

Summary: Gauss introduced the MLSq and Helmert completed its development whereas Bessel made important discoveries in astronomy and geodesy but was often extremely inattentive. Gauss' final condition of least variance led to effective estimators of the unknowns sought, jointly effective in case of the normal distribution of the observational errors. Gauss' memoir of 1823 leads to the principle of least squares much easier than generally thought.

Keywords: principle and method of least squares, sample variance, adjustment of triangulation, personal equation, deviation from normality.

DOI: 10.15611/sps.2014.12.01

1. Introduction

I (Sheynin 1979) have somewhat dwelt on Gauss' investigations in probability proper. He was a tireless collector of statistical data, even of non-essential nature, and successfully managed the widows' fund of the Göttingen University. His correspondence and scientific legacy include a study of the mortality of newly-born and of the members of tontines (of closed societies of mutually insured persons, see § 2.1.3). In the theory of probability, he left the inversion formula for the Fourier transform of the density function.

Gauss also solved the first problem in the metric theory of numbers. He considered the expansion of a number M ($0 < M < 1$) into a continued fraction with unit numerators and investigated the probability $P(n; x)$ that, beginning with its $(n + 1)$ -st convergent, the "tail" of this fraction was less than x . If all the permissible values of M were equally probable or more or less so, then, as he explained his problem in a letter of 1812 to Laplace (*Werke*, Bd. 10/1, pp. 371, 372), $P(0; x) = x$ and

$$\lim P(n; x) = \frac{\ln(1+x)}{\ln 2}, n \rightarrow \infty.$$

Nevertheless, he was not quite satisfied with his solution and asked Laplace to have a look at the problem. He, Gauss, was sure that Laplace

will find a *plus complete* solution, – a pre-limiting expression. A phrase from Gauss' *Mathematisches Tagebuch* written in 1800 (p. 552 of the *Werke*) testifies that Gauss had already then derived the equality above – and had then been satisfied with his work.

Stäckel (Gauss, *Werke*, Bd. 10/1, pp. 554–556) and then Kuzmin (1928) proved this equality and the latter also derived an asymptotic expansion for $P(n; x)$.

Here, I also repeat in a few words Gauss' general opinion (*Werke*, Bd. 12, pp. 201–204) about the applications of the theory of probability as described by W.E. Weber in one of his letters of 1841. If only based on numbers, Gauss reasoned, such applications could be greatly mistaken; the nature of the studied subject ought also to be taken into account. However, probability provides clues when nothing except numbers is known, as for example when dealing with annuities; and in jurisprudence, and it can determine the desired number of witnesses and jurors (but hardly without allowing for “the nature” of law courts).

2. The method of least squares before 1809

It had been indirectly and inaccurately applied from the mid-18th century (§ 6.3.2) and its peculiar version was possibly known even earlier.¹ When some point P was graphically intersected from three or more given stations, a triangle, or a polygon of errors appeared on the surveyor's table sheet and it was apparently natural to select the position of P by eye in such a manner that the sum of the squares of its distances from the sides of the triangle (of the polygon) was minimal. To a certain extent I can justify my opinion by mentioning an experimental smoothing of a broken line by eye (Tutubalin 1973, p. 27): on the whole, the curves thus drawn were as accurate as if having been determined by the MLSq.

Euler (1778) can be considered as Gauss' predecessor in the heuristic sense, but Gauss possibly did not see that memoir (a commentary on Daniel Bernoulli's memoir of the same year). However, in his letters he expressed his surprise at the fact that the principle of least squares was not discovered earlier.

2.1. Huber

Many authors, for example Merian (1830, p. 148), stated that somewhat before 1802 the Swiss mathematician and astronomer Huber had

¹ This term should only be applied to the method as substantiated by Gauss in 1823; until then, strictly speaking, the *principle* of least squares ought to be thought of.

discovered the principle of least squares, but that, living far from scientific centres, he had not reported his finding to anyone. However, Dutka (1990), who referred to a forgotten paper (Spieß 1939), concluded otherwise. It occurs that Spieß quoted Huber himself who had mentioned “Legendre’s criterion [Maßstab] of least squares”.

2.2. Legendre

Legendre (1805, pp. 72, 73) introduced the principle of least squares:

Of all the principles that can be proposed [for solving redundant systems of linear equations], I think there is none more exact, or easier to apply, than that which we have used in this work; it consists of making the sum of the squares of the errors [of the residuals] a minimum. This method establishes a kind of equilibrium among the errors, which, since it prevents the extremes from dominating, is appropriate for revealing the state of the system which most nearly approaches the truth.

Translation by Stigler (1986, p. 13). Legendre also indicated that the absolute values of the extremes [again: of the residuals] should be confined within the shortest possible interval. He had not added that it was the minimax principle (§ 6.3.2) rather than his innovation that ensured his desire.

2.3. Adrain

The American mathematician Adrain (1809) justified the principle of least squares and the [normal distribution]² at about the same time as Gauss did and applied it to the solution of several problems, see below (Dutka 1990). He also indicated that the lack of space prevented him to discuss the adjustment of pendulum observations. About ten years later he (1818a) published that study in which he revealed two mistakes in Laplace’s pertinent calculations (1798–1825, t. 2, § 42 of Livre 3). The same year his derivation of the length of the larger semi-axis of the Earth’s ellipsoid of revolution (1818b) appeared. Incidentally, that length (6378.629 km) was sufficiently close to a modern determination of 1940 by F.N. Krasovsky (6378.245 km).

² Adrain included his work in a periodical published by himself for the year 1808; however, its pertinent issue appeared only in 1809 (Hogan 1977). Adrain’s library included a copy of Legendre’s memoir (Coolidge 1926) in which, however, the normal distribution was lacking; furthermore, it is unknown when had Adrain obtained the memoir. The term *normal distribution* appeared in 1873 (Kruskal 1978) and was definitively introduced by K. Pearson (1894).

Adrain's main paper was first mentioned much later (C. Abbe 1871) but his second article had become known to Olbers who (Schilling 1909, p. 711) informed Gauss about it. An American author, wrote Olbers to Gauss, had mentioned his previous paper and "ascribed" the MLSq to himself. Gauss hardly made any comment; the priority strife with Legendre was apparently enough for him. Here are Adrain's derivations of the normal distribution.

a) Lines a and b are measured in the field with errors x and y respectively and

$$x/a = y/b \quad (1)$$

and the total error is fixed:

$$x + y = c. \quad (2)$$

Introducing the density of the observational errors φ and tacitly assuming their independence, Adrain applied the principle of [maximum likelihood]

$$\varphi(x; a) \varphi(y; b) = \max$$

so that, after allowing for conditions (1) and (2),

$[\varphi'(x; a)/\varphi(x; a)]dx + [\varphi'(y; b)/\varphi(y; b)]dy = 0$, $\varphi'(x; a)/\varphi(x; a) = mxa$, etc.

b) Suppose that for linear measurements

$$x^2 + y^2 = r^2,$$

then

$$W = \varphi(x) \varphi(y) - \lambda(x^2 + y^2) = \max, \varphi'(x)\varphi(y) - 2\lambda x = 0,$$

$$\varphi(x)\varphi'(y) - 2\lambda y = 0,$$

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

Adrain then wrote out the joint distribution of both these errors and indicated that the appropriate contour lines were ellipses (ellipses of errors, as they were later called in the theory of errors).

Conditions (1) and (2) hardly conform to reality; thus, the former describes the action of systematic errors. Also arbitrary is the condition applied in the second justification. Nevertheless, John Herschel (1850), Maxwell (1860), Thomson & Tait (1867, p. 314) and Krylov (1950, Chapt. 8) repeated that demonstration without any references (Sheynin 1965). Later on Kac (1939) and Linnik (1952) weakened the condition of independence.

Adrain was now able to prove quite simply that the arithmetic mean of direct measurements was optimal; this, of course, conformed to the principle of least squares in case of several unknowns. Finally, Adrain showed how to adjust a traverse (a polygon with measured sides and bearings) by the principle of least squares and, what is also remarkable, he calculated corrections to directly measured magnitudes rather than to their functions which were not independent from each other.

2.4. Gauss

Gauss (1809a; 1809b, § 186) applied the principle of least squares from 1794 or 1795. In the second instance, he called it “our principle”: *Unser Princip, dessen wir uns seit dem Jahre 1795 bedient haben*, and in both cases he mentioned Legendre. Much later Gauss (1823a) once more mentioned Legendre, although not in the memoir itself (1823b, § 17). Legendre (letter to Gauss 31.5.1809, see Gauss, *Werke*, Bd. 9, p. 380) was badly offended. He rightfully stated that priority is only established by publication. Gauss did not answer him, and Legendre (1820, pp. 79, 80) charged him with appropriating the MLSq.

Many authors commented on this episode, and I quote May (1972, p. 309) and Biermann (1966, p. 18):

Gauss cared a great deal for priority. [...] But to him this meant being first to discover, not first to publish; and he was satisfied to establish his dates by private records, correspondence, cryptic remarks in publications. [...] Whether he intended it so or not, in this way he maintained the advantage of secrecy without losing his priority in the eyes of later generations.

