp. 21 – 72

1. Introduction

Some of my papers [125 – 131] were at least largely devoted to the prehistory or early history of the theory of errors. I [133] have also described the work of Laplace who (non-rigorously) created the theory of treating a large number of observations. Here, I am concerned with Carl Friedrich Gauss (1777 – 1855) who studied the treatment of a finite number of observations. The classical theory of errors had thus been born. Many authors [58, 79, 123] have contributed to my subject, but this paper is much more detailed and some new findings are in §§ 2.4 and 5.8. I repeatedly refer to the correspondence of Gauss; their selection is in his *Werke*[1].

2. The principle of least squares before 1809

2.1. Daniel Bernoulli [39] and Euler [68]. Euler had all but introduced the principle of least squares [128, § 1.3; 131, p. 123]. Moreover, taken together, [39] and [68] contain ideas sufficient for Gauss' first derivation of that principle (§ 3.2)[2]. Gauss never referred to these works and the appropriate volume of the *Acta Acad. Petrop.* is not mentioned in the (unfortunately incomplete) list of library books he borrowed during his student's years [64, pp. 398 – 404][3].

Gauss was surprised that the principle of least squares was not discovered earlier (G – O, W-8, p. 140) and was quite prepared to accept that possibility (G - S [30, Bd. 6, p. 89]). During his later years; Gauss, however, decided that he had no immediate predecessors[4].

2.2. Legendre. In 1805 Legendre [96; 131, pp. 123-124] introduced the principle of least squares, declaring that it established *une sorte d'équilibre* among the errors and prevented *les extrêmes* [*erreurs*] *de prévaloir*. The first reference to Legendre's principle appeared in the same year (1805) [116, pp. 137-141; 74]. The ending of that statement was wrong: it is the minimax method which prevents etc.

2.3. Adrain. In 1808[5] the American mathematician Adrain published an article [32], see also [55] and [125], which contained
(1) Two derivations of the normal law of random errors[6].
(2) A derivation of the principles of least squares and arithmetic mean. Adrain considered the case of normally distributed errors, taking as his basis the principle of maximum likelihood.
(3) An application of the principle of least squares to the solution of problems in navigation and surveying.

In 1818 Adrain [33; 34] applied the principle of least squares to the deduction of the size and figure of the earth.

It is possible that Adrain did not arrive at the principle of least squares independently, for he had Legendre's book in his library [55], but since when? Still, the substantiation of this principle (and the principle of the arithmetic mean) as well as his other studies constitute

Adrain's indisputable contribution to probability theory. However, the level of his mathematics was very low and, more than that, for a long time his articles remained unknown, completely or mostly. See also § 2.6.2.

2.4. Gauss. He first used the method of least squares (MLSq) in astronomical calculations, in 1794 or 1795, and he used it regularly from 1801 or 1802 onward. Gauss himself pointed out these facts (see below) which are corroborated by his correspondence[7] and by Olbers's evidence [112, p. 192n]:

Gauss bereits im Junius 1803 die Güte hatte, mir diese Methode [the MLSq], *als längst von ihm gebraucht, mitzuteilen und mich über die Anwendung derselben zu belehren.*

An indirect support for Gauss's claim is provided by his ability, at the very beginning of the 19th century, to calculate the orbit of the new planet, Ceres, which disappeared from observation after its first discovery.

But how did Gauss arrive at the MLSq? I begin with a few passages from his writings.

(1) *Selbstanzeige* (1809) [11, p. 59]:

Die Grundsätze, welche hier ausgeführt werden, und welche von dem Verfasser schon seit 14 Jahren angewandt ... führen zu derjenigen Methode, welche auch Legendre … vor einigen Jahren unter dem Namen <u>Méthode des moindres carrés</u> aufgestellt hat: die Begründung der Methode, welche von dem Verfasser gegeben wird, ist diesem ganz eigenthümlich.

(2)*Theoria motus* (1809, § 186):

Übrigens ist unser Princip, dessen wir uns schon seit dem Jahre 1795 bedient haben, kürzlich auch von Legendre ... aufgestellt worden.

(3) *Selbstanzeige* (1821) [5, p. 98]:

Der Verfasser ... welcher im Jahr 1797 diese Aufgabe [the combination of observations] *nach den Grundsätzen der Wahrscheinlichkeitsrechnung zuerst untersuchte, fand bald, dass die Ausmittelung der wahrscheinlichsten Werthe der unbekannten Große unmöglich sei, wenn nicht die Function, die die Wahrscheinlichkeit der Fehler darstellt, bekannt ist. In sofern sie dies aber nicht ist, bleibt nichts übrig, als hypothetisch eine solche Function anzunehmen. Es schien ihm das natürlichste, zuerst den umgekehrten Weg einzuschlagen und die Function zu suchen, die zum Grunde gelegt werden muss, wenn eine allgemein als gut anerkannte Regel … daraus hervorgehen soll, die nemlich, dass das arithmetische Mittel … als der wahrscheinlichste betrachtet werden müsse.*

Es ergab sich daraus [the normal distribution and] *dann gerade diejenige Methode, auf die er schon einige Jahre zuvor durch andere Betrachtungen gekommen war*[8]*, allgemein nothwendig werde. Diese Methode, welche er nachher besonders seit 1801 bei allerlei astronomischen Rechnungen fast täglich anzuwenden Gelegenheit hatte, und auf welche auch Legendre inzwischen gekommen war, ist jetzt unter dem Namen <u>Methode der kleinsten Quadrate</u> im allgemeinen Gebrauch.*

The last two passages evidently mean that Gauss, like Legendre, first discovered the MLSq as a practical procedure, then substantiated it by theoretical considerations.

2.4.1. Calculus probabilitatis contra La Place defensus (1798). The title of this subsection is a phrase which Gauss wrote in 1798 in his *Tagebuch* (W-10/1, p. 533). He explained his note in a letter to Olbers dated 24.3.1807 [28, No. 1, p. 329]: The principle of least squares, he wrote,

Ist ... dem La Place'sehen vorzuziehen, nach welchem die Summe jener Differenzen = 0, *und die Summe derselben Differenzen, aber sämmtlich positiv genommen, ein Minimum sein soll. Man kann zeigen, dass* [dies] *nach den Gründen der Probabilitätsrechnung nicht zulässig ist, sondern auf Widersprüche führt.*

Gauss returned to this point in 1812 (G-O, 24.1.1812; Ibidem, pp. 493-494):

Ich im Juni 1798 … zuerst La Place's Methode gesehen, und die Unverträglichkeit derselben mit den Grundsätzen der Wahrscheinlichkeitsrechnung in einem kurzen Notizen-Journal ... angezeigt habe[9].

Gauss's criticism evidently relates to one of Laplace's early memoirs [92]; see [130, § 1.3.2][10]. Laplace's or, rather, the Boscovich-Laplace method of adjusting indirect observations leads to a number of zero residuals (§ 3.3), a fact which Gauss considered unfavourable (§ 3.1). It is this point which Gauss evidently had in mind in his letters to Olbers. But then, it is hardly possible to refute Laplace's principle for any distribution $\varphi(x, x_0)$ of errors whatsoever or even for any unimodal and symmetric distribution. Thus, for distribution

$$\varphi(x, x_0) = C\exp[-h^2|x - x_0|]$$

the maximum likelihood estimator of the parameter x_0 is the sample median, to which Laplace's principle also leads [130, § 1.3.4].

2.5. Dispute over priority. Laplace [94, p. 353] objectively, see [115, p. 290], described the discovery of the MLSq. He indicated that Gauss was the first to use this method while Legendre first published it in his book[11]. What Laplace did not add is that the substantiation of the MLSq and a study of numerous related problems are due to Gauss alone.

These facts are unquestionable. Still, Gauss used one careless phrase in his *Theoria motus* (*unser Princip, dessen wir uns schon seit dem Jahre 1795 bedient haben*, see § 2.4) that provoked an attack by Legendre. Quoting this expression (letter to Gauss dated 31.5.1809; W-10/1, p. 380), he indicated in strong wording that priority in scientific discoveries can be established only by publication.

In 1820, having received no answer, Legendre [97, pp. 79-80] launched a full-scale assault against Gauss, see [131, p. 124n; 136] adding for good measure a similar accusation concerning the theory of numbers. Once again, no answer from Gauss followed[12].

2.6. Peculiar features of Gauss's creative work. Gauss would have hardly considered his own words careless or inopportune, regardless of the dispute over priority.

2.6.1. Delays in publication. As a rule, Gauss always needed much time to prepare his apparently completed researches for publication[13], and even Bessel's efforts to convince him that such delays were extremely undesirable proved of no avail[14]. Bessel's first reproach (B – G, 28.5.1837 [27, pp. 516-520; 64, p. 216]):

Sie haben nie die Verpflichtung anerkannt, durch zeitige Mitteilung eines dem ganzen angemessenen Theils Ihrer Forschungen die gegenwärtige Kenntnis der Gegenstände derselben zu befördern; Sie leben für die Nachwelt. Dieses ist aber ganz gegen meine Ansicht.

Gauss answered (G-B, 28.2.1839; [27, pp. 523-529]) that he kept nothing to himself (!) but was pressed for time to prepare his works for publication. The same complaint is found elsewhere. Thus (G-G, 29.12.1839; [29, p. 591]) Gauss indicated that he had to rewrite the *Supplementum* [4] three or four times over; and (G – O, 14.4.1819 [28, No. 1, p. 720] he spoke of linguistic difficulties:

Die spröde lateinische Sprache widersteht oft dem leichten natürlichen Ausdruck des Gedankens.

Lastly (G-S, 9.1.1841; [30, vol. 4, p. 29]) Gauss referred to an additional argument: it is reasonable, he wrote, to postpone publication until getting acquainted with similar work by other authors.

Disregarding Gauss's explanations, Bessel resumed his admonitions (B - G, 28.6.1839; [27, pp. 526-529]): referring to the imperfection of some works of Lagrange (and, by implication, of Euler) he asked rhetorically:

Sollte nicht die Hauptidee selbst, hervortretend in anständiger, wenn auch nicht das Maximum erreichender Darstellung, die Wissenschaft schneller fördern als ihre Vertagung auf die Zeit, welche ihrer allergediegendsten Erscheinung günstig ist? ... Sie können Sich nicht verbergen dass Sie auch das, was Ihnen nicht genommen wird, in die Gefahr des gänzlichen Verlustes bringen.

Gauss's other friends, for example, Olbers, shared Bessel's opinion. In a letter to Bessel dated 25.1.1825 [64, p. 216; 48, p. 12] Olbers maintained:

Ist unser Gauss oft selbst schuld, wenn ihm Andere mit Erfindungen zuvorkommen, die auch er gemacht hat. … Gauss scheint mir aber immer erst selbst die schönsten Früchte pflücken zu wollen ... ehe er Andern denselben zeigt. Ich halte dies für eine kleine Schwachheit des sonst so großen Mannes, um so weniger zu erklären, da er bei seinem unermesslichen Reichthum an Ideen so Vieles wegzuschenken hat.

2.6.2. Attitude toward the work of other authors. Gauss did not consistently follow his own implicit intention to read other authors (§ 2.6.1). Thus, over the years, Olbers informed him that

(1) *Auch ein Amerikaner* [33, p. 122] *schreibt sich ... die Erfindung der Methode der kleinsten Quadrate zu*. (O – G, 24.2.1819; [28, No. 1, p. 711].

(2) An article of T. Young partly devoted to the derivation of the central limit theorem had come out (28.9.1819; Ibidem, pp. 749-751.)

(3) (The first part of) Poisson's article *Sur la probabilité des résultats moyens*, etc. had just been published. (28.1.1825; [28, No. 2, p. 370]).

Gauss made no comment on any of these cases though he should have been directly interested in two of the statements at least. Then, in 1850, Encke [67, p. 333] erroneously attributed the MLSq to Lagrange [90, § 17 from problem 5]. On June 21 Schumacher [30, vol. 6, p. 87] passed the news to Gauss. On June 24 (Ibidem, p. 89) Gauss admitted that, though he was aware of the existence of Lagrange's memoir, he had not read it. Gauss added that he would do so when an opportunity presented itself and that in any case he did not attach any great importance to bare ideas, e. g., to an unsubstantiated introduction of the principle of least squares.

This last assertion can be somewhat disproved by another of Gauss's pronouncements (G-O, 31.12.1814; [28, No. 1, p. 567]):

Der Aufsatz von La Place [95] *ist meinem Urtheile nach dieses großen Geometers ganz unwürdig. Ich finde zwei verschiedene, sehr arge Missgriffe darin. Ich hatte mir bisher immer vorgestellt, dass bei den Geometern vom ersten Range der Kalkul immer nur das Kleid sei, in dem sie das, was nicht durch Kalkul, sondern durch Meditation über die Sache selbst geschaffen, vorführen. Dieser Aufsatz beweist, dass die Regel doch Ausnahmen leidet.*

I do not study Gauss's influence on the development of traditions of intuitionalism; still, it is worth noting that his work on the MLSq before 1809 (§ 2.4) seems to correspond to his general scheme of "meditation - substantiation".

Schilling, the author of the book on Olbers [28], supplies a reference to Gauss's review [14] of Laplace's memoir [95][15]. Laplace had explained the absence of comets with hyperbolic orbits but Gauss found two errors in that work.

Lastly, I quote one phrase from Gauss's *Tagebuch* (1796) [26, p. 66]:

Ein Gesetz ist entdeckt: wenn es auch noch bewiesen sein wird, werden wir das System zur Vollendung geführt haben.

It seems that here too Gauss referred to the same general scheme "meditation-substantiation".

I do not mention Legendre's work [96], which Gauss, so as not to disrupt the sequence of his own ideas, did not try to obtain [12], pp. 275-277]: it does not follow from this statement that he would deliberately refuse to read Legendre until after publishing the *Theoria motus*.

Gauss's reluctance to recognize Legendre's official priority seems to be thus explained. Besides, I must recall also that Gauss was not in the habit of referring to others. He did not mention Lagrange in his basic work on conformal mapping; during a long twenty years he never referred either to K. Jacobi or Dirichlet [48, pp. 17-18]; presenting the *Intensitas vis magneticae terrestris* etc., Gauss

Typically acknowledged the help of Weber but did not include him as joint author [106, p. 305].

Gauss (G-S, 6.7.1840; [30, vol. 3, pp. 385 and 388]) himself admitted that he referred to other authors only if they completely deserved to be mentioned.

As to special *literarische Recherchen* which become necessary in this connection, he, Gauss, was [always] pressed for time and, moreover, felt no inclination for them[15a].

2.6.3. One Conclusion. Any author whose creative work is characterized by the peculiar features described above will apparently seem unattractive. However [48, p. 18],

Was einem normalen Autor verboten ist, einem Gauss wohl gestattet werden muss, zumindest müssen wir seine Gründe respektieren.

Still, I prefer the more overt opinion of May [106, p. 309]:

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.

Later generations spare only Gausses! And this is just what Biermann means.

3. Theoria motus [1]

3.1. Preliminary considerations. Bearing in mind calculations of the orbits of celestial bodies, Gauss (§ 172) notes that the mathematical treatment of a large number of observations is tantamount to combining them properly. In § 173 he indicates that as far as possible the programme of observations should ensure the mutual cancellation of their random errors and that since there was no reason to prefer one or another result, the arithmetic mean of various observations should be adopted.

Lastly, Gauss (§ 174) considers the case of redundant indirect observations:

Da nun kein Grund vorhanden ist, weshalb man … diese oder jene sechs [Beobachtungen] *als absolut genau annehmen soll, sondern da man vielmehr nach den Principien der Wahrscheinlichkeit*[srechnung] *bei allen ohne Unterschied größere oder kleinere Fehler als gleich möglich voraussetzen muss, und da ferner im allgemeinen geringere Fehler häufiger begangen werden als gröbere, so ist es offenbar, dass eine solche Bahn, welche zwar sechs Daten vollkommen befriedigt, von den übrigen aber mehr oder weniger abweicht, für eine mit den Principien der Wahrscheinlichkeitsrechnung weniger übereinstimmende zu halten ist, als eine andere, welche zwar auch von jenen sechs Daten um ein Geringes unterschieden ist, desto besser aber mit den übrigen zusammenstimmt.*

Let the number of mutually independent observations be n ($n > 6$), denote their errors by $x_1, x_2, \ldots, x_n$, and suppose that the density of these errors $\varphi(x)$ is an even and unimodal function. The probability that the series $x_1, x_2, \ldots, x,$ occurs is proportional to $\varphi(x_1)\ \varphi(x_2) \ldots \varphi(x_n)$ and Gauss's reasoning seems to mean that, in general, the condition $x_1 = x_2 = \ldots = x_6 = 0$ decreases this probability. See also §§ 3.2 and 3.3.

