

Oscar Sheynin

History of Statistics

Berlin, 2012

ISBN 978-3-942944-20-5

© Oscar Sheynin 2012

NG Verlag Berlin

Contents

0. Introduction

- 0.1. General information
- 0.2. The pertinent disciplines and their stages

1. Prehistory

- 1.1. Randomness
- 1.2. Probability
- 1.3. Cause vs chance
- 1.4. Expectation
- 1.5. Astronomy
- 1.6. Astrology
- 1.7. Treatment of observations

2. Early history

- 2.1. Stochastic ideas in science and society
- 2.2. Mathematical investigations

3. Jakob Bernoulli and the law of large numbers

- 3.1. Bernoulli's works
- 3.2. *Art of Conjecturing*: part four
- 3.3. Bernoulli's contemporaries

4. De Moivre and the De Moivre – Laplace limit theorem

- 4.1. *The Measurement of Chance* (1712)
- 4.2. Life insurance
- 4.3. *The Doctrine of Chances* (1718, 1738, 1756)
- 4.4. The De Moivre – Laplace limit theorem

5. Bayes

- 5.1. The Bayes formula and induction
- 5.2. The limit theorem
- 5.3. Additional remark

6. Other investigations before Laplace

- 6.1. Stochastic investigations
- 6.2. Statistical investigations
- 6.3. Treatment of observations

7. Laplace

- 7.1. Theory of probability
- 7.2. Theory of errors
- 7.3. Critical considerations

8. Poisson

- 8.1. Probability and statistics
- 8.2. Theory of errors
- 8.3. Criminal statistics
- 8.4. Statistical physics
- 8.5. Medical statistics

9. Gauss, Helmert, Bessel

- 9.1. Gauss
- 9.2. Helmert
- 9.3. Bessel

10. The second half of the nineteenth century

- 10.1. Cauchy
- 10.2. Bienaymé
- 10.3. Cournot
- 10.4. Buniakovsky
- 10.5. Quetelet
- 10.6. Galton
- 10.7. Statistics
- 10.8. Statistics and natural sciences
- 10.9. Natural scientists

11. Bertrand and Poincaré

- 11.1. Bertrand: general information
- 11.2. Bertrand: the random chord
- 11.3. Poincaré

12. Chebyshev

- 12.1. His contributions
- 12.2. His lectures
- 12.3. Some general considerations

13. Markov, Liapunov, Nekrasov

- 13.1 Markov: personal traits
- 13.2. Markov: scientific issues
- 13.3. Markov: main investigations
- 13.4. Liapunov
- 13.5. Nekrasov

14. The birth of mathematical statistics

- 14.1. Stability of statistical series
- 14.2. The Biometric school
- 14.3. The merging of the two streams?

Supplement: axiomatization**Bibliography****Index of names**

0. Introduction

I do feel how wrongful it was to work for so many years at statistics and neglect its history
K. Pearson (1978, p. 1)

0.1. General Information. This book is intended for those interested in the history of mathematics or statistics and more or less acquainted with the latter. It will also be useful for statisticians. My exposition is based, in the first place, on my own investigations published over some 35 years and monograph (2009) and I stop at the axiomatization of probability and at the birth of the real mathematical statistics, i.e., at Fisher. In § 9.1.3 I succeeded in greatly simplifying Gauss' mature justification of least squares.

Among the preceding literature I single out the great work of Hald (1900; 1998). However, his second book does not touch on the contribution of the Continental direction of statistics (see my § 14.1) and only describes everything from a modern point of view. It is therefore only intended for highly qualified readers (my account is much easier to understand). Second, Hald does not describe the contents of any given contribution and the reader will not know what exactly was contained in, say, any of Laplace's memoirs.

At least in my field the situation is greatly worsened by bad reviewing. Reputable publishers sometimes reprint literature without consulting their authors so that the unsuspecting reader gets dated information. Gnedenko & Sheynin (1978/1992) reprinted in 2001 in the volume put out by Birkhäuser is a good example. Many bad books are also appearing because their manuscripts had not been properly reviewed. Subsequent comments are often no more than sweet nothings or downright misleading. An ignorant author who stated that Poisson had proved the strong law of large numbers was praised as a scholar of the first rank.

In some cases the cause of such facts is apparently well described by the saying *Scratch my back, and I'll scratch yours*. Incidentally, this is a consequence of publishers supplying free copies of their new books to editors of periodicals. Then, abstracting journals are as a rule publishing whatever they get from their reviewers. But, first and foremost, the scientific community wrongly does not set high store on that most important work. Even worse: *Truth is dismissed as an old-fashioned superstition*. This conclusion (Truesdell 1984, p. 292) which concerned scientific work in general fell on deaf ears.

About 1985 the then Editor of *Historia Mathematica* visited Moscow and made a report at the (now, Vavilov) Institute for History of Natural Sciences and Technology. Answering a question, he said that only a few readers of his periodical read Russian. I do not think that that situation had changed, much to the detriment of science. For that matter, students of the humanities certainly become versed in older masters, but not in modernity. How, indeed, can we otherwise explain why did a young offspring of the British Royal family go to a fancy-dress party clad as a German officer of SS?

With sincere gratitude I recall the late Professors Youshkevitch, without whose assistance I, living in Moscow, would have been unable to publish abroad, and Truesdell, the Editor of the *Archive for History of Exact Sciences*, who had to busy himself with my English and compelled me to

pay due attention to style. In 1991, after moving to Germany, I became able to continue my work largely because of Professor Pfanzagl's warm support. He secured a grant for me (which regrettably dried up long ago) from Axel-Springer Verlag. Professor Strecker essentially helped me to prepare and publish both English editions of my Russian book (1990). In my papers, I had acknowledged the help of many colleagues including the late Doctors Chirikov (an able mathematician whose bad health thwarted his scientific career) and Eisenhart. Professor Herbert A. David (Iowa State University) and especially Professor Ulrich Krengel provided useful comments on this text.

A final remark. According to some clever regulation, Bernstein, who published many contributions abroad and spelled his name in that way, should now be called Bernshtein. This is ugly and it corrupts his pen-name (if not real name). And why then are we not ordered to spell Markof, Chuprof? When I began publishing abroad, I had not chosen the best spelling of my name, but it became my pen-name, and I refuse to change it.

Some explanation

Abbreviation

- CLT – central limit theorem
- LLN – law of large numbers
- MLSq – method of least squares
- W-i – Gauss, *Werke*, Bd. i
- W/Erg-i – Gauss, *Werke*, Ergänzungsreihe, Bd. i

Notation

- $[ab] = a_1b_1 + \dots + a_nb_n$ (introduced by Gauss).
- $\ln x = \log \text{ nat } x$, $\lg x = \log_{10} x$

References in text

A double page number, e.g. 59/216, means that either the pertinent source has double paging, or a reference to a later edition, or that it was translated from Russian into English with p. 59 of the original contribution corresponding to p. 216 of the translation.

0.2. The Pertinent Scientific Disciplines and Their Stages

My subject covers a great chronological period and is very wide since it includes the theory of probability and statistics, which are difficult to separate, while statistics itself is a vast subject which ought to be subdivided. I also have to explain the relation of the theory of errors to statistics. In addition, I subdivide the history of the development of these disciplines into stages to help the readers grasp at once their general outline.

Theory of Probability

1. Its prehistory (from Aristotle to the mid-17th century).
2. Its early history (from Pascal and Fermat to Jakob Bernoulli).
3. The creation of its initial version (completed by Jakob Bernoulli, De Moivre and Bayes).
4. Its development as an applied mathematical discipline (from Bayes to Laplace and Poisson to Chebyshev).
5. A rigorous proof of its limit theorems (Chebyshev, Markov, Liapunov) and its gradual transition to the realm of pure mathematics.

6. Axiomatization.

In the second half of the 19th century and the first decades of the 20th mathematicians barely recognized probability theory and perhaps to our day all but ignore the Gaussian theory of errors.

Mathematical Statistics. It originated in the early years of the 20th century in the Biometric school and the Continental direction of statistics and Fisher moved it to the realm of pure mathematics. Its aim is the systematizing, processing and utilizing statistical data, – information on the number of the specified objects (Kolmogorov & Prokhorov 1988/1990, p. 138). Unlike *theoretical statistics*, it does not include collection of data or exploratory data analysis which means revealing general structures in the data (e. g., blunders, systematic influences, deception).

The Statistical Method. Usually, *statistics* is meant to study population and the term *statistical method* is applied in all other instances. The statistical method underwent three stages. At first, conclusions were being based on qualitative regularities conforming to the essence of ancient science. Indeed, a Roman scholar Celsus (1935, p.19) stated:

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.

During the second stage (Tycho in astronomy, Graunt in demography and medical statistics) statistical data became available. Conclusions were made by means of simple stochastic ideas and methods or even directly, as before. At the present stage, which dates back to Poisson, inferences are being checked by quantitative stochastic rules.

The Theory of Errors. From its origin in the mid-18th century and until the 1920s the stochastic theory of errors had been a most important chapter of probability theory (P. Lévy 1925, p. vii) and mathematical statistics borrowed from it its principles of maximum likelihood and minimal variance. It is the application of the statistical method to the treatment of observations.

The determinate error theory examines the process of measurement without applying stochastic reasoning and is related to exploratory data analysis and experimental design. Consequently, it studies systematic errors. Its application began in ancient astronomy (§ 1.5) but its real development was due to the differential calculus which ensured the study of functions of measured magnitudes. Gauss and Bessel assumed that each instrument was faulty unless and until the ensuing random and systematic errors were minimized. Thus originated a new stage in experimental science.

The theory of errors has its own stages. Ancient astronomers were dealing with observations as they saw fit. At the second stage, beginning with Tycho Brahe, observations ceased to be *private property*, but their treatment was not yet corroborated by quantitative considerations. This happened during the third stage (T. Simpson, Lambert), and the final, fourth stage was the completion of the classical theory of errors (Laplace and especially Gauss) although Helmert fruitfully continued the relevant investigations.

1. Prehistory

I trace the prehistory of statistics until Kepler and Galileo inclusively and describe the appearance of randomness and probability as philosophical notions. Statistical considerations were mostly based on general impressions. The arithmetic mean appeared in astronomy as a universal estimator. Kepler rejected the Ptolemaic system of the world.

Key words: randomness, probability, cause vs chance, qualitative correlation, expectation

1.1. Randomness

Is an infinite (a much more difficult question: a finite) number sequence random or not? This is a fundamental problem. Another point is the role of randomness in natural sciences, for example in evolution of species or the kinetic theory of gases. Then, in statistics, a random variable should be statistically stable, but in natural science this restriction is not necessary, cf. Poincaré (1896/1912, p. 3), so how to check stability? All this exonerates the need to study the history of randomness, and, incidentally, to see how a philosophical concept becomes a mathematical notion.

Early scientists threw light upon randomness. Aristotle's examples of random events are a sudden meeting of two acquaintances (*Phys.* 196b30) and a sudden unearthing of a buried treasure (*Metaphys.* 1025a). Lack of aim or intersection of chains of events is also seen in Hobbes' remark (1646/1840, p. 259):

When a traveller meets with a shower, the journey had a cause, and the rain had a cause [...], but because the journey caused not the rain, nor the rain the cause, we say that they were contingent one to another.

Cournot (1843, § 40) revived the first example due to Aristotle as an intersection of two independent chains of events and both illustrate one of Poincaré's interpretations of randomness (1896/1912, p. 4): if equilibrium was unstable, a small cause determined a considerable effect. Again, an event was random if its causes were complicated and numerous.

I continue to dwell on Aristotle, but leave aside several other ancient philosophers because their understanding of randomness seems difficult to explain. Aristotle's special example (*Phys.* 199b1; also see *De generatione animalium* 767b5) mentioned deviations from law, monstrosities. The first departure of nature from the type is *that the offspring should become female instead of male; [...] as it is possible for the male sometimes not to prevail over the female. [...]* He did not consider such events random; indeed, he (e. g., *De Caelo* 283b) stated that chance did not occur always or usually. Possibly, however, the sex of the offspring is determined either by small, or by complicated and numerous causes, so that the birth of a female (or a male) is a random event.

An addition is necessary. A chaotic process engendered by a small corruption of the initial conditions of motion can lead to exponential deviation of the appropriate path. A coin toss has a constant number of outcomes whose probabilities persist, whereas chaotic motions imply a

rapid increase of their instability with time and countless positions of their possible paths.

According to Aristotle (e. g., *Metaphys.* 1064b15), *none of the traditional sciences busies itself about the accidental* [...]. Neither does the theory of probability consider the accidental, but rather studies the laws of randomness. Randomness was indirectly mentioned in Indian philosophy as intersection of chains of events (Belvalkar et al 1927, p. 458):

The crow had no idea that its perch would cause the palm-branch to break, and the palm-branch had no idea that it would be broken by the crow's perch; but it all happened by pure Chance.

In medicine, we find randomness occurring when equilibrium is unstable (Galen, 2nd century/1951, p. 202):

In those who are healthy [...] the body does not alter even from extreme causes, but in old men even the smallest causes produce the greatest change.

Chance was recognized in biology as an intrinsic feature of nature. Thus, Harvey (1651/1952, p. 338) stated that spontaneous generation occurred accidentally and even Lamarck (1809/1873, p. 62) kept to the same opinion. He (1815, p. 133) also maintained that the deviations from the divine lay-out of the tree of animal life had been occasioned by a *cause accidentelle*.

The Old Testament also contains statements concerning randomness, for example: *A certain man drew his bow at a venture and struck the King of Israel* (1 Kings 22:34, 2 Chronicles 18:33). Kepler (1606/2006, p. 163) denied it, called it *an idol*, but his laws of planetary motion were unable to justify the values of the eccentricity of their orbits. He (1618 – 1621, 1620/1952, p. 932) had to consider them random, caused by disturbances, deviations from (Divine) laws of nature and Poincaré (1896/1912, p. 1) formulated the dialectical link between randomness and necessity (but did not mention regularity of mass random events):

There exists no domain where precise laws decide everything, they only outline the boundaries within which randomness may move. In accordance with this understanding, the word randomness has a precise and objective sense.

Kant (1755/1910, p. 337) repeated Kepler's pronouncement about deviations from laws. And, in spite of Newton's proof that the eccentricities were determined by the planets' velocities, Laplace (1796/1884, p. 504, note 7) followed suit.

1.2. Probability

Aristotle (*Anal. Priora* 70a0) also reasoned about logical or subjective probability which is *A generally approved proposition* and (*Rhetorica* 1376a19) recommended the use of probabilities in law courts. In the Talmud, the part of forbidden food should not have exceeded certain limits and Maimonides, in the 12th century (Rabinovitch 1973, p. 41), listed seven relevant ratios, i.e., seven different probabilities of eating it. His works also

contain an embryo of a random variable (Ibidem, p. 74): *Among contingent things some are very likely, other possibilities are very remote, and yet others are intermediate.* In the new time one of the first to follow suit in natural science was Maupertuis (1745/1756, pp. 120 – 121) who explained instances when a child resembled one of his remote ancestors, as well as mutations by *non-uniform* randomness.

Subjective probability can lead to sophisms. Here is the opinion of Rabbi Shlomo ben Adret, 1235 – 1310 (Rabinovitch 1973, p. 40). There are several pieces of meat, all of them kosher except one. Eating the first one is allowed, since it is likely kosher; the same with the second one etc, and when only two pieces of meat are left, the forbidden piece was likely already eaten and they are also allowed.

1.3. Cause vs Chance.

In jurisprudence, attempts to separate necessity (divine punishment) from chance were made in ancient India (Bühler 1886/1967, p. 267): if a witness in law-suits pertaining to loans within seven days after he had given evidence experienced a misfortune, *he shall be made to pay the debt and a fine*, – he was considered a liar. An attempt to separate divine design and chance was the main aim of De Moivre's *Doctrine of Chances* (§ 4.3).

In the Old Testament we also find a separation of necessity and chance: Job (9:24 and 21:17 – 18) decided that the world was *given over to the wicked* [this being the cause] since *their lamp was put out rarely*. The Talmud (Taanid 3⁴) decides whether deaths in a town were normal events or occasioned by the beginning of a plague epidemic and it seems likely that a disregarded probability of 1/8 meant an occurrence of the first alternative.

Galileo (1613) managed to separate cause (regular rotation of the newly discovered sunspots with the Sun itself) and chance (their random proper motion relative to the Sun's disc). The same goal is still with us in mathematical statistics, e. g., in clinical trials. Galen (1946, p. 113) indirectly mentioned it:

What is to prevent the medicine which is being tested from having a given effect on two [of three] hundred people and the reverse effect on twenty others, and that of the first six people who were seen at first and on whom the remedy took effect, three belong to the three hundred and three to the twenty without your being able to know which three belong to the three hundred and three to the twenty [...]. You must needs wait until you see [...] very many people in succession.

Qualitative correlation, corresponding to the qualitative nature of ancient science, was introduced and served to separate cause from chance. Here is an example (Hippocrates, flourished 400 BC, 1952, No. 44): *fat men are apt to die earlier than those who are slender.* Or, Aristotle (*Problemata* 892a0): *Why is it that fair men and white horses usually have grey eyes?* Statements amounting to qualitative correlation can be found in contributions of many ancient scientists. Again, conforming to the nature of ancient science, conclusions had been made by issuing from general impression. Thus, climatic belts were introduced in antiquity without any quantitative support. In 1817, Humboldt introduced them anew, but based them on mean yearly temperatures.

1.4. Expectation

Expectation was introduced on a layman's level much earlier than in science. Maimonides (Rabinovitch 1973, p. 164) noted that a marriage settlement (providing for a widow or a divorced wife) of 1000 *zuz* can be sold for 100 [of such monetary units], but a settlement of 100 can be sold only for less than 10. It follows that there existed a more or less fixed expected value of a future possible gain. Large payments were thus valued comparatively higher and this psychologically determined subjective attitude can also be traced in later lotteries up to our days (Cohen et al 1970; 1971).

A marriage settlement is a particular case of insurance; the latter possibly existed in an elementary form even in the 20th century BC (Raikher 1947, p. 40). Another statement of Maimonides (Rabinovitch 1973, p. 138) can also be linked with jurisprudence and might be considered as an embryo of Jakob Bernoulli's (1713, part 4) thoughts about arguments:

One should not take into account the number of doubts, but rather consider how great is their incongruity and what is their disagreement with what exists. Sometimes a single doubt is more powerful than a thousand other doubts.

Expectation was indirectly mentioned in ancient India (Al-Biruni (973 – 1048) 1887, vol. 2, pp. 158 – 160): in law-courts, in certain cases many kinds of oaths had been demanded in accordance with the value of the object of the claim. The probability of lying with impunity multiplied by that value was the expectation of fraudulent gain.

Expectation is connected with mean values, and, in moral issues, with mean behaviour. Aristotle (for example, *Ethica Nicomachea* 1104a24) believed that mean behaviour, moderation possessed optimal properties. Analogous statements had appeared even earlier in ancient China; the doctrine of means is attributed to a student of Confucius (Burov et al 1973, pp. 119 – 140). Again, a similar teaching existed in the Pythagorean school (Makovelsky 1914, p. 63), and Nicomachus of Gerasa, ca. 100 BC (1952, p. 820) stated that a perfect number was a mean between numbers the sum of whose divisors was less, and greater than the number itself; was between excess and deficiency. In medicine the mean was considered as the ideal state (of health). Thus (Galen 1951, pp. 20 – 21): *A good constitution is a mean between extremes*. In games of chance the (arithmetic) mean was believed to possess certain stochastic properties (§ 2.1.1). In the new time, the arithmetic mean became the main estimator of the constants sought in the theory of errors and has been applied in civil life.

The Talmud (Jerus. Talmud/Sangedrin 1⁴) was also concerned with the redemption of the first born by lot. Moses wrote *Levite* on 22, 273 ballots and added 273 more demanding five shekels each. Only 22,000 *Levite* ballots were needed so that Moses ran the risk of losing some of the required money. Nevertheless, the losing ballots turned up at regular intervals, which was regarded as a miracle. The existence of the superfluous ballots was not explained; the Israelites were apparently mistakenly thinking that the last 273 of them to draw the lots will be the losers, see a similar example in Tutubalin (1972, p. 12).

1.5. Astronomy

Ancient astronomers did not mention randomness, but they knew that some, and only some errors acted systematically (for example, refraction), e. g., Ptolemy, 2nd c. (1984, IX, 2). Lloyd (1982, p. 158, n66) noted that Ptolemy had a special term for *significant and noteworthy* differences. I myself cite Ptolemy (1956, III, 2, p. 231):

Horoscopic instruments [...] are frequently capable of error, the solar instruments by the occasional shifting of their positions or of their gnomons, and the water clocks by stoppages and irregularities in the flow of the water from different causes and by mere chance.

All this (and the following) directly bears upon the determinate error theory.

Ancient astronomers had been able to *select best conditions* (time of observation) for given errors to influence the results as little as possible. Hipparchus, 1st c. BC (Toomer 1974, p. 131) was aware of that fact and Aaboe & De Solla Price (1964, pp. 2 and 3) concluded that

In the pre-telescopic era there is [...] a curious paradox that even a well-graduated device for measuring celestial angles [...] is hardly a match for the naked and unaided eye judiciously applied.

They even mentioned *qualitative measurements* in the title of their paper. Neugebauer (1948/1983, p. 101) more carefully remarked that in antiquity

All efforts were concentrated upon reducing to a minimum the influence of the inaccuracy of individual observations with crude instruments [...].

The second feature of ancient astronomy was the *determination of bounds* for the constants sought, *a well known technique, practiced for instance by Aristarchus, Archimedes and Eratosphenes* (Toomer 1974, p. 139). The third and last feature was the practice of regular observations. Neugebauer (1975, p. 659) credited Archimedes and Hipparchus with systematic observations of the apparent diameters of the sun and the moon. And Hipparchus could have otherwise hardly been able to compile his star catalogue.

Al-Biruni (1967, pp. 46 – 51), rejected four indirect observations of the latitude of a certain town in favour of its single and direct measurement. He (1967) tells us about his own regular observations, in particular (p. 32) for predicting dangerous landslides (which was hardly possible; even latitude was determined too crudely).

Levi ben Gerson (Goldstein 1985, pp. 29, 93 and 109) indirectly but strongly recommended regular observations. In the first two cases he maintained that they proved to him that the declinations of the stars and the lunar parallax respectively were poorly known.

Al-Biruni (1967, p. 152) was the first to consider, although only qualitatively, the propagation of computational errors and the combined effect of observational and computational errors:

The use of sines engenders errors which become appreciable if they are added to errors caused by the use of small instruments, and errors made by human observers.

His statement (Ibidem, p. 155) on the observation of lunar eclipses for determining the longitudinal difference between two cities testified to his attempt to exclude systematic influences from final results: Observers of an eclipse should

Obtain all its times [phases] so that every one of these, in one of the two towns, can be related to the corresponding time in the other. Also, from every pair of opposite times, that of the middle of the eclipse must be obtained.

Contrary to modern notion, ancient astronomers regarded their observations as their private property, did not report rejected results or explain their methods of treating them (Pannekoek 1961/1989, pp. 339 – 340). It is possible that, when selecting point estimates for the constants sought, they had been choosing almost any number within some appropriate bounds. Indeed, modern notions about treating observations, whose errors possess a *bad* distribution, justify such an attitude, which, moreover, corresponds with the qualitative nature of ancient science.

1.6. Astrology

It was practised in good faith by the most celebrated astronomers, and qualitative correlation was present there as well. Kepler considered himself the founder of scientific astrology, of a science of correlational rather than strict influence of heaven on men and states. Thus (Kepler 1619/1997, book 4, pp. 377 – 378), his heavenly bodies were not Mercury, but Copernicus and Tycho Brahe, and the constellations at his birth only woke rather than heartened his spirit and the abilities of his soul. And (1610/1941, p. 200), *heaven and earth are not coupled as cog-wheels in a clockwork*. Before him Tycho likely held the same view (Hellman 1970, p. 410). As an astrologer, Ptolemy (1956, I, 2 and I, 3), also believed that the influence of the heaven was a tendency rather than a fatal drive, that astrology was to a large extent a science of qualitative correlation, and Al-Biruni (1934, p. 232) likely thought the same way: *The influence of Venus is towards [...], The moon tends [...]*. Maimonides (1977, pp. 118 – 129) was an exception: *The theories of the astrologists are devoid of any value*.

For Kepler, the main goal of astrology was not the compilation of horoscopes concerning individuals, but the determination of tendencies in the development of states for which such circumstances as geographical position, climate, etc, although not statistical data, should also be taken into account.

1.7. Treatment of Observations

The treatment of direct measurements is studied by the theory of errors (see below), but it had to be done from most ancient times. In § 1.5, I mentioned the qualitative approach to it by ancient astronomers. In Kepler's time, and possibly even somewhat earlier, the arithmetic mean became the generally accepted estimator of such 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 demanding 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 admissible, so that either the observations or the underlying theory were faulty. Kepler reasonably trusted Tycho's observations and his inference was obvious.

I am unaware of any sound discussion of Tycho's observations and a particular pertinent question also suggests itself. Temporary removals of at least one of his instruments had been certainly necessary. This would have likely led to systematic shifts in the mean measurements, so how did he manage in such cases?

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.

In astronomy, numerous observations distributed over years and even centuries are necessary for determining astronomical constants and estimating, say, the proper motion of stars. In other branches of natural sciences the situation is not so straightforward. Boyle (1772/1999, p. 376), the cofounder of scientific chemistry and co-author of the Boyle – Mariotte law, when discussing experiments rather than observations, kept to his own rule:

Experiments ought to be estimated by their value, not their number; [...] a single experiment [...] may as well deserve an entire treatise [...]. As one of those large and orient pearls may outvalue a very great number of those little [...] pearls, that are to be bought by the ounce [...].

Flamsteed's attitude would have also been advisable to describe. This is, however, difficult, but at least I am referring to Baily (1835, p. 376) and Rigaud (1841, pp. 129 – 131).