What is forbidden for usual authors, ought to be allowed for Gausses and in any case we must respect his [Gauss’] initial considerations.

It seems that Legendre could have stated in 1820 that he was the inventor of the MLSq and that, in spite of Gauss’ claim, everyone will agree with him. Regrettably, this did not happen. Laplace (1812/1886, p. 353) objectively described the discovery of the MLSq but did not add that Legendre had not really substantiated it. Then, he offered his own version of the theory of errors only suitable for a large number of observations and for other conditions necessary for the central limit theorem, as it was named by Polya in 1920, and other French mathematicians including Poisson followed him and Legendre and did not even mention Gauss. This seriously diminished the importance of their work.

In his letter to Gauss, Legendre also noticed that Euler rather than Laplace, as Gauss (1809, § 177) stated, calculated the integral of the

exponential function of the negative square. Later Gauss (*Monatliche Correspondenz*, Bd. 21, p. 280) explained that he noticed his mistake when the printing of his book was almost completed and that it was Laplace who provided the final formulation of the result. The Editors of Gauss (1887), who noticed that explanation did not supply the year of that Bd. 21.

I (Sheynin 1999a; 1999b) described the possible cases in which Gauss could have applied the MLSq before 1805 and named many of his colleagues and friends to whom he had communicated his discovery. Unexpectedly, it occurred that von Zach, who allegedly refused to testify to Gauss' priority, had not until 1805 known the formulation of the principle of least squares, and, furthermore, that he (1813, p. 98n) indirectly agreed with the latter's statements by repeating them without any qualification remark:

The celebrated Dr Gauss was in possession of that method since 1795 and he advantageously applied it when determining the elements of the elliptical orbits of the four new [minor] planets as it can be seen in his excellent work [Theoria motus].

Regrettably, "it" is not seen there.

This passage is even more important than Zach's editorial acceptance of Gauss' priority (noticed by Dutka 1996, p. 357). In 1809, Zach's periodical, *Monatliche Correspondenz*, carried an anonymous review of Gauss' *Theoria motus*, and there, on p. 191, Gauss' pertinent claim was repeated.

Gauss' claim about his early use of the MLSq is not generally accepted, see for example Marsden (1995, p. 185) who nevertheless had not mentioned the opposite opinion of Brendel (1924) and Galle (1924, p. 9) or of Gauss' contemporaries³. In any case, Gerardy (1977),

³ Their opinion should not be forgotten. Here is another example. Encke (1851, p. 2) believed that Gauss had applied the MLSq when determining the orbit of Ceres, the first-discovered minor planet (Gauss did not comment). In Note 20 to Chapter 6 I mentioned an inadmissible free and easy manner adopted by a certain author (Stigler 1986) with respect to Euler. His attitude towards Gauss was not better. Here are his statements: Legendre "immediately realized the method's potential" (p. 57), but "there is no indication that [Gauss] saw its great potential before he learned of Legendre's work" (p. 146); then (p. 143), only Laplace saved Gauss's argument [his first justification of the MLSq] from joining "an accumulating pile of essentially ad hoc constructions"; and, finally (p. 145), Gauss "solicited reluctant testimony from friends that he had told them of the method before 1805". I (Sheynin 1999a, b) had refuted these astonishing declarations which Stigler (1999), the first ever slanderer of the great man, repeated slightly less impudently, also see § 1.4. Regrettably, no-one supported me; on the contrary, Stigler's first book met with universal approval al-

drawing on archival sources, discovered that Gauss, in 1802–1807, had participated in land surveying (in part, for his own satisfaction) and concluded, on p. 19 (note 16) that Gauss started using the method not later than in 1803. Regrettably, Gerardy concentrated on describing Gauss' simple calculations and his statement mentioned just above was not quite definite.

There are many other instances including that mentioned by von Zach (above) in which Gauss could have well applied his invention at least for preliminary, trial calculations, or short cuts. For him, the MLSq was not a cut and dry procedure, see § 5-3. Then, possible mistakes in the data (Ibidem) and weighing the observations could have made justification impossible.

As to the communication of his discovery, I proved that among those whom Gauss had informed before 1805 were Bessel and Wolfgang Bolyai (the father of the cofounder of the non-Euclidean geometry, Janos or Johann Bolyai), – and Olbers which was known long ago. In 1812, Olbers promised Gauss to state publicly that he came to know about the MLSq from Gauss a few years before Legendre's publication, but he only did so in 1816. In 1812–1815, Olbers just did not publish anything suitable for inserting such a remark, see *Catalogue of Scientific Literature* of the Royal Society.

3. *Theoria Motus* (1809b)

In accordance with the publisher's demand, this book appeared in Latin. Its German original is lost and Gauss' correspondence (letter from Olbers of 27.6.1809, see Schilling (1900, p. 436)) proves that, while translating, he essentially changed its text. The treatment of observations occupies only a small part of the book.

1) The Boscovich method (see § 6.3.2). Suppose that n equations (1.2) in m unknowns ($n > m$) are adjusted by that method. Then, as Gauss (§ 186) remarked, equation (6.13) meant that exactly m residual free terms will be zero. Somewhat below, in the same § 186, Gauss qualified his statement by taking into account the other Boscovich equation (6.12) but mistakenly attributed it to Laplace. In § 174 he stated that the formulated corollary was undesirable although in §§ 188 and 189 he apparently agreed that the Boscovich method might ensure

though he, in addition, left aside the ancient history as well as such scholars as Kepler, Lambert and Helmert. Hald (1998, p. xvi), whose outstanding contribution deserves highest respect, called Stigler's book "epochal". I am unable to understand suchlike opinions.

a first approximation. His remark, that can be easily proved, means that he knew an important theorem in linear programming.

2) The [normal distribution] (§§ 175–177). Gauss (§ 177) assumed “as an axiom” that the arithmetic mean of many observations was the most probable value of the measured constant “if not absolutely precisely, then very close to it”. He (§ 175) derived the density φ of observational errors believing that it was [unimodal] and “in most cases” even; this, then, was his understanding of the properties of random errors. Finally, in order to justify the principle of [maximal likelihood], Gauss (§ 176) proved the “fundamental principle” of inverse probability, see my § 7.1-1, for the case of equal probabilities of the various hypotheses. However, the principle of the arithmetic mean (above) already implied his restriction (Whittaker & Robinson 1949, p. 219).

And so, if the observations are denoted by $x_i, i = 1, 2, \dots, n$, then, according to the principle of maximal likelihood,

$$[\varphi'(x_1 - a)/\varphi(x_1 - a)] + [\varphi'(x_2 - a)/\varphi(x_2 - a)] + \dots + [\varphi'(x_n - a)/\varphi(x_n - a)] = 0,$$

where a is the estimator sought, coinciding, as stipulated, with the arithmetic mean x_0 . If

$$x_i = x_1 - nN, i = 2, 3, \dots, n,$$

then

$$x_1 + (x_2 + x_3 + \dots + x_n) = x_1 + (n-1)x_1 - n(n-1)N, \\ N = (x_1 - x_0)/(n-1), x_i - x_0 = -N,$$

$$\varphi'(x_1 - x_0)/\varphi(x_1 - x_0) = (1-n)\varphi'(-N)/\varphi(-N) = -(1-n)\varphi'(N)/\varphi(N), \\ \varphi'[N(n-1)]/\{(1-n)\varphi[N(n-1)]\} = -\varphi'(N)/\varphi(N), \\ \varphi'(x)/x\varphi(x) = \text{Const},$$

$$\varphi(x) = (h/\sqrt{\pi})\exp(-h^2x^2), h > 0. \quad (3)$$

Gauss (§ 178) called h the “measure of precision” (*gradus praecisionis*). It might be supposed that, from the very beginning, he was not satisfied with his derivation. His wording of the principle of the arithmetic mean and of the properties of the density of observational errors contained qualification remarks whereas the obtained principle of least squares (see below § 2-3) occurred to be an axiom. Again, it is difficult to believe that Gauss was pleased with the appearance of a universal law of error. Later he (1821/1887, pp. 193 and 194; 1823a/1887, p. 196) remarked that his derivation had depended on a hypothetically assumed distribution. And here is Bertrand’s opinion (1888, p. XXXIV):

Gauss had not claimed to establish the “vérité”, he attempted to search for. Bertrand (pp. 180, 181) also remarked that the mean of the values of some function did not coincide with the mean value of its arguments, which, in his opinion, testified against the principle of arithmetic mean. Gauss, however, considered direct measurements. Note also that he (his letter to Encke of 1831; *Werke*, Bd. 8, pp. 145, 146) “not without interest” acquainted himself with the attempt of his correspondent to justify the arithmetic mean by deterministic analytical axioms. Many authors made similar efforts and Zoch (1935) concluded that, although they were unsuccessful, the postulate of the arithmetic mean can nevertheless be established without stochastic considerations. His finding was unrelated to the theory of errors, but the pertinent investigations apparently served as the point of departure for the theory of invariant statistical hypotheses and estimators (Lehmann 1959, Chapt. 6).