3.2. The normal law and the principle of least squares. Formal mathematics begins only in § 175. Starting from the *postulatum* [41,

p. 176] on the arithmetic mean, Gauss (§ 177) proved that among unimodal, symmetric and differentiable distributions there is a unique distribution (the normal) for which the maximum likelihood estimator $\hat{x}$ of the location parameter x_0 coincides with the arithmetic mean.

His proof is as follows: Let M_1, M_2, … be the observations, μ in number, and *p*, their arithmetic mean. Then the likelihood equation is

$$\varphi'(M_1 - \hat{x}) + \varphi'(M_2 - \hat{x}) + \ldots = 0.$$

Here, $\varphi'(\Delta) = d\varphi(\Delta)/\varphi(\Delta)d\Delta$ possesses a (unique) solution *p*:

$$\varphi'(M_1 - p) + \varphi'(M_2 - p) + \ldots = 0.$$

Supposing, then, that $M_1 = M_2 = \ldots = M_1 - \mu N$, Gauss arrived at

$$\varphi'[N(\mu - 1)] = (1 - \mu)\varphi'(-N),\ \varphi'(\Delta)/\Delta = k,\ k < 0$$

for any natural μ, so

$$\varphi(\Delta) = C\exp(k\Delta^2/2). \qquad (3.2.1)$$

Obviously the maximum likelihood of a given series of observations corresponds to the minimum of the squared sum of the discrepancies between the observations and the "true" value of the constant sought. Indeed, Gauss used this simple corollary, but he did so in the general case of adjusting indirect observations (§ 179): if the density law of errors Δ_i is

$$\varphi(\Delta) = \frac{h}{\sqrt{\pi}}\mathrm{C}\exp(h^2\Delta^2),$$

then, as Gauss noted, the function

$$\Omega = h^{\mu}\pi^{-\mu/2}\exp[-h^2(v_1^2 + v_2^2 + \ldots + v_{\mu}^2)]$$

where μ is the number of observations and v_i are the differences between the observed and calculated values of given linear forms of the unknowns sought, attains its maximum value if

$$v_1^2 + v_2^2 + \ldots + v_{\mu}^2 = \min.$$

Gauss added that the principle of least squares *muss überall ... als Axiom gelten*. I am inclined to perceive here a certain deviation from his main train of thought. Gauss also extended that principle to include observations of unequal precision, and noted that it could be applied to the adjustment of heterogeneous magnitudes. He repeated that latter remark (G-G, 2.4. 1840; W-8, pp. 153-154) in connection with the problem of adjusting geodetic networks with measured angles and sides.

In 1829 Gauss [10, p. 28] noticed the similarity between the principle of least squares and the mechanical principle of least constraint:

Es ist sehr merkwürdig, dass die freien Bewegungen, wenn sie mit den nothwendigen Bedingungen nicht bestehen können, von der Natur gerade auf dieselbe Art modificiert werden, wie der rechnende Mathematiker, nach der Methode der kleinsten Quadrate, Erfahrungen ausgleicht, die sich auf unter einander durch nothwendige Abhängigkeit verknüpfte Großen beziehen.

The similarity between the adjustment of direct observations and the determination of the centre of gravity of a system of material points became known in the early 18th century (Cotes). In the 20th century the analogy between geodetic and mechanical systems was exploited, and various versions of the method of geodetic relaxation [124] due to Gauss (§ 6.4) were worked out.

Now I consider in more detail some aspects of Gauss's deduction.

3.2.1. Random Errors. In his *Theoria motus* Gauss did not yet distinguish between random and systematic errors. He (§ 175) considered unimodal and, *im allgemeinen*, errors possessing symmetric density functions. His derivation of the normal law was meant for such errors. He used their properties indirectly, by means of the principle of the arithmetic mean[16]. For his part, Merriman [107, p. 165] noted that the density function arrived by Gauss was not strictly *a law of facility of error* but only a law of distribution of residuals (i. e., of calculated errors). Czuber [58, p. 108] repeated this criticism without referring to Merriman.

3.2.2. The principle of the arithmetic mean. A number of scholars used and even formulated the principle of the arithmetic mean before Gauss did [131, pp. 122-123]; moreover, Simpson and Lagrange, respectively, proved the advantage of that mean over a single observation for two types of distributions and for a whole series of them [129, §§ 1.2.2 and 2].

Gauss himself [7, p. 143], see [131, p. 112] repeated his reasoning on the mean, though less distinctly, and added that the initial observations should be independent.

Many authors, beginning, as it seems, with Encke [65], attempted to reduce the principle of arithmetic mean to more obvious premises. The constructive aspect of such attempts laid the foundation for the modern theory of invariant tests and estimators [98, Chap. 6].

That the arithmetic mean affords the most probable value, as Gauss (§ 175) put it, is not exactly true. However, see above, he restricted himself to unimodal and symmetric densities, for which his reasoning is correct [131, p 123]. See however § 3.2.1 for a qualification remark and, also, §§ 5.2 and 5.9 for a discussion of Gauss's terminology.

3.2.3. The Principle of maximum likelihood. Gauss assumed that, with a uniform prior distribution of the location parameter sought, its estimator should be the mode of the posterior unimodal distribution of the observational errors. To put it otherwise, Gauss had suggested the principle of maximum likelihood which was introduced initially, if rather imperfectly, by Lambert [126] and then, much better, by Daniel Bernoulli [128, § 1.2]. In the same way as Bernoulli, Gauss justified this principle by the fundamental, to quote Laplace [133, p. 6], principle of inverse probability.

3.3. Various methods of adjusting indirect observations. In § 186 Gauss briefly outlined the adjustment of indirect observations under conditions

$$v_1^{2n} + v_2^{2n} + ... + v_\mu^{2n} = \min \qquad (3.3.1)$$

(n is either a small natural number or $n \to \infty$) or

$$w = |v_1| + |v_2| + \ldots + |v_n| = \min. \qquad (3.3.2)$$

He maintained that for finite $n = 2, 3, ...$ condition (3.3.1) leads to involved computations; that the case in which $n \to \infty$ is tantamount to the minimax principle; and that condition (3.3.2) means that, for k unknowns ($k < \mu$) exactly k of the v_i's will be equal to zero.

I [133, p. 50] have commented on the first two statements, and now I discuss the third one[17]. Let

$$a_ix + b_iy + c_iz + \ldots + l_i = v_i,\ i = 1, 2, ...\ \mu,$$

and the number of unknowns be k. Suppose that condition (3.3.2) ought to be satisfied. Change (3.3.2) for 3^n conditions [38]

$$w = \delta_1 v_1 + \delta_2 v_2 + \ldots + \delta_\mu v_\mu = \min$$

in which $\delta_i = -1$ if $v_i < 0$, $\delta_i = 0$ if $v_i = 0$ and $\delta_i = 1$ if $v_i > 0$. Then solve (at least in principle) all the 3^n corresponding problems in linear programming and a set of residuals ($v_1, v_2, \ldots, v_n$) satisfying condition (3.3.2) will be thus chosen.

On the other hand, the solutions of each of these 3^n problems contain exactly k zero v_i's etc. Thus Gauss knew an important theorem in linear programming, but I do not know how he managed to prove it.

4. Bestimmung der Genauigkeit der Beobachtungen [6]

Referring to his *Theoria motus*, Gauss (§ 1) supposes that the frequency function of the errors of observation is

$$p(\Delta) = \frac{h}{\sqrt{\pi}} \exp(-h^2\Delta^2). \qquad (4.1)$$

Here, h is the unknown *Maas der Genauigkeit* [1, § 178], and the goal of Gauss's memoir is to evaluate it.

4.1. The true value of the measure of precision and the probable error of the observation. Setting

$$\theta(t) = \frac{2}{\sqrt{\pi}} \int_0^t \exp(-z^2)dz, \qquad (4.1.1)$$

Gauss indicates that the probability of obtaining m errors equal to

α, β, γ, … in a series of m observations, which is proportional to

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

takes its maximal value when

$$h = H\sqrt{\frac{m}{2(\alpha^2+\beta^2+\gamma^2+\ldots)}}. \quad (4.1.2)$$

Gauss apparently supposes that exactly this H is the true (*wahr*) value of h. [See Sheynin (2007).] However, he continues, the probability of inequalities $H + \lambda \le h \le H + \lambda + d\lambda$ is

$$P = K\exp(-\lambda^2 m/H^2)d\lambda,\ K\int \exp(-\lambda^2 m/H^2)d\lambda = 1.$$

The limits of integration are infinite because of the rapid decrease of the integrand. Then the probability that the true value of h lies in the interval $[H - \lambda, H + \lambda]$ is, see (4.1.1),

$$P = \Theta(\lambda\sqrt{m}/H).$$

In particular, for $P = 1/2$ the corresponding interval is

$$[H(1 - \rho/\sqrt{m}; H(1 + \rho/\sqrt{m}]$$

(for $\Theta(\rho) = 1/2$ the value of ρ is approximately 0.477).
Lastly, calling

$$r = \rho/h \quad (4.1.3)$$

the probable error, Gauss determines the corresponding interval, which he calls "probable", for the true value of r.

The study just described comprises the first four sections of Gauss's memoir. In § 5 he points out that in these sections α, β, γ, … were *bestimmte und gegebene Grossen* but that now he will suppose them to comply to *irgend einem bestimmten Wahrscheinlichkeitsgesetze*. I note that if α, β, γ, … are the "true" errors, then the most probable value of h, formula (4.1.2), will coincide with its mean value, while the corresponding expression for the mean square error of the observations will become

$$\sigma = \sqrt{\frac{\alpha^2+\beta^2+\gamma^2+\ldots}{m}}. \quad (4.1.4)$$

But, if the true errors are *bestimmte und gegebene Grossen*, then the stochastic essence of formulas similar to (4.1.2) seems to disappear.

Somewhat later Laplace [133, §§ 8.2 and 9] studied another measure of precision of observations, but his extremely interesting results directly related only to triangulation. Laplace should also be credited with a formula of the type of (4.1.4). He derived it in a rather roundabout way a year before Gauss.

4.2. Derivation of the probable error of observations

4.2.1. Sums of natural powers of absolute errors. Suppose that a large number (m) of errors α, β, γ, … possess a density function $\varphi(x)$ and write

$$|\alpha|^n + |\beta|^n + |\gamma|^n + \ldots = S_n, \quad \int_{-\infty}^{\infty} x^n \varphi(x) = K_n. \qquad (4.2.1.1)$$

Gauss[18] points out that the most probable value of S_n is mK_n and that the probability for the true value of S_n to belong to the interval $[mK_n - \lambda, mK_n + \lambda]$ is

$$P = \Theta\left(\frac{\lambda}{\sqrt{2m(K_{2n} - K_n^2)}}\right). \qquad (4.2.1.2)$$

His first assertion is not exactly correct: mK_n is the *mean* value of S_n. The second statement, which Gauss did not prove (§ 4.3), is one of the main results of his memoir.

Restricting his study to the case of the normal density (4.1), Gauss arrives at an expression for the most probable value of S_n:

$$mK_n = \bar{S}_n = m\frac{\Pi[(n-1)/2]}{h^n\sqrt{\pi}}, \ \Pi(x) = \Gamma(x+1),$$

Obviously, though Gauss did not say so, H, in formula (4.1.2), should be used here instead of h. However, it is also possible to derive h (or, rather, r) from S_n, and this is just what Gauss does. The probable error r, see formula (4.1.3), is thus

$$r = \rho\sqrt[n]{\frac{S_n\sqrt{\pi}}{m\Pi[(n-1)/2]}}.$$

Gauss also derives the probable intervals for this r Comparing them for various values of n[19], he notes that for $n = 2$ a hundred observations ensure a result as *zuverlässig* as 114 observations in case $n = 1$, as 109 observations in case $n = 3$, etc.

4.2.2. Median of absolute errors. Lastly, Gauss indicates that another, more opportune, though *beträchtlich weniger genau* method of deducing r is possible for normally distributed errors: denote the mean absolute error (or the arithmetic mean of the two middlemost absolute errors) by M; then its most probable value can be taken for r while the probable interval for r will be

$$M[1 \mp \sqrt{\frac{\pi}{8m}}\exp(\rho^2)] = M[1 \mp \frac{0.7520974}{\sqrt{m}}].$$

Gauss did not prove these assertions either but Helmert (1875; 1876) and Lipschitz [100] did. Then Cramér [57, § 28.2] indicated that it is a particular case of the central limit theorem.

Gauss's first statement is evident in regard to the mean value of M. As to the asymptotic coincidence of the mean and most probable values of this quantity, it is a simple corollary of the second statement.

Let

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x_\lambda}\exp(-z^2/2)dz = \lambda = 0.75.$$

Then $x_\lambda = x_{0.75} = 0.6745$. Also let the sample size (m) be such that $m\lambda = 0.75m$ is an integer[20] and the frequency function of the errors be

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp(-x^2/2\sigma^2).$$

Then, according to a rather particular case of a theorem due to Cramér [111, § 2.2], the sample statistic

$$g(x_\lambda)\sqrt{\frac{m}{\lambda(1-\lambda)}}(x_{(m\lambda)} - x_\lambda)$$

possesses an asymptotic distribution

$$g(x_\lambda)\sqrt{\frac{m}{2\pi\lambda(1-\lambda)}}\exp[\frac{-mg^2x_\lambda}{2\lambda(1-\lambda)}]x_{(m\lambda)} - x_\lambda)^2. \quad (4.2.2.2)$$

Integrate the density (4.2.2.2) from – α to α, suppose that the thus appearing probability is equal to 1/2, and note that

$$g(x_\lambda) = \frac{1}{\sigma\sqrt{2\pi}}\exp(-r^2/2\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}\exp(-\rho^2).$$

Then

$$\alpha \approx 0.6123272\frac{r\exp\rho^2}{\sqrt{m}} \approx 0.768782/\sqrt{m}.$$

It seems that in this case Cramér's theorem is not so effective as Gauss's formula (4.2.2.1); actually, though, Gauss's numerical computations are erroneous, and his formula should be written as

$$M[1\mp\sqrt{\frac{\pi}{8m}}\exp(\rho^2)] = M[1\mp\frac{0.786716}{\sqrt{m}}].$$

Dirichlet [62] was corrected Gauss's error and proved the formula in question. Encke [66] first published Dirichlet's study.

Using results obtained by Peters [113], Jordan [80] modified Gauss's formulas in my § 4.2 and applied them to the case of most probable errors.

Jordan also stated that according to Gauss [3, § 37] true errors are discussed in the research just described. In the source to which Jordan refers Gauss mentions his memoir [6] and studies the transition from true to most probable errors but does nothing more. However, taking into account also Gauss's *Selbstanzeige* [5, pp. 103-104], I concur with Jordan: after all, Gauss did mean true errors.

4.3. Addendum: Life and death of the probable error. From the very origin of the theory of probability chances for occurrences and non-occurrences of events were of course compared with each other. The equality of these chances was usually singled out for special study and, in particular, the concept of a probable duration of life (though not the term itself) thus appeared in 1669, in the correspondence of Lodewijk and Christiaan Huygens [134, p. 248].

The probable error was introduced by Bessel [107]. First, in 1815, he used this error as a measure of precision of observations [42, p. 267; 43]. Then, in 1816, he formally introduced the probable error [44], pp. 141-142] and used it in an astronomical context.