So are series of observations always needed? All depends on the order of the random errors, their law of distribution, on the magnitude of systematic influences, the precision and accuracy required (the first term concerns

random errors, the second one describes systematic corruption) and on the cost of observation. In any case, it is hardly advisable to dissolve a sound observation in a multitude of worse measurements. Specifically, the danger of systematic corruption demands that a programme of its elimination be drawn up and this means that the number of observations should be known beforehand. Sequential analysis is ruled out.

2. The Early History

In the 17th century gambling led to the development of the nascent probability theory; in jurisprudence, a bit later stochastic ideas began to be applied for objectively solving civil cases. Some studies pertaining to insurance of life had been based on probabilities, mortality tables appeared and elements of population statistics and statistics itself had emerged.

Key words: games of chance, theory of probability, mortality, population statistics, expectation

2.1. Stochastic Ideas in Science and Society

2.1.1. Games of Chance. They fostered the understanding of the part of chance in life whereas mathematicians discovered that such games provided formulations of essentially new problems. Pascal (1654b/1998, p. 172) suggested a remarkable term for the nascent theory, – *Aleae geometria, La Géométrie du hazard*. Later Huygens (1657/1920, pp. 57 – 58) prophetically remarked that it was not a simple *jeu d'esprit* and that it laid the foundation *d'une spéculation fort intéressante et profonde*. Leibniz (1704/1996, p. 506) noted that he had repeatedly advocated the creation of a *new type of logic* so as to study *the degrees of probability* and recommended, e. g., in 1703, in a letter to Jakob Bernoulli, to examine in this connection all kinds of games of chance. Actually, Bernoulli began studying them in 1675 (Biermann 1955).

Even in antiquity games of chance provided examples of stochastic considerations (Aristotle, *De caelo* 292a30 and 289b22):

Ten thousand Coan throws [whatever that meant] in succession with the dice are impossible; it is difficult to conceive that the pace of each star should be exactly proportioned to the size of its circle, –

their invariable mutual arrangement cannot be random.

The theory of probability had originated in the mid-17th century rather than earlier. Exactly then influential scientific societies came into being, scientific correspondence became usual and games of chance provided models for posing natural and properly formulated stochastic problems. In addition, they were in the social order of the day. Previously, they had not been sufficiently conducive to the development of stochastic ideas because of the absence of *combinatorial ideas* and of the notion of chance events, of superstition and moral or religious *barriers* (M. G. Kendall 1956/1970, p. 30).

Montmort (1708/1713, p. 6) had testified to the superstition of gamblers; Laplace (1814/1995, pp. 92 – 93) and Poisson (1837a, pp. 69 – 70) repeated his statement (and adduced new examples). Illusions exist even in our time although Bertrand (1888a, p. XXII) had remarked that the roulette wheel had *ni conscience ni mémoire*. Even a just game (with a zero expectation of loss for each participant) is ruinous and is therefore based on superstition. Petty (1662/1899, p. 64) stated that lotteries were *properly a Tax upon unfortunate self-conceited fools* and Arnauld & Nicole (1662/1992, p. 332) indicated that large winnings in a lottery were illusory. They came out against hoping for unlikely favourable events. On the contrary, it is

reasonable to be guarded against unlikely unfavourable events which is the rationale behind the institution of insurance.

On the other hand, gamblers had been noticing interesting regularities. Apparently during 1613 – 1623 Galileo (ca. 1613 – 1623, publ. 1718/1962) wrote a note about the game of dice. He calculated the number of all the possible outcomes of a throw of three dice and testified that gamblers were believing that 10 or 11 points turned out more often than 9 or 12. If only these events are considered (call them *A* and *B* respectively), then the difference between their probabilities

$$P(A) = 27/52, P(B) = 25/52, \Delta P = 1/26 = 0.038$$

can be revealed thus strengthening the trust in mean values (§ 1.4).

Galileo had predecessors, Fournival, the probable author of the *De Vetula* (Bellhouse 2000), and Cardano, *De Vetula*, written in the mid-13th century, considered the throws of three dice. Bellhouse concluded that it had led to *elementary probability calculations being established and known in Europe from about 1250*. He also provided an English translation of its mathematical lines. At the time and even earlier elements of combinatorial mathematics had been certainly known outside Europe; a recent source about ancient India is Raju (2010).

Bellhouse (2005) believes that Cardano had based his stochastic reasoning on *De Vetula*. Cardano (Ore 1953; Hald 1990, pp. 36 – 41) compiled a book on games of chance (dicing, including playing with imagined dice having 3 – 5 sides, and card games) only published in 1663. He enumerated the possible outcomes of throws of three dice and effectively applied the *classical* definition of probability; true, he worked with odds rather than probabilities.

At the end of his life Cardano (1575) compiled his biography which contained a chapter called *Things of worth which I have achieved* (pp. 215 – 219 of the English edition of 1962) where he only mentioned a particular stochastic problem but formulated it incomprehensively. Nevertheless, his was the first discussion of stochastic methods and he (see my § 3.2.3) applied, as other scientists then did, the simplest formula pertaining to the prehistory of the LLN.

2.1.2. Jurisprudence. One of the first tests for separating chance from necessity was provided for the administration of justice (§ 1.3). It seems, however, that the importance of civil suits and stochastic ideas in law courts increased exactly in the mid-17th century. A comparison of the attitudes of Kepler (1610/1941, p. 238) and Jakob Bernoulli (1713, pt. 4, Chapt. 2) is instructive. Kepler refused to answer someone whether his absentee friend was alive or not. Bernoulli, however, was quite prepared to weigh the probabilities of such facts against each other (which was just what Nikolaus Bernoulli did, see § 3.3.2). Descartes (1644/1978, pt. 4, § 205, p. 323) put moral certainty into scientific circulation, apparently bearing in mind jurisprudence. See § 3.2.2 on Jakob Bernoulli's statements about that notion.

Niklaus Bernoulli (§ 3.3.2), in the beginning of the 18th century, devoted his dissertation to the application of the *art of conjecturing* to jurisprudence. Leibniz (1704/1996, pp. 504 – 505) mentioned degrees of proofs and doubts in law and in medicine and indicated that

our peasants have since long ago been assuming that the value of a plot is the arithmetic mean of its estimates made by three groups of appraisers.

That mean was considered as an approximation to the expected value of the plot, cf. § 1.4.

2.1.3. Insurance of Property and Life Insurance. Marine insurance was the first essential type of insurance of property but it lacked stochastic ideas or methods. 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. *Publicke Acte* No. 12 of 1601 (*Statutes of the Realm*, vol. 4, pt. 2, pp. 978 – 979) mentioned *policies of assurance* in marine insurance:

It hathe bene tyme out of mynde [...] in this realme and in forraine nacyons to have assurance of goodes, merchandizes, ships and things adventured.

Life insurance 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 1423 when a Papal bull officially allowed them (Du Pasquier 1910, pp. 484 – 485). 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), the annuitant's age was not usually taken into consideration. Otherwise they had been allowed for only in a generalized way (Kohli & van der Waerden 1975, pp. 515 – 517; Hald 1990, p. 119). The situation began to change at the end of the 17th century.

However, in the 18th, and even in the mid-19th century, life insurance hardly essentially depended on stochastic considerations; moreover, the statistical data collected by the insurance societies as well as their methods of calculations remained secret and honest business based on statistics of mortality barely superseded cheating before the second half of the 19th century. Nevertheless, beginning at least from the 18th century, life insurance strongly influenced the theory of probability, see §§ 4.2 and 6.1.1-3.

De Witt (1671) distinguished four age groups and without proof assumed that the chances of death increased in a definite way from one group to the next one but remained constant within each of them. According to his calculations, the cost of an annuity for *young* men should have been 16 times higher than the yearly premium (not 14, as it was thought). Eneström (1896/1897, p. 66) noted that De Witt's proposed chances of death were contrary to what was calculated and that his risk of dying concerned an infant and was explained misleadingly. But still, a likely corollary of De Witt's work was that the price of annuities sold in Holland in 1672 – 1673 depended on the age of the annuitants (Commelin 1693, p. 1205). De Witt's appendix to the main text (Hendriks 1853, pp. 117 – 118) contained an interesting observation belonging to the prehistory of the LLN. Examining *considerably more than a hundred different classes, each class consisting of about one hundred persons*, he found that a purchaser of ten or more life

annuities will certainly gain profit. Several authors mentioned the practice, possibly justified by intuitive stochastic reasoning, of insuring a number of healthy infants cf. § 3.2.3.

In the same year, De Witt (Hendriks 1853, p. 109) calculated the cost of annuity on several lives (an annuity that should be paid out until the death of the last person of the group, usually, of a married couple) and thus determined the distribution of the maximal term of a series of observations [obeying a uniform law]. Kohli & van der Waerden (1975) described the history of life insurance including the work of De Witt and Huygens (§ 2.2.2).

The first estimation of the present worth of life annuities, based on a table of expectations of life, was made by the Praetorian Prefect Ulpianus (170 – 228), see Hendriks (1852) and Greenwood (1940 and 1941 – 1943). His sources are not known, neither is it clear whether his *expectation* coincided with our present notion, but at least methodologically his table constituted the highest achievement of demographic statistics until the 17th century.

Leibniz (MS 1680?/1872) described his considerations about state insurance, see Sofonea (1957a). He had not studied insurance as such, but maintained that the *princes* should care about the poor, that the society ought to be anxious for each individual etc. Much later Süßmilch (§ 6.2.2) formulated similar ideas.

Tontines constituted a special form of mutual insurance. Named after the Italian banker Laurens Tonti, 1630 – 1695 (Hendriks 1863), they, acting as a single body of participants, 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) proposed flexible tontines with variable ages of their members as well as their initial contributions. Such tontines would then become perpetual bodies rather than remaining only for a few decades in existence. Apparently for the same reason his proposal had not been adopted.

2.1.4. Population Statistics. The Old Testament (Numbers, Chapter 1) reports on a census of those able to bear arms and, accordingly, the Talmud estimated the population of towns only by the number of soldiers *brought forth* [when needed]. In China, in 2238 BC or thereabouts, an estimation of the population was attempted and the first census of the warrior caste in Egypt occurred not later than in the 16th century BC (Fedorovitch 1894, pp. 7 – 21). In Europe, even in 15th century Italy, for all its achievements in accountancy and mathematics (M. G. Kendall 1960),

counting was by complete enumeration and still tended to be a record of a situation rather than a basis for estimation or prediction in an expanding economy.

Only Graunt (1662) and, to a lesser extent, Petty (1690) can be called the fathers of population statistics. 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. However, neither Petty, nor his followers ever introduced any definition of political arithmetic. 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 shows him as a philosopher of science congenial in some respects with Leibniz (pp. 39 – 40):

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 Observations [like Graunt!]) Dedicated this Discourse ...* 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 newly born 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 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 definitely introduced by Süßmilch and Euler (§ 6.2.2).

In spite of the meagre and sometimes wrong information, Graunt was able to compile the first mortality 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.

Huygens (§ 2.2.2) made use of Graunt's mortality table and so did, indirectly, Jakob Bernoulli (§ 3.2.2).

Halley (1694a; 1694b), a versatile scholar and an astronomer in the first place, compiled the next mortality 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 were 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 newly born 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 *would 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. Halley's second note is interesting as a reasoning on the welfare of the population. Thus, he emphasized the need to help the poor, especially by finding them jobs.

Success came immediately. K. Pearson (1978, p. 78) indicated that Halley had made *all the use that a modern actuary could* of his data and that he had computed his life-table *as we should do it today*. Sofonea (1957b, 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 (1725/1756) introduced the 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, see § 0.2.

In 1680 – 1683 Leibniz wrote several manuscripts mostly pertaining to statecraft (§ 6.2.1) and published in 1866 (Leibniz 1986, pp. 340 – 349, 370 – 381, and 487 – 491). He recommended the compilation of *state tables* (numerical or not?) of remarkable facts and their comparison, year with year, or state with state, by a special recording office. He thought it advisable to collect information about scientific achievements, *clever ideas* and medical and meteorological observations, and to establish *sanitary boards* for compiling data on a wide range of subjects (meteorology, medicine, agriculture). One of Leibniz' manuscripts (Leibniz 1986, pp. 456 – 467, or, with a German translation, 2000, pp. 428 – 445) was devoted to political arithmetic. There, he introduced the *moyenne longueur de la vie humaine*, necessary, as he remarked, for calculating the cost of annuities; assumed without substantiation several regularities, for example, that the ratio of mortality to population was equal to 1:40; and wrongly stated that the mortality law for each age group including infants was uniform. Following Arnauld & Nicole (1662/1992, pp. 331 and 332), he discussed *apparence* or *degré de la probabilité* and *apparence moyenne* [expectation]. When discussing the game of dice, Leibniz made two elementary mistakes. Much worse, he argued that the birth-rate could be nine or ten times higher than it actually was.

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.

2.2. Mathematical Investigations

2.2.1. Pascal and Fermat. In 1654 Pascal and Fermat exchanged several letters (Pascal 1654a) 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. About 1400 an anonymous Italian author (Franklin 2001, pp. 294 – 296) correctly solved a particular case of the same problem, but did not introduce expectation.

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, can not be equated to any temporal advantage. [...] We should never balance them with anything worldly. [...] 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 [...].

2.2.2. 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

$$\frac{pa + qb}{p + q}.$$

(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)

$$P(B) = \sum P(A_i)P(B/A_i), i = 1, 2, \dots, n.$$

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 (J. Bernoulli 1713/1999, pp. 167 – 168; De Moivre 1712/1984, Problem 14 and 1718/1756, Problem 20) appear in connection with statistical inspection of mass production.

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 (1895), see Kohli & van der Waerden (1975), discussed stochastic problems connected with mortality and life insurance. Issuing from Graunt's mortality table (§ 2.1.4), Huygens (pp. 531 – 532) introduced the probable duration of life (but not the term itself). On p. 537 he specified that expected life ought to be used in calculations of annuities and the former for betting on human lives. Huygens 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 $0 \leq x \leq 100$.

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 Shoemith (1986).

2.2.3. Newton. 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 Approximatio [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.

On De Moivre see Chapt. 4 and *Bayes theorem* is a misnomer (§ 5.1). Then, Newton never mentioned mean values. In 1971, answering my question on this point, the Editor of his 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 Newton was perhaps unconsciously thinking what De Moivre put into words.

Indeed, I have since found that 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, cf. also my § 7.1-3, 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 idea of a divine reformation of the system of the world was later abandoned, but his recognition of the existence and role of its random disturbances is very important. At the same time Newton (1958, pp. 316 – 318) denied randomness and explained it by ignorance of causes. The future theologian Bentley, in 1693, expressed his thoughts after discussing them with Newton. The texts of two of his sermons, of Newton's letters to him, and an article on Newton and Bentley are in Newton (1958).

Newton (MS 1664 – 1666/1967, pp. 58 – 61) generalized the notion of expectation and 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). K. Pearson (1928) described Newton's later indication of the sources of his estimate and dwelt on Voltaire's adjoining remarks, and, especially, on the relevant work of Condorcet.

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 ...

Most important were the optical experiments; the main sources are Newton's *Lectures* (1669 – 1671) and his *Papers and Letters* (1958).

2.2.4. Arbuthnot. He (1712) assembled the existing data on baptisms in London during 1629 – 1710. He 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.

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 have 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 a finite sequence) 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 H. A. David & Edwards (2001, pp. 9 – 11). Arbuthnot was also the first to publish a trick equivalent to the application of a generating function of the binomial distribution although only for its particular case. Jakob Bernoulli (§ 3.1.2)

actually applied a generating function before Arbuthnot did, but his book only appeared in 1713.

Bellhouse (1989) described Arbuthnot's manuscript written in 1694. There, the author examined the game of dice, attempted to study chronology and to a certain extent anticipated his published note of 1712.

3. Jakob Bernoulli and the Law of Large Numbers

I consider Bernoulli's main work, the *Ars conjectandi* (AC), which blazed a new trail by proving that statistical probability can be considered on a par with the theoretical probability. Also described is the work of his contemporaries.

Key words: law of large numbers, statistical probability, moral certainty, stochastic arguments

3.1. Bernoulli's Works

3.1.1. The Diary (*Meditationes*). There, Bernoulli studied games of chance and the stochastic side of civil law. He (1975, p. 47) noted that the probability of a visitation of a plague in a given year was equal to the ratio of the number of these visitations during a long period of time to the number of years in that period. He thus applied the definition of probability of an event (of statistical probability!) rather than making use of chances. On p. 46, in a marginal note, he wrote out the imprint of a review of Graunt's book (§ 2.1.4) which Bernoulli possibly had not seen. But the most important in the *Meditationes* is a (fragmentary) proof of the LLN which means that Bernoulli proved it not later than in 1690.

3.1.2. The Art of Conjecturing (1713). Its Contents. Niklaus Bernoulli compiled a Preface (J. 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 and 3 of the AC (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.

Part 1 is a reprint of Huygens' tract (§ 2.2.2). Bernoulli also compiled a table which enabled him to calculate the coefficients of x^m in the development of $(x + x^2 + \dots + x^6)^n$ for small values of n . That polynomial to the power of n was the generating function of the binomial $(p + qx)$ with $p = q$.

Part 2 dealt with combinatorial analysis and introduced the *Bernoulli numbers*.

Part 4 contained the LLN. There also is an informal *classical definition of probability* (which Bernoulli had not applied when formulating that law), a reasoning on the aims of the art of conjecturing (determination of probabilities for choosing the best solutions of problems, apparently in civil life) and elements of stochastic logic.

Bernoulli likely considered the art of conjecturing as a mathematical discipline based on probability as a measure of certainty and on expectation and including (the not yet formally introduced) addition and multiplication theorems and crowned by the LLN.

In a letter of 3 Oct. 1703 Bernoulli (Kohli 1975b, p. 509) informed Leibniz about the progress in his work. He had been compiling it for many

years with repeated interruptions caused by his *innate laziness* and worsening of health; the book still lacked its *most important part*, the application of the art of conjecturing to civil life; nevertheless, he, J. B., had already shown his brother [Johann] the solution of a difficult problem, special in its own way [§ 3.2.3], that justified the applications of the art of conjecturing.

Leibniz, in his own letters to Bernoulli, never agreed that observations could secure moral certainty but his arguments were hardly convincing. Thus, he in essence repeated Arnauld & Nicole (1662/1992, pp. 304 and 317) in that the finite (the mind; therefore, observations) can not always grasp the infinite (God, but also, as Leibniz stated, any phenomenon depending on innumerable circumstances).

He understood randomness as something *whose complete proof exceeds any human mind* (Leibniz 1686/1960, p. 288) which does not contradict a modern approach to randomness founded on complexity and he was also right in the sense that statistical determinations can not definitively corroborate a hypothesis. Cf. Cicero (1991, Buch 2, § 17, p. 149): *Nothing is more opposed to calculation and regularity than chance*. Leibniz had also maintained that the allowance for the circumstances was more important than subtle calculations.

Gauss (§§ 9.1.3 and 9.1.5) stated that the knowledge of the essence of the matter was extremely important. Later Mill (1843/1886, p. 353) contrasted the consideration of circumstances with *elaborate application* of probability, but why contrasting rather than supplementing? Anyway, more than a half of Chapter 4 of Part 4 of the AC in essence coincided with passages from Bernoulli's letters to Leibniz.

In 1714, in a letter to one of his correspondents, Leibniz (Kohli 1975b, p. 512) softened his doubts about the application of statistical probabilities. For some reason he added that the late Jakob Bernoulli had *cultivated* the theory of probability in accordance with his, Leibniz' *exhortations*.

3.2. The Art of Conjecturing, Part 4

3.2.1. Stochastic Assumptions and Arguments. Bernoulli used the addition and the multiplication theorems for combining various arguments. Unusual was the non-additivity of the probabilities. Thus, *something* possesses $\frac{2}{3}$ of certainty but its opposite has $\frac{3}{4}$ of certainty; both possibilities are probable and their probabilities are as 8:9. See Shafer (1978) and Halperin (1988). Shafer also studied non-additive probabilities in Lambert's *Architectonic* (1771). Koopman (1940) resumed the study of such probabilities whose sources can be found in the medieval doctrine of *probabilism* that considered the opinion of each theologian as probable. Franklin (2001, p. 83) dated the origin of probabilism as 1611 or (p. 74) even as 1577. Similar pronouncements on probabilities of opinion go back to John of Salisbury (the 12th century) and even to Cicero (Garber & Zabell 1979, p. 46).

Bernoulli (1713/1999, p. 233) wrote *ars conjectandi sive stochastice*, and Bortkiewicz (1917, p. x) put that Greek word into circulation. Prevost & Lhuillier (1799, p. 3) anticipated him, but apparently their attempt was forgotten. The *Oxford English Dictionary* included it with a reference to a source published in 1662.

3.2.2. Statistical Probability and Moral Certainty. 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, 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. He also maintained that in ordinary life people ought to choose what is more probable. This idea goes back to Cicero (1997, Book 1, § 12, p. 7): *Many things are probable and [...] though these are not demonstrably true, they guide the life of the wise man [...]*. A similar statement was formulated in China in the 4th century BC (Burov et al 1972, p. 203):

Who even before battle gains victory by military estimation, has many chances. [...] Who has many chances gains victory, who has few chances does not gain victory. All the less he who has no chances at all.

3.2.3. The Law of Large Numbers. Bernoulli proved a proposition that, beginning with Poisson, is called the LLN. Let r and s be natural numbers, $t = r + s$, n , a large natural number, $v = nt$, the number of independent trials in each of which the studied event occurs with theoretical probability r/t , μ – the number of the occurrences of the event (of the successes). Then Bernoulli proved without applying mathematical analysis that

$$P\left(\left|\frac{\mu}{v} - \frac{r}{t}\right| \leq \frac{1}{t}\right) \geq 1 - \frac{1}{1+c} \quad (1)$$

and estimated the value of v necessary for achieving a given $c > 0$. In a weaker form Bernoulli's finding meant that

$$\lim_{v \rightarrow \infty} P\left(\left|\frac{\mu}{v} - \frac{r}{t}\right| < \varepsilon\right) = 1, \quad (2)$$

where, as also in (1), μ/v was the statistical probability.

Markov (*Treatise*, 1924, pp. 44 – 52) improved Bernoulli's estimate mainly by specifying his intermediate inequalities and K. Pearson (1925), by applying the Stirling formula, achieved a practically complete coincidence of the Bernoulli result with the estimate that makes use of the normal distribution as the limiting case of the binomial law. Pearson (p. 202) considered Bernoulli's estimate of the necessary number of trials in formula (1) *crude* and leading to the ruin of those who would have applied it. He also inadmissibly compared Bernoulli's law with the wrong Ptolemaic system of the world (and De Moivre with Kepler and Newton):

Bernoulli saw the importance of a certain problem; so did Ptolemy, but it would be rather absurd to call Kepler's or Newton's solution of planetary motion by Ptolemy's name!

The very fact described by formulas (1) and (2) was, however, extremely important for the development of probability and statistics; and, anyway, should we deny the importance of existence theorems? Bernoulli's result proved that, given a large number of observations, statistical probability provided moral certainty and was therefore not worse than the theoretical probability. His main aim was to discover whether the limit (2) existed and whether it was indeed unity rather than a lesser positive number. The latter would have meant that induction (from the v trials) was inferior to deduction.

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 law of large numbers* instead of the Bernoulli proposition that allegedly impeded the development of statistics. His own 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 (1912, p. 22; 1924, pt 1, p. 15; 1961, p. 127) kept to a similar view.

That elementary understanding of the LLN has its prehistory, see the statements of De Witt (§ 2.1.3) and Halley (§ 2.1.4). Again, it was thought that the number of successes in n Bernoulli trials with probability p was approximately equal to np . Cardano applied this formula in calculations connected with games of dice (Ore 1953/1963, pp. 152 – 154 and 196).

In astronomy, the arithmetic mean became the universal estimator of the constant sought (§ 1.7). Recall also (§ 2.1.3) the practice of buying annuities *upon several young lives*. Boscovich (1758, § 481) had somewhat vaguely maintained that the sum (not the mean!) of random magnitudes decreased with an increase in the number of terms (Gower 1993, p. 272). My correction also applies to the other statements above to which I am now adding Kepler (Sheynin 1973c, p. 120). He remarked that the total weight of a large number of metal money of the same coinage did not depend on the inaccuracy in the weight of the separate coins. Even Helmert (1905/1993, p. 200) had to refute that mistake.

3.2.4. Randomness and Necessity. Apparently not wishing to encroach upon theology, Bernoulli (beginning of Chapter 1) refused to discuss the notion of randomness and subjectively described the *contingent* but at the beginning of Chapter 4 explained randomness by the action of numerous complicated causes, cf. § 11.3. The last lines of his book stated that some kind of necessity was present even in random things. He referred to Plato who had indeed taught that after a countless number of centuries everything will return to its initial state at the moment of creation. In accordance with that archaic notion of the *Great Year*, Bernoulli thus unjustifiably generalized the boundaries of his law.

It is noteworthy that Kepler (1596) believed that the end of the world was unlikely but his reasoning is difficult to understand. In 1621, in the second edition of that book, he substantiated his conclusion by stating, in essence like Oresme (1966, p. 247) did before him, that two randomly chosen numbers were *probably* incommensurable. Without mentioning the end of the world, Levi ben Gerson (1999, p. 166) stated that the heavenly bodies will be unable to return to their initial position if their velocities were incommensurable. However, I do not see any connection between astronomical distances or velocities and that notion.

3.3. Bernoulli's Contemporaries

3.3.1. Arnauld. Arnauld & Nicole anonymously put out the *Art of Reasoning* (1662). Arnauld, who was the main author, had applied the term *probabilité*, although without a formal definition, and expressed ideas later repeated by Bernoulli (who cited him).

3.3.2. Niklaus Bernoulli. He 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 to use it 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 (cf. § 2.1.2), but he borrowed separate passages from the *Ars* and even from the *Meditationes* (Kohli 1975c, p. 541), never intended for publication. His general references to Jakob do not excuse his plagiarism.