Gauss (1845/1873, p. 143) left a lesser known statement about the arithmetic mean. He remarked that the random variations corrupting observations mostly compensate one another so that the mean becomes ever more reliable as the number of observations increases. This is “generally absolutely right”, and often led to “splendid results” in natural sciences. However, Gauss continued, an important condition, often overlooked and difficult to check, was that the disordered variations ought to be entirely independent from each other, cf. § 4-4.

3) The principle of least squares (§ 179) followed immediately. Gauss, however, added that, similar to the principle of the arithmetic mean, it should be considered an axiom [considered as a corollary of an axiom?]. A special point here is that, instead of the real errors the principle of least squares was formulated with regard to residual free terms. Helmert (1872, p. 75) indicated this fact but paid scant attention to it and had not mentioned Gauss. Apparently he had not realized that the normal law was [stable] (cf. §§ 7.2-6 and 2-7).

4) The precision of the arithmetic mean. Gauss, naturally, restricted his attention to the case of the [normal distribution]. Later he (§ 4) abandoned this restriction.

5) The precision of a random sum (marginal note to § 183, included not in the German translation, but in Gauss’ *Werke*, Bd. 7). Suppose that

$$x = a + b + c + \dots,$$

then

$$h_x = 1 \div [(1/h_a^2) + (1/h_b^2) + (1/h_c^2) + \dots]^{1/2}.$$

Gauss did not explain his note; it might be supposed that the terms above were normally distributed since he only introduced h for that law. However, he may well have derived this formula in the general case.

6) The precision of the [estimators of the] unknowns (§ 182; 1811, § 13). Suppose that these estimators are determined by solving a system of normal equations in accordance with the Gauss method of successive eliminations. Then, assuming that the precision of a direct measurement is unity, the precision of the estimator of the last unknown is equal to the root of its coefficient in the last reduced equation. Also see my § 4-5.

4. “Determining the Precision of Observations” (1816)

1) The precision of the measure of precision h in formula (3). Suppose that the errors of m [independent] observations are $\alpha, \beta, \gamma, \dots$. Then the most probable value of that magnitude is determined by the condition

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

and is therefore equal to

$$h_0 = \{m/[2(\alpha^2 + \beta^2 + \gamma^2 + \dots)]\}^{1/2} = 1/\sigma\sqrt{2}.$$

In the last expression, which is my own, σ is the mean square error of an observation. Gauss also indicated that

$$P(h_0 - \lambda \leq h \leq h_0 + \lambda) = \theta(\lambda\sqrt{m}/h_0), \quad \theta(t) = (2/\sqrt{\pi}) \int_0^t \exp(-z^2) dz,$$

so that, for $P = 1/2$, $\lambda = \rho h_0/\sqrt{m}$, $\rho \approx 0.477$.

In addition, for distribution (3),

$$P(|x| \leq \rho\sqrt{h}) = 1/2, \quad \text{and } r = \rho/h$$

is the probable error formally introduced by **Bessel** (1816, pp. 141 – 142).

Let

$$S_n = |\alpha|^n + |\beta|^n + |\gamma|^n + \dots, \quad K_n = \int_{-\infty}^{\infty} x^n \varphi(x) dx,$$

then, for large values of m ,

$$P(-\lambda \leq S_n - mK_n \leq \lambda) = \theta\{\lambda/[2m(K_{2n} - K_n^2)]^{1/2}\}, \quad (4)$$

where mK_n is the most probable [the mean] value of S_n . Actually, Gauss treated absolute moments and the formula for K_n should be corrected. Formula (4) was proved by Helmholtz (§ 8) and then by Lipschitz (1890),

but Cramér (1946, § 28.2) noted that it was a particular case of the central limit theorem.

Finally, Gauss derived a formula for the absolute moments of the normal law

$$mK_n = S_{n0} = m\Pi[(n-1)/2]/h^n\sqrt{\pi}, \quad \Pi(x) = \Gamma(x+1),$$

so that h (and therefore r) could have been estimated by S_{n0} , the mean value of S_n . Comparing the probable intervals of r for different n , Gauss concluded that $n = 2$ secured its best estimator.

In one of his letters of 1825 Gauss (*Werke*, Bd. 8, p. 143) objected to the probable error as “depending on a hypothesis” [on the law of distribution]. Still, again in his correspondence, he applied it quite a few times (Sheynin 1994a, p. 261). Natural scientists, for example Mendeleev (§ 10.10.3) and Newcomb (§ 10.9.4), followed suit and Bomford (1971, pp. 610, 611) “reluctantly” changed from probable to mean square error in the last (!) edition of his book.

2) Denote $1/h\sqrt{2} = \alpha$ and let $n = 2$. Then

$$[m(K_4 - K_2^2)]^{1/2} = \alpha^2 \sqrt{2m}$$

and, in accordance with formula (4), the sum of squares S_2 is distributed normally $N [m\alpha^2; \alpha^2 \sqrt{2m}]$. This is the asymptotic chi-squared distribution, cf. Cramér (1946, § 20.2).

5. “The theory of combinations” (1823–1828)

I consider the main part of this memoir in which Gauss provided his definitive justification of the MLSq by the principle of maximum weight [of minimal variance], and I add a few words about its supplement (1828).

1) Random errors and the density of observational errors. Gauss (§§ 1–3) distinguished between random and systematic errors. He (§ 4) then repeated (see my § 2-2) the definition of density and listed its properties. The mean value of the errors (§ 5) was equal to zero; otherwise, as Gauss additionally remarked, it determined the action of constant errors.

2) The measure of precision. Gauss (§ 6) introduced a measure of precision [the variance]

$$m^2 = \int_{-\infty}^{\infty} x^2 \varphi(x) dx$$

calling it the mean error to be feared, – *des mittleren zu befürchtenden Fehler; errorum medium metuendum* (1821/1887, p. 194; 1823b, § 7). Gauss (§ 7 and his letter to Bessel of 28.2.1839, *Werke*, Bd. 8, pp. 146, 147) stressed that an integral measure of precision was preferable to a local measure. He (1823b, § 6) also indicated that the quadratic function was the simplest [from integral measures], and in 1821 he (1887, p. 192) dwelt on his choice in more detail: it was also connected with “some other, extremely important advantages which no other function possesses. However, any other even degree could have been selected as well ...” Could have been chosen in spite of the advantages of the variance? Bienaymé (1853/1867, pp. 167–169) proved that a formula of the type of (5), see below, was not valid for any other even exponent; a clear exposition of this proof is due to Idelson (1947, pp. 269–271). Therefore, Bienaymé continued, the choice of the variance was unavoidable. I doubt, however, that, as he believed (p.169), Gauss was here mistaken. The sample variance (see § 4–6) is distribution-free.

3) An inequality of the **Bienaymé–Chebyshev** type. Gauss (§ 9) examined the probability

$$\mu = P(|\xi| \leq \lambda m) = \int_{-\lambda m}^{\lambda m} \varphi(x) dx$$

for a [unimodal] density of observational errors ξ having variance m^2 and proved (§ 10) that

$$\lambda \leq \mu\sqrt{3} \text{ for } \mu \leq 2/3 \text{ and } \lambda \leq 2/3\sqrt{1-\mu} \text{ for } 2/3 \leq \mu \leq 1.$$

Cramér (1946, § 15.7 and Example 4 to Chapt. 15–20) more easily proved this “remarkable” theorem, as Gauss called it, whereas Seal (1967/1970, p. 210) indicated that Gauss had wished to abandon the universality of the normal distribution since it occurred that, anyway, $P(|\xi| \leq 2m) \geq 0.89$. But should we forget his own, although indirect, arguments and doubts?

4) Independence. Gauss (§ 18) indicated that, if some observation was common for two functions of observational results, the errors of these functions will not be independent from one another and the mean value of their product will not therefore vanish⁴. In one of his examples,

⁴ It is not amiss to add that the primary triangulation of the Soviet Union consisted of chains independent one from another in the Gauss’ sense. This, together with other conditions, enabled the geodesists to estimate realistically the precision of the whole great net (Sakatow 1950/1957, pp. 438–440). And in general, geodesists, not necessarily mentioning Gauss, were keeping to his opinion. I also note that Kapteyn (1912), who had not cited

Gauss calculated the variance of a linear form of independent random variables.

Gauss (1809b, § 175; 1823b, § 15) mentioned independence even earlier but without explanation, and, later he (1826/1887, p. 200; 1828, § 3) described the mutual dependence of magnitudes known from observation by the existence of functional connections between them. This meant, for example, that the adjusted angles of a triangle, since their sum was equal to 180° plus the spheroidal excess, were dependent on one another. See also end of § 2-2.

His reasoning heuristically resembles the definition of independence of events in the axiomatic theory: events are independent if the probability of their product is equal to the product of their probabilities. Now, in mathematical statistics the definition of independence is different. An orthogonal transformation of independent and normally distributed magnitudes leads to their as though “adjusted” values, – to their linear forms of a certain type, which are nevertheless independent (the Fisher lemma; Cramér (1946, § 29.2)). Here is K. Pearson’s appropriate statement (1920/1970, p. 187) which I do not however understand: for Gauss

The observed variables are independent, for us [they] are associated or correlated. For him the non-observed variables are correlated owing to their known geometrical relations with observed variables; for us, [they] may be supposed to be uncorrelated causes, and to be connected by unknown functional relations with the correlated variables.

5) The principle of maximum weight for [unbiased] estimators. Gauss described this subject ponderously. For that matter, Helmert (1872) and Idelson (1947) are in general much better understood. Suppose that, without loss of generality, the initial equations are

$$a_i x + b_i y = G_i = g_i + \varepsilon_i, i = 1, 2, \dots, n$$

where ε_i is the error of the free term g_i . The estimators of the unknowns might be represented by linear forms, for example by $x = [\alpha G]$ with unknown coefficients α_i so that

$$m_x^2 = [\alpha\alpha]m^2 \quad (5)$$

where m^2 is the variance of an observation.

Gauss and was unsatisfied with the then originating correlation theory, proposed to estimate quantitatively the dependence between series or functions of observations by issuing from the same notion of independence, see Sheynin (1984, § 9.2.1). His article went unnoticed.

It is easy to prove that $[a\alpha] = 1$, $[b\alpha] = 0$ and the condition of maximal weight will be

$$W = [a\alpha] - 2Q_{11}[a\alpha] - 2Q_{12}[b\alpha] = \max$$

where Q_{11} and Q_{12} are the Lagrange multipliers. Similar considerations, and, in particular, an estimation of precision resembling formula (5), are also possible for the other unknowns. It occurs that the estimators of the unknowns are determined from the normal equations and their weights are calculated by means of the Lagrange multipliers of the type of Q_{ii} which, like the other multipliers Q_{ij} , are determined from the same normal equations with partly unit and partly zero free terms. Thus, in formula (5) $[a\alpha] = Q_{11}$. According to the above, it follows that such formulas can be made use of even before observation; the general layout of the geodetic network and the crude values of its angles obtained during reconnaissance make it possible to calculate the Q_{ij} . And (what Gauss had not known) these multipliers are connected with covariations; thus, $Q_{12} = E(xy)$.

6) The estimator of the sample [variance]. Gauss (§§ 37, 38) proved that, for n observations and k unknowns, the unbiased sample variance and its estimator were, respectively,

$$m^2 = E[vv]/(n - k), \quad (6a)$$

$$m_o^2 = [vv]/(n - k) \quad (6b)$$

where v_i were the residual free terms of the initial equations. Instead of the mean value, the sum of squares $[vv]$ itself has to be applied. Coupled with the principle of maximal weight (of least variance), formulas (6) provide effective estimators, as they are now called. Gauss (1823a/1887, p. 199) remarked that the acceptance of his formula (6b) instead of the previous expression (§ 7.2-5), whose denominator was equal to n , was demanded by the “dignity of science”.

7) The precision of the estimator of the sample variance. Gauss (§ 40) directly calculated the boundaries of the $\text{var } m_o^2$ by means of the fourth moment of the errors and indicated that for the normal distribution

$$\text{var } m_o^2 = 2m^4/(n - k). \quad (6c)$$

He somehow erred in calculating the abovementioned boundaries, see (15); in addition, his formulas should have included the unknown magnitude $E\varepsilon_i^2$ (ε_i were the observational errors) rather than m^2 . Formula

(6c) shows that m_0^2 is a consistent estimator of the sample variance; this persists in the general case, see formulas (15).

8) Other topics. Gauss also determined the variance of a linear function of the estimators of the unknowns (which are not independent) and provided expedient procedures for further calculations after additional data become known or after the weights of some observations have to be changed.

9) Another manner of adjusting observations. In the supplement (1828) to his memoir Gauss described the adjustment of observations by the MLSq according to the pattern of conditional observations. In geodetic practice, it is often expedient to issue from the directly measured magnitudes and *conditional equations* rather than from observational equations (1.2). Sometimes both kinds of equations are made use of at the same time, but I leave this case aside and consider now a (later) typical chain of, say, 10 triangles of triangulation. Each angle is measured as are the lengths of two extreme sides (baselines) whose directions (azimuths) are determined by astronomical observations. The observational errors are such that both the baselines and the azimuths might be considered exact; only the angles are adjusted. Each measured angle q_i provides an equation

$$x_i - q_i = v_i, \quad (7)$$

where the first term is the real value of the angle and the right side is the sought correction. Now, the condition of closing the first triangle (I disregard its excess) is

$$x_1 + x_2 + x_3 - 180^\circ = 0. \quad (8)$$

Extremely simple is also the condition that demands that the azimuth of the first baseline plus the algebraic sum of the appropriate angles be equal to the azimuth of the second baseline. The sine theorem is however needed for the transition from the first baseline to the second one, but a first approximation is achieved by introducing the measured angles so that the required trigonometric equation is linearized. It follows that all the conditions can be written as

$$[av] + w_1 = 0, [bv] + w_2 = 0, \text{ etc.} \quad (9)$$

Formed by means of equations (7), they should be exactly fulfilled and the number of the terms in the square brackets is either three, as in equations of the type of (8), or more, depending on the number of the triangles in the chain. The adjustment proper consists in determining the conditional minimum of $[vv]$ with the usual application of the

Lagrange multipliers and the corrections v_i are determined through these multipliers. Strangely enough, only Helmert (1872, p. 197) was the first to provide such an explanation.

6. Additional considerations

Having substantiated the MLSq, Gauss nevertheless deviated from rigid rules; one pertinent example is in § 6.3.2. Here, I have more to say.

1) The number of observations. In his time, methods of geodetic observations were not yet perfected. Gauss himself was successfully developing them and he understood that a formal estimation of precision could describe the real situation only after all the conditions (§ 4-9) were allowed for, i.e., only after all the field work was done. It is no wonder, then, that Gauss continued to observe each angle at each station until being satisfied that further work was useless, see Note 11 in Chapt. 6.

2) Rejection of outliers. This delicate operation does not yield to formal investigation since observations are corrupted by systematic errors, and, in general, since it is difficult to distinguish between a blunder and a “legitimate” large error. Statistical tests, which had appeared in the mid-19th century, have not been widely used in the theory of errors. Gauss himself (letter to Olbers of 1827, *Werke*, Bd. 8, pp. 152, 153) had indicated that, when the number of observations was not very large, and a sound knowledge of the subject was lacking, rejection was always doubtful.

3) Calculations. Without even a comptometer, Gauss was able to carry out difficult calculations; once he solved a system of 55 normal equations (letter to Olbers of 1826; *Werke*, Bd. 9, p. 320). For other examples see Sheynin (1979, p. 53). His preparatory work (station adjustment; compilation of the initial equations, see § 4-9, and of the normals themselves) had to be very considerable as well.

Sometimes Gauss applied iterative calculations (letter to Gerling of 1823; *Werke*, Bd. 9, pp. 278–281), also see Forsythe (1951) and Sheynin (1963). The first to put on record this fact, in 1843, was Gerling himself. Then, Gauss (1809b, § 185) left an interesting qualitative remark stating that “it is often sufficient” to calculate approximately the coefficients of the normal equations. The American astronomer Bond (1857) had applied Gauss’ advice and Newcomb (1897, p. 31) followed suit.

As a calculator of the highest calibre (Maennchen 1918/1930, p. 3), *Gauss was often led to his discoveries by means of mentally agonizing precise calculations [...]; we find [in his works] substantial*

tables whose compilation would in itself have occupied the whole working life of some calculators of the usual stamp.

I ought to add that Gauss made some mistakes in his computations possibly because, first, he had not invariably checked them, see for example Gerardy (1977) or his own methodological note (1823c) where the signs of dx and dy were wrong. Second, Gauss calculated “unusually fast” (Maennchen 1918/1930, p. 65ff).

Maennchen did not study Gauss’ geodetic calculations possibly because in his time the solution of systems of linear equations had not yet attracted the attention of mathematicians.

For my part, I note that, when compiling a certain table of mortality, Gauss (*Werke*, Bd. 8, pp. 155, 156) somehow calculated the values of exponential functions b^n and c^n for $n = 3$ and $7(5)97$ with $\lg b = 0.039097$ and $\lg c = -0.0042225$.

Here, now, is Subbotin’s conclusion (1956, p. 297) about the determination of the orbits of celestial bodies but applicable to my subject as well: Lagrange and Laplace

Restricted their attention to the purely mathematical aspect [of the problem] whereas Gauss had 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.

4) Estimation of precision (Sheynin 1994, pp. 265, 266). In his letters to Bessel (in 1821) and Gerling (in 1844 and 1847) Gauss stated that the estimation of precision based on a small number of observations was unreliable. In 1844 he combined observations made at several stations and treated them as a single whole, cf. Laplace’s attitude (§ 2-7). And in 1847 Gauss maintained that, lacking sufficient data, it was better to draw on the general knowledge of the situation.

7. More about the method of least squares

1) In spite of Gauss’ opinion, his first justification of the MLSq became generally accepted (Sheynin 1995b, § 3.4), in particular because the observational errors were (and are) approximately normal whereas his mature contribution (1823b) was extremely uninviting; and the work of Quételet (§ 10.5) and Maxwell (§ 10.9.5) did much to spread the idea of normality. Examples of deviation from the normal law were however accumulating both in astronomy and in other branches of natural sciences as well as in statistics (Sheynin 1995b, § 3.5; again Quetelet and Newcomb, see § 10.9.4). And, independently from that fact, several authors came out against the first substantiation. Markov

(1899), who referred to Gauss himself (to his letter to Bessel, see my § 4-2), is well known in this respect but his first predecessor was Ivory (§ 10.10-1).

The second justification was sometimes denied as well. Thus, Bienaymé (1852, p. 37) declared that Gauss had provided considerations rather than proofs; see also Poincaré's opinion in § 11.2-7.

2) When justifying the MLSq in 1823 in an essentially different way, Gauss called the obtained estimators *most plausible* (*maxime plausibiles*, or, in his preliminary note (1821), *sicherste*, rather than as before, *maxime probable*, *wahrscheinlichste*. For the case of the normal distribution, these are jointly effective among unbiased regular estimators⁵.

The second substantiation of the MLSq can be accomplished by applying the notions of multidimensional geometry (Kolmogorov 1946; Hald 1998, pp. 473, 474). Kolmogorov (p. 64) also believed that the formula for m^2 (6a) should, after all, be considered as its definition. Much earlier Tsinger (1862, § 33) stated that it already "concealed" the MLSq.

3) Mathematicians had not paid due attention to Gauss' work on the MLSq (§§ 2-3 and 13.2-7), and neither did statisticians, see the Epigraph to this book which apparently complements the following passage (Eisenhart 1978, p. 382):

When Karl Pearson and G. Udny Yule began to develop the mathematical theory of correlation in the 1890s, they found that much of the mathematical machinery that Gauss devised [...] was immediately applicable. [...] Gauss' contributions to the method of least squares embody mathematics essential to statistical theory and its applications in almost every field of science today.

I really think that K.P. and Yule only discovered Gauss at a late stage of their work.

8. Helmert

It was Helmert who mainly completed the development of the classical Gaussian theory of errors; furthermore, some of his findings were interesting for mathematical statistics. With good reason Schumann (1917, p. 97) called him "Master of both the *niedere* [surveying and applied] geodesy and higher [triangulation etc., gravimetry, figure of the Earth] geodesy". Until the 1930s, Helmert's treatise (1872) remained

⁵ Concerning this rarely mentioned concept see Cramér (1946, § 32.6).

the best source for studying the error theory and the adjustment of triangulation.

Indeed, its third, posthumous edition of 1924 carried a few lines signed by a person (H. Hohenner) who explained that, upon having been asked by the publishers, he had stated that the treatise still remained the best of its kind. His opinion, he added, convinced the publishers.

Helmert (1886, pp. 1 and 86) was the first to consider appropriate geodetic lines rather than chains of triangulation, and this innovation, developed by Krasovsky, became the essence of the method of adjustment of the Soviet primary triangulation (see Note 19 to Chapt. 6 and Sakatow 1950, § 91). Another of his lesser known contributions (Helmert 1868) was a study of various configurations of geodetic systems. Quite in accordance with the not yet existing linear programming, he investigated how to achieve necessary precision with least possible effort, or, to achieve highest possible precision with a given amount of work. Some equations originating in the adjustment of geodetic networks are not linear, not even algebraic; true, they can be linearized (§ 4-9), and perhaps some elements of linear programming could have emerged then, in 1868, but this had not happened. Nevertheless, Helmert noted that it was expedient to leave some angles of a particular geodetic system unmeasured, cf. § 2-1, but this remark was only academic: all angles have always been measured at least for securing a check upon the work as a whole.

I will describe now Helmert's stochastic findings.

1) The chi-square distribution (E. Abbe 1863; M.G. Kendall 1971). Abbe derived it as the distribution of the sum of the squares of normally distributed errors. He wished to obtain a test for revealing systematic errors, and he required, in particular, the distribution of the abovementioned function of the errors since it was indeed corrupted by those errors. Exactly his test rather than the distribution obtained was repeatedly described in the geodetic literature whereas Linnik (1958/1961, pp. 109–113) introduced a modified version of the Abbe test.

Helmert (1876b) provided his own derivation of the χ^2 distribution which he first published without justification (1875a). Neither then nor much later (see § 10.6-2) did he mention Abbe. Actually, he continued after Gauss (1816), see § 3, by considering observational errors $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ and the sum of their powers $\sum \varepsilon_i^n$ for the uniform and the [normal] distributions and for an arbitrary distribution as $n \rightarrow \infty$. In the last instance, he proved the Gauss formula (4) and then specified it for the abovementioned distributions. He derived the χ^2 distribution by

induction beginning with $n = 1$ and 2; Hald (1952, pp. 258–261) provided a modernized derivation.

2) Much later Helmert (1905) offered a few tests for revealing systematic influences in a series of errors which he wrote down as

$$v_1\varepsilon_1 + v_2\varepsilon_2 + \dots + v_n\varepsilon_n$$

with $v_i = 1$ or -1 and $\varepsilon_i > 0$. He issued from the formula

$$P(|\xi - E\xi| \leq m) \approx 0.68 \quad (10)$$

where m was the mean square error of ξ (and thus restricted his attention to the normal law): if the inequality in the left side of (10) did not hold, then, as he thought, systematic influences were present. When deriving his tests, Helmert considered Σv_i , $|\Sigma v_i|$, runs of signs of the v_i and functions of the errors ε_i themselves and in this last-mentioned case he provided a somewhat modified version of the Abbe test.

3) The Peters formula (1856) for the mean absolute error. For n normally distributed errors it was

$$\theta = \Sigma |v_i| / \sqrt{n(n-1)}, \quad 1 \leq i \leq n \quad (11)$$

with v_i being the deviations of the observations from their arithmetic mean. Helmert (1875b) derived formula (11) anew because Peters had tacitly and mistakenly assumed that these deviations were mutually independent. Passing over to the errors ε_i , Helmert calculated the appropriate integral applying for that purpose the Dirichlet discontinuity factor. However, since the normal distribution is stable, it is possible to say now at once (David 1957) that formula (11) is correct because

$$E\Sigma |v_i| = \sqrt{n(n-1)} / h\sqrt{\pi},$$

where h is the appropriate parameter [measure of precision] of the initial normal distribution and, as it should be, $\theta = 1/h\sqrt{\pi}$.

Helmert also attempted to generalize the Peters formula by considering indirect measurements with k unknowns ($k > 1$). He was unable to derive the appropriate formula but proved that a simple replacement of $(n - 1)$ in formula (10) by $(n - k)$ resulted in underestimating the absolute error.

4) Helmert (1876b) calculated the variance of the estimator (11). His main difficulty here was the derivation of $E|v_i v_j|$, $i < j$, but he was able to overcome it and obtained

$$\{\pi/2 + \arcsin[1/(n-1)] - n + \sqrt{n(n-2)}\} / \pi n h^2.$$

Later Fisher (1920, p. 761) independently derived this formula.

5) In the same paper Helmert investigated the precision of the Gauss formula (6b). For direct measurements it can be replaced by the expression for the mean square error

$$m = \sqrt{\frac{[v\bar{v}]}{n-1}}.$$

Helmert derived it for the normal distribution by the principle of maximum likelihood, but had not remarked that the estimator obtained (which, however, directly followed from (6a) and was always applied in practice in geodesy) was, unlike the Gauss formula, biased.

Denote the observational errors by ε_i and their mean by ε , then

$$v_i = \varepsilon_i - \varepsilon$$

and the probability that these errors had occurred, as Helmert indicated in the context of his proof, was equal to

$$P = n(h/\sqrt{\pi})^n \exp[-h^2([v\bar{v}] + n\varepsilon^2)] dv_1 dv_2 \dots dv_{n-1} d\varepsilon. \quad (12)$$

This formula shows that, for the normal distribution, $[v\bar{v}]$, – and, therefore, the variance as well, – and the arithmetic mean are independent. Helmert had thus proved the important Student–Fisher theorem although without paying any attention to it.

A special feature in Helmert's reasoning was that, allowing for (6c), he wrote down the Gauss formula (6b) for the case of direct measurements (and, to repeat, for the normal distribution) as

$$m_o^2 = \frac{[v\bar{v}]}{n-1} \left[1 \pm \frac{\sqrt{2}}{\sqrt{n-1}} \right]; \quad (13)$$

that is, he considered the variance together with its mean square error.

Formula (13) also indirectly indicated the relative mean square error; Czuber (1891, p. 460) testified that Helmert had thought that $\text{var } m_o^2 / m_o^2$ was more important than $\text{var } m_o^2$ by itself and Eddington (1933, p. 280) expressed the same opinion. Czuber also proved that, for the normal distribution, that relative error was minimal for the estimator (6b).

In addition, Helmert noted that for small values of n the $\text{var } m_o^2$ did not estimate the precision of formula (6b) good enough and derived the following formula

$$E\left[m - \frac{[v\bar{v}]}{\sqrt{n-1}}\right]^2 = (1/h^2) \left\{ 1 - \sqrt{2} \frac{\Gamma(n/2)}{\Gamma[(n-1)/2]} \sqrt{n-1} \right\}. \quad (14)$$

He issued from the probability of the values of v_i , $i = 1, 2, \dots, (n - 1)$,

$$P = \sqrt{n(h/\sqrt{\pi})^{n-1}} \exp(-h^2[vv]) dv_1 dv_2 \dots dv_{n-1}$$

that follows from formula (12), noted that the probability $P(\varepsilon \leq [vv] \leq \varepsilon + d\varepsilon)$ was equal to the appropriate integral, and introduced new variables

$$t_1 = \sqrt{2}(v_1 + 1/2v_2 + 1/2 v_3 + 1/2v_4 + \dots + 1/2v_{n-1}),$$

$$t_2 = \sqrt{3/2} (v_2 + 1/3v_3 + 1/3v_4 + \dots + 1/3v_{n-1}),$$

$$t_3 = \sqrt{4/3} (v_3 + 1/4v_4 + \dots + 1/4v_{n-1}), \dots,$$

$$t_{n-1} = \sqrt{n/(n-1)} v_{n-1}.$$

Note that $[vv] = [tt]$ where, however, the first sum consisted of n terms and the second one, of $(n - 1)$ terms, and the Jacobian of the transformation was \sqrt{n} . The derivation of formula (14) now followed immediately since Helmert knew the χ^2 distribution. Taken together, the transformations from $\{\varepsilon\}$ to $\{v\}$ and from $\{v\}$ to $\{t\}$ are called after him.

Kruskal (1946) transformed formula (12) by introducing a bivariate “Helmert distribution” with variables

$$s = \sqrt{[vv]/n}, u = x - \mu,$$

where x was the arithmetic mean of n normally distributed observations $N(\mu; \sigma)$, and replaced h by σ . He mentioned several authors who had derived that new distribution by different methods, determined it himself by induction and indicated that the Student distribution followed from it, see Hald (1998, p. 424).

Finally, Helmert corrected the boundaries of the estimator (6b). As indicated by Gauss they were

$$2(v_4 - 2s^4)/(n - k); [1/(n - k)] (v_4 - s^4) + (k/n) (3s^4 - v_4),$$

where v_4 was the fourth moment of the errors and $s^2 = Em^2$. Helmert had discovered that the lower boundary was wrong and Kolmogorov et al. (1947) independently repeated his finding. Here is the final result; Maltzev (1947) proved that the lower bound was attainable: for non-negative and non-positive $(v_4 - 3s^4)$ the variance $\text{var } m_0^2$ is contained within, respectively,

$$[(v_4 - s^4)/(n - k) - (k/n) (v_4 - 3s^4)/(n - k); (v_4 - s^4)/(n - k)], \quad (15a)$$

$$[(v_4 - s^4)/(n - k); (v_4 - s^4)/(n - k) + (k/n) (3s^4 - v_4)/(n - k)]. \quad (15b)$$

9. Other topics

Gauss and Bessel were the originators of a new direction in practical astronomy and geodesy which demanded a thorough examination of the instruments and investigation of the plausibility of observational methods.

I mentioned Bessel in § 7 as well as in §§ 3-1 and 4-2. His achievements in astronomy and geodesy are well known; in addition to those already cited, I name the determination of astronomical constants; the first determination of a star's parallax; the discovery of the personal equation; the development of a method of adjusting triangulation; and the derivation of the parameters of the Earth's ellipsoid of revolution. He also determined the density of the total observational error made up of many heterogeneous components, but a rigorous solution of such problems became possible, with a doubtful exception of one of Cauchy's memoir (§ 10.1), only much later (§ 13.1-4)⁶.

The personal equation is the systematic difference of the moments of the passage of a star through the cross-hairs of an astronomical instrument as recorded by two observers. When studying this phenomenon, it is necessary to compare the moments fixed by the astronomers at different times and, consequently, to take into account the correction of the clock. Bessel (1823) had indeed acted appropriately, but in one case he failed to do so, and his pertinent observations proved useless. He made no such comment; furthermore, without any justification, he greatly overestimated their precision.

Bessel (1838, §§ 1 and 2) determined the densities of two functions of a continuously and uniformly distributed [random variable], and, unlike Laplace, he clearly formulated this problem. Nevertheless, he erred in his computations of the pertinent variances and probable errors⁷.

It became customary to measure each angle of a chain of triangulation an equal number of times and, which was more important, to secure their mutual independence so as to facilitate the treatment of the observations, – to separate the station adjustment from the adjustment of the chain as a whole. Bessel, however, did not keep to the

⁶ In 1839 Gauss informed Bessel (*Werke*, Bd. 8, pp. 146, 147) that he had read the latter's memoir with interest although the essence of the problem had been known to him for many years.

⁷ I (Sheynin 2000) discovered 33 mistakes in arithmetical and simple algebraic operations in Bessel's contributions collected in his *Abhandlungen* (1876). Not being essential, they testify to his inattention and undermine the trust in the reliability of his more involved calculations.

abovementioned condition (and had to adjust all the observations at once). There are indications that the actual rejection of his method annoyed him⁸.

References

The following abbreviations are used:

AHES – *Archive for History of Exact Sciences*

IMI – *Istoriko-Matematicheskie Issledovania*

ISI – *International Statistical Institute*

ISR – *International Statistical Revue*

JNÖS – *Jahrbücher f. Nationalökonomie u. Stat.*

OC – *Oeuvr. Compl.*

R – in Russian

UMN – *Uspekhi Matematicheskikh Nauk*

Abbe C. (1871), Historical note on the method of least squares. *Amer. J. Sci. Arts*, ser. 3, vol. 1, pp. 411–415.

Abbe, E. (1863), Über die Gesetzmässigkeit in der Verteilung der Fehler bei Beobachtungssreihen. *Ges. Abh.*, Bd. 2, 1989, pp. 55–81.

Adrain R. (1809), Research concerning the probabilities of the errors which happen in making observations. All three of his papers (see below) are reprinted in Stigler (1980, vol. 1).

– (1818a), Investigation of the figure of the Earth and of the gravity in different latitudes. *Trans. Amer. Phil. Soc.*, new ser., vol. 1.

– (1818b), Research concerning the mean diameter of the Earth. *Ibidem*.

– (1769, in Latin), Manuscript, same title as in 1778. In English, in *Festschrift for Lucien Le Cam*. New York, 1997, pp. 358–367.

– (1778; 1982, pp. 361–375, in Latin), The most probable choice between several discrepant observations and the formation therefrom of the most likely induction. *Biometrika*, vol. 48, 1961, pp. 3–13, with translation of Euler (1768). Reprint: E.S. Pearson & Kendall (1970, pp. 155–172).

– (1982; 1987), *Werke*, Bde 2–3. Basel.

Bertrand J. (1888), *Calcul des probabilités*. 2nd ed., 1907. Reprints: New York, 1970, 1972. Second edition practically coincides with the first one.

Bessel F.W. (1816), Untersuchungen über die Bahn des Olbersschen Kometen. *Abh. Preuss. Akad. Berlin*, math. Kl. 1812–1813, pp. 119–160. Bessel (1876) only contains a passage from this contribution.

– (1823), Persönliche Gleichung bei Durchgangsbeobachtungen. In Bessel (1876, Bd. 3, pp. 300–304).

– (1838), Untersuchung über die Wahrscheinlichkeit der Beobachtungsfehler. In Bessel (1876, Bd. 2, pp. 372–391).

– (1876), *Abhandlungen*, Bde 1–3. Leipzig.

⁸ In 1825, Gauss had a quarrel with Bessel but no details are known (Sheynin 2001, p. 168). Even in 1817 Olbers (Erman 1852, Bd. 2, p. 69) regretted that the relations between Bessel and Gauss were bad. In 1812, in a letter to Olbers, Bessel (*Ibidem*, Bd. 1, p. 345) had called Gauss “nevertheless” the inventor of the ML_{Sq}, but in 1844, in a letter to Humboldt (Sheynin 2001, p. 168), he stressed Legendre’s priority.

- Bienaymé I.J. (1852), Sur la probabilité des erreurs d'après la méthode des moindres carrés. *J. Math. Pures Appl.*, sér. 1, t. 17, pp. 33–78. Also: *Mém. pres. Acad. Sci. Inst. France*, sér. 2, t. 15, 1858, pp. 615–663.
- (1853), Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés. *C. r. Acad. Sci. Paris*, t. 37, pp. 309–324. Also: *J. Math. Pures Appl.*, sér. 2, t. 12, 1867, pp. 158–176.
- Biermann K.-R. (1966), Über die Beziehungen zwischen Gauss und Bessel. *Mitt. Gauss-Ges. Göttingen*, Bd. 3, pp. 7–20.
- Bomford G. (1971), *Geodesy*. Oxford. First two eds: 1952, 1962.
- Bond G.P. (1857), On the use of equivalent factors in the method of least squares. *Mem. Amer. Acad. Arts Sciences*, new ser., vol. 6, pt. 1, pp. 179–212.
- Brendel M. (1924), Über die astronomische Arbeiten von Gauss. In Gauss, *Werke*, Bd. 11, Tl. 2, Abt. 3. Separate paging.
- Coolidge J.L. (1926), Adrain and the beginnings of American mathematics. *Amer. Math. Monthly*, vol. 33, no. 2, pp. 61–76.
- Cramér H. (1946), *Mathematical Methods of Statistics*. Princeton. 13th printing, 1974.
- Czuber E. (1891), Zur Kritik einer Gauss'schen Formel. *Monatshefte Math. Phys.*, Bd. 2, pp. 459–464.
- David H.A. (1957), Some notes on the statistical papers of Helmert. *Bull. Stat. Soc. New South Wales*, vol. 19, pp. 25–28. First publ. Ibidem, 1954.
- Dutka J. (1990), R. Adrain and the method of least squares. *AHES*, vol. 41, pp. 171–184.
- (1996), On Gauss' priority in the discovery of the method of least squares. *AHES*, vol. 49, pp. 355–370.
- Eddington A.S. (1933), Notes on the method of least squares. *Proc. Phys. Soc.*, vol. 45, pp. 271–287.
- Eisenhart C. (1978), Gauss. In Kruskal & Tanur (1978, vol. 1, pp. 378–386).
- Encke J.F. (1851), Über die Bestimmung der elliptischen Elemente bei Planetenbahnen. *Abh. Kgl. Akad. Wiss. Berlin* for 1849, pp. 1–68 of second paging.
- Euler L. (1778, in Latin). [Commentary to D. Bernoulli (1778).] English transl.: 1961, and its reprint 1970, are together with the transl. and reprint of D. Bernoulli.
- Fisher R.A. (1920), Mathematical examination of the methods of determining the accuracy of an observation. *Monthly Notices Roy. Astron. Soc.*, vol. 80, pp. 758–770.
- Galle E. (1924), Über die geodätischen Arbeiten von Gauss. In Gauss, *Werke*, Bd. 11/2, Abt. 1. Separate paging.
- Gauss, C. F. (1809a), Preliminary author's report about Gauss (1809b). In Gauss (1887, pp. 204–205).
- (1809b, in Latin), *Theorie der Bewegung*, Book 2, Section 3. German transl. Ibidem, pp. 92–117.
- (1811, in Latin), Aus der Untersuchung über die elliptischen Elemente der Pallas. Ibidem, pp. 118–128.
- (1816), Bestimmung der Genauigkeit der Beobachtungen. Ibidem, pp. 129–138.
- (1821), Preliminary author's report about Gauss (1823b, pt. 1). Ibidem, pp. 190–195.
- (1823a), Preliminary author's report about Gauss (1823b, pt. 2). Ibidem, pp. 195–199.
- (1823b, in Latin), Theorie der den kleinsten Fehlern unterworfenen Combination der Beobachtungen, pts 1–2. Ibidem, pp. 1–53.
- (1823c), Anwendung der Wahrscheinlichkeitsrechnung auf eine Aufgabe der praktischen Geometrie. *Werke*, Bd. 9. Göttingen–Leipzig, 1903, pp. 231–237.
- (1826), Preliminary author's report about Gauss (1828). Ibidem, pp. 200–204.
- (1828, in Latin), Supplement to Gauss (1823b). German transl.: Ibidem, pp. 54–91.
- (1845; Nachlass), Anwendung der Wahrscheinlichkeitsrechnung auf die Bestimmung der Bilanz für Witwenkassen. *Werke*, Bd. 4, 1873, pp. 119–183.

- (1855), *Méthode des moindres carrés*. Paris.
- (1863–1930), *Werke*, Bde 1–12. Göttingen a.o. Reprint: Hildesheim, 1973–1981.
- (1887), *Abhandlungen zur Methode der kleinsten Quadrate*. Hrsg. A. Börsch & P. Simon. Latest ed.: Vaduz, 1998.
- Gerardy T. (1977), Die Anfänge von Gauss' geodätische Tätigkeit. *Z. f. Vermessungswesen*, Bd. 102, pp. 1–20.
- Hald A. (1952), *Statistical Theory with Engineering Applications*. New York, 1960.
- (1998), *History of Mathematical Statistics from 1750 to 1930*. New York.
- Helmert F.R. (1868), Studien über rationelle Vermessungen im Gebiete der höhern Geodäsie. *Z. Math. Phys.*, Bd. 13, pp. 73–120, 163–186.
- (1872), *Ausgleichsrechnung nach der Methode der kleinsten Quadrate*. Leipzig. Subsequent eds: 1907 and 1924.
- (1875a), Über die Berechnung des wahrscheinlichen Fehlers aus einer endlichen Anzahl wahrer Beobachtungsfehler. *Z. Math. Phys.*, Bd. 20, pp. 300–303.
- (1875b), Über die Formeln für den Durchschnittsfehler. *Astron. Nachr.*, Bd. 85, pp. 353–366.
- (1876a), Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichen Beobachtungsfehlers directer Beobachtungen. *Ibidem*, Bd. 88, pp. 113–132.
- (1876b), Über die Wahrscheinlichkeit der Potenzsummen der Beobachtungsfehler. *Z. Math. Phys.*, Bd. 21, pp. 192–218.
- (1886), *Lotabweichungen*, Heft 1. Berlin.
- (1905), Über die Genauigkeit der Kriterien des Zufalls bei Beobachtungsreihen. *Sitz. Ber. Kgl. Preuss. Akad. Wiss.*, Phys.-Math. Cl., Halbbd. 1, pp. 594–612. Also in Helmer (1993, pp. 189–208).
- (1993), *Akademie-Vorträge*. Frankfurt am Main. Reprints of author's reports.
- Herschel J. (1850), [Review of] Quételet (1846). *Edinb. Rev., or, Critical J.*, vol. 92, No. 185, pp. 1–57. Publ. anonymously.
- Heyde C.C., Seneta E. Editors (2001), *Statisticians of the Centuries*, Springer, New York.
- Hogan E.R. (1977), R. Adrain: American mathematician. *Hist. Math.*, vol. 4, pp. 157–172.
- Idelson N.I. (1947), *Sposob naimenshikh kvadratov i teoria matematicheskoi obrabotki nabludeny* (Method of Least Squares and the Theory of Math. Treatment of Observations). Moscow.
- Kac M. (1939), On a characterization of the normal distribution. In author's book *Probability, Number Theory and Statistical Physics*. Cambridge (Mass.), 1979, pp. 77–79.
- Kapteyn J.C. (1912), Definition of the correlation-coefficient. *Monthly Notices Roy. Astron. Soc.*, vol. 72, pp. 518–525.
- Kendall M.G. (Sir Maurice) (1971), The work of Ernst Abbe. *Biometrika*, vol. 58, pp. 369–373. Reprint: M.G. Kendall & Plackett (1977, pp. 331–335).
- Kendall M.G., Plackett R.L., Editors (1977), *Studies in the History of Statistics and Probability*, vol. 2. London. Collected reprints of papers.
- Kolmogorov A.N. (1931, in Russian), The method of median in the theory of errors, *Matematicheskyy Sbornik* vol. 38, No. 3-4, pp. 47-49, English translation in 1992, pp. 115-117
- (1946), Justification of the method of least squares. *UMN*, vol. 1, pp. 57–71. English transl. in Kolmogorov (1992, pp. 285–302. (R)
- (1985–1986, in Russian), *Selected Works*, vols. 1–2. Dordrecht, 1991–1992.
- Kolmogorov A.N., Petrov A.A., Smirnov Yu.M. (1947), A formula of Gauss in the method of least squares. *Izvestia Akad. Nauk SSSR*, ser. math., vol. 11, pp. 561–566. English transl. in Kolmogorov 1992, pp. 303–308). (R)
- Kruskal W. (1946), Helmert's distribution. *Amer. Math. Monthly*, vol. 53, pp. 435–438.
- (1978), Formulas, numbers, words: Statistics in prose. In: *New Directions for Methodology of Social and Behavioral Science*. San Francisco, 1981. Editor, D. Fiske, pp. 93–102.

- Kruskal W., Tanur J.M., Editors (1978), *International Encyclopedia of Statistics*, vols 1–2. New York.
- Krylov A.N. (1950), *Lektsii o Priblizhennykh Vychisleniakh* (Lectures on Approximate Calculations). Moscow.
- Kuzmin R. (1928), Sur un problème de Gauss. *Atti Congr. Intern. Matem. Bologna 1928*. Bologna, 1932, t. 6, pp. 83–89. (R)
- Laplace P.S. (1798–1825), *Traité de mécanique céleste*. OC, t. 1–5. Paris, 1878–1882. English transl. of vols 1–4 by N. Bowditch: *Celestial Mechanics* (1829–1839). New York, 1966.
- (1812), *Théorie analytique des probabilités*. OC, t. 7, No. 1–2. Paris, 1886. Consists of two parts, an Introduction (1814) and supplements, see below. Theory of probability proper is treated in pt. 2.
 - (1816), *Théor. anal. prob., Supplément 1*. OC, t. 7, No. 2, pp. 497–530.
 - (1818), *Théor. anal. prob., Supplément 2*. Ibidem, pp. 531–580.
 - (ca. 1819), *Théor. anal. prob., Supplément 3*. Ibidem, pp. 581–616.
 - (1827), Sur le flux et reflux lunaire atmosphérique. OC, t. 13. Paris, 1904, pp. 342–358.
- Legendre A.M. (1805), *Nouvelles méthodes pour la détermination des orbites des comètes*. Paris.
- (1814), Méthode des moindres carrés. *Mém. Acad. Sci. Paris, Cl. Sci., Math. et Phys.*, t. 11, pt. 2, année 1810, pp. 149–154.
 - (1820), *Nouvelles méthodes* [...]. Suppl. 2. Paris.
- Lehmann E.L. (1959), *Testing Statistical Hypotheses*. [New York, 1997].
- Linnik Yu.V. (1952), Remarks concerning the classical derivation of the Maxwell law. *Doklady Akad. Nauk SSSR*, vol. 85, pp. 1251–1254. (R)
- (1958), *Method of Least Squares and Principles of the Theory of Observations*. Oxford, 1998. Reference in text to English edition of 1961. (R)
- Lipschitz R. (1890), Sur la combinaison des observations. *C. r. Acad. Roy. Sci. Paris*, t. 111, pp. 163–166.
- Maennchen Ph. (1918), Gauss als Zahlenrechner. In Gauss (1863–1930, Bd. 10/2, Abt. 6, 1930, separate paging).
- Maltsev A.I. (1947), Remark on the work of Kolmogorov et al (1947). *Izvestia Akad. Nauk SSSR, Ser. math.*, vol. 11, pp. 567–578. (R)
- Markov A.A. (1899), The law of large numbers and the method of least squares. In Markov (1951, pp. 231–251). [1, pp. 130–142].
- Marsden B.G. (1995), 18- and 19-th century developments in the theory and practice of orbit determination. In Taton & Wilson (1995, pp. 181–190).
- Maxwell J.C. (1860), Illustrations of the dynamical theory of gases. Ibidem, pp. 377–410.
- May K.O. (1972), Gauss. *Dict. Scient. Biogr.*, vol. 5, pp. 298–315.
- Merian P. (1830), D. Huber. In *Verh. Allg. Schweiz. Ges. ges. Naturwiss. in ihrer 16. Jahresversammlung. 1830*. St. Gallen, pp. 145–152.
- Newcomb S. (1897), A new determination of the precessional constant. *Astron. Papers*, vol. 8, pp. 1–76.
- Pearson E.S., Kendall M.G., Editors (1970), *Studies in the History of Statistics and Probability* [vol. 1]. London. Collection of reprints.
- Pearson K. (1894), On the dissection of asymmetrical frequency curves. *Phil. Trans. Roy. Soc.*, vol. A185, pp. 71–110.
- (1920), Notes on the history of correlation. *Biometrika*, vol. 55, pp. 459–467. Reprinted in E.S. Pearson & Kendall (1970, pp. 185–205).
 - (1978), *History of Statistics in the 17th and 18th Centuries against the Changing Background of Intellectual, Scientific and Religious Thought*. Lectures 1921–1933. Editor E.S. Pearson. London.

- Peters [C.A.F.] (1856), Über die Bestimmung des wahrscheinlichen Fehlers einer Beobachtung. *Astron. Nachr.*, Bd. 44, pp. 29–32.
- Sakatov P.S. (1950), *Lehrbuch der höheren Geodäsie*. Berlin, 1957. Later Russian editions, 1953, 1964. (R)
- Schilling C., Editor (1900–1909), *Olbers. Sein Leben und sein Werk*, Bd. 2, Abt. 1–2 [Briefwechsel zwischen Gauss und Olbers.]. Reprinted as Gauss, *Werke*, Ergänzungsreihe, Bd. 4. Hildesheim, 1976.
- Schumann R. (1917), F. R. Helmert. *Österr. Z. f. Vermessungswesen*, Jg. 15, No. 718, pp. 97–100.
- Seal H.L. (1967), The historical development of the Gauss linear model. *Biometrika*, vol. 54, pp. 1–24. Reprinted in E.S. Pearson & Kendall (1970, pp. 207–230).
- Sheynin O.B. (1965), On the work of Adrain in the theory of errors. *IMI*, vol. 16, pp. 325–336.
- (1979), Gauss and the theory of errors. *AHES*, vol. 20, pp. 21–72.
 - (1984), On the history of the statistical method in astronomy. *AHES*, vol. 29, pp. 151–199.
 - (1994), Gauss and geodetic observations. *AHES*, vol. 46, pp. 253–283.
 - (1995a), Helmert’s work in the theory of errors. *AHES*, vol. 49, pp. 73–104.
 - (1995b), Density curves in the theory of errors. *Ibidem*, pp. 163–196.
 - (1999a), Discovery of the principle of least squares. *Hist. Scient.*, vol. 8, pp. 249–264.
 - (1999b), Gauss and the method of least squares. *JNÖS*, Bd. 219, pp. 458–467.
 - (2000), Bessel: some remarks on his work. *Hist. Scient.*, vol. 10, pp. 77–83.
 - (2001a), Gauss. In Heyde & Seneta (2001, pp. 119–122).
 - (2001b), Gauss, Bessel and the adjustment of triangulation. *Hist. Scient.*, vol. 11, pp. 168–175.
 - (2007), Euler’s work in probability and statistics. In *Euler Reconsidered. Tercentenary Essays*. Ed., R. Baker. Heber City, Uta, pp. 281–316.
- Spieß W. (1939), Kann man für D. Huber Ansprüche als Erfinder der “Methode der kleinste Quadrate” geltend machen? *Schweiz. Z. Vermessungswesen u. Kulturtechnik*, Bd. 37, pp. 11–17, 21–23.
- Stigler S.M. Editor (1980), *American Contributions to Mathematical Statistics in the 19th Century*, vols 1–2. New York. Reprints of papers of early American authors. No single paging.
- (1986), *History of Statistics*. Cambridge (Mass.).
 - (1999), *Statistics on the Table*. Cambridge (Mass.).
- Subbotin M.F. (1956), Astronomical and geodetic work of Gauss. In *Gauss. Memorial volume*. Moscow, pp. 243–310. (R)
- Thomson W., Tait P.G. (1867), *Treatise on Natural Philosophy*, vol. 1. Oxford. Later edition: New York, 2002. References in text to edition of 1867.
- Tsinging V.Ya. (1862), *Sposob naimenshikh kvadratov* (Method of Least Squares). Dissertation. Moscow.
- Tutubalin V.N. (1973), *Statisticheskaja obrabotka riadov nabliudenii* (Statistical Treatment of Series of Observations). Moscow.
- Whittaker E.T., Robinson G. (1949), *Calculus of Observations*. London. First edition 1924.
- Zach F.X. von (1813), Sur le degré du méridien. *Mém. Acad. Imp. Sci., Littérature, Beaux-Arts Turin* pour 1811–1812. *Sci. math. et phys.*, pp. 81–216.
- Zoch R.T. (1935), On the postulate of the arithmetic mean. *Annals Math. Stat.*, vol. 6, pp. 171–182.

C.F. GAUSS I METODA NAJMNIEJSZYCH KWADRATÓW

Nr 12 (18)

Streszczenie: Metodę najmniejszych kwadratów wprowadził Gauss, Helmert w pełni zakończył jej rozwój, Bessel zaś dokonał ważnych odkryć w astronomii i geodezji, był w tym jednak wyjątkowo niedbały. Gaussowski warunek o najmniejszej wariancji doprowadził do efektywnych estymatorów poszukiwanych nieznanymi wielkościami. W przypadku gdy błędy obserwacji mają rozkład normalny, estymatory są łącznie efektywne. Pamiętniki Gaussa z 1823 r. prościej prowadzą do zasady najmniejszych kwadratów, niż się powszechnie sądziło.

Słowa kluczowe: zasada i metoda najmniejszych kwadratów, wariancja próbkowa, wyrównywanie triangulacji, odchylenia od normalności.