Nevertheless, investigations connected with the use of the probable error were mainly due to Gauss (§§ 4.1- 4.2). Two of his later statements (after 1816) are in his correspondence (Gauss-Encke, 25.2.1819, W-12, pp. 200-201; G-S, 2.2. 1825, W-8, p. 143):

(1) *Ist dieses Resultat* [the formula for the mean square error of observations (§ 5.8)] ... *auch von dem Fehlergesetz unabhängig. Allein die Bestimmung der dem wahrscheinlichen Fehler selbst beizulegenden Genauigkeit ist es nicht, dies ist auch eine nicht ganz leichte Aufgabe. Sehr merkwürdig aber ist, dass wenn die Formel* $\exp(-h^2x^2)$ *angenommen wird jene Bestimmung des wahrscheinlichen Beobachtungsfehlers gerade eben so zuverlässig ist, als wäre sie auf* $(n-m)$ *wirklich bekannte Beobachtungsfehler gegründet.*

(2) *Die sogenannten wahrscheinlichen Fehler wünsche ich eigentlich, als von Hypothese abhängig, ganz proscribirt; man mag sie aber berechnen, indem man die mittlern mit 0.6744897 multiplicirt.*

(The second half of the latter assertion is true only for the normal law.)

It seems that, after all, the probable error did not take root in either the German or the Russian literature. However, British and possibly American geodesists used it until recently as witnessed by the first two editions (1952 and 1962) of the fundamental treatise of Bomford [51]. Only in the next edition (1971, pp. 610-611) he reluctantly changed from probable to mean square error.

Nevertheless, by the end of the 19th century both Newcomb and Mendeleev still applied the probable error (Sheynin 2017, p. 191). Gauss himself [19] somehow sticked to it and applied it as though his observations obeyed the normal law (which he had not even mentioned). Traditions die hard!

5. Theoria Combinationis [3]

I do not discuss the *Supplementum* [4] devoted to the adjustment of conditioned observations according to the principle of least squares. Theoretically this subject does not present any essentially new ideas or methods, but it is extremely important in practice and the very fact of its study by Gauss ought to be pointed out.

5.1. Random and systematic errors. Galileo was the first to describe the stochastic properties of usual random errors [104, Chap. 1, § 5]. In the middle of the 18th century random errors were independently studied by Lambert [126, §§ 3.2-3.3], but a distinct bifurcation of errors into random (normally distributed) and systematic (constant) is due to Daniel Bernoulli [127, § 5.2].

Gauss, who hardly knew about Bernoulli's work, distinguished between random (*irregulares seu fortuiti*) and systematic (*constantes seu regulares*) errors (§ 1) and indicated (§§ 2 and 17) that he will not be concerned with the latter, or even with random errors which contain a constant component.

According to him (§§ 1-3), random errors are those unyielding to calculation and caused by the imperfection of human organs of sense or instruments, or brought about by external reasons (e. g., by the *Wallen der Luft*)[21, 22].

Gauss (§ 4) supposed that the density function of errors is unimodal and *in den meisten Fällen* even. Therefore (§ 5),

$$\int_{-\infty}^{\infty} x\varphi(x)dx = 0 .$$

5.2. Measure of Precision. Gauss (§ 6) introduced the variance

$$m^2 = \int_{-\infty}^{\infty} f(x)\varphi(x)dx = \int_{-\infty}^{\infty} x^2(x)dx$$

as a measure of precision and called m (§ 7) the *mittleren zu befürchtendem Fehler, oder einfach den mittleren Fehler* (*errorem medium metuendum, sive simpliciter errorem medium*). He also defined the quantities inversely proportional to m and m^2 as the precision (*Genauigkeit*) and the weight (*Gewicht*) of the observations respectively[23].

Here, Gauss (§§ 6 and 7), did not yet explain that the condition of minimal variance will be the cornerstone of his theory of mathematical treatment of observations. Nevertheless, he (§ 7) indicated that it was expedient to introduce integral measures of precision (or, to put it otherwise, integral measures of error), and pointed out the arbitrariness involved in the selection of $f(x)$.

More definitely Gauss pronounced the same opinion in his letters to Encke and Bessel dated 23.8.1831 and 28.2.1839 respectively (W-8, pp. 145-146 and 146-147):

(1) *Genau besehen hat aber eben deshalb solcher wahrscheinlichster Werth*[24] *nur wenig praktisches Interesse, viel weniger als derjenige Werth, wobei der zu befürchtende Irrthum im Durchschnitt am wenigsten schädlich ist, daher ich* (*außer andern freilich eben so*

wichtigen oder noch viel wichtiger Gründen)[25] *dieses zweite mit dem ersten ja nicht zu verwechselnde Princip vorgezogen habe*.
(2) *Ich müsse es nemlich in alle Wege für weniger wichtig halten, denjenigen Werth einer unbekannten Große auszumitteln, dessen Wahrscheinlichkeit die größte ist, die ja doch immer nur unendlich klein bleibt, als vielmehr denjenigen, an welchen sich haltend man das am wenigsten nachtheilige Spiel hat; oder wenn f*(*a*) *die Wahrscheinlichkeit des Werths a für die Unbekannte x bezeichnet, so ist weniger daran gelegen, dass f*(*a*) *ein Maximum werde, als daran, dass* [the integral of] *f*(*x*)*F*(*x – a*)*dx ausgedehnt, durch alle möglichen Werthe des x, ein Minimum werde, indem für F eine Function gewählt wird, die immer positiv und für größere Argumente auf eine schickliche Art immer großer wird.*

Gauss (§ 7) reasonably noted that the function F (or f, as in the *Theoria combinationis*) should be such that the measure of error increases more rapidly than the error itself. Lastly, he (§ 6) indicated that Laplace [133, § 11.2] hat *die Sache zwar auf eine ähnliche Weise betrachtet*, but that his measure of error is awkward because it *gegen die Stetigkeit verstösst*[26].

5.3. An inequality of Bienaymé – Chebyshev type for unimodal distributions. Gauss (§ 9) studied the probability

$$\mu = P(|\xi| \le \lambda m) = \int_{-\lambda m}^{\lambda m} \varphi(x)dx,\ \lambda > 0,$$

where, in my notation, ξ is the random error of observation[27]. He (§ 10) indicated that for unimodal functions $\varphi(x)$

$$\lambda \le \mu\sqrt{3} \text{ if } \mu \le 2/3;\ \lambda \le \frac{2}{\sqrt{1-\mu}} \text{ if } 2/3 < \mu < 1.$$

This *merkwürdiger Lehrsatz*, as Gauss called it, was formulated by Cramér [57, § 15.7] as

$$P[|\xi - x_0| \ge k\tau] \le \frac{4}{9k^2},\ k, \tau > 0.$$

Here x_0 is the mode of $\varphi(x)$ and $\tau^2 = \sigma^2 + (x_0 - E\xi)^2$ is the second moment relative to the mode (according to Gauss $x_0 = E\xi = 0$). Cramér outlined the proof in Ex. 4 to Chapters 15-20 of his book.

In the opinion of Seal [123, p. 210] Gauss's inequality holds for continuous distributions symmetrical about their single modes; but, then, neither Gauss nor Cramér introduced the condition of symmetry. Seal also supposes that Gauss's willingness to discard the normal [distribution] is explained by his discovery of inequality

$$P(|\xi| \le 2m) \ge 0.89.$$

Seal's argument is indeed interesting, but I still suppose that, at best, it played a subsidiary role for Gauss's change of heart.

5.4. An inequality for the fourth moment of errors. Without substantiation, Gauss (§ 11) formulated the following statement: for unimodal, or, at any rate, for non-increasing density functions

$$\int_{-\infty}^{\infty} x^4\varphi(x)dx \div (\int_{-\infty}^{\infty} x^2\varphi(x)dx)^2 \geq 9/5$$

(the left side assumes its minimal value, 9/5, in the case of the uniform distribution).

This, and even more general statements have been repeatedly proved by many authors, although not always successfully. See von Mises [138, § 2] and Krafft [84] who refer to their predecessors, Winckler, 1866, Krüger, 1897 and Faber, 1922. See also Kendall & Stuart [81, p. 92, Ex. 3.18].

5.5. Distribution of functions of random variables. Laplace time and time again derived the density functions of random magnitudes to solve various concrete problems [129, § 3.5]. For his part, Gauss (§ 11) derived a formula for the density $\psi(x)$ of some function[28] of errors of observations $x_1, x_2, \ldots, x_n$. Supposing that these errors possess a continuous density $\varphi(x)$, he noted that

$$P(0 \leq y \leq \eta) = I = \int_0^{\eta} \psi(z)dz =$$

$$\iint \ldots \int \varphi(x_1)\varphi(x_2)\ldots\varphi(x_n)dx_1dx_1\ldots dx_n,\ 0 \leq y \leq \eta.$$

For $x_1 = f(y, x_2, x_3, \ldots, x_n)$, call it (5.5.1), he also derived, again for $0 \leq y \leq \eta$,

$$\psi(y) = \frac{d}{d\eta}\int_0^{\eta} \psi(z)dz,\ (\eta=y)\ = I.$$

The errors $x_2, \ldots, x_n$ can assume any values for which f exists.

The use of this formula tacitly presupposes that function (5.5.1) is single-valued and that $\partial f/\partial y > 0$. Nasimov [109] considered the general case, in which these restrictions need not hold[29]. Pointing out that it is difficult to calculate separate values of the function ψ, Gauss deduces the mean value of y:

$$Ey = \int_{-\infty}^{\infty} x\psi(x)dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} y\varphi(x_1)\varphi(x_2)\ldots\varphi(x_n)dx_1dx_2\ldots dx_n.$$

In his main example he (§ 15) is concerned with

$$y = \frac{x_1^2 + x_2^2 + \ldots + x_n^2}{s},$$

calculates $Ey = m^2$ and states that the probability of a small difference $|y - Ey|$ increases with s. However, instead of this probability, Gauss calculates the mean square deviation

$$\sqrt{E(y-Ey)^2} = \sigma_y = \sqrt{\frac{n^4 - m^4}{s}},\ n^4 = \int_{-\infty}^{\infty} x^4 \varphi(x)dx.$$

He concludes that[30]

$$m \approx \sqrt{\frac{[xx]}{s}},\ \sigma_m = \sigma_y,\ n^4 \approx \frac{x_1^4 + x_2^4 + ... + x_s^4}{s}.$$

Now, limvar $y = 0$ as $s \to \infty$, so that y is a consistent estimator of m^2. Also, the probability of various values of $|y - Ey|$ can of course be deduced according to Chebyshev's inequality.

5.6. Adjustment of observations. Let $e_1, e_2, ..., e_n$ be the errors of mutually independent observations with $Ee_i = 0$ and $Ee_i^2 = m_i^2$. Then (§ 18) the linear function

$$\lambda_1 e_1 + \lambda_2 e_2 + \ldots + \lambda_n e_n$$

will possess mean error

$$M = \sqrt{\lambda_1^2 m_1^2 + \lambda_2^2 m_2^2 + ... + \lambda_n^2 m_n^2}.$$

In particular, if $m_i = m$ = Const, the mean error will take its minimal value when $[\lambda\lambda]$ = min.

Now, if (§ 20) p unknowns $x, y, z, ...$ are determined from a system of inconsistent equations

$$a_i x + b_i y + \ldots + l_i = 0,\ i = 1, 2, \ldots, \pi\ (\pi > \rho)$$

the zeros on the right side of these equations should be changed for some residuals v_i. Suppose that the first unknown, x, is a linear function of v_i,

$$x = k + [\theta v].$$

It is required to deduce multipliers θ_i for which the mean error of x is minimal (see above), $[\theta\theta]$= min.

Gauss's solution of this problem is somewhat ponderous[31]. Many authors [79, Chap. 5; 123] have described it and I say straight away that it is tantamount to the solution by the MLSq (§ 21). More precisely, the MLSq leads to minimal variance of each unknown (§ 24).

5.7. Adjustment of observations: related problems. Gauss then studies the adjustment of direct observations (§ 22) and derives relations between functions of true and most probable errors (§§ 26 - 27). He also obtains formulas for the second moments Ex^2, Ey^2, Ez^2, …, Exy, Exz, Eyz, … (§ 28) and for the weight of a linear function of unknowns (§§ 29 and 34), and devises a procedure to allow for the change of weight of the initial (the "observational") equations, and for an addition of new equations, without complete recalculation (§§ 35 - 36). Lastly, Gauss notes that the case of dependent normal equations cannot occur in the adjustment of real-life observations (§ 23).

The exposition of these problems is too abstract. However, the subject-matter is well known [79, Chap. 5], and I take up only the last problem. The unknowns can be deduced from the system of normal equations if and only if these *von einander unabhängig sind*. Presupposing this fact, Gauss easily concludes that the determinant of the system of dependent normals is zero; or, rather, he does not introduce any determinants but gives only a general explanation: pointing out that the residuals v_i of dependent equations are invariant in regard to multiplication of the unknowns by one and the same non-zero number, Gauss just excludes such systems from his study.

5.8. The mean square error. Referring to his §§ 15-16 (§ 5.5), Gauss (§ 37) additionally notes that the formula for the mean square error

$$m = \sqrt{\frac{[\lambda\lambda]}{\pi}} \tag{5.8.1}$$

(n is the number of observations) is only valid if λ_i, $i = 1, 2, ..., n$ are true errors. If, however, λ_i are the most probable errors, this formula overestimates precision. The correct formula should then have $\pi - \rho$ in the denominator, call it (5.8.2). Substantiations due to Gauss and subsequent scholars are well known [79, Chap. 5, § 12]. I [133, § 7.1] have described the preceding results due to Laplace as well as Gauss's reasoning about formulas (5.8.1) and (5.8.2).

Continuing his research, Gauss proceeded to study the error of the mean square error (5.8.2) and obtained

$$\frac{2\nu^4 - 4m^4}{\pi - \rho} \le \operatorname{var} m^2 \le \frac{2m^4}{\pi - \rho}. \tag{5.8.3}$$

Here, ν^4 is the fourth moment of the errors, and Gauss naturally did not use the *var*. His somewhat tedious investigation is sufficiently clear. But then, Helmert [76, p. 959; 77, p. 585] noted that

Aus irgendwelchen Gründen nimmt er [Gauss] *die Grenzen* [for varm^2] *nicht so eng, als es möglich ist.*

Kolmogorov et al. [83] voiced a more definite opinion. They repeated Gauss's study in the language of linear algebra and explained the weakness of his estimate by a mere oversight. They, as Helmert before them, also showed that formula (5.8.3) should be rewritten thus[32]:

$$\frac{\nu^4 - m^4}{\pi - \rho} \le \operatorname{var} m^2 \le \frac{2m^4}{\pi - \rho}.$$

A more accurate formula for the variance of the square of the mean square error (5.8.1) is also available [63, p. 226]. Naturally, a similar formula for the variance of the square of the unbiassed estimator (5.8.2) can be also derived.

Formula (5.8.3) is the only one of the whole *Theoria combinationis* into which a parameter of the frequency enters.

5.9. Substantiation of the MLSq. Gauss arrived at the MLSq in § 21 of his memoir (§ 5.6); in his § 17 he remarked that this method might be substantiated in various ways. Gauss added that the choice of a suitable procedure for adjusting observations to ensure the calculation of the unknowns

Mit den kleinsten Fehlern behafteten [ist] *bei der Anwendung der Mathematik auf die Naturwissenschaften eine der wichtigsten Aufgaben.*

He also recalled that according to Laplace [133, §§ 5.2, 6.3 and 6.6] the principle of least squares ought to be preferred for any distribution if only the number of observations is large[33]. Gauss continued:

Wir hoffen ... den Mathematikern einen Dienst zu erweisen[34]*, indem wir bei dieser neuen Behandlung des Gegenstandes zeigten, dass die Methode der kleinsten Quadrate die beste von allen Combinationen liefere, und zwar nicht angenähert, sondern unbedingt, welches auch das Wahrscheinlichkeitsgesetz für die Fehler, und welches auch die Anzahl der Beobachtungen sei, wenn man nur die Definition des mittleren Fehlers nicht im Sinne von Laplace, sondern so, wie es von uns in den Art. 5 und 6. geschehen ist, festgestellt*[35].

Gauss offered a similar description of the new substantiation in his *Selbstanzeige* [5, pp. 99-100], and mentioned this subject in his letters to Encke dated 25. 2.1819 and to Schumacher dated 25. 11.1844 (W-12, pp. 200-201 and W-8, pp. 147-148 respectively). In the former letter Gauss wrote:

Ich beschäftige mich jetzt mit Untersuchungen aus der Wahrscheinlichkeitsrechnung, wodurch die sogenannte Methode der kleinsten Quadrate auf eine neue Art begründet wird, unabhängig von dem Gesetz der Fehler und der Voraussetzung einer großen Zahl der Beobachtungen.

Writing to Schumacher, Gauss enumerated the three possible substantiations. The MLSq, he explained, can be used just because of its convenience, or it can be derived by starting either from the principles of maximum likelihood and arithmetic mean, or from the principle of minimal variance[36]. He concluded:

Nach meiner Überzeugung [the third substantiation] *ausschließlich einzige zulässige ist.*

Introducing the third substantiation (1823), Gauss changed his terminology. In his previous works, he used the term *maxime probabile* (e. g., [1, § 177]: a *maxime probabile* system of values)[37] but throughout the *Theoria combinations,* beginning from § 21 and including the *Supplementum* [4], Gauss employed the expression *maxime plausibiles* (e. g., in § 21, *maxime plausibiles* values)[38]. The third substantiation did not become universally accepted all at once[39].

Even during Gauss's lifetime the so-called theory of elementary errors came to the fore. According to this theory, the error of each observation is composed of a large number of elementary ones. Therefore, the proponents of the theory [73] concluded, the various (non-rigorously proved) versions of the central limit theorem lead to the normality of the errors and, consequently, to the possibility of the second substantiation (1809) of the MLSq.

I also note that while a number of scholars preferred the third substantiation, they did not take such a stand on the point of principle as had Gauss. Thus, in his *Vorwort* to Gauss's collected memoirs [1] Helmert was content to refer to the opinion of Gauss[40]. Similarly, Bertrand [41, p. 268] stated that *La théorie nouvelle* [of 1823] *semble préférable*, but he advanced no clear argument to support his idea.

In 1899 Markov [105, p. 246] come to support definitely but lamely the third substantiation:

I regard only this derivation ... as rational; it was indicated by Gauss. … Keeping to this derivation, we do not hold that the MLSq most probable ... results; we only consider this method as a general procedure by means of which the approximate values of the unknowns as well as a conditionally accepted estimate of the results obtained are furnished. [However] *astronomers prefer another derivation* [of 1809].

Quoting Gauss's letter to Bessel dated 28.2.1839 (§ 5.2), Markov continued (p. 247):

It takes one to be too obstinate to continue to adhere ... to most probable hypotheses, Gauss's opinion just referred to notwithstanding.

Markov also criticized the presupposition of the normal distribution of errors (p. 248):

First, it is difficult to establish the fulfilment of this supposition. Second, the [central limit] *theorem on the limit of probability can be derived only subject to many restrictions, and, third, the conception of an error as a sum of many independent* [elementary] *errors should be properly attributed to the realm of fancyland*[41].

Lastly, Markov disagrees also with Laplace (whom he does not mention) and indicates that in practice the number of observations is always finite[42]. Neyman [110], p. 595] mistakenly attributed the third substantiation to Markov. David & Neyman [60] made the same mistake and even aggravated the situation by proving an "extension of the Markov theorem" actually due to Gauss [123, p. 212]. The role of Markov in the third substantiation of the MLSq from a purely mathematical point of view is now disputed [114, p. 460; 123, p. 212], and correctly so. However, Markov should be undoubtedly named as the scholar who revived Gauss's opinion[43].

6. Geodesy

I briefly describe Gauss's field work and remark on the importance of his geodetic computations.

6.1. Triangulation: general features. In 1802-1807 Gauss [71a] accomplished a micro-triangulation, all by himself and for his own pleasure. He measured the angles by a sextant, and he possibly used the principle of least squares for the adjustment of the coordinates of the intersected points. (Gerardy mostly discussed elementary intersections and was not quite definite on this most important point.)

That work was only a *Vorübung*. Some fifteen years later Gauss became responsible for the accomplishment of, and directly involved in every technological operation connected with, the triangulation in Hannover. See his correspondence and official reports, W-9 and Galle [71].

The Hannover triangulation suffered from shortcomings, especially from the complexity of the system of triangles [37, p. 11]. The reason, as Gauss [22, p. 425] himself indicated, was that the original goal of the triangulation fell short of subsequent, much more ambitious plans.

6.1.1. Preparatory work. Gauss regarded every technical operation with due attention. Thus (G - B, 15. 11.1822; W-9, p. 353), he pointed out:

Immer machte ich es mir zum Gesetz, mit der Rechnung[44] *allen Messungen, wie ich sie erhalten hatte, gleichen Schritt zu halten* (*bis auf die allerletzte Zeile*), *und nur dadurch ist es möglich geworden, alle Durchhaue mit der äußersten Präcision so durchzuführen, dass auch nicht Ein Stamm ohne Noth gefällt ist, oder die Unmöglichkeit der Durchhaue so früh wie möglich bestimmt zu erkennen*.

It looks as if Gauss not only considered economic factors but took care of nature.

6.1.2. Form of triangles. In his letter G – O, 8.7.1824 (W-9, p. 371) Gauss indicated that he sometimes tolerated rather acute angles in his triangles if only the corresponding sides were not "transitional"[45]. Gauss added that he tried to avoid such angles not because he hoped

Dadurch an Genauigkeit etwas zu gewinnen, sondern aus dem wohl verzeihlichen Wunsche, dem System so viel möglich, außer dem inneren Gehalt, auch Schönheit und Rundung zu geben.

These words likely characterize Gauss and his creative work in general. The problem of acute angles can hardly be solved in an abstract way: Gauss himself (Ibidem) began to doubt the correctness of his opinion while authors of this, 20th century either repeat Gauss [51, p. 7] or put forward more general considerations and claim [85, pp. 72-73] that any side of a triangulation can become a transitional side of a new network and therefore should be determined with sufficient accuracy.

6.1.3. Precision. On a number of occasions Gauss advocated that triangulation should be measured with utmost precision (G – G, 5.10.1821, W-9, p. 380) and letter to Spehr dated 18.11.1828 (W-12, p. 98). See also his contribution on the Hannover triangulation [21, p. 404].

In the last instance he maintained:

Bei der trigonometrischen Vermessung eines Landes ist es ... in mehreren Rücksichten allerdings rathsam, die Genauigkeit in der Bestimmung der gegenseitigen Lage der Hauptpunkte so weit zu *treiben, wie es der Zustand der Kunst und die Umstände nur zulassen, zumal da es dann in unzähligen Fällen möglich wird, hinreichend genau abgeleitete Bestimmungen secundärer Punkte mit äußerst geringer Arbeit und durch Methoden zu gewinnen, die ohne jene Voraussetzung ins Wilde führen würden. … Je mehr die in verschiedenen Theilen von Europa ausgeführten trigonometrischen Messungen mit einander in Verbindung kommen und nach und nach sich einem großen Ganzen nähern werden, desto mehr erhalten die einzelnen Bestandtheile den Charakter eines kostbaren Gemeinguts von einem für alle Zeiten bleibenden Werthe, und desto wichtiger wird es, alle wesentlichen Momente derselben in solcher Vollständigkeit aufzubewahren, dass ihre Zuverlässigkeit im Ganzen wie im Einzelnen stets geprüft werden könne*.

Gauss also spoke out on the international importance of triangulation in his letters G - O 13.1.1821 (W-9, p. 368) and to Bonenberg dated 16 11 1823 Ibidem, p. 365).

6.1.4. Permanence. For a long time now, the permanence of triangulation has been ensured by special marks fixed into concrete blocks buried under the triangulation stations. It is the centres of such marks that the coordinates of the stations are related to.

In Gauss's times no precautions of this kind had been taken but Gauss [22, pp. 414, 420 and 424-426] made it a rule to include also permanent local structures, mostly churches and bell towers, in his triangulation. Thus, on p. 426 he reported:

Kirchthürme werden im ganzen [Hannoverschen] *Königreiche nicht viele ohne Bestimmung geblieben sein*.

6.2. Errors in measurement of angles. Apparently by the end of the 18th century the so-called repeating theodolite came into use [54, p. 13]. The method of repetition was introduced and, consequently, the accuracy of triangulation increased considerably. It seems that everyone concerned was satisfied except Gauss, who detected a small systematic error inherent in this method. But then, Gauss also proposed an effective procedure to eliminate this error from the final results. According to one source, he (G-G, 8.4.1844; [29, p. 677]) discovered this procedure in 1825 though he had noticed the error not later than 1824 (G-O, 12.11.1824; [28, No. 2, p. 356]).

Then, in 1825, Gauss described the behaviour of this systematic error (G-O, July 1825; W-9, pp. 490-491; G-S, 14.8.1825; W-9, pp. 493 - 494); in discussing his geodetic work of 1821-1825 he mentioned his procedure (G-B, 29.10.1843; Ibidem, pp. 494-495). Lastly, 15.8.1844 Gauss (Ibidem, pp. 498-499) communicated to Bessel some additional measures[46].

Gauss also studied other essential errors, both random and systematic.

(1) He informed Olbers (G-O, July 1825 and 14.5.1826; W-9, pp. 491-492 and 320) and Schumacher (G-S, 14.8.1825; Ibidem, p. 493) about the systematic influence of lateral refraction.

(2) In two letters to Schumacher he described his inquiry into the errors of graduation of the measuring circle. The first letter was written 10.7.1826 and the second one sometime between July 14 and September 8 of the same year [30, Bd. 2, pp. 59 and 65]. In the former

he outlined his ideas on the elimination of these errors by an expedient programme of observations.
(3) In his letters G - S 12.1.1824 [30, Bd. 1, pp. 360-363] and G – O, July 1825 (W-9, p. 491) Gauss discussed systematic errors of sighting.

An inescapable fact is that the end results of measurements always were, and are, more or less distorted by random and systematic errors. Naturally, observers should be guided by some specifications; and, speaking about triangulation, geodesists must not quit their stations until obtaining sufficiently plausible measurements. But when can an observer consider his measurements complete?

Apparently Gauss never adhered to any definite programme of measuring angles. At one of his stations he measured six angles with weights varying from six to seventy-eight (G-G, 26. 12.1823; W-9, pp. 278-281)[47]. To put it otherwise, one of the angles was measured six times, another one, seventy-eight times. This striking difference was likely necessitated by different conditions of measuring the corresponding directions (in particular, by different random and systematic influences) rather than by Gauss's desire to ensure a formal equality of the relative variances. A similar conclusion is suggested by observational records at other stations of the Hannover triangulation (W-9, pp. 263 and 268-269). Note, however, that other observers at least sometimes confined themselves to the rule of measuring each angle at a given station with an equal number of sets (Ibidem, p. 273).

By the end of the 19th century triangulations had spread over vast territories of the world and in any case in some countries (India, United States)[48] rigid programmes of observation were introduced at that time or, possibly, later [51, p. 24]. The fixing of one or another definite programme was likely occasioned by several reasons:
(1) The scatter of measurements at a given station does not sufficiently characterize the influence of systematic errors.
(2) Some systematic errors, as, for example, errors in the graduation of the measuring circle, are eliminated to a larger extent if the number of measurements is known beforehand and the observations correspondingly planned.
(3) Other systematic errors (e.g., the systematic influence of the lateral refraction) are excluded in a greater measure if observations are protracted over a certain period of time. This fact, which is especially true for "unfavourable" directions, by itself makes it desirable to fix a certain minimal number of measurements.

6.3. Rejection of outlying observations. In astronomy and geodesy, rejection is a delicate and important procedure whose general history I have described in preceding articles [131, § 3; 133, § 12.1]. In particular, I quoted Gauss's opinion (G – O, 3.5.1827, W-8, pp. 152-153).

The essence of his idea was that all observations should be recorded and that, lacking an *allemal umfassende Sachkenntniss*, rejection is

Immer misslich, wenn nicht die Anzahl der vorhandenen Beobachtungen sehr groß ist,

and risky and leads to overestimating the accuracy of observations[49]. Gauss expressed his views in reply to Olbers's request (O – G, 28.4.1827; [28, No. 2, p. 477]) to formulate a definite rule for rejection. Gauss's opinion proved characteristic of at least the beginning and middle of the 19th century[50].

6.4. Adjustment of observations. Gauss devoted a few writings [20], [13], separate sections of two memoirs [2, §§ 13-15; 4, §§ 23-24] and a number of posthumously published manuscripts (W-9) to the adjustment of geodetic networks and astronomical observations.

Besides, in his letters to Gerling he discussed various problems of the adjustment of triangulation. Thus Gauss described station adjustment and the treatment of geodetic systems with measured sides and angles (G- G, 26.12.1823 and 29.12.1839; W-9, pp. 278-281, [29, pp. 588-592]) and calculated the number of conditions arising in networks (G-G, 5.6.1838; W-9, pp. 323-324). The first of these three letters describes a method for solving a system of linear algebraic equations by a version of the method of successive approximations, or by the method of relaxation as it is now called [69][51].

In two places Gauss explained how to fix the weights of observations (G-O, 20.4.1812, W-12, p. 247; G-S, 22.2.1850; [30, Bd. 6, pp. 64-67]). In the first instance he indicated that the structure of his formula for calculating the most probable value x_0 from given observations x_1, x_2, ...

$$x_0 = \frac{e_1^2 x_1 + e_2^2 x_2 + ...}{e_1^2 + e_2^2 + ...}$$

(*Theoria motus*, see § 5.2) depended on the presupposed density law of errors (the normal law) but that in no case linear functions should be applied in that fraction.

Gauss did not prove his second assertion. However, it is not difficult to show that even for two observations with density laws

$$y = \sigma_i^{-1}\varphi[\sigma_i^{-1}(x - x_0)],\ i = 1, 2$$

the principle of maximum likelihood cannot lead to linear functions for x_0.

Thus, in 1812 Gauss was prepared to admit the existence of various density functions of errors, but, for example, in 1845 he returned to the principle of the arithmetic mean (§ 3.2.2).

Some information on the volume of adjustment calculations carried out by Gauss could be gleaned from his correspondence and other sources, from [19] in particular[52] [sources are indicated in my original text of 1979, p. 53]:

1823, 76 directions
1826, 150 directions, 55 normal equations
1828, 171 observations, 46 normal equations
1837, ca. 40 triangles
1839 or 1840, 62 conditions

In Gauss's time all this work, to say nothing of the preliminary calculations, such as, for example, the station adjustment, demanded considerable efforts indeed. Describing his work of 1826, Gauss justly remarked: *Es hat vielleicht noch niemals jemand eine so complicirte Elimination* [solution of normal equations] *ausgeführt*.

Still, it is much more important to note that Gauss's method of solving systems of linear algebraic equations became canonical. It was practically the only one used before electronic computers were introduced, and it is widely used even now[53].

Largely owing to the exceptionally convenient notation, also due to Gauss [2, § 13; 4, § 5 et seq.] this method is elegant and simple. He used symbols [*ab*], [*ac*] etc. to denote the coefficients

$$a_1b_1 + a_2b_2 + \ldots + a_nb_n,\ a_1c_1 + a_2c_2 + \ldots + a_nc_n, \text{ etc.} \qquad (6.4.1)$$

of the normal equations. Even now, natural as it is to treat quantities (6.4.1) as scalar products of corresponding vectors, Gauss's notation did not fall into disuse. But besides this notation Gauss [2, § 13] introduced special symbols to stand for the quantities which appear in the (Gauss's method of) solution of normal equations

$$[bc,1] = [bc] - [ab][ac]:[aa],\ [cd,2] = [cd,1] - [bc,1][bd,1]:[bb,1].$$

These symbols, only without the commas, are also used nowadays at least in geodetic literature.

6.5. The master of experimental science. Even before the Hannover triangulation was under way Bessel regarded Gauss as a master of experimental science (B - G 15.6.1818; [27, p. 272]):

Wir verdanken Ihnen den größten Teil der heutigen Verfeinerung der Astronomie, nicht nur wegen Ihrer kleinsten Quadrate, sondern auch wegen der Erweckung des Sinns für Feinheit der seit Bradley's Zeit[54] *von der Erde verschwunden zu sein schien und erst seit 18 Jahren wieder erschien. Wir sind erst jetzt auf den Punct gekommen, kleinen Fehlern oder Abweichungen außer den Grenzen der Wahrscheinlichkeit*[55] *mit derselben Aufmerksamkeit nachzuspüren als früher großen*.

Subbotin [137, pp. 246, 248] voiced a similar opinion:

All his [Gauss's] *activities … testify to the fact that, like Newton, he was not only a mathematician, but in a no lesser degree a natural scientist, and that he felt the necessity of direct contacts with nature, with real life*[55]. …

Gauss never thought of problems connected with accumulation of extensive observational data[56], *such as compilation of star catalogues, determination of fundamental astronomical constants, etc. He was interested in methods to analyse instruments, to determine and to allow for instrumental errors. Gauss and Bessel are the originators of a new trend in astrometry.*

Astronomers of old aspired to amass observations rather than to increase their precision[57] *and very often relied upon the adjustment of instruments, but after Gauss and Bessel everything is based on analysis of the instrument, on the fullest possible determination of its* *errors and on allowing for the influence these errors may have upon the results of observation*.

Lastly, Subbotin (p. 297) notes a trait peculiar to Gauss as an astronomer:

The apparently striking underestimation and almost complete oblivion [until the end of the 19th century] *of the works of Lagrange and Laplace* [on the determination of planetary and cometary orbits from a minimal number of observations] *was caused by the fact that these authors restricted themselves by the purely mathematical aspect of the problem whereas Gauss thoroughly worked out his solution from the point of view of computations taking into account all the conditions of the work of astronomers and* [even] *their habits*[58].

Subbotin could have added a similar remark about adjustment calculations in geodesy. That geodesists no longer use Laplace's work related to error theory is largely due to Gauss's careful elaboration of the practical side of problems discussed, to his exceptionally successful notation (see above) etc[59]. Bessel's opinion (B – G, 12.12.1826; [27, p. 468]) can be quoted here:

Unter Ihren jetzigen Arbeiten wird mich kaum eine andere so sehr interessiren als die Abhandlung über die Anwendung der Methode der kleinsten Quadrate auf die geodätischen Messungen ([20]?).

But then, Bessel also reproached Gauss with an apparently superfluous passion for field geodetic work (B – G, end of 1822; [27, p. 415]):

Ein Dreieck oder zwei wären genug, um alles kennen zu lernen, was etwa entgangen wäre, und das übrige müsste N. N. machen und nicht Gauss.

Gauss (G-B, 15. 11.1822; W-9, pp. 355-356) answered:

Ich fühle oft … bei dieser wie bei so vielen andern Gelegenheiten, wie meine äußern Verhältnisse mich an weitaussehenden theoretischen Arbeiten hindern. Wenn solche ganz gedeihen sollen muss man sich ihnen ganz hingeben können und nicht durch so heterogene Arbeiten wie Collegia lesen, alles kleinliche Detail beim Observiren und Rechnen der Beobachtungen, etc. etc. stündlich gehindert werden.

Besides (p. 357), everyday calculations

Immer einige Unterhaltung gab [while] *das Bemerken, Ausmitteln und Berechnen eines neuen Kirchthurms wohl ebenso viel Vergnügen machte, wie das Beobachtungen eines neuen Gestirns.*

Later Gauss (G-B, 14.3.1824; [27, p. 428]) added:

Alle Messungen in der Welt wiegen nicht ein Theorem auf, wodurch die Wissenschaft der ewigen Wahrheiten wahrhaft weiter gebracht wird.

However, time for important studies is all the same lacking while observations would not be accomplished without him (Gauss) and he needs take into account the vital requirements of a numerous family[60].

Nevertheless, Gauss did intend to withdraw gradually from field work [23, p. 488]:

Was nun aber die Messungen ersten Rangen betrifft, die ich bisher allein auf mich genommen habe, so hoffe ich, dass es späterhin *möglich sein wird, auch die andern Officiere nach und nach zu solchen feiner Arbeiten einzuüben.*

7. Theory of probability and demography

7.1. Theory of probability: general features. Gauss lectured on probability theory, and his lectures [61, p. 305] included

Eine überaus klare und durch originelle Beispiele erläuterte Entwicklung der Grundbegriffe und der Hauptsätze der Wahrscheinlichkeitsrechnung.

It also seems (Ibidem) that his course on probability embraced a study of the theory of definite integrals[61]. Gauss undoubtedly knew the theory of probability as developed in his time[62]. Besides, his correspondence and *Nachlass* contain extremely interesting contributions to this theory (§§ 7.2-7.5). Lastly, Gauss paid due attention to the principles governing the application of stochastic reasoning to natural science. Thus, in 1845, in a letter to Benzenberg he [47] spoke out against a stochastic proof of the diurnal rotation of the earth. Even an extremely high probability of the rotation, Gauss indicated, is no substitute for a deterministic proof.

In the same letter Gauss voiced his reservation about the principle of inverse probability[63]:

Wenn man aus der Wahrscheinlichkeit des Eintretens eines Ereignisses auf die Wahrscheinlichkeit der Ursachen zurückschließt, so ist dies ein schlüpfriger Boden.

Writing to Fries 12.2.1841, Weber (W-12, pp. 201-204) described some thoughts on probability which Gauss confided to him[64]. Weber explained here, in more detail than Gauss himself in the subsequent letter to Benzenberg (see above), that stochastic reasoning is admissible only when nothing is known about the essence of the phenomenon studied.

Er [Gauss] *gab Ihnen gleich von Anfang darin Recht, dass in den Anwendungen der Wahrscheinlichkeitsrechnung sehr gefehlt werden könne, wenn man nur auf die Zahlen bauet, welche wiederholte Beobachtungen geben, und nicht jeder andern Kenntnis, die man sich von der Natur der Sache und deren Verhältnissen verschaffen kann, ihr Recht widerfahren lässt, so schwer dies oft auch sei. ... Die französischen Mathematiker hätten wohl diese Vorsicht nicht immer genug beobachtet*[65]*. Gauss ... hat beim Vortrag immer vorausgeschickt: die Wahrscheinlichkeitsrechnung habe den Zweck nur in solchen Fällen eine bestimmte Auskunft zu geben, wo man außer den Beobachtungszahlen nichts weiter von der Sache wisse oder berücksichtigen wolle. ... Der hohe Werth der Wahrscheinlichkeitsrechnung besteht ... darin, dass sie gerade in den Fällen, wo gar keine andern Kenntnisse vorliegen ... irgend eine Richtschnur an die Hand gibt: z. B. bei der Einrichtung einer Leibrentenanstalt. Ebenso kann die Wahrscheinlichkeitsrechnung dem Gesetzgeber eine Richtschnur für die Bestimmung der Zahl der Zeugen und der Richter geben, wenn sie auch für den einzelnen Fall nichts lehrt.*

And Poisson (1837) had indeed studied the administration of justice to determine the jurors' optimal majority vote needed for condemning the accused.

Weber also pointed out that the recurrence of a phenomenon (e. g., of the daily sunrise) brings about a better knowledge of the law which governs it. But, Weber continued, laws of organic (?) life are unknown so that

Dort [in astronomy] *folgt aus dem Ausbleiben einer erwarteten Erscheinung dass man ein in der Natur wirkendes Element übersehen hat: wir würden also die Wahrscheinlichkeit eines solchen Übersehens vorher zu schätzen haben. Ganz anders verhält es sich z. B. mit der Verbindung der Thiere, aus der junge Thiere hervorgehen, man weiß nicht wie. Hier hält man sich bloß an die wiederholte Beobachtung des Factums, und die Wahrscheinlichkeit wächst mit der Wiederholung*.

A few comments.

(1) The theory of probability deals with the laws of mass phenomena rather than with events whose essence remains unknown. This, however, does not mean that Gauss's opinion about the recurrence of sunrises is erroneous.

(2) The criticism of Laplace's study of cometary orbits is rather slight. See Cournot [56, Chap. 12] who proved that in this instance Laplace did not reason out the classification of events into remarkable and ordinary ones.

(3) Gauss did not reject out of hand applications of probability to jurisprudence.

7.2. The Inversion formula for the Fourier transform. A posthumously published note [15] includes an inversion formula for the Fourier transform of the density function. The note was possibly written after the appearance of the corresponding works of Fourier, Cauchy and Poisson, and, in the opinion of the editor, in any case not before 1814. The title of the note (*Schönes Theorem der Wahrscheinlichkeitsrechnung*) is indeed provocative [122, p. 79].

7.3. The first problem in the metric theory of numbers. In a letter to Laplace dated 30.1.1812 (W-10/1, pp. 371-374) Gauss formulated a problem, the first one in the metric theory of numbers [72]. A certain number M, $0 < M < 1$, is expanded into a continued fraction

$$1/a_1 + \ldots$$

Required is the probability $P(n, x)$ for the tail of this fraction

$$1/a_{n+1} + \ldots$$

to be less than x.

If $P(0,x) = x$, i. e., if all permissible values of M are equally probable, then, according to Gauss,

$$\lim P(n,x) = \frac{\log \text{nat}(1+x)}{\log \text{nat} 2},\ n \to \infty. \qquad (7.3.1)$$

However, he was unable to deduce an asymptotic formula for $P(n,x)$.

Gauss mentioned this problem in 1789 and again in 1800; in the second instance he wrote [26, p. 77]:

Das Problem aus der Wahrscheinlichkeitsrechnung hinsichtlich der Kettenbrüche, das einstmals vergeblich untersucht worden ist, haben wir gelöst.

This phrase, or, rather, its Latin original, is the title under which Gauss's problem is published in his *Werke* [24].

Formula (7.3.1) was proved by Stäckel (W-10/1, pp. 554-556) and then by Kuzmin [88], [89], who also derived an asymptotic expression for $P(n,x)$.

7.4. Elements of the theory of random arrangements. Gauss's *Nachlass* includes a short note [17] which now relates to the theory of random arrangements[66]. Perhaps he became interested in this theory when studying the distribution of cards (e. g., aces), discernible or not, between players, see § 7.6.

Supposing that the arrangements are "purely random" and denoting the number of places by $p = 1/x$ and the number of objects by m, Gauss calculated probabilities (m, n) that the objects occupy $m – n$ places, i. e. that $[p – (m – n)]$ places remain empty. He calculated

(2.0), (2,l), (3,0), (3.1), (3,2), (4,0), (4,l), (4,2) and (4,3)

Then Gauss discovered the rule for deducing the coefficients of these products and wrote out an equation in finite differences for the expected value of n,

$$En = (m, 1) + 2(m,2) + 3(m,3) + n(m, n)$$

He solved it:

$$En = [(1 – x)^m – (1 – mx)]:x.$$

7.5. Expected values of functions of random variables. In another short note Gauss [16] discussed binomial trials. Suppose event E happens at each trial with probability p. Then, in n independent trials event E appears μ times ($0 \le \mu \le n$) and $E\mu = pn$. Indicating these generally known facts, Gauss also pointed out that

$$E[\mu(\mu – 1)] = n(n – 1)p^2, \qquad (7.5.1)$$
$$E[\mu(\mu – 1)(\mu – 2)] = n(n – 1)(n – 2)p^3, \qquad (7.5.2)$$
$$E[(\mu – pn)^2] = pqn,\ q = 1 – p. \qquad (7.5.3)$$

Formula (7.5.1) can be derived by starting from the expression for the variance

$$\text{var}\mu = pqn = E\mu^2 – (E\mu)^2 = E\mu^2 – p^2n^2,\ E\mu^2 = pqn + p^2n^2, \ldots$$

It is also easy to deduce formula (7.5.2). Indeed,

$$P[\mu = k) \equiv p_k = C_n^k p^k q^{n-k},\ k = 0,1,2,...,n,$$

$$\mathrm{E}[\mu(\mu-1)(\mu-2)] = \sum_{k=3}^{n} k(k-1)(k-2)p^k = \sum_{k=3}^{n} k(k-1)(k-2)C_n^k p^k q^{n-k} =$$

$$n(n-1)(n-2)p^3 \sum_{k=3}^{n} C_{n-3}^{k-3} p^{k-3} q^{n-k} =$$

$$n(n-1)(n-2)p^3 \sum_{\alpha=0}^{\beta} C_\beta^\alpha p^\alpha q^{\beta-\alpha} = n(n-1)(n-2)p^3.$$

This equation seems evident; however, it can also be derived by use of the generating function

$$P(s) = p_0 + p_1 s + p_2 s^2 + \ldots + p_n s^n$$

of the quantity μ:

$$E[\mu(\mu - 1)(\mu - 2) = P'''(1).$$

A similar derivation of (7.5.1) is of course possible. Lastly, formula (7.5.3) needs no substantiation: it is the known expression for the variance varμ.

Thus Gauss derived the mean values of some functions of a random magnitude distributed according to the binomial law. Did he conduct similar studies for other distributions? This is unknown.

7.6. Collection of statistical data. Over the years Gauss felt a strong predilection to compile statistical data, see G – O 26.10.1802 [28, No. 1, p. 106] and his letter to Humboldt dated 14.4.1846 [31, pp. 92-97]. In the second instance Gauss maintained that in mortality statistics, as well as generally in science, important progress can be achieved if research were not confined to requirements of direct applications. Bearing this in mind, Gauss continued, he was mainly interested in mortality of infants (reasons for the death of infants are more evident) and of the very old. He also made known his (academic) desire to obtain data on deaths caused by lightning and on the number of lightning bolts per year per unit of the earth's surface[67].

Von Waltershausen [139, p. 89] reported that Gauss had collected data on the longevity, expressed in days, of many distinguished persons, his late friends included, and recorded the dates of storms; moreover his study of economic and financial statistics brought him a lot of money[68].

Lastly, Gauss kept a special register of the distribution of cards in games which he often played with his friends [64, p. 227].

Studying one or another aspect of Gauss's work, commentators seldom refer to his teachers which is quite natural because his creative work is original and profound. Still, concerning Gauss's interest in the collection of statistical data, it seems worthwhile to mention the name of Professor E. A. W. Zimmermann (1743-1815) from the Brunswick Collegium Carolinum [35]. Zimmermann delivered lectures on mathematics, physics, natural history and physical geography while his scientific activities also included statistics. In 1849, recalling his years at the Collegium, Gauss gratefully remembered

Vor allem aber der väterlichen Freundschaft des edlen, alle seine wissenschaftlichen Bestrebungen auf jede mögliche Weise befördernden Zimmermann.

At least one of Zimmermann's statistical works appeared in an English translation, and Yule [141] credited him with the introduction of the words *statistics* and *statistical* into English.

7.7. Study of the laws of mortality

7.7.1. Infant mortality. Gauss introduced two empirical laws of mortality. The first of these (G-S, 12.7.1847; W-12, pp. 71-72) relates to the number of newly born (x) who live to be n months old and is based on the data collected by Quetelet [118, p. 170] in Belgium:

$$x = 100{,}000 – A\sqrt{n},\ \log A = 3.98273.$$

Here 100,000 is the initial number of children born. As Gauss himself noted, his formula is similar to the one due to Moser[69].

[My original text of 1979 includes a table of differences between statistical data for months of life 1(1)6, 12 and 18 and the formula above. For the first six months of life those differences do not exceed 44, but then they amount to 470 and 1448.]

Gauss maintained that his formula ensured a good approximation, and noted that for other values of A it might be used for other countries. His assertion seems justified because the initial differences up to and including 12 months hardly surpass inevitable errors of the initial data.

7.7.2. Mortality of members of tontines (*source unspecified*). Denoting the number of persons living until age 3 and n by a and a/x respectively, Gauss [18] supplied a table of $\log x$

$$\log x = A + Bb^n – Cc,$$
$$\log B = 4.66231,\ \log C = 1.67925,$$
$$\log b = 0.039097,\ \log c= – 0.0042225$$

and adduced the values of A_n computed for n = 3 and 7(5)97.

[My original text of 1979 includes a table of differences between the statistical data and the formula above.]

I checked the calculation of A_n; the difference between my results and the values arrived at by Gauss do not surpass two units of the last digit. The mean value of A_n is 0.48301 whereas Gauss provided 0.48213; he apparently failed to include A_{82} = 0.49766 in his calculations. But is there any sense in that A_n? Gauss did not explain the method of calculating the four parameters of his law.

Gauss's law is a particular case of the formula due to Lazarus (1867) [101][71]. As stated above, Gauss did not indicate the source of his data. I think he used a mortality table compiled for ages 3, 4, ..., 95 *sur les listes des* [French] *tontines 1689 et 1696* [61a, table 13] I compare the number of living persons (N) for ages 3, 7, 12, ..., 97 according to Deparcieux [61a] and Gauss. For the latter

$\log N = \log a - \log x.$

Deparcieux begins his table with $a = 1{,}000$, and I assume exactly this value for the calculation of N. Ages, as understood by Gauss and Deparcieux, seem to differ by a half-year; for this reason age (n – 1/2) rather than age n is chosen for comparison.

Calculating his mortality table, Deparcieux adjusted the data on the two tontines [61a, tables 6 and 7] but did not elaborate. In turn, Gauss somewhat adjusted Deparcieux's table. For example, the former assumed that $N_{97} > 0$, which is not the case with the latter.

[See the table of the results of both authors in my original text of 1979. Here, I note that their results differ by 5 – 7 deaths for ages 42 – 77 and even less for 82, …]

7.8. Life insurance. One of Gauss's posthumously published writings [7], which I mentioned in § 3.2.2, is a report on the activities of the widow's fund at the University of Göttingen. Drawing on statistical data from various sources, Gauss managed to solve a number of important practical problems related to the work of the fund.

Besides, some of the suppositions Gauss used (for example, those concerning probabilities of marriage) *zuweilen auch heute noch angewendet wird* [135, p. 65*].

Gauss [8] also compiled a table for the cost of life annuities.

7.9. Gauss and Quetelet. In the mid-19th century statistics developed under the dominating influence of Quetelet, whose writings contain numerous subtle remarks on the effect of various causes on social phenomena, formulations of profound statistical problems[72], advancement of social statistical theories and detection of important facts of social life[73].

Quetelet regarded the lack of unified statistical data as the main obstacle for the development of statistics [119, pp. 362-364], and accordingly he attached paramount importance to the collection of statistical data on a worldwide scale[74]. However, the history of statistics evidently proved the inadequacy of that statement: for statistics, most important it also was to become a mathematically based science.

Gauss, for all his interest in statistics, was not concerned with its mathematical structure[75]. Even in his correspondence he did not touch this subject, which is all the more disappointing in the light of Quetelet's evidence [121, p. 655]:

Il paraît que, vers la même époque [1847] *Gauss (?) et Schumacher s'occupaient avec un vif intérêt de la théorie des probabilités appliquée aux lois sociales, car, dans une lettre que m'écrivait Schumacher, en juillet 1846, il me parlait de l'intention qu'il avait de donner une traduction de mes Lettres* [119].

Being concerned with meteorological statistics (§ 7.6), Gauss apparently did not comment on the first volume of Quetelet's work on the climate of Belgium [120][76]. This work is crammed with statistical data and contains certain conclusions largely derived by stochastic rules of thumb[77].

Gauss would not have failed to single out climatology as an object for application of the statistical method, as a discipline, I may add, whose scientific requirements demanded the development of ideas and methods of mathematical statistics even in the mid-19th century, i. e. before the same need was felt in biological and physical quarters.

8. General conclusions

Gauss was the first to use and substantiate the MLSq, while Legendre should be credited for the first introduction of this method in a published work, in which he recommended it as a fit procedure for mathematical treatment of observations.

Complaining about lack of time and referring to the need to prepare carefully the manuscripts intended for the press, Gauss usually delayed long in publishing apparently completed researches. He seemed to have been satisfied to establish his dates by private records and correspondence.

Considering the solution of systems of equations

$$a_i x + b_i y + c_i z + \ldots + l_i = v_i \quad i = 1, 2, \ldots, \mu$$

with the number of unknowns $k < \mu$ in his *Theoria motus*, and assuming a unimodal density law for the errors of observation, Gauss maintained that the case of k zero residuals v_i is unacceptable. He also noted that this case is brought about by an additional restriction

$$|v_1| + |v_2| + \ldots + |v_\mu| = \min$$

which is the main condition for the adjustment of observations according to the method of Boscovich – Laplace. Thus, Gauss formulated an important theorem in linear programming, and exactly his opinion, mentioned above, explains a phrase from Gauss's diary (1798): *Calculus probabilitatis contra La Place defensus*.

However, in accord with notions of mathematical statistics, that condition is preferable to the principle of least squares for certain (even unimodal) distributions of errors.

Furthermore in the *Theoria motus* Gauss proved that among unimodal, symmetric and differentiable distributions there is a unique distribution (the normal) for which the maximum likelihood estimator of the location parameter coincides with the arithmetic mean.

Assuming the normal distribution and once again using the principle of maximum likelihood, Gauss arrived at the MLSq.

In 1816 Gauss [6] determined the most probable value of the measure of precision h for the density law

$$P(\Delta) = \frac{h}{\sqrt{\pi}} \exp(-h^2 \Delta^2)$$

and showed how to estimate the probable error of observations by the absolute moments of the errors and sums of natural powers of absolute errors. For the normal distribution Gauss pointed out a method of estimating the probable error from the median of the absolute errors. He indicated some of these results without proof.

In his *Theoria combinationis* Gauss chose the variance as a measure of precision and introduced the adjustment of observations according to the principle of least variance which led to the principle of least squares.

The new substantiation of the MLSq squares did not depend on normality and Gauss spoke out decidedly in favour of his new approach. However, this substantiation only recently came to be generally recognized.

The *Theoria combinationis* also contains an inequality of the Bienaymé – Chebyshev type for unimodal distributions, an inequality for the fourth moment of errors, a study of the distribution of a function of random variables, a bilateral estimate of the variance of the sample variance (subsequently strengthened by Helmert) and a number of practically important formulas.

Gauss's achievements befittingly concluded the construction of the classical theory of errors, and this very reason in the 20th century impeded application of mathematical statistics to the treatment of observations.

The "geodetic" period of Gauss's life lasted for about ten years. He was responsible for the accomplishment of, and was directly involved in, the triangulation in Hannover. He pointed out that triangulations of separate countries when connected with one another acquire international importance.

Gauss introduced measures to eliminate errors from angle measurements. He was the master or experimental science who revived the aspiration of astronomers of the first half of the 18th century (Bradley) toward the highest possible accuracy. In connection with his geodetic work Gauss had to solve large systems of linear algebraic equations and had to perform extensive computations without even a simplest machine.

Gauss formulated and partly solved the first problem in the metric theory of numbers; he should also be credited with a first study related to the theory of random arrangements. Gauss collected statistical data on demography, meteorology and economics, and proposed formulas for describing infant mortality and mortality of tontine members.

Acknowledgements. I am grateful to Professors R. L. Plackett and S. M. Stigler, who pointed out the existence of some commentaries on § 11 of the *Theoria combinationis*. D. H. L. Harter, the late Dr. K. O. May, Dr. E. Seneta and Professor Stigler sent me reprints of their work and photostat copies of necessary literature. Quite a few persons are now working on the history of the theory of errors. I personally feel that I have picked up right where Dr. C. Eisenhart regrettably stopped. Or has he? Dr. M. V. Chirikov noticed that Gauss used various kinds of probability (cf. for example, §§ 3.2 and 7.3).

Notes

1. In these instances notation such as G-B, 24.1.1812 stands for letter from Gauss to Bessel dated Jan. 24, 1812. Other abbreviations are: G-G, Gauss to Gerling; G-O, Gauss to Olbers; G-S, Gauss to Schumacher. Lastly, references W-9 (or W-8 etc.)

mean Gauss's *Werke*, Bd. 9 (or Bd. 8 etc.).
2. Euler did not support Bernoulli's suppositions, so that if Gauss had read their writings, it would have been necessary for him to separate the ideas and consider the whole problem from the very beginning.
3. In 1805 an anonymous author, possibly von Zach [140], the editor of the *Monatl. Correspondenz*, described Bernoulli's memoir and Euler's commentary, 27 years after their publication. The author took for granted Euler's understanding of Bernoulli's reasoning about the weights of observations [128, § 1.2], a fact which corroborates my opinion (Ibidem) on the singularity of this reasoning.
4. This opinion should be qualified by Gauss's own confession (G-S, 12.2.1841; [30, vol. 4, p. 9]) to the effect that he was never able to remember what he read. (This, of course, is an overstatement.) As to Legendre, Gauss did not consider him as his precursor (see § 2.5). Agreeing with Gauss from a purely scientific point of view, I repeat that in a sense Bernoulli and Euler were his forerunners.
5. Hogan [78, p. 170] found out that Adrain's article was published in 1809.
6. Here and throughout I use the term normal law (distribution) which had not yet appeared during Gauss's lifetime.
7. Plackett [115] published passages from this correspondence. I do not repeat his comments.
8. Obviously Gauss meant the principle of least variance.
9. Writing to Laplace on Jan. 30, 1812 (W-10/1, p. 373) Gauss restricted himself to a neutral formula:

Je trouve dans mes papiers, que le mois de Juin 1798 est l'époque où je l'ai [the MLSq is meant] *rapprochée aux principes du calcul des probabilités*.
10. I [133, § 8.3] corrected an error made in the latter source in the proof of one of Laplace's statements. The same subsection of this article [133] describes a similar method of adjusting indirect observations from t. 2 of the *Mécanique céleste* (published, however, after Gauss formulated his objections against Laplace) and the *Théor. anal. prob.* In the latter source Laplace used the term *méthode de situation*.
11. Schumacher's opinion, voiced in his letter to Gauss dated March 3, 1832 [30, Bd. 6, p. 299], is not convincing. Discussing rules for establishing priority, he wrote:

Es giebt von jeder Regel glänzende Ausnahmen, und eine solche ist hier. ... Sie der Methode der kleinsten Quadrate zuerst öffentlich erwähnt haben (where?) *... nur eine öffentliche Erwähnung, nicht grade eine Entwickelung verlangt wird*.

Gauss's letter to von Zach dated 24.8.1799 (W-8, p. 136) is likely meant. But Gauss referred to an unspecified method.
12. However, one of Gauss's pronouncements quoted in § 2.4 was published soon after 1820. Also, as testified by von Waltershausen [139, p. 43],

Er [Gauss] *hat sich ein Mal, den Streit erwähnend, gegen uns mit den Worten ausgesprochen „Die Methode der kleinsten Quadrate ist nicht die größte meiner Entdeckung." Ein anderes Mal hat er gegen einige Zuhörer nur die Worte betont: „Man hätte mir wohl glauben können"*
13. In the opinion of Klein [82, pp. 11-12]

Oft hat er [Gauss] *seine schönsten Errungenschaften nicht veröffentlicht. Was mag dies seltsame Stillstehen dicht vor dem Ziele veranlasst haben? Vielleicht ist der Grund in einer gewissen Hypochondrie zu suchen, die Gauss offenbar zuweilen mitten im erfolgreichsten Schaffen überfiel*.

Klein also suggests that the reasons for Gauss's hypochondria were the

Drückende Elend des Alltags and the *Rückschlag gegen die übergroße Intensität* of his scientific production.
14. See also Biermann [48; 50, pp. 8-10] who largely excuses Gauss.
See also § 2.6.3.
15. Bearing in mind this work, I [132, p. 164] have mistakenly referred to another of Laplace's memoirs.
15a. At the same time Gauss [106, p. 304] put a high value on both Jacobi and Dirichlet and bitterly lamented over Legendre's fate (G-S, 17.10.1824; [30, Bd. 1, p. 413]):

Mit Unwillen und Betrübnis habe ich ... gelesen, dass man dem alten Legendre, der eine Zierde seines Landes und seines Zeitalters ist, die Pension gestrichen hat.
16. For example, Gauss indicates that the case $c > 0$ (formula (3.2.1)) is impossible because function $\varphi(\Delta)$ will not then attain its maximum value. Thus this restriction

follows not from the properties of random errors but rather from most general properties of density functions.
17. Gauss apparently made mistakes in the wording of two obvious corollaries to his last statement; besides, his remark on the use of condition (3.3.2) by Boscovich and Laplace is not altogether correct [130, § 1.3.3].
18. Actually K_n should be the corresponding absolute moments, see below.
19. For $n = 2$ this interval naturally coincides with the one derived in the first part of the memoir.
20. This restriction is not really necessary.
21. Omitted.
22. Compare with a modern definition [36, p. 78]: *Random errors of observations are such errors, indeterminate both in magnitude and nature, as caused by reasons depending on the measuring apparatus … and also on external conditions. In the theory of errors random errors are considered as random quantities*.
23. Gauss noted (G-O, 14.4.1819; W-8, pp. 150-151) that

Gewicht ist übrigens immer dem Quadrate der Genauigkeit direct, oder dem Quadrate des sogenannten wahrscheinlichen Fehlers umgekehrt proportional; welches aber kein Lehrsatz, sondern bloß die Definition des Worts Gewicht ist.

He repeated the first half of this definition in 1843 (G-S, 1.4.1843; W-12, p. 292); but *Genauigkeitsgrad* is first found in the *Theoria motus* (§ 173): if errors inversely proportional to e_1, e_2, ... occur in observations *gleich leicht*, then, Gauss maintained, the *Genauigkeitsgrad* of these observations should be proportional to e_1, e_2, ... while

$$\frac{e_1^2 x_1 + e_2^2 x_2 + ...}{e_1^2 + e_2^2 + ...}$$

will be the *mittlere wahrscheinlichste Werth* (*valor medius maxime probabilis*) of the constant sought. Explaining this fact, Gauss referred to the *unter anzugebenden Principien*; apparently, he thought of the generalized principle of least squares [1, § 179], see § 3.2.
24. See letter to Bessel just below.
25. Apparently Gauss referred to the invariance of the combination of observations with regard to the law of distribution of their errors, see § 5.6.
26. Evidently following the spirit of his time, Gauss did not regard a combination of arcs (segments) of various curves (straight lines) as a continuous function.
27. Without noting the change in notation as compared with the *Theoria motus*, Gauss supposed in one of his examples that

$$\varphi(x) = \frac{1}{h\sqrt{\pi}} \exp(-x^2/h^2).$$

28. Gauss restricted himself to the case of a rational function.
29. I quote the beginning of his article:

This year I decided to use the first part of Gauss's excellent memoir [*Theoria combinationis*] *for instruction in the theory of least squares.*

30. For the notation [*xx*] see the end of § 6.4.
31. Gauss poses this problem quite formally; he does not even mention any "observations".
32. Mal'tsev [103] proved the attainability of the upper estimate. In his *Supplementum* Gauss [4, § 17] carried out a similar investigation for the case of conditioned observations. His final result is for some reason restricted to normally distributed errors, and he published only one intermediate formula. Allowing for the restriction, this formula coincides with the appropriate formula from § 39 of the *Theoria combinationis*.
33. This of course is not exactly so.
34. Such expressions, which separate their author from mathematicians, were characteristic of Laplace [133, p. 12n; 132, §§ 2.4 and 2.6], who evidently did not regard himself as a pure-blooded géomètre.

35. Gauss formulated another important remark about Laplace's [133, §§ 5.2 and 6.3] substantiation of the MLSq (G - O, 22.2.1819; W-8, pp. 142-143):

Die Generalisierung seines Schlusses von zwei unbekannten Großen auf jede Anzahl noch nicht die nöthige Evidenz zu haben scheint.

The same possibly independent reservation is due to Czuber [58, p. 252]. Much more interesting though is the lack of any such remark in the *Theoria combinationis*, so did not Gauss himself manage to achieve that generalization?

36. Laplace's mode of substantiation (see above) should also be mentioned.

37. The German term *wahrscheinlichste*, or *wahrscheinlichste Werthe* is used in that part of the *Selbstanzeige* [5] in which Gauss recalls his *Theoria motus*. Besides (§ 5.2, note 23) and § 4.2), in these earlier writings Gauss even equated mean and most probable values.

38. Or, as in the *Selbstanzeige* [5, p. 101], the German term *sicherste Werthe.*

39. Moreover, some authors (Encke [65], p. 74], Merriman [107, pp. 165 and 174]) did not accept any of these three substantiations. For example, Merriman (p. 174) somehow managed to conclude in regard to the principle of minimal variance that

It is but little more than a begging of the question to assume that the mean of the squares of the errors is a measure of precision.

Tsinger [133], p. 53] voiced almost the same argument.

40. Actually adhering to the same opinion in his own tract [75], Helmert here did no better.

41. The last statement seems too strong.

42. I also hold that the point of view expressed by Tsinger (1862) [133, p. 53], who recognized only the Laplacean substantiation, is refuted by the practice of mathematical treatment of observations during the last one and a half centuries. Even much worse: his statement shows that he did not read Gauss and attributed nonsense to him.

43. I do not agree with Linnik et al. [99, p. 637] who maintained that Markov

Essentially introduces ... notions, equivalent to the notions of unbiased and efficient estimators.

The same (mistaken) assertion can just as well stated about Gauss. Cf. Neyman [110, p. 593]:

The importance of the work of Markov concerning the best linear estimates consists, I think, chiefly in a clear statement of the problem [not in the formal introduction of that concept].

I especially notice the stubborn existence of the mysterious Gauss – Markov theorem (due to Gauss alone). Already Neyman himself (1938) admitted his mistake (1934) in overestimating Markov’s role.

44. Obviously, Gauss here means preliminary calculations.

45. Not necessary for the calculation of subsequent sides.

46. Gauss (G-B, 27.1.1819; Ibidem, p. 515) also described an error committed by Laplace [94, Suppl. 2, pp. 559-564] in his study of the method of repetition. It is not my intention to comment on Gauss's remark, a rather short one at that.

47. It seems that Gauss did not consider this case unusual; he offered no comment whatsoever.

48. Also in the USSR.

49. Strictly speaking, Gauss did not object to rejection based on yet unknown stochastic criteria. On a number of occasions he reproached geodesists with failing to register rejected observations; see his *Selbstanzeige* [5, p. 106], his letters (G-B, 15.11.1822; W-9, p. 353; G-B, 12. 3.1826; Ibidem, p. 361; G-S, 20. 12.1823; [30, Bd. 1, p. 349]: G-O, 14.5 1826: W-9, p. 321) and his review published in 1830 [9, p. 372]. Many authors had proposed criteria for rejection but finally Barnett & Lewis (1978, p. 360) concluded that we still do not know what is an outlier or how to deal with it.

50. To my earlier description of the history of the problem in question [131, § 3.5] Bessel's point of view (B – G, Apr. 1, 1819; [27, p. 295]) can be added.

51. Gauss's use of the method of relaxation is a good example of his ingenuity in calculations.

52. For calculations of a smaller extent see the references mentioned at the beginning of this subsection.

53. In 1930 Maennchen [102] described Gauss's merits in the field of calculation. He did not mention any geodetic computations, evidently because at that time the solution of systems of linear algebraic equations was not yet included in numerical analysis. Maennchen (p. 3) noted that Gauss was often led to his discoveries

Durch peinlich genaues Rechnen. ... *Wir finden* [in Gauss's writings] *ganzen Tafeln, deren Herstellung allein die Lebensarbeit manches Rechners vom gewöhnlichen Schlage ausfüllen würde*.

54. See my preceding article [131, p. 110].

55. Omitted.

56. See however § 7.6.

57. Bessel's description above is apparently more accurate.

58. I add one short phrase (Krylov [87, p. 287]):

Being an astronomer and geodesist, Gauss also introduced unprecedented accuracy into magnetic observations.

Krylov failed to honour Weber alongside Gauss.

59. But then, the explanation of some problems in the *Theoria combinationis* is too abstract (see my § 5.6). The same feature characterizes all the mathematical work of Gauss [86, p. 42]:

Die Art der Darstellung ist in den Disquisitiones [*arithmeticae*] *wie überhaupt in den Gaussischen Arbeiten, die Euklidische. Er stellt die Sätze auf und beweist sie, wobei er geradezu mit Fleiß jede Spur der Gedankengänge verwischt, die ihn zu seinen Resultaten geführt haben. In dieser dogmatischen Form ist gewiss auch der Grund dafür zu suchen, dass sein Werk so lange unverstanden blieb*.

Gauss [3] is a prime example. In spite of the unsurmountable difficulty of understanding its exposition, it is possible (although Gauss had not even hinted at that possibility) to derive the principle of least squares by bypassing those difficulties, see Sheynin (2017, pp. 148 – 149). The protracted general acceptance of the 1809 justification of least squares can be ended.

60. Gauss also maintained that studies in astronomy were difficult exactly because Bessel had *zuvor gekommen … und den meisten Desideraten bereits … abgeholfen haben*.

61. To recall, the first section, a lengthy one at that, of the course of lectures on probability delivered by Chebyshev in 1879-1880 [53] was devoted to definite integrals.

62. He was not as a rule interested in the *Fachliteratur* (§ 2.6.2); however, it is likely that at least the main initial ideas and results arrived at in probability theory were not foreign to him. Thus (G – B 28.2.1839, W-8, pp. 146-147):

Ihren Aufsatz [45] … *über die Annäherung des Gesetzes für die Wahrscheinlichkeit aus zusammengesetzten Quellen entspringender Beobachtungsfehler an die Formel* $\exp(-x^2/h^2)$ *habe ich mit großem Interesse gelesen; doch bezog sich, wenn ich aufrichtig sprechen soll, dieses Interesse weniger auf die Sache selbst, als auf Ihre Darstellung. Denn jene ist mir seit vielen Jahren familiär, während ich selbst niemals dazu gekommen bin, die Entwickelung vollständig auszuführen.*

63. In 1809 Gauss used exactly this principle (§ 3.2) though for mathematical rather than general scientific purposes.

64. Working on his future book [70], Fries had applied to Gauss for his opinion on general principles of probability, and Weber replied to him instead of Gauss. At the end of his letter Weber explained the situation (see also § 2.6.1):

Gauss hätte selbst wohl einige Zeilen beigelegt, wenn er etwas zu sagen gehabt, dessen Ausdruck, um nichts an Präcision zu verlieren, schwieriger gewesen wäre.

Is it far-fetched to compare, on this occasion, Weber and Gauss with Bentley and Newton, respectively?

65. The only example which follows concerns Laplace, who made a rather elementary error in his study [92] of the mutual positions of the planes of cometary orbits.

Gauss first pointed out this error in 1813 (G-O, 25.7.1813; [28, No. 1, p. 527]). I note also that in 1812 Olbers (O – G July 18, 1812; [28, No. 1, pp. 506-509]) concluded that the large inclination of Pallas is hardly accidental and that Laplace (who repeatedly studied the inclinations of planets and comets [131, §§ 3.1 and 3.4]) agreed with him.

66. This theory originated in our time; however, J. Bernoulli [40, pt. 3, pp. 13-18] studied a problem which would now be attributed #to the theory.
67. Gauss related the latter problem to meteorology. Nowadays it would likely be attributed to meteorological statistics. The editor of the source [31] notes that first statistical data on lightning appeared in 1937 [47].
68. Gauss would have made an excellent minister of finance, von Waltershausen (p. 90) opines, and adds that luckily no such transformation took place. The author does not seem to account for quite a number of qualities Gauss happily never possessed.
69. Moser [108, p. 281] used $n^{1/4}$.
70. Quetelet also supplied separate data for towns and rural districts and for boys and girls. As Gauss provided no additional information, I suppose that he used generalized data for Belgium as a whole, which is entered in my table. Gauss referred to a German edition of Quetelet.
71. For the empirical laws of mortality introduced by Gompertz, Makeham and Lazarus see Czuber [59, pp. 312-314].
72. One of these problems, likely characteristic of Quetelet's time, was to determine how railroad construction affects population, industry and the price of land [119, p. 351]. See Sheynin (2017, § 10.5) for a general description of his work.
73. For example, variations in mortality and birth-rate are connected with fluctuations in the price of bread [117, p. 210].
74. A most important result of the activities of statisticians of those times was the adoption, in 1875, of the metric system by seventeen countries. It is opportune to recall (§ 6.1.3) Gauss's opinion about the possible international importance of separate triangulations and to mention his (and Weber's) work on the introduction of the absolute system of units. Gauss's thoughts concerning the metric system (G-O, 8.12.1817; [28, No. 1, p. 674]) are also quite relevant:

Sehr interessant ist mir die Aussicht einer vielleicht allgemeinen Einführung des französischen Maassystems. Höchst bequem finde ich dieses System, und ich bediene mich desselben gern überall, und glaube, dass alles oder das meiste, was man gegen allgemeine Einführung gesagt hat, auf Vorurtheilen beruht. Nur bei den allerfeinsten Messungen, glaube ich, entstehen große Inkonvenienzen aus der Einführung eines natürlichen Maassystems, und man muss daneben immer irgend ein Maassindividuum haben. ... Jede Gradmessung direkt oder indirekt den Zweck hat, das Meter zu suchen; gibt man ihn nach Metern an, so bedeutet da Meter nicht 1 :10,000,000 Erdquadrant, sondern die Länge desjenigen Stückes Eisen. ... Es ist also ein nie aufhörendes Schwanken.

75. The establishment of empirical laws of mortality (§ 7.7) has nothing, or almost nothing to do with rendering statistics mathematical.
76. The second volume appeared after Gauss's death.
77. Quetelet discovered many facts just by compiling relevant data (and exactly this reason likely explains why he did not care for mathematisation of statistics). A convincing example concerning criminal statistics is contained in his *Lettres* [119, pp. 358-359].

References

L, M = Leningrad, Moscow
W-i = *Werke*, Bd. i.

C. F. Gauss

1. *Theoria motus* ... (1809). German transl.: *Aus der Theorie der Bewegung der Himmelkörper* etc. Excerpt In *Abh. zur Methode der kleinsten Quadrate*. Hrsg. A. Börsch und P. Simon. Berlin, 1887, 92-117. Vaduz, 1998. *Theory of motion*, 1865. Boston, 2009.
2. Disquisitio de elementis ellipticis Palladis ... (1811). German transl. Aus der Untersuchung über die elliptischen Elemente der Pallas etc. Ibidem, 118-128.
3. Theoria combinationis ... (1823). German transl.: Theorie der den kleinsten Fehlern unterworfenen Combination der Beobachtungen. Ibidem, 1-53. English translation: Stewart, G. W. Philadelphia, 1995.
4. Supplementum theoriae combinationis ... (1828). German transl. Theorie der ... Combination etc., Ergänzung. Ibidem, 54-91.
5. Theoria combinationis ..., and Supplementum ..., Selbstanzeigen. (1821-1826).

W 4. Göttingen, 1880, 95-108.
6. Bestimmung der Genauigkeit der Beobachtungen. (1816). Ibidem, 109-117.
7. Anwendung der Wahrscheinlichkeitsrechnung auf die Bestimmung der Bilanz für Witwenkassen. (1845-1851; Nachlass). Ibidem, 119-169.
8. Tafeln zur Bestimmung des Zeitwerthes von einfachen Leibrenten und von Verbindungsrenten. (Nachlass). Ibidem, 170-183.
9. Review of *Opérations géodésiques et astronomiques ... exécutées en Piémont et en Savoie ...en 1821, 1822, 1823*, tt. 1-2. Milan, 1825-1827. (1830). Ibidem, 370-381.
10. Über ein neues allgemeines Grundgesetz der Mechanik. (1829). *W*-5. Göttingen, 1877, 25-28.
11. Theoria motus ... Selbstanzeige. (1809). W-6. Göttingen, 1874, 59-60.
12. Zweiter Comet von 1805. (1806). Ibidem, 270-277.
13. Chronometrische Längenbestimmungen. (1826). Ibidem, 455-459.
14. Review of [95] (1815). Ibidem, 581-586.
15. Schönes Theorem der Wahrscheinlichkeitsrechnung. (Nachlass). W-8. Göttingen-Leipzig, 1900, p. 88.
16. [Aufgabe aus der Wahrscheinlichkeitsrechnung]. (Nachlass). Ibidem, p. 133.
17. [Aufgabe]. (Nachlass). Ibidem, 134-135.
18. [Eine Ausgleichsformel für Mortalitätstafeln]. (Nachlass). Ibidem, 155-156.
19. Bestimmung des Breitenunterschiedes zwischen den Sternwarten von Göttingen und Altona etc. (1828). W-9. Göttingen-Leipzig, 1903, 5-62. **S, G,** 72.
20. Anwendung der Wahrscheinlichkeitsrechnung auf eine Aufgabe der practischen Geometrie. (1823). Ibidem, 231-237.
21. [Plan und Anfang zum Werke über die trigonometrischen Messungen in Hannover]. (Nachlass). Ibidem, 401-405.
22. Auszüge aus Berichten über die Triangulierung an das hannoversche Cabinets-Ministerium. (1822-1844). Ibidem, 406-427.
23. [Über die bei der Landestriangulierung erforderlichen Instrumente]. (Nachlass). Ibidem, 487-489.
24. [Problema e calculo probabilitatis circa fractiones continuas olim frustra tentatum solvimus]. (1800; Nachlass). W-10/1. Göttingen-Leipzig, 1917, 552-554.
25. *Izbr. geod. soch.* (Sel. geod. Works), vols. 1-2. M., 1957-1958.
26. *Math. Tagebuch*. (1796-1814). Leipzig, 1976. (*Ostwald's Klassiker* No. 256).

Gauss's Correspondence

27. *Briefwechsel zwischen Gauss und Bessel*. Leipzig, 1880.
28. [*Briefwechsel zwischen Gauss und Olbers*]. Schilling C., *W. Olbers. Sein Leben und sein Werk*, Bd. 2, Abt. 1-2. Berlin, 1900 und 1909.
29. *Briefwechsel zwischen Gauss und Gerling.* Berlin, 1927.
30. *Briefwechsel zwischen Gauss und Schumacher*, Bd. 1-6. Altona, 1860-1865.
31. *Briefwechsel zwischen von Humboldt und Gauss*. Berlin, 1977.
Reprint of correspondence: *Werke, Ergänzungsreihe*, Bde. 1 – 5. Hildesheim. G – B, Bd. 1, 1975; G – G, Bd. 3, 1977; G – O, Bd. 4, 1976; G – S, Bd. 5, 1975

Other Authors

32. Adrain, R., Research concerning the probabilities of the errors which happen in making observations, 1808 [actually, 1809]. All of his papers, see below, reprinted in Stigler (1980, vol. 1).
33. ---, Investigation of the figure of the earth and of the gravity in different latitudes. 1818.
34. ---, Research concerning the mean diameter of the earth, 1818.
35. Anonymous, E. A. W. von Zimmermann. *Allg. deutsche Biogr.*, Bd. 45. Leipzig, 1899, 256-258.
36. Anonymous, Errors of observations. *Fizich. enz. slovar* (Phys. Enc. Dict.), vol. 4. M, 1965, 77-78.
37. Bagratuni, G. V., Gauss's geodetic work [25, vol. 1, 3-18].
38. Bejar, J., Regresión en mediana y la programación lineal. *Trabajos de estadística*, t. 7, 1956, 141-158.
39. Bernoulli, Daniel, The most probable choice between several discrepant observations etc. (1778, in Latin). *Biometrika*, vol. 48, No. 1-2, 1961, 3-13. Repr.: Pearson E. S., Kendall (1970, 157-167).

40. Bernoulli, Jacob, *Wahrscheinlichkeitsrechnung* (*Ars conjectandi*). (1713, in Latin). Leipzig, 1899; Frankfurt/Main, 1999.
41. Bertrand, J., *Calcul des probabilités*. Paris, 1888, 1907. New York, 1970, 1972.
42. Bessel, F. W., Untersuchung der Große und des Einflusses des Vorrückens der Nachtgleichen, 1815. *Abh.*, Bd. 1. Leipzig, 1875, 262-285.
43. ---, Bemerkung über Veränderlichkeit der Passageninstrumente, 1815. *Abh.*, Bd. 2. Leipzig, 1876, p. 19.
44. ---, Untersuchungen über die Bahn des Olbersschen Kometen. *Abh. Preuss. Akad.* [Berlin], math. Kl., 1812-1813 (1816), 119-160.
45. ---, Untersuchungen über die Wahrscheinlichkeit der Beobachtungsfehler, 1838. *Abh.*, Bd. 2, 372-391.
45a. ---, *Abhandlungen*, Bde 1-3. Leipzig, 1876.
46. Biermann, K.-R., Aus der Entstehung der Fachsprache der Wahrscheinlichkeitsrechnung. *Forschungen und Fortschritte*, Bd. 39, No. 5, 1965, 142-144.
47. ---, Die Probleme der Schwereänderung etc. in einem Brief von C. F. Gauss an A. von Humboldt. Ibidem, No. 12, 357-361.
48. ---, Über die Beziehungen zwischen Gauss und Bessel. *Mitt. Gauß-Ges. Göttingen*, Bd. 3, 1966, 7-20. **S, G,** 72.
49. ---, Über die statistischen Zahlenregister von Gauss. *Proc. 13th Intern. Congr. Hist. Sci*., sect. 5. (Moscow, 1971). M., 1974, 150-157.
50. ---, Historische Einführung [26, 7-20].
51. Bomford, G., *Geodesy*. Oxford, 1971. Other editions: 1952, 1962, 1980.
52. Cauchy, A. L., Sur les coefficientes limitateurs ou restricteurs. (1853). *Oeuvr. Compl.*, sér. 1, t. 12. Paris, 1900, 79-94.
53. Chebyshev, P. L., *Teoriya veroiytnostei* (Theory of Prob. Lectures delivered in 1879-1880 as written down by A. M. Liapunov). Ed. by A. N. Krylov. M. – L., 1936. A great lot of misprints.
54. Clarke, A. R., *Geodesy*. Oxford, 1880.
55. Coolidge, J. L., R. Adrain and the beginnings of American mathematics. *Amer. Math. Monthly*, vol. 33, No. 2, 1926, 61-76.
56. Cournot, A. A., *Exposition de la théorie des chances et des probabilités*. Paris, 1843. Paris, 1984. Editor B. Bru. **S, G,** 54.
57. Cramér, H., *Mathematical methods of statistics*. Princeton, 1946.
58. Czuber, E., *Theorie der Beobachtungsfehler*. Leipzig, 1891.
59. ---, *Mathematische Bevölkerungstheorie*. Leipzig-Berlin, 1923.
60. David, F. N., Neyman, J., Extension of the Markoff theorem on least squares. *Stat. Res. Mem.*, vol. 2. London, 1938, 105-117.
61. Dedekind, R., Gauss in seiner Vorlesung über die Methode der kleinsten Quadrate, 1901. *Ges. math. Werke*, Bd. 2. Braunschweig, 1931, 293-306.
61a. Deparcieux, A., *Essai sur les probabilités de la durée de la vie humaine*. Paris, 1746.
62. Dirichlet, P. G. L., Über einen von Dirichlet herrührenden Beweis aus der Wahrscheinlichkeitsrechnung, 1834. *Werke*, Bd. 2. Berlin, 1897, 368-372.
63. Dunin-Barkovski, I. V., Smirnov, N. V., *Teoriya veroyatnostei i matematicheskaiya statistika v tekhnike* (Theory of Prob. and math. Stat.: engineering Applications). M., 1955.
64. Dunnington, G. W., *Gauss: Titan of Science*. New York, 1955.
65. Encke, J. F., Über die Begründung der Methode der kleinsten Quadrate. *Abh. Kgl. Akad. Wiss. zu Berlin*, math. Kl, 1831 (1832), 73-78.
66. ---, Uber die Methode der kleinsten Quadrate, 1834. *Astron. Abh.*, Bd. 1. Berlin, 1866. Orig. paging (249-312) is retained, but no page numbers of the book are given.
67. ---, Über die Anwendung der Wahrscheinlichkeitsrechnung auf Beobachtungen. *Berliner Astron. Jahrbuch* für 1853 (1850), 310-351.
68. Euler, L, Observations on the foregoing dissertation of Bernoulli. (1778, in Latin; E (Eneström) 488). *Biometrika*, vol. 48, No. 1-2, 1961, 13-18. Repr.: Kendall, Plackett (1977, 167-172).
69. Forsythe, G. E., Gauss to Gerling on relaxation. *Math. Tables and Other Aids to Computation*, vol. 5, No. 36, 1951, 255-258.

70. Fries, J. F, *Versuch einer Kritik der Principien der Wahrscheinlichkeits-rechnung*. Braunschweig, 1842. *Sämtl. Schriften*, Bd. 14. Halen, 1974.
71. Galle, A., Über die geodätischen Arbeiten von Gauss. W-11/2, Abh. 1. Berlin, 1924. Separate paging.
71a. Gerardy, T., Die Anfänge von Gauss' geodätischer Tätigkeit. *Z. Vermessungs-wesen*, Bd. 102, No. 1, 1977, 1-20.
72. Gnedenko, B. W., Über die Arbeiten von Gauss zur Wahrscheinlichkeitsrech-nung. In: Gauss. *Gedenkband.* Leipzig, 1957, 194-204. Orig. publ. in Russian (1956).
73. Hagen, G., *Grundzüge der Wahrscheinlichkeitsrechnung.* Berlin, 1867 (other editions 1837, 1882).
74. Harter, H. L., The method of least squares and some alternatives, pt. 1. *Intern. stat. rev*., vol. 42, No. 2, 1974, 147-174.
75. Helmert, F. R., *Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate*. Leipzig, 1872, 1907, 1924.
76. ---, Zur Ableitung der Formel von Gauss für den mittleren Beobachtungs-fehler und ihrer Genauigkeit. *Sitz.-Ber. Kgl. Preuss. Akad. Wiss*. Berlin, 1904, Halbbd. 1, 950-964.
77. ---, Shorter version of same. *Z. Vermessungswesen*, Bd. 33, 1904, 577-587.
78. Hogan, E. R., R. Adrain: American mathematician. *Hist. Math.*, vol. 4, 1977, 157- 172.
79. Idelson, N. I., *Sposob naimenshikh kvadratov* etc. (Method of least squares etc.). M., 1947. **S, G,** 58 (Chapter 1).
80. Jordan, W., Über die Bestimmung der Genauigkeit mehrfach wiederholter Beobachtungen einer Unbekannten. *Astron. Nachr*., Bd. 74, No. 1766-1767, 1869, 209-226.
81. Kendall, M. G., Stuart A., *Advanced Theory of Statistics*, vol. 1. London, 1969.
82. Klein, F., *Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert*, Tl. 1. Berlin, 1926.
83. Kolmogorov, A. N., Petrov, A. A., Smirnov, Yu. V., Gauss's formula from the theory of least squares. *Izv. AN SSSR* (Proc. Acad. Sci. USSR), ser. math., vol. 11, No. 6, 1947, 561- 566. In Russian.
84. Krafft, M., Zwei Sätze aus Gauss' Theoria combinationis etc. *Deutsch. Math*., Bd. 2, No. 5, 1937, 624-630.
85. Krassovski, F. N., *Izbr. Soch.* (Sel. works), vol. 3. M., 1955.
86. Kronecker, L., *Vorlesungen über Zahlentheorie*, Bd. 1. Leipzig, 1901.
87. Krylov, A. N., C. F. Gauss. (1934). *Sobr. trudov* (Works), vol. 1/2. M.-L., 1951, 279- 297.
88. Kuzmin, R., Sur un problème de Gauss. *Atti Congr. intern. matem. Bologna* 1928, t. 6. Bologna, 1932, 83-89.
89. ---, Russian version of same. *Dokl: AN SSSR* (Reports Acad. sci. USSR), ser. A, 1928, No. 18-19, 375-380.
90. Lagrange, J. L., Sur l'utilité de la méthode de prendre le milieu etc., 1776. *Oeuvr.*, t. 2. Paris, 1868, 173-234.
91. Laplace, P. S., Sur la probabilité des causes par les événements, 1774. *Oeuvr. compl*, t. 8. Paris, 1891, 27-65.
92. ---, Sur l'inclinaison moyenne des orbites des comètes etc, 1773 (1776). Ibidem, 279-321.
93. ---, Sur les degrés mesurés des méridiens etc, 1789 (1792). Ibidem, t. 11. Paris, 1895, 493-516. This memoir comprises §§ 8-14 of Laplace's Sur quelques points du système du monde.
94. ---, *Théorie analytique des probabilités*. 1812, 1814, 1820. Ibidem, t. 7. Paris, 1886.
95. ---, Sur les comètes, 1816 (1813). Ibidem, t. 13. Paris, 1904, 88-97.
96. Legendre, A. M., *Nouvelles méthodes pour la détermination des orbites des comètes*. Paris, 1805.
97. ---, *Nouvelles méthodes*, etc., suppl. 2. Paris, 1820.
98. Lehmann, E. L., *Testing statistical hypotheses.* New York-London, 1959.

99. Linnik, Yu. V., Sapogov, N. A., Timofeev, V. N., Essay on the work of Markov on number theory and theory of probability. In: Markov, A. A., *Izbr. Trudy* (Sel. works). No place, 1951, 614-640.
100. Lipschitz, R., Sur la combinaison des observations. *C. r. Acad. sci. Paris,* t. 111, 1890, 163-166.
101. Loewy, A., Die Gauss'sche Sterbeformel. *Z. für die ges. Versicherungswiss.*, Bd. 6, No. 3, 1906, 517-519.
102. Maennchen, Ph., Gauss als Zahlenrechner, 1918. W-12, Abh. 6. Göttingen, 1930. Separate paging.
103. Mal'tsev, A. I., Remark on the work of Kolmogorov et al. *Izv. AN SSSR* (Proc. Acad. sci. USSR), ser. math, vol. 11, No. 6, 1947, 567-578. In Russian.
104. Maistrov, L. E, *Probability theory. A historical sketch*. New York-London, 1974. Orig. publ. in Russian (1967).
105. Markov, A. A., The law of large numbers and the method of least squares (1899). *Izbr. trudy* (Sel. works). No place, 1951, 231-251. **S, G,** 5.
106. May, K. O., Gauss, C. F. *Dict. scient. biogr.*, vol. 5, 1972, 298-315.
107. Merriman, M., *List of writings relating to the method of least squares* etc., 1877. Reprint: Stigler (1980, vol. 1).
108. Moser, L., *Die Gesetze der Lebensdauer*. Berlin, 1839.
109. Nasimov, P. S., Über eine Gauss'sche Abhandlung. *Varsh. Univ. Izv., Nachr. Warsch. Univ*., No. 4, 1889. Separate paging, 8 pp. In Russian.
110. Neyman, J., On two different aspects of the representative method. *J. Roy. Stat. Soc*., vol. 97, 1934, 558-625.
111. Ogawa, J., Distribution and moments of order statistics. In: *Contributions to order statistics*. Eds., A. E. Sarhan, B. G. Greenberg. New York-London, 1962, 11-19.
112. Olbers, W., Über den veränderlichen Stern im Halse des Schwans. *Z. für Astron. und verw. Wiss*., Bd. 2, 1816, 181-198.
113. Peters, C. A. F., Über die Bestimmung des wahrscheinlichen Fehlers etc. *Astron. Nachr*., Bd. 44, 1856, 29-32.
114. Plackett, R. L, Historical note on the method of least squares. *Biometrika*, vol. 36, No. 3-4, 1949, 458 - 460.
115. ---, Discovery of the method of least squares. Ibidem, vol. 59, No. 2, 1972, 239-251. Repr.: Kendall, Plackett (1977, 279-291).
116. Puissant, L., *Traité de géodésie*. Paris, 1805.
117. Quetelet, A., Population de la Belgique etc. *Corr. math. et phys*., t. 7, 1832, 208-210. (Published anonymously.)
118. ---, *Sur l'homme* etc., t. 1. Paris, 1835.
119. ---, *Lettres ... sur la théorie des probabilités* etc. Bruxelles, 1846.
120. ---, *Sur le climat de Belgique*, t. 1. Bruxelles, 1849.
121. ---, *Sciences mathématiques et physiques au commencement du XIXe siècle*. Bruxelles, 1867.
122. Seal, H. L, The historical development of the use of generating functions in probability theory, 1949. Reprint: Kendall, Plackett (1977, 67-86).
123. ---, The historical development of the Gauss linear model. *Biometrika*, vol. 54, 1967, 1-24. Repr.: Pearson E. S., Kendall (1970, 207-230).
124. Sheynin, O. B, Adjustment of a trilateration figure by frame structure analogue. *Surv. Rev*., vol. 17, No. 127, 1963, 55-56. **S, G,** 109.
125. ---, On the work of R. Adrain in the theory of errors. *Istoriko-matematicheskie issledovania* (Hist.-math. studies), vol. 16, 1965, 325-336. **S, G,** 111.
126. ---, J. H. Lambert' s work on probability. *Arch. Hist. Ex. Sci*., vol. 7, No. 3, 1971, 244-256.
127. ---, Daniel Bernoulli's work on probability, 1972. Repr.: Kendall, Plackett (l977, 105-132).
128.---, On the mathematical treatment of observations by L. Euler. *Arch. Hist. Ex. Sci*., vol. 9, No. 1, 1972, 45-56.
129. ---, Finite random sums. Ibidem, No. 4/5, 1973, 275-305.
130. ---, R. J. Boscovich's work on probability. Ibidem, 306-324. **S, G,** 120.
131. ---, Mathematical treatment of astronomical observations (historical essay). Ibidem, vol. 11, No. 2-3, 1973, 97-126. **S, G,** 115.
132. ---, P. S. Laplace's work on probability. Ibidem, vol. 16, No. 2, 1976,

137-187.
133. ---, P. S. Laplace's theory of errors. Ibidem, vol. 17, No. 1, 1977, 1-61.
134. ---, Early history of the theory of probability. Ibidem, No. 3, 201-259.
S, G, 116.
135. Sofonea, T., Gauss und die Versicherung. *Het Verzerkerings-Archief*, vol. 32, 1955, 57*-69*.
136. Stigler, S. M., An attack on Gauss, published by Legendre in 1820. *Hist. Math.*, vol. 4, 1977, 31-35.
137. Subbotin, M. F. Gauss's astronomical and geodetic work. In: *K. F. Gauss* (C. F. Gauss). M., 1956, 243-310. In Russian.
138. von Mises, R. Über einige Abschätzungen von Erwartungswerten. (1931). *Sel. papers*, vol. 2. Providence, 1964, 135-148.
139. Sartorius von Waltershausen, W., *Gauss zum Gedächtnis*, 1856. Wiesbaden, 1965.
140. von Zach, F. X., Versuch einer auf Erfahrung gegründeten Bestimmung terrestrischer Refraktionen. *Monatl. Corr.*, Bd. 11, 1805, 389 – 415, 485 – 504.
141. Yule, G. U., Introduction of the words *statistics, statistical* into the English language. *J. Roy. Stat. Soc.*, vol. 68, 1905, 391 – 396.

David F. N. (1938), *Lectures and conferences on mathematical statistics and probability*. Washington, 1952.

Barnett V., Lewis T. (1978), *Outliers in statistical data*. Chichester,1984.

Helmert F. R. (1875), Über der Wahrscheinlichkeit der Potenzsummen der Beobachtungsfehler. *Z. Math., Phys.*, Bd. 20, 300 – 303.

--- (1876), Same title. Ibidem, Bd. 21, 192 – 218.

Kendall M. G., Plackett R. L. (1977), *Studies in the history of statistics and probability*, vol. 2. London.

Neyman J. (1938), *Lectures and conferences on mathematical statistics and probability*. Washington, 1952.

Pearson E. S., Kendall M. G. (1970), *Studies in the history of statistics and probability*, [vol. 1]. London.

Poisson S. D. (1837, 2003), *Recherches sur la probabilité des jugements* etc. Paris. **S, G,** 53.

Sheynin O. (2017), *Theory of probability. Historical essay*. Berlin. **S, G,** 10.

Stigler S. M. (1980), *American contributions to mathematical statistics in the 19th century*, vols. 1 – 2. New York. Collection of reprints, only single paging.