3.3.3. Montmort. He published an anonymous book (1708), important in itself and because of its influence upon De Moivre (Chapter 4) as well as on Niklaus Bernoulli, the correspondence with whom Montmort included in 1713 in the second edition of his work. In the Introduction he noted that, being unable to formulate appropriate *hypotheses*, he was not studying the applications of [stochastic] methods to civil life.

Henny (1975) and Hald (1990) examined Montmort's findings. The latter listed Montmort's main methods: combinatorial analysis, recurrent formulas and infinite series; and the method (the formula) of inclusion and exclusion

$$P(\sum A_i) = \sum P(A_i) - \sum P(A_i \cdot A_j) + \sum P(A_i \cdot A_j \cdot A_k) - \dots$$

(3)

where A_1, A_2, \dots, A_n were events and $i < j < k < \dots$. This formula is a stochastic corollary of a proposition about arbitrarily arranged sets. Here are some problems solved by Montmort:

1) The problem of points. Montmort arrived at the negative binomial distribution and returned to this problem in his correspondence with Niklaus Bernoulli.

2) A study of throwing s points with n dice, each having f faces. Montmort applied the combinatorial method and formula (3).

3) A study of arrangements and, again, of a game of dice. Montmort arrived at the multivariate hypergeometric, and the multinomial distributions.

4) A study of occupancies. Tickets numbered $1, 2, \dots, n$, are extracted from an urn one by one without replacement. Determine the probability that at least one ticket with number k , $1 \leq k \leq n$, will occur at the k -th extraction. Montmort derived the appropriate formulas

$$P_n = 1 - 1/2! + 1/3! - \dots + (-1)^{n-1}/n!, \quad \lim P_n = 1 - 1/e, \quad n \rightarrow \infty.$$

Niklaus Bernoulli and De Moivre returned to this problem, see H. A. David & Edwards (2001, pp. 19 – 29).

3.3.4. Montmort and Niklaus Bernoulli: Their Correspondence. I outline their correspondence of 1710 – 1713 (Montmort 1708/1713, pp. 283 – 414).

1) The strategic game *Her* (Hald 1990, pp. 314 – 322) depending both on chance and decisions made. The modern theory of games studies it by means of the minimax principle. For his part, Bernoulli indicated that the gamblers ought to keep to [mixed strategies].

2) The gambler's ruin. Montmort wrote out the results of his calculations for some definite initial conditions whereas Bernoulli indicated, without derivation, the appropriate formula (an infinite series). Hald believes that he obtained it by means of formula (3). On this point and on the appropriate findings of Montmort and De Moivre see also Thatcher (1957), Takács (1969) and Kohli (1975a).

3) The sex ratio at birth. I only dwell on Bernoulli's indirect derivation of the normal distribution. Let the sex ratio be m/f , n , the total yearly number of births, and μ and $(n - \mu)$, the numbers of male and female births in a year. Denote

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

and let $s = 0(\sqrt{n})$. Then Bernoulli's formula (Montmort 1708/1980, pp. 388 – 394) can be presented as

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

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

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 $\sqrt{2/\pi}$. A. P. Youshkevich (1986) reported that at his request three mathematicians had examined that reasoning and concluded that Bernoulli came close to the local theorem but they did not mention the missing factor. The very fact that it took three mathematicians to deal with that subject is noteworthy.

4) The Petersburg game. In a letter to Montmort, Bernoulli (Ibidem, p. 402) described his invented game. *B* throws a die; if a six arrives at once, he receives an *écu* from *A*, and he obtains 2, 4, 8, ... *écus* if a six only occurs at the second, the third, the fourth, ... throw. Determine the expectation of *B*'s gain. Gabriel Cramer insignificantly changed the conditions of the game; a coin appeared instead of the die, and the occurrence of heads (or tails) has been discussed ever since. The expectation of gain became

$$E\xi = 1 \cdot 1/2 + 2 \cdot 1/4 + 4 \cdot 1/8 + \dots = \infty,$$

(4)

although a reasonable man will never pay any considerable sum in exchange for it.

Additional conditions were being introduced; for example, suggestions were made to neglect unlikely gains, i.e., to truncate the series (4); to restrict beforehand the possible payoff; and to replace expectation by *moral expectation* (§ 6.1.1). Daniel Bernoulli published his memoir in Petersburg, hence the name of the invented game. In addition, Condorcet (1784, p. 714) noted that the possibly infinite game provided only one trial and that only many such games can lead to an expedient solution. Indeed, Freudenthal (1951) proposed to consider a number of games with the role of the gamblers in each of them to be decided by lot. Finally, the Petersburg game caused Buffon (1777, § 18) to carry out the apparently first statistical experiment. He conducted a series of 2048 games; the mean payoff was 4.9 units, and the longest duration of play (in six cases), nine throws. The game introduced a random variable with an infinite expectation.

Spieß (1975) dwelt on the early history of the Petersburg game and Dutka (1988) described later developments and adduced the results of its examination by statistical simulation. However, I especially mention Jorland (1987) who provided a vast relevant picture.

4. De Moivre and the De Moivre – Laplace Limit Theorem

De Moivre contributed to insurance of life by introducing the uniform law of mortality, discussed the initial concepts of probability theory and solved several important stochastic problems. His main achievement was the proof of the first version of the central limit theorem.

Key words: central limit theorem, chance and design, gambler's ruin

4.1. *The Measurement of Chance (1712)*

In his first probability-theoretic work, a preliminary version of his later contributions, De Moivre (1712/1984) justified the notion of expected random gain by common sense rather than defining it formally as has been done later, cf. § 2.2.2; introduced the classical definition of probability usually attributed to Laplace and the multiplication theorem for chances (mentioning independence of the events) and applied the addition theorem, again for chances; and, in solving one of his problems (No. 26), applied the formula (3.3) of inclusion and exclusion. I describe some of his problems; I have mentioned Problem 14 (repeated in De Moivre's *Doctrine of chances*) in § 2.2.2.

1) Problem No. 2. Determine the chances of winning in a series of games for two gamblers if the number of remaining games is not larger than n , and the odds of winning each game are a/b . De Moivre notes that the chances of winning are as the sums of the respective terms of the development of $(a + b)^n$.

2) Problem No. 5. The occurrence of an event has a chances out of $(a + b)$. Calculate the number of trials (x) after which it will happen, or not happen, with equal probabilities. After determining x from the equation

$$(a + b)^x - b^x = a^x,$$

De Moivre assumed that $a/b = 1/q$, $q \rightarrow \infty$, and obtained

$$1 + x/q + x^2/2q^2 + x^3/6q^3 + \dots = 2, \quad x = q \ln 2, \quad (1)$$

which resembles the Poisson distribution.

3) A lemma. Determine the number of chances for the occurrence of k points in a throw of f dice each having n faces. Later De Moivre (1730, pp. 191 – 197; 1718, Problem No. 3, Lemma) solved this problem by means of a generating function of a sequence of possible outcomes of a throw of one die.

4) Problem No. 9 (cf. Pascal's problem from § 2.2.2). Gamblers A and B have p and q counters, and their chances of winning each game are a and b , respectively. Determine the odds of their ruining. By a clever trick that can be connected with the notion of martingale (Seneta 1983, pp. 78 – 79) De Moivre obtained the sought ratio:

$$P_A/P_B = a^q (a^p - b^p) \div b^p (a^q - b^q). \quad (2)$$

He left aside the elementary case of $a = b$.

5) Problem No. 25. Ruining of a gambler during a finite number of games played against a person with an infinite capital. De Moivre described the solution in a generalized way; its reconstruction is due to Takacz (1967, pp. 2 – 3) and Hald (1990, pp. 358 – 360).

4.2. Life Insurance

De Moivre first examined life insurance in the beginning of the 1720s. Issuing from Halley's table (§ 2.1.4), 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 and he solved a number of pertinent problems by applying the integral calculus.

Here is an example (p. 324). Determine the probability of one person outliving another one if the complements of their lives are n and p , $n > p$. Let the random durations of the lives of A and B be ξ and η . Then, since at some moment x the complement of A 's life is $(n - x)$,

$$P(\xi \geq x, \eta = x) = \frac{(n - z)dz}{pn}, \quad P(\xi > \eta) = \int_0^p \frac{(n - z)dz}{pn} = 1 - \frac{p}{2n}.$$

Probabilities of the type of $P(\xi > x)$ easily lead to integral distribution functions.

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 De Moivre. Hald (p. 546) concluded that Simpson's relevant results represented *an essential step forward*.

4.3. The Doctrine of Chances (1718, 1738, 1756)

This work published in three editions, in 1718, 1738, and, posthumously, in 1756, was De Moivre's main achievement. He developed it from his previous memoir (§ 4.1) and intended it for gamblers so that many results were provided there without proof. Then, following the post-Newtonian tradition, De Moivre did not use the symbol of integration; Todhunter inadequately described De Moivre's main finding (§ 4.4) and Laplace (1814/1995, p. 119) did not sufficiently explain it. All this caused his book, whose translation into French contemplated both Lagrange and Laplace, see Lagrange (1776b), to remain barely known for many decades. I refer to its last edition.

In his Introduction, De Moivre listed his main methods: combinatorial analysis, recurrent sequences (whose theory he himself developed) and infinite series; in particular, he applied appropriately truncated divergent series. Also in the Introduction, on pp. 1 – 2, he once more provided the *classical* definition of probability but kept to the previous reasoning on expectation (§ 4.1) and even introduced the *value of expectation* (p. 3), formulated the multiplication theorem for probabilities (not for chances, as previously) and, in this connection, once more mentioned independence. Two events, A and B , were independent, if, as he stated,

$$P(B) = P(B/A), \quad P(A) = P(A/B)$$

(modern notation here and below). For dependent events (p. 6), three in number (say),

$$P(A \cdot B \cdot C) = P(A) P(B/A) P(C/A \cdot B). \quad (3)$$

I list now some of the problems from the *Doctrine* mentioned by Hald (1990, pp. 409 – 413) without repeating those described in § 4.1 and, for the time being, leaving aside the normal distribution.

1) The Huygens additional Problem No. 4 (§ 2.2.2) including the multivariate case. The appearance of the hypergeometric distribution: Problems NNo. 20 and 26.

2) Runs of successes in n Bernoulli trials including the case of $n \rightarrow \infty$: Problems NNo. 34 and 74.

3) Coincidences. A generalization of Montmort's findings (§ 3.3.3) by the method of inclusion and exclusion: Problems 35 and 36.

4) The gambler's ruin: Problems 58 – 71.

5) Duration of game: Problems 58 – 64, 68 – 71.

For the general reader the main merit of the *Doctrine* was the study of many widely known games whereas De Moivre himself, in dedicating its first edition to Newton (reprinted in 1756 on p. 329), perceived his main goal, i. e., the aim of the theory of probability, in working out

A Method of calculating the Effects of Chance [...] and thereby 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.

4.4. The De Moivre – Laplace Theorem

In 1730 De Moivre published his *Miscellanea analytica* and later appended two supplements. He printed the second one (1733) in a small number of copies and sent it out to his colleagues. I only call it a supplement for the sake of tradition; its extant copies in large libraries are bound to that book. In 1738 De Moivre translated it into English and included in the second, and then, in an extended form, in the third edition of the *Doctrine*. Its title includes the words *binomial* $(a + b)^n$ so that, although studying the symmetric binomial, De Moivre thus thought about the general case. He also stated that the transition to the general case was not difficult. On the first page of the Latin original De Moivre noted that he had concluded (at least its mathematical part) about 12 years earlier.

1) In Book 5 of the *Misc. anal.* De Moivre determined the ratio of the middle term of the symmetric binomial to the sum of all of its terms, and in the first supplement to that work he derived, independently from, and simultaneously with Stirling, the so-called Stirling formula. Only the value of the constant, $\sqrt{2\pi}$, the latter communicated to him.

In the same supplement De Moivre included a table of $\lg n!$ for $n = 10$ (10) 900 with 14 decimals; reprinted: (1718/1756, p. 333). Eleven or twelve decimals were correct; a misprint occurred in the value of $\lg 380!$.

2) In the same Book, De Moivre calculated the logarithm of the ratio of the middle term of the binomial $(1 + 1)^n$ to the term removed by l from it,

but only in the second supplement he calculated the ratio of the sum of the terms between the middlemost and the one removed from it by l to the sum of all the terms. It was equal to

$$\frac{2}{\sqrt{2\pi n}} \left(l - \frac{2l^3}{1 \cdot 3n} + \frac{4l^5}{2 \cdot 5n^2} - \dots \right).$$

He calculated this sum either by numerical integration, or, for $l < \sqrt{n}/2$, by leaving only a few of its first terms. For $n \rightarrow \infty$ his main result can be written as

$$\lim P\left[a \leq \frac{\mu - np}{\sqrt{npq}} \leq b\right] = \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{z^2}{2}\right) dz. \quad (4)$$

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

This is the integral De Moivre – Laplace theorem (see § 7.1-3), as Markov (1900/1924, p. 53) called it, – a particular case of the CLT, a term introduced by Polya (1920). Neither De Moivre, nor Laplace knew about uniform convergence with respect to a and b that takes place here.

In 1812, Laplace (§ 7.1-3) proved (4) simpler and provided a correction term allowing for the finiteness of n . De Morgan (1864) was the first to notice the normal distribution in (4). However, he 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 $\tan \infty = \cot \infty = \mp \sqrt{-1}$.

De Moivre (1718/1756, p. 252) mentioned the study of the sex ratio at birth (§ 2.2.4) and illustrated it by imagined throws of dice. His reasoning (and his general considerations) meant that, for him, the binomial distribution was a divine law of nature, stochastic only because of possible deviations from it. De Moivre thus recognized the mutual action of necessity and randomness, cf. § 1.1.

5. Bayes

Bayes proved the inverse law of large numbers by assuming that an unknown constant was a random variable with an unknown law of distribution. His result completed the first version of the theory of probability.

Key words: law of large numbers, inverse law of large numbers, first version of probability theory

5.1. The Bayes Formula and Induction

I dwell on the posthumous memoir (Bayes 1764 – 1765) complete with the commentaries by 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

$$P(A_i/B) = \frac{P(B/A_i)P(A_i)}{\sum_{j=i}^n P(B/A_j)P(A_j)} \quad (1)$$

and it were I. W. Lubbock & J. E. Drinkwater-Bethune who first applied that term, as noted by David & Edwards (2001, p. 215), and Cournot (1843, § 88) followed suit. Bayes had in essence introduced induction into probability and his approach that assumed the existence of prior probabilities or distributions greatly influenced the development of mathematical statistics.

A modern encyclopaedia (Prokhorov 1999) contains 14 items mentioning him, for example, Bayesian estimator, Bayesian approach. There also, on p. 37, the author of the appropriate entry mistakenly attributes formula (1) to Bayes.

Bayes studied an imaginary experiment, a ball falling on point r situated in a unit square $ABCD$, to the left or to the right of some straight line MN parallel to, and situated between AB and CD . If, after $(p + q)$ trials, the point r occurred p times to the right of MN and q times, to the left of it, then

$$P(b \leq r \leq c) = \int_b^c u^p (1-u)^q du \div \int_0^1 v^p (1-v)^q dv \quad (2)$$

where bc is a segment within AD . Bayes derived the denominator of (2) 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 (2) is now known to be equal to the difference of two values of the incomplete beta-function

$$I_c(p + 1, q + 1) - I_b(p + 1, q + 1).$$

Thus, given the results of the experiment, and assuming a uniform prior distribution of the location of MN and r , which represented ignorance, he

determined the appropriate theoretical probability. Nevertheless, it would be wrong to apply formula (2) for determining, say, the probability that some far digit in the development of π equals 4 (Neyman 1938/1967, p. 337). A constant is not a random variable.

Bayes himself had not stated that his distribution was uniform, but this assumption is necessary (K. Pearson 1978, p. 364). Without providing any explanation, Mises (1919, § 9.2) remarked that Bayes had considered the general case as well. Following Czuber, Mises proved that the influence of non-uniformity weakened with the increase in the number of observations.

In his covering letter to the Bayes memoir, Price provided purely methodical illustrations; one of them required the probability of the next sunrise observed 10^6 times in succession. Formula (2) indirectly answers his question if $b = 1/2$ and $c = 1$ are chosen; it also provides the probability of the contrary event if $b = 0$ and $c = 1/2$. Price (Bayes 1764/1970, pp. 149 and 150 – 151) also solved the same question for $p = 1$ and $q = 0$ and obtained $P = 3/4$ which is doubtful: knowing nothing about the essence of a phenomenon we should have got $P = 1/2$ (cf. Poisson's reasoning in § 8.1.4). In this case, formula (2) is wrong. The actual probability of the next sunrise is

$$P = \frac{\int_0^1 x^{p+1} dx}{\int_0^1 x^p dx} = \frac{p+1}{p+2}$$

and Polyá (1954, p. 135) remarked that each consecutive success (sunrise) provided ever less justification for the next one.

Cournot (1843, § 93) considered a similar problem: A woman gave birth to a boy; determine the probability that her next child will also be a boy. Without justification, he stated that *perhaps* the odds were 2:1 but that it was impossible to solve that problem. See the opinions of Laplace (§§ 7.1-1) and Chebyshev (§ 12.2-5) about the Bayesian approach. Another point concerned the Bayesian treatment of an unknown constant r in formula (2) as a random variable, see above.

Beginning with the 1930s and perhaps for three decades English and American statisticians had been denying Bayes. 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, cf., however, Laplace's general statement about rectifying hypotheses (§ 7.2-1). The *inverse probability* defined by formula (1) is tantamount to conditional probability given that the stipulated condition has indeed been fulfilled.

5.2. The Limit Theorem

Bayes had not expressly discussed the case of $n = (p + q) \rightarrow \infty$. Price, however, remarked that, for a finite n , De Moivre's results were not precise. In another posthumous note published in 1764, Bayes warned mathematicians about the danger of applying divergent series. He had not named De Moivre, but apparently had in mind the derivation of the De Moivre – Laplace theorem (4.4) 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, p. 175) stated that

that trick was possible. De Moivre considered the series included in the Stirling formula.

Timerding, the Editor of the German translation of the Bayes memoir, nevertheless went on to consider the limiting case. He issued from Bayes' calculations made for large but finite values of p and q . Applying a clever trick, he proved that, as $n \rightarrow \infty$, the probability α of the ball falling to the right of MN obeyed the proposition

$$\lim P\left\{ \frac{|\alpha - a|}{\sqrt{pq/n^{3/2}}} \leq z \right\} = \frac{1}{\sqrt{2\pi}} \int_0^z \exp(-w^2/2) dw, \quad (3)$$

where (not indicated by Timerding) $a = p/n = E\alpha$, $pq/n^{3/2} = \text{var}\alpha$.

The functions in the left sides of formulas (4.4) and (3) are random variables, centred and normed in the same way; Bayes, without knowing the notion of variance, apparently understood that (4.4) was not sufficiently precise for describing the problem inverse to that studied by De Moivre. Anyway, Price (Bayes 1764/1970, p. 135) stated that he knew

of no person who has shewn how to deduce the solution of the converse problem [...]. What Mr De Moivre has done therefore cannot be thought sufficient ...

Jakob Bernoulli maintained that his formulas were also fit for solving the inverse problem – but how precisely? De Moivre (1718/1756, p. 251) also stated that he had proved the inverse problem as well:

Conversely, if from numberless observations we find the Ratio of the Events to converge to a determinate quantity [...], then we conclude that this ratio expresses the determinate Law according to which the Event is to happen.

This insufficiently known problem due to Bayes is very important. Together with the integral De Moivre – Laplace theorem it completed the creation of the first version of the theory of probability and could have stimulated Mises (who did not notice that possibility).

5.3. Additional Remark

In 1983, Stigler quoted a curious statement (Hartley 1749, pp. 338 – 339) and interpreted it as a testimony against Bayes' priority. After referring to De Moivre, Hartley wrote, in part:

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

Later Stigler (1986, pp. 98, 132) recalled Hartley and his own earlier paper of 1983, but did not definitively repeat his previous inference. Then, however, he (1999, pp. 291 – 301) reprinted that paper and added a tiny footnote brushing aside all the criticism published by that time.

Stigler inferred that the author of the Bayes' theorem was Saunderson (1682 – 1739), and by applying formula (1), he even found that his

conclusion was three times more probable than the former opinion. However, he assumed that the prior probabilities of the authorship of Bayes and Saunderson were the same. This means that the extra-mathematical arguments (for example, the evidence of Price, a close friend of Bayes) were not considered at all. In addition, not only a honest personality as Saunderson, but almost any pretender will be able to claim equal prior rights with an established author (or a politician) of the past. For my part, I think that it was Bayes himself who communicated to Hartley the solution *of the inverse problem*.

6. Other Investigations before Laplace

I consider the work of several scientists (Daniel Bernoulli in the first place), geometric statistics and applications of statistics. Together with De Moivre, Bernoulli was the main predecessor of Laplace.

Key words: the Buffon needle, the Ehrenfests' model, moral expectation, inoculation of smallpox

6.1. Stochastic Investigations

6.1.1. Daniel Bernoulli. He published a number of memoirs pertaining to probability and statistics, and, before that, he (1735) provided a simple stochastic reasoning on the structure of the Solar system. I consider some of Bernoulli's memoirs and postpone the study of his other work until §§ 6.2.3 and 6.3.1.

In a letter of 1742 he left a curious but unclear statement (Fuss 1843/1968, t. 2, p. 496):

I believe that mathematics can also be rightfully applied in politics. [...] An entirely new science will emerge provided that as many observations are made in politics as in physics.

Mathematics here likely meant probability theory to which I am turning now.

1) Moral expectation. While attempting to explain the paradoxical nature of the Petersburg game (§ 3.3.4), Bernoulli (1738) suggested that the gain y of a gambler was determined by his winnings x in accord with the differential equation (the first such equation in probability)

$$dy = cdx/x, c > 0, \text{ so that } y = f(x) = c \ln(x/a)$$

where a was the gambler's initial capital. The logarithmic function also appears in the celebrated Weber – Fechner psychophysical law and is applied in the theory of information.

Bernoulli also proposed that the expected winnings $\sum p_i x_i / \sum p_i$ where p_i were the appropriate probabilities be replaced by their *moral expectation* $\sum p_i f(x_i) / \sum p_i$. He indicated but had not proved (see § 7.1-9) that even a *just* game with a zero expected loss for each participant became disadvantageous because the moral expectation of winnings, again for each, was negative, and that the paradoxical infinite expected gain in the Petersburg game (3.4) can be replaced by a finite moral expectation. Applying his innovation to a study of marine shipping of freight, he maintained (again, without proof, see same subsection) that the freight should be evenly distributed among several vessels.

Bernoulli appended the text of a letter of 1732 from Gabriel Cramer to Nikolaus Bernoulli which contained his (not Daniel's) term *moral expectation*. Cramer also indirectly suggested to select

$$f(x) = \min(x; 2^{24}) \text{ or } f(x) = \sqrt{x}.$$

Moral expectation had become popular and Laplace (1812/1886, p. 189) therefore proposed a new term for the previous *usual* expectation calling it *mathematical*; his expression regrettably persists at least in the French and Russian literature. At the end of the 19th century, issuing from Bernoulli's idea, economists began to develop the theory of marginal utility thus refuting Bertrand's opinion (1888a, p. 66) that moral expectation was useless:

The theory of moral expectation became classical, and never was a word applied more exactly. It was studied and taught; it was developed in books really celebrated. With that, the success came to a stop; no application was made, or could be made, of it.

2) A limit theorem. While studying the same problem concerning the sex ratio at birth (§§ 2.2.4, 3.3.4, 4.4), 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

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

He calculated this fraction not by the Wallis formula but by means of differential equations. After deriving $q(N - 1)$ and $q(N + 1)$ and the two appropriate values of Δq , he arrived at

$$dq/dN = -q/(2N + 2), dq/dN = -q/(2N - 1)$$

and, *in the mean*, $dq/dN = -q/(2N + 1/2)$. Assuming that the solution of this equation passed through point $N = 12$ and $q(12)$ as defined above, he obtained

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

Application of differential equations was Bernoulli's usual method in probability, also see item 1.

Bernoulli also determined the probability of the birth of approximately m boys (see below):

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

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

$$Em = M = \frac{2Na - b}{a + b} \approx \frac{2Na}{a + b}$$

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}):

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

$$-\frac{d\pi}{\pi} = \frac{\mu + 1 + \mu a/b}{m + \mu + 1} d\mu.$$

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

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

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

Issuing from some statistical data, he compared two possible pertinent ratios a/b but had not made a final choice in favour of either of them. He also determined such a value of μ that the sum of probabilities (1), beginning from $\mu = 0$, equalled one half. Applying summation rather than integration, he had not therefore arrived at an integral limit theorem and (also see above) he did not refer to, and apparently had not known about De Moivre's findings. This shows, once again (cf. § 4.4), that they had for a long time been forgotten.

3) Urn problems. I consider two of these. An urn contains n pairs of white and black stripes. Determine the number (here and below, actually, the expected number) of paired stripes left after $(2n - r)$ extractions without replacement. By the combinatorial method Bernoulli (1768a) obtained

$$x = r(r - 1)/(4n - 2); \text{ and } x = r^2/4n \text{ if } n = \infty.$$

He derived the same result otherwise: when r decreases by dr the corresponding dx is either zero $[(r - 2x)$ cases] or dr ($2x$ cases) so that

$$dx = [(r - 2x) \cdot 0 + 2x \cdot dr]/r, \quad x = r^2/4n \text{ since } r = 2n \text{ if } x = n.$$

Bernoulli then considered unequal probabilities of extracting the stripes of different colours and (1768b) applied his findings to study the duration of marriages, a subject which was directly linked with insurance of joint lives.

Suppose now that each of two urns contains an equal number n of balls, white and black, respectively. Determine the number of white balls in the first urn after r cyclic interchanges of one ball. Bernoulli (1770) solved this problem by the same two methods. Thus, the differential approach led him to

$$dx = -xdr/n + [(n - x)/n]dr \text{ so that } x \approx (1/2)n [1 + e^{-2r/n}].$$

Bernoulli then combinatorially considered the case of three urns with balls of three different colours. He noted that the number of white balls in

the first urn was equal to the sum of the first, the fourth, the seventh, ... terms of the development of $[(n - 1) + 1]^r$ divided by n^{r-1} . For the other urns he calculated, respectively, the sums of the second, the fifth, the eighth, ..., and the third, the sixth, the ninth, ... terms. For the first urn he obtained

$$A = \frac{1}{n^{r-1}} [(n - 1)^r + C_r^3 (n - 1)^{r-3} + C_r^6 (n - 1)^{r-6} + \dots] \approx ne^{-r/n} S. \quad (2)$$

The expression designated by S obeyed the differential equation

$$Sdr^3/n^3 = d^3S$$

and was therefore equal to

$$S = ae^{r/n} + be^{-r/2n} \sin(r\sqrt{3}/2n) + ce^{-r/2n} \cos(r\sqrt{3}/2n)$$

where, on the strength of the initial conditions, $a = 1/3$, $b = 0$, $c = 2/3$.

Bernoulli derived similar expressions for the other urns, calculated the number of extractions leading to the maximal number of white balls in the first urn, and noted the existence of a limiting state, of an equal number of balls of each colour in each urn. This can be easily verified by referring to the theorem on the limiting transition matrix in homogeneous Markov chains and his problem anticipated the celebrated Ehrenfests' model (1907), the beginning of the history of stochastic processes.

Bernoulli obtained formula (2) by issuing also from differential equations

$$dx = -xdr/n + [n - (x + y)]dr/n, \quad dy = -ydr/n + xdr/n$$

where x , y , and $[n - (x + y)]$ were the numbers of white balls in the urns after r interchanges. I return to this problem in §7.1-3; here, I note that Todhunter (1865, pp. 231 – 234) simplified Bernoulli's solution and made it more elegant. He wrote the differential equations as

$$dx = (dr/n)(z - x), \quad dy = (dr/n)(x - y), \quad dz = (dr/n)(y - z)$$

and noted that the sum S was equal to

$$S = (1/3)[e^{\alpha r/n} + e^{\beta r/n} + e^{\gamma r/n}]$$

with α , β , γ being the values of $\sqrt[3]{1}$.

Bernoulli's x in his first problem, and his S and A from (1) depend on discrete time r/n , which is characteristic of stochastic processes with non-homogeneous time.

Lagrange (1777) solved such and other stochastic problems by means of partial difference equations.

6.1.2. D'Alembert. In the theory of probability, he is mostly known as the author of patently wrong statements. Thus, he (1754) maintained that the probability of heads appearing twice in succession was equal to $1/3$ rather than to $1/4$. Then, he (1768a) reasoned on the difference between *mathematical* and *physical* probabilities, stating without justification that,

for example, after one of two contrary events had occurred several times in succession, the appearance of the other one becomes physically more probable. He was thus ridden by prejudices which Montmort had already mentioned and which Bertrand later refuted by a few words (§ 2.1.1). At the same time, D'Alembert recommended to determine probabilities experimentally but had not followed his own advice (which saved him from revealing his mistakes). Finally, he (1768b) denied the difference (perfectly well understood by Huygens, § 2.2.2) between the mean, and the probable durations of life and even considered its existence as an (additional) argument against the theory of probability itself.

It is opportune to recall Euler's opinion as formulated in one of his private letters of 1763 (Juskevic et al 1959, p. 221): D'Alembert tries *most shamelessly to defend all his mistakes*. Anyway, D'Alembert (1768d, pp. 309 – 310) did not ascribe the theory of probability to a *precise and true calculus with respect either to its principles or results*.

On the other hand, D'Alembert thought that, in a single trial, rare events should be considered unrealizable (Todhunter 1865, § 473) and that absolute certainty was qualitatively different from *the highest probability*. It followed from the latter statement that, given a large number of observations, an unlikely event might happen (cf. the strong law of large numbers), and, taken together, his considerations meant that the theory of probability ought to be applied cautiously. D'Alembert also reasonably objected to Daniel Bernoulli's work on prevention of smallpox and formulated his own pertinent ideas (§ 6.2.3). I ought to add that D'Alembert was indeed praiseworthy for his work in other branches of mathematics (and in mechanics); note also that Euler had not elaborated his likely correct remark.

On D'Alembert's work see also Yamazaki (1971). He published many contributions on probability and its applications and it is difficult to organize them bibliographically; on this point see Paty (1988).

6.1.3. Lambert. He was the first follower of Leibniz in attempting to create a doctrine of probability as a component of a general teaching of logic. Like D'Alembert, Lambert explained randomness by ignorance of causes, but he also stated that all digits in infinite decimal developments of irrational numbers were equally probable, which was an heuristic approach to the notion of normal numbers, and he formulated a modern-sounding idea about the connection of randomness and disorder (Lambert 1771, § 324; 1772 – 1775). His thoughts were forgotten until Cournot (1851/1975, § 33, Note) noted them, and only Chuprov (1909/1959, p. 188) mentioned them afterwards.

Lambert did not go out of the confines of *uniform randomness*. The philosophical treatises of the 18th century testify to the great difficulties experienced in generalizing the notion of randomness, also see § 2.2.4. Even in the 19th century, many scientists, imagining that randomness was only uniform, refused to recognize the evolution of species.

6.1.4. Buffon. He (1777) is mostly remembered for his definitive introduction of geometric probabilities (§ 6.1.6). He experimentally studied the Petersburg game (§ 3.3.4), proposed the value 1/10,000 as a (non-existing) universally negligible probability, wrongly solved the problem of the probability of the next sunrise (§ 5.1) and compiled tables of mortality which became popular.

Negligible, as he thought, was the probability of death of a healthy man aged 56 during the next 24 hours, but his figure was apparently too low; K. Pearson (1978, p. 193) thought that 1/1,000 would have been more appropriate. In addition, negligibility ought to be only chosen for a particular event rather than assigned universally. All the above is contained in Buffon's main work (1777). There also (§ 8, Note) he published the text of his letter of 1762 to Daniel Bernoulli which contained an embryo of Quelelet's celebrated Average man (see my § 10.5):

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.

6.1.5. Condorcet. He attempted to apply the theory of probability to jurisprudence in the ideal and tacitly assumed case of independent judgements made by jurors or judges. He also estimated the trustworthiness of testimonies and critically considered electoral problems.

His main method was the application of difference equations. Todhunter (1865, pp. 351 – 410) described the work of Condorcet in detail and concluded (p. 352) that in many cases it was *almost impossible to discover* what he had meant: *The obscurity and self contradiction are without any parallel [...]* He, Todhunter, will provide some illustrations, *but no amount of examples can convey an adequate impression of the extent of the evils.* At the very least, however, Laplace and Poisson continued to apply probability to jurisprudence and certainly profited to some extent from the work of Condorcet. Poisson (1837a, p. 2) mentioned his ideas quite favourably.

I note however that, while discussing games of chance, Condorcet (1785/1847, p. 561) expressed himself rather unfortunately, and stated on the next page without any justification that Daniel Bernoulli had not removed all the objections to the *rule* of expectation which was allegedly achieved by D'Alembert. In 1772, in a letter to Turgot, he (Henry 1883/1970, pp. 97 – 98) told his correspondent that he was *amusing himself* by calculating probabilities, had compiled *a booklet* [which remains unknown] *on that subject* and was keeping to the opinions of D'Alembert. On Condorcet see also Yamazaki (1971).

6.1.6. Geometric Probabilities. These were decisively introduced in the 18th century although the definition of the notion itself, and, for that matter, only on a heuristic level, occurred in the mid-19th century (§ 10.3). Newton (§ 2.2.3) was the first to think about geometric probability; Daniel Bernoulli (§ 6.1.1) tacitly applied it in 1735 as did somewhat later De Moivre (1725/1756, p. 323), T. Simpson (1757) (§ 6.3.1) and Bayes (§ 5.1). Dealing with the continuous uniform distribution, De Moivre assumed, for example, that if $0 < \xi < b$ and $0 < a < b$, then $P(0 < \xi < a) = [0; a] \div [0; b]$.

The Michell problem (1767) became classical: Determine the probability that two stars from all of them, uniformly distributed over the celestial sphere, were situated not farther than 1° from each other. Choose an arbitrary point (A) on a sphere with centre O and imagine a circle perpendicular to OA having distance 1° from A . The probability sought is the ratio of the surface of the spherical segment thus obtained to that of the sphere.

Newcomb and Fisher calculated the expected number of closely situated stars (§ 10.8-4) and general issues were also debated by others. Thus, Proctor (1874, p. 99) wished to determine *what peculiarities of distribution might be expected to appear among a number of points spread over a plane surface at random*. His was a question now belonging to mathematical statistics and concerning the deviations of an empirical density curve from its theoretical counterpart. Bertrand (1888a, pp. 170 – 171) remarked that without studying other features of the sidereal system it was impossible to decide whether stars were arranged randomly.

Buffon (§ 6.1.4) expressly studied geometric probability; 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

$$P = 4r/\pi a$$

(3)

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 (TAP, p. 366) who noted that formula (3) enabled to determine [with a low precision] the number π .

6.2. Statistical Investigations

6.2.1. Staatswissenschaft (Statecraft, University Statistics). 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. 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 (1752/1756, Intro.) also 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 [Reiche].

Achenwall's student Schlözer (1804, p. 86) figuratively stated that *History is statistics flowing, and statistics is history standing still*. For those keeping to *Staatswissenschaft* this pithy saying became the definition of statistics which was thus not compelled to study causal connections in society or discuss possible consequences of innovations. Then, only political arithmetic was mostly interested in studying population; finally, wordy descriptions rather than numbers lay at the heart of the works of the Göttingen school.

Knies (1850, p. 24) 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 has to do with the limitations of mathematics in general.

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. And *In a multitude of people is the glory of a king, but without people a prince is ruined* (Proverbs 14:28).

Tabular statistics which had originated with Anchersen (1741) could have served as an intermediate link between words and numbers, 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).

By the end of the 19th century the scope of *Staatswissenschaft* narrowed, although it still exists, at least in Germany, in a new form: it includes numerical data and studies causes and effects and it is the application of the statistical method to various disciplines and a given state, but statistics, in its modern sense, owed its origin to political arithmetic.

6.2.2. Population 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 (§ 10.5); 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 1765 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 the multiplication of mankind was a divine commandment and that 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 proportion as the appropriate law.

Euler left no serious contribution to the theory of probability, but he published a few memoirs on population statistics. He did not introduce any stochastic laws, but such concepts as increase in population and the period of its doubling are 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 (§ 14.2). 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 issued 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.

6.2.3. Medical Statistics. It originated in the 19th century, partly because of the need to combat the devastating visitations of cholera. At the end of the 18th century Condorcet (1795/1988, p. 542) advocated collection of medical observations and Black (1788, pp. 65 – 68) even compiled a *Medical catalogue of all the principal diseases and casualties by which the Human Species are destroyed or annoyed* that reminded of Leibniz' thoughts (§ 2.1.4). He also appended to his book a *Chart of all the fatal diseases and casualties in London during [...] 1701 – 1776*. By means of such charts, he (p. 56) stated, *we shall [...] be warned to make the best disposition and preparation for defence*. In an earlier publication Black (1782), however, expressed contradictory views.

D'Alembert (1759/1821, pp. 163 and 167) arrogantly declared that

Systematic medicine is a real scourge of mankind. Multiple and detailed observations, conforming to each other, this [...] is what the reasoning in medicine ought to be reduced.

A physician is a blind man armed with a club. He lifts it without knowing who will he hit. If he hits the disease, he kills it; if he hits Nature, he kills Nature.

The physician most deserving to be consulted, is that who least believes in medicine.

All this is contained in the second edition of his book, but was written not later than in 1783, the year of his death.

Especially important was the study of prevention of smallpox (Condamine 1759, 1763, 1773; Karn 1931). Condamine (1759) listed the objections against inoculation, both medical and religious. Indeed, an approval from theologians was really needed. White (1896/1898) described the *warfare of science with theology* and included (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 distinguished between theology, the opposing force, and *practical* religion. Condamine (1773) included his correspondence, in particular with Daniel Bernoulli, to whom he had given the data on smallpox epidemics which the latter used in his research.

Karn began her article by stating 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 Daniel 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 introduced necessarily crude parameters of smallpox epidemics and assumed that the inoculation itself proved fatal in 0.5% of cases. He formed and solved the appropriate differential equation and thus showed the relation between the age, the number of people of the same age, and of those of them who had not contacted smallpox. Also by means of a differential equation he derived a similar formula for a population undergoing inoculation. It occurred that inoculation lengthened the mean duration of life by 3 years and 2 months and was therefore extremely useful. Vaccination, *the inestimable discovery by Jenner, who had 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 final success had not however ruled out statistical studies. Simon (1887, vol. 1, p. 230) concluded that only comprehensive national statistics could duly compare it with inoculation.

D'Alembert (1761; 1768c) criticized Daniel Bernoulli, see Todhunter (1865, pp. 265 – 271, 277 – 278 and 282 – 286). 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. Without denying the benefits of that procedure, 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 applicable to studies of even unpreventable diseases. Dietz et al (2000; 2002) described Bernoulli's and D'Alembert's investigations on the level of modern mathematical epidemiology and mentioned sources on the history of inoculation. For his part, K. Pearson (1978, p. 543) stated that inoculation was *said to have been a custom in Greece in the 17th century and was advocated [...] in the Philosophical Transactions of the Royal Society in 1713*. Also see Sheynin (1972/1977, pp. 114 – 116; 1982, pp. 270 – 272).

6.2.4. Meteorology. Leibniz (§ 2.1.4) recommended regular meteorological observations. Indeed (Wolf 1935/1950, p. 312),

Observations of barometric pressure and weather conditions were made at Hanover, in 1678, and at Kiel, from 1679 to 1714, at the instigation of Leibniz.

The *Societas meteorologica Palatina* in Pfalz (a principality in Germany) was established in 1780, and, for the first time in the history of experimental science, it organized cooperation on an international scale. At about the same time the *Société Royale de Médecine* (Paris) conducted observations in several European countries (Kington 1974) and even in the 1730s – 1740s they were carried out in several towns in Siberia in accordance with directions drawn up by Daniel Bernoulli in 1733 (Tikhomirov 1932). In the second half of the 18th century several scholars (the meteorologist Cotte, Lambert and Condorcet) proposed plans for comprehensive international meteorological studies.

Lambert (1773) studied the influence of the Moon on the air pressure and Daniel Bernoulli encouraged him (Radelet de Grave et al 1979, p. 62): if the influence of the Moon on the *air* is similar to its influence on the seas, it should be observable, because the Moon's distance varies, but the elasticity of air and its weak inertia should be allowed for. And, further:

Your considerations [...] are quite justified; publish them without hesitating [...] whatever are the results [...]. Only try to establish them properly.

Toaldo (1775; 1777) statistically studied the connections between phenomena concerning meteorology and stated that the weather depended on the configurations of the Moon. His opinion was not abandoned until the mid-19th century (Muncke 1837, pp. 2052 – 2076).

6.3. Treatment of Observations

It became necessary after regular astronomical observations had begun (since Tycho Brahe). A problem of determining the Earth's figure presented itself in the second half of the 18th century. Newton proved that the Earth was an ellipsoid of revolution with its equatorial radius (*a*) larger than its polar radius (*b*), and attempts were being made to prove or disprove this conclusion by meridian arc measurements. Their lengths were indirectly calculated by triangulation. Two such measurements are needed for calculating the parameters of the ellipsoid (although local deviations of the figure of the Earth corrupt the results) whereas redundant measurements

lead to systems of linear equations in these unknowns which can then be derived more precisely. Nowadays, according to the Krasovsky ellipsoid of 1940 (Sakatov 1950, p. 364), it is held that $a = 6356.8 \text{ km}$ and $b = 6356.8 \text{ km}$, so that $2\pi a$ is approximately equal to $40,000 \text{ km}$ corresponding to the initial definition of the metre. However, in 1960 the metre was defined in terms of the wavelength of light. Not a and b were actually derived, but rather a and the *flattening* $(a - b)/a$. That parameter had also been determined by pendulum observations, cf. § 10.9.1.

The introduction of the metric system, and the demands of cartography, physics and chemistry led to the advancement of the treatment of observations. Scientists recognized the common character of adjusting direct and indirect observations: in both cases the unknowns were called *Mittel* (Lambert 1765b, § 6) or *milieu* (Maire & Boscovich 1770, pp. 484 and 501).

6.3.1. Direct Measurements. The first to touch on this case was Cotes (1722), see Gowing (1983, p. 107). Without any justification he advised to regard the weighted arithmetic mean, which he compared with the centre of gravity of the system of points, – of the observations, – as the *most probable* estimator of the constant sought:

Let p be the place of some object defined by observation, q, r, s the places of the same object from subsequent observations. Let there also be weights P, Q, R, S reciprocally proportional to the displacements arising from the errors in the single observations, and which are given by the limits of the given errors; and the weights P, Q, R, S are conceived as being placed at p, q, r, s , and their centre of gravity Z is found; I say the point Z is the most probable place of the object.

Cotes appended a figure (perhaps representing a three-dimensional picture) showing nothing except these four points. He had not explained what he meant by *most probable*, nor did he describe his statement clearly enough. Nevertheless, his authority gave support to the existing common feeling (§ 2.1.1). Without mentioning Cotes Condamine (1751, p. 223) recommended to apply that estimator. Then, Laplace (1814/1995, p. 121) stated that *all calculators* followed the Cotes rule. Even before Cotes Picard (1693/1729, pp. 330, 335, 343) called the arithmetic mean *véritable*.

T. Simpson (1756), see also Shoesmith (1985b), 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, cf. end of § 1.7. Simpson assumed that the chances of observational errors

$$-v, -v + 1, \dots, -2, -1, 0, 1, 2, \dots, v - 1, v$$

were equal [proportional] either to

$$r^{-v}, r^{-v+1}, \dots, r^{-2}, r^{-1}, 1, r, r^2, \dots, r^{v-1}, r^v$$

or to

$$r^{-v}, 2r^{-v+1}, \dots, (v-1)r^{-2}, vr^{-1}, (v+1), vr, (v-1)r^2, \dots, 2r^{v-1}, r^v.$$

Taking $r = 1$ he thus introduced the uniform and the triangular discrete distributions. Denote the observational errors by ε_i , and by N , the number of some chances. Then, as Simpson noted,

$N(\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_n = m)$ was the coefficient of r^m in the expansions of

$$\begin{aligned} (r^{-v} + \dots + r^0 + \dots + r^v)^n &= r^{-vn}(1-r)^{-n}(1-r^{2v+1})^n, \\ (r^{-v} + 2r^{-v+1} + \dots + (v+1)r^0 + \dots + 2r^{v-1} + r^v)^n &= r^{-vn}(1-r)^{-2n}(1-r^{v+1})^{2n}. \end{aligned}$$

The left sides of these two equalities were generating functions with unit coefficients in the first case, and coefficients 1, 2, ..., $v+1$, ... 2, 1 in the second instance.

For both these cases Simpson determined 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. Consequently, 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. He himself (1740, Problem No. 22), and earlier Montmort (§ 3.3.3), although without introducing generating functions, and De Moivre (1730, pp. 191 – 197) had studied the game of dice. 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 to which he passed over by assuming that $|v| \rightarrow \infty$ with $(m/n)/v$ remaining constant. Here, m/n was the admissible error of the mean. 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.

Without mentioning Simpson, Lagrange (1776a) studied the error of the mean for several other and purely academic distributions, also by applying generating functions (even for continuous laws, thus anticipating the introduction of characteristic functions). He was the first to use integral transformations, and, in Problem 6, he derived the equation of the multivariate normal distribution. In his § 18 he introduced the term *courbe de la facilité des erreurs*. A possible though inadequate reason for ignoring Simpson was the heated dispute over priority between De Moivre and him: Lagrange apparently had not wanted to be even indirectly involved in it. De Moivre was a scholar of a much higher calibre (a fact clearly recognized by Simpson) and 43 years the senior. At least on several important occasions Simpson did not refer to him and, after being accused by De Moivre (1725; only in edition of 1743, p. xii) of *mak[ing] a Shew of new Rules, and works of mine*, Simpson (posthumous publication, 1775, p. 144)

appeal[ed] to all mankind, whether in his treatment of me he has [not] discovered an air of self-sufficiency, ill-nature, and inveteracy, unbecoming a gentleman.

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) quantities. 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 $\varphi(x - x_0)$, where x_0 was the sought parameter of location. Denote the observations by x_1, x_2, \dots, x_n , and, somewhat simplifying his reasoning, write his likelihood function as

$$\varphi(x_1 - x_0) \varphi(x_2 - x_0) \dots \varphi(x_n - x_0).$$

When differentiating it, Lambert had not indicated that the argument here was the parameter x_0 , etc.

In a few years 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*.

In a letter of 1971 E. S. Pearson informed me that *curiously* his father's *Lectures* (1978), – then not yet published, – omitted Lambert. He explained:

It was not because [Lambert's] writings were in German of which my father was an excellent scholar. I suppose [...] that he selected the names of the personalities he would study from a limited number of sources, e.g.,

Todhunter, and that these did not include Lambert's name. [Todhunter did refer to Lambert but had not described his work.] Of course, K. P. was over 70 by the time his history lectures passed the year 1750, and no doubt his exploration was limiting itself to the four Frenchmen, Condorcet, D'Alembert, La Grange and Laplace.

Johann III Bernoulli (1785) published a passage from a manuscript of Daniel Bernoulli (1769/1997) which he had received in 1769 but written, as its author had told him, much earlier. There, Daniel assumed the density law of observational errors as a semi-ellipse or semi-circumference of some radius r ascertained by assigning a reasonable maximal error of observation and the location parameter equal to the weighted arithmetic mean with posterior weights

$$p_i = r^2 - (\bar{x} - x_i)^2. \quad (4)$$

Here, x_i were the observations and \bar{x} , the usual mean. The first to apply weighted, or generalized arithmetic means was Short (1763). This estimator demanded a subjective selection of weights and it 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

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

where x was the unknown abscissa of the centre of the semi-circumference, and x_1, x_2, x_3, \dots , 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

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

so that the maximum likelihood estimate is

$$x_0 = \frac{[px]}{\sum p_i}, p_i = \frac{1}{r^2 - (x_0 - x_i)^2} \quad (5; 6)$$

with unavoidable use of successive approximations. For some inexplicable reason these formulas are lacking in Bernoulli's memoir although the posterior weights (6) were the inverse of the weights (4) 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.

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. Euler then (§ 7) remarked that there was no need

to have recourse to the principle of the maximum, since the undoubted precepts of the theory of probability are quite sufficient to resolve all questions of this kind.

Instead of the arithmetic mean Euler recommended the estimate (5) with posterior weights (4) and mistakenly assumed that Bernoulli had chosen these same weights.

Euler denoted the observations by $\Pi + a$, $\Pi + b$, $\Pi + c$, ... and (§11) remarked that his estimate can be obtained from the condition

$$[r^2 - (x_0 - a)^2]^2 + [r^2 - (x_0 - b)^2]^2 + [r^2 - (x_0 - c)^2]^2 + \dots = \max. \quad (7)$$

The quantities in parentheses are the deviations of observations from the estimate sought and their fourth powers are negligible so that condition (7) is equivalent to the requirement

$$(x_0 - a)^2 + (x_0 - b)^2 + (x_0 - c)^2 + \dots = \min, \quad (8)$$

whence follows the arithmetic mean. Condition (7) is heuristically similar to the principle of least squares (which in case of one unknown leads to the arithmetic mean) and resembles the Gaussian principle of maximum weight, § 9.1.3. A small corruption of condition (8) does exist, it is caused by inevitable deviations of the observations from the proposed (or assumed) symmetrical law. Bernoulli noted this fact, and actually proposed the general arithmetic mean.

In his last memoir devoted to pendulum observations Daniel Bernoulli (1780) separated, for the first time ever, observational errors into random (*momentaneorum*) and systematic (*chronicarum*), although not for observations in general. He indicated that these errors acted proportionally to the square root of, and to the time itself respectively. Making use of his previous findings (§ 6.1.1), Bernoulli justified his investigation by the normal distribution which thus first occurred in the theory of errors, although only as a limiting law.

From $2N$ daily vibrations of a pendulum, as Bernoulli assumed, $(N + \mu)$ were slower, and $(N - \mu)$ faster than stipulated, with periods of $(1 + \alpha)$ and $(1 - \alpha)$ respectively. His pattern meant that the number of positive (say)

errors possessed a symmetric binomial distribution and that the error of the pendulum accumulated after a large number of vibrations will have a normal distribution.

Bernoulli had not investigated the more general pattern of an unequal number of the slower and the faster vibrations although it corresponded to the case of unequal probabilities of male and female births, also studied by him (§ 6.1.1). Neither had he said anything about the possible dependence between the periods of successive vibrations.

In his previous work Bernoulli (1770 – 1771) noted that, for $N = 10,000$ and $\mu = 47.25$

$$\frac{2}{\sqrt{\pi N}} \int_0^{\mu} \exp\left(-\frac{x^2}{N}\right) dx = \frac{1}{2}.$$

Now, having $N = 43,200$, he obtained, for the same probability of 1/2,

$$\mu = 47.25 \sqrt{4.32} \approx 100.$$

It was this calculation that caused his conclusion (above) about the behaviour of random errors. Already in the 19th century, however, it became known that such errors can possess other laws of distribution (end of § 10.5).

Note also that Bernoulli came close to introducing the probable error; to recall (§ 2.2.2), Huygens discussed the probable duration of life. Bernoulli was also the first to introduce elementary errors. I do not, however, set high store by this fact; indeed, this notion is not necessary for proving the CLT.

6.3.2. Indirect measurements. Here, I consider the adjustment of redundant systems

$$a_i x + b_i y + \dots + s_i = v_i, \quad i = 1, 2, \dots, n \quad (9)$$

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

1) In case of two unknowns (cf. beginning of § 6.3) astronomers usually separated systems (9) 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 (9) was some weighted mean of these partial solutions (Whittaker & Robinson 1924/1949, p. 251).

In 1757 and later Boscovich, see Cubranic (1961, pp. 90 – 91) and Maire & Boscovich (1770, pp. 483 – 484), applied this method but it did not satisfy him, see below. In the first case he (Cubranic 1961, p. 46) derived the arithmetic mean of four latitudinal differences in an unusual way: he first calculated the half-sums of all six pairs of differences and then took their mean. He apparently attempted to reveal the unavoidable systematic errors and to ensure a (qualitative) estimation of the order of random errors.

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

$$\sum v_i = 0,$$

where i indicated the number of an equation; if the first group included the first nine of them, then $i = 1, 2, \dots, 9$.

In a letter of 1850 Gauss (W/Erg-5, Tl. 6, p. 90) remarked that Mayer had only calculated by means of primitive combinations. He referred to Mayer's manuscripts, but it is likely that Mayer's trick was almost the same in both cases. And Gauss himself, in an earlier letter of the same year (Ibidem, pp. 66 – 67), recommended a similar procedure for calibrating an aneroid. Anyway, 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) The Boscovich method. He (Maire & Boscovich 1770, p. 501) adjusted systems (9) under additional conditions

$$v_1 + v_2 + \dots + v_n = 0, |v_1| + |v_2| + \dots + |v_n| = \min, \quad (10; 11)$$

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 Boscovich' method with the median. Indeed, he adjusted systems (9) by constructing a straight line whose slope was equal to the median of some fractions. In 1809, Gauss noted that (11) 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 (11) in the case in which the magnitudes such as v_i were positive by definition. Just the same, W. 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.

Here is W. Herschel's earlier reasoning (1783/1912, p. 120):

We ought [...] to resolve that which is common to all the stars [...] into a single real motion of the Solar system, as far as that will answer the known

facts, and only to attribute to the proper motions of each particular star the deviations from the general law the stars seem to follow ...

This statement resembles Newton's *Rules of Reasoning in Philosophy* (1729, Book 3): admit no more causes *than such that are both true and sufficient*. Even Ptolemy (1984, III, 4, p. 153) maintained that *a simpler hypothesis would seem more reasonable*.

When treating direct measurements W. Herschel (1806) preferred the median rather than the arithmetic mean (Sheynin 1984a, pp. 172 – 173).

4) The minimax method. Kepler (§ 1.7) had apparently made use of some elements of this method although it did not ensure optimal, in any sense, results. 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

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

5) Euler (1749, 1755, 1770) had to treat indirect measurements as well. At least in the first two instances his goal was much more difficult than that outlined in § 1.7 where the underlying theory was supposed to be known. Concerning the first of his contributions, Wilson (1980, p. 262n) remarked that Euler was

Stymied by the finding that, for certain of the variables, the equations led to wildly different values.

Euler did not attempt to build a general theory, he wished to achieve practical results and turned in some cases to the minimax principle. On the last occasion Euler did not keep to any definite method and combined equations in a doubtful manner. So as to eliminate one unknown, he subtracted each equation from (say) the first one, thus assigning it much more weight.

Stigler (1986, pp. 27 – 28) called Euler's memoir (1749) a *statistical failure* and, in his opinion, Euler was a mathematician who *distrusted* the combination of equations. Not understanding the main goal of the method of minimax, he mentioned a classic in a free and easy manner, so that his statement was absolutely inadmissible, see also item 6 below. In his second book Stigler (1999, pp. 317 – 318) unblushingly called Euler a great statistician but did not notice his inadequate reasoning concerning deviating observations (§ 6.3.1).

6) In the 18th century practitioners at least sometimes experienced difficulties when deciding how to adjust their observations (Bru 1988, pp. 225 – 226). Indeed, Maupertuis (1738/1756, p. 160; 1756b, pp. 311 – 319) calculated his triangulation 12 times (taking into account differing sets of observations), selected two of his results and adopted their mean value.

At the turn of that century Laplace and Legendre refused to adjust a triangulation chain laid out between two baselines. Likely fearing the propagation of large errors, they calculated each half of the chain starting from its own baseline. Much later Laplace (ca. 1819/1886, pp. 590 – 591) defended their decision by previous ignorance of the *vraie théorie* of

adjustment and added that his (not Gauss'!) justification of the MLSq had changed the situation.

7. Laplace

Laplace devoted a number of memoirs to the theory of probability and later combined them in his *Théorie analytique des probabilités* (TAP) (1812). He often issued from the non-rigorously proved CLT by applying characteristic functions and the inversion formula, calculated difficult integrals, applied Hermite polynomials, introduced the Dirac function and (after Daniel Bernoulli) the Ehrenfests' model, studied sampling, but left probability on its previous level. His theory of errors was impractical. However, 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. Many commentators reasonably stated that his contributions made difficult reading.

Key words: CLT, criminal statistics, theory of errors, absolute expectation, stochastic processes

7.1. Theory of Probability

I describe the second *Livre* of the TAP; in the first one he studied the calculus of generating functions with application to the solution of ordinary and partial difference equations and the approximate calculation of integrals. In many instances he treated the problems involved in his earlier memoirs. As indirectly seen in his *Essay* (1814/1995, pp. 2, 43 – 44), Laplace thought that the main aim of probability theory was to discover the laws of nature.

1) In **Chapter 1** Laplace provided the *classical* definition of probability (introduced by De Moivre, see my § 4.1), formulated the addition and multiplication theorems for independent events as well as theorems concerning conditional probabilities. Elsewhere, he (1814/1995, p. 10), added to this general subject matter the so-called Bayes theorem (5.1), calling it a principle.

2) In **Chapter 2** Laplace solved a number of problems by means of difference, and partial difference equations. I consider three other problems (§§ 13, 15, 15).

a) In an astronomical context Laplace studied sampling with replacement. Tickets numbered from 0 to n are extracted from an urn. Determine the probability that the sum of k numbers thus extracted will be equal to s . While solving this problem, he applied a discontinuous factor $(1 - l^{m+1})^m$ with $l = 0$ or 1 and $m = 1, 2, \dots$

Laplace considered the case of $s, n \rightarrow \infty$ and his derived formula for the distribution of the sum of independent, continuous variables obeying the uniform law on interval $[0; 1]$ corresponded with modern literature (Wilks 1962, § 8.3.1) which does not, however, demand large values of s and n .

b) Non-negative random variables t_1, t_2, \dots, t_k with differing laws of distribution $\varphi_i(t)$ are mutually independent and their sum is s . Determine the integral

$$\int \psi(t_1; t_2; \dots; t_k) \varphi_1(t) \varphi_2(t) \dots \varphi_k(t) dt_1 dt_2 \dots dt_k$$

over all possible values of the variables; ψ was yet to be chosen. Laplace then generalized his very general problem still more and derived the Dirichlet formula even in a more general setting.

c) 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

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

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

3)The **third Chapter** is devoted to the integral *De Moivre – Laplace* theorem and to several interesting problems connected with the transition to the limit. Here is that theorem:

$$P(|\mu - np - z| \leq l) = \frac{2}{\sqrt{\pi}} \int_0^{l\sqrt{n/2xx'}} \exp(-t^2) dt + \sqrt{\frac{n}{2\pi xx'}} \exp\left(-\frac{l^2 n}{2xx'}\right),$$

p was the probability of success in a single Bernoulli trial, μ , the total number of successes in n trials, $q = 1 - p$, z was unknown but $|z| < 1$, $x = np + z$, and $x' = nq - z$.

In proving it, Laplace applied the Euler – MacLaurin summation formula, and his remainder term allowed for the case of large but finite number of trials. He indicated that his theorem was applicable for estimating the theoretical probability given statistical data, cf. the Bayes theorem in § 5.2, but his explanation was not clear. Molina (1930, p. 386) quoted Laplace's memoir (1786/1894, p. 308) where he (not clearly enough) had contrasted the *appraisals* admitted in probability with certainty provided in analysis.

Already Daniel Bernoulli (§ 6.1.1) solved one of Laplace's problem: There are two urns, each containing n balls, some white and the rest black; on the whole, there are as many white balls as black ones. Determine the probability u that the first urn will have x white balls after r cyclic interchanges of one ball. The same problem was solved by Lagrange (1777/1869, pp. 249 – 251), Malfatti (Todhunter 1865, pp. 434 – 438) and Laplace (1811; and in the same way in the TAP). Laplace worked out a partial difference equation, *mutilated it most unsparingly* (Todhunter 1865, p. 558) and expressed its solution in terms of functions related to the

[Chebyshev –] Hermite polynomials (Molina 1930, p. 385). Hald (1998, p. 339) showed, however, that Todhunter's criticism was unfounded.

Markov (1915b) generalized this problem by considering the cases of $n \rightarrow \infty$ and $r/n \rightarrow \infty$ and $n \rightarrow \infty$ and $r/n = \text{const}$ and Steklov (1915) proved the existence and uniqueness of the solution of Laplace's differential equation with appropriate initial conditions added whereas Hald (2002) described the history of those polynomials. Hostinský (1932, p. 50) connected Laplace's equation with the Brownian movement and thus with the appearance of a random process (Molina 1936).

Like Bernoulli, Laplace discovered that in the limit, and even in the case of several urns, the expected numbers of white balls became approximately equal to one another in each of them irrespective of the initial distribution of the balls. Finally, Laplace (1814/1995, p. 42) added that nothing changed even if new urns, again with arbitrary distributions of the balls, were placed among the original urns. He declared, apparently too optimistically, that

These results may be extended to all naturally occurring combinations in which the constant forces animating their elements establish regular patterns of action suitable to disclose, in the very mist of chaos, systems governed by these admirable laws.

The Daniel Bernoulli – Laplace problem coincides with the celebrated Ehrenfests' model (§ 6.1.1).

4) I touch on **Chapter 4** in § 7.2-3. Laplace devoted **Chapter 5** to the detection of constant causes (forces) in nature. Thus, he attempted to estimate the significance of the daily variation of the atmospheric pressure. K. Pearson (1978, p. 723) remarked that nowadays the Student distribution could be applied in such cases, that some of Laplace's assumptions proved to be wrong and that Laplace unjustifiably rejected those days during which the variation exceeded 4 mm.

Laplace solved Buffon's problem on geometric probability anew. To repeat (§ 6.1.6), a needle of length $2r$ falls from above on a set of parallel lines a distance $a \geq 2r$ apart and the probability p that the needle intersects a line is

$$p = 4r/\pi a.$$

Without proof Laplace mistakenly stated that, for $a = 1$, $2r = \pi/4$ was the optimal length of the needle for statistically determining π although he had the correct answer, $2r = 1$, in the first edition of the TAP. A much easier justification than provided by commentators, Todhunter (1865, pp. 590 – 591) and Gridgeman (1960), is possible: $|d\pi| = (4r/p^2)dp$ so that p , and therefore r , ought to be maximal, and according to the condition of the problem, $r = a/2 = 1/2$.

5) In **Chapter 6** Laplace solved some problems by means of the Bayes approach (see § 5.1) although without referring to him; true, he mentioned Bayes elsewhere (1814/1995, p. 120). Here is one of them. Denote the unknown probability that a newly born baby is a boy by x and suppose that during some time p boys and q girls were born. Then the probability of that *compound* event will be proportional to

$$(1) \quad y = x^p(1-x)^q,$$

$$P(a \leq x \leq b) = \int_a^b y dx \div \int_0^1 y dx, \quad 0 < a < b < 1.$$

For large values of p and q Laplace expressed that probability by an integral of an exponential function of a negative square.

For the curve (1) the point of its maximum

$$(2) \quad \alpha = p/(p+q)$$

seems to be chosen by Laplace as a natural estimator of x , but $E\xi$, or, more precisely, the expectation of a random variable ξ with distribution

$$x^p(1-x)^q \div \int_0^1 x^p(1-x)^q dx,$$

does not coincide with (2): the latter is only an asymptotically unbiased estimator of x . This expectation is evidently

$$E\xi = \frac{p+1}{p+q+2}.$$

After discussing the bivariate case Laplace solved another problem. Suppose that the inequality $p > q$ persisted during a number of years. Determine the probability that the same will happen for the next hundred years (under invariable social and economic conditions!).

He also considered the celebrated problem about the probability of the next sunrise. Finally, Laplace determined the population of France given sampling data, and, for the first time ever, estimated the precision of (his version of) sampling. Suppose that N and n are the known numbers of yearly births in France as a whole and in some of its regions and m is the population of those regions. Laplace naturally assumed that $M = (m/n)N$. He then had to estimate the fraction, see Hald (1998, p. 288),

$$\int_0^1 x^{N+n} (1-x)^{m-n+M-N} dx \div \int_0^1 x^n (1-x)^{m-n} dx.$$

K. Pearson (1928) achieved a reduction of the variance of his result; it should have been multiplied by $[(N-n) \div (N+n)]^{1/2}$. Here are his two main remarks. First, Laplace considered (m, n) and (M, n) as independent samples from the same infinite population whereas they were not independent and the very existence of such a population was doubtful. Second, Laplace chose for the magnitude sought an absolutely inappropriate uniform prior distribution (as is usual when keeping to the Bayesian approach).

The first remark had to do with Laplace's supplementary urn problem. Suppose that an urn contains infinitely many white and black balls. After n drawings with replacement m white balls were extracted; a second sample of

an unknown volume ξ (with expected value nr/m) provided r white balls. Laplace derived a limit theorem

$$P(|\xi - nr/m| < z) = 1 - 2 \int (m^3 / \sqrt{\pi S}) \exp(-m^3 z^2 / S) dz, \quad S = 2nr(n-m)(m+r).$$

The limits of integration, as Laplace formally assumed, were z and ∞ .

6) In **Chapter 7** Laplace studied the influence of a possible inequality of probabilities assumed equal to each other. For example, when tossing a coin the probability of heads can be $p = (1 \pm a)/2$ with an unknown a . Supposing that both signs were equally probable, Laplace derived the probability of throwing n heads in succession; for $n > 1$ it was greater than $1/2^n$. Then he considered the general case: the probability was $(p + z)$, $|z| \leq a$, with density $\varphi(z)$.

Turning to urn problems, Laplace assumed that the probabilities of extracting tickets from them were not equal to one another. However, the inequalities will be reduced had the tickets been put into the urn according to their random extraction from an auxiliary urn, and still more reduced in case of additional auxiliary urns. Laplace justified this statement by a general principle: randomness diminished when subjected to more randomness. This is perhaps too general, but Laplace's example was tantamount to reshuffling a deck of cards (to events connected into a Markov chain), and his conclusion was correct (Feller 1950, § 9 of Chapter 15).

7) **Chapter 8** was devoted to the mean durations of life and marriages. Laplace did not apply there any new ideas or methods. However, he studied anew the Daniel Bernoulli model of smallpox (§ 6.2.3), adopted more general assumptions and arrived at a more general differential equation (Todhunter 1865, pp. 601 – 602).

8) In **Chapter 9** Laplace considered calculations made in connection with annuities and introduced the *Poisson* generalization of the Jakob Bernoulli theorem (Molina 1930, p. 372). Suppose that two contrary events signify a gain v and a loss μ and can occur in each independent trial i with probabilities q_i and p_i respectively, $q_i + p_i = 1$, $i = 1, 2, \dots, s$. For constant probabilities q and p the expected gain after all these trials will be $s(qv - p\mu)$, as Laplace for some reason concluded in a complicated way. He then considered the case of a large s by means of his theorem of § 7.1-3. Then, generalizing the result obtained to variable probabilities, he introduced the characteristic function of the final gain

$$[p_1 + q_1 \exp(v_1 \omega i)] [p_2 + q_2 \exp(v_2 \omega i)] \dots [p_s + q_s \exp(v_s \omega i)],$$

applied the inversion formula and obtained the normal distribution, all this similar to the derivation of the law of distribution of a linear function of observational errors (§ 7.2-3).

9) In **Chapter 10** Laplace described moral expectation (§ 6.1.1). If the physical capital of a gambler is x , his moral capital will be

$$y = k \ln x + \ln h, \quad h, x > 0.$$

Let Δx take values a, b, c, \dots with probabilities p, q, r, \dots . Then

$$E_y = k[p \ln(x + a) + q \ln(x + b) + \dots] + \ln h,$$

$$E\Delta y < E\Delta x.$$

Thus, even a just game of chance ($E\Delta x = 0$) was disadvantageous. He then considered shipping of freight and proved that it should be evenly distributed among several vessels, see my paper (1972/1977, pp. 112 – 113). On this occasion Laplace (1814/1995, p. 89) expressed himself in favour of insurance of life and compared a nation with an association *whose members mutually protect their property by proportionally supporting the costs of this protection*.

10) In the **eleventh**, the last, **Chapter**, and, in part, in **Supplement 1** to the TAP, Laplace examined the probability of testimonies. Suppose that an urn contains 1,000 numbered tickets. One of them is extracted, and a witness states that that was ticket number i , $1 \leq i \leq 1,000$. He may tell the truth and be deceived or not; or lie, again being deceived or not. Laplace calculated the probability of the fact testified by the witness given the probabilities of all the four alternatives. In accordance with one of his corollaries, the witness's mistake or lie becomes ever more probable the less likely is the fact considered by itself.

Laplace next introduced the prior probability of a studied event confirmed by m witnesses and denied by n others. If it was $1/2$ and the probability of the truthfulness of each witness was p , then the probability of the event was

$$P = \frac{p^{m-n}}{p^{m-n} + (1-p)^{m-n}}.$$

He derived several other pertinent formulas, for example, describing the probability of an event reported by a chain of witnesses, and examined verdicts brought in by s independent judges (jurors) assuming that each of them decides justly with probability $p > 1/2$.

Then, if the probability of a just verdict reached by each judge (juror) was unknown, and p judges condemned, and q of them acquitted the defendant, the indirect probability of a just final verdict was

$$\int_{1/2}^1 u^p(1-u)^q du \div \int_0^1 v^p(1-v)^q dv.$$

In passing, Laplace (1816, p. 523) stated that the verdicts were independent.

7.2. Theory of Errors

In the 18th century, Laplace was applying the comparatively new tool, the density, and trying out several rules for the selection of estimators for the constants sought. His equations proved too complicated and he had to keep to the case of three observations. Later Laplace proved (not rigorously) several versions of the CLT and was able to drop his restriction, but he had to adopt other conditions. Bienaymé (1853/1867, p. 161) remarked that

For almost 40 years Laplace had been presenting [...] memoirs on probabilities, but [...] did not want to combine them into a general theory, –

until the (non-rigorously proved) CLT enabled him to compile his TAP. His main achievements in error theory belong to the 19th century. Laplace (1774/1891, p. 56) noted the appearance of *un nouveau genre de problème sur les hasards* and even (1781/1893, p. 383) of *une nouvelle branche de la théorie des probabilités* (of mathematical statistics).

He (1814/1995, p. 37) provided examples of *statistical determinism*, – of the stability of the number of dead letters and of the profits enjoyed by organizers of lotteries. He (1796/1884, p. 504) qualitatively (and wrongly, see § 7.3) explained irregularities in the Solar system by the action of random causes. Elsewhere he (1812/1886, p. 361) stated that a certain magnitude, although having been indicated by [numerous] observation[s], was neglected by most astronomers, but that he had proved its high probability and then ascertained its reality. Thus, Laplace understood that in general, unavoidable ignorance concerning a single random event becomes a cognizable regularity.

1) **In the 18th century**, he published two interesting memoirs (1774; 1781) hardly useful from the practical side. Thus, he introduced, without due justification, two academic density curves. Already then, in 1781, Laplace offered his main condition for adjusting direct observations: the absolute expectation of error should be minimal. In the 19th century, he applied the same principle for justifying the MLSq, which was only possible for the case of normal distribution (existing on the strength of his non-rigorous proof of the CLT when the number of observations was large).

In 1781, Laplace proposed, as a density curve,

$$\varphi(\alpha x) = 0, x = \infty; \varphi(\alpha x) = q \neq 0, x \neq \infty, \alpha \rightarrow 0.$$

His deliberations might be described by the Dirac delta-function which had already appeared in Euler's works (Truesdell 1984, p. 447, Note 4, without an exact reference). One of his conclusions was based on considering an integral which has no meaning in the language of generalized functions, but his finding is extremely interesting on the *physical level*.

Also in 1781 he first discussed the problem 7.1-2c. Some of Laplace's assumptions were not really justified, and he (Laplace 1814/1995, p. 116; 1798 – 1825/1878 – 1882, t. 3, p. xi) 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*.

2) **The Years 1810 – 1811.** Laplace (1810a) considered n [independent] discrete random magnitudes uniformly distributed on interval $[-h; h]$. After applying a particular case of characteristic functions and the inversion formula, he proved, carelessly and non-rigorously, that, in modern notation, as $n \rightarrow \infty$,

$$\lim P\left(\frac{|\sum \xi_i|}{n} \leq s\right) = \frac{\sqrt{3}}{\sigma\sqrt{2\pi}} \int_0^s \exp\left(-\frac{x^2}{2\sigma^2}\right) dx, i = 1, 2, \dots, n,$$

where $\sigma^2 = h^2/3$ was the variance of each ξ_i . He then generalized his derivation to identically but arbitrarily distributed variables possessing

variance. When proving the CLT he (1810a/1898, p. 304) made use of an integral of a complex-valued function, remarked that he hoped to interest *géomètres* in that innovation and thus separated himself from (pure) mathematicians, see also similar reservations elsewhere (Laplace 1774/1891, p. 62; 1812/1886, p. 365).

In a supplement to the first-mentioned memoir Laplace (1810b), apparently following Gauss, justified the principle of least squares without making any assumptions about the arithmetic mean (cf. § 9.1.3), but he had to consider the case of a large number of observations and to suppose that the means of their separate groups were also normally distributed. Soon enough Laplace (1811) returned to least squares. This time he multiplied the observational equations in one unknown

$$a_i x + s_i = \varepsilon_i, i = 1, 2, \dots, n,$$

where the right sides were errors rather than residuals, by indefinite multipliers q_i and summed the obtained expressions:

$$[aq]x + [sq] = [\varepsilon q].$$

The estimator sought was

$$x_0 = -[sq]/[aq] + [\varepsilon q]/[aq] \equiv -[sq]/[aq] + m.$$

Tacitly assuming that all the multipliers q_i were of the same order, Laplace non-rigorously proved a version of the local CLT for $n \rightarrow \infty$:

$$P(m = \alpha) = \frac{1}{\sigma_m \sqrt{2\pi}} \exp\left(-\frac{\alpha^2}{2\sigma_m^2}\right), \sigma_m^2 = k'' \frac{[qq]}{[aq]^2}, k'' = \int_{-\infty}^{\infty} x^2 \psi(x) dx$$

where $\psi(x)$ was an even density of observational errors possessing variance. Then Laplace determined the multipliers q_i by demanding minimal absolute expectation of error which led him to the principle of least squares (in the case of one unknown) so that $x = [as]/[aa]$.

Finally, Laplace generalized his account to the case of two unknowns. The derived principle of least squares essentially depended on the existence of the normal distribution. No wonder that Laplace's theory was not practically useful.

3) **Chapter 4 of the TAP.** Laplace non-rigorously proved the CLT for sums and sums of absolute values of independent, identically distributed errors restricted in value as well as for the sums of their squares and for their linear functions. All, or almost all of this material had already been contained in his previous memoirs although in 1811 he only proved the local theorem for linear functions of errors. In § 23 Laplace formulated his aim: to study the mean result of *observations nombreuses et non faites encore*. This was apparently the first direct statement about general populations.

4) In **Supplement 1 to the TAP** Laplace (1816) considered observational equations in (say) two unknowns

$$a_i x + b_i y + l_i = v_i, i = 1, 2, \dots, s.$$

Suppose that Δx and Δy are the errors of the least-squares estimators of the unknowns, denote the even density of the observational errors by $\varphi(u/n)$ with $|u| \leq n$, the moments by the letter k with appropriate subscripts, $\xi = \Delta x\sqrt{s}$, $\eta = \Delta y\sqrt{s}$,

$$\beta^2 = \frac{k}{kk_4 - 2k_2^2}, Q^2 = \sum_{i=1}^s (a_i\xi + b_i\eta)^2 \text{ and } t = \frac{[vv]}{\sqrt{s}} - \frac{2k_2n^2\sqrt{s}}{k}.$$

Laplace calculated

$$P(\xi; \eta) \sim \exp\{-Q^2(2[vv] - 2t\sqrt{s})\}, P(t) \sim \exp\{-\frac{\beta^2}{4n^4} [t + (Q^2/s\sqrt{s})]^2\}.$$

He also obtained $P(\xi; \eta; t)$ which showed that t was independent of $\xi; \eta$; or, that the sample variance was independent from the estimators of the unknowns, cf. § 9.2; to repeat, the observational errors were assumed to possess an even distribution, – and a normal distribution in the limit. For a proof see Meadowcroft (1920).

Finally, Laplace derived a formula for estimating the precision. Without explanation (which appeared on p. 571 of his *Supplement 2*) he approximated the squared sum of the real errors by the same sum of the residuals and arrived at an estimator of the variance $m = ([vv]/s)^{1/2}$. Interestingly, Laplace (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*.

5) In **Supplement 2 to the TAP** Laplace (1818) adopted the normal law as the distribution of observational errors themselves and not only as the law for their means. Indeed, as he noted, the new *repeating* theodolites substantially reduced the influence of the error of reading and thus equated its order with that of the second main error of the measurement of angles in triangulation, the error of sighting.

Laplace studied the precision of triangulation without allowing for two conditions corresponding to the existence of two baselines and, possibly, two astronomical azimuths. In line with Bayes (§ 5.1), he tacitly assumed that the parameter of precision of the normal distribution was a random variable. Laplace also discussed the Boscovich method of adjusting meridian arc measurements (§ 6.3.2-3) and showed, basing his derivation on variances rather than on absolute expectations as before, that the Boscovich method was preferable to the MLSq if, and only if,

$$4\varphi^2(0) > \frac{1}{2k''}, k'' = \int_0^\infty x^2\varphi(x)dx$$

where $\varphi(x)$ was the even density of observational errors. In Supplement 3 Laplace again applied the variance as the main measure of precision of the observations.

According to Kolmogorov (1931), the median is preferable to the arithmetic mean if for the population median m

$$\frac{1}{2\sigma\varphi(m)} < 1, \sigma^2 = 2k''.$$

While translating Laplace's *Mécanique Céleste* into English, Bowditch (Laplace 1798 – 1825/1832, vol. 2, § 40, Note) stated:

The method of least squares, when applied to a system of observations, in which one of the extreme errors is very great, does not generally give so correct a result as the method proposed by Boscovich [...]; the reason is, that in the former method, this extreme error [like any other] affects the result in proportion to the second power of the error; but in the other method, it is as the first power.

In other words, the robustness of the Boscovich method is occasioned by its connection with the median.

7.3. Critical Conclusions

Laplace (1814/1995, p. 2) stated that, for a mind, able to *comprehend* all the natural forces, and to *submit these data to analysis*, there would exist no randomness *and the future, like the past, would be open* to it. Nowadays, this opinion cannot be upheld because of the recently discovered phenomenon of chaos (§ 1.1). Then, such a mind does not exist so that he actually recognized randomness, and Maupertuis (1756a, p. 300) and Boscovich (1758, §385) had anticipated him. Laplace (1776/1891, pp. 144 – 145) did not formally recognize randomness and explained it by ignorance of the appropriate causes, or by the complexity of the studied phenomenon; in such cases, we nevertheless recognize randomness (§ 1.1). Without indicating the dialectical link of randomness and necessity, he even declared that the theory of probability was indebted for its origin to the weakness of the mind.

He applied an unsuitable model when calculating the population of France and he insisted on his own impractical justification of the MLSq and virtually neglected Gauss. He had not even heuristically introduced the notion of *random variable* and was unable to study densities or characteristic functions as mathematical objects. His theory of probability therefore remained an applied mathematical discipline unyielding to development which necessitated its construction anew. 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.

It is opportune to conclude by quoting Fourier (1829, pp. 375 – 376):

We cannot affirm that it was his destiny to create a science entirely new, like Galileo and Archimedes; to give to mathematical doctrines principles original and of immense extent, like Descartes, Newton and Leibniz; or, like Newton, to be the first to transport himself into the heavens, and to extend to all the universe the terrestrial dynamics of Galileo: but Laplace was born to perfect everything, to exhaust everything, and to drive back every limit, in order to solve what might have appeared incapable of solution. He would have completed the science of the heavens, if that science could have been completed.

Noteworthy in my context was also Fourier's opinion (p. 379) about Laplace (1796):

If he writes the history of great astronomical discoveries, he becomes a model of elegance and precision. No leading fact ever escapes him. [...] Whatever he omits, does not deserve to be cited.

So, did Fourier fail to notice Laplace's mistake, or was he ignorant of Newton's discovery as well?

8. Poisson

Poisson introduced the concepts of random variable and distribution function. He contributed to limit theorems and brought into use the law of large numbers proving it for the case of *Poisson trials*, studied criminal statistics and systematically determined the significance of empirical discrepancies which proved essential for the development of statistics and stressed the difference between subjective and objective probabilities.

Key words: law of large numbers, criminal statistics, medical statistics, null hypothesis

8.1. Probability and Statistics

8.1.1. General Statements. Libri Carruci et al (1834, p. 535) advocated the need for basing statistics on the theory of probability:

The most sublime problems of the arithmétique sociale can only be resolved with the help of the theory of probability.

For their part, Double et al (1835, p. 174) noted its connection with treating mass phenomena:

Statistics carried into effect always is, after all, the functioning mechanism of the calculus of probability, necessarily concerning infinite [?] masses, an unrestricted number of facts; and (p. 176) [with respect to the applicability of mathematics] the state of the medical sciences is not worse than, not different from the situation with all the physical and natural sciences, jurisprudence, moral and political sciences etc.

Social arithmetic denoted demography, medical statistics and actuarial science. In both cases Poisson was co-author.

He (1837a, pp. 1 and 36) remarked that in the 18th century, probability became a main branch of mathematics with respect both to the number and utility of its applications and the kind of analysis which it engendered. And, too categorically: No other part of mathematics is capable of more or more immediately useful applications.

On pp. 35 – 36 Poisson defined the aims of the theory of probability as determining the cause for believing that a thing is true, and for comparing those causes for widely differing problems. He added that *These principles should be regarded as a necessary supplement to logic*. The connection of probability with logic originated with Leibniz and Lambert and was developed in the 19th century by Boole, Jevons and Venn.

Poisson (1837a) consistently demanded to check the significance of empirical discrepancies, for example between results of different series of observations; along with Bienaymé, he was therefore the Godfather of the Continental direction of statistics (§ 14.1). True, his approach was restricted (§ 8.5).

Poisson's programme (1837c) of probability calculus paid serious attention to population and medical statistics and statistics of financial institutions. He also participated in reviewing the desirability of establishing a tontine (§ 2.1.3) (Fourier et al 1826). The reviewers opposed that project.

Poisson (1837a, pp. 140 – 141) introduced a discrete random variable but called it by a provisional term, *chose A*, then (p. 254) went over to a continuous random variable. There (1837a, p. 274), and earlier he (1833, p. 637) corroborated the transition from discrete to continuous by a trick that can be described by Dirac's delta-function. When considering density $\varphi(x)$ equal to zero everywhere excepting a finite number of points $c_i, i = 1, 2, \dots, n$, and such that

$$\int_{c_i-\varepsilon}^{c_i+\varepsilon} \varphi(x) dx = g_i, \sum g_i = 1, \varepsilon \rightarrow 0,$$

Poisson had thus introduced that function of the type of $\varphi(x) = \sum g_i \delta(x - c_i)$.

He (1829, § 1; 1837b, pp. 63 and 80) also defined the distribution function as $F(x) = P(\xi < x)$ and the density as the derivative of $F(x)$. Such functions only became widely used in the 20th century although Davidov (1885) noted Poisson's innovation (Ondar 1971).

8.1.2. The De Moivre – Laplace Theorem. Poisson (p. 196) derived the integral De Moivre – Laplace theorem with a correction term as an asymptotic corollary of his own formula (p. 189) for the probability of an event occurring not less than m times in $\mu = m + n$ Bernoulli trials with probability p of its happening in each trial and $q = 1 - p$

$$P = p^m \left\{ 1 + mq + \frac{m(m+1)}{2!} q^2 + \dots + \frac{m(m+1) \dots (m+n-1)}{n!} q^n \right\}.$$

For small values of q Poisson (p. 205) then derived the approximation

$$P \approx e^{-\omega} (1 + \omega + \omega^2/2! + \dots + \omega^n/n!)$$

where $mq \approx \mu q = \omega$. He had not provided the expression $P(\xi = m) = e^{-\omega} \omega^m/m!$.

His approximation had been all but ignored, for example, by Cournot (1843), until Bortkiewicz (1898a) introduced his law of small numbers, allegedly a breakthrough. Kolmogorov (1954), however, identified it as the Poisson formula, but did not justify his statement, and I (2008) proved it.

8.1.3. Poisson Trials. Poisson (1837a) generalized the law introduced by Jakob Bernoulli (§ 3.2.3) on the case of variable probabilities of success in different trials. Suppose that contrary events A and B occur in trial j with probabilities p_j and q_j ($p_j + q_j = 1$). Poisson (p. 248) determined the probability that in s trials event A occurred m times, and event B , n times ($m + n = s$). He wrote out the generating function of the random variable m (or, the bivariate generating function of m and n) as

$$X = (up_1 + vq_1) (up_2 + vq_2) \dots (up_s + vq_s)$$

so that the probability sought was the coefficient of $u_m v_n$ in the development of X . His further calculations (lacking in Chapter 9 of Laplace's TAP) included transformations

$$u = e^{ix}, v = e^{-ix}, up_j + vq_j = \cos x + i(p_j - q_j) \sin x = \rho_j \exp(ir_j),$$

$$\rho_j = \{\cos^2 x + [(p_j - q_j) \sin x]^2\}^{1/2}, r_j = \arctan[(p_j - q_j) \tan x].$$

Excluding the case of p_j or q_j decreasing with an increasing s , and without estimating the effect of simplifications made, Poisson (pp. 252 – 253) derived the appropriate local and integral limit theorems. They were, however, complicated and their importance apparently consisted in extending the class of studied random variables.

8.1.4. Subjective Probabilities. In connection with a generally known game Poisson (1825 – 1826) studied sampling without replacement. Cards were extracted one by one from six decks shuffled together as a single whole until the sum of the points in the sample obtained was in the interval [31; 40]. The sample was not returned and a second sample of the same kind was made. It was required to determine the probability that the sums of the points were equal. Poisson was able to solve this difficult analytical problem; from the statistical angle, it was interesting in that the probabilities involved were assumed the same for both samples which is characteristic for the subjective viewpoint, cf. § 1.2. Another problem (Poisson 1837a, p. 47) concerned an urn containing a finite number of white and black balls in an unknown proportion. The subjective probability of extracting a white ball occurred to be equal to 1/2. In the theory of information, it would have meant least possible information. Just the same, the unknown probability of each outcome of a coin toss can be proved to be 1/2.

8.1.5. The Law of Large Numbers. Here is how he (1837a, p. 7) defined this law in his Préambule:

Things of every kind obey a universal law that we may call the law of large numbers. Its essence is that if we observe a very large number of events of the same nature, which depend on constant causes and on causes that vary irregularly, sometimes in one manner sometimes in another, i.e., not progressively in any determined sense, then almost constant proportions will be found among these numbers.

He went on to state qualitatively that the deviations from his law became ever smaller as the number of observations increased. Bortkiewicz (1904, p. 826, Note 13) remarked that the Préambule was largely contained in Poisson's previous work (1835). Poisson (1837a, pp. 8 – 11) illustrated his vague definition by various examples, which, however, did not adequately explain the essence of the law but were interesting indeed. Thus (pp. 9 and 10), the LLN explains the stability of the mean sea level and the existence of a mean interval between molecules. Beginning with 1829, Poisson's contributions had been containing many direct or indirect pronouncements on molecular conditions of substance, local parameters of molecular

interactions, etc. sometimes connected with the LLN (Sheynin 1978b, p. 271, note 25).

Poisson then (pp. 138 – 142) formulated but did not prove three propositions characterizing the LLN. These were based on the standard formula (which Poisson had not written out)

$$P(B) = \sum P(A_i) P(B/A_i).$$

In actual fact, he studied the stability of statistical indicators by means of the CLT, see Hald (1998, pp. 576 – 582). Poisson described his law in a very complicated way; no wonder that Bortkiewicz (1894 – 1896, Bd. 8, p. 654) declared that *There hardly exists such a theorem that had met with so many objections as the law of large numbers*. Here, in addition, is a passage from Bortkiewicz' letter to Chuprov of 1897 (Sheynin 1990a/2011, p. 60):

Or take [...] my last three-hour talk with Markov about the law of sm. [small] numbers [§ 14.1.2]. It caused me nothing but irritation. He again demanded that I change the title. With respect to this topic we got into conversation about the law of l. nn. It happens that Markov (like Chebyshev) attributes this term to the case when all the probabilities following one another in n trials are known beforehand. [...] In concluding, Markov admitted that perhaps there did exist 'some kind of ambiguity' in Poisson's reasoning, but he believed that it was necessary to take into account the later authors' understanding of the term 'law of l. nn.'

It is indeed difficult to examine Poisson's considerations on that point, but at least one of his examples (p. 148ff) is clear. It deals with a throw of many coins of the same denomination and *mode de fabrication*. And, although Poisson (p. 147) argued that the probability of (say) heads could be established statistically, it seems that his example had to do with unknown probabilities. Other examples mentioned above (sea level and interval between molecules) can only be understood to include unknown probabilities.

The LLN was not recognized for a long time. In 1855 Bienaymé declared that it contained nothing new (§ 10.2-2) which apparently compelled Cournot (1843) to pass it over in silence; Bienaymé came to this view even in 1842 (Heyde & Seneta 1977, pp. 46 – 47). Even much later Bertrand (1888a, pp. XXXII and 94) considered it unimportant and lacking in rigour and precision. However, already Bessel (1838b, especially § 9) guardedly called the LLN a *principle* of large numbers, Buniakovsky (1846, p. 35) mentioned it and Davidov (1854?; 1857, p. 11) thought it important.

There is a lesser known aspect of the LLN (§ 5.2). Bernoulli, De Moivre and Poisson (the *Poisson form of the LLN*) alleged that their findings were just as applicable for the inverse case, in which the probability p (or probabilities p_i) was (were) unknown and had to be estimated by the observed frequency. Even more: Bernoulli and Poisson (1836; 1837a) thought that even the existence of p (or p_i) was (were) not necessary. The former provided an example of an individual being taken ill by an infectious disease, Poisson mentioned several such cases as stability of the mean sea level, of the mean interval between molecules of a body, and (1837a, § 59) of the sex

ratio at birth. Nevertheless, it was Bayes (§ 5.2) who investigated the inverse case.

Poisson proved his LLN by issuing from the CLT which he (1837a, pp. 254 – 271) was yet unable to justify rigorously; he had not even stated the imposed conditions clearly enough. He also applied the CLT for estimating the significance of discrepancies between indicators obtained from different series of observations and Cournot (1843, Chapters 7 and 8) borrowed his findings without mentioning him. Poisson proved the CLT even earlier (1824; 1829). In the first instance he introduced the Cauchy distribution and found out (1824, §§ 4 and 6) that it was stable.

Statisticians only recognized the LLN for the case of Bernoulli trials, and only when the probability of the studied event existed, otherwise they refused to turn to the theory of probability at all (§ 3.2.3). Even worse, as a rule, they only understood the LLN in a loose sense (Ibidem).

8.2. Theory of Errors

In the theory of errors Poisson offered his proof of the CLT and a distribution-free test for the evenness of the density of observational errors (1829, § 10). When discussing the precision of firing, Poisson (1837b, p. 73) stated that the less was the scatter (the appropriate variance) of hit-points, the better was the gun. He thus made a step towards recognizing Gauss' choice of least variance as the criterion for adjusting observations (§ 9.1.3), but he followed Laplace and never mentioned the Gauss theory of errors partly since French mathematicians had been reasonably angered by Gauss' attitude towards Legendre (§ 9.1.1). One of his problems (1837b, § 7) consisted in determining the distribution of the square of the distance of some point from the origin given the normal distributions of the point's distances from the two coordinate axes. He thus was perhaps the first to treat clearly the densities as purely mathematical objects.

8.3. Criminal Statistics

Unlike Laplace, Poisson introduced the prior probability of the defendant's guilt, not to be applied in individual cases. One of Poisson's statements (1837a, pp. 375 – 376) is debatable: he thought that the rate of conviction should increase with crime. At the same time he (p. 21) recognized that criminality represented *l'état moral de notre pays*.

The application of probability theory to jurisprudence had been criticized time and time again. Poinsot (Poisson 1836, p. 380) called it *une fausse application de la science mathématique* and unwisely quoted Laplace (1814/1886, p. XI) who had remarked that the theory of probability was very delicate. *Unwisely*, because the same *Essai* contained a page (p. LXXVIII) entitled *Application du calcul des probabilités aux sciences morales* where Laplace declared that such applications were the *effets inévitables du progrès des lumières*. The same *Essai* also contained three chapters devoted to such applications to say nothing of Laplace's own work on criminal statistics.

Then, Mill (1843/1886, p. 353) had called the application of probability to jurisprudence an *opprobrium* [disgrace] of mathematics.

In 1899, Poincaré (Sheynin 1991a, p. 167) approvingly cited him in connection with the notorious Dreyfus case and later (1896/1912, p. 20) stated that people act like the *moutons de Panurge*.

Leibniz, in his letters to Jakob Bernoulli in the very beginning of the 18th century, Mill, and several modern authors, rather stressed the importance of the pertinent circumstances. Nevertheless, Heyde & Seneta (1977, pp. 28 – 34) discussed criminal statistics and noticed, on p. 31, that *there was a surge of activity stimulated by Poisson*. Gelfand & Solomon (1973) reviewed Poisson's study and included information about the French legal system of his time.

8.4. Statistical Physics

Poisson qualitatively connected his LLN with the existence of a stable mean interval between molecules (Gillispie 1963, p. 438). The creators of the kinetic theory of gases could have well mentioned this opinion as also his important related considerations, but nothing of the sort actually happened.

8.5. Medical Statistics

Double et al (1835, p. 173, 174 and 176) with Poisson as co-author stated that the application of statistics to medicine was quite possible. Anyway, the statistical method did gnaw its way into that science. First, population statistics was closely connected with medical problems. Leibniz (§ 2.1.4) advocated the compilation of various pertinent data. Halley (Ibidem) compiled the first (after Graunt's not really reliable finding) mortality table for a closed population and estimated populations from data on births and deaths. Daniel Bernoulli (§6.2.3) and Lambert (§ 6.2.2) studied mortality, birth rates and sicknesses and their work belongs to the history of probability and of medicine.

Second, the range of application of the statistical method greatly widened after the emergence, in the mid-19th century, of public hygiene (largely a forerunner of ecology) and epidemiology. Third, about the same time surgery and obstetrics, branches of medicine proper, yielded to the statistical method (§ 10.8.1). Fourth and last, in 1825 a French physician Louis (§ 10.8) introduced the so-called numerical method (actually applied much earlier in various branches of science) of studying symptoms of various diseases. His proposal amounted to the use of the statistical method without involving stochastic considerations. Discussions about the work of Louis lasted at least a few decades. Gavarret (1840) noted the shortcomings of the numerical method, popularized the formulas of probability theory and advised to check the null hypothesis. Before taking to medicine, Gavarret had graduated from the Ecole Polytechnique. There, he studied under Poisson whose influence he (1840, p. XIII) sincerely acknowledged. Many authors repeated his recommendations but he was not mentioned in the literature pertaining to the breakthrough in surgery that took place in the mid-century, the introduction of anaesthesia and antiseptic measures (Sheynin 1982, § 6.1). Indeed, numerous observations, advocated by Poisson and him, were needed in other branches of medicine (epidemiology).

Thus, Liebermeister (ca. 1877, pp. 935 – 940) argued that in therapeutics large numbers of observations were lacking and that, anyway, recommendations based on several (reliable) observations should be adopted as well:

Theoreticians rather often categorically tell us, practical physicians, that all our inferences about the advantages or shortcomings of some methods of treatment, so far as they are based on results which have really taken place, simply remain up in the air if only we do not apply rigorous rules of the theory of probability. [...] Physicians have until now applied that theory so seldom not so much because they sometimes did not attach proper significance to it, but mainly since its analytical arsenal was too imperfect and awkward. [...] Mathematicians say: If you, physicians, wish to arrive at plausible conclusions, you must invariably work with large numbers; you ought to collect thousands and hundred thousands observations. [...] This, however, is impossible for statistics of a general practitioner. And, nevertheless, if this condition is fulfilled, it will often be doubtful whether the theory of probability is necessary in the same pressing manner. [...] Gavarret somewhat arbitrarily presumed, as Poisson also did in several problems, that 0.9953 or 212:213 [...] is a sufficient measure of probability. [...] Suppose that the successes of two methods of treatment are only as 10:1, would not that be sufficient for preferring the first one?

Liebermeister's criticism is still valid. Then, beginning with 1863 and even earlier astronomers and geodesists had begun to offer tests for rejecting outliers quite in vein with his reasoning. Previous practitioners had also made plausible inferences on the strength of scarce data (Bull 1959) whereas Niklaus Bernoulli (1709/1975, p. 302) thought that an absentee ought to be declared dead once his death becomes only twice as probable as his remaining alive.

Liebermeister studied the possibility of distinguishing between equality and inequality of success probabilities in two small series of binomial trials. Starting from a Laplacean formula based on the existence of uniform prior distribution, and assuming that the two probabilities coincided, he considered the size of the tail probability (of the hypergeometric distribution). His main formula had hardly ever reappeared. See Seneta (1994).

9. Gauss, Helmert, Bessel

Gauss introduced the MLSq and Helmert completed its development whereas Bessel made important discoveries in astronomy and geodesy but was often extremely inattentive. Gauss' final condition of least variance led to effective estimators of the unknowns sought, to jointly effective in case of the normal distribution of the observational errors.

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

9.1. Gauss

His correspondence and scientific legacy include a study of the mortality of newly-born and of the members of tontines, but his main achievement was the development of the MLSq.

9.1.1. Adjustment of Observations. Denote the observations of a constant sought by

$$x_1, x_2, \dots, x_n, x_1 \leq x_2 \leq \dots \leq x_n. \quad (1)$$

It is required to determine its value, optimal in some sense, and estimate the residual error, its deviation from the *real value* of the unknown constant, i. e., to adjust *direct observations*. The classical theory of errors considers independent observations and, without loss of generality, they might be regarded as of equal weight.

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

$$a_i x + b_i y + c_i z + \dots + s_i = 0 \quad (2)$$

whose coefficients are given by the appropriate theory and the free terms are measured. The approximate values of x, y, z, \dots were usually known, hence the linearity of (2). The equations are linearly independent (a later notion), so that such systems are inconsistent and were *solved* by allowing small enough residual free terms (denote them by v_i). This procedure is called *adjustment of indirect observations*.

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

$$W = [vv] = v_1^2 + v_2^2 + \dots + v_n^2 = \min \quad (3)$$

among all possible values, so that

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

Equations (4) easily lead to a system of *normal equations*

$$[aa]\hat{x} + [ab]\hat{y} + \dots + [as] = 0, [ab]\hat{x} + [bb]\hat{y} + \dots + [bs] = 0, \dots, \quad (5)$$

having a positive definite and symmetric matrix. For direct measurements the same condition (3) leads to the arithmetic mean. Gauss (1828) devised another no less important and barely known to statisticians pattern of adjusting indirect observations.

9.1.2. The Priority Strife. Legendre (1805, pp. 72 – 73) was the first to state publicly the *condition* of least squares. Declaring that the extreme errors without regard to sign should be contained within as narrow limits as possible (which is achieved by the minimax principle rather than by least squares!), he, as translated by Hald (1998, p. 119), continued:

Among all the principles [of adjusting observations] I think there is no one more general, more exact and more easy to apply than that which we have made use of in the preceding researches [in the same contribution], and which consists in making the sum of the squares of the errors a minimum. In this way there is established a sort of equilibrium among the errors, which prevents the extremes to prevail and is well suited to make us know the state of the system most near the truth.

Here, Legendre made a mistake: he should have mentioned not errors, but residuals. Those shortcomings did not deter Stigler (1986, p. 13) who called Legendre's exposition *One of the clearest and the most elegant introduction of a new statistical method in the history of statistics*. And on p. 146 Stigler wrongly praised Legendre as opposed to Gauss.

Gauss publicly derived the *principle* of least squares in 1809, but stated (1809, § 186) that he had applied the *condition* of least squares from 1794 or 1795 and called it his own. His statement offended Legendre (letter to Gauss of 31 May 1809, see Gauss, W-9, p. 309) as well as other French mathematicians although not Laplace.

Gauss (letter of 17.10.1824 to H. C. Schumacher), see W/Erg-5, Tl. 1, p. 413) bitterly lamented over Legendre's fate:

With indignation and distress I have [...] read that the pension of the elderly Legendre, an ornament to his country and his epoch, was cancelled.

May (1972, p. 309) formulated a likely opinion about the problem of priority as approached by Gauss:

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

Here is another comment (Biermann 1966, p. 18): *What is forbidden for usual authors, ought to be allowed for Gausses and in any case we must respect his [Gauss'] initial considerations.*

And of course Legendre's qualitative statement (and, at that, not quite correct) was not comparable to Gauss' deliberations.

Robert Adrain was an American scientist who also derived the normal distribution of observational errors, see Coolidge (1926). His not at all rigorous work was published in an obscure periodical in 1809 (Hogan 1977) rather than in 1808 as stated there.

9.1.3. The Two Justifications of Least Squares (Gauss 1809; 1823b).

In 1809 Gauss (§ 177) assumed *as an axiom* that the arithmetic mean of many observations was the most probable value of the measured constant *if not absolutely precisely, then very close to it*. Together with the principle of maximal likelihood (§ 6.3.1), his axiom or *postulate* (Bertrand 1888a, p. 176) led to the normal distribution of the observational errors as the only possible law. He 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. In his letter to Bessel of 1839 (Plackett 1972/1977, p. 287) he stated that the *largest probability* of the value of an unknown parameter was still infinitely small so that he preferred to rely on the *least disadvantageous game*, on minimal variance.

In 1823 Gauss provided his final justification of the principle of least squares by the principle of maximum weight [of minimal variance]

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

where $\varphi(x)$ was an even unimodal function. He (§§ 18 and 19) also introduced independence of linear functions: they should not contain common observations. Then Gauss (§§ 37 – 38) proved that, for n observations and k unknowns, the unbiased sample variance and its estimator were, respectively,

$$m^2 = E[vv]/(n - k), \quad m_0^2 = [vv]/(n - k). \quad (6a, b)$$

Instead of the mean value, the sum of squares $[vv]$ itself has to be applied. Coupled with the principle of maximal weight, formulas (6) provide effective estimators, as they are now called. Without mentioning Laplace (§ 7.2-4), Gauss (1823b, §§ 37 – 38) noted that the previously known formula was not good enough. Elsewhere, he (1823a/1887, p. 199) stated that the change was also necessary for the *dignity of science*.

Gauss (§ 40) calculated the boundaries of the var m_0^2 by means of the fourth moment of the errors but made a mistake, see § 9.2. *I note that the derivation of (6a) does not depend on the principle of least squares which now follows immediately.*

The unavoidable presence of systematic errors meant that formula (6b) should only be applied after completing all the necessary work. For a triangulation chain closures of the triangles as well as the discrepancies between the baselines situated at the ends of the chain and between its astronomically fixed end lines are computed and provide the v_i 's thus

revealing the influence of systematic errors as much as possible. In particular, during observations at a given station, formula (6b) should not be relied upon; indeed, Gauss observed each angle at each triangulation station until being satisfied that further work was useless. The rejection of outliers remains a most delicate problem; at best, statistical criteria are only marginally helpful.

Gauss' opinion notwithstanding, his first justification of the principle of least squares became generally accepted, in particular because the observational errors were (and are) approximately normal and the work of Quetelet (§ 10.5) and Maxwell (§ 10.8.5) did much to spread the idea of normality whereas his mature contribution (1823b) was extremely uninviting. However, the proof of formula (6a), from which the condition of least squares immediately follows, is not difficult; Gauss himself provided it; it demands linearity and independence of the initial equations (2) and unbiasedness of the sought estimators of the unknowns. Therefore, it is methodically possible to disregard Gauss' main extremely difficult justification of that condition, to rest content with the actual second substantiation. Then, it will not be practically necessary to restrict the description of least squares by his initial reasoning of 1809.

Why then did not Gauss himself change his description accordingly? At least he could have additionally mentioned the alternative. May (1972/1977, p. 309) provided a general comment which likely answers my question: In particular, by *careful and conscious removal from his writings of all trace of his heuristic methods* [Gauss] *maintained an advantage that materially contributed to his reputation*. Much earlier, Kronecker (1901, p. 42) voiced the first part of May's pronouncement.

Examples of deviation from the normal law were accumulating both in astronomy and in other branches of natural sciences as well as in statistics, see the same § 10.5 (and the missed opportunity mentioned in § 9.3), which supported the rejection of the first substantiation of the principle of least squares.

Tsinger (1862, p. 1) wrongly compared the importance of the Gaussian and the Laplacean approaches:

Laplace provided a rigorous [?] and impartial investigation [...]; it can be seen from his analysis that the results of the method of least squares receive a more or less significant probability only on the condition of a large number of observations; [...] Gauss endeavoured, on the basis of extraneous considerations, to attach to this method an absolute significance [...]. With a restricted number of observations we have no possibility at all to expect a mutual cancellation of errors and [...] any combination of observations can [...] equally lead to an increase of errors as to their diminution.

Tsinger lumped together both justifications of the principle of least squares due to Gauss. Then, practice demanded the treatment of a finite (and sometimes a small) number of observations rather than limit theorems. Tsinger's high-handed attitude towards Gauss (and his blind respect for Laplace) was not an isolated occurrence, see § 12.2-5. Even a recent author (Eisenhart 1964, p. 24) noted that the existence of the second Gaussian

approach *seems to be virtually unknown to almost all American users of least squares except students of advanced mathematical statistics.*

Gauss (1823b) called his estimators *most plausible*. In mathematical statistics, they are called *consistent* and *effective*, meaning that they converge in probability to the respective unknowns and that, among such estimators, their variance is minimal. In case of the normal distribution, they are *jointly effective* (Petrov 1954) which means that the joint distribution of two (say) estimators has the least variance among such distributions of any other two estimators (Cramér 1946, § 32.6). All this is in spite of Markov (1899a/1951, p. 246) who defended Gauss' second justification of the principle of least squares but declared that the MLSq was not optimal in any sense (hence, did not need any justification!).

9.1.4. The True Value of a Measured Constant. Astronomers, geodesists, metrologists and other specialists making measurements have always been using this expression. 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.

Incidentally, 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 more than a century before him. He determined the *véritable objet de la recherche* (the constant sought, or its *real value*) as the limit of the arithmetic mean of n appropriate observations as $n \rightarrow \infty$. Incidentally, he thus provided the Gauss *postulate* with a new dimension.

Many authors, beginning perhaps with Timerding (1915, p. 83) [and including Mises (1919/1964a, 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 *real 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 so that the term *limit* in the Fourier definition (which is in harmony with the Mises definition of probability) must not be understood literally.

9.1.5. Did Gauss Really Apply Least Squares before 1805? I (Sheynin 1999b; 1999d) described the possible cases and named many of his colleagues and friends to whom he had communicated his discovery. Among them were Olbers, Bessel (1832/1848, p. 27), and Wolfgang Bolyai (father of the better known Janos Bolyai, a cofounder of the non-Euclidean geometry).

Unexpectedly, it occurred that von Zach, who allegedly refused to testify to Gauss' priority, had not until 1805 known the formulation of the principle of least squares, and, furthermore, that he (1813, p. 98n) indirectly agreed with the latter's statements by repeating them without any qualification remark:

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

Regrettably, it is not seen there. This passage is even more important than Zach's editorial acceptance of Gauss' priority (Dutka 1996, p. 357): in 1809, Zach's periodical, *Monatliche Correspondenz*, carried an anonymous review of Gauss' *Theoria motus*, and there, on p. 191, Gauss' pertinent claim was repeated.

The case of Olbers is special. 4 Oct. 1809 Gauss asked him: *Do you still remember [...] that [...] in 1803 I talked with you about the principle ...* Olbers apparently did not answer (or answered through a third party). Then, on 24 Jan. 1812 Gauss asked Olbers whether he will publicly attest to the same fact. Yes, *with pleasure*, answered Olbers 10 March 1812 and, indeed, fulfilled his promise in 1816. All this is documented by Plackett (1972/1977, pp. 283 – 285).

However, Stigler (1986, p. 145), for the first time ever, questioned Gauss' integrity: *Gauss solicited reluctant testimony from friends that he had told them of the method before 1805*. And in 1999, on p. 322, repeating his earlier (of 1981) statement of the same ilk: *Olbers did support Gauss's claim [...] but only after seven years of repeated prodding by Gauss*. Grasping at straws, Stigler adds an irrelevant reference to Plackett (1972). But why did Olbers wait several years (1812 – 1816)? Because, during that time he had not published anything suitable, see the appropriate volume of the Royal Society's *Catalogue of Scientific Literature*.

Much later, 3 Dec. 1831, in a letter to H. C. Schumacher (W/Erg-5, Tl. 1, p. 292) Gauss remarked that, had he known (better: recalled) Olbers' intention, he would have objected to it. He apparently became sick and tired with the entire business. Still later Encke (1851, p. 2) stated that Gauss had applied the condition of least squares when determining the orbit of the first-discovered minor planet (in 1802). Gauss did not comment.

Stigler made many other unwarranted and absolutely inadmissible remarks humiliating Gauss (and Euler). Here is one of them (Stigler, 1986, p. 146), appropriate with respect to a suspected rapist, but not to Gauss:

Although Gauss may well have been telling the truth about his prior use of the method, he was unsuccessful in whatever attempts he made to communicate it before 1805.

Gauss' claim about his early use of least squares is not generally accepted, see for example Marsden (1995, p. 185), who nevertheless had not mentioned the opposite opinion of Brendel (1924) and Galle (1924, p. 9) or of Gauss' contemporaries. Gerardy (1977), drawing on archival sources, discovered that Gauss, in 1802 – 1807, had participated in land surveying (in part, for his own satisfaction) and concluded, on p. 19 (note 16) that Gauss started using the method not later than in 1803. Regrettably, Gerardy concentrated on describing Gauss' simple calculations and his statement mentioned just above was not quite definite. Concerning these testimonies, it is not amiss to recall Gauss' opinion (W-14, pp. 201 – 204) about the application of the theory of probability as discussed in a letter of 1841 by W. E. Weber: An approach only based on numbers could be greatly mistaken, the nature of the studied subject also ought to be taken into account.

There are many other instances including that mentioned by von Zach (above) in which Gauss could have well applied his invention at least for preliminary, trial calculations, or short cuts. For him (Gauss 1809, § 185), least squares were not a cut and dry procedure; he allowed himself approximate calculations. Then, possible mistakes in calculations and weighing the observations could have made justification impossible.

9.2 Helmert

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 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 M. G. 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 $\text{var } m_0^2 / m_0^2$ was more important than $\text{var } m_0^2$ by itself and Eddington (1933, p. 280) expressed the same opinion. Czuber also proved that, for the normal distribution, that relative error was minimal for the estimator (6b).

In addition, Helmert noted that for small values of n the $\text{var } m_0^2$ did not estimate the precision of formula (6b) good enough. His considerations led him to the so-called Helmert transformations.

Finally, Helmert (1904) corrected the boundaries of the estimator (6b). Kolmogorov et al (1947) independently repeated his finding and wrote the correct formula more properly whereas Maltzev (1947) proved that the lower bound was attainable.

9.3. 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 (§ 10.1), only much later (§ 12.1-3). Incidentally, Gauss was familiar with the pertinent problem. In the same letter to Bessel of 1839 (§9.1.3), 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 is the systematic difference of the moments of the passage of a star through cross-hairs of the eyepiece of an astronomical instrument as recorded by two observers. When studying this phenomenon, it is necessary to compare the moments fixed by the astronomers at different times and, consequently, to take into account the correction of the clock. Bessel (1823) had indeed acted appropriately, but in one case he failed to do so, and his pertinent observations proved useless; he made no such comment. When studying Bradley's observations, Bessel (1818) 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 2001b, p. 168) stressed Legendre's priority in his dispute. 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.

10. The Second Half of the 19th Century

Many scientists participated in developing the treatment of observations and statistics whose scope had greatly widened. New scientific disciplines inseparably connected with it had originated and some discoveries concerning the theory of probability were also made although its general level did not change.

Key words: new chapters of statistics, new scientific disciplines, theory of evolution, kinetic theory of gases

Here, I consider the work of several scholars, statistics, and its application to various branches of natural sciences. The findings of some natural scientists are discussed separately since it proved difficult to describe them elsewhere but I included Helmert in Chapter 9.

10.1. Cauchy

He published not less than 10 memoirs devoted to the treatment of observations and the theory of probability. In particular, he studied the solution of systems of equations by the principle of minimax (§ 1.7), proved the theorem in linear programming known to Gauss (§ 6.3.2-3) and applied the method of averages (§ 6.3.2-2). Linnik (1958/1961, § 14.5) found out that the pertinent estimators were unbiased and calculated their effectiveness for the cases of one and two unknown(s).

Cauchy (1853b) derived the even density of observational errors demanding that the probability for the error of one of the unknowns, included in equations of the type of (6.9), to remain within a given interval, was maximal. Or, rather, he derived the appropriate characteristic function

$$\varphi(\theta) = \exp(-c\theta^{\mu+1}), \quad c, \theta > 0$$

and noted that the cases $\mu = 1$ and 0 led to the normal law and to the *Cauchy distribution*, cf. § 8.1.

In two memoirs Cauchy (1853c; 1853d) proved the CLT for the linear function $A = [m\varepsilon]$ of independent errors ε_i having an even density on a finite interval. In both cases he introduced characteristic functions of the errors and of the function A , obtained for the latter

$$\Phi(\theta) = \exp(-s\theta^2)$$

where $2s$ was close to σ^2 , the variance of A , and arrived at

$$P(|A| \leq \alpha) \approx \frac{\sqrt{2}}{\sigma\sqrt{\pi}} \int_0^{\alpha} \exp(-\frac{x^2}{2\sigma^2}) dx.$$

He had also estimated the errors due to assumptions made and Freudenthal (1971, p. 142) declared that his proof was rigorous; see, however, Heyde & Seneta (1977, pp. 95 – 96).

Cauchy devoted much thought to interpolation of functions, and, in this connection, to the MLSq, but, like Poisson, he never cited Gauss. In one case he (1853a/1900, pp. 78 – 79) even indicated that the MLSq provided most probable results only in accordance with the Laplacean approach [that is, only for the normal distribution] and apparently considered this fact as an essential shortcoming of the method.

10.2. Bienaymé

Heyde & Seneta (1977) described his main findings; I abbreviate their work as HS. Bru et al (1997) published two of Bienaymé's manuscripts and relevant archival sources.

1) The Liapunov inequalities (Bienaymé 1840b; HS, pp. 111 – 112). Without proof, Bienaymé indicated that the absolute initial moments of a discrete random variable obeyed inequalities which could be written as

$$(E|\xi|^m)^{1/m} \leq (E|\xi|^n)^{1/n}, 0 \leq m \leq n.$$

Much later Liapunov (1901a, § 1) proved that

$$(E|\xi|^m)^{s-n} < (E|\xi|^n)^{s-m} < E(|\xi|^s)^{m-n}, s > m > n \geq 0.$$

He applied these inequalities when proving the CLT.

2) The law of large numbers. Bienaymé (1839) noted that the fluctuation of the mean statistical indicators was often greater than it should have been in accordance with the Bernoulli law, and suggested a possible reason: some causes acting on the studied events, as he thought, remained constant within a given series of trials but essentially changed from one series to the next one. Lexis and other *Continental* statisticians took up this idea without citing Bienaymé (Chapter 14) but it was also known in the theory of errors where systematic errors can behave in a similar way. Bienaymé, in addition, somehow interpreted the Bernoulli theorem as an attempt to study suchlike patterns of the action of causes. He (1855/1876) repeated this statement and, on p. 202, he mistakenly reduced the Poisson LLN to the case of variable probabilities whose mean value simply replaced the constant probability of the Bernoulli trials, also see HS, § 3.3.

3) The Bienaymé – Chebyshev inequality (Bienaymé 1853; HS, pp. 121 – 124; Gnedenko & Sheynin 1978/2001, pp. 258 – 262). This is the name of the celebrated inequality

$$P(|\xi - E\xi| < \beta) > 1 - \text{var}\xi/\beta^2, \beta > 0.$$

Differing opinions were pronounced with regard to its name and to the related method of moments. Markov touched on this issue four times. In 1912, in the Introduction to the German edition of his *Treatise* (1900a), he mentioned *the remarkable Bienaymé – Chebyshev method*. At about the same time he (1912b, p. 218) argued that

Nekrasov's statement [that Bienaymé's idea was exhausted in Chebyshev's works] *is refuted by indicating a number of my papers which contain the extension of Bienaymé's method* [to the study of dependent random variables].

Then, Markov (1914/1981, p. 162) added that the *starting point* of Chebyshev's second proof of Poisson's LLN *had been* [...] *indicated by* [...] *Bienaymé* and that in 1874 Chebyshev himself called this proof a consequence of the new method that Bienaymé gave. Nevertheless, Markov considered it *more correct* to call the method of moments after both Bienaymé and Chebyshev, and *sometimes* only after the latter, since *it only acquires significance through Chebyshev's work* [especially through his work on the CLT]. Finally, Markov (*Treatise*, 1924, p. 92) stated that Bienaymé had indicated the main idea of the proof of the inequality, although restricted by some conditions, whereas Chebyshev was the first to formulate it clearly and to justify it.

Bienaymé (1853/1867, pp. 171 – 172) considered a random sum, apparently (conforming to the text of his memoir as a whole) consisting of identically distributed terms, rather than an arbitrary magnitude ξ , as in the formula above. This is what Markov possibly thought of when he mentioned some conditions. HS, pp. 122 – 123, regarded his proof, unlike Chebyshev's substantiation [§ 12.1-2], *short, simple, and* [...] *frequently used in modern courses* ... Yes, Hald (1998, p. 510) repeated it in a few lines and then got rid of the sum by assuming that it contained only one term. Gnedenko (1954/1973, p. 198) offered roughly the same proof but without citing Bienaymé.

Bienaymé hardly thought that his inequality was important (Gnedenko & Sheynin 1978/2001, p. 262; Seneta 1998, p. 296). His main goal was to prove that only the variance was an acceptable estimator of precision in the theory of errors and, accordingly, he compared it with the fourth moment of the sums of random [and independent] errors. Consequently, and the more so since he never used integrals directly, I believe that Chebyshev (1874); see also Gnedenko & Sheynin (1978/2001, p. 262) overestimated the part of his predecessor in the creation of the method of moments. Here are his words:

The celebrated scientist presented a method that deserves special attention. It consists in determining the limiting value of the integral [...] given the values of the integrals...

The integrand in the first integral mentioned by Chebyshev was $f(x)$ and the limits of integration were $[0; a]$; in the other integrals, $xf(x)$, $x^2f(x)$, ... with the same limits of integration, $f(x) > 0$ and $A > a$.

4) Runs up and down (Bienaymé 1874; 1875; HS, pp. 124 – 128). Suppose that n observations of a continuous random variable are given. Without proof Bienaymé indicated that the number of intervals between the points of extremum (almost equal to the number of these points) is distributed approximately normally with parameters

mean ... $(2n - 1)/3$, variance ... $(16n - 29)/90$.

5) The MLSq (Bienaymé 1852; HS, pp. 66 – 71). Bienaymé correctly remarked that least variance for each estimator separately was not as important as the minimal simultaneous confidence interval for all the estimators (as joint efficiency!). He assumed that the distribution of the observational errors was known, made use of its first moments and even introduced the first four cumulants and the multivariate Gram – Charlier series (Bru 1991, p. 13; Hald 2002, pp. 8 – 9). He determined that confidence interval by applying the principle of maximum likelihood, introducing the characteristic function of the vector of the errors and making use of the inversion formula. True, he restricted his choice of the confidence region, but derived here the χ^2 distribution.

6) A branching process (Bienaymé 1845; HS, pp. 117 – 120). Bienaymé had formulated the properties of criticality of a branching process while examining the problem of the extinction of noble families that became attributed to Galton. D. G. Kendall (1975) reconstructed Bienaymé’s proof and reprinted his note.

7) When investigating the stability of statistical frequencies (see also item 2 above), Bienaymé (1840a; HS, pp. 108 – 110) expressed ideas that underlie the notion of sufficient estimators.

10.3. Cournot

He intended his main contribution (1843) for a broader circle of readers. However, almost completely declining the use of formulas, he hardly achieved his goal. Recall also (§ 8.1) that Cournot passed over in silence the LLN. I describe his work as a whole; when referring to his main book, I mention only the appropriate sections. Chuprov (1925a/1926, p. 227) called Cournot *the real founder of the modern philosophy of statistics*. This seems to be exaggerated. He did not *substantiate* and *canonically* prove the LLN (Chuprov 1905/1960, p. 60; 1909/1959, pp. 166 – 168), did not even formulate that law.

1) The aim of the theory of probability. It was *The creation of methods for assigning quantitative values to probabilities* (p. 181). He thus moved away from Laplace (§ 7.1) who had seen the theory as a means for revealing the laws of nature.

2) The probability of an event. Cournot’s definition (§ 18) included geometric probability, which had been lacking any formula, and combined it with the classical case. He (§ 113) also introduced probabilities unyielding to measurement and (§§ 233 and 240.8) called them *philosophical*. They might be related to expert estimates whose treatment is now included in the province of mathematical statistics.

3) The term *médiane*. This is due to Cournot (§ 34).

4) The notion of randomness. Cournot (§ 40) repeated its ancient connection with the intersection of chains of events (my § 1.1), and, in § 43, indirectly connected randomness with unstable equilibrium by remarking that a right circular cone, when stood on its vertex, fell in a *random* direction. Cournot (1851, § 33, Note 38; 1861, § 61, pp. 65 – 66) also recalled Lambert’s attempt to study randomness (see my § 6.1.3), and (1875/1979, pp. 177 – 179) applied Bienaymé’s test (§ 10.2-4) for investigating whether the digits of the number π were random but reasonably abstained from a final conclusion.

5) Dependence between the decisions of judges and/or jurors. Cournot (1838; 1843, §§ 193 – 196 and 206 – 225) gave thought to this issue but his study was hardly successful in the practical sense.

6) A critical attitude towards statistics; a description of its aims and applications (Chapters 7 and 8 and §§ 103 – 120). Statistics (§ 105) should have its theory, rules, and principles, it ought to be most widely applied; its main goal was to ascertain *the knowledge of the essence of things*, to study the causes of phenomena (§ 120) and the *principe de Bernoulli* was its only pertinent sound foundation (§ 115). Statistics had blossomed exuberantly and [the society] should be on guard against its *premature and wrong* applications which might discredit it for some time (§ 103).

7) Explanation of known notions and issues (§§ 64 – 65, 73 – 74).

10.4. Buniakovsky

His treatise (1846) was the first comprehensive Russian contribution so that Struve (1918) called him *a Russian student of the French mathematical school*. A list of his contributions is in *Materialy* (1917).

1) The theory of probability. Buniakovsky (1846, p. I) correctly attributed it to applied mathematics.

2) Moral expectation (see § 6.1.1). Buniakovsky (1846, pp. 103 – 122) independently proved Daniel Bernoulli's conclusion that an equal distribution of the cargo on two ships increased the moral expectation of the freightowner's capital as compared with transportation on a single ship. Later he (1880) considered the case of unequal probabilities of the loss of the ships.

3) Geometric probabilities (§ 6.1.6). Buniakovsky (1846, pp. 137 – 143) generalized the Buffon problem by considering the fall of the needle on a system of congruent equilateral triangles. His geometric reasoning was, however, complicated and his final answer, as Markov (*Treatise*, 1900/1924, p. 270) maintained, was wrong. Earlier Buniakovsky (1836 – 1837) remarked that the solution of such problems might help to determine the values of special transcendental functions.

4) Statistical control of quality. Buniakovsky (1846, Adendum) proposed to estimate military losses by sample data but his study was hardly useful. He (1846, pp. 468 – 469) also indicated that his findings might facilitate the acceptance of *a very large number of articles and supplies* only a fraction of which was actually examined.

5) The history of the theory of probability. Buniakovsky was one of the first after Laplace to consider this subject and made a few of factual mistakes.

6) Population statistics. Buniakovsky (1846, pp. 173 – 213) described various methods of compiling mortality tables, studied the statistical effect of a weakening or disappearance of some cause of death (cf. § 6.2.3), calculated the mean and the probable durations of marriages and associations and, following Laplace, solved several other problems.

After 1846, Buniakovsky continued these investigations. He compiled mortality tables for Russia's Orthodox believers and tables of their distribution by age (1866a; 1866b; 1874) and estimated the number of Russian conscripts ten years in advance (1875b). No one

ever verified his forecast and the comments upon his tables considerably varied. Bortkiewicz (1898b) sharply criticized them. Finally, Novoselsky (1916, pp. 54 – 55) indicated that Buniakovsy's data were inaccurate and incomplete (as Buniakovsky himself had repeatedly stressed) but called his tables *a great step forward*.

7) Buniakovsky's urn problem (1875a) was connected with partition of numbers. An urn contains n balls numbered from 1 through n . All at once, m balls ($m < n$) are extracted; determine the probability that the sum of the numbers drawn was equal to s . This problem is due to Laplace (TAP, Chapter 2) and Laurent (1873) who referred to Euler (1748, Chapter 16).

Markov (1914/1981, p. 162) considered Buniakovsky's treatise (1846) *a beautiful work* and Steklov (1924, p. 177) believed that it was *complete and outstanding*. Buniakovsky did not, however, pay attention to the work of Chebyshev; after 1846, he actually left probability for statistics.

10.5. Quetelet

At the beginning of his scientific career Quetelet visited Paris and I think that Fourier (1821 – 1829) had 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 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), he 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*.

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 never mentioned Darwin and (1846, p. 259) declared that *the plants and the animals have remained as they were when they left the hands of the Creator*. He collected and systematized meteorological observations and described the tendency of the weather to persist by elements of the theory of runs, cf. § 10.8.3. 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 postal charges (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; 1832b, p. 1; 1848b, p. 38), inclinations to crime (1832b, p. 17; 1836, t. 2, p. 171 and elsewhere) and marriage (1848a, p. 77; 1848b, p. 38), – actually, the appropriate probabilities, – and for mistaken (Rehnsch 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 (§ 6.2.2) in originating moral statistics. Quetelet (1848a, p. 82; 1869, t. 2, p. 327) indicated that the inclination to crime of a given person might differ considerably from the apparent mean tendency and (1848a, pp. 91 – 92) 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 (1882) or Block (1886), only applied arithmetic.

The Average man, as he thought, was the type of the nation and even of entire mankind. Reasonable objections were levelled against this concept. Thus, he (1846, p. 216) only mentioned the Poisson LLN in connection with the mean human stature. The Average man was physiologically impossible (the averages of the various parts of the human body were inconsistent one with another), and 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; 1869, t. 2, pp. 304 and 347) 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.

10.6. Galton

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 it. 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 applying rank correlation. Then, he (1863) devised an expedient system of

symbols for weather charts and immediately discovered the existence of previously unknown anticyclones. He (K. 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.

Galton, in 1892, 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 (§ 14.2), and Galton's close relations with Pearson were an important cause of its successes.

10.7. Statistics

Delambre (1819, p. LXVII) argued that statistics was hardly ever engaged in discussions or conjectures and did not aim at perfecting theories, and that political arithmetic ought to be *distinguished* from it. Under statistics he understood geodetic, meteorological and medical data, mineralogical descriptions and even art expositions.

The London Statistical Society declared that statistics *does not discuss causes, nor reason upon probable effects* (Anonymous 1839, p. 1). True, they denied that *the statist* [!] *rejects all deductions*, or that *statistics consists merely of columns of figures* and stated that *all conclusions shall be drawn from well-attested data and shall admit of mathematical demonstration*. This announcement was thus ambiguous; the Society attempted to adhere to its former statement, but in vain. Woolhouse (1873, p. 39) testified that *These absurd restrictions have been necessarily disregarded*. Indeed, that statistics should explain the present state of a nation by considering its previous states was declared a century before (Gatterer 1775, p. 15). And the very title of Dufau (1840) called statistics *The theory of studying the laws according to which the social events are developing*.

During the 19th century the importance of statistics had been considerably increasing. Graunt (1662/1939, p. 79) was not sure whether his work would be *necessary to many, or fit for others, than the Sovereign, and his chief Ministers* [...] and the investigations of the sex ratio at birth (§§ 2.2.4, 3.3.4, 4.4, 6.1.1) had not found direct applications. However, by the mid-19th century it became important to foresee how various transformations will influence society and Quetelet (§ 10.5) repeatedly stressed this point. Then, at the end of the 19th century censuses of population, answering an ever widening range of questions, began to be carried out in various countries. However,

1) Public opinion was not yet studied, nor was the quality of mass production checked by statistical methods.

2) Sampling had been considered doubtful. Cournot (1843) passed it over in silence and Laplace's sample determination of the population of France (§ 7.1-5) was largely forgotten. Quetelet (§ 10.5) 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 (§ 10.8).

3) The development of the correlation theory began at the end of the 19th century (§§ 10.6, 14.2), 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*. See, however, § 13.2-4.

4) Variance began to be applied in statistics only after Lexis (§ 14.1.1), 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.

5) Preliminary data analysis (generally recognized only a few decades ago) is necessary, and should be the beginning of the statistician's work. Halley, in 1701, see § 2.1.4, drew lines of equal magnetic declinations over North Atlantic, also see § 10.8.3, which was a splendid example of such analysis.

6) Econometrics originated only in the 1930s.

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

7) The absence of *equally possible* cases whose existence is necessary for understanding the classical notion of probability. Statisticians repeatedly mentioned this cause, also see § 3.2.3. Lexis (1874/1903, pp. 241 – 242; 1886, pp. 436 – 437; 1913, p. 2091) wavered; he had no integral viewpoint.

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

9) 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. Nevertheless, statisticians 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).

Statisticians did not trust mathematics; see § 3.2 concerning the LLN. They never mentioned Daniel Bernoulli who published important statistical memoirs, see Bibliography, almost forgot insurance, barely understood the treatment of observations (see § 9.1), did not notice Quetelet's mistakes or his inclinations to crime and to marriage (see § 10.5).

Two circumstances explained the situation. First, mathematicians often did not show how to apply their findings in practice. Poisson (1837a) is a good example; his student Gavaret (1840) simplified his formulas, but still insisted that conclusions should be based on a large number of observations which was often impossible (see § 8.5). Second, student-statisticians barely studied mathematics and, after graduation, did not trust it; see § 3.2 re the LLN.

It is not amiss to mention here the 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. The term *mathematical statistics* is apparently due to him.

10.8. Statistics and Natural Sciences

The statistical method gave rise to stellar statistics, epidemiology, public hygiene (the forerunner of ecology), climatology, medical statistics, geography of plants, zoogeography, biometry, and kinetic theory of gases. Opposition to it can be explained by attachment of mean indicators to individuals (Comte 1830 – 1842, t. 3/1893, No. 40, p. 329). Then, Louis (1825) introduced the so-called numerical method by calculating the frequencies of the symptoms of various diseases so as to facilitate diagnosing. He (pp. xvii – xviii) even thought that, given observations, any physician ought to make the same conclusion. Bouillaud (1836) favourably described the numerical method. D’Alembert (§ 6.2.3) offered astounding and patently wrong statements on this subject and Greenwood (1936, p. 139) excessively praised it:

Some heart-breaking therapeutic disappointments in the history of tuberculosis and cancer would have been avoided if the method of Louis had been not merely praised, but generally used during the last fifty years.

Compilation of data does not contradict statistics; the numerical method has its place in science. True, as Gavarret (1840, p. x) remarked, it was not in itself scientific and was not based on *general philosophy*. D’Amador (1837, p. 12) wrongly attributed the numerical method to probability theory. The numerical method can be traced back to the 18th century (see below) and my description (§§ 10.8. 1 – 10.8.4) shows that it continued in existence for many decades. Furthermore, empiricism had been a feature of the Biometric school (§ 14.2). It originated with Anchersen (1741) when statisticians have begun to describe states in a tabular form (and thus facilitated the use of numbers), see § 6.2.1. Recall (§ 2.1.4), moreover, that Leibniz recommended compilation of *Staatstafeln*.

In statistics proper, Fourier’s *Recherches* (1821 – 1829) concerning Paris and the Département de la Seine almost exclusively consisted of statistical tables with data on demography, industry, commerce, agriculture and meteorology. True, empiricism was not sufficient even for compiling tables. Then, the abundance of materials led to the wrong idea that a mass of heterogeneous data was better than a small amount of reliable observations (§ 10.8.1).

10.8.1. Medicine. In 1835, Double et al (§ 8.5) indicated that statistics might be applied in medicine. Surgery occurred to be the first branch of medicine to justify their opinion. Already in 1839 there appeared a (not convincing) statistical study of the amputation of limbs. J. Y. Simpson (1847 – 1848/1871, p. 102) mistakenly attempted to obtain reliable results

by issuing from materials pertaining to several English hospitals during 1794 – 1839:

The data I have adduced [...] have been objected to on the ground that they are collected from too many different hospitals, and too many sources. But, [...] I believe all our highest statistical authorities will hold that this very circumstance renders them more, instead of less, trustworthy.

I ought to add, however, that Simpson (Ibidem, p. 93) stated that only a statistical investigation could estimate the ensuing danger. Soon afterwards 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 in § 10.5.

At about the same time Pirogov began to compare the merits of the conservative treatment of the wounded versus amputation. Much 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.

Farr's study of cattle plague of 1866 (Brownlee 1915) methodically belonged to epidemiology. Here is his reasoning. Denote the number of attacks of the plague during a period of four weeks by s and time by t . He noted that the third differences of $\ln s$ were constant, so that

$$s = C \exp\{\delta t[t + m]^2 + n\}, C > 0, \delta < 0.$$

It was Brownlee who supplied this formula because Farr was unable to insert it in his newspaper letter. Farr's calculated values of s did not agree with actual figures, but at least he correctly predicted a rapid decline of the epidemic. Enko (1889) provided the first mathematical model (of measles) in epidemiology proper (Dietz 1988).

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. Pettenkofer (1886 – 1887) published a monstrous collection of statistical materials pertaining to cholera, but was unable to process them. He (1865, p. 329) stressed that cholera epidemics were impossible at a certain moment without a local *disposition* to it which does not contradict modern ideas about necessary threshold values. However, he did not believe in contemporary bacteriological studies.

Seidel (1865 – 1866) investigated the dependence of the monthly cases of typhoid fever on the level of subsoil water, and then on both that level and the rainfall and quantitatively (although indirectly and with loss of information) estimated the significance of the studied connections.

Already Leibniz (§ 2.1.4) recommended to collect and apply information concerning a wide range of issues, which pertained to public hygiene. Condorcet (1795/1988, pp. 316 and 320) described the aims of *mathématique sociale* [political arithmetic] and mentioned the study of the influence of temperature, climate, properties of soil, food and general habits on the ratio of men and women, birth-rate, mortality and number of marriages. M. Lévy (1844) attempted to consider these causes.

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!

Pettenkofer (1873) estimated the financial loss of the population of Munich ensuing from such diseases as typhoid fever and his booklet can be attributed to this discipline.

10.8.2. Biology. The attempts to connect the appearance of leaves, flowers and fruits on plants of a given species with the sums of mean daily temperatures began in the 18th century (Réaumur 1738) and Quetelet (1846, p. 242) proposed to replace those sums by the sums of squares, but he was still unable to compare both procedures quantitatively. Also in the 19th century, vast statistical materials describing the life of plants were published (DeCandolle 1832), and Babbage (1857) compiled a statistical questionnaire for the class of mammalia. In Russia, Baer (1860 – 1875) with associates conducted a large-scale statistical investigation of fishing.

Humboldt created the geography of plants (Humboldt & Bonpland 1815; Humboldt 1816) which was based on collecting and estimating statistical data. Darwin had to study various statistical problems, for example on cross-fertilization of plants (§ 10.6), the life of earthworms (§ 11.2) and on the

inheritance of a rare deformity in humans (1868/1885, vol. 1, p. 449). Statistical tables and summaries with qualitative commentaries occur in a number of Darwin's writings and he also collected statistical data.

The stochastic essence of the evolution hypothesis was evident both for its partisans and the opponents; Boltzmann, however, was an exception (§ 10.8.5). I reconstruct now Darwin's model of evolution. 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 distances 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.

Darwin (1859/1958, p. 77) vividly described the difficulties of his hypothesis (and at the same time offered one of his differing explanations of randomness as the effect of complicated causes):

Throw up a handful of feathers and all fall to the ground according to definite laws; but how simple is the problem where each shall fall compared with problems in the evolution of species.

The main mathematical argument against Darwin's hypothesis was that a purposeful evolution under *uniform* randomness was impossible or at least demanded enormous time. Only Mendel's contributions (1866; letters of 1866 – 1873, published in 1905), forgotten until the beginning of the 20th century, and then the study of mutation allowed to answer such criticisms. Many objections and problems still remain, but Darwin had transformed biology as a science. In addition, his work was responsible for the appearance of the Biometric school (§ 14.2).

Mendel only applied the binomial distribution in an elementary way, but his memoir marked the origin of genetics and provided an example of a fundamental finding achieved by elementary means. His experiments became the object of discussions with regard to his subjective and objective honesty. Fisher (1936) and van der Waerden (1968) participated in the debates, and all doubts have possibly blown over the more so since Mendel's life and his meteorological observations and investigations testify in his favour. According to a communication from Prof. Walter Mann, a grandson of Mendel's nephew Alois Schindler, and the latter's report of 1902, Mendel was German. In 1945 – 1946 the descendants of his relatives were driven out of the then Czechoslovakia.

10.8.3. Meteorology. Humboldt (1818, p. 190) stressed the importance of studying the mean state of the atmosphere:

To discover the laws of nature [in meteorology] we ought to determine the mean state of the atmosphere and the constant type of its variations before examining the causes of the local perturbations.

In general, he (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 latter case he mentioned *the sole decisive method* [in natural sciences], *that of the mean numbers*. He himself (1817, p. 466) introduced isotherms and climatic belts (§ 1.3) and thus separated climatology from meteorology; he (1845 – 1862, Bd. 4, p. 59) had borrowed the idea of contour lines from Halley (§ 2.1.4). When defining climate, he (1831, p. 404) nevertheless had not directly linked it with mean states as later scholars did (Körber 1959, p. 296).

Köppen (1874, p. 3) believed that *the introduction of the arithmetic mean in meteorology was the most important step*, but that it was not sufficient all by itself. Indeed, Dove (1839, p. 285) formulated the aims of meteorology as the *determination of mean values* [of temperature], *derivation of the laws of [its] periodic changes and indication of rules for [determining its] irregular changes*. Later he (1850, p. 198) introduced monthly isotherms. Buys Ballot (1850, p. 629) stated that the study of deviations from mean values (mean states) constituted the second stage in the development of meteorology. He (1847, p. 108) noted that a similar process was going on in astronomy and in all sciences that did not admit experimentation.

Meteorological observations multiplied, and they had been published almost uselessly. Biot (1855, pp. 1179 – 1180) had opposed that practice and Mendeleev (1876/1946, p. 267) remarked that the prevailing *collecting* school of meteorologists needed nothing but *numbers and numbers*. Later he (1885/1952, p. 527) decided that a new meteorology was being born and that *little by little* it had begun, [still] basing its work on statistical data, to *master, synthesize, forecast*.

Lamont (1867, p. 247) maintained that the irregular temporal changes of the atmosphere were not random *in the sense of the calculus of probability* and (p. 245) recommended, instead, simultaneous observations made at different localities. Quetelet (1849, t. 1, Chapt. 4, p. 53) remarked that the differences of such observations conformed to accidental errors.

Lamarck occupied himself with physics, chemistry and meteorology. In meteorology, he is remembered for his *pioneer work in the study of weather* (Shaw & Austin 1926/1942, p. 130). He 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 (Ibidem, t. 4, pp. 153 – 154) 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. Quetelet (1846, p. 275) contended that meteorology was alien to statistics and cited other *alien* sciences, such as physical geography, mineralogy, botany. His statement was correct insofar as statistical meteorology, stellar statistics etc. belong to the appropriate sciences.

The study of densities of the distributions of meteorological elements began in the mid-19th century; Meyer (1891, p. 32), when mentioning that fact, stated that the theory of errors was not applicable to meteorology. However, K. Pearson (1898) made use of Meyer's material for illustrating his theory of asymmetric curves.

Lamarck (1800 – 1811) was one of the first scholars to note the dependence of the weather on its previous state, see for example t. 5, pp. 5 and 8 and t. 11, p. 143 of that source.

Quetelet (1852; 1853, p. 68; 1849 – 1857, 1857, pt 5, pp. 29 and 83) analysed lasting periods of fair or foul weather by applying elementary stochastic considerations and concluded that the chances of the weather persisting (or changing) were not independent. Köppen's analysis (1872) was more mathematically oriented. Quetelet also compiled and systematized meteorological observations. In many letters of 1841 – 1860 Faraday (1991 – 2008), see for example vol. 3, No. 1367 and vol. 4, No. 2263, praised Quetelet's observations of atmospheric electricity. In the first instance he wrote:

You are indeed a worthy example in activity & power to all workers in science and, if I cannot imitate your example, I can at least appreciate & value it.

10.8.4. Astronomy. Already Daniel Bernoulli (§ 6.1.1) and Laplace (§ 7.1-2) stochastically studied regularities in the Solar system. They actually considered planets as elements of a single population, and this approach was vividly revealed in the later investigations of the asteroids. Newcomb (1861a and elsewhere) compared the theoretical (calculated in accordance with the uniform distribution) and the actual parameters of the orbits of asteroids but was yet unable to appraise quantitatively his results. Concerning their distribution, he (1862; 1881) seems to have intuitively arrived at the following proposition: a large number of independent points $A_1 = (B_1 + b_1t)$, $A_2 = (B_2 + b_2t)$, ... where t denoted time, and the other magnitudes were constant, will become almost uniformly distributed over a circumference.

In 1881 Newcomb remarked that the first pages of logarithmic tables wore out *much faster* than the last ones and set out to derive the probability that the first significant digits of empirically obtained numbers will be n_1, n_2, \dots . Without any proof he indicated that, if numbers s_1, s_2, \dots, s_n were selected at random, the positive fractional parts of the differences $(s_1 - s_2)$, $(s_2 - s_3)$, ... will tend, as $n \rightarrow \infty$, to a uniform distribution over a circumference, and that the empirical magnitudes, to which these differences conform, will have equally probable mantissas of their logarithms. Newcomb's reasoning heuristically resembled the Weyl celebrated theorem that states that the terms of the sequence $\{nx\}$, where x is irrational, $n = 1, 2, \dots$, and the braces mean *drop the integral part*, are uniformly distributed on a unit interval. In the sense of the information theory, Newcomb's statement means that each empirical number tends to provide one and the same information. Several authors independently one from another proved that Newcomb was right. One of them called his statement *an inspired guess* but reasonably noted that it was not universally valid (Raimi 1976, p. 536).

By the mid-century, after processing observations made over about a century, a rough periodicity of the number of sunspots was established. Newcomb (1901), who studied their observations from 1610 onward, arrived at $T = 11.13$ years which did not, however, essentially differ from the previous results. The present-day figure is $T \approx 11$ years but a strict periodicity is denied. In any case, it might be thought that the numbers of

sunspots constitute a time series, an object for stochastic studies. I note that Newcomb considered the maxima and the minima of that phenomenon as well as half the sums of the numbers of the sunspots *corresponding to the year of minimum and the following maximum, or vice versa* (p. 4). He determined the four appropriate values of T and their mean without commenting on the possible dependence between them.

Variation of the terrestrial latitudes is known to be caused by the movement of the pole about some point along a curve resembling a circumference with period 1.2 years. Newcomb (1892) checked the then proposed hypothesis that the movement was periodic with $T = 1.17$ years. He assumed that the pole moved uniformly along a circumference. Some of his calculations are doubtful and in any case not sufficiently detailed (a feature peculiar to many of his works) but he correctly concluded that the hypothesis was [apparently] valid.

In 1767 Michell (§ 6.1.6) determined the probability that two stars were close to each other. By applying the Poisson distribution, Newcomb (1859 – 1861, vol. 2, 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 *random 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. See also Newcomb (1904a). He (1861b) also determined the probability of the distance between the poles of two great circles randomly situated on a sphere. Issuing from other initial considerations, Laplace (1812/1886, p. 261) and Cournot (1843, § 148) earlier provided solutions differing both from each other and from Newcomb’s answer (Sheynin 1984a, pp. 166 – 167).

About 1784 William Herschel started counting the number of stars situated in different regions of the sky. He thought that his telescope was able to penetrate right up to the boundaries of the (finite) universe and hoped to determine its configuration. In one section of the Milky Way he (1784/1912, p. 158) counted the stars in six fields selected *promiscuously* and assumed the mean number of them as an estimate for the entire section. Later Herschel (1817) proposed a model of a uniform spatial distribution of the stars. He fixed the boundaries for the distances of the stars of each magnitude but allowed the stars to be randomly distributed within these boundaries and thus provided an example of randomness appearing alongside necessity, cf. Poincaré’s statement in § 1.1.

When estimating the precision of his model for the stars of the first seven magnitudes, Herschel calculated the sum of the deviations of his model from reality. For the first four magnitudes the sum was small although the separate deviations were large. Recall (§ 6.3.2-3) that, when adjusting

observations, Boscovich applied a similar test with respect to absolute deviations and that Herschel independently (1805) made use of it when determining the Sun's movement (again § 6.3.2-3).

Herschel (1817/1912, p. 579) wrongly indicated that *any star promiscuously chosen [...] out of [14,000 stars of the first seven magnitudes] is not likely to differ much from a certain mean size of them all*. With regard to size, the stars are incredibly different; that mean value is a worthless quantity, and, in general, stochastic statements, made in the absence of data, are hardly useful. However, it occurred that the stars, even earlier than the asteroids, had been considered as elements of a single population (in the last-mentioned instance, wrongly).

Stellar statistics really originated in the mid-19th century with the study of the proper motions of hundreds of stars (until 1842, when astronomers started to use the Doppler's invention, only in the directions perpendicular to the appropriate lines of sight). The calculated mean proper motions for stars of a given magnitude proved, however, almost meaningless since magnitudes depended on distances. Beginning with W. Herschel, astronomers thought that the proper motions were random, but they understood randomness in different ways. Newcomb (1902) assumed that their projections on an arbitrary axis were normally distributed. He derived, although without providing any calculations, the density laws of their projections on an arbitrary plane and their own distribution. Both were connected with the χ^2 distribution.

The general statistical study of the starry heaven became more important than a precise determination of the parameters of some star (Hill & Elkin 1884, p. 191):

The great Cosmical questions to be answered are not so much what is the precise parallax of this or that particular star, but – What are the average parallaxes of those of the first, second, third and fourth magnitude respectively, compared with those of lesser magnitude? [And] What connection does there subsist between the parallax of a star and the amount and direction of its proper motion or can it be proved that there is no such connection or relation?

Then, Kapteyn (1906b; 1909) described a stochastic picture of the stellar universe by the laws of distribution of the (random!) parameters, parallaxes and peculiar motions, of the stars. He (1906a) also initiated the study of the starry heaven by [stratified] sampling; here is a passage from a letter that he received in 1904 on this subject from one of his colleagues and inserted on his p. 67:

As in making a contour map, we might take the height of points at the corners of squares a hundred meters on a side, but we should also take the top of each hill, the bottom of each lake, [...], and other distinctive points.

In statistics, sampling became recognized at about the same time, although not without serious resistance (You Poh Seng 1951) and its most active partisan was Kiaer, also see § 10.7-2.

The compilation of vast numerical materials (catalogues, yearbooks) was also of a statistical nature. Sometimes this direction of work had been

contrasted to theoretical constructions. Thus, Proctor (1872) plotted 324 thousand stars on his charts attempting to leave aside any theories on the structure of the stellar system, but the development of astronomy proved him wrong.

Calculation and adjustment of observations, their reasonable comparison has always been important for astronomy. Here, I again ought to mention, in the first place, Newcomb. Benjamin (1910) and many other commentators stated that he had to process more than 62 thousand observations of the Sun and the planets and that his work included a complete revision of the constants of astronomy. I add that he discussed and compared observations obtained at the main observatories of the world and that he hardly had any aids except for logarithmic tables. In addition, he published some pertinent theoretical studies. He was of course unable to avoid the perennial problem of the deviating observations. At first he regarded them with suspicion, then (1895, p.186), however, became more tolerant. If a series of observations did not obey the normal law, Newcomb (1896, p. 43) preferred to assign a smaller weight to the *remote* observations, or, in case of asymmetric series, to choose the median instead of the arithmetic mean. He had not mentioned Cournot (§ 10.3-3), and, in two memoirs published at the same time, he (1897a; 1897b) called the median by two (!) other, nowadays forgotten, terms.

Mendeleev (§ 10.9.3) objected to combining different summaries of observations; Newcomb, however, had to do it repeatedly, and in such cases he (1872), hardly managing without subjective considerations, assigned weights to individual astronomical catalogues depending on their systematic errors. Interestingly enough, he then repeated such adjustments with weights, depending on random errors.

After determining that the normal law cannot describe some astronomical observations made under unfavourable conditions, Newcomb (1886) proposed for them (and, mistakenly, for all astronomical observations altogether) a generalized law, a mixture of normal laws with differing measures of precision occurring with certain probabilities. The measure of precision thus became a discrete random variable, and the parameters of the proposed density had to be selected subjectively. He noted that his density led to the choice of a generalized arithmetic mean with weights decreasing towards the *tails* of the variational series which was hardly better than the ordinary arithmetic mean (§ 6.3.1).

He had also introduced some simplifications, and later authors noted that they led to the choice of the location parameter by the principle of maximum likelihood. Newcomb hardly knew that his mixture of normal laws was not normal (Eddington 1933, p. 277). In turn, two authors generalized Newcomb's law, but their work was of little practical importance.

Like Mendeleev (§ 10.9.3), Newcomb (1897b, p. 165) thought that the discrepancy between two empirical magnitudes was essential if it exceeded the sum of the two appropriate probable errors, and it seems that this rigid test had been widely accepted in natural sciences. Here is Markov's relevant pronouncement from a rare source (Sheynin 1990b, pp. 453 – 454): he

Like[d] very much Bredikhin's rule according to which 'in order to admit the reality of a computed quantity, it should at least twice numerically

exceed its probable error'. I do [he does] not know, however, who established this rule or whether all experienced calculators recognized it.

In other words, the difference between zero and a *real* non-zero quantity must twice exceed its probable error, a statement that conformed to Mendeleev's and Newcomb's opinion. But still, Newcomb several times indicated that some quantity a determined by him had mean square error b even when the latter much exceeded the former including the case (1901, p. 9) of $a = 0.05$ and $b = 0.92$!

Repeatedly applying the MLSq, Newcomb sometimes deviated from strict rules; cf. my comment in § 9.1.5. In another case he (1895, p. 52) thought that small coefficients in a system of normal equations might be neglected, but he had not provided any quantitative test. Newcomb realized that, when forming normal equations, the propagation of round-off errors could result in their interdependence, and he reasonably concluded that in such cases the calculations should be made with twice as many significant digits. This is what he (1867) did when studying the calculations of the Kazan astronomer Kowalski, who had noted that, out of the four normal equations which he formed, only two were independent. It is now known that ill-conditioned observational equations should rather be processed without forming normal equations, – for example, by successive approximations.

Newcomb's calculation (1874, p. 167) presents a special case. Having 89 observational equations in five unknowns, he formed and solved the normal equations. Then, however, he calculated the residual free terms of the initial equations and somehow solved them anew (providing only the results of both solutions). He apparently wished to exclude systematic influences as much as possible, but how?

Newcomb (1895, p. 82; 1897b, p. 161) mistakenly stated, although mentioning earlier the definitive Gaussian justification of the MLSq, that the method was inseparable from the normal law. I note also his unfortunate reasoning (Newcomb & Holden 1874, p. 270) similar to the one made by Clausius (§ 10.8.5): for systematic error s and random errors r_1 and r_2 , as he went on to prove, and only for the normal law, by considering the appropriate double integral, that

$$E[(s + r_1)(s + r_2)] = s^2.$$

Newcomb necessarily remained more or less within the boundaries of the classical theory of errors and simple stochastic patterns but the extant correspondence between him and K. Pearson during 1903 – 1907 (Sheynin 2002b, § 7.1) testifies that he wished to master the then originating mathematical statistics. Here is a passage from his letter of 1903 to Pearson:

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.

I mention finally Newcomb's statistical contribution (1904b) in which he examined the classical problem of the sex ratio at birth (see §§ 2.2.4, 3.3.4, 4.4 and 6.1.1). He assumed that there existed three kinds of families numbered, say, m , n , and n , for whom the probabilities of the birth of a boy

were p , $p + \alpha$ and $p - \alpha$ respectively and he studied, in the first place, the births of twins. The sex of the embryo, as he thought, became established only after the action of a number of successive causes made it ever more probable in either sense.

10.8.5. Physics. 1) The kinetic theory of gases originated in mid-19th century as the result of the penetration of the statistical method into physics. Truesdell (1975) discussed its early history; thus (p. 28), it was Waterson who, in 1843, introduced the mean free path of a molecule, but his innovation was not published. Clausius likely published the first memoir (1849) belonging to physics (but did not deal with the molecular hypothesis) and contained ideas and methods of the theory of probability.

After Poisson's death that theory sank into oblivion. No wonder that Clausius (1889 – 1891, p. 71) made a point to prove the equality $E(\xi/E\xi) = 1$ for the velocity ξ of a molecule and Boltzmann (1896/1909, p. 570) stated that the normal law followed from equal probabilities of positive and negative elementary errors of the same absolute value. His was of course an unworthy formulation of the CLT.

I ought to add that Boltzmann respected the theory of probability. Thus (1872/1909, p. 317),

An incompletely proved theorem whose correctness is questionable should not be confused with completely proved propositions of the theory of probability. Like the results of any other calculus, the latter show necessary inferences made from some premises.

And again (1895/1909, p. 540): *the theory of probability is as exact as any other mathematical theory if, however, the concept of equal probabilities, which cannot be determined from the other fundamental notions, is assumed.*

Maxwell twice mentioned Laplace (Sheynin 1985, pp. 364 and 366n), although without providing any definite references, whereas Boltzmann, who cited many scholars and philosophers in his popular writings, never recalled him. Khinchin (1943/1949, p. 2) maintained that