Swan Song

I am 90 +. For a long time now, I am mostly translating papers from English to Russian (some, from Russian to English or even from German to Russian). Apart from possible short notes, the MS below is apparently my last original contribution.

Statistics

1. Early History

1.1. University Statistics. In the 1660s, Hermann Conring originated a new discipline, the Staatswissenschaft, or university statistics, and by the beginning of the eighteenth century it was taught all over Germany (Lazarsfeld 1961, p. 291). He modestly named Aristotle, Strabo and Ptolemy as the coauthors of the new discipline (Fedorovich 1894, p. 17).

Then, in mid-18th century Achenwall created the Göttingen school of Staatswissenschaft which described the climate, geographical situation, political structure and economics of separate states and estimated their population by issuing from data on births and mortality but did not study relations between quantitative variables. Wordy descriptions rather than numbers lay at the heart of the works of the Göttingen school, but Achenwall advised state measures fostering the multiplication of the population and recommended censuses without which (1763/1779, p. 187) a probable estimate of the population could be still got, see above. He also appropriately defined *the so-called statistics* as the Staatswissenschaft of separate states (Achenwall 1749, p. 1) and (1752/1756, Intro.) left an indirect definition of statistics:

In any case, statistics is not a subject that can be understood at once by an empty pate. It belongs to a well digested philosophy, it demands a thorough knowledge of European state and natural history taken together with a multitude of concepts and principles, and an ability to comprehend fairly well very different articles of the constitutions of present-day kingdoms.

On Achenwall see Schiefer (1916). It is appropriate to mention that in a letter of 1742 Daniel Bernoulli (Fuss 1843/1968, t. 2, p. 496) stated that *mathematics can also be rightfully applied in politics*. Citing Maupertuis' approval, he continued: *An entirely new science will emerge if only as many observations will be made in politics as in physics*. But did he understand politics just as Achenwall did later? Or, as Laplace (1814/1995, p. 62), who urged that the method based on observation and calculus should be applied to the political and moral sciences?

Achenwall's student Schlözer (1804, p. 86) figuratively stated that *History is statistics flowing, and statistics is history standing still.* Obodovsky (1839, p. 48) suggested a similar maxim: Statistics is to history as painting is to poetry. For those keeping to Staatswissenschaft Schlözer's pithy saying became the definition of statistics which was thus not compelled to study causal connections in society or discuss possible consequences of innovations.

Staatswissenschaft still exists, at least in Germany, in a new form. It applies quantitative data and studies causes and effects. It is the application of the statistical method to the life of a state or a region.

Knies (1850, p. 24) and John (1883, p. 670) quoted unnamed German authors who had believed, in 1806 and 1807, that the issues of statistics ought to be the national spirit, love of freedom, the talent and the characteristics of the great and ordinary people of a given state. This critic had to do with the limitations of mathematics in general. Here, however, is an ancient example of uniting description with approximate numbers:

Moses (Numbers 13: 17 – 20), who sent out spies to the land of Canaan, wished to find out *Whether the people who dwell in it are strong or weak*, *whether they are few or many*, – wished to know both numbers (roughly) and moral strength.

Tabular statistics which had originated with Anchersen (1741) could have served as an intermediate link between words and numbers (between Staatswissenschaft and political arithmetic, see below), but Achenwall (1752, Intro.) had experienced a public attack against the first edition of that book (published in 1749 under a previous title) by Anchersen. Tabular statisticians continued to be scorned, they were called *Tabellenfabrikanten* and *Tabellenknechte* (slaves of tables) (Knies 1850, p. 23). In 1734, I. K. Kirilov (Ploshko and Eliseeva 1990, pp. 65 – 66) compiled a tabular description of Russia, but it was only published in 1831.

In the beginning of the 1680s Leibniz compiled several manuscripts on political arithmetic and Staatswissenschaft which were only published in 1866. Now, they are available in his collected writings on insurance and finance mathematics (2000). In one of those manuscripts he (1680 – 1683/2000, pp. 442 and 443) adopted unfounded premises about population statistics including a simply fantastic statement: the birth rate can be nine or ten times higher than it is.

In his manuscripts devoted to Staatswissenschaft, Leibniz had recommended the compilation of state tables containing information useful for the state and the comparison of those of them which pertained to different states or times; the compilation of medical sourcebooks of observations made by physicians, of their recommendations and aphorisms; and the establishment of sanitary commissions with unimaginably wide tasks. He mentioned inspection of shops and bakeries, registration of the changes in the weather, fruit and vegetable yields, prices of foodstuffs, magnetic observations and, the main goal, recording of diseases and accidents affecting humans and cattle.

Leibniz (1682) also compiled a list of 56 questions (actually, of 58 since he made two mistakes in numbering them). He left them in an extremely raw and disordered state and a few are even incomprehensible. Their main topics were population statistics in a wide sense; money circulation; cost of living; morbidity. Incidentally, for some strange reason population statistics at least up to the 20th century had largely shunned medical problems. Graunt was a remarkable exception and Poisson (§ 5) treated them in his lectures.

1.2. Political Arithmetic. Statistics, in its modern sense, owed its origin to political arithmetic founded by Petty and Graunt. One of its main problems belonged with demography.

They studied population, economics, and commerce and discussed the appropriate causes and connections by means of elementary stochastic considerations. Petty called the new discipline *political arithmetic* and its aims were to study from a socio-economic point of view states and separate cities (or regions) by means of (rather unreliable) statistical data on population, industry, agriculture, commerce etc. Petty (1690/1899, p. 244) plainly formulated his denial of *comparative and superlative Words* and attempted to express himself *in Terms of Number, Weight, or Measure …*; Graunt undoubtedly did, if not said the same.

Petty (1927, vol. 1, pp. 171 – 172) even proposed to establish a *register generall of people, plantations & trade of England*, to collect the accounts of all the *Births, Mariages, Burialls* […] *of the Herths, and Houses* […] *as also of the People, by their Age, Sex, Trade, Titles, and Office*. The scope of that Register was to be wider than that of *our existing Register office* (Greenwood 1941 – 1943/1970, p. 61).

At least 30 Petty's manuscripts (1927) pertained to political arithmetic. This source (pp. 39 – 40) shows him as a philosopher of science congenial in some respects with Leibniz:

What is a common measure of Time, Space, Weight, & motion? What number of Elementall sounds or letters, will […] *make a speech or language? How to give names to names, and how to adde and subtract sensata, & to ballance the weight and power of words; which is Logick & reason*.

Graunt (1662) studied the weekly bills of mortality in London which began to appear in the 16th century and had been regularly published since the beginning of the 17th century. His contribution had been (but is apparently not anymore) attributed to Petty who perhaps qualifies as co-author. For my part, I quote his *Discourse* (1674): *I have also* (*like the author of those <u>Observations</u>* [like Graunt!]) <u>Dedicated this Discourse ..</u>.

Graunt used the fragmentary statistical data to estimate the population of London and England as well as the influence of various diseases on mortality and he attempted to allow for systematic corruptions of the data. Thus, he reasonably supposed that the number of deaths from syphilis was essentially understated out of ethical considerations. His main merit consisted in that he attempted to find definite regularities in the movement of the population. Thus, he established that both sexes were approximately equally numerous (which contradicted the then established views) and that out of 27 newborn babies about 14 were boys. When dealing with large numbers, Graunt did not doubt that his conclusions reflected objective reality which might be seen as a fact belonging to the prehistory of the law of large numbers (LLN). The ratio 14:13 was, in his opinion, an estimate of the ratio of the respective probabilities.

Nevertheless, he had uncritically made conclusions based on a small number of observations as well and thought that the population increased in an arithmetical progression, since replaced by the geometrical progression definitively introduced by Süssmilch and Euler (§ 1.3).

In spite of the meagre and sometimes wrong information, Graunt was able to compile the first life table (common for both sexes). He somehow calculated the relative number of people dying within the first six years and within each next decade up to age 86. According to his table, only one person out of a hundred survived until that age. The very invention of the mortality table was the main point here. The indicated causes of death were also incomplete and doubtful, but Graunt formulated some important conclusions as well (although not without serious errors). His general methodological (but not factual) mistake consisted in that he assumed, without due justification, that statistical ratios during usual years (for example, the per cent of yearly deaths) were stable. Graunt had influenced later scholars (Huygens, letter of 1662/1888 – 1950, 1891, p. 149; Hald 1990, p. 86):

1. *Grant's* [!] *discourse really deserves to be considered and I like it very much. He reasons sensibly and clearly and I admire how he was able to elicit all his conclusions from these simple observations which formerly seemed useless*.

2. *Graunt reduced the data from several great confused Volumes into a few perspicuous Tables and analysed them in a few succinct Paragraphs which is exactly the aim of statistics.*

1.3. Population Statistics. I discuss medical and juridical statistics separately (§§ 2.2 and 2.3), but I emphasize that those fields are fundamentally important for population statistics.

Halley (1693), a versatile scholar and an astronomer in the first place, compiled the next life table. He made use of statistical data collected in Breslau, a city with a closed population. Halley applied his table for elementary stochastic calculations and thus laid a mathematical foundation of actuarial science. He was also able to find out the general relative population of the city. Thus, for each thousand infants aged less than a year, there remained 855 children from one to two years of age, …, and, finally, 107 persons aged 84 – 100. After summing up all these numbers, Halley obtained 34 thousand (exactly) so that the ratio of the population to the newborn babies occurred to be 34. Until 1750 his table remained the best one (K. Pearson 1978, p. 206).

The yearly rate of mortality in Breslau was 1/30, the same as in London, and yet Halley considered that city as a statistical standard. If such a notion is appropriate, standards of several levels ought to be introduced. Again, Halley thought that the irregularities in his data will rectify themselves, were the number of years [of observation] much more considerable. Such irregularities could have been produced by systematic influences, but Halley's opinion shows the apparently wide-spread belief in an embryo of the LLN.

Sofonea (1957, p. 31*) called Halley's contribution the beginning of the entire development of modern methods of life insurance, and Hald (1990, p. 141) stated that it became of great importance to actuarial science. Drawing on Halley, De Moivre (1725) introduced the continuous uniform law of mortality for ages beginning at 12 years.

In 1701 Halley (Chapman 1941, p. 5) compiled a chart of Northern Atlantic showing the lines of equal magnetic declinations so that he (and of course Graunt) might be called the founders of exploratory data analysis.

It might be thought that statistics and statistical method are equivalent notions (see however § 9), but it is normal to apply the former term when studying population and to use the latter in all other instances and especially when applying statistics to natural sciences. Nevertheless, there also exist such expressions as *medical* and *stellar statistics*, and theory of errors.

Three stages may be distinguished in the history of the statistical method. At first, conclusions were being based on (statistically) noticed qualitative regularities, a practice which conformed to the qualitative essence of ancient science. Here, for example, is the statement of the Roman scholar Celsus (1935, p.19):

Careful men noted what generally answered the better, and then began to prescribe the same for their patients. Thus sprang up the Art of medicine.

The second stage (Tycho in astronomy, Graunt in demography and medical statistics) was distinguished by the availability of statistical data. Scientists had then been arriving at important conclusions either by means of simple stochastic ideas and methods or even directly, as before. During the present stage, which dates back to the end of the 19th century, inferences are being checked by quantitative stochastic rules.

In the 18th century, statisticians had been attempting to bring into conformity the speedy increase in population with the Biblical command (Genesis 1:28), *Be fruitful and multiply and fill the earth and subdue it*, and K. Pearson (1978, p. 337) severely criticized them:

Instead of trying, in the language of Florence Nightingale, to interpret the thought of God from statistical data, [they] *turn the problem around and twist their data to suit what they themselves consider the will of the Creator.*

And, on the same page, again about those statisticians who paved the way for the Malthusians if not Malthus himself:

While the Creator would not approve of starvation for thinning humanity, He would have no objection to plague or war.

The most renown statistician of the second half of the 18th century was Süssmilch although Pearson (p. 347) called Struyck a more influential forerunner in the field of vital statistics. Süssmilch (1741) adhered to the tradition of political arithmetic. He collected data on the movement of population and attempted to reveal pertinent divine providence but he treated his materials loosely. Thus, when taking the mean of the data pertaining to towns and rural districts, he tacitly assumed that their populations were equally numerous; in his studies of mortality, he had not attempted to allow for the differences in the age structure of the populations of the various regions etc.

Nevertheless, his works paved the way for Quetelet; in particular, he studied issues which later came under the province of moral statistics (e.g., illegitimate births, crime, suicides) and his tables of mortality had been in use even in the beginning of the 19th century, see Birg

(1986) and Pfanzagl & Sheynin (1997). After A. M. Guerry and Quetelet the domain of moral statistics essentially broadened and includes now, for example, philanthropy and professional and geographical mobility of the population.

Like Graunt, Süssmilch discussed pertinent causes and offered conclusions. Thus, he (1758) thought of examining the dependence of mortality on climate and geographical position and he knew that poverty and ignorance were conducive to the spread of epidemics.

Süssmilch's main contribution, the *Göttliche Ordnung*, marked the origin of demography. Its second edition of 1761 − 1762 included a chapter *On the rate of increase and the period of doubling* [of the population]; it was written jointly with Euler and served as the basis of one of Euler's memoirs (Euler 1767). Süssmilch thought that, since multiplication of mankind was a divine commandment, rulers must take care of their subjects. He condemned wars and luxury and indicated that the welfare of the poor was to the advantage of both the state, and the rich. His pertinent appeals brought him into continual strife with municipal (Berlin) authorities and ministers of the state (Prussia). He would have likely agreed with a much later author (Budd 1849, p. 27) who discussed cholera epidemics:

By reason of our common humanity, we are all the more nearly related here than we are apt to think. […] *And he that was never yet connected with his poorer neighbour by deeds of Charity or Love, may one day find, when it is too late, that he is connected with him by a bond which may bring them both, at once, to a common grave*.

Süssmilch's collaboration with Euler and frequent references to him in his book certainly mean that Euler had shared his general social views. Malthus (1798) picked up one of the conclusions in the *Göttliche Ordnung*, viz., that the population increased in a geometric progression (still a more or less received statement). Euler compiled three tables showing the increase of population during 900 years beginning with Adam and Eve. His third table based on arbitrary restrictions meant that each 24 years the number of living increased approximately threefold. Gumbel (1917) proved that the numbers of births, deaths and of the living in that table were approaching a geometric progression and noted that several authors since 1600 had proposed that progression as the appropriate law.

Note, however, that it was Gregory King (1648 – 1712) who first discussed the doubling of population (K. Pearson 1978, p. 109).

Euler left no serious contributions to the theory of probability, but he published a few elegant and methodically important memoirs on population statistics. He did not introduce any stochastic laws, but the concept of increase in population is due to him, and his reasoning was elegant and methodically interesting, in particular for life insurance (Paevsky 1935).

Lambert published a methodical study in population statistics (1772). Without due justification he proposed there several laws of mortality (belonging to types IX and X of the Pearson curves). Then, he formulated the problem about the duration of marriages, studied children's mortality from smallpox and the number of children in families (§ 108). See Sheynin (1971b) and Daw (1980) who also appended a translation of the smallpox issue.

When considering the last-mentioned subject, Lambert started from data on 612 families having up to 14 children, and, once more without substantiation, somehow adjusted his materials. He arbitrarily increased the total number of children by one half likely attempting to allow for stillbirths and the death of children. Elsewhere he (§ 68) indicated that statistical inquiries should reveal irregularities.

Population statistics owed its later development to the general problem of isolating randomness from Divine design. Kepler and Newton achieved this aim with regard to inanimate nature, and scientists were quick to begin searching for the laws governing the movement of population (and attempting to fit them to the Biblical command). Moreover, De Moivre thought that exactly that problem constituted the main aim of his philosophy. He dedicated the first edition of his *Doctrine of Chances* (1718/1756, p. 329) to Newton, and here are a few pertinent lines. His wish was to work out

A Method of calculating the Effects of Chance […] *and thereby* [of] *fixing certain rules, for estimating how far some sort of Events may rather be owing to Design than Chance* […] [so as to learn] *from your Philosophy how to collect, by a just Calculation, the Evidences of exquisite Wisdom and Design, which appear in the Phenomena of Nature throughout the Universe.*

De Moivre thus believed that the (future) theory of probability should be applied in natural sciences, but he rigorously demonstrated his theorems. Studies of various distributions had not yet begun. Chance had been certainly separated from design in everyday life. Bühler (1886/1967, p. 267) described an appropriate (for us, unreasonable) example pertaining to the administration of justice in ancient India. Horrible trials with red-hot iron had been widespread.

1.3.1. The sex ratio at birth. The solution of this problem was not practically needed, but the subject itself attracted scientists and provided a possibility of applying mathematical methods.

1.3.1-1. Arbuthnot. He (1712) assembled the existing data on baptisms in London for 1629 – 1710, noted that during those 82 years more boys (m) were invariably born than girls (f) and declared that that fact was *not the Effect of Chance but Divine Providence, working for a good End.* Boys and men, as he added, were subject to greater dangers and their mortality was higher than that of the females. Even disregarding both that unsubstantiated statement and such [hardly exhibited] regularities as the *constant Proportion m:f and fix'd limits of the difference* ($m - f$), *the Value of Expectation* of a random occurrence of the observed inequality was less than $(1/2)^{82}$, he stated.

Arbuthnot could have concluded that the births of both sexes obeyed the binomial distribution, which, rather than the inequality $m > f$, manifested Divine design; and could have attempted to estimate its parameter. Then, baptisms were not identical with births. Graunt (1662, end of Chapt. 3) stated that during 1650 – 1660 less than half of the general [Christian] population had believed that baptism was necessary; Christians perhaps somehow differed from other people, London was perhaps an exception. Note however, that during the 18th century philosophers almost always understood randomness in the uniform sense.

One more point. Denote a year by *m* or *f* if more boys or girls were respectively born. Any combination of the *m*'s and *f*'s in a given order has the same probability (2^{-82} in Arbuthnot's case). However, if the order is of no consequence, then those probabilities will greatly differ. Indeed, in a throw of two dice the outcome "1 and 2" in any order is twice as probable as "1 and 1". It is this second case which Arbuthnot likely had in mind.

I note Laplace's inference (1776/1891, p. 152; 1814/1995, p. 9) in a similar case: a sensible word would hardly be composed by chance from separate letters. Poisson (1837a, p. 114) provided an equivalent example and made a similar conclusion. However, a definition of a random sequence (and especially of its finite variety) is still a subject of subtle investigations.

Freudenthal (1961, p. xi) called Arbuthnot the author of the first publication on mathematical statistics, see also Shoesmith (1987) and David & Edwards (2001, pp. 9 – 11).

1.3.1-2. Nicolaus Bernoulli. While discussing the same subject, he indirectly derived the normal distribution. Let the sex ratio be *m/f*, *n*, the total yearly number of births, and μ and (n – μ), the numbers of male and female births in a year. Denote

$$n/(m + f) = r,\ m/(m + f) = p,\ f/(m + f) = q,\ p + q = 1,$$

and let $s = 0(\sqrt{n})$. Then Bernoulli's formula (Montmort 1713/1980, pp. 388 – 394) can be presented as

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

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

It is not an integral theorem since *s* is restricted (see above) and neither is it a local theorem; for one thing, it lacks the factor $\sqrt{2/\pi}$.

The context of De Moivre's paper (1733) in which he proved the first version of the central limit theorem, CLT (a term introduced by Polya (1920)) shows that he intended it for studying that same problem, the sex ratio at birth.

1.3.1-3. While investigating the same problem, **Daniel Bernoulli** (1770 – 1771) first assumed that male and female births were equally probable. It followed that the probability that the former constituted a half of $2N$ births will be

$$P = \frac{1\cdot 3\cdot 5\cdot ...(2N-1)}{2\cdot 4\cdot 6\cdot ...\cdot 2N} = q(N).$$

He calculated this fraction not by the Wallis formula but by means of differential equations and obtained

$$q = \frac{1.12826}{\sqrt{4N+1}}.$$

Application of differential equations was Bernoulli's usual method in probability. Bernoulli also determined the probability of the birth of approximately *m* boys (see below):

$$P(m = N \pm \mu) = q\exp(-\mu^2/N) \text{ with } \mu \text{ of the order of } \sqrt{N}. \quad (1)$$

In the second part of his memoir Bernoulli assumed that the probabilities of the birth of both sexes were in the ratio of *a:b*. Equating the probabilities of *m* and $(m + 1)$ boys being born, again being given $2N$ births, he thus obtained the [expected] number of male births

$$\mathrm{E}m = M = \frac{2Na-b}{a+b} \approx \frac{2Na}{a+b}$$

which was of course evident. More interesting was Bernoulli's subsequent reasoning for determining the probability of an arbitrary *m* (for μ of the order of $\sqrt{N}$):

$$P(m = M + \mu + 1) - P(m = M + \mu) \equiv d\pi =$$

$$\pi - \pi\frac{a(2N-M-\mu)}{b(M+\mu+1)}d\mu, \quad -\frac{d\pi}{\pi} = \frac{\mu+1+\mu a/b}{m+\mu+1}d\mu.$$

The subsequent transformations included the expansion of $\ln[(M + 1 + \mu)/(M + 1)]$ into a power series. Bernoulli's answer was

$$P(m = M \pm \mu) = \pi = P(m = M)\exp[-\frac{(a+b)\mu^2}{2bM}],$$

hence (1). Note that Bernoulli had not applied the local De Moivre (– Laplace) theorem.

2. Statistical Startups, Not Yet Explored Topics, Difficulties

Graunt (1662) was not sure whether anyone except *the Sovereign and his chief Ministers* needed statistics, but since then the situation has essentially changed, and especially with the creation of the welfare state and government decision making. Great changes have occurred with regard to natural sciences as well. Mostly in the 19th century a number of new disciplines linked to statistics have originated: medical statistics (especially epidemiology), public hygiene (the forerunner of ecology), geography of plants, zoogeography, biometry, climatology, stellar statistics, and kinetic theory of gases. Many fundamental problems, such as the influence of solar activity on terrestrial phenomena have been studied statistically.

Just to illustrate the widest scope of statistics I mention two papers: Thornberg (1929) about the trade union movement (which showed an unusual aspect of the application of statistics in industry) and Thorp (1948) who described the use of statistics in foreign relations.

During the first five decades of the 19th century, statistical institutions and/or national statistical societies came into being in the main states of Europe and America. International statistical congresses aiming at unification of official statistical data had been held from 1851 onward, and in 1885 the still active International Statistical Institute was established instead.

Throughout the 19th century the importance of statistics had been considerably increasing. By the mid-19th century it became important to foresee how various transformations will influence society and Quetelet (§ 7.1) repeatedly stressed this point. Then, at the end of that century censuses of population, answering an ever widening range of questions, began to be carried out in various countries. However,

● Public opinion was not yet studied.

● Sampling had been considered doubtful. Cournot (1843) passed it over in silence and Laplace's sample determination of the population of France was largely forgotten. Quetelet opposed sampling. Much later Bortkiewicz (1904, p. 825) and Czuber (1921, p. 13) called sampling *conjectural calculation* although already the beginning of the century witnessed *legions* of new data (Lueder 1812, p. 9) and the tendency to amass sometimes useless or unreliable data revealed itself in various branches of natural sciences.

I adduce two barely known statements. In 1904, Newcomb had sent a letter to the Carnegie Institution urging it to establish an *institute or a bureau of exact sciences* for developing methods of dealing with the *great mass of existing observations* (*Method* 1905, p. 180). Neither he, nor Pearson (p. 184), one of the several scientists whom the Carnegie Institution asked to comment on Newcomb's proposal, mentioned sampling. Pearson argued that the situation was certainly bad and held that *at least 50 per cent of the observations made and the data collected are worthless*. Either the conditions necessary for testing a theory were not met or collectors or observers were *hopelessly ignorant* of the conditions required for accurate work. Owing to various difficulties, Newcomb's proposal was not adopted.

In 1915 or 1916, Chuprov mentioned the need to organize after the end of the world war, under the (Russian) Academy of Sciences, the studies of population and its productive forces (Sheynin 1990/2011, p. 130).

On the history of sampling, whose most active partisan was Kiaer, see You Poh Seng (1951) and Tassi (1988). Kapteyn (1906) initiated an international stratified sampling of the starry heaven.

● The development of the correlation theory began at the end of the 19th century, but even much later Kaufman (1922, p. 152) declared that *the so-called method of correlation adds nothing essential to the results of elementary analysis*.

● Variance began to be applied in statistics only after Lexis, but even later Bortkiewicz (1894 – 1896, Bd. 10, pp. 353 – 354) stated that the study of precision was a luxury, and that the statistical flair was much more important. This opinion had perhaps been caused by the presence of large systematic corruptions in the initial materials.

● Preliminary (or exploratory) data analysis (generally recognized only a few decades ago) was necessary, and should have been the beginning of the statistician's work.

● Statistical quality control had not been applied until the 1920s.

● Econometrics only originated in the 1930s (Frisch 1933, p. 1): the main object of the just established Econometric Society was to promote *unification of the theoretical-quantitative and the empirical-quantitative approach to economic problems* and to foster *constructive and rigorous thinking similar to that which has come to dominate in the natural sciences*.

Poincaré, in an undated letter kept in his *Dossier* at the Paris Academy of Sciences (Sheynin 2009, p. 117, No. 619) quite positively described the work of H. Laurent both in probability and actuarial science and noted his *Traité* [1902] on mathematical political economy and lectures *dans un cours libre* at the Sorbonne on the same subject. Poincaré called this discipline a *science nouvelle crée par Walras et ses disciples*. So are Walras and Laurent the forerunners of econometrics?

I can also mention Petty and Bortkiewicz. Petty's essays on political arithmetic were *econometric in its methodological framework, even from the modern point of view* (Strotz 1978, p. 188). And Bortkiewicz

Made the necessary modifications that rendered the Marxian scheme of surplus values and prices consistent. However, his dry presentation prevented the Marxists (*except for Klimpt* [1936]) *from accepting his method.* And he had made *the lonely effort to construct a Marxian econometrics* [without applying statistical data] (Gumbel 1978, pp. 25 and 26).

Strotz (p. 189) also argues that econometrics *is disappearing as a special branch of economics*.

In conformity with the situation in the Soviet Union (Sheynin 1998) econometrics had hardly existed there. At an economic conference in 1960 Kolmogorov (Birman 1960, p. 44) stated that

The main difficult but necessary aim is to express the desired optimal state of affairs in the national economy by a single indicator.

Indeed, the prices of commodities had only been administratively established.

I list now the difficulties, real and imaginary, of applying the theory of probability to statistics.

● The absence of *equally possible* cases whose existence is necessary for understanding the classical notion of probability. Statisticians repeatedly mentioned this cause.

● Disturbance of the constancy of the probability of the studied event and/or of the independence of trials. Before Lexis (1879) statisticians had only recognized the Bernoulli trials; and even much later, again Kaufman (1922, pp. 103 – 104), argued that the theory of probability was applicable only to these trials, and, for that matter, only in the presence of equally possible cases.

● The abstract nature of the (not yet axiomatized) theory of probability. The history of mathematics testifies that the more abstract it became, the wider had been the range of its applicability.

In the beginning of the 1820s in a letter to Quetelet, Fourier maintained that the *statistical sciences* will only progress insofar as they were supported by *mathematical theories* (Quetelet 1826, p. 177). Soon, however, Quetelet (1828, p. i) called the calculus of probability (not just mathematics) the *basis of observational sciences*, and later (1869, t. 1, p. 134) *a most reliable and most indispensable companion* of statistics. Bortkiewicz (1904) expressed similar ideas.

For most statisticians all these pronouncements remained alien (see also § 8.3.3). They had not expected any help from the theory of probability. Block (1878/1886, p. 134) thought that it was too abstract and should not be applied *too often*, and Knapp (1872, p. 115) called it difficult and hardly useful beyond the sphere of games of chance and insurance. In 1911, G. von Mayr declared that mathematical formulas were not needed in statistics and privately told Bortkiewicz that he was unable to bear mathematics (Bortkevich & Chuprov 2005, Letter 109 of 1911). Bortkiewicz (1904, p. 822) also mentioned Guerry (1864, p. XXXIII ff) as an opponent of the application of the theory of probability and therefore his opponent as well.

I have noted (§ 2.1) that in 1835 several scientists including Poisson had stressed the connection between statistics and probability. A bit earlier three scholars, again including Poisson (Libri-Carrucci et al 1834, p. 535), declared that *the most sublime problems of the* <u>*arithmétique sociale*</u> [see § 5] *can only be resolved with the help of the theory of probability.*

Nevertheless, statisticians never mentioned Daniel Bernoulli who published important statistical memoirs, almost forgot insurance, barely understood the treatment of observations, did not notice either Quetelet's mistakes or his inclinations to crime and to marriage (§ 7.1). After his death in 1874 they all the more turned away from probability.

Two circumstances worsened the situation. First, mathematicians often did not show how to apply their findings in practice. Poisson (1837a) is a good example; his student Gavarret (1840) simplified his formulas, but still insisted that conclusions should be based on a large number of observations which was often impossible (§ 2.1). Second, student-statisticians barely studied mathematics and, after graduation, did not trust it.

It is not amiss to mention here a pioneer attempt to create mathematical statistics (Wittstein 1867). He compared the situation in statistics with the *childhood* of astronomy and stressed that statistics (and especially population statistics) needed a Tycho and a Kepler to proceed from reliable observations to regularities. Specifically, he noted that statisticians did not understand the essence of probability theory and never estimated the precision of the results obtained.

Knies (1850 [p. 163]) was *strongly in favour of adopting the name* <u>*statistics*</u> *for* <u>*political arithmetic*</u> *called by him* <u>*mathematical statistics*</u> (John 1883, p. 677). Hardly proper, but the term *mathematical statistics* was apparently thus first pronounced.

● A few words about astronomy and meteorology. In astronomy asteroids were understood to form a statistical population: their orbital parameters were studied statistically (Newcomb). From the mid-18th century (William Herschel) statistical reasoning was also applied to studying the arrangement and (later) the movement of stars and Kapteyn (1906) initiated an international plan for a sampling study of the stellar universe. In meteorology, Humboldt (1817) used statistical data on air temperatures to construct isotherms on a world-wide scale and thus to isolate the Earth's main climatic belts (more precisely, to confirm quantitatively their existence, qualitatively suggested by ancient geographers) and originate climatology. The introduction of contour lines for representing statistical information (a brilliant example of exploratory data analysis) was due to Halley (§ 1.3).

In general, Humboldt (1845 – 1862, Bd. 1, pp. 18 and 72; Bd. 3, p. 288) conditioned the investigation of natural phenomena by examination of mean states. In the last-mentioned case he mentioned *the sole decisive method* [in natural sciences], *that of the mean numbers* which (1845, Bd. 1, p. 82) *show us the constancy in the changes*. In other words, he stressed the importance of statistical studies.

Lamarck, the most eminent biologist of his time, seriously occupied himself with physics, chemistry and meteorology. In meteorology, his merits had for a long time been ignored (Muncke 1837), but he is now remembered for his *pioneer work in the study of weather* (Shaw & Austin (1926/1942, p. 130). He repeatedly applied the term *météorologie statistique* (e.g., 1800 – 1811, t. 4, p. 1) whose aim (Ibidem, t. 11, p. 9 – 10) was the study of climate, or, as he (Ibidem, t. 4, pp. 153 – 154) maintained elsewhere, the study of the climate, of regularities in the changes of the weather and of the influence of various meteorological phenomena on animals, plants and soil.

He preferred a *reasoned* rather than an empirical meteorology (Ibidem, t. 5, p. 1) with its own theory, general principles and aphorisms (Ibidem, t. 3, p. 104). At the time, such an approach was impossible but the development of statistical physics in the 19th century somewhat changed the situation (Angström 1929, p. 229).

Buys Ballot (1850, p. 629) noted the appearance of the second stage in the development of meteorology, the study of the deviations of meteorological elements from their mean values or states. He could have mentioned a few other sciences as well (for example, astronomy and geodesy, − and statistics!).

2.1. Medical Statistics. Interestingly enough, the expression *medical probability* appeared not later than in the mid-18th century (Mendelsohn 1761, p. 204). At the end of that century Condorcet (1795/1988, p. 542) advocated collection of medical observations and Black (1788, p. 65) even compiled a possibly forgotten *Medical catalogue of all the principle diseases and casualties by which the Human Species are destroyed or annoyed* that reminded of Leibniz' thoughts. Descriptions belonging to other branches of natural sciences as well have actively been compiled (mostly later) and such work certainly demanded preliminary statistical efforts. Some authors mistakenly stated that their compilations ruled out the need for theories and, in addition, until the beginning of the 20th century the partisans of complete descriptions continued to deny sampling in statistics proper.

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

Public hygiene began statistically studying problems connected with the Industrial Revolution in England and, in particular, by the great infant mortality (Chadwick 1842/1965, p. 228). Also, witness Farr (ca. 1857/1885, p. 148): *Any deaths in a people exceeding 17 in a 1,000 annually are unnatural deaths*. Unnatural, but common!

Epidemiology was properly born when cholera epidemics had been ravaging Europe. Snow (1855) compared mortality from cholera for two groups of the population of London, whose drinking water was either purified or not, ascertained that purification decreased mortality by eight times, and thus discovered how did cholera epidemics spread. The need to combat the devastating visitations of cholera was of utmost importance.

No less important was the study of prevention of smallpox. The history of smallpox epidemics and inoculation, the communication of a mild form of smallpox from one person to another, is described in various sources (Condamine 1759, 1763, 1773; Karn 1931). In his first memoir, Condamine listed the objections against inoculation, both medical and religious.

Indeed, White (1896/1898) described the *warfare of science with theology* including, in vol. 2, pp. 55 – 59, examples of fierce opposition to inoculation (and, up to 1803, to vaccination of smallpox). Many thousands of Canadians perished in the mid-19th century only because, stating their religious belief, they had refused to be inoculated. White clearly distinguished between theology, the opposing force, and "practical" religion.

Karn stated at the very beginning of her article that

The method used in this paper for determining the influence of the death-rates from some particular diseases on the duration of life is based on suggestions which were made in the first place by D. Bernoulli.

Daniel Bernoulli (1766) justified inoculation. That procedure, however, spread infection, was therefore somewhat dangerous for the neighbourhood and prohibited for some time, first in England, then in France. Referring to statistical data, but not publishing it, Bernoulli specified the yearly rates of the occurrence of smallpox in those who have not had it before and of the corresponding mortality and assumed that the inoculation itself proved fatal in 0.5% of cases.

He formed the appropriate differential equation whose solution showed the relation between age in years and the number of people of that age and, in addition, of those who had not contacted smallpox. Also by means of a differential equation he derived a similar formula for a population undergoing inoculation, that is, for its 99.5% which safely endured it and were not anymore susceptible to the disease. It occurred that inoculation lengthened the mean duration of life by 3 years and 2 months and that it was therefore, in his opinion, extremely useful. The Jennerian vaccination, – *the inestimable discovery by Jenner, who has thereby become one of the greatest benefactors of mankind* (Laplace 1814/1995, p. 83), – was introduced at the end of the 18th century. Its magnificent success had not however ruled out statistical studies. Thus, Simon (1887, vol. 1, p. 230) formulated a question about the impermanence of protection against post-vaccinal smallpox and concluded that only comprehensive national statistics could have provided an answer.

D'Alembert (1761; 1768) criticized Daniel Bernoulli. Not everyone will agree, he argued, to lengthen his mean duration of life at the expense of even a low risk of dying at once of inoculation; then, moral considerations were also involved, as when inoculating children. D'Alembert concluded that statistical data on smallpox should be collected, additional studies made and that the families of those dying of inoculation should be indemnified or given memorial medals.

He also expressed his own thoughts, methodologically less evident but applicable to studies of even unpreventable diseases. Dietz & Heesterbeek (2002) described Bernoulli's and D'Alembert's investigations on the level of modern mathematical epidemiology and mentioned sources on the history of inoculation.

Seidel (1865; 1866), a German astronomer and mathematician, quantitatively estimated the dependence of the number of cases of typhoid fever on the level of subsoil water and precipitation but made no attempt to generalize his study, to introduce correlation.

Already in 1839 there appeared (an unconvincing) statistical study of the amputation of limbs. J. Y. Simpson (1847 – 1848/1871, p. 102) mistakenly attempted to obtain reliable results about that operation by issuing from materials pertaining to several English hospitals during 1794 – 1839.

Indeed, physicians learned that the new procedure, anaesthesia, could cause complications, and began to compare statistically the results of amputations made with and without using it.

Simpson (1869 – 1870/1871, title of contribution) also coined the term *Hospitalism* which is still in vogue. He compared mortality from amputations made in various hospitals and reasonably concluded, on the strength of its monotonous behaviour, that mortality increases with the number of beds; actually (p. 399), because of worsening of ventilation and decrease of air space per patient. Suchlike justification of conclusions was not restricted to medicine, cf. Quetelet's study of probabilities of conviction of defendants (§ 7.1).

In the mid-19th century Pirogov began to compare the merits of the conservative treatment of the wounded versus amputation. Later he (1864, p. 690) called his time *transitional*:

Statistics shook the sacred principles of the old school, whose views had prevailed during the first decades of this century, – and we ought to recognize it, – but it had not established its own principles.

Pirogov (1849, p. 6) reasonably believed that the application of statistics in surgery was in *complete agreement* with the latter because surgical diseases depended incomparably less on individual influences but he indicated that medical statistics was unreliable, that (1864/1865

– 1866, p. 20) a general impression based on sensible observation of cases was better. He (1879/1882, p. 40) singled out *extremely different circumstances* and stressed (1871, pp. 48 – 49) the importance of *efficient administration.* Pirogov participated in the Crimean war, in which Florence Nightingale, on the other side, showed her worth both as a medical nurse and a statistician. She would have approved of Pirogov's conclusion above.

Pirogov was convinced in the existence of regularities in mass phenomena. Thus (1850 – 1855/1961, p. 382), each epidemic disease as well as each *considerable* operation had a constant mortality rate, whereas war was a *traumatic epidemic* (1879/1882, p. 295). This latter statement apparently meant that under war conditions the sickness rate and mortality from wounds obeyed statistical laws. Then (1854, p. 2), the skill of the physicians [but not of witch doctors] hardly influenced the total result of the treatment of many patients.

A French physician Louis (1825) introduced the so-called numerical method of studying symptoms of various diseases. His proposal had been applied much earlier in various branches of science. It amounted to the use of the statistical method without involving stochastic considerations which. Quantitative data were also collected in agriculture, meteorology, astronomy etc.; astronomical catalogues, for example, fall in the same category. Nevertheless, this line of development was not sufficient. See also § 2, bullet point 2.

Discussions about the numerical method lasted at least a few decades. Thus, d'Amador (1837) attacked Louis wrongly attributing to him a recommendation to use the theory of probability.

Gavarret (1840), a former student of Poisson, noted the shortcomings of the numerical method and adduced examples on the comparison of competing methods of medical treatment as also an advice on the check of the null hypothesis (as it is now called), see p. 194. Thus, apart from popularizing probability theory, Gavarret's main achievement was the introduction of the principle of the null hypothesis and the necessity of its check into medicine (actually, in natural science in general).

Laplace (1798 – 1825/1878 – 1882, t. 3, ca. 1804, p. xi; 1814/1995, p. 116) argued that the adopted hypotheses ought to be *incessantly rectified by new observations* until *veritable causes or at least the laws of the phenomena* be discovered. Cf. Double et al (1835, p. 176 – 177): the main means for revealing the *vérité* were induction, analogy and hypotheses founded on facts and *incessantly verified and rectified by new observations*.

Gavarret's contribution became generally known and many authors repeated his recommendations. The time for mathematical statistics or for its application in medicine was not yet ripe, but at least the Poisson – Gavarret tradition led to the existence, in medicine, of a lasting drive towards the use of probability based on numerous observations (and the skill of the physician). Indeed, Poisson (1837a, Note to Annotated Contents) stated that *Medicine will not become either a science or an art if not based on numerous observations, on the tact and proper experience of the physicians …*

A few years earlier Double, Poisson et al (1835) maintained that statistics was *the functioning mechanism of the calculus of probability, necessarily concerning infinite* [?] *masses …*

A large number of observations! However, at least from the mid-18th century (Bull 1959, p. 227) valuable medical conclusions had been based on very small numbers of them, but it was Liebermeister (ca. 1877) who vigorously opposed Gavarret and Poisson. He argued that it was impossible, in therapeutics, to collect vast observations and, anyway, recommendations based on several (reliable) observations should be adopted as well. Statisticians have only quite recently discovered his paper written as though by a specialist in mathematical statistics. Then, at least from Gossett (Student) onwards small samples are necessary for statistics. For his life and work see E. S. Pearson (1990).

2.2. Juridical Statistics. Niklaus Bernoulli published a dissertation on the application of the art of conjecturing to jurisprudence (1709/1975). It contained the calculation of the mean duration of life and recommended its use for ascertaining the value of annuities and estimating the probability of death of absentees about whom nothing is known; methodical calculations of expected losses in marine insurance; calculation of expected losses in the celebrated Genoese lottery and of the probability of truth of testimonies; the determination of the life expectancy of the last survivor of a group of men (pp. 296 – 297), see Todhunter (1865, pp. 195 – 196). Assuming a continuous uniform law of mortality (the first continuous law in probability theory), he calculated the expectation of the appropriate order statistic and was the first to use, in a published work, both this distribution and an order statistic.

Bernoulli’s work undoubtedly fostered the spread of stochastic notions in society, but he borrowed separate passages from the *Ars Conjectandi* and even from the *Meditationes* (Kohli 1975, p. 541), never intended for publication. His general references to Jakob, his late uncle, do not excuse his plagiarism the less so since he dedicated his work not to the memory of Jakob, but to his father Johann.

Condorcet, Laplace and Poisson actively studied the application of probability and statistics to jurisprudence. Todhunter (1865, p. 352) concluded that *The obscurity and self contradictions* in the work of Condorcet *are without any parallel*, but Poisson (1837a, pp. 2 and 5) favourably mentioned his ideas. As to Laplace, it seems that his main achievement consisted in drawing Poisson’s attention to the administration of justice.

Poisson (1837a, pp. 1 – 2) thought that the study of the probabilities of verdicts and, in general, of majority decisions, was a most important application of the calculus of probability. He (p. 17) perceived his main goal in that field as an examination of the stability of the rate of conviction and of the probability of miscarriage of justice as well as in the comparison of judicial statistics of different countries and (p. 7) in proving the applicability of mathematical analysis *to things that are called moral*.

Poisson was mainly interested in studying criminal offences. Unlike Laplace, he (p. 4 and § 114, p. 318) introduced a positive probability of the defendant's guilt (not to be taken into account in any individual case). One of his statements (§ 136, pp. 375 – 376) is debatable: he thought that the rate of conviction should increase with crime.

Poisson estimated the (beneficial or otherwise) changes in the rate of conviction following changes in the administration of justice (in the number of jurors, in the majority vote needed for conviction). I do not know, however, whether his calculations had been taken into account.

Neither Condorcet, nor Poisson mentioned that they had assumed that the jurors reach decisions independently from each other whereas Laplace (1816, p. 523) only said so in passing.

The application of the theory of probability to jurisprudence continued to be denied. Here are the two most vivid pertinent statements (Mill 1843/1886, p. 353; Poincaré 1896/1912, p. 20):

1) *Misapplications of the calculus of probability* […] *made it the real opprobrium of mathematics. It is sufficient to refer to the applications made of it to the credibility of witnesses, and to the correctness of the verdicts of juries*.

2) People *influence each other* and act like *the moutons de Panurge*.

The higher is a scientist's standing, the more reserved he ought to be when invading an alien field. Even Mill, not to mention Poincaré, should not have categorically condemned a subject of which he was ignorant.

It is opportune to cite Gauss whose opinion was voiced by W. E. Weber in a letter of 1841 to J. F. Fries (Gauss, W-12, pp. 201 – 204): probability can serve as a guide line for determining the desired number of jurors and witnesses. Fries had then been preparing his book on the principles of the theory of probability; it appeared in 1842. Then, juridical statistics effectively applied the notion of errors of both kinds.

2.3. Insurance of Property and Life Insurance. Marine insurance was the first essential type of insurance of property but it lacked stochastic ideas or methods. In particular, there existed an immoral and repeatedly prohibited practice of betting on the safe arrivals of ships. Anyway, marine insurance had been apparently based on rude and subjective estimates.

And here is a quote from the first English Statute on assurance (Publicke Acte No. 12, 1601; *Statutes of the Realm*, vol. 4, pt. 2, pp. 978 – 979):

And whereas it hathe bene tyme out of mynde an usage amongste merchants, both of this realme and of forraine nacyons, when they make any great adventure, […] *to give some consideracion of money to other persons* […] *to have from them assurance made of their goodes, merchandizes, ships, and things adventured,* […] *whiche course of dealinge is commonly termed a policie of assurance* […].

Life insurance came into its own not by a front-door entrance, but by the marine insurance porthole (O'Donnell 1936, p. 78) … It exists in two main forms. Either the insurer pays the policy-holder or his heirs the stipulated sum on the occurrence of an event dependent on human life; or, the latter enjoys a life annuity. Annuities were known in Europe from the 13th century onward although later they were prohibited for about a century until a Papal bull officially allowed it in

1423 (Du Pasquier 1910, pp. 484 – 485). The annuitant's age was not usually taken into consideration either in the mid-17th century (Hendriks 1853, p. 112), or even, in England, during the reign of William III [1689 – 1702] (K. Pearson 1978, p. 134). Otherwise, as it seems, the ages had been allowed for only in a very generalized way (Kohli & van der Waerden 1975, pp. 515 – 517; Hald 1990, p. 119). At the end of the 17th century the situation began to change.

In the 18th, and even in the mid-19th century, life insurance still hardly essentially depended on stochastic considerations; moreover, the statistical data collected by the insurance societies as well as their mortality tables and methods of calculations remained secret. And more or less honest business based on statistics of mortality hardly superseded downright cheating before the second half of the 19th century. Nevertheless, beginning at least from the 18th century, the institute of life insurance which essentially depended on studies of mortality strongly influenced the theory of probability and turned the attention of scholars to medical and social problems.

Tontines constituted a special form of mutual insurance. Named after the Italian banker Laurens Tonti (Hendriks 1863), they, acting as a single body, distributed the total sums of annuities among their members still alive, so that those, who lived longer, received considerable moneys. Tontines were neither socially accepted nor widespread on the assumed rationale that they are too selfish and speculative (Hendriks 1853, p. 116). Nevertheless, they did exist in the 17th century. Euler (1776) devised a tontine with flexible moments of entering it, and flexible ages of its members and of their contributions (therefore, of their annual income as well). Such a tontine could have theoretically existed forever rather than disappearing with the death of its last member. Euler's innovation was apparently never taken up.

De Moivre first examined life insurance in the beginning of the 1720s and became the most influential author of his time in that field. Issuing from Halley's table, he (1725/1756, pp. 262 – 263) assumed a continuous uniform law of mortality for all ages beginning with 12 years and a maximal duration of life equal to 86 years.

Hald (1990, pp. 515 – 546) described in detail the work of De Moivre and of his main rival, Simpson (1775), in life insurance. Simpson improved on, and in a few cases corrected the former's findings. After discussing one of the versions of mutual insurance, Hald (p. 546) concluded that Simpson's relevant results represented an essential step forward; however, his attitude to De Moivre showed him as an *unblushing liar* (K. Pearson 1978, p. 184).

Daniel Bernoulli (1768b) investigated the duration of marriages for differing ages of man and wife which was important for insurance on two lives. He based his analysis on another study (1768a) of extracting pairs of white and black stripes from an urn with the respective probabilities being equal or unequal.

Laplace (1814/1995, p. 89) compared *free people* to *an association whose members mutually protect their property* and went on to praise institutions based on the probabilities of human life. Markov collaborated with pension funds (Sheynin 1997) and in 1906 he destructively criticized a proposed official scheme for insuring children (reprinted in same article).

Actuarial science inevitably led to the compilation of life tables and their improvement. Quetelet & Smits (1832, p. 33) stated that separate tables for men and women had *only recently* begun to be published. However (Nordenmark 1929, p. 250), Wargentin compiled such separate tables for Sweden in 1766.

Then, many authors noted that the expectation of life of the general male (say) population is either larger or smaller than that of men from selected populations (e. g., from monks). Note a related remark made by Buffon (1777/1954, § 8, note) in 1762, in a letter to Daniel Bernoulli and thus to some extent foreshadowing Quetelet's Average man:

Mortality tables are always concerned with the average man; that is, with people in general, feeling themselves quite well or ill, healthy or infirm, robust or feeble.

Andersson (1929, p. 239) voiced a serious complaint:

The State does not [do] *much* […] *to protect and forward the sound practice of insurance*. […] *State statistics should pay regard to all the desires of insurance and try to meet them.* […] *No country has as yet suitable fire statistics, no shipping statistics are* [is] *being performed with due attention to the special demands of marine insurance*. […] *The insurance itself* […] *still has not given the due place to statistics in the scientific insurance work*.

2.4. Earliest Stochastic and Statistical Investigations

2.4.1. Pascal and Fermat. In 1654 Pascal and Fermat exchanged several letters (Pascal 1654) which heralded the beginning of the formal history of probability. They discussed several problems; here is the most important of them which was known even at the end of the 14th century. Two or three gamblers agree to continue playing until one of them scores *n* points; for some reason the game is interrupted and it is required to divide the stakes in a reasonable way. Both scholars solved this *problem of points*, see Takácz (1994), by issuing from one and the same rule: the winnings of the gamblers should be in the same ratio(s) as existed between the expectations of their scoring the *n* points. The actual introduction of that notion, expectation, was their main achievement. They also effectively applied the addition and the multiplication theorems.

The methods used by Pascal and Fermat differed from each other. In particular, Pascal solved the above problem by means of the arithmetic triangle (Edwards 1987) composed, as is well known, of binomial coefficients of the development $(1 + 1)^n$ for increasing values of n. Pascal's relevant contribution (1665) was published posthumously, but Fermat was at least partly familiar with it. Both there, and in his letters to Fermat, Pascal made use of partial difference equations (Hald 1990, pp. 49 and 57).

The celebrated Pascal wager (1669/2000, pp. 676 – 681), also published posthumously, was a discussion about choosing a hypothesis. Does God exist, rhetorically asked the devoutly religious author and answered: you should bet. If He does not exist, you may live calmly [and sin]; otherwise, however, you can lose eternity. In the mathematical sense, Pascal's reasoning is vague; perhaps he had no time to edit his fragment. Its meaning is, however, clear: if God exists with a fixed and however low probability, the expectation of the benefit accrued by believing in Him is infinite. Pascal died in 1662 and the same year Arnauld & Nicole (1662/1992, p. 334) published a similar statement:

Infinite things, like eternity and salvation, cannot be equated to any temporal advantage. […] *We should never balance them with any worldly benefit*. […] *The least degree of possibility of saving oneself is more valuable than all the earthly blessings taken together, and the least peril of losing that possibility is more considerable than all the temporal evils* […].

2.4.2. Huygens. Huygens was the author of the first treatise on probability (1657). Being acquainted only with the general contents of the Pascal – Fermat correspondence, he independently introduced the notion of expected random winning and, like those scholars, selected it as the test for solving stochastic problems. He went on to prove that the value of expectation of a gambler who gets a in p cases and b in q cases was

$$\frac{pa+qb}{p+q}. \qquad (1)$$

Jakob Bernoulli (1713/1999, p. 9) justified the expression (1) much simpler than Huygens did: if each of the p gamblers gets a, and each of the q others receives b, and the gains of all of them are the same, then the expectation of each is equal to (1). After Bernoulli, however, expectation began to be introduced formally: expressions of the type of (1) followed by definition.

Huygens solved the problem of points under various initial conditions and listed five additional problems two of which were due to Fermat, and one, to Pascal. He solved them later, either in his correspondence, or in manuscripts published posthumously. They demanded the use of the addition and multiplication theorems, the introduction of conditional probabilities and the formula (in modern notation)

$$P(B) = \Sigma P(A_i)P(B/A_i),\ i = 1, 2, \ldots, n.$$

Problem No. 4 was about sampling without replacement. An urn contained 8 black balls and 4 white ones and it was required to determine the ratio of chances that in a sample of 7 balls 3 were, or were not white. Huygens determined the expectation of the former event by means of a partial difference equation (Hald 1990, p. 76). Nowadays such problems leading to the hypergeometric distribution (Jakob Bernoulli 1713/1999, pp. 167 – 168; De Moivre 1712/1984, Problem 14 and 1718/1756, Problem 20) appear in connection with statistical quality control.

Pascal's Problem No. 5 was the first to discuss the gambler's ruin. Gamblers A and B undertake to score 14 and 11 points respectively in a throw of 3 dice. They have 12 counters each and it is required to determine the ratio of the chances that they be ruined. The stipulated numbers of points occur in 15 and 27 cases and the ratio sought is therefore $(5/9)^{12}$.

In 1669, in a correspondence with his brother, Huygens (1888 – 1950, 1895), see Kohli & van der Waerden (1975), discussed stochastic problems connected with mortality and life insurance. Issuing from Graunt's mortality table, Huygens (pp. 531 – 532) introduced the probable duration of life (but not the term itself). He also showed that the probable duration of life could be determined by means of the graph (plate between pp. 530 and 531) of the function $y = 1 - F(x)$, where, in modern notation, $F(x)$ was a remaining unknown integral distribution function with admissible values of the argument being $0 \le x \le 100$.

In the same correspondence Huygens (p. 528) examined the expected period of time during which 40 persons aged 46 will die out; and 2 persons aged 16 will both die. The first problem proved too difficult, but Huygens might have remarked that the period sought was 40 years (according to Graunt, 86 years was the highest possible age). He mistakenly solved a similar problem by assuming that the law of mortality was uniform and that the number of deaths will decrease with time, but for a distribution, continuous and uniform in some interval, n order statistics will divide it into $(n + 1)$ approximately equal parts and the annual deaths will remain about constant. In the second problem Huygens applied conditional expectation. When solving problems on games of chance, Huygens issued from expectations which varied from set to set rather than from constant probabilities and was compelled to compose and solve difference equations. See also Shoesmith (1986).

2.4.3. Newton left interesting ideas and findings pertaining to probability, but more important were his philosophical views (K. Pearson 1926):

Newton's idea of an omnipresent activating deity, who maintains mean statistical values, formed the foundation of statistical development through Derham, Süssmilch, Niewentyt, Price to Quetelet and Florence Nightingale […]. *De Moivre expanded the Newtonian theology and directed statistics into the new channel down which it flowed for nearly a century. The cause which led De Moivre to his* <u>Approximatio</u> [1733] *or Bayes to his theorem were more theological and sociological than purely mathematical, and until one recognizes that the post-Newtonian English mathematicians were more influenced by Newton's theology than by his mathematics, the history of science in the 18th century – in particular that of the scientists who were members of the Royal Society – must remain obscure*.

Bayes theorem is a misnomer (§ 2.4.7). Then, Newton never mentioned mean values. In 1971, answering my question on this point, the Editor of his future book (1978), E. S. Pearson, stated:

From reading [the manuscript of that book] *I think I understand what K. P. meant*. […] *He had stepped ahead of where Newton had to go, by stating that the laws which give evidence of Design, appear in the stability of the mean values of observations. i. e.*, [he] *supposed* *that Newton was perhaps unconsciously thinking what De Moivre put into words.*

Indeed, K. Pearson (1978, pp. 161 and 653) had attributed to De Moivre (1733/1756, pp. 251 – 252) the Divine *stability of statistical ratios, that is, the original determination of original design* and referred to Laplace who (1814/1995, p. 37) had formulated a related idea:

In an infinitely continued sequence of events, the action of regular and constant causes ought, in the long run, to outweigh that of irregular causes.

However, Laplace never mentioned Divine design. And here is Newton's most interesting pronouncement (1704/1782, Query 31):

Blind fate could never make all the planets move one and the same way in orbs concentrick, some inconsiderable irregularities excepted, which may have risen from the mutual actions of comets and planets upon one another, and which will be apt to increase, till this system wants a reformation. Such a wonderful uniformity in the planetary system must be allowed the effect of choice. And so must the uniformity in the bodies of animals.

Newton's recognition of the existence and role of random disturbances is very important. At the same time Newton (1958, pp. 316 – 318) denied randomness and explained it by ignorance of causes.

Newton (MS 1664 – 1666/1967, pp. 58 – 61) was the first to mention geometric probability: *If the Proportion of the chances* […] *bee irrational, the interest may bee found after ye same manner.* Newton then considered a throw of an irregular die. He remarked that [nevertheless] *it may bee found how much one cast is more easily gotten than another*. He likely bore in mind statistical probabilities. Newton (1728) also applied simple stochastic reasoning for correcting the chronology of ancient kingdoms:

The Greek Chronologers […] *have made the kings of their several Cities* […] *to reign about 35 or 40 years a-piece, one with another; which is a length so much beyond the course of nature, as is not to be credited. For by the ordinary course of nature Kings Reign, one with another, about 18 or 20 years a-piece; and if in some instances they Reign, one with another, five or six years longer, in others they reign as much shorter: 18 or 20 years is a medium*.

Newton derived his own estimate from other chronological data and his rejection of the twice longer period was reasonable. Nevertheless, a formalized reconstruction of his decision is difficult: within one and the same dynasty the period of reign of a given king directly depends on that of his predecessor. Furthermore, it is impossible to determine the probability of a large deviation of the value of a random variable from its expectation without knowing the appropriate variance (which Newton estimated only indirectly and in a generalized way). And here is the opinion of Whiteside (private communication, 1972) about his thoughts concerning errors of observation:

Newton in fact (*but not in explicit statement*) *had a precise understanding of the difference between random and structurally 'inbuilt' errors. He was certainly, himself, absorbed by the second type of 'inbuilt' error, and many theoretical models of differing types* *of physical, optical and astronomical phenomena were all consciously contrived so that these structural errors should be minimized. At the same time, he did, in his astronomical practice, also make suitable adjustment for 'random' errors in observation …*

2.4.4. Arbuthnot. See § 1.3.1-1.

2.4.5. Jakob Bernoulli. His *Ars Conjectandi* (1713) appeared posthumously; Niklaus Bernoulli compiled a Preface (Jakob Bernoulli 1975, p. 108) where, for the first time ever, the term *calculus of probability* (in Latin) had appeared. The book itself contained four parts. Interesting problems are solved in parts 1 (a reprint of Huygens' tract, see § 1.3) and 3 (the study of random sums for the uniform and the binomial distributions, a similar investigation of the sum of a random number of terms for a particular discrete distribution, a derivation of the distribution of the first order statistic for the discrete uniform distribution and the calculation of probabilities appearing in sampling without replacement). The author's analytical methods included combinatorial analysis and calculation of expectations of winning in each set of finite and infinite games and their subsequent summing.

In the beginning of pt. 4 Bernoulli explained that the theoretical number of cases was often unknown, but what was impossible to obtain beforehand, might be determined by observations. In his Diary (*Meditationes*), whose stochastic considerations were only published in Bernoulli (1975), he indirectly cited Graunt and reasoned how much more probable was it that a youth will outlive an old man than vice versa.

Bernoulli maintained that moral certainty ought to be admitted on a par with absolute certainty. His theorem will show, he declared, that statistical probability was a morally certain (a *consistent*) estimator of the theoretical probability. It was Descartes (1644/1978, pt. 4, No. 205, 483°, p. 323) who introduced moral certainty *for regulating our morals* (moeurs).

Actually, Bernoulli proved a proposition that, beginning with Poisson, is called the LLN. Denote the statistical probability of the occurrence of the studied event in a trial by $\hat{p}$ and the theoretical probability of the event by p; assume that n independent Bernoulli trials in which p = Const are made. Then, as $n \to \infty$,

$$\lim P(\hat{p} - p) = 1. \qquad (1)$$

This is an existence theorem and Bernoulli properly stated that it signified that [for the Bernoulli trials] induction (the trials) was not worse than deduction (the theoretically determined p). Had the right side of (1) be a proper fraction, he added, induction would have been worse.

His direct LLN thus determined $\hat{p}$ whereas, as stated above, he initially stated that $\hat{p}$ was a morally certain estimate of p; moreover, he even adduced an appropriate example in which p did not even exist. This initial statement is called the *inverse* LLN, and Bernoulli mistakenly believed that any version of that law led to the other version.

Bernoulli also estimated the rapidity of the convergence of one probability to the other; however, not knowing the later discovered Stirling theorem, his estimation was not good enough. Without noticing the existence theorem K. Pearson (1925) denied Bernoulli's great achievement and even compared it with the wrong Ptolemaic system of the world.

As Cournot (1843, § 86) emphasized, although not definitely enough, stochastic reasoning was now justified beyond the province of games of chance, at least for the Bernoulli trials. Strangely enough, statisticians for a long time had not recognized this fact. Haushofer (1872, pp. 107 – 108) declared that statistics, since it was based on induction, had no intrinsic connections with mathematics based on deduction. And Maciejewski (1911, p. 96) introduced a statistical LLN instead of the Bernoulli proposition that allegedly impeded the development of statistics. His new law qualitatively asserted that statistical indicators exhibited ever lesser fluctuations as the number of observations increased and his opinion likely represented the prevailing attitude of statisticians. Bortkiewicz (1917, pp. 56 – 57) thought that the LLN ought to denote a quite general fact, unconnected with any stochastic pattern, of a degree of stability of statistical indicators under constant or slightly changing conditions and a large number of trials. Even Romanovsky (1961, p. 127) kept to a similar view.

2.4.6. De Moivre. For $n \to \infty$ De Moivre's main result concerning Bernoulli trials can be written as

$$\lim P\ [a \le \frac{\mu - np}{\sqrt{npq}} \le b] = \frac{1}{\sqrt{2\pi}} \int_a^b \exp(-\frac{z^2}{2}) dz. \qquad (2)$$

Here μ is the number of successes, $np = E\mu$ and $npq = \text{var}\mu$.

This is the integral De Moivre – Laplace theorem, as Markov (1900/1924, p. 53) called it, – a particular case of the CLT. Neither De Moivre, nor Laplace knew about uniform convergence with respect to a and b that takes place here. De Moivre proved (2) in a short Latin memoir of 1733 which he sent around to his colleagues and then translated it into English and incorporated in the editions of 1738 and 1756 of his *Doctrine of Chances.*

Laplace (1812) proved (2) simpler and provided a correction term allowing for the finiteness of n. De Morgan (1864) was the first to notice the normal distribution in (2) but he also made unbelievably wrong statements about the appearance of negative probabilities and those exceeding unity. More: in a letter of 1842 he (Sophia De Morgan 1882, p. 147) declared that $\tan\infty = \cot\infty = \mp\sqrt{-1}$.

2.4.7. Bayes. I dwell on the posthumous memoir (Bayes 1764 – 1765) complete with the commentaries by R. Price. In its first part Bayes introduced his main definitions and proved a few theorems; note that he defined probability through expectation. There was no hint of the so-called Bayes theorem introduced by Laplace (1812/1886, p. 183)

$$P(A_i/B) = \frac{P(b/a_j)P(A_i)}{\sum_{j=1}^{n} P(b/a_j)P(A_i)} \qquad (3)$$

as Cournot (1843, § 88), actually following predecessors, called it.

Here is the real Bayes' theorem in a simplified description (Gnedenko 1954, p. 366). It is required to determine the unknown probability r having continuous uniform density on interval [0, 1] if after $n = p + q$ (independent) trials it occurred p times and failed q times. Answer:

$$P(a \le r \le b) = \int_a^b u^p (1-u)^q du \div \int_0^1 u^p (1-u)^q du. \qquad (4)$$

Here, [a, b] is a segment within [0, 1]. Bayes derived the denominator of (4) obtaining the value of the [beta-function] $B(p+1; q+1)$ and spared no effort in estimating its numerator, a problem that remained difficult until the 1930s. The right side of (4) is now known to be equal to the difference of two values of the incomplete beta-function

$$I_b(p + 1; q + 1) – I_a(p + 1; q + 1).$$

Beginning with the 1930s and perhaps for three decades English and American statisticians had been denying Bayes after which his theorem *has returned from the cemetery* (Cornfield 1967).

The first and the main critic of the Bayes *theorem* or formula was Fisher (1922, pp. 311 and 326). It seems that he disagreed with the introduction of hardly known prior probabilities and/or with the assumption that they were equal to one another.

Bayes had not expressly discussed the case of $n \to \infty$. In another posthumous note published in 1764 he warned mathematicians about the danger of applying divergent series. He had not named De Moivre, but apparently had in mind his derivation of the De Moivre – Laplace theorem as well. De Moivre and his contemporaries had indeed employed convergent parts of divergent series for approximate calculations, and about a century later Poisson (1837a, § 68, p. 175) stated that that trick was possible. Note that divergent series are now included in the province of mathematics.

Timerding, the Editor of the German translation of the Bayes memoir (1908), managed to consider the limiting case without applying divergent series. He issued from Bayes' calculations made for large but finite values of p and q. Applying a clever trick, he proved that, as $n \to \infty$, the probability of the studied event obeyed the proposition

$$\lim P\{-z \le \frac{r-a}{\sqrt{pq}n^{3/2}} \le z\} = \frac{1}{\sqrt{2\pi}} \int_0^z \exp(-\frac{w^2}{2})dw, \qquad (5)$$

where (not indicated by Timerding) $a = p/n = \mathrm{E}r$, $pq/n^{3/2} = \mathrm{var}r$ so that $r \approx \mathrm{E}r = p/n$.

The assumption of a uniform density is not a restriction; according to information theory, it is tantamount to a statement of ignorance. The influence of a non-uniform density taking place apparently decreases with the increase of n (with the increase in posterior information). The functions in the left sides of formulas (2) and (5) are different random variables, centred and normed in the same way; Bayes, without knowing the notion of variance, apparently understood that (2) was not sufficiently precise for describing the problem inverse to that studied by De Moivre, who (1718/1756, p. 251) mistakenly thought otherwise (as Jakob Bernoulli also did). Note that, unlike the direct law, its inverse counterpart has less initial data (the theoretical probability is unknown) which qualitatively explains the situation.

A modern encyclopaedia (Prokhorov 1999) contains 14 items mentioning Bayes, for example, Bayesian estimator, Bayesian approach (and many more items are mentioned elsewhere). There also, on p. 37, the author of the appropriate entry mistakenly attributes formula (3) to Bayes. For my part, I believe that, since Bayes had correctly interpreted the inverse LLN, he thus completed the first stage of the theory of probability. He also was the main predecessor of Mises (who never acknowledged it). And when a statistician starts working, he invariably has to issue from some statistical probability. If and when justifying this step, he refers to Mises, but he could have mentioned Bayes instead.

3. Treatment of Observations

3.1. The following **explanation** will be needed below. Denote the observations of a constant sought by

$$x_1, x_2, \ldots, x_n, \; x_1 \le x_2 \le \ldots \le x_n. \qquad (1)$$

It is required to determine its value, optimal in some sense, and estimate the residual error. The classical theory of errors considers independent observations and, without loss of generality, they might also be regarded as of equal weight. This problem is called *adjustment of direct observations*.

Suppose now that k unknown magnitudes $x, y, z, \ldots$ are connected by a redundant system of n physically independent equations ($k < n$)

$$a_i x + b_i y + c_i z + \ldots + s_i = 0 \qquad (2)$$

whose coefficients are given by the appropriate theory and the free terms are measured. The approximate values of $x, y, z, \ldots$ were usually known, hence the linearity of (2). The equations are linearly independent (a later notion), so that the system is inconsistent (which was perfectly well understood). Nevertheless, a solution had to be chosen, and it was done in such a way that the residual free terms (call them v_i) were small enough. The values of the unknowns thus obtained are called their estimates ($\hat{x}, \hat{y}, \ldots$) and this problem is called *adjustment of indirect measurements*.

Since the early 19th century the usual condition for solving (2) was that of least squares

$$W = \sum v_i^2 = [vv] = v_1^2 + v_2^2 + \ldots + v_n^2 = \min, \quad (3)$$

so that

$$\partial W/\partial x = \partial W/\partial y = \ldots = 0. \quad (4)$$

Conditions (4) easily lead to a system of normal equations

$$[aa]\hat{x} + [ab]\hat{y} + \ldots + [as] = 0, [ab]\hat{x} + [bb]\hat{y} + \ldots + [bs] = 0, \ldots,$$

with a positive definite and symmetric matrix. For direct measurements the same condition (3) leads to the arithmetic mean.

There also existed a determinate branch of the theory of errors now partly superseded by experimental design. It studies the process of measurement without applying stochastic reasoning. Here is a simplest example: determine the form of a geodetic figure ensuring optimal (in some sense) results. The real development of the determinate error theory was due to the differential calculus which ensured the study of the sought functions of measured magnitudes, but even Hipparchus was aware that, under favourable conditions, a given error of observation can comparatively little influence the unknown sought (Toomer 1974, p. 131), see also below.

Gauss and Bessel originated a new direction in practical astronomy and geodesy. They demanded and carried out thorough examinations of the instruments and investigations of the plausibility of the methods of observation. This direction belonged to the determinate error theory.

Now, the design of experiments is a branch of mathematical statistics dealing with the rational organization of measurements subject to random errors (*Enc. Math.* 1977 – 1985/1988 – 1994, vol. 3, p. 66). Finney (1960), however, argued that this new discipline does not entirely belong to the *mathematical theory of statistics*, but did not elaborate. I would say, belongs to theoretical statistics, see § 7.2.

The design of experiments ought to include the choice of optimal methods and circumstances of observation, design of instruments capable of using such methods etc. (Box 1964). Many of such problems have nothing to do with randomness; and they undoubtedly belong to the determinate error theory.

Some Russian authors (Romanovsky 1955; Bolshev 1989) state that the stochastic theory of errors belongs to statistics, but it seems more natural to define it as the application of the statistical method to the treatment of observations in experimental science, see § 9. Romanovsky excluded systematic errors from their consideration; Bolshev agreed and attributed their study to a special discipline, the processing (the treatment) of observations. I categorically deny such opinions. Observers have to take care of both random and systematic errors which cannot therefore be attributed to separate branches (or twigs) of science.

3.2. Ancient astronomers apparently selected point estimates for the constants sought by choosing almost any number within reasonable bounds. According to modern notions, such an attitude is quite proper if the errors of observations are large; moreover, it fits in with the qualitative nature of ancient science.

It was Daniel Bernoulli (1780) who introduced, although in a restricted sense, the notions of random and systematic errors, but ancient astronomers obviously acquired some understanding of both. The influence of refraction, for example, was systematic.

3.3. In **Kepler's** time, and possibly even somewhat earlier, the arithmetic mean became the generally accepted estimator of measurements. Indeed, Kepler (1609/1992, p. 200/63), when treating four observations, selected a number as the *medium ex aequo et bono* (in fairness and justice). A plausible reconstruction assumes that it was a generalized arithmetic mean with differing weights of observations. More important, the Latin expression above occurred in Cicero, 106 – 43 BC (*Pro A. Caecina oratio*), and carried an implication *Rather than according to the letter of the law*, an expression known to lawyers. In other words, Kepler, who likely read Cicero, called the ordinary arithmetic mean the letter of the law, i.e., the universal estimator of the parameter of location.

Kepler repeatedly adjusted observations. How had he convinced himself that Tycho's observations were in conflict with the Ptolemaic system of the world? I believe that Kepler applied the minimax principle which demanded that the residual free term of the given system of equations, maximal in absolute value, be the least from among all of its possible solutions. He (1609/1992, p. 286/113) apparently determined such a minimum, although only from among some possibilities, and found out that that residual was equal to 8′ which was inadmissible. Any other solution would have been even less possible, so that either the observations or the underlying theory were faulty. Kepler reasonably trusted Tycho's observations and his inference was obvious. Note that this principle did not ensure optimal, in any sense, results.

When adjusting observations, Kepler (Ibidem, p. 334/143) corrupted them by small arbitrary corrections. He likely applied elements of what is now called statistical simulation, but in any case he must have taken into account the properties of usual random errors, i.e., must have chosen a larger number of small positive and negative corrections and about the same number of the corrections of each sign. Otherwise, Kepler would have hardly achieved success.

3.4. Direct Observations. I am now entering the 18$^{\text{th}}$ century and, after discussing Lambert, begin with the treatment of direct observations.

3.4.1. The term *Theory of errors* (*Theorie der Fehler*) is due to **Lambert** (1765a, Vorberichte and § 321) who defined it as the study of the relations between errors, their consequences, circumstances of observation and the quality of the instruments. He isolated the aim of the *Theory of consequences* as the study of functions of observed (and error-ridden) magnitudes. In other words, he introduced the determinate error theory and devoted to it §§ 340 – 426 of his contribution. Neither Gauss, nor Laplace ever used the new terminology, but Bessel (1820, p. 166; 1838b, § 9) applied the expression *theory of errors* without mentioning anyone and by the mid-19th century it became generally known.

Lambert studied the most important aspects of treating observations and in this respect he was Gauss' main predecessor. He (1760, §§ 271 – 306) described the properties of usual random errors, classified them in accordance with their origin (§ 282), unconvincingly proved that deviating observations should be rejected (§§ 287 – 291) and estimated the precision of observations (§ 294), again lamely but for the first time ever. He then formulated an indefinite problem of determining a [statistic] that with maximal probability least deviated from the real value of the constant sought (§ 295) and introduced the principle of maximal likelihood, but not the term itself, for a continuous density (§ 303), maintaining, however (§ 306), that in most cases it will provide estimates little deviating from the arithmetic mean. The translator of Lambert's contribution into German left out all this material claiming that it was dated.

Lambert introduced the principle of maximum likelihood for an unspecified, more or less symmetric and unimodal curve (as shown on his figure), call it $\varphi(x-\hat{x})$, where $\hat{x}$ was the sought parameter of location. Denote the observations by $x_1, x_2, \ldots, x_n$, and, somewhat simplifying his reasoning, write his likelihood function as

$$\varphi(x_1-\hat{x})\ \varphi(x_2-\hat{x})\ \ldots\ \varphi(x_n-\hat{x}).$$

When differentiating it, Lambert had not indicated that the argument here was the parameter $\hat{x}$, etc.

When Lambert (1765a) returned to the treatment of observations, he attempted to estimate the precision of the arithmetic mean, but did not introduce any density and was unable to formulate a definite conclusion. He also partly repeated his previous considerations and offered a derivation of a density law of errors occurring in pointing an instrument (§§ 429 – 430) in accordance with the principle of insufficient reason: it was a semi-circumference (with an unknown radius) simply because there were no reasons for its *angularity*.

3.4.2. Simpson (1756), see also Shoesmith (1985), applied, for the first time ever, stochastic considerations to the adjustment of measurements by assuming that observational errors obeyed some density law and thus extended probability to a new domain and effectively introduced random observational errors. He aimed to refute some unnamed authors who had maintained that one good observation was as plausible as the mean of many of them. Simpson considered errors obeying the discrete uniform and triangular distributions and effectively applied the proper generating functions.

For both these cases he founded out that the probability that the absolute value of the error of the arithmetic mean of *n* observations was less than some magnitude, or equal to it. He decided that the mean was always [stochastically] preferable to a separate observation and thus arbitrarily and wrongly generalized his proof. Simpson also indicated that his first case was identical with the determination of the probability of throwing a given number of points with n dice each having $(v + 1)$ faces. Note that in the continuous case, Simpson's distributions can be directly compared with each other: their respective variances are $v^2/3$ and $v^2/6$.

Soon Simpson (1757) reprinted his memoir adding to it an investigation of the continuous triangular distribution. However, his graph showed the density curve of the error of the mean which should have been near-normal but which did not possess the distinctive form of the normal distribution.

3.4.3. Daniel Bernoulli (1769) assumed the density law of observational errors as a *semi-ellipse* or semi-circumference of some radius r which he ascertained by assigning a reasonable maximal error of observation and the location parameter equal to the weighted arithmetic mean with posterior weights

$$p_i = r^2 - (\bar{x} - x_i)^2. \qquad (5)$$

Here, x_i were the observations and $\bar{x}$, the usual mean. The first to apply weighted, or generalized arithmetic means was Short (1763). Such estimators demanded a subjective selection of weights and only provided a correction to the ordinary arithmetic mean which tended to vanish for even density functions.

In his published memoir Daniel Bernoulli (1778) objected to the application of the arithmetic mean which (§ 5) only conformed to an equal probability of all possible errors and was tantamount to shooting blindly. K. Pearson (1978, p. 268), however, reasonably argued that small errors were more frequent and had their due weight in the mean. Instead, Bernoulli suggested the maximum likelihood estimator of the location parameter. Listing reasonable restrictions for the density curve (but adding the condition of its cutting the abscissa axis almost perpendicularly), he selected a semi-circumference with radius equal to the greatest possible, for the given observer, error. He then (§ 11) wrote out the likelihood function as

$$\{[r^2 - (x - x_1)^2]\ [r^2 - (x - x_2)^2]\ [r^2 - (x - x_3)^2]\ldots\}^{1/2},$$

where x was the unknown abscissa of the centre of the semi-circumference, and $x_1, x_2, x_3, \ldots$, were the observations. Preferring, however, to ease calculation, he left the semi-circumference for an arc of a parabola but he had not known that the variance of the result obtained will therefore change.

For three observations his likelihood equation was of the fifth degree. Bernoulli numerically solved it in a few particular instances with some values of x_1, x_2, and x_3 chosen arbitrarily (which was admissible for such a small number of them). I present his equation as

$$\frac{x - x_1}{r^2 - (x - x_1)^2} + \frac{x - x_2}{r^2 - (x - x_2)^2} + \ldots = 0$$

so that the maximum likelihood estimate is

$$\hat{x} = \frac{[px]}{\sum p_i},\ p_i = \frac{1}{r^2 - (\hat{x} - x_i)^2} \qquad (6;\ 7)$$

with unavoidable use of successive approximations. For some inexplicable reason these formulas are lacking in Bernoulli's memoir although the posterior weights (7) were the inverse of the weights (5) from his manuscript and heuristically contradicted his own preliminary statement about shooting skilfully. It is now known, however, that such weights are expedient in case of some densities.

3.4.4. Euler (1778, § 6) objected to the principle of maximum likelihood. He argued that the result of an adjustment should barely change whether or not a deviating observation was adopted, but that the value of the likelihood function essentially depended on that decision. His remark should have led him to the median although he (§ 7) selected the estimate (6) with posterior weights (5) and mistakenly assumed that Bernoulli had chosen these same weights.

It is not regrettably known whether Gauss had read these two contributions. Indeed, an intermediate formula of Euler heuristically resembled Gauss' choice of least variance as a criterion for treating observations.

3.5. Indirect Measurements. Here, I consider the adjustment of redundant systems

$$a_ix + b_iy + \ldots + s_i = v_i,\ i = 1, 2, \ldots, n \qquad (8)$$

in k unknowns ($k < n$) and residual free terms v_i.

3.5.1. In case of two unknowns astronomers usually separated systems (8) into all possible **groups of two equations** each and averaged the solutions of these groups. As discovered in the 19th century, the least-squares solution of (8) was some weighted mean of these partial solutions (Whittaker & Robinson 1924/1949, p. 251).

3.5.2. For three unknowns that method becomes unwieldy. In an astronomical context, **Mayer** (1750) had to deal with 27 equations in three unknowns. He calculated three particular solutions (see below), and averaged them. The plausibility of the results thus obtained depended on the expediency of the separation and it seems that Mayer had indeed made a reasonable choice. Being mostly interested in only one unknown, he included the equations with its greatest and smallest in absolute value coefficients in the first, and the second group respectively. Note also that Mayer believed that the precision of results increased as the number of observations, but in his time this mistake was understandable.

Mayer solved each group of equations under an additional condition

$$\Sigma v_i = 0,$$

where i indicated the number of an equation; if the first group included the first nine of them, then $i = 1, 2, \ldots, 9$. Laplace (1812/1886, pp. 352 – 353) testified that the best astronomers had been following Mayer. A bit earlier Biot (1811, pp. 202 – 203) reported much the same.

The condition above determines the method of averages and Lambert's recommendation (1765b, § 20) about fitting an empirical straight line might be interpreted as its application. Lambert separated the points (the observations) into two groups, with smaller and larger abscissas, and drew the line through their centres of gravity, and into several groups when fitting curves.

3.5.3. The Boscovich Method. He (Maire & Boscovich 1770, p. 501) adjusted systems (8) under additional conditions

$$v_1 + v_2 + \ldots + v_n = 0,\ |v_1| + |v_2| + \ldots + |v_n| = \min, \qquad (9;\ 10)$$

the first of which can be allowed for by summing all the equations and eliminating one of the unknowns from the expression thus obtained. The second condition linked the Boscovich' method with the median. Indeed, he adjusted systems (8) by constructing a straight line whose slope was equal to the median of some fractions. In 1809, Gauss noted that (10) led exactly to k zero residuals v_i, which follows from an important theorem in the then not yet known theory of linear programming.

Galileo (1632), see Hald (1990, § 10.3), and Daniel Bernoulli (1735/1987, pp. 321 – 322) applied condition (10) in the case in which the magnitudes such as v_i were positive by definition. Just the same, Herschel (1805) determined the movement of the Sun by issuing from the apparent motion of the stars. The sum of these motions depends on the former and its minimal value, as he assumed, estimated that movement. Herschel's equations were not even algebraic, but, after some necessary successive approximations, they might have been considered linear. In those times the motion of a star could have been discovered only in the plane perpendicular to the line of vision. When treating direct measurements Herschel (1806) preferred the median rather than the arithmetic mean (Sheynin 1984a, pp. 172 – 173).

3.5.4. The Minimax Method. Kepler (§ 3.3) had apparently made use of some elements of this method. Laplace (1789/1895, pp. 493, 496 and 506 and elsewhere) applied it for preliminary investigations. This method corresponds, as Gauss (1809, § 186) remarked, and as it is easy to prove, to the condition

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

Below, I describe the subsequent history of the theory of errors, but right now I emphasize that beginning with Simpson and until the 1930s it had been the main field of application of the theory of probability and that mathematical statistics had borrowed two main principles from the theory of errors, those of maximal likelihood and of least variance.

4. Laplace

He devoted a number of memoirs to the theory of probability and later combined them in his *Théorie analytique des probabilités* (TAP) (1812). He made use of characteristic functions and the inversion formula, calculated difficult integrals, applied Hermite polynomials, introduced the Dirac function and (after Daniel Bernoulli) the

Ehrenfests' model and studied sampling. Issuing from observations, Laplace proved that the Solar system will remain stable for a long time and completed the explanation of the movement of its bodies in accordance with the law of universal gravitation.

He had not even heuristically introduced the notion of *random variable* and was unable to study densities or characteristic functions as mathematical objects, did not bother to prove rigorously his theorems (for example, often issued from non-rigorously proved versions of the CLT, not even properly formulated) which was contrary to the attitude of his predecessors. His theory of probability therefore became an applied mathematical discipline unyielding to development and it had to be constructed anew. Here, indeed, is Poisson (1837a, § 84) who methodically followed Laplace: *There exists a very high probability that these unknown chances little differ from the ratio …*

Then, Laplace insisted on his own impractical justification of the method of least squares and virtually neglected Gauss. Many commentators reasonably stated that his contributions made difficult reading.

Here is an interesting problem from *Chapter 2* of the TAP. An interval *OA* is divided into equal or unequal parts and perpendiculars are erected to the intervals at their ends. The number of perpendiculars is *n*, their lengths (moving from *O* to *A*) form a non-increasing sequence and the sum of these lengths is given. Suppose now that the sequence is chosen repeatedly; what, Laplace asks, will be the mean broken line connecting the ends of the perpendiculars? The mean value of a current perpendicular? Or, in the continuous case, the mean curve? Each curve might be considered as a realization of a stochastic process and the mean curve sought, its expectation. Laplace was able to determine this mean curve and to apply this finding for studying expert opinions.

Suppose that some event can occur because of *n* mutually exclusive causes. Each expert arranges these in an increasing (or decreasing) order of their [subjective] probabilities, which, as it occurs, depend only on *n* and the number of the cause, *r,* and are proportional to

$$\frac{1}{n}+\frac{1}{n-1}+...+\frac{1}{n-r+1}.$$

The comparison of the sums of these probabilities for each cause also shows the mean opinion about its importance. To be sure, different experts will attribute differing perpendiculars to one and the same cause.

In *Chapter 6* Laplace applied the Bayesian approach to problems in population statistics. First, he wrote out formula (5) from § 2.4.7 with *r* being the unknown probability of a male birth and *p* and *q*, the very large numbers of male and female births. He expressed the integrals of *functions of very large numbers* (as Laplace called them) by integrals of an exponential function of a negative square.

In the same way Laplace estimated the population of France (*M*) by issuing from sampling, from the known number of yearly births in

France and in some of its regions (*N* and *n*) and the population of those regions (*m*). K. Pearson (1928) remarked that Laplace had mistakenly considered (*m*, *n*) and (*M*, *N*) as independent samples from the same infinite population (whose very existence was doubtful) and that his estimate of the achieved precision of sampling (the first of its kind) was somewhat erroneous.

Laplace's theory of errors was based on several versions of the CLT (whose conditions he never really formulated!) and therefore required, first of all, a large number of observations. In geodesy, that number was barely sufficient, and the errors in long series of astronomical observations hardly obeyed one and the same law of distribution. And only the normal distribution became worthy of attention.

Without explanation which appeared in his *Supplement* 2 (1818/1886, p. 571) Laplace (1816) approximated the squared sum of the real errors by the same sum of the residuals and, for the case of *s* observations, arrived at an estimator of their variance $m = \sqrt{[vv]/s}$. Interestingly, he (1814/1995, p. 45) stated that *the weight of the mean result increases like the number of observations divided* [*divisé*] *by the number of parameters*. See below the more precise formula due to Gauss (§ 6) and note that *variance* is a modern term.

Curiously, Laplace (1796/1884, p. 504), actually attributed the planetary eccentricities to randomness:

Had the Solar system been formed perfectly orderly, the orbits of the bodies composing it would have been circles whose planes coincided with the plane of the Solar equator. We can perceive however that the countless variations that should have existed in the temperatures and densities of the diverse parts of these grand masses gave rise to the eccentricities of their orbits and the deviations of their movement from the plane of that equator.

Curiously, since Newton had proved that the eccentricities were determined by the planets' initial velocities. However, did Newton get rid of randomness? No, not at all: those velocities seem to be random.

5. Poisson

He introduced the concepts of random variable and distribution function. He contributed to limit theorems and brought into use the LLN, proving it for the case of *Poisson trials*. He devoted much attention to the study of juridical statistics (§ 2.2) and systematically determined the significance of empirical discrepancies. Poisson stressed the difference between subjective and objective probabilities. Cournot (1843) kept to the same attitude and even introduced non-numerical probabilities. They as well as the subjective probabilities are being applied as expert estimates (cf. § 4).

Since Poisson (1837a) consistently checked the significance of empirical discrepancies, for example between results of different series of observations, he, along with Bienaymé, can be called the Godfather of the Continental direction of statistics (Lexis, Bortkiewicz, Chuprov, Markov, Bohlmann, see § 8.3) that mostly studied population. True, his approach was definitely restricted as it became apparent in medicine (§ 2.1).

Poisson's generally known formula (1837a, § 81, p. 206)

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

for an event having probability $q = 1 - p \approx 0$ to occur not more than n times in a large number μ of Bernoulli trials had been all but ignored until Bortkiewicz (1898) introduced his law of small numbers, allegedly a breakthrough extremely important for statistics. However, Whitaker (1914) and then, Kolmogorov (1954) had identified it as the Poisson formula. They did not justify that statement, and I (2008) proved it, see also § 8.3.3.

Poisson's (1837a) LLN is his best known innovation. It generalized the Bernoulli trials on the case of variable probabilities p_i of success although many authors have reasonably noted that his proof was not rigorous. For him, the LLN was rather a principle whose scope he exaggerated. Still, he (p. 10) qualitatively connected it with the existence of a stable mean interval between molecules (Gillispie 1963, p. 438). The founders of the kinetic theory of gases had not regrettably noticed Poisson's conclusion.

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

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

That programme also mentioned insurance establishments, annuities, tontines, savings banks and *emprunts* (loans or perhaps bonds). Social arithmetic therefore meant population statistics, at least some medical statistics and insurance.

Following Laplace, Poisson (see § 4) had often left demonstrations without indicating the boundaries of possible errors and his theory of probability still belonged to applied mathematics. One of his examples (1837a, § 11) led to a subjective probability of the studied event equal to 1/2, and, in conformity with the future information theory, he (Ibidem, § 4) properly remarked that such results illustrate *la perfaite perplexité de notre esprit*.

Poisson (1825 – 1826) applied subjective probability when investigating a game of chance. Cards are extracted one by one from six decks shuffled together as a single whole until the sum of the points in the sample obtained will be in the interval [31; 40]. The sample is not returned and a second sample of the same kind is made. It is required to determine the probability that the sums of the points are equal. Like the gamblers and bankers, Poisson tacitly assumed that the second sample was extracted as though from the six initial fresh decks. Actually, this was wrong, but the gamblers thought that, since they did not know what happened to the initial decks, the probability of drawing some number of points did not change.

When blackjack is played, bankers are duty bound to act the same wrong way: after each round the game continues without the used cards, and, to be on the safe side, they ought to stop at 17 points. A gambler endowed with a retentive memory can certainly profit from this restriction.

Catalan (1877; 1884) even formulated the following principle: If the causes, on which the probability of an event depended, changed in an unknown way, that probability remains unaltered.

6. Gauss

He was the real, although not the formal discoverer of the method of least squares (MLSq) first publicly proposed by Legendre (1805). Indeed, Gauss had applied it from 1794 or 1795, informed his colleagues about it before 1805 and justified it. Legendre, however, only put forward reasonable arguments and, even so, actually and mistakenly stated that the MLSq also ensured a minimax solution of redundant systems of equations.

Three circumstances greatly impeded the dissemination of Gauss' ideas. *First*, although citing Legendre, he (1809, § 186) mentioned *our principle* (of least squares) which insulted the much older French scientist. That same year, Legendre (Gauss, W-9, p. 380) wrote a letter to Gauss stating that priority is only established by publication. A withdrawn person that he was, Gauss did not answer; for the time being, Legendre could have dropped the subject and repeated his proper remark at the first occasion.

As it happened, however, Legendre, as well as all the other French mathematicians interested in the treatment of observations except Laplace, became infuriated and, to their own detriment, for at least a few decades had continued to ignore Gauss' contributions to the theory of errors.

Second, Laplace (1812, § 24) properly described the situation, but kept to his own version of the theory of errors. *Third*, Laplace somehow eclipsed Gauss. Innumerable geodetic textbooks only described the MLSq according to Gauss (1809), but even so many scientists barely noticed that work. Tsinger (1862, p. 1), who obviously did not even read Gauss, was the worst perpetrator:

Laplace provided a rigorous [?] *and impartial investigation* […]. *On the basis of extraneous considerations, Gauss endeavoured to attach to* [the MLSq] *an absolute significance* etc.

So what had Gauss achieved in 1809? Gauss (1809, § 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*. Together with the principle of maximal likelihood, his axiom or *postulate* (Bertrand 1888a, p. 176) led to the normal distribution of the observational errors as the only possible law. Gauss was hardly satisfied with his derivation. His axiom contained qualification remarks, other laws of error were possible and maximum likelihood was worse than an integral criterion. It is somewhat strange that Gauss himself only mentioned the last item and only in a few letters. In his letter to Bessel of 1839 (Plackett 1972/1977, p. 287) he stated that the *highest probability* of the value of an unknown parameter was still infinitely low so that he preferred to rely on the *least disadvantageous game*, on maximum weight or minimal variance.

Indeed, Gauss (1823b, § 6) introduced the variance

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

where the density φ(x) was an even unimodal function (which conformed with the properties of *usual* random errors) and selected its minimal value as the criterion for adjusting observations.

He (§§ 18 and 19) also introduced independence of linear functions: they should not contain common observations. Then Gauss (§§ 37 – 38) proved what was practically necessary: for n observations and k unknowns, the unbiased sample variance and its estimator were, respectively,

$$m^2 = \mathrm{E}[vv]/(n-k),\ \hat{m}^2 = [vv]/(n-k). \qquad \text{(1a, b)}$$

Instead of the mean value, the sum of squares [*vv*] itself has to be applied. Coupled with the principle of maximal weight, formulas (1) provide effective estimators, as they are now called. Without mentioning Laplace, see above, Gauss (1823b, §§ 37 – 38) noted that his formula was not good enough. Elsewhere, he (1823a/1887, p. 199) stated that its correction was also necessary for the *dignity of science*.

The necessary restrictions for the derivation of (1a) are linearity of the equations (1) of § 3.1, independence of their free terms (of the results of observation), and the unbiasedness of the estimators $\hat{x}, \hat{y}, ...$ of the unknowns. An extremely important corollary follows: the immediately appearing principle of least squares can be derived without recourse to sections 7 – 38 of the memoir. Gauss had thus derived the principle of least squares by two independent ways: by the method which he described in those intermediate sections and by the just outlined method.

The first method is so complicated that perhaps up to the second half of the 20th century textbook authors invariably introduced the MLSq in accordance with Gauss' first memoir of 1809, which he no longer acknowledged. Now, however, after my discovery outlined above (Sheynin 2012), the situation has changed.

Gauss (§ 40) calculated the boundaries of the var $\hat{m}^2$ by means of the fourth moment of the errors but made a mistake later corrected by Helmert and then by Kolmogorov et al.

But why did not Gauss even hint at the described possibility? I can only quote Kronecker (1901, p. 42):

The method of exposition in the Disquisitiones [*Arithmeticae* of 1801] *as in his works in general is Euclidean. He formulates and proves theorems and diligently gets rid of all the traces of his train of thoughts which led him to his results. This dogmatic form was certainly the reason for his works remaining for so long incomprehensible*.

Later commentators expressed the same opinion. It remains to illustrate the former difficulties which led to the choice of the memoir of 1809 over Gauss' final memoir of 1823: the very existence of that final memoir (Eisenhart 1964, p. 24)

Seems to be virtually unknown to all American users of Least Squares, except students of advanced mathematical statistics.

Here, indeed, is Fisher (1925, p. 260):

In the cases to which it is appropriate, this method [of least squares] *is a special application of the method of maximum likelihood, from which it may be derived*.

Quite recently Nikulin & Poliscuk (1999) failed to mention that final memoir. Petrov (1954) perhaps still provides the best description of the properties of estimators derived by the MLSq.

6.1. There are important **additional considerations**: the determination of the necessary number of observations, the rejection of outliers and the so-called true values of the unknowns. Owing to the unavoidable presence of systematic errors, the number of observations is not really determined by the formulas (1). For the same reason statistical criteria for rejecting outliers are hardly useful and this latter problem remains delicate.

Astronomers, geodesists, metrologists and other specialists making measurements have always been using the expression *true value*. Mathematical statistics has done away with true values and introduced instead parameters of densities (or distribution functions), and this was a step in the right direction: the more abstract was mathematics becoming, the more useful it proved to be.

Fisher was mainly responsible for that change; indeed, he (1922, pp. 309 – 310) defined the notions of consistency, efficiency and sufficiency of statistical estimators without any reference to true values. But then, on p. 311, he accused the Biometric school of applying the same names to *the true value which we should like to know* […] *and to the particular value at which we happen to arrive*… So the true value was then still alive and even applied, as in the lines above, to objects having no existence in the real world.

The same can be said about Gauss (1816, §§ 3 and 4) who repeatedly considered the true value of a measure of precision of observations. And Hald (1998) mentioned *true value* many times in Chapters 5 and 6; on p. 91 he said: *the estimation of the true value, the location parameter*…

So what is a true value? Markov (1900/1924, p. 323) was the only mathematician who cautiously, as was his wont, remarked: *It is necessary in the first place to presume the existence of the numbers whose approximate values are provided by observations*. This phrase first appeared in the 1908 edition of his *Treatise* (and perhaps in its first edition of 1900). He had not attempted to define *true value*, but this is exactly what Fourier (1826/1890, p. 534) had done about a century before him. He determined the *véritable objet de la recherche* (the constant sought, or its *true* value) as the limit of the arithmetic mean of *n* appropriate observations as $n \to \infty$. Incidentally, he thus provided the Gauss *postulate* with a new dimension.

Many authors, beginning perhaps with Timerding (1915, p. 83) [and including Mises (1919/1964b, pp. 40 and 46)], without mentioning Fourier and independently from each other, introduced the same definition. One of them (Eisenhart 1963/1969, p. 31) formulated the unavoidable corollary: the mean residual systematic error had to be included in that *true* value:

The mass of a mass standard is […] *specified* […] *to be the mass of the metallic substance of the standard plus the mass of the average volume of air adsorbed upon its surface under standard conditions.*

However, even leaving systematic influences aside, the precision of observations is always restricted and the number of observations finite, so that the term *limit* in the Fourier definition (which is in harmony with the Mises definition of probability) must not be understood literally.

Statistics moved from true values to parameters of densities or distribution functions, but still does not entirely abandon them.

6.2. Chronologically, **Helmert** belongs to the second half of the 19th century, but it is better to mention him here. He mainly completed the development of the classical Gaussian theory of errors and some of his findings were interesting for mathematical statistics. Until the 1930s, Helmert's treatise (1872) remained the best source for studying the error theory and the adjustment of triangulation. When adjusting a complicated geodetic net, Helmert (1886, pp. 1 and 86) temporarily replaced chains of triangulation by geodetic lines. His innovation had been applied in the Soviet Union. The chains of the national primary triangulation were there situated between baselines and astronomically determined azimuths. Before the general adjustment of the entire system, each chain was replaced by the appropriate geodetic line; only they were adjusted, then the chains were finally dealt with independently one from another.

Elsewhere Helmert (1868) studied various configurations of geodetic systems. 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, 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 particular geodetic systems unmeasured, but his remark was purely academic: all angles ought to be measured at least for checking the work as a whole.

Abbe (1863) derived the chi-square distribution, see also Sheynin (1966) and Kendall (1971), as the frequency of the sum of the squares of n normally distributed errors. Helmert (1875; 1876) derived the same distribution by induction beginning with $n = 1$ and 2 and Hald (1952/1960, pp. 258 – 261) provided a modernized derivation. Much later Helmert (1905) offered a few tests for revealing systematic influences in a series of errors. Among other results, I note that he (1876) derived a formula that showed that, for the normal distribution, $[vv]$, – and, therefore, the variance as well,– and the arithmetic mean were independent. He had thus proved the important Student – Fisher theorem although without paying any attention to it.

Czuber (1891, p. 460) testified that Helmert had thought that $\operatorname{var}\hat{m}^2/\hat{m}^2$ was more important than $\operatorname{var}\hat{m}^2$ by itself and Eddington (1933, p. 280) independently expressed the same opinion. Czuber also proved that, for the normal distribution, that relative error was minimal for the estimator (1b).

In addition, Helmert noted that for small values of *n* the var $\hat{m}^2$ did not estimate the precision of formula (1b) good enough. His considerations led him to the so-called Helmert transformations.

6.3. Bessel. His achievements in astronomy and geodesy include 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 (1838a) also proved the CLT, but its rigorous proof became possible, with a doubtful exception of one of Cauchy's memoirs, only much later. Incidentally, Gauss was familiar with the pertinent problem. In the letter to Bessel of 1839 mentioned above, he stated that he had read that proof *with great interest*, but that

This interest was less concerned with the thing itself than with your exposition. For the former has been familiar to me for many years, though I myself have never arrived at carrying out the development completely.

The personal equation of an observer is his systematic error of registering the moments of the passage of a star through the cross-hairs of the eyepiece of an astronomical instrument. When studying this phenomenon, it is possible to compare the moments fixed by two astronomers at different times. Although Bessel did not explain the situation, it followed from the context that he and another astronomer had only one clock. Consequently, it was necessary to take into account its correction. Bessel (1823), who discovered the existence of the personal equation, had indeed acted appropriately, since apparently (he did not explain the situation) both observers had been using the same clock.

However, in one case he mistakenly presumed that the rate of the clock was constant, and his pertinent observations proved useless; he made no such comment. When studying Bradley's observations, Bessel (1818; 1838a, § 11) failed to note the deviation of their errors from normality. And 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.

That Gauss had been familiar with the derivation of the CLT could have angered Bessel. Anyway, in 1844, in a letter to Humboldt he (Sheynin 2001c, p. 168) reversed his previous opinion and stressed Legendre's priority in the dispute over the discovery of the MLSq. Moreover, in 1825 Bessel met Gauss and quarrelled with him, although no details are known (Ibidem) and even in 1822 Olbers in a letter to Bessel (Erman 1852, Bd. 2, p. 69) regretted that the relations between the two scholars were bad. Gerling (1861), a former student of Gauss, described Bessel's unwarranted attempts made in 1843 to establish his priority over Gauss in the adjustment of triangulation. See also Sheynin (2001c, pp. 171 – 172).

Bessel's posthumously published collected reports (1848) include an item on the theory of probability (pp. 387 – 407), this being his report to a physical society, written on a low scientific level (apparently occasioned by the poor knowledge of his listeners). Among the applications of the theory of probability Bessel only dwelt on astronomy, but he did not say a single word about the discovery of the minor planets, about the MLSq or Gauss. A distressing impression!

Bessel (p. 401) stated that the *great Lambert* had objected to the use of the arithmetic mean. Actually, Lambert (1760) introduced the principle of maximum likelihood but noted, certainly without proving it that the appearing estimate does not deviate much from the arithmetic mean, the mean which he never denied. Worse is to come. Bessel (1843) stated that Herschel had discovered the planet Uranus, saw its disc. Actually, Herschel only saw a moving body and thought that it was a comet. It follows that Bessel did not know the true story and falsely reconstructed it. Then, he (1845), without any statistical data, invented a false picture about Native Americans.

A great scholar and a deep-rooted fabricator! A case for a psychologist.

7. The Second Half of the 19th Century

7.1. At the beginning of his scientific career **Quetelet** visited Paris and I think that Fourier had mostly inspired him. Quetelet tirelessly treated statistical data and attempted to standardize statistics on an international scale. He was co-author of the first statistical reference book (Quetelet & Heuschling 1865) on the population of Europe (including Russia) and the USA that contained a critical study of the initial data; in 1853, he (1974, pp. 56 – 57) served as chairman of the *Conférence maritime pour l'adoption d'un système uniforme d'observation météorologiques à la mer* and the same year he organized the first *International Statistical Congress*. K. Pearson (1914 – 1930, 1924, vol. 2, p. 420) praised Quetelet for *organizing official statistics in Belgium and* […] *unifying international statistics*. About 1831 – 1833 Quetelet had successfully suggested the formation of a Statistical Society in London, now called the *Royal Statistical Society*.

Quetelet's writings (1869; 1871) contain many dozen of pages devoted to various measurements of the human body, of pulse and respiration, to comparisons of weight and stature with age, etc. and he extended the applicability of the normal law to this field. Following Humboldt's advice, Quetelet (1870; 1871) introduced the term *anthropometry* and thus curtailed the boundaries of anthropology. He was influenced by Babbage (1857), an avid collector of biological data. In turn, Quetelet impressed Galton (1869, p. 26) who called him *the greatest authority on vital and social statistics*. While discussing that contribution, K. Pearson (1914 – 1930, vol. 2, 1924, p. 89) declared:

We have here Galton's first direct appeal to statistical method and the text itself shows [that the English translation of Quetelet (1846)] *was Galton's first introduction to the* […] *normal curve.*

Quetelet (1846) recommended the compilation of questionnaires and the preliminary checking of the data; maintained (p. 278) that too many subdivisions of the data was a *charlatanisme scientifique*, and, what was then understandable, opposed sampling (p. 293). Darwin (1887, vol. 1, p. 341) approvingly cited that contribution whereas

Quetelet (1846, p. 259) declared that *the plants and the animals have remained as they were when they left the hands of the Creator*. Lamarck was the first who attempted to construct a theory of evolution, and Quetelet's statement proves that his thoughts had been more or less discussed. However, Quetelet never mentioned either Lamarck, or Wallace, or Darwin.

He collected and systematized meteorological observations and described the tendency of the weather to persist by elements of the theory of runs. Köppen (1875, p. 256), an eminent meteorologist, noted that *ever since the early 1840s* the Belgian meteorological observations *proved to be the most lasting* [in Europe] *and extremely valuable*.

Quetelet discussed the level of postage rates (1869, t. 1, pp. 173 and 422) and rail fares (1846, p. 353) and recommended to study statistically the changes brought about by the construction of telegraph lines and railways (1869, t. 1, p. 419). He (1836, t. 2, p. 313) quantitatively described the monotone changes in the probabilities of conviction of the defendants depending on their personality (sex, age, education) and Yule (1900/1971, pp. 30 – 32) called it the first attempt to measure association.

Quetelet is best remembered for the introduction of the Average man (1832a, p. 4 and elsewhere), inclinations to crime (1832b, p. 17 and elsewhere) and marriage (1848a, p. 77 and elsewhere), – actually, the appropriate statistical probabilities, – and for mistaken (Rehnisch 1876) statements about the constancy of crime (1829, pp. 28 and 35 and many other sources) whose level he (1836, t. 1, p. 10) connected with the general organization of the society. The two last-mentioned items characterized Quetelet as the follower of Süssmilch in originating moral statistics. Quetelet (1848a, p. 82 and elsewhere) indicated that the inclination to crime of a given person might differ considerably from the apparent mean tendency and (pp. 91 – 92) and related these inclinations to the Average man, but statisticians did not notice that reservation and denied inclinations and even probability theory. True, many of them, e. g., Haushofer (1872) or Block (1878), only applied arithmetic. After Quetelet's death statisticians (mostly in Germany) had simply discarded him.

The Average man, as Quetelet thought, was the type of the nation and even of entire mankind. Reasonable objections were levelled against this concept. Thus, the Average man was physiologically impossible (the averages of the various parts of the human body were inconsistent one with another). Then, Quetelet (1846, p. 216) only mentioned the Poisson LLN in connection with the mean human stature. Bertrand (1888a, p. XLIII) ridiculed Quetelet:

In the body of the average man, the Belgian author placed an average soul. He has no passions or vices [wrong, see above], *he is neither insane, nor wise, neither ignorant nor learned.* […] [He is] *mediocre in every sense. After having eaten for thirty-eight years an average ration of a healthy soldier, he has to die not of old age, but of an average disease that statistics discovers in him*.

However, that concept is useful at least as an average producer and consumer; Fréchet (1949) replaced him by a closely related *typical* man.

Quetelet (1848a, p. 80 and elsewhere) noticed that the curves of the inclinations to crime and to marriage plotted against ages were exceedingly asymmetric. He (1846, pp. 168 and 412 – 424) also knew that asymmetric densities occurred in meteorology and he (1848a, p. viii) introduced a mysterious *loi des causes accidentelles* whose curve could be asymmetric (1853, p. 57)! No wonder Knapp (1872, p. 124) called him *rich in ideas, but unmethodical and therefore unphilosophical*. Nevertheless, Quetelet had been the central figure of statistics in the mid-19th century.

7.2. Being influenced by his cousin, Darwin, **Galton** began to study the heredity of talent (1869). In a letter of 1861 Darwin (1903, p. 181) favourably mentioned that contribution. Darwin (1876/1878, p. 15) also asked Galton to examine his investigation of the advantages of cross-fertilization as compared with spontaneous pollination. Galton solved that problem by effectively applying rank correlation. Then, he (1863) devised an expedient system of symbols for weather charts and immediately discovered the existence of previously unknown anticyclones. This was the third (after Halley and Humboldt, see § 2) example of a wonderful application of a preliminary or exploratory data analysis, the comparatively new stage of statistical investigations. See Andrews (1978) who refers to many authors including J. W. Tukey. In particular, this analysis aims at discovering patterns in the data (including systematic influences). Tukey (1962/1986, p. 397) remarked on an important feature of that stage:

Data analysis, and the parts of statistics which adhere to it, must […] *take on the characteristics of a science rather than those of mathematics*.

Kolmogorov (1948a, p. 216) unfortunately, as I think, stated that mathematical statistics comprised theoretical statistics and a (preliminary) descriptive part devoted to systematizing mass data and to calculating the appropriate means, moments, etc. He himself (Anonymous 1954, pp. 46 – 47) later maintained that theoretical statistics comprises mathematical statistics and *some technical methods of collecting and treating statistical methods*. Many statisticians seem to share this opinion but he belittled these *technical methods* and denied theoretical statistics. Anyway, I cannot agree with Anscombe (1967, p. 3n) who called mathematical statistics *a grotesque phenomenon*.

Galton (Pearson 1914 – 1930, vol. 2, Chapter 12) also invented *composite photographs* of people of a certain nationality or occupation, or criminals, all of them taken on the same film with an appropriately shorter exposure. Such photographs heuristically showed Quetelet's Average man.

In 1892, Galton became the main inventor of fingerprinting. Another of Galton's invention (1877) was the so-called *quincunx*, a device for demonstrating the appearance of the normal distribution as the limiting case of the binomial law which also showed that the normal law was stable. Galton's main statistical merit consisted, however, in the introduction of the notions of regression and correlation. The development of correlation theory became one of the aims of the Biometric school, and Galton's close relations with Pearson were an important cause of its successes.

7.3. I reconstruct now **Darwin's** model of evolution (1859). Introduce an *n*-dimensional (possibly with $n = \infty$) system of coordinates, the body parameters of individuals belonging to a given species (males and females should be treated separately), and the appropriate Euclidean space with the usual definition of distance between its points. At moment t_m each individual is some point of that space and the same takes place at moment t_{m+1} for the individuals of the next generation. Because of the *vertical* variation, these, however, will occupy somewhat different positions. Introduce in addition point (or subspace) V, corresponding to the optimal conditions for the existence of the species, then its evolution will be represented by a discrete stochastic process of the approximation of the individuals to V (which, however, moves in accordance with the changes in the external world) and the set of individuals of a given generation constitutes the appropriate realization of the process. Probabilities describing it (as well as estimates of the influence of habits, instincts, etc) are required for the sake of definiteness, but they are of course lacking.

Mendel's discovery was only unearthed at the very end of the 19th century, and it certainly changed the picture of evolution. Then, the importance of mutation became known (De Vries 1905).

Darwin and his teaching inspired the founders of the Biometric school (§ 8.1).

7.4. In 1855 **Bertrand** had translated Gauss' works on the MLSq into French. The title-page of this translation carried a phrase *Translated and published avec l'autorisation de l'auteur*, but Bertrand himself (*C. r. Acad. Sci. Paris*, t. 40, 1855, p. 1190) indicated that Gauss, who had died that same year, was only able to send him *quelques observations de détail*.

Bertrand's own work on probability began in essence in 1887 – 1888 when he published 25 notes in one and the same periodical as well as his main treatise (1888a), written in great haste and carelessly, but in a very good literary style. I take up its main issues and state right now that it lacks a systematic description of its subject.

1) Statistical probability and the Bayesian approach. Heads appeared $m = 500{,}391$ times in $n = 10^6$ tosses of a coin (p. 276). The statistical probability of that event is $p = 0.500391$; it is unreliable, not a single of its digits merits confidence. After making this astonishing declaration, Bertrand compared the probabilities of two hypotheses, namely, that the probability was either $p_1 = 0.500391$, or $p_2 = 0.499609$. However, instead of calculating

$$[p_1^m p_2^n] \div [p_2^m p_1^n],$$

he applied the De Moivre – Laplace theorem and only indicated that the first probability was 3.4 times higher than the second one. So what should have the reader thought?

As I understand him, Bertrand (p. 161) *condemned* the Bayes *principle* only because the probability of the repetition of the occurrence of an event after it had happened once was too high. This conclusion was too hasty, and the reader was again left in suspense: what might be proposed instead? Note that Bertrand (p. 151) mistakenly thought that the De Moivre – Laplace theorem precisely described the inverse problem, the estimation of the theoretical probability given the statistical data, cf. § 2.4.7.

2) Mathematical treatment of observations. Bertrand paid much attention to this issue, but his reasoning was amateurish and sometimes wrong. Even if, when translating Gauss (see above), he had grasped the essence of the MLSq, he barely remembered that subject after more than 30 years. Thus, he (pp. 281 – 282) attempted to prove that the sample variance (1) of § 6 might be replaced by another estimator of precision having a smaller variance. He failed to notice, however, that, unlike the Gauss' statistic, his new estimator was biased. Furthermore, when providing an example, Bertrand calculated the variance for the normal distribution instead of applying the Gauss additional formula for that case.

At the same time Bertrand also formulated some sensible remarks. He (p. 248) expressed a favourable opinion about the second Gauss justification of the MLSq but indicated (p. 267) that, for small errors, the even distribution

$$\varphi(x) = a + bx^2$$

can be approximately represented by an exponential function of a negative square, – that the first substantiation of the method was approximately valid.

3) Several interesting problems dwell on a random composition of balls in an urn; on sampling without replacement; on the ballot problem; and on the gambler's ruin.

a) White and black balls are placed in the urn with equal probabilities and there are N balls in all. A sample made with replacement contained m white balls and n black ones. Determine the most probable composition of the urn (pp. 152 – 153). Bertrand calculated the maximal value of the product of the probabilities of the sample and of the hypotheses on the composition of the urn.

b) An urn has sp white balls and sq black ones, $p + q = 1$. Determine the probability that after n drawings without replacement the sample will contain $(np – k)$ white balls (p. 94). Bertrand solved this problem applying the [hypergeometric distribution] and obtained, for large values of s and n, an elegant formula

$$P = \frac{1}{\sqrt{2\pi pqn}}\sqrt{\frac{s}{s-n}}\exp[-\frac{k^2 s}{2pqn(s-n)}].$$

He published this formula earlier without justification and noted that the variable probability of extracting the balls of either colour was *en quelque sorte un régulateur*.

c) Candidates *A* and *B* scored *m* and *n* votes respectively, $m > n$ and all the possible chronologically differing voting records were equally probable. Determine the probability *P* that, during the balloting, *A* was always ahead of *B* (p. 18). Following André (1887), who provided a simple demonstration, Bertrand proved that

$$P = (m - n)/(m + n), \qquad (1)$$

see also Feller (1950, § 1 of Chapter 3). Actually, Bertrand was the first to derive formula (1) by a partial difference equation. This *ballot problem* has many applications (Feller, Ibidem). Takácz (1982) traced its history back to De Moivre. He indicated that it was extended to include the case of $m \geq \mu n$ for positive integral values of μ and that he himself, in 1960, had further generalized that extended version.

d) I select one out of the few problems on the gambler's ruin discussed by Bertrand (pp. 122 – 123). Gambler *A* has *m* counters and plays with an infinitely rich partner. His probability of winning any given game is *p*. Determine the probability that he will be ruined in exactly *n* games ($n > m$). Bertrand was able to solve this problem by applying formula (1). Calculate the probability that *A* loses $(n + m)/2$ games and wins $(n - m)/2$ games; then, multiply it by the probability that during that time *A* will never have more than *m* counters, that is, by m/n. Conforming to common sense, Bertrand's derived formula shows that in case of a very high *p* the game will last exceedingly long.

In a brief chapter he largely denied everything done in the *moral applications* of probability by Condorcet (and did not refer to Laplace or Poisson).

In two of his notes Bertrand (1888b; 1888c) came close to proving that for a sample from a normal population the mean and the variance were independent (to the Student – Fisher theorem).

4) I take up Bertrand's celebrated problem about a random chord of a circle in § 7.6.1.

Taken as a whole, Bertrand's treatise is impregnated with its non-constructive negative (and often unjustified) attitude towards the theory of probability and treatment of observations. And at least once he (pp. 325 – 326) wrongly alleged that Cournot had supposed that judges decided their cases independently one from another. I ought to add, however, that Bertrand exerted a strong (perhaps too strong) influence upon Poincaré, and, its spirit and inattention to Laplace and Bienaymé notwithstanding, on the revival of the interest of French scientists in probability (Bru & Jongmans 2001).

7.5. In the theory of probability, **Poincaré** is known for his treatise (1896); I refer to its extended edition of 1912. I note first of all that he had passed over in silence not only the Russian mathematicians, but even Laplace and Poisson, and that his exposition was imperfect.

Following Bertrand, Poincaré (p. 62) called the expectation of a random variable its probable value; denoted the measure of precision of the normal law either by *h* or by $\sqrt{h}$; made use of loose expressions such as *z lies between z and z + dz* (p. 252).

Several times Poincaré applied the formula

$$\lim \frac{\int \varphi(x)\,\Phi^n(x)dx}{\int \psi(x)\,\Phi^n(x)dx} = \frac{\varphi(x_0)}{\psi(x_0)},\ n \to \infty \qquad (2)$$

where Φ(x) was a restricted positive function, x_o, the only point of its maximum, and the limits of integration could have been infinite (although only as the result of a formal application of the Bayesian approach). Poincaré (p. 178) only traced the proof of (2) and, for being true, some restrictions should perhaps be added. To place Poincarè's trick in the proper perspective, see Erdélyi (1956, pp. 56 – 57). I discuss now some separate issues mostly from Poincaré's treatise.

1) The theory of probability. Poincaré (p. 24) reasonably stated that a satisfactory definition of prior probability was impossible. Strangely enough, he (1902/1923, p. 217) declared that *all the sciences* were nothing but an *unconscious application* of the calculus of probability, that the theory of errors and the kinetic theory of gases were based on the LLN (wrong about the former) and that the calculus of probability *will evidently ruin them* (*les entrainerait évidemment dans sa ruine*). Therefore, as he concluded, the calculus was only of practical importance. Another strange pronouncement is in his treatise (p. 34). As I understand him, he maintained that a mathematician is unable to understand why forecasts concerning mortality figures come true.

In a letter of ca. 1899 partly read out at the hearing of the notorious Dreyfus case (*Le procès* 1900, t. 3, p. 325; Sheynin 1991, pp. 166 – 167) Poincaré followed Mill (§ 2.2) and even generalized his statement to include *moral sciences* and declared that the appropriate findings made by Condorcet and Laplace were senseless. And he objected to a stochastic study of handwriting for identifying the author of a certain document.

The interest in application of probability to jurisprudence is now revived. Heyde & Seneta (1977, p. 34) had cited several pertinent sources published up to 1975; to these I am adding Zabell (1988), Gastwirth (2000) and Dawid (2005) who emphasized the utmost importance of interpreting background information concerning stochastic reasoning.

2) Poincaré (1892a) had published a treatise on thermodynamics which Tait (1892) criticized for his failure to indicate the statistical nature of this discipline. A discussion followed in which Poincaré (1892b) stated that the statistical basis of thermodynamics did not satisfy him since he wished to remain *entirely beyond all the molecular hypotheses however ingenious they might be*; in particular, he therefore passed the kinetic theory of gases over in silence. Soon he (1894/1954, p. 246) made known his doubts: he was not sure that that theory can account for all the known facts. In a later popular booklet Poincaré (1905/1970, pp. 210 and 251) softened his attitude: physical laws will acquire an *entirely new aspect* and differential equations will become statistical laws; laws, however, will be shown to be imperfect and provisional.

3) The binomial distribution. Suppose that m Bernoulli trials with probability of success p are made and the number of successes is α. Poincaré (pp. 79 – 84), in a roundabout and difficult way, derived (in

modern notation) $E(\alpha - mp)^2$ and $E|\alpha - mp|$. In the first case he could have calculated $E\alpha^2$; in the second instance he obtained

$$E|\alpha - mp| \approx 2mpq\, C_m^{mp}\, p^{mp} q^{mq},\ q = 1 - p.$$

4) Without mentioning Gauss (1816, § 5), Poincaré (pp. 192 – 194) derived the moments of the [normal] distribution

$$\varphi(y) = \sqrt{h/\pi}\, \exp(-\, hy^2) \tag{3}$$

obtaining

$$Ey^{2p} = \frac{(2p)!}{h^p\, p!\, 2^{2p}} \tag{4}$$

and proved, by issuing from formula (2), that the density function whose moments coincided with the respective moments of the [normal] law was [normal]. This proposition was, however, due to Chebyshev (1887), see also Bernstein (1945/1964, p. 420).

Then Poincaré (pp. 195 – 201) applied his investigation to the theory of errors. He first approximately calculated $E\,\bar{y}^{\,2p}$ for the mean $\bar{y}$ of a large number n of observations having $Ey_i = 0$ and $Ey_i^2 = \text{Const}$, equated these moments to the moments (4) and thus expressed h through Ey_i^2. This was a mistake: $\bar{y}$, being a mean, had a measure of precision nh rather than h. Poincaré (p. 195) also stated that Gauss had calculated $E\,\bar{y}^{\,2}$; actually, Gauss (1823b, §15) considered the mean value of $\sum y_i^2/n$.

The main point here and on pp. 201 – 206, where Poincaré considered the mean values of $(y_1 + y_2 + \ldots + y_n)^{2p}$ with identical and then non-identical distributions and $Ey_i = 0$, was a non-rigorous proof of the CLT: for errors of *sensiblement* the same order and constituting *une faible part* of the total error, the resulting error follows *sensiblement* the Gauss law (p. 206). For Poincaré, the theory of probability was still an applied science as he himself actually stated, see item 1) above.

Also for proving the normality of the sum of errors Poincaré (pp. 206 – 208, only in 1912) introduced characteristic functions which did not conform to their modern definition. Nevertheless, he was able to apply the Fourier formulas for passing from them to densities and back. These functions were

$$f(\alpha) = \Sigma p_x e^{\alpha x},\ f(\alpha) = \int \varphi(x) e^{\alpha x} dx \tag{5}$$

and he noted that

$$f(\alpha) = 1 + \alpha Ex/1! + \alpha^2 Ex^2/2! + \ldots \tag{6}$$

5) Homogeneous [Markov chains]. Poincaré provided interesting examples which can be interpreted in the language of these chains.

a) He (p. 150) assumed that all the asteroids moved along one and the same circular orbit, the ecliptic, and explained why they were uniformly scattered across it. Denote the longitude of a certain minor planet by $l = at + b$ where a and b are random and t is the time, and, by $\varphi(a; b)$, the continuous joint density function of a and b. Issuing from the expectation

$$\mathrm{E}e^{iml} = \iint \varphi(a;b)\, e^{im(at + b)}dadb$$

(which is the appropriate characteristic function in the modern sense), Poincaré not very clearly proved his proposition that resembled the celebrated Weyl theorem (the terms of the sequence $\{nx\}$ where x is irrational and $n = 1, 2, \ldots$ and the braces mean *drop the integral part* are uniformly distributed on a unit interval). The place of a planet in space is only known with a certain error, and the number of all possible arrangements of the asteroids on the ecliptic can therefore be assumed finite whereas the probabilities of the changes of these arrangements during time period $[t; t + 1]$ do not depend on t. The uniform distribution of the asteroids might therefore be justified by the ergodic property of homogeneous Markov chains having a finite number of possible states.

b) The game of roulette. A circle is alternately divided into a large number of congruent red and black sectors. A needle is whirled with force along the circumference of the circle, and, after a great number of revolutions, stops in one of the sectors. Experience proves that the probabilities of *red* and *black* coincide and Poincaré (p. 148) attempted to justify that fact. Suppose that the needle stops after travelling a distance s ($2\pi < s < A$). Denote the corresponding density by $\varphi(x)$, a function continuous on $[2\pi; A]$ with a bounded derivative on the same interval. Then, as Poincaré demonstrated, the difference between the probabilities of *red* and *black* tended to zero as the length of each red (and black) arc became infinitesimal (or, which is the same, as s became infinitely large). He based his proof on the method of arbitrary functions (Khinchin 1961/2004, pp. 421 – 422; von Plato 1983) and sketched its essence. Poincaré also indicated that the rotation of the needle was unstable: a slight change in the initial thrust led to an essential change in the travelled distance (and, possibly, to a change from *red* to *black* or vice versa).

c) Shuffling a deck of cards (p. 301). In an extremely involved manner, by applying hypercomplex numbers, Poincaré proved that after many shuffling all the possible arrangements of the cards tended to become equally probable.

6) Mathematical treatment of observations. In a posthumously published *Résumé* of his work, Poincaré (1921/1983, p. 343) indicated that the theory of errors *naturally* was his main aim in the theory of probability. His statement reflected the situation in those times. In his treatise he (pp. 169 – 173) derived the normal distribution of observational errors mainly following Gauss; then, like Bertrand, he changed the derivation by assuming that not the most probable value of the estimator of the [location parameter] coincided with the arithmetic mean, but its mean value. He (pp. 186 – 187) also noted that,

for small absolute errors $x_1, x_2, \ldots, x_n$, the equality of $f(z)$ to the mean value of $f(x_i)$, led to z, the estimate of the real value of the constant sought, being equal to the arithmetic mean of x_i. It seemed to him that he thus corroborated the Gauss postulate. In the same context Poincaré (p. 171) argued that everyone believed that the normal law was universal: experimenters thought that that was a mathematical fact and mathematicians believed that it was experimental. Poincaré referred to the oral statement of Lippmann, an author of a treatise on thermodynamics.

Finally, Poincaré (p. 188) indicated that the [variance] of the arithmetic mean tended to zero with the increase in the number of observations and referred to Gauss (who nevertheless had not stated anything at all about the case of $n \to \infty$. *Nothing*, however, followed since other linear means had the same property, as Markov (1899a/1951, p. 250) stated. Poincaré himself (pp. 196 – 201 and 217) twice proved the [consistency] of the arithmetic mean. In the second case he issued from a characteristic function of the type of (5) and (6) and passed on to the characteristic function of the arithmetic mean. He noted that, if that function could not be represented as (6), the consistency of the arithmetic mean was questionable, and he illustrated that fact by the Cauchy distribution. Perhaps because of all this reasoning on the mean Poincaré (p. 188) declared that Gauss' rejection of his first substantiation of the MLSq was *assez étrange* and corroborated this conclusion by remarking that the choice of the [parameter of location] should not be made independently from the distribution. Gauss (1823b) came to the opposite conclusion, but he restricted his attention to practically occurring distributions.

Poincaré (pp. 217 – 218) also stated that very small errors made it impossible to obtain absolute precision as $n \to \infty$. More properly, this fact is explained by the non-evenness of the law of distribution, the variability of that law and some interdependence of the observations.

7) Randomness. See § 7.6.2.

Poincaré's almost total failure to refer to his predecessors except Bertrand testifies that he was not duly acquainted with their work. Furthermore: in 1912 he was already able to, but did not apply Markov chains. At the same time, however, he became the author of a treatise that for about 20 years had remained the main writing on probability in Europe. Le Cam's declaration (1986, p. 81) that neither Bertrand, nor Poincaré *appeared to know* the theory was unjust: he should have added that, at the time, Markov was apparently the only one who did master probability.

7.6. Supplement to § 7.4. I ought to discuss Bertrand's problem about the random chord and I seize the opportunity to introduce geometric probability and the notion of randomness.

7.6.1. Geometric Probabilities. These were decisively introduced in the 18th century although the definition of the notion itself only occurred in the mid-19th century. Newton (§ 2.4.3) was the first to think about geometric probability. Beginning with Nikolaus Bernoulli (1709/1975, pp. 296 – 297), see also Todhunter (1865, pp. 195 – 196), each author dealing with continuous laws of distribution effectively applied geometric probability. The same can be said about

Boltzmann (1868/1909, p. 49) who defined the probability of a system being in a certain phase as the ratio of the time during which it is in that time to the whole time of the motion. Ergodic theorems can be mentioned, but they are beyond our boundaries.

However, it was Buffon who expressly studied the new notion. The first report on his work likely written by him himself was Anonymous (1735). Here is his main problem: A needle of length $2r$ falls *randomly* on a set of parallel lines. Determine the probability P that it intersects one of them. It is seen that

$$P = 4r/\pi a \quad (7)$$

where $a > 2r$ is the distance between adjacent lines. Buffon himself had, however, only determined the ratio r/a for $P = 1/2$. His main aim was (Buffon 1777/1954, p. 471) *to put geometry in possession of its rights in the science of the accidental*. Many commentators described and generalized the problem above. The first of them was Laplace (1812/1886, p. 366) who noted that formula (7) enabled to determine [with a low precision] statistically the number π.

A formal definition of the new concept was only due to Cournot (1843, § 18). More precisely, he offered a general definition for a discrete and a continuous random variable by stating that probability was the ratio of the *étendue* of the favourable cases to that of all the cases. We would now replace *étendue* by *measure* (in particular, by area).

Michell (1767) attempted to determine the probability that two stars were close to each other. By applying the Poisson distribution, Newcomb (1859 – 1861, 1860, pp. 137 – 138) calculated the probability that some surface with a diameter of 1° contained s stars out of N scattered "at random" over the celestial sphere and much later Fisher (Hald 1998, pp. 73 – 74) turned his attention to that problem. Boole (1851/1952, p. 256) reasoned on the distinction between a uniform and any other law of distribution:

A random distribution meaning thereby a distribution according to some law or manner, of the consequences of which we should be totally ignorant; so that it would appear to us as likely that a star should occupy one spot of the sky as another. Let us term any other principle of distribution an indicative one.

His terminology is now unsatisfactory, but his statement shows that Michell's problem had indeed led to deliberations of a general kind.

Determine the probability that a random chord of a given circle is shorter than the side of an inscribed equilateral triangle (Bertrand 1888a, p. 4). This celebrated problem had been discussed for more than a century and several versions of *uniform randomness* were studied. Bertrand himself offered three different solutions, and it was finally found out that, first an uncountable number of solutions was possible, and, second, that the proper solution was *probability equals* 1/2 and I note that it corresponded to *la perfaite perplexité de notre esprit* (§ 5). Thus ended the protracted discussion.

For a modern viewpoint on geometric probability see Kendall & Moran (1963); in particular, following authors of the 19th century (e.g., Crofton 1869, p. 188), they noted that it might essentially simplify the calculation of integrals. Then, Ambartzumian (1999) indicated that geometric probability and integral geometry are connected with stochastic geometry.

7.6.2. Randomness is a fundamental notion which inevitably enters statistics. For a popular discussion of recent mathematical efforts to define it, see Chaitin (1975). The history of that notion begins in antiquity; Aristotle and other early scientists and philosophers attempted to define, or at least to throw light upon randomness. His examples of random events are a sudden meeting of two acquaintances (*Phys*. 196b30) and a sudden unearthing of a buried treasure (*Metaphys*. 1025a). In both cases the event occurred without being aimed at and in addition they illustrate one of Poincaré's explanations (interpretations) of randomness (1907), then incorporated in his popular book (1908) and in his treatise (1912/1987, p. 4): if equilibrium is unstable,

A very small cause which escapes us determines a considerable effect […] *and we say that that effect is due to chance*.

Many authors had been repeating Aristotle's first example and Cournot's (1843, § 40) explanation can also be cited:

Events occurring as a combination or meeting of phenomena which apparently belong to independent series [but] *happening as ordered by causality, are called fortuitous, or results of hazard*.

Poincaré could have mentioned a coin toss. His deliberations (also see below) heralded the beginning of the modern period of studying randomness. However, Poincaré certainly had predecessors who only failed to mention directly randomness. Among them was the ancient physician Galen (1951, p. 202): I*n old men even the slightest causes produce the greatest change*; Pascal (1669/2000, p. 675): *Had Cleopatra's nose been shorter, the whole face of the Earth would have changed*; and Maxwell (1873a/1882, p. 364) who referred to the unstable refraction of rays within biaxial crystals. Elsewhere he (1859/1927, p. 295 – 296) left a most interesting statement:

There is a very general and very important problem in Dynamics. [...] *It is this: Having found a particular solution of the equations of motion of any material system, to determine whether a slight disturbance of the motion indicated by the solution would cause a small periodic variation, or a total derangement of the motion*.

Given a large number of births, regularities of such mass random events will, however, certainly reveal themselves but Aristotle did not connect such events with randomness. *Corruption of*, or *deviation from laws of nature* also means randomness, and this idea can be traced at least until Lamarck who stated that the deviations from the divine lay-out of the tree of animal life had been occasioned by a *cause accidentelle* (Lamarck 1815, p. 133).

There also, on p. 173, he indicated that the spontaneous generation of organisms was caused by a *très-irrégulière* force but did not mention randomness. When considering the state of the atmosphere, Lamarck (1800 – 1811/1800, p. 76) stated that it was disturbed by two kinds of causes, including *variables, inconstantes et irrégulières*. Again, no mention of randomness, but then he (1810 – 1814/1959, p.

632) denied it: no *part of nature* disobeys *invariable laws*; therefore *that, which is called chance*, does not exist.

Louis Pasteur definitively disproved spontaneous generation, but until then it was apparently always considered random. Witness indeed Harvey (1651/1952, p. 338):

Creatures that arise spontaneously are called automatic […] *because they have their origin from accident, the spontaneous act of nature*.

Harvey did not say anything about the essence of accidents, but it seems that he thought them aimless, identified them with lack of law. Many other scientists denied randomness as Lamarck did.

I will now mention Laplace (1814/1995, p. 9) who stated that the arrangement of printed letters in the word *Constantinople is not due to chance*; all arrangements are equally unlikely, but that word has a meaning and it is *incomparably more probable* that someone had written it on purpose. He equated randomness with lack of purpose. This example shows that human judgement is needed for supplementing mathematical reasoning about randomness; intersection of events (above) can be additionally interpreted as lack of purpose.

Poincaré (1896/1912, p. 1) also formulated a dialectical statement about determinism and randomness much broader than the one following from *deviation from laws of nature*: it legitimizes randomness and indirectly defines it but does not say anything about regularities of mass random events:

In no field [of science] *do exact laws decide everything, they only trace the boundaries within which randomness is permitted to move. According to this understanding, the word* <u>*randomness*</u> *has a precise and objective meaning*.

He thus restricted the action of his pattern *small cause – considerable effect*. Exact laws tolerate randomness, cf. Newton's statement about the system of the world (§ 2.4.3). He recognized randomness, although this time only in its *uniform* version as witnessed by the expression *blind fate*. Whether in English, or in equivalent French and German terms, scientists of the 17th and 18th centuries, if discussing randomness, mostly understood it in this sense. For example, Arbuthnot (§ 1.3.1-1), only compared Design with a discrete uniform distribution of the sexes of the newborn babies.

Maupertuis (1745/1756, pp. 120 – 121) indicated that the seminal liquid of *chaque individu* most often contained *parties* similar to those of their parents, but he (p. 109) also mentioned rare cases of a child resembling one of his remote ancestors as well as mutations (p. 121, a later term). It seems that Maupertuis thus recognized randomness with a multinomial distribution, but, when discussing the origin of eyes and ears in animals, he (1751/1756, p. 146) only compared *une attraction uniforme & aveugle* [blind] and *quelque principe d'intelligence* (and came out in favour of design).

A chaotic process engendered by a small corruption of the initial conditions of motion can lead to its exponential deviation. Only in a sense this may be understood as an extension of Poincaré's pattern *small cause – considerable effect*. However complicated and protracted is a coin toss, it has a constant number of outcomes whose probabilities persist, whereas chaotic motions imply rapid increase of their instability with time and countless positions of their possible paths. Their importance in mechanics and physics is unquestionable. My explanation of the comparatively new concept is only qualitative, but I have not seen any better.

In statistics, a random variable should be statistically stable, but in natural science this restriction is not necessary. Lamarck (see above) provided a good example of the latter phenomena: the deviations from the divine lay-out of the tree of animal life. Kolmogorov (1983/1992, p. 515) properly stated:

We should distinguish between randomness in the wider sense (absence of any regularity) and stochastic random events (*which constitute the subject of probability theory*).

There seems to be no quantitative criteria of statistical stability which apparently characterizes observations belonging to a single law of distribution, to a single population. However, practice often has to work in its absence; example: sampling estimation of the content of the useful component in a deposit. Choose other sample points, and it will be unclear whether they possess the same statistical properties (Tutubalin 1972/2011, § 1.2). But, according to scientific folklore, pure science achieves the possible by rigorous methods, whereas applications manage the necessary by possible means.

I provide now an example of a false conclusion caused by lack of statistical stability of the considered deviations. William Herschel (1817/1912, p. 579), who certainly knew nothing either about the size of stars or of their belonging to different spectral classes, decided that the size of a randomly chosen star will not much differ from the mean size of all of them. The sizes of stars are enormously different and their mean size is a purely abstract notion. There are stars whose radii are greater than the distance between the Sun and the Earth. Again, *ex nihilo nihil fit*.

Earlier, De Moivre (1733/1756, pp. 251 – 252) refused to admit randomness in the wide sense:

Absurdity follows, if we should suppose the Event not to happen according to any Law, but in a manner altogether desultory and uncertain; for then the Event would converge to no fixt Ratio at all.

8. The First Half of the 20th Century

8.1. Karl Pearson (1857 – 1936) was an applied mathematician and philosopher and the creator of biometry, the main branch of what later became mathematical statistics.

Pearson studied physics on which he expressed some extremely interesting ideas. Thus, *negative matter* exists in the universe (1891, p. 313); *all atoms in the universe of whatever kind appear to have begun pulsating at the same instant* (1887, p. 114) and *physical variations effects* were perhaps *due to the geometrical construction of our space* (Clifford 1885/1886, p. 202). He did not, however, mention Riemannian spaces whereas it is nowadays thought that the curvature of space-time is caused by forces operating in it. Remarkable also was Pearson's idea (1892, p. 217) about the connection of time and space subjectively expressed as:

Space and time are so similar in character, that if space be termed the breadth, time may be termed the length of the field of perception.

Mach (1897), in his Introduction, mentioned K. P. in the first edition of his book which appeared after 1892:

The publication [of the *Grammar of Science*] *acquainted me with a researcher whose* *erkenntnisskritischen* [Kantian] *ideas on every important issue coincide with my own notions and who knows how to oppose, candidly and courageously, extra-scientific tendencies in science*.

Again in the same contribution we find Pearson's celebrated maxim (1892, p. 15): *The unity of all science consists alone in its method, not in its material*. I return to this statement in § 9. Here, I indicate that Pearson, a Fellow of the Royal Society since 1896, was unable to take up the invitation of Newcomb, the president of the forthcoming International Congress of Arts and Sciences (St. Louis, 1904), to deliver there a talk on methodology of science (Sheynin 2002, p. 163, note 8).

At the very end of the 19th century, by founding the celebrated *Biometrika*, Galton, Pearson and Weldon (who died in 1906) established the Biometric school which aimed at the creation of methods of treating biological observations and of studying statistical regularities in biology. Pearson became the chief (for many years, the sole) editor of that periodical. In the Editorial, in its first issue of 1902, we find a reference to Darwin:

[*E*]*very idea of Darwin – variation, natural selection* […] *– seems at once to fit itself to mathematical definition and to demand statistical analysis.*

K. P. compiled contributions on Weldon (1906) and on Galton's life and achievements, a fundamental and most comprehensive tribute to any scholar ever published (1914 – 1930). Incidentally, Chr. Bernoulli (1841, p. 389) had coined the word *Biometric* (in German) which referred to mass observations.

The immediate cause for establishing *Biometrika* seems to have been scientific friction and personal disagreement between Pearson and Weldon on the one hand, and biologists, especially Bateson, on the other hand, who exactly at that time had discovered the unnoticed Mendel. It was very difficult to correlate Mendelism and biometry: the former studied discrete magnitudes while the latter investigated continuous quantitative variations. However, in 1926 Bernstein (Kolmogorov 1938, § 1) proved that under wide assumptions the Galton law of inheritance of quantitative traits was a corollary of the Mendelian laws.

The speedy success of the Biometric school had been to a large extent prepared by the efforts of Edgeworth (1845 – 1926); his collected writings appeared in 1996. Pearson's results in statistics include the development of the elements of correlation theory and contingency; introduction of the *Pearsonian curves* for describing empirical distributions; and a derivation of a most important chi-squared test for checking the correspondence of experimental data with one or other law of distribution, as well as the compilation of many important statistical tables.

Pearson's posthumously published lectures (1978) examined the development of statistics in connection with religion and social conditions of life. On the very first page we find the statement about the importance of the history of science: *I do feel how wrongful it was to work for so many years at statistics and neglect its history*. However, he provided a false appraisal of the Bernoulli LLN (§ 2.4.5).

Pearson attempted, often successfully, to apply the statistical method, and especially correlation theory, in many branches of science. Here is his interesting pronouncement (1907, p. 613):

I have learnt from experience with biologists, craniologists, meteorologists, and medical men (who now occasionally visit the biometricians by night!) that the first introduction of modern statistical method into an old science by the layman is met with characteristic scorn; but I have lived to see many of them tacitly adopting the very processes they began by condemning.

It is interesting to note the different views held of K. P. by other scientists. Kolmogorov (1947, p. 63) stated that

The modern period in the development of mathematical statistics began with the fundamental works of English statisticians (*K. Pearson, Student, Fisher*) *which appeared in the 1910s, 1920s and 1930s. Only in the contributions of the English school did the application of probability theory to statistics cease to be a collection of separate isolated problems and became a general theory of statistical testing of stochastic hypotheses (i. e., of hypotheses about laws of distribution) and of statistical estimation of parameters of these laws.*

Kolmogorov (p. 64 of same paper) had not then duly appreciated Fisher, and here is his possible explanation:

The investigations made by Fisher, the founder of the modern British mathematical statistics, were not irreproachable from the standpoint of logic. The ensuing vagueness in his concepts was so considerable, that their just criticism led many scientists (*in the Soviet Union, Bernstein*) *to deny entirely the very direction of his research*.

A year later Kolmogorov (1948b/2002, p. 68) criticized the Biometric school:

Notions held by the English statistical school about the logical structure of the theory of probability which underlies all the methods of mathematical statistics remained on the level of the eighteenth century.

Fisher (1922, p. 311) expressed similar criticisms as did Chuprov (Sheynin 1990/2011, p. 149); Chuprov (Ibidem) informed his correspondents that Continental statisticians (especially Markov) did not wish to recognize Pearson.

Here are some other opinions about Pearson.

1) Bernstein (1928/1964, p. 228), when discussing *a new cycle of problems in the theory of probability which comprises the theories of distribution and of the general non-normal correlation*, wrote:

From the practical viewpoint the Pearsonian English school is occupying the most considerable place in this field. Pearson fulfilled an enormous work in managing statistics; he also has great theoretical merits, especially since he introduced a large number of new concepts and opened up practically important paths of scientific research. The *justification and criticism of his ideas is one of the central problems of current mathematical statistics. Charlier and Chuprov, for example, achieved considerable success here whereas many other statisticians are continuing Pearson's practical work, definitely losing touch with probability theory …*

2) Fisher, letter of 1946 (Edwards 1994, p. 100):

He was singularly unreceptive to and often antagonistic to contemporary advances made by others in [his] *field*. [Otherwise] *the work of Edgeworth and of Student, to name only two, would have borne fruit earlier*.

Fisher (1937, p. 306) also accused Pearson: his *plea of comparability* [between the methods of moments and maximum likelihood] *is* […] *only an excuse for falsifying the comparison* […]. Pearson died in 1936, but his son, Egon, kept silent.

3) But there are also testimonies of a contrary nature: Mahalanobis, in a letter of 1936 (Ghosh 1994, p. 96): he *always looked upon* [K. P.] as his *master,* and upon himself*, as one of his humble disciples*. And Newcomb, who had never been Pearson's student, wrote in a letter of 1903 to him (Sheynin 2002, p. 160):

You are the one living author whose production I nearly always read when I have time and can get at them, and with whom I hold imaginary interviews while I am reading.

4) Hald (1998, p. 651) offered a reasonable general description of one aspect of the Biometric school:

Between 1892 and 1911 he [Pearson] *created his own kingdom of mathematical statistics and biometry in which he reigned supremely, defending its ever expanding frontiers against attacks*. […] He *was not a great mathematician, but he effectively solved the problems head-on by elementary methods*.

5) Fisher (1956/1990, p. 3), however, ungenerously criticized Pearson for the *weakness of his mathematical and scientific work*.

In Russia, Chuprov and Slutsky defended Pearson's work against Markov's opposition (Sheynin 1990/2011, §§ 7.4 and 7.6). Chuprov wished to unite the Continental direction of statistics with biometry, but did not achieve real success.

Lenin's criticism of Pearson was in itself a sufficient cause of the negative Soviet attitude towards Pearson. Maria Smit's statement (1934, pp. 227 – 228) was its prime example: his curves are based

On a fetishism of numbers, their classification is only mathematical. Although he does not want to subdue the real world as ferociously as it was attempted by […] *Gaus* [Smit's spelling], *his system nevertheless only rests on a mathematical foundation and the real world cannot be studied on this basis at all*.

In 1931 this troglodyte (Corresponding member of the Soviet Academy of Sciences since 1939!) declared: *The crowds of arrested saboteurs are full of statisticians* (Sheynin 1998, p. 533, literal translation). She likely participated in enlarging that *crowd*.

However, the tone of the item *Pearson*, in the third edition of the *Great Sov. Enc.* (vol. 19, 1975/English edition: same volume, 1978, p. 366) was quite different: he *considerably contributed to the development of mathematical statistics* and Lenin had only criticized his *subjective-idealistic interpretation of the nature of scientific knowledge*.

8.2. Markov is known to have opened up a new direction of probability theory dealing with dependent events, and in particular, to have studied the *Markov chains*. At the same time, he refused to apply his chains to problems in natural sciences, did not apply the allegedly meaningless term *random magnitude* (as it is still called in Russia) and, similarly, the expressions *normal law* and *coefficient of correlation* were absent in his works. Like a student of Chebyshev that he was, he underrated the then emerging axiomatic approach to probability as well as the theory of functions of a complex variable (A. A. Youshkevich 1974, p. 125).

During his last years, in spite of extremely difficult conditions of life in Russia and his worsened health, he completed (perhaps not entirely) the last posthumously published edition of his *Treatise* but insufficiently described there the findings of the Biometric school; such scholars as Yule and Student (Gosset) were not mentioned and he (1900/1924, pp. 10, 13 – 19 and 24) even wrongly stated that he transferred probability to the realm of pure mathematics just by proving the addition and multiplication theorems. Actually, to some extent he became a victim of his own rigidity; he failed, or did not wish to notice the new tide of opinion in statistics, or even probability theory, see Sheynin (2006) and the text above.

Markov (1888) compiled a table of the normal distribution which gave it to 11 digits for the argument x = 0 (0.001) 3 (0.01) 4.8. Two such tables, one of them Markov's, and the other, published ten years later, remained beyond compare up to the 1940s (Fletcher et al 1946/1962).

Markov included some innovations in the last edition of his *Treatise*: a study of statistical series, linear correlation. He determined the parameters of lines of regression, discussed random variables possessing certain densities and included a reference to Slutsky (1912), whom he previously barely recognized, but paid no attention either to the chi-squared test or to the Pearsonian curves.

The so-called Gauss – Markov theorem *invented* by Lehmann (1951), who followed Neyman's mistake (which he himself later acknowledged), never existed.

8.3. The Continental Direction of Statistics. At the end of the 19th, and in the beginning of the 20th century, statistical investigations on the Continent were chiefly restricted to the study of population whereas in England scientific statistics was mostly applied to biology. The so-called Continental direction of statistics originated as the result of the work of Lexis whose predecessors had been Poisson, Bienaymé, Cournot and Quetelet. Poisson and Cournot examined the significance of statistical discrepancies for a large number of observations without providing examples. Cournot also attempted to reveal dependence between the decisions reached by judges (or jurors). Bienaymé (1839) was interested in the change in statistical indicators from one series of trials to the next one and Quetelet (§ 7.1) investigated the connections between causes and effects in society, attempted to standardize statistical data worldwide and, following Süssmilch (§ 1.3), created moral statistics.

At the same time statisticians held that the theory of probability was only applicable to statistics if *equally possible cases* were in existence, and the appropriate probability remained constant (§ 2.4.5).

8.3.1. Lexis (1879) proposed a distribution-free test for the equality of probabilities in different series of observations; or, a test for the stability of statistical series. Suppose that there are m series of n_i observations, $i = 1, 2, \ldots, m$, and that the probability of success p was constant throughout. If the number of successes in series i was a_i, the variance of these magnitudes could be calculated by two independent formulas (Lexis 1879, § 6)

$$\sigma_1^2 = pqn, \quad \sigma_2^2 = [vv]/(m-1) \qquad (1; 2)$$

where n was the mean of n_i, v_i, the deviations of a_i from their mean, and $q = 1 - p$. Formula (2) was due to Gauss, see (§ 6); he also knew formula (1), see Gauss, W-8, p. 133. The frequencies of success could also be calculated twice. Note however that Lexis applied the probable error rather than the variance and mistakenly believed that the relation between the mean square error and the probable error was distribution-free. Lexis (§ 11) called the ratio

$$Q = \sigma_2/\sigma_1$$

the *coefficient of dispersion*. For him, the case $Q = 1$ corresponded to normal dispersion (with admissible random deviations from unity); he called the dispersion supernormal, and the stability of the observations subnormal if $Q > 1$ (and indicated that the probability p was not then constant); finally, Lexis explained the case $Q < 1$ by dependence between the observations, called the appropriate variance subnormal, and the stability, supernormal. He did not, however, pay attention to this possibility. His coefficient was the ratio of the appearance of the studied event as calculated by the Gauss formula to that peculiar to the binomial distribution.

Lexis hardly thought about calculating the mean value and variance of Q (and in any case that was a serious problem). In 1916, Markov, and, much better, Chuprov derived EQ and, in a manuscript of 1916 or 1917, Chuprov derived the mean square error of Q.

8.3.2. In 1910 Markov and **Chuprov**, in their letters to each other (Ondar 1977), proved that some of the Lexian considerations were wrong. Then, in 1918 – 1919, Chuprov formulated the shortcomings of Q as a criterion but, strangely enough, he somehow kept to the Lexian theory until 1921. Indeed, in a letter of 30 Jan. 1921 to a friend Chuprov wrote:

One of the most important doctrines of theoretical statistics, which I until now entirely accepted and professed, the Lexian theory of stability of statistical figures is to a large extent based on a mathematical misunderstanding.

Concerning this paragraph see Sheynin (1990/2011, pp. 140 – 143).

The refutation of those Lexian considerations was apparently barely noticed. Bernstein (1928/1964, p. 224) called them *the first important step of the scientific treatment of statistical materials* and even much later Särndal (1971, pp. 376 – 377) who described this subject did not mention any criticisms of Lexis. As it seems, Bernstein also positively although obliquely referred to the non-existing Bortkiewicz' law of small numbers (1898): *Poisson's investigations had been recently specified and essentially supplemented.* Yes, Lexis thought of basing statistical investigations on a stochastic foundation (although so did Jakob Bernoulli), and he also made a forgotten attempt to define stationarity and trend.

In a paper devoted to the application of probability theory to statistics, Lexis (1886, pp. 436 – 437) stated that the introduction of equipossibility led to the subjectivity of the theory of probability. He did not say that the existence of equally possible cases was not necessary. These cases haunted him (Lexis 1913, p. 2091).

8.3.3. Bortkiewicz had introduced his own test, Q', not coinciding with the Lexian Q, and equal to the ratio of two dependent random variables, call them ξ and η. Unlike Q, Q' could not be less than 1 (1898, p. 31). Later Bortkiewicz (1904, p. 833) noted that $EQ = Q'$ but mistakenly justified this equality by believing that, for those dependent variables, $E\xi/\eta = E\xi/E\eta$. Then, he (1918, p. 125n) unjustifiably admitted that the equality was only insignificantly approximate. Chuprov (1922) devoted a paper to that subject.

See my discussion (1990/2011, pp. 59 – 62) of the Lexian innovation. In particular, I quoted Bortkiewicz' letter to Chuprov of 29 March 1911: *Poisson cannot at all be considered the own father of the law of large numbers* since he, Bortkiewicz, did not regard a low level of the probability of the studied event as the *decisive point.* Rarity, he continued, can mean a small number of occurrences of that event when the number of trials was also small. He thus undermined his alleged law! Delicate Chuprov did not comment.

8.3.4. The Two Statistical Streams. Bauer (1955, p. 26) investigated how the Biometric school and the Continental direction of statistics had been applying analysis of variance and concluded (p. 40) that their work was going on side by side but did not tend to unification. For more details about Bauer`s study see Heyde & Seneta (1977, pp. 57 – 58) where it also correctly indicated that, unlike the Biometric school, the Continental direction had concentrated on nonparametric statistics. Chuprov can be certainly mentioned here. He achieved some important results; for example, he discovered finite exchangeability (Seneta 1987).

However, his formulas, *being of considerable theoretical interest*, were *almost useless* due to complicated calculations involved (Romanovsky 1930, p. 216). In addition, he had not paid due attention to notation. Thus, in one case he (1923, p. 472) applied two-storey superscripts and two-storey subscripts in the same (five-storey!) formula. Hardly has any other author (not even Bortkiewicz) allowed himself to take such liberties, to expect his readers to understand suchlike monsters.

For his part, Bortkiewicz just had not respected his readers. Winkler (1931, p. 1030) quoted his letter (but did not provide its date) in which Bortkiewicz mentioned that he expected to have five readers of his (unnamed by Winkler) publication. Statisticians had not been mathematically educated and despised mathematics; for them, Bortkiewicz remained an alien body.

I myself (Gnedenko & Sheynin 1978/2001, p. 275), probably following other authors, suggested that mathematical statistics properly originated as the coming together of the two streams. However, now I correct myself. At least until the 1920s, say, British statisticians had continued to work all by themselves. E. S. Pearson (1936 – 1937), in his study of the work of his father, had not commented on Continental statisticians and the same is true about other such essays (Mahalanobis 1936; Eisenhart 1974). I believe that English, and then American statisticians for the most part only accidentally discovered some findings already made by the Continental school. Furthermore, the same seems to happen nowadays as well. Even Hald (1998) called his book *History of Mathematical Statistics*, but barely studied the work of that school.

In 1919 there appeared in *Biometrika* an editorial entitled *Peccavimus*! (we were guilty). Its author, Pearson, corrected his mathematical and methodological mistakes made during several years and revealed mostly by Chuprov (Sheynin 1990a/2011, p. 75) but he had not taken the occasion to come closer to the Continental statisticians. In 2001, five essays were published in *Biometrika*, vol. 88, commemorating its centenary. They were devoted to important particular issues, but nothing was said in that volume about the history of the Biometric school, and certainly nothing about Continental statisticians.

8.3.5. Statistics and Sociology in the Soviet Union. Concerning the general situation there, see Sheynin (1998).

Sociology studies society, its institutions, population, existing tendencies and attempts to discern possible developments. Statistics is certainly essential for such investigations, and many statisticians from Graunt to Quetelet to modern specialists can be cited here. Here, I am only concerned with the year 1954 and begin by quoting two authors (Schlözer 1804, p. 51) and Truesdell (1981/1984, pp. 115 – 117) who invented two terms, *plebiscience* which describes modern times and *prolescience* of the future:

Statistics and despotism are incompatible.

Prolescience will *confirm and comfort the proletariat in all that will by then have been ordered to believe*. […] *That will be mainly social science*.

Süssmilch attempted to reveal divine order in demography, and official Soviet statistics regarded statistics as a discipline reduced to corroborate quantitatively Marxist propositions. Many participants in a statistical conference held in Moscow in 1954 voiced that opinion (Anonymous 1954; see also Kotz 1965; Sheynin 1998, pp. 540 – 541).

Only *the revolutionary Marxist theory* is the basis for developing statistics as a social science (p. 41); statistics does not study mass random phenomena (p. 61) which anyway possess no special features

(p. 74); the LLN is not a mathematical proposition (p. 64); probability is not the necessary basis of statistics, the theory of stability of statistical series is a *bourgeois theory* and even honest *bourgeois* statisticians are compelled to violate their professional duty (p. 46, the notorious Maria Smit, see § 8.1).

K. V. Ostrovitianov (p. 82), the vice-president of the Academy of Sciences, ignorantly declared that Lenin had *completely subordinated* [adapted] *the statistical methods of research* […] *to the class analysis of the rural population*. And, as he menacingly continued, the same scientific methods cannot be used in astronomy and economics.

His latter statement directly contradicted Kolmogorov's (pp. 46 – 47) definition of mathematical statistics who also mentioned several *safe* areas of application of the statistical method (studies of the work of telephone exchanges, management of life insurance, determination of necessary stocks of foodstuffs) but omitted population statistics. This subject was dangerous. The census of 1937 was proclaimed worthless and followed by a decimation of the Central Statistical Directorate: it revealed a demographic catastrophe occasioned by arbitrary rule, uprooting of millions, mass hunger and savage witch-hunt. And the war losses had to be hushed up.

Much later, still in accord with the resolution of the conference, Riabushkin (1980, p. 498) argued that statistical descriptions should be inseparably bound with life's *qualitative content*. Ten more years had to pass before Orlov (1990) *rejected* the decisions of that conference, revealed the falsifications of Soviet statistics and its backwardness (certainly known abroad).

9. The Unity of Statistics Consists Alone in Its Method.

Schlözer (1804) called his book *Theory of statistics*, but it did not contain any theory in our sense. Bearing in mind other authors of the first half of the 19th century, I believe that in those times theory of statistics meant a systematic arrangement of statistical data according to reasonably chosen indicators.

For that matter, even Achenwall had a theory (of Staatswissenschaft) in that same sense, and, as it seems, so did Delambre (1819, p. LXVII) and the London Statistical Society (Anonymous 1839, p. 1). Delambre argued that statistics ought not to engage in discussions or conjectures or to aim at perfecting theories, and that Society declared that statistics does not *discuss causes* nor *reason upon probable effects*. True, these *absurd restrictions* have been necessarily *disregarded* (Woolhouse 1873, p. 39), − I would say, they became obsolete, but no theory of statistics had yet emerged.

The very title of Dufau (1840) called statistics the theory of studying the laws to which the social events are developing. And, without mentioning any theories, a kindred idea was pronounced much earlier (Gatterer 1775, p. 15): Just as in history it is necessary *to investigate not only the Pourquoi, but also the Pourquoi of the Pourquoi, so it is necessary in statistics to explain the present state of a nation by its previous states*.

This *Pourquoi of the Pourquoi* likely came from Sophie Charlotte, Queen of Prussia, apparently from her letter to Leibniz (Krauske 1892, p. 682). Cf. Cournot (1843, § 106):

The essential goal of the statistician, just like of any other observer, is to penetrate as deeply as possible into the knowledge of the essence of things.

Perhaps Cauchy (1845/1896, p. 242) can also be cited: statistics was *infallible* in judging doctrines and institutions.

Here is how Chuprov's student and the last representative of the Continental direction, Anderson (1932, p. 243), described the previous situation of the application of probability in statistics:

Our (*younger*) *generation of statisticians is hardly able to imagine that mire in which the statistical theory had got into after the collapse of the Queteletian system, or the way out of it which only Lexis and Bortkiewicz have managed to discover*.

But did they (or Chuprov, whom Anderson later added to them) really overcome the occurring difficulty? Did they convince statisticians? In any case, the situation changed only gradually. Only in the mid-20th century Neyman (1950, p. 4), Mises (1964a, posthumous, p. 1) and Kendall (1978, p. 1093) stated that mathematical statistics (a section of the theory of probability, as the two first authors held) was the mathematical theory of statistics. The relations between probability theory and mathematical statistics does not directly bear on statistics and I only note that Kolmogorov (1948a, p. 216) thought that *the theory of probability must be considered the structural part of mathematical statistics,* but that (p. 218) *statistics only gradually ceases to be the applied theory of probability*. And (p. 216) mathematical statistics is a *science of the mathematical methods of studying mass phenomena.*

Later, however, Kolmogorov (Anonymous 1954, pp. 46 – 47) only declared that mathematical statistics is not an applied theory of probability. Then, *mass phenomena* is too restrictive. Anyway, much later Kolmogorov provided quite another definition of mathematical statistics, see below.

The following two definitions should perhaps be altered by substituting *theory of statistics* instead of *statistics* and *statistical data* instead of *mass observations*; they both will then be in line with the definitions above.

Fisher (1925, p. 1) argued that statistics *is a branch of applied mathematics and may be regarded as mathematics, applied to observational data*. K. Pearson (1978, p. 3) stated that *statistics is the application of mathematical theory to the interpretation of mass observations*.

Alph. De Candolle (1833, p. 334) and Chaddock (1925, p. 26) thought that statistics is a branch of mathematics. Here also, this rather incomplete definition can be altered to conform to those of Neyman, Mises and Kendall.

According to the comparatively new definition of Kolmogorov & Prokhorov (1988/1990, p. 138),

Mathematical statistics is a branch of mathematics devoted to systematizing, processing and utilizing statistical data, or information on the number of objects in some more or less extensive collection that have some specific properties.

They (p. 139) also argued that *the method of research, characterized as the discussion of statistical data,* […] *is called statistical* and consists in calculating the number of objects *in some group or other*, in discussing the distribution of quantitative indicators, applying the method of sampling and estimating the adequacy of the number of observations etc (p. 139).

Kolmogorov & Prokhorov's definition apparently excluded the theory of errors and in addition it remains unclear whether the *information* was raw or corrected, either initially or during *systematization* by means of exploratory data analysis, − whether they incorporated that stage of work into mathematical statistics. See § 7.2 on the difference between mathematical and theoretical statistics.

Many definitions are more or less akin to theirs, although their authors sometimes discuss statistics instead of theory of statistics or mathematical statistics. Thus (Butte 1808, p. XI),

Statistics is a science of the art [science and art] *of the knowledge and due estimation of statistical data, of their collection and systematic analysis*.

Zhuravsky (1846, p. 173): statistics is a calculus of categories, which distributes objects among the categories and counts them in each category. He thought that *statistics is a special and very wide science*. Maxwell (1871/1890, vol. 2, p. 253; 1877, p. 242) defined the statistical method as an *estimation of an average condition of a group of atoms*, as *a study of the probable number of bodies in each group* under investigation.

Some modern definitions have been offered by Egon Pearson (Bartholomew 1995, p. 7), Kendall (1950, p. 130), Kendall & Buckland (1971), Marriot (1991), Bancroft (1966, p. 530), Kruskal (1978, p. 1072), Wilks (1968, p. 162), anonymous authors (1968, p. 166; 1985, p. 230) and Dodge (2003, p. 388).

The first two definitions are rather abstract as also, to a lesser extent, is the fourth one; the others have much in common with Kolmogorov & Prokhorov's. And here is Dodge: *statistics is a science of collecting, analysing and interpreting* the data (the *numerical* information *relating to an aggregate of individuals*).

Several authors have preferred a narrower and therefore hardly sufficient definition of statistics. Chuprov, in his unpublished thesis of 1896 (Sheynin 1990/2011, p. 118), as well as Lindley (1984, p. 360) and Stigler (1986, p. 1) believed that it measures our ignorance or uncertainty. And Chernoff & Moses (1959, p. 1) even stated that

Today's statistician will be more likely to say that statistics is concerned with decision making in the face of uncertainty (than with *processing* of data).

Cf. Mahalanobis' statement of 1950 (Rao 1993, p. 339): *The aim of statistics is to reach a decision on a probabilistic basis, on available evidence*. And Bancroft (1966), remarked that *statistical inferences are made in the face of uncertainty*.

Several authors held that statistics is only a method (Fox 1860, p. 331; Miklashevsky 1901, p. 476). Alph. De Candolle (1873, p. 12) reversed his own much earlier opinion, agreed with that statement and even contrasted statistics with mathematics mistakenly arguing that the latter (only) provided deterministic conclusions.

It is time to formulate my own conclusions.

1. Statistics and statistical method: at the end of § 2 I noted that these terms are (sometimes) understood as synonyms. More precisely, the statistical method is almost the same as mathematical statistics or theory of statistics.

2. Such expressions as *stellar* or *medical statistics* mean the application of the statistical method to stellar astronomy or medicine.

3. Statistical theory or mathematical statistics rather than statistics as a whole may perhaps be likened to a statistical method or a series of statistical procedures.

4. Sociology or the science of the life of groups of men in a society essentially applies the statistical method.

5. The stochastic theory of errors is the application of the statistical method to the treatment of observations. This statement contradicts the definition of Kolmogorov & Prokhorov, but I believe that their understanding of statistical data may well be generalized to include results of observations or measurements.

6. K. Pearson (§ 8.1) stated that the *unity of all science consists alone in its method* … To a certain extent this maxim is borne out by the essence of statistical method. Kruskal (1978, p. 1082) thought that *statistics has a neighbourly relation with philosophy of science*, but I will argue that *statistics* ought to be replaced here by *statistical method*. Recall also Achenwall (beginning of § 1.1): statistics *belongs to a well digested philosophy*.

7. For statistics, the axiomatized theory of probability is useless.

Bibliography

Abbreviation: AHES = *Arch. Hist. Exact Sci*.

JNÖS = *Jarbücher f. Nationalökonomie u. Statistik*

S, G, n means that the source in question is available, either in its original Russian or English, or in English translation at my website sheynin.de or at Google, Oscar Sheynin, in Document n.

Gauss *W*-i = Gauss, *Werke*, Bd. i

Gauss W/Erg-i = Gauss, *Werke*, Ergänzungsreihe, Bd. i.

Gauss (1816/1887, p. 130) means that I refer to the edition of 1886 of the memoir of 1816.

My main contributions on the present subject are Sheynin (2009a, 2009b, 2011 and 2013).

Abbe E. (1863), Über die Gesetzmäßigkeit in der Verteilung der Fehler bei Beobachtungsreihen. *Ges. Abh.*, vol. 2, 1989, pp. 55 – 81.
Achenwall G. (1752), *Staatsverfassung der europäischen Reiche im Grundrisse*. Göttingen. The first edition (Göttingen, 1749) was called *Abriß der neuesten Staatswissenschaft* etc. A large number of later editions up to 1798, but in 1768 the title was again changed.
--- (1763), *Staatsklugheit und ihren Grundsätzen*. Göttingen. Fourth edition, 1779.
Ambartzumian R. V. (1999), Stochastic geometry. In Prokhorov (1999, p. 682).
Anchersen J. P. (1741), *Descriptio statuum cultiorum in tabulis*. Copenhagen – Leipzig.
Anderson O. (1932), Ladislaus von Bortkiewicz. *Z. f. Nationalökonomie*, Bd. 3, pp. 242 – 250.
Andersson T. (1929), Statistics and insurance. *Nordic Stat. J.*, vol. 1, pp. 235 – 240.

André D. (1887), Solution directe du problème résolu par Bertrand. *C. r. Acad. Sci. Paris*, t. 105, pp. 436 – 437.
Andrews D. F. (1978), Data analysis, exploratory. In Kruskal & Tanur (1978, pp. 97 – 107).
Angström A. (1929), Statistics and meteorology. *Nordic Stat. J.*, vol. 1, pp. 228 – 234.
Anonymous (1735), Géométrie. *Hist. Acad. Roy. Sci. avec les Mém. Math. Phys.*, pp. 43 – 45 of the *Histoire.* The author was certainly Buffon.
Anonymous (1839), Introduction. *J. Stat. Soc. London*, vol. 1, pp. 1 – 5.
Anonymous (1954, in Russian), Account of an All-Union statistical conference. *Vestnik Statistiki*, No. 5, pp. 39 – 95. Also in *Vestnik Ekonomiki*, No. 12, pp. 75 – 111. The conference in Moscow, 1954, was organized by the Academy of Sciences, the Ministry for Higher Education, and the Central Statistical Agency of the USSR.
Anonymous (1968), Statistics, mathematical. In *Enc. Brit*., vol. 21, pp. 166 – 170.
Anonymous (1985), Statistics. In *New Enc. Brit*., vol. 28, pp. 230 – 239.
Anscombe F. J. (1967), Topics in the investigation of linear relations fitted by the method of least squares. *J. Roy. Stat. Soc*., vol. B29, pp. 1 – 52.
Arbuthnot J. (1712), An argument for Divine Providence taken from the constant regularity observed in the birth of both sexes. In Kendall & Plackett (1977, pp. 30 – 34).
Aristotle (1908 − 1930, 1954), *Works*, vols 1 – 12. London. I am referring to many treatises from this source. There is also a new edition of Aristotle (Princeton, 1984, vols 1 – 2).
Arnauld A., Nicole P. (1662), *L'art de penser.* Paris, 1992.
Babbage C. (1857), On tables of the constants of nature and art. *Annual Rept Smithsonian Instn* for 1856, pp. 289 – 302. An abstract appeared in 1834.
Bancroft T. A. (1966), Statistics. *Enc. Amer*., vol. 25, pp. 530 – 536a.
Bartholomew D. J. (1995), What is statistics? *J. Roy. Stat. Soc*., vol. A158, pp. 1 – 20.
Bauer R. K. (1955), Die Lexische Dispersionstheorie in ihren Beziehungen zur modernen statistischen Methodenlehre. *Mitteilungsbl. f. math. Statistik u. ihre Anwendungsgebiete*, Bd. 7, pp. 25 – 45.
Bayes T. (1764 – 1765), An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Roy. Soc.*, vols 53 – 54 for 1763 – 1764, pp. 360 – 418, 296 – 325. German translation: Leipzig, 1908. First part of memoir reprinted: *Biometrika*, vol. 45, 1958, pp. 293 – 315, also E. S. Pearson & Kendall (1970, pp. 131 – 153).
Bernoulli Chr. (1841), *Handbuch der Populationistik oder der Volkes- und Menschenkunde*. Ulm, Stettin.
Bernoulli Daniel (1735), Recherches physiques et astronomiques … Quelle est la cause physique de l'inclinaison des plans des planètes … In author's book (1987, pp. 303 – 326).
--- (1766/1982, pp. 235 – 267), Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir.
--- (1768/1982, pp. 276 – 287), De usu algorithmi infinitesimalis in arte coniectandi specimen.
--- (1768b/1982, pp. 290 – 303), De duratione media matrimoniorum etc.
--- (1768c, in Russian), Letter to Euler. *Priroda*, № 5, 1982, p. 103 – 104. Translated from Latin by A. P. Youshkevich.
--- (Manuscript 1769, in Latin), Same title as the memoir of 1778. English translation in *Festschrift for Lucien Le Cam.* New York, 1997, pp. 358 – 367.
--- (1770/1982, pp. 306 – 324), Disquisitiones analyticae de nouo problemata coniecturale.
--- (1770 – 1771/1982, pp. 326 – 338, 341 – 360), Mensura sortis ad fortuitam successionem rerum naturaliter contingentium applicata.
--- (1778, in Latin/1982, pp. 361 – 375), The most probable choice between several discrepant observations and the formation therefrom of the most likely induction. *Biometrika*, vol. 48, 1961, pp. 3 – 13, together with translation of Euler (1778). Reprint: E. S. Pearson & Kendall (1970, pp. 155 – 172).
--- (1780/1982, pp. 376 – 390), Specimen philosophicum de compensationibus horologicis, et veriori mensura temporis.
--- (1982, 1987), *Werke*, Bde 2 – 3. Basel.

Bernoulli Jakob (1713), *Ars conjectandi*. In author's book (1975, pp. 107 – 259). German translation: *Wahrscheinlichkeitsrechnung*. Leipzig, 1899 and Thun – Frankfurt/Main, 1999. English translation of part 4: *On the Law of Large Numbers*. Berlin, 2005 and at **S, G,** 8. Never trust the English translation of 2006 of the entire book by Sylla.
--- (1975), *Werke*, Bd. 3. Basel. Contains reprints of related materials and comments.
Bernoulli Niklaus (1709), *De usu artis conjectandi in jure*. In Bernoulli Jakob (1975, pp. 289 – 326). Translation: *Dissertation on the Use of the Art of Conjecture in Law*. Google.
Bernstein S. N. (1928, in Russian), The present state of the theory of probability. In Bernstein (1964, pp. 217 – 232). **S, G,** 7.
--- (1945, in Russian), On Chebyshev's work on the theory of probability. In Bernstein (1964, pp. 409 – 433). **S, G,** 5.
--- (1964), *Sobranie Sochineniy* (Coll. Works), vol. 4. N. p.
Bertrand J. (1888a), *Calcul des probabilités*. Second edition 1907. Reprint of first edition: New York, 1970.
--- (1887b), Solution d'un problème. *C. r. Acad. Sci. Paris*, t. 105, p. 369.
--- (1887c), Sur les épreuves répétées. Ibidem, pp. 1201 – 1203.
Bessel F. W. (1816), Untersuchungen über die Bahn des Olbersschen Kometen. *Abh. Preuss. Akad. Berlin*, math. Kl. 1812 – 1813, pp. 119 – 160. Only an abstract is included in Bessel (1876).
--- (1818), *Fundamenta astronomiae*. Königsberg.
--- (1820), Beschreibung des auf des Königsberger Sternwarte. *Astron. Jahrb.* (Berlin) für 1823, pp. 161 – 168.
--- (1823), Persönliche Gleichung bei Durchgangsbeobachtungen. In Bessel (1976, Bd. 3, pp. 300 – 304).
--- (1838a), Untersuchung über die Wahrscheinlichkeit der Beobachtungsfehler. In Bessel (1876, Bd. 2, pp. 372 – 391).
--- (1838b), *Gradmessung in Ostpreußen*. Berlin. Also in Bessel (1876, Bd. 3).
--- (1848), *Populäre Vorlesungen über wissenschaftliche Gegenstände*. Hamburg, Pertes-Besser u. Mauke. The report on probability theory is reprinted in Bessel (1876, Bd. 3).
--- (1843), Sir William Herschel. Ibidem, pp. 468 – 478.
--- (1845), Übervölkerung. Ibidem, pp. 483 – 486.
--- (1876), *Abhandlungen*, Bde 1 – 3. Leipzig.
Bienaymé I. J. (1839), Théorème sur la probabilité des résultats moyens des observations. *Soc. Philomat. Paris, Extraits*, sér. 5, pp. 42 – 49. Also in *L'Institut*, t. 7, No. 286, pp. 187 – 189.
Biermann K. R. (1966), Über die Beziehungen zwischen Gauss und Bessel. *Mitt. Gauss-Ges. Göttingen*, Bd. 3, pp. 7 – 20.
Biot J. B. (1811), *Traité élémentaire d'astronomie physique*, t. 2. Paris – St. Pétersbourg. Second edition.
Birg S., Editor (1986), *Ursprünge der Demographie in Deutschland. Leben und Werke J. P. Süssmilch's.* [Coll. papers.] Frankfurt/Main.
Birman I. (1960, in Russian), Scientific conference on application of mathematics in economics and planning. *Vestnik Statistiki*, No. 7, pp. 41 − 52.
Black W. (1788), *Comparative View of the Mortality of the Human Species*. London.
Block M. (1878), *Traité théorique et pratique de statistique*. Paris, 1886.
Bolshev L. N. (1989), Errors, theory of. In *Enc. Math*. (1988 – 1994, vol. 3, pp. 416–417). Also in *Fizichesky Enziklopedichesky Slovar* (1963, vol. 3, p. 577). Moscow.
Boltzmann L. (1868), Studien über das Gleichgewicht der lebenden Kraft. In author's book (1909, Bd. 1, pp. 49 – 96).
--- (1909), *Wissenschaftliche Abhandlungen*. Bde 1 – 3. Leipzig.
Boole G. (1851), On the theory of probabilities. In author's book (1952, pp. 247 – 259).
--- (1952), *Studies in Logic and Probability*, vol. 1. London.
Bortkevich V. I. & Chuprov A. A. (2005), *Perepiska* (Correspondence) *1895 – 1926*. Berlin. **S, G,** 9.
Bortkiewicz, L. von (1894 – 1896), Kritische Betrachtungen zur theoretischen Statistik. JNÖS, Bde 8, 10, 11, pp. 641 – 680, 321 – 360, 701 – 705.
--- (1898), *Das Gesetz der kleinen Zahlen*. Leipzig.

--- (1904), Anwendung der Wahrscheinlichkeitsrechnung auf Statistik. In *Enc. Math. Wiss.* Leipzig, Bd. 1, pp. 822 – 851.
--- (1917), *Die Iterationen*. Berlin.
--- (1918), Der mittlere Fehler der zum Quadrat erhobenen Divergenzkoeffizienten. *Jahresber. der deutschen Mathematiker-Vereinigung*, Bd. 27, pp. 71 – 126.
Box G. E. P. (1964). Errors, theory of. In *Enc. Brit.*, vol. 8, pp. 688 – 689.
Bru B., Jongmans F. (2001), Bertrand. In Heyde et al (2001, pp. 185 – 189).
Budd W. (1849), *Malignant Cholera*. London.
Buffon G. L. L. (1777), *Essai d'arithmétique morale*. In author's book (1954, pp. 456 – 488).
--- (1954), *Œuvres philosophiques.* Paris. Editors, J. Piveteau, M. Fréchet, C. Bruneau.
Bühler G., Editor (1886), *Laws of Manu*. Oxford, 1967.
Bull J. P. (1959), Historical development of clinical therapeutic trials. *J. Chronic Diseases*, vol. 10, pp. 218 – 248.
Butte W. (1808), *Die Statistik als Wissenschaft*. Landshut.
Buys Ballot C. H. D. (1850), Die periodischen […] Änderungen der Temperatur. *Fortschritte Phys.*, Bd. 3 für 1847, pp. 623 – 629.
Campbell L., Garnett W. (1882), *Life of Maxwell*. London. [London, 1884; New York – London, 1969.]
Catalan E. C. (1877), Un nouveau principe de probabilités. *Bull. Acad. Roy. Sci., Lettres, Beaux-Arts Belg*., 2me sér., 46e année, t. 44, pp. 463 – 468.
--- (1884), Application d'un nouveau principe de probabilités. Ibidem, 3me sér., 53e année, t. 3, pp. 72 – 74.
Cauchy A. L. (1845), Sur les secours que les sciences du calcul peuvent fournir aux sciences physiques on même aux sciences morales. *Oeuvr. compl*., sér. 1, t. 9. Paris, 1896, pp. 240 – 252.
Celsus (1935, in English), *De medicina*, vol. 1. London. Written in the first century AD.
Chaddock R. E. (1925), *Principles and Methods of Statistics.* Boston.
Chadwick E. (1842), *Report on the Sanitary Condition of the Labouring Population.* Edinburgh, 1965.
Chaitin G. J. (1975), Randomness and mathematical proof. *Scient. American*, vol. 232, pp. 47 – 52.
Chapman, S. (1941), *Halley as a Physical Geographer.* London.
Chebyshev P. L. (1887, in Russian), Sur les résidus intégraux qui donnent des valeurs approchées des intégrales. *Acta Math*., t. 12, 1888 – 1889, pp. 287 – 322.
Chernoff H. & Moses L. E. (1959), *Elementary Decision Theory*. New York, Wiley.
Chuprov (Tschuprow) A. A. (1906), Statistik als Wissenschaft. *Arch. Sozialwiss. u. Sozialpolitik*, Bd. 5, pp. 647 – 711.
--- (1922, in Russian), On the expectation of the ratio of two mutually dependent random variables. *Trudy Russk. Uchenykh za Granizsei*, vol. 1. Berlin, pp. 240 – 271. **S, G,** 2.
--- (1923), On the mathematical expectation of the moments of frequency distributions in the case of correlated observations. *Metron*, t. 2, pp. 461 – 493, 646 – 683.
Clifford, W. K. (1885). *Common Sense of the Exact Sciences*. London, Appleton, 1886, this being the first posthumous edition essentially extended by K. Pearson. There were several later editions, for example, New York, Dover, 1955.
Condamine C. M. de la (1759), Sur l'inoculation de la petite vérole. *Hist. Acad. Roy. Sci. Paris* 1754 *avec Mém. math. et phys*., pp. 615 – 670 of the *Mémoires*.
--- (1763), Second mémoire sur l'inoculation […]. Ibidem, pp. 439 – 482 of the *Mémoires*.
--- (1773), *Histoire de l'inoculation*. Amsterdam.
Condorcet M. J. A. N. (1795), *Esquisse d'un tableau historique des progrès de l'esprit humain*. Paris, 1988.
Cornfeld J. (1967), The Bayes theorem. *Rev. Intern. Statistical Inst.*, t. 35, pp. 34 – 49.
Cournot A. A. (1843), *Exposition de la théorie des chances et des probabilités*. Reprinted, Paris, 1984. *Exposition of the Theory of Chances and Probabilities.* Berlin, 2013 and **S, G,** 54.

Crofton M. W. (1869), On the theory of local probability applied to straight lines drawn at random in a plane. *Phil. Trans. Roy. Soc.*, vol. 158 for 1868, pp. 181 – 199.
Czuber E. (1891), Zur Kritik einer Gauss'schen Formel. *Monatshefte Math. Phys.*, Bd. 2, pp. 459 − 464.
--- (1921), *Die statistischen Forschungsmethoden.* Wien.
D'Alembert J. Le Rond (1761), Sur l'application du calcul des probabilités à l'inoculation de la petite vérole. *Opusc. Math.*, t. 2. Paris, pp. 26 – 95.
--- (1768), Sur un mémoire de M. Bernoulli concernant l'inoculation. Ibidem, pp. 98 – 105.
D'Amador R. (1837), *Mémoire sur le calcul des probabilités appliqué à la médecine*. Paris.
Darwin, C. (1859), *Origin of Species*. Cambridge, Mass., 1964. [Manchester, 1995.]
--- (1876), *The Effects of Cross and Self-Fertilisation in the Vegetable Kingdom.* London, 1878. [London, 1989.]
--- (1887), *Life and Letters*, vols 1 – 2. New York – London, 1897. [New York, 1969.]
--- (1903), *More Letters*, vol. 1. London.
David H. A. & Edwards A. W. F. (2001), *Annotated Readings in the History of Statistics*. New York, Springer.
Daw R. H. (1980), J. H. Lambert, 1728 – 1777. *J. Inst. Actuaries*, vol. 107, pp. 345 – 350.
Dawid Ph. (2005), Statistics on trial. *Significance*, vol. 2, No. 1, pp. 6 – 8.
De Candolle Alph. (1833), Revue des progrès de la statistique. *Bibl. Universelle*, Cl. Litt., année 18, t. 52, pp. 333 – 354.
--- (1873), *Histoire des sciences*. Genève – Bale. German translations: 1911, 1921.
Delambre J. B. J. (1819), Analyse des travaux de l'Académie … pendant l'année 1817, partie math. *Mém. Acad. Roy. Sci. Inst. de France*, t. 2 pour 1817, pp. I – LXXII of the *Histoire*.
De Moivre A. (1712, in Latin), De mensura sortis or the measurement of chance. *Intern. Stat. Rev.*, vol. 52, 1984, pp. 236 – 262. Commentary (A. Hald): Ibidem, pp. 229 – 236.
--- (1718), *Doctrine of Chances*. Third edition, 1756, reprinted: New York, 1967.
--- (1725), *Treatise on Annuities on Lives*. Later editions 1743 and 1756 (incorporated in the third edition of the *Doctrine*, pp. 261 – 328).
--- (1733, in Latin), A method of approximating the sum of the terms of the binomial $(a + b)^n$ expanded into a series from whence are deduced some practical rules to estimate the degree of ascent which is to be given to experiments. Translated by author, incorporated in the second edition of the *Doctrine* (1738) and in extended form in its third edition, pp. 243 – 254.
De Morgan A. (1864), On the theory of errors of observation. *Trans. Cambr. Phil. Soc.*, vol. 10, pp. 409 – 427.
De Morgan Sophia Elizabeth (1882), *Memoir of Augustus De Morgan*. London.
Descartes R. (1644), *Les principes de la philosophie. Œuvres*, t. 9, pt. 2. Paris, 1978 this being a reprint of the edition of 1647.
De Vries H. (1905), The evidence of evolution. *Annual Rept Smithsonian Instn* for 1904, pp. 389 – 396.
Dietz K., Heesterbeek J. A. P. (2002), Daniel Bernoulli's epidemiological model revisited. *Math. Biosciences*, vol. 180, pp. 1 – 21.
Dodge Y. (2003), *Oxford Dictionary of Statistical Terms*. Oxford, University Press.
Double F. J., Dulong P. L., Larrey F. H., Poisson S. D. (1835), Report on a manuscript by J. Civiale, "Researches de statistique sur l'affection calculeuse." *C. r. Acad. Sci. Paris*, t. 1, pp. 167 – 177.
Dufau P. A. (1840), *Traité de statistique ou théorie de l'étude des lois, d'après lesquelles se développent des faits sociaux.* Paris.
Du Pasquier L. G. (1910), Die Entwicklung der Tontinen bis auf die Gegenwart. *Z. schweiz. Stat.*, 46. Jg, pp. 484 – 513.
Eddington A. S. (1933), Notes on the method of least squares. *Proc. Phys. Soc.*, vol. 45, pp. 271 – 287.
Edgeworth F. Y. (1996), *Writings in Probability, Statistics and Economics*, vols 1 – 3. Cheltenham. Editor, C. R. McCann, Jr.

Edwards A. W. F. (1987), *Pascal's Arithmetic Triangle*. Baltimore, 2002.
--- (1994), R. A. Fisher on Karl Pearson. *Notes and Records Roy. Soc. London*, vol. 48, pp. 97 – 106.
Eisenhart C. (1963), Realistic evaluation of the precision and accuracy of instrument calibration. In Ku (1969, pp. 21 – 47).
--- (1964), The meaning of "least" in least squares. *J. Wash. Acad. Sci.*, vol. 54, pp. 24 – 33.
--- (1974), Pearson. *Dict. Scient. Biogr.*, vol. 10, pp. 447 – 473.
Encyclopedia of Mathematics (1988 – 1994), vols. 1 – 10. Dordrecht, Kluwer. Originally published in Russian, 1977 – 1985, in 5 vols.
Erdélyi A. (1956), *Asymptotic Expansions.* New York.
Erman A., Editor (1852), *Briefwechsel zwischen W. Olbers und F. W. Bessel*, Bde 1 – 2. Leipzig.
Euler L. (1767), Recherches générales sur la mortalité et la multiplication du genre humain. *Opera Omnia*, ser. 1, t. 7. Leipzig − Berlin, 1923, pp. 79 – 100. There also, pp. 545 – 552, is his manuscript *Sur multiplication du genre humaine*.
--- (1776), Eclaircissements sur les établissements publics en faveur tant des veuves que des morts, avec la description d'une nouvelle espèce de Tontine aussi favorable au Public qu'utile à l'Etat. Ibidem, pp. 181 – 245.
--- (1778, in Latin), Observations on the foregoing dissertation of Bernoulli. See Bernoulli Daniel (1778).
Farr W. (1885), *Vital Statistics*. London.
Fedorovich L. V. (1894), *Istoria i Teoria Statistiki* (History and Theory of Statistics). Odessa.
Feller W. (1950), *Introduction to Probability Theory and Its Applications*, vol. 1. New York – London. Third edition, 1968.
Finney D. J. (1960), *An Introduction to the Theory of Experimental Design*. Chicago, Univ. of Chicago Press. I refer to the preface, written in 1967, of the Russian edition, Moscow, 1970.
Fisher R. A. (1922). On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc.*, vol. 222, pp. 309 – 368.
--- (1925), *Statistical Methods for Research Workers*. In Fisher (1990, separate paging, reprint of edition of 1973).
--- (1937), Professor Karl Pearson and the method of moments. *Annals of Eugenics*, vol. 7, pp. 303 – 318.
--- (1956), *Statistical Methods and Statistical Inference*. In Fisher (1990). Reprint of edition of 1973, separate paging.
--- (1990) *Statistical Methods, Experimental Design and Scientific Inference.* Oxford, Oxford University Press.
Fletcher A., Miller J. C. P. et al (1946), *Index of Mathematical Tables*, vol. 1. Oxford, 1962.
Fourier J. B. J., Editor (1821 – 1829), *Recherches statistiques sur la ville de Paris et de département de la Seine*, tt. 1 – 4. Paris.
--- (1826), Sur les résultats moyens déduits d'un grand nombre d'observations. *Œuvres*, t. 2. Paris, 1890, pp. 525 – 545.
Fox J. J. (1860), On the province of the statistician. *J. Stat. Soc. London*, vol. 23, pp. 330 – 336.
Fréchet M. (1949), Réhabilitation de la notion statistique de l'homme moyen. In author's book *Les mathématiques et le concret.* Paris, 1955, pp. 317 – 341.
Freudenthal H. (1961), 250 years of mathematical statistics. In *Quantitative Methods in Pharmacology.* Amsterdam, 1961, pp. xi – xx. Editor H. De Jonge.
Frisch A. (1933), Editorial. *Econometrica*, vol. 1, pp. 1 – 4.
Fuss P. N. (1843), *Correspondance mathématique et physique de quelques célèbres géomètres du XVIII siècle*, tt. 1 – 2. New York – London, 1968.
Galen C. (1951), *Hygiene*. Springfield, Illinois.
Galilei G. (1632, in Italian), *Dialogue concerning the Two Chief World Systems*. Berkeley – Los Angeles, 1967.
Galton F. (1863), *Meteorographica.* London – Cambridge.
--- (1869), *Hereditary Genius*. London – New York, 1978.
--- (1877), Typical laws of heredity. *Nature*, vol. 15, pp. 492 – 495, 512 – 514, 532 – 533.
Gastwirth J. L., **Editor** (2000), *Statistical Science in the Courtroom.* New York.

Gatterer J. C. (1775), *Ideal einer allgemeinen Weltstatistik*. Göttingen.
Gauss C. F. (1809, in Latin), *Theorie der Bewegung*, Book 2, Section 3. In Gauss (1887, pp. 92 – 117).
--- (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.
--- (1855), *Méthode des moindres carrés.* Paris.
--- (1863 – 1930), *Werke*, Bde 1 – 12. Göttingen. Reprint: Hildesheim, 1973 – 1981.
--- (1887), *Abhandlungen zur Methode der kleinsten Quadrate*. Hrsg. A. Börsch, P. Simon. Vaduz, 1998.
--- (1900 – 1909), *Briefwechsel zwischen Gauss und Olbers*. W/Erg-4, No. 1 – 2. Hildesheim, 1976.
--- (1975 – 1987), *Werke*, Ergänzungsreihe, Bde 1 – 5. Hildesheim.
--- (1995, Latin and English), *Theory of Combination of Observations Least Subject to Error*. Includes Gauss (1823a) in German and English. Translated with Afterword by G. W. Stewart. Philadelphia.
Gavarret J. (1840), *Principes généraux de statistique médicale.* Paris. German translation: Erlangen, 1844.
Gerling Ch. L. (1861), Notiz in Betreff der Prioritätsverhältnisse in Beziehung auf die Methode der kleinsten Quadrate. *Nachr. Georg-Augusts-Univ. und Kgl. Ges. Wiss. Göttingen*, pp. 273 – 275.
Ghosh J. K. (1994), Mahalanobis and the art and science of statistics: the early days. *Indian J. Hist. Sci*., vol. 29, pp. 89 – 98.
Gillispie C. (1963), Intellectual factors in the background of analysis of probabilities. In Crombie A. C. *Scientific Change*. New York, pp. 431 – 453.
Gnedenko B. V. (1950), *Kurs Teorii Veroiatnostei* (Course in the Theory of Probability). Moscow, 1954. [*Theory of Probability*. Providence, RI, 2005.]
Gnedenko B. V. & Sheynin O. B. (1978, in Russian), Theory of probability. Chapter in *Matematika XIX Veka*. Moscow. Editors, A. N. Kolmogorov, A. P. Youshkevich. *Mathematics of the 19th Century,* vol. 1. Basel, Birkhäuser, 1992, 2001, pp. 211 – 288.
Graunt J. (1662), *Natural and Political Observations Made upon the Bills of Mortality*. Baltimore, 1939. Editor, W. F. Willcox.
Great Books (1952), *Great Books of the Western World*, vols 1 – 54. Chicago.
Greenwood M. (1936), Louis and the numerical method. In author's *Medical Dictator.* London, pp. 123 – 142.
--- (1941 – 1943), Medical statistics from Graunt to Farr. *Biometrika*. Reprint: E. S. Pearson & Kendall (1970, pp. 47 – 120).
Guerry A. M. (1864), *Statistique morale de l'Angleterre comparée avec la statistique morale de la France.* Paris, Baillière et fils.
Gumbel E. J. (1917), Eine Darstellung statistischer Reihe durch Euler. *Jahresber. Deutschen Mathematiker-Vereinigung*, Bd. 25, pp. 251 – 264.
--- (1978), Bortkiewicz. In Kruskal & Tanur (1978, pp. 24 – 27).
Hald A. (1952), *Statistical Theory with Engineering Applications.* New York, 1960.
--- (1990), *History of Probability and Statistics and Their Applications before 1750.* New York.
--- (1998), *History of Mathematical Statistics from 1750 to 1930*. New York.
Halley E. (1693), *An Estimate of the Degree of Mortality of Mankind. Baltimore*, 1942**.**
Harvey W. (1651 in Latin), *Anatomical Exercises in the Generation of Animals*. In *Great Books* (1952, vol. 28, pp. 329 – 498).
Haushofer D. M. (1872), *Lehr- und Handbuch der Statistik*. Wien.
Helmert F. R. (1868), Studien über rationelle Vermessungen im Gebiete der höhern Geodäsie. *Z. Math. Phys*., Bd. 13, pp. 73 – 120, 163 – 186.
--- (1872), *Ausgleichungsrechnung nach der Methode der kleinsten Quadrate.* Later editions: 1907 and 1924 (Leipzig).
--- (1875), Über die Berechnung des wahrscheinlichen Fehlers aus einer endlichen Anzahl wahrscheinlicher Beobachtungsfehler. *Z. Math. Phys*., Bd. 20, pp. 300 – 303.

--- (1876), 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. In author's book (1993, pp. 189 – 208).
--- (1993), *Akademie-Vorträge.* Frankfurt am Main.
Hendriks F. (1852 – 1853), Contributions to the history of insurance. *Assurance Mag.*, vol. 2, pp. 121 – 150, 222 – 258; vol. 3, pp. 93 – 120.
--- (1863), Notes on the early history of tontines. *J. Inst. Actuaries*, vol. 10, pp. 205 – 219.
Herschel W. (1805), On the direction and motion of the Sun. In author's book (1912, vol. 2, pp. 317 – 331).
--- (1806), On the quantity and velocity of the solar motion. Ibidem, pp. 338 – 359.
--- (1817), Astronomical observations and experiments tending to investigate the local arrangement of celestial bodies in space. Ibidem, pp. 575 – 591.
--- (1912), *Scientific Papers*, vols 1 – 2. London. [Bristol, 2003.]
Heyde C. C., Seneta E. (1977), *Bienaymé.* Berlin.
---, **Editors** (2001), *Statisticians of the Centuries*. New York.
Humboldt A. (1817), Des lignes isothermes. *Mém. Phys. Chim. Soc. d'Arcueil*, t. 3, pp. 462 – 602.
--- (1845 – 1862), *Kosmos*, Bde 1 – 5 (1845, 1847, 1850, 1858, 1862). Stuttgart. [Bde 1 – 4: Stuttgart, 1877.]
Huygens C. (1657), De calcul dans les jeux de hasard. In author's book (1888 – 1950, t. 14, pp. 49 – 91).
--- (1888 – 1950), *Oeuvres complètes*, tt. 1 – 22. La Haye. Volumes 4, 6, 10 and 14 appeared in 1891, 1895, 1905 and 1920 respectively.
John V. (1883, in German), The term *statistics*. *J. Roy. Stat. Soc.*, vol. 46, pp. 656 – 679.
Johnson N. L., Kotz S., Editors (1997), *Leading Personalities in Statistical Sciences.* New York.
Kapteyn J. C. (1906), *Plan of Selected Areas.* Groningen.
Karn M. Noel (1931), An inquiry into various death-rates and the comparative influence of certain diseases on the duration of life. *Annals of Eugenics*, vol. 4, pp. 279 – 326.
Kaufman A. A. (1922), *Teoria i Metody Statistiki* (Theory and Methods of Statistics). Moscow. Fourth ed. Fifth, posthumous edition, Moscow, 1928. German edition: *Theorie und Methoden der Statistik*. Tübingen, 1913.
Kendall M. G. (Sir Maurice) (1950), The statistical approach. *Economica*, vol. 17, pp. 127 – 145.
--- (1971), The work of Ernst Abbe. *Biometrika*, vol. 58, pp. 369 – 373. Reprint: Kendall & Plackett (1977, pp. 331 – 335).
--- (1978), The history of the statistical method. In Kruskal & Tanur (1978, vol. 2, pp. 1093 – 1102).
Kendall M. G. & Buckland W. R. (1971). Statistics. In *Dictionary of Statistical Terms.* Edinburgh, Oliver and Boyd, p. 145.
Kendall M. G., Moran P. A. P. (1963), *Geometrical Probabilities.* London.
Kendall M. G., Plackett R. L., Editors (1977), *Studies in the History of Statistics and Probability*, vol. 2. London, Griffin.
Kepler J. (1609, in Latin), *New Astronomy*. Cambridge, 1992.
Khinchin A. Ya. (1961, in Russian), The Mises frequency theory and modern ideas of the theory of probability. *Vopr. Filosofii*, No. 1 and 2, pp. 91 – 102 and 77 – 89. English translation: *Science in Context*, vol. 17, 2004, pp. 391 – 422.
Klimpt W. (1936), *Mathematische Untersuchungen*. Berlin.
Knapp G. F. (1872), Quetelet als Theoretiker. JNÖS, Bd. 18, pp. 89 – 124.
Knies C. G. A. (1850), *Die Statistik als selbstständige Wissenschaft.* Kassel. [Frankfurt, 1969.]
Kohli K. (1975), Kommentar zur Dissertation von N. Bernoulli. In Bernoulli Jakob. (1975, pp. 541 – 556).
Kohli K., van der Waerden B. L. (1975), Bewertung von Leibrenten. Ibidem, pp. 515 – 539.

Kolmogorov A. N. (1938, in Russian), The theory of probability and its applications. In *Matematika i Estestvoznanie v SSSR* (Math. and Nat. Sci. in the USSR). Moscow, pp. 51 – 61. **S, G,** 7.
--- (1947, in Russian), The role of Russian science in the development of the theory of probability. *Uchenye Zapiski Moskovsk. Gosudarstvenn. Univ.*, No. 91, 53 – 64. **S, G,** 7.
--- (1948a, in Russian), The main problems of theoretical statistics. Abstract. In *Vtoroe Vsesoiuznoe Soveschanie po Matematicheskoi Statistike 1948* (Second All-Union Conf. on Math. Stat., 1948). Tashkent, pp. 216 – 220.
--- (1948b, in Russian). E. E. Slutsky. *Math. Scientist*, vol. 27, 2002, pp. 67 – 74.
--- (1954, in Russian). Address to a statistical conference. In Anonymous (1954, pp. 46 – 47).
--- (1983, in Russian), On the logical foundations of probability theory. *Sel. Works*, vol. 2. Dordrecht, 1992, pp. 515 – 519.
Kolmogorov A. N., Prokhorov Yu. V. (1988, in Russian), Mathematical statistics. *Enc. Math.*, vol. 6, 1990, pp. 138 – 142.
Köppen W. (1875, in Russian), On the observation of periodic phenomena in nature. *Zapiski Russk. Geografich. Obshchestvo po Obshchei Geografii*, vol. 6, No. 1, pp. 255 – 276.
Kotz S. (1965), Statistics in the USSR. *Survey*, vol. 57, pp. 132 – 141.
Krauske O. (1892), Sophie Charlotte. *Allg. Deutsche Biogr.*, Bd. 34, pp. 676 – 684.
Kronecker L. (1901), *Vorlesungen über Zahlentheorie*, Bd. 1. Leipzig.
Kruskal W. H. (1978), Statistics: the field. In Kruskal & Tanur (1978, vol. 2, pp. 1071 – 1093).
Kruskal W. H. and Tanur J. M., Editors (1978), *International Encyclopedia of Statistics*, vols. 1 – 2. New York, McMillan.
Ku H. H., **Editor** (1969), *Precision Measurement and Calibration.* Nat. Bureau Standards Sp. Publ. 300, vol. 1. Washington.
Lamarck J. B. (1800 − 1811), *Annuaire météorologique*, tt. 1 − 11. Paris.
--- (1810 – 1814, manuscript), *Aperçu analytique des connaissances humaines*. Partly published: Vachon M., et al (1972), *Inédits de Lamarck*. Paris, pp. 69 – 141. Russian translation of entire work to which I refer is in author's *Izbrannye Proizvedenia* (Sel. Works), vol. 2. Moscow, 1959, pp. 93 – 662.
--- (1815), *Histoire naturelle des animaux sans vertèbres*, t. 1. Paris.
Lambert J. H. (1760), *Photometria.* Augsburg.
--- (1765a), Anmerkungen und Zusätze zur practischen Geometrie. In author's book *Beyträge zum Gebrauche der Mathematik und deren Anwendung*, Tl. 1. Berlin, 1765, pp. 1 – 313.
--- (1765b), Theorie der Zuverlässigkeit der Beobachtungen und Versuche. Ibidem, pp. 424 – 488.
--- (1772), Anmerkungen über die Sterblichkeit, Todtenlisten, Geburthen und Ehen. Ibidem, Tl. 3. Berlin, 1772, pp. 476 – 569.
Laplace P. S. (1776), Recherches sur l'intégration des équations différentielles. *Oeuvr. Compl.*, t. 8. Paris, 1891, pp. 69 – 197.
--- (1789), Sur quelques points du système du monde. *Oeuvr. Compl.*, t. 11. Paris, 1895, pp. 477 – 558.
--- (1796, 1798/1799, 1808, 1813, 1835), *Exposition du système du monde. Oeuvr. Compl.*, t. 6. Paris, 1884.
--- (1798 – 1825), *Traité de mécanique céleste. Oeuvr. Compl.*, tt. 1 – 5. Paris, 1878 – 1882. English translation (N. Bowditch): *Celestial Mechanics* (1832), vols 1 – 4. New York, 1966.
--- (1812), *Théorie analytique des probabilités. Oeuvr. Compl.*, t. 7. Paris, 1886.
--- (1814, in French), *Philosophical Essay on Probabilities*. New York, 1995.
--- (1816), *Théor. Anal. Prob., Supplément 1. Oeuvr. Compl.*, t. 7, No. 2, pp. 497 – 530.
--- (1818), *Théor. Anal. Prob., Supplément 2.* Ibidem, pp. 531 – 580.
Laurent P. H. (1902), *Petit traité d'économie politique mathématique.* Paris, Schmid.
Le Cam L. (1986), The central limit theorem around 1935. *Stat. Sci.*, vol. 1, pp. 78 – 96.
Lazarsfeld P. F. (1961), Notes on the history of quantification in sociology. *Isis*, vol. 52, pp. 277 – 333.

Legendre A. M. (1805), *Nouvelles méthodes pour la détermination des orbites des comètes.* Paris.
Lehmann E. L. (1951), A general concept of unbiasedness. *Annals Math. Stat.*, vol. 22, pp. 587 – 592.
Leibniz G. W. (manuscript 1680 – 1683), Essay de quelques raisonnemens nouveau sur la vie humaine. In Leibniz (2000, pp. 428 – 445).
--- (manuscript 1682), Quaestiones calculi politici circa hominum vitam. In Leibniz (2000, pp. 520 – 523, Latin and German).
--- (2000), *Hauptschriften zur Versicherungs- und Finanzmathematik.* Berlin. Editors, E. Knobloch et al.
Le procès (1900), *Le procès Dreyfus devant le Conceil de guerre de Rennes*, tt. 1 – 3. Paris.
Lexis W. (1879), Über die Theorie der Stabilität statistischer Reihen. JNÖS, Bd. 32, pp. 60 – 98. Reprinted in author's book (1903, pp. 170 – 212).
--- (1886), Über die Wahrscheinlichkeitsrechnung und deren Anwendung auf die Statistik. JNÖS, Bd. 13 (47), pp. 433 – 450.
--- (1903), *Abhandlungen zur Theorie der Bevölkerungs- und Moralstatistik.* Jena.
--- (1913), Review of A. A. Kaufmann (1913). *Schmollers Jahrbuch f. Gesetzgebung, Verwaltung u. Volkswirtschaft in Deutschen Reiche*, Bd. 37, pp. 2089 – 2092.
Libri-Carrucci G. B. I. T., rapporteur (1834), Au nom d'une Commission. *Procès verbaux des séances. Acad. Sci. Paris*, t. 10, pp. 533 – 535. Report on a manuscript submitted by I. J. Bienaymé. Members of the Commission: S. F. Lacroix, S. D. Poisson.
Liebermeister C. (ca. 1876), Über Wahrscheinlichkeitsrechnung in Anwendung auf therapeutische Statistik. *Sammlung klinischer Vorträge* No. 110 (Innere Medizin No. 39). Leipzig, pp. 935 – 961.
Lindley D. V. (1984), Prospects for the future. The next 50 years. *J. Roy. Stat. Soc*., vol. A147, pp. 359 – 367.
Louis P. C. A. (1825), *Recherches anatomico-pathologiques sur la phtisie*. Paris.
Lueder A. F. (1812), *Kritik der Statistik und Politik*. Göttingen.
Mach E. (1897), *Die Mechanik in ihrer Entwicklung*. Leipzig. Third edition.
Maciejewski C. (1911), *Nouveaux fondements de la théorie de la statistique.* Paris.
Mahalanobis P. C. (1936), Note on the statistical and biometric writings of K. Pearson. *Sankhya*, vol. 2, pp. 411 – 422.
Maire [C.], Boscovich [R. J.] (1770), *Voyage astronomique et géographique dans l'État de l'Église.* Paris.
Malthus T. R. (1798), *Essay on the Principle of Population. Works*, vol. 1. London, 1986.
Markov A. A. (1888), *Table des valeurs de l'intégrale …* St. Pétersbourg.
--- (1899, in Russian), The law of large numbers and the method of least squares. In Markov (1951, pp. 231 – 251). **S, G,** 5.
--- (1900), [*Treatise*] *Ischislenie Veroiatnostei* (Calculus of Probabilities). Subsequent editions: 1908, 1913, and (posthumous) 1924. German translation: Leipzig – Berlin, 1912.
--- (1951), *Izbrannye Trudy* (Sel. Works). No place.
Marriot F. H. C. (1991), Statistics. In *Dictionary of Statistical Terms*. Harlow (Essex), New York, p. 196**.**
Maupertuis P. L. M. (1745), Venus physique. *Oeuvres*, t. 2. Lyon, 1756, pp. 1 – 133.
--- (1751), Système de la nature. Ibidem, pp. 135 – 184.
Maxwell J. C. (1859), On the stability of the motion of Saturn's rings. *Scient. Papers*, vol. 1, pp. 288 – 376.
--- (1871), Introductory lecture on experimental physics. Ibidem, vol. 2, pp. 241 – 255.
--- (read 1873, 1873a), Does the progress of physical science tend to give any advantage to the opinion of necessity […] over that of contingency of events. In Campbell & Garnett (1882, pp. 357 – 366).
--- (1873b, manuscript; publ. 1882), Discourse on molecules. In Campbell & Garnett (1882, pp. 272 – 274).
--- (1877), Review of H. W. Watson, *Treatise on the Kinetic Theory of Gases*. Oxford, 1876. *Nature*, vol. 16, pp. 242 – 246.

--- (1890), *Scientific Papers*, vols 1 – 2. Cambridge. Reprints: Paris, 1927, New York, 1965.
Mayer T. (1750), Abhandlung über die Umwälzung des Mondes um seine Axe. *Kosmograph. Nachr. u. Samml.* für 1748, pp. 52 – 183.
Mendelsohn M. (1761), Über die Wahrscheinlichkeit. *Phil. Schriften*, Tl. 2. Berlin, pp. 189 – 228.
Methods (1905), Methods for promoting research in the exact sciences. *Carnegie Instn of Washington*, *Yearbook* No. 3 for 1904, pp. 179 – 193.
Michell J. (1767), An inquiry into the probable parallax and magnitude of the fixed stars. *Phil. Trans. Roy. Soc. Abridged*, vol. 12, 1809, pp. 423 – 438.
Miklashevsky I. N. (1901, in Russian). Statistics. In *Enziklopedichesky Slovar*, F. A. Brockhaus and I. A. Efron, Editors. Petersburg, Halfvolume 62, pp. 476 – 505.
Mill J. S. (1843), *System of Logic*. London, 1886. Many more editions, e. g., *Coll. Works*, vol. 8. Toronto, 1974.
Mises R. von (1919), Fundamentalsätze der Wahrscheinlichkeitsrechnung. *Math. Z.*, Bd. 4, pp. 1 – 97.
--- (1964a), *Mathematical Theory of Probability and Statistics*. Edited and completed by Hilda Geiringer. New York.
--- (1964b), *Selected Papers*, vol. 2. Providence, RI.
Montmort P. R. (1708), *Essay d'analyse sur les jeux de hazard.* Second edition, 1713. New York, 1980.
Muncke G. W. (1837), Meteorologie. *Gehler's Phys. Wörterbuch*, Bd. 6/3. Leipzig, pp. 1817 – 2083.
Newcomb S. (1859 – 1861), Notes on the theory of probability. *Math. Monthly*, vol. 1, pp. 136 – 139, 233 – 235, 331 – 335; vol. 2, pp. 134 – 140, 272 – 275; vol. 3, pp. 119 – 125, 343 – 349.
Newton I. (1704), *Opticks.* London, 1931. *Queries* were added later, from 1717 onward, and the edition of 1931 (reprinted in 1952) was based on that of 1730.
--- (1728), *Chronology of Ancient Kingdoms Amended.* London. [London, 1770.]
--- (1958), *Papers and Letters on Natural Philosophy*. Cambridge.
--- (1967), *Mathematical Papers*, vol. 1. Cambridge.
Neyman J. (1950), *First Course in Probability and Statistics*. London, Griffin.
Nikulin M. S. & Poliscuk V. I. (1999), Theory of errors. In Prokhorov (1999, pp. 439 – 440).
Nordenmark N. V. E. (1929), P. W. Wargentin. *Nordic Stat. J.*, vol. 1, pp. 241 – 252.
O'Donnell T. (1936), *History of Life Insurance*. Chicago.
Obodovsky A. (1839), *Teoria Statistiki* (Theory of Statistics). Petersburg.
Ondar Kh. O., Editor (1977, in Russian), *Correspondence between Markov and Chuprov on the Theory of Probability and Mathematical Statistics*. New York, 1981.
Orlov A. (1990, in Russian), On reforming the statistical science and on its application. *Vestnik Statistiki*, No. 1, pp. 65 – 71.
Paevsky V. V. (1935, in Russian), Euler's work in population statistics. In *Euler. Memorial volume.* Moscow – Leningrad, pp. 103 – 110. **S, G,** 5.
Pascal B. (1654, in French), [Correspondence with P. Fermat.] *Oeuvr. Compl.*, t. 1, pp. 145 – 166.
--- (1665), *Traité du triangle arithmétique.* Ibidem, pp. 282 – 327.
--- (1669), Pensées, fragments. *Oeuvr. Compl.*, t. 2, pp. 543 – 1046.
--- (1998 – 2000), *Oeuvres complètes*, tt. 1 – 2. Paris.
Pearson E. S. (1936 – 1937), K. Pearson: an appreciation of some aspects of his life and work. *Biometrika*, vol. 28, pp. 193 – 257; vol. 29, pp. 161 – 248.
--- (1990), *"Student". A Statistical Biography of W. S. Gossett*. Edited and augmented by R. L. Plackettt assisted by G. A. Barnard. Oxford.
Pearson E. S., Kendall M. G., Editors (1970), *Studies in the History of Statistics and Probability* [vol. 1]. London, Griffin.
Pearson K. (1887), On a certain atomic hypothesis. *Trans. Cambr. Phil. Soc.*, vol. 14, pp. 71 – 120.
--- (1891), Atom squirts. *Amer. J. Math.*, vol. 13, pp. 309 – 362.
--- (1892), *The Grammar of Science*. London, Griffin. [New York, 2004.]
--- (1906), W. F. R. Weldon, 1860 – 1906. *Biometrika*, vol. 5, pp. 1 – 50.
--- (1907), On correlation and methods of modern statistics. *Nature*, vol. 76, pp. 577 – 578, 613 – 615, 662.

--- (1914 – 1930), *Life, Letters and Labours of Fr. Galton*, vols 1, 2, 3A, 3B. Cambridge.
--- (1925), James Bernoulli's theorem. *Biometrika*, vol. 17, pp. 201 – 210.
--- (1926), A. De Moivre. *Nature*, vol. 117, pp. 551 – 552.
--- (1928), On a method of ascertaining limits to the actual number of marked individuals … from a sample. *Biometrika*, vol. 20A, pp. 149 – 174.
--- (1978), *The History of Statistics in the 17th and 18th Centuries*. Editor E. S. Pearson. London, Griffin. Lectures given in 1921 – 1933.
Petrov V. V. (1954, in Russian), The method of least squares and its extreme properties. *Uspekhi Matematich. Nauk*, vol. 1, pp. 41 – 62.
Petty W. (1674), *Discourse Read before the Royal Society*. London.
--- (1690), Political Arithmetic. In author's book (1899, vol. 2, pp. 239 – 313).
--- (1899), *Economic Writings*, vols 1 – 2. Editor C. H. Hull. Cambridge. [London, 1997.]
--- (1927), *Papers*, vols 1 – 2. London, 1997.
Pfanzagl J., Sheynin O. (1997), Süssmilch. In Johnson & Kotz (1997, pp. 73 – 75). Incorporated in second edition (2006) of Kotz & Johnson but somehow appeared there anonymously.
Pirogov N. I. (1849, in Russian), On the achievements of surgery during the last five years. *Zapiski po Chasti Vrachebnykh Nauk Med.-Khirurgich. Akad.*, Year 7, pt. 4, sect. 1, pp. 1 – 27.
--- (1850 – 1855, in Russian), Letters from Sevastopol. In Pirogov (1957 – 1962, vol. 8, 1961, pp. 313 – 403).
--- (1854), *Statistischer Bericht über alle meine im Verlauf eines Jahres ... vorgenommenen oder beobachteten Operationsfälle*, this being the author's booklet *Klinische Chirurgie*, No. 3. Leipzig.
--- (1864), *Grundzüge der allgemeinen Kriegschirurgie.* Leipzig. Russian version: 1865 – 1866.
--- (1871), *Bericht über die Besichtigung der Militär-Sanitäts-Anstalten in Deutschland, Lothringen und Elsas im Jahre 1870*. Leipzig.
--- (1879, in Russian), *Das Kriegs-Sanitäts-Wesen und die Privat-Hilfe auf dem Kriegsschauplätze etc.* Leipzig, 1882.
--- (1957 – 1962), *Sobranie Sochineniy* (Coll. Works), vols 1 – 8. Moscow.
Plackett R. L. (1972), The discovery of the method of least squares. *Biometrika*, vol. 59, pp. 239 – 251. Reprinted in Kendall & Plackett (1977, pp. 279 – 291).
Plato J. von (1983), The method of arbitrary functions. *Brit. J. Phil. Sci.*, vol. 34, pp. 37 – 47.
Ploshko V. G. and Eliseeva I. I. (1990), *Istoria Statistiki* (History of Statistics). Moscow.
Poincaré H. (1892a), *Thermodynamique.* Paris.
--- (1892b), Réponse à P. G. Tait. *Œuvres*, t. 10. Paris, 1954, pp. 236 – 237.
--- (1894), Sur la théorie cinétique de gaz. Ibidem, pp. 246 – 263.
--- (1896), *Calcul des probabilités*. Paris. Second edition, 1912, reprint Sceaux, 1987.
--- (1902), *La science et l'hypothèse.* Paris, 1923. [Paris, 1968.]
--- (1905), *La valeur de la science.* Paris, 1970.
--- (1907), Le hasard. *La rev. du mois*, t. 3, pp. 257 – 276.
--- (1908), *Science et méthode.* Paris.
--- (1921), Résumé analytique [of his own contributions]. In *Mathematical Heritage of H. Poincaré.* Providence, Rhode Island, 1983. Editor F. E. Browder, pp. 257 – 357.
Poisson S.-D. (1825 – 1826), Sur l'avantage du banquier au jeu de trente-et-quarante. *Annales math. pures et appl.*, t. 16, pp. 173 – 208.
--- (1837a), *Recherches sur la probabilité des jugements, principalement en matière criminelle et en matière civile*. Paris. [Paris, 2003.] *Researches into the Probabilities of Judgements in Criminal and Civil Cases*. Berlin, 2013. Also at **S, G,** 53.
--- (1837b), Programmes de l'enseignement de l'Ecole Polytechnique … pour l'année scolaire 1836 – 1837. Paris.
Polya G. (1920), Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem. *Math. Z.*, Bd. 8, pp. 171 – 181.
Prokhorov Yu. V., Editor (1999), *Veroiatnost i Matematicheskaia Statistika. Enziklopedia* (Probability and Math. Statistics. Enc.). Moscow.
Quetelet A. (1826), À M. Villermé etc. *Corr. Math. et Phys*., t. 2, pp. 170 – 178.

--- (1828), *Instructions populaires sur le calcul des probabilités*. Bruxelles.
--- (1829), Recherches statistiques sur le Royaume des Pays-Bas. *Mém. Acad. Roy. Sci., Lettres et Beaux-Arts Belg*., t. 5. Separate paging.
--- (1832a), Recherches sur la loi de la croissance de l'homme. Ibidem, t. 7. Separate paging.
--- (1832b), Recherches sur le penchant au crime. Ibidem. Separate paging.
--- (1836), *Sur l'homme et le développement de ses facultés, ou essai de physique sociale*, tt. 1 – 2. Bruxelles.
--- (1846), *Lettres ... sur la théorie des probabilités.* Bruxelles.
--- (1848a), *Du système social et des lois qui le régissent*. Paris.
--- (1848b), Sur la statistique morale. *Mém. Acad. Roy. Sci., Lettres et Beaux-Arts Belg.*, t. 21. Separate paging.
--- (1853), *Théorie des probabilités.* Bruxelles.
--- (1869), *Physique sociale*, tt. 1 – 2. Bruxelles. A revised edition of the authors' book of 1836. [Bruxelles, 1997.]
--- (1870), Des lois concernant le développement de l'homme. *Bull. Acad. Roy. Sci., Lettres, Beaux-Arts Belg*., 39[e] année, t. 29, pp. 669 – 680.
--- (1871), *Anthropométrie*. Bruxelles.
--- (1974), *Mémorial.* Bruxelles.
Quetelet A., Heuschling X. (1865), *Statistique internationale* (*population*). Bruxelles.
Quetelet A. & Smits Ed. (1832), *Recherches sur la reproduction et la mortalité de l'homme*. Bruxelles.
Rao C. R. (1993), Statistics must have a purpose: the Mahalanobis dictum. *Sankhya*, vol. A55, pp. 331 – 349.
Rehnisch E. (1876), Zur Orientierung über die Untersuchungen und Ergebnisse der Moralstatistik. *Z. Philos. u. philos. Kritik*, Bd. 69, pp. 43 – 115.
Riabushkin T. V. (1980, in Russian), Statistics. In *Great Sov. Enc.*, English ed., vol. 24, pp. 497 – 499.
Romanovsky V. I. (1930), *Matematicheskaia Statistika* (Math. Statistics). Moscow – Leningrad.
--- (1955, in Russian), Errors, theory of. In *Great Sov. Enc*., second edition, vol. 31, pp. 500 – 501.
--- (1961), Same title as in 1930, book 1. Tashkent. Editor, T. A. Sarymsakov.
Särndal C.-E. (1971), The hypothesis of elementary errors and the Scandinavian school in statistical theory. *Biometrika*, vol. 58, pp. 375 – 392. Reprint: Kendall & Plackett (1977, pp. 419 – 435).
Schiefer P. (1916), *Achenwall und seine Schule*. München.
Schlözer A. L. (1804), *Theorie der Statistik nebst Ideen über das Statium der Politik überhaupt*. Göttingen.
Seidel L. (1865), Über den … Zusammenhang … zwischen der Häufigkeit der Typhus-Erkrankungen und dem Stande des Grundwassers. *Z. Biol.*, Bd. 1, pp. 221 – 236.
--- (1866), Vergleichung der Schwankungen der Regenmengen mit der Schwankungen in der Häufigkeit des Typhus. Ibidem, Bd. 2, pp. 145 – 177.
Seneta E. (1987), Chuprov on finite exchangeability, expectation of ratios and measures of association. *Hist. Math.*, vol. 14, pp. 243 – 257.
Shaw N., Austin E. (1926), *Manual of Meteorology*, vol. 1. Cambridge, 1942.
Sheynin O. B. (1966), Origin of the theory of errors. *Nature*, vol. 211, pp. 1003 – 1004.
--- (1971a), Newton and the theory of probability. AHES, vol. 7, pp. 217 – 243.
--- (1971), Lambert's work in probability. AHES, vol. 7, pp. 244 – 256.
--- (1977), Early history of the theory of probability. AHES, vol 17, pp. 201 – 259.
--- (1979), Gauss and the theory of errors. AHES, vol. 20, pp. 21 – 72.
--- (1984a), On the history of the statistical method in astronomy. AHES, vol. 29, pp. 151 – 199.
--- (1984b), On the history of the statistical method in meteorology. AHES, vol. 31, pp. 53–95.
--- (1986), Quetelet as a statistician. AHES, vol. 36, pp. 281–325.
--- (1990, in Russian), *Aleksandr A. Chuprov: Life, Work, Correspondence*. Göttingen, 2011. Revised translation.
--- (1991), Poincaré's work in probability. AHES, vol. 42, pp. 137 – 172.

--- (1997, in Russian), Markov and insurance of life. *Math. Scientist*, vol. 30, 2005, pp. 5 – 12.
--- (1998), Statistics in the Soviet epoch. JNÖS, Bd. 217, pp. 529 – 549.
--- (2000), Bessel: some remarks on his work. *Historia Scientiarum*, vol. 10, pp. 77 – 83.
--- (2001a), Pirogov as a statistician. Ibidem, pp. 213 – 225.
--- (2001b), Gauss. In Heyde & Seneta (2001, pp. 119 – 122).
--- (2001c), Gauss, Bessel and the adjustment of triangulation. *Historia Scientiarum*, vol. 11, pp. 168 – 175.
--- (2002), Newcomb as a statistician. Ibidem, vol. 12, pp. 142 – 167.
--- (2006), Markov's work on the treatment of observations. Ibidem, vol. 16, pp. 80 – 95.
--- (2008), Bortkiewicz' alleged discovery: the law of small numbers. Ibidem, vol. 18, pp. 36 – 48.
--- (2009a), *Theory of Probability. Historical Essay*. Berlin. **S, G,** 10.
--- (2009b), *Theory of Probability and Statistics As Exemplified in Short Dictums.* Berlin. **S, G,** 23.
--- (2011), Statistics, history of. In *Intern. Enc. of Statistical Science*. Göttingen, pp. 1493 – 1504.
--- (2012), New exposition of Gauss' final justification of least squares. *Math. Scientist*, vol. 37, pp. 147 – 148.
--- (2013), Extended version of Sheynin (2009a), in Russian. **S, G,** 11.
Shoesmith E. (1985), T. Simpson and the arithmetic mean. *Hist. Math.*, vol. 12, pp. 352 – 355.
--- (1986), Huygens' solution to the gambler's ruin problem. Ibidem, vol. 13, pp. 157 – 164.
--- (1987), The Continental controversy over Arbuthnot's argument for Divine Providence. Ibidem, vol. 14, pp. 133 – 146.
Short J. (1763), Second paper concerning the parallax of the Sun. *Phil. Trans. Roy. Soc.*, vol. 53, pp. 300 – 342.
Simon J. (1887), *Public Health Reports*, vols 1 – 2. London.
Simpson J. Y. (1847 – 1848), Anaesthesia. *Works*, vol. 2. Edinburgh, 1871, pp. 1 – 288.
Simpson J. Y. (1869 – 1870), Hospitalism. Ibidem, pp. 289 – 405.
Simpson T. (1756), On the advantage of taking the mean of a number of observations. *Phil. Trans. Roy. Soc.*, vol. 49, pp. 82 – 93.
--- (1757), Extended version of same. In author's book *Miscellaneous Tracts on Some Curious… Subjects…* London, pp. 64 – 75.
--- (1775), *Doctrine of Annuities and Reversions*. London.
Slutsky E. E. (1912), *Teoria Korreliatsii* (Correlation theory). Kiev.
Smit Maria (1934, in Russian), Against idealistic and mechanistic theories in the theory of Soviet statistics. *Planovoe Khoziastvo*, No. 7, pp. 217 – 231.
Snow J. (1855), On the mode of communication of cholera. In *Snow on Cholera.* New York, 1965, pp. 1 – 139.
Sofonea T. (1957), E. Halley (1656 – 1742) und seine Sterbetafel 300 Jahre nach seiner Geburt. *Het Verzerkerings Archief*, t. 34, pp. 31* – 42*.
Stigler S. M. (1986), *The History of Statistics.* Cambridge, Mass. Harvard Univ. Press.
Strotz R. H. (1978), Econometrics. In Kruskal & Tanur (1978, pp. 188 – 197).
Süssmilch J. P. (1741), *Die Göttliche Ordnung in den Veränderungen des menschlichen Geschlechts, aus der Geburt, dem Tode und der Fortpflanzung desselben.* Berlin, 1761. Several later editions.
--- (1758), Gedancken von dem epidemischen Krankheiten. In Wilke J., Editor (1994), *Die königliche Residenz und die Mark Brandenburg im 18. Jahrhundert*. Berlin, pp. 69 – 116.
Tait P. G. (1892), Poincaré's *Thermodynamics*. *Nature*, vol. 45, pp. 245 – 246.
Takácz (Takács) L. (1982), Ballot problems. In Kotz S., Johnson N. L., Editors (1982 – 1989), *Enc. of Statistical Sciences*, vols 1 – 9 + Supplement volume. New York, vol. 1, pp. 183 – 188). Second edition: Hoboken, New Jersey, 2006.
--- (1994), The problem of points. *Math. Scientist*, vol. 19, pp. 119 – 139.
Tassi P. (1988), De l'exhaustif au partiel. Un peu d'histoire sur le développement des sondages. *J. Soc. Stat. Paris*, t. 129, No. 1 − 2, pp. 116 – 132.

Thornberg A. (1929), Statistics and [the] trade union movement. *Nordic Stat. J.*, vol. 1, pp. 33 – 35.
Thorp W. L. (1948), Statistics and foreign policy. *J. Amer. Stat. Assoc.*, vol. 43, pp. 1 – 11.
Timerding H. E. (1915), *Analyse des Zufalls.* Braunschweig.
Todhunter I. (1865), *History of the Mathematical Theory of Probability from the Time of Pascal to That of Laplace.* New York, 1949, 1965.
Toomer G. J. (1974), Hipparchus on the distances of the sun and moon. AHES, vol. 14, pp. 126 – 142.
Truesdell C. (1984), *An Idiot's Fugitive Essays on Science.* New York. Collected reprints of author's prefaces and reviews on history and philosophy of natural sciences.
Tsinger V. Ya. (1862), *Sposob Naimenshikh Kvadratov* (Method of Least Squares). Dissertation. Moscow.
Tukey J. W. (1962), The future of data analysis. *Coll. Works*. Monterey, Calif., 1986, vol. 3, pp. 391 – 484.
Tutubalin V. N. (1972), *Teoria Veroiatnostei v Estestvoznanii.* (Theory of probability in natural science). Moscow. **S, G,** 45.
Whitaker Lucy (1914), On the Poisson law of small numbers. *Biometrika*, vol. 10, pp. 36 – 71.
White A. D. (1896), *History of the Warfare of Science with Theology in Christendom*, vols 1 – 2. London – New York, 1898. Many later editions up to 1955.
Whittaker E. T., Robinson G. (1924), *Calculus of observations.* London, 1949. [London, 1952.]
Wilks S. S. (1968), Statistics. In *Enc. Brit*., vol. 21, pp. 162 – 166.
Winkler W. (1931), Ladislaus von Bortkiewicz. *Schmollers Jahrbuch f. Gesetzgebung, Verwaltung u. Volkswirtschaft im Deutschen Reich,* 55. Jg., pp. 1025 – 1033.
Wittstein Th. (1867), *Mathematische Statistik.* Hannover.
Woolhouse W. S. B. (1873), On the philosophy of statistics. *J. Inst. Actuaries*, vol. 17, pp. 37 – 56.
You Poh Seng (1951), Historical survey of the development of sampling theories and practice. *J. Roy. Stat. Soc*., vol. A114, pp. 214 – 231. Reprint: Kendall & Plackett (1977, pp. 440 – 458).
Youshkevitch A. A. (1974), Markov. *Dict. Scient. Biogr*, vol. 9, pp. 124 – 130.
Yule G. U. (1900), On the association of attributes in statistics. *Phil. Trans. Roy. Soc.*, vol. A194, pp. 257 – 319. Also in author's book *Statistical Papers.* London, 1971, pp. 7 – 69.
Zabell S. L. (1988), The probabilistic analysis of testimony. *J. Stat. Planning and Inference*, vol. 20, pp. 327 – 354.
Zhuravsky D. P. (1846), *Ob Istochnikakh i Upotreblenii Statisticheskikh Svedenii* (On the Sources and Use of Statistical Materials). Kiev.

Acknowledgement. A very short version of same has appeared in
Wiley StatsRef. Statistics Reference Online, pp. 1 – 13.
I am unable to see it since, for me, its appropriate address is too complicated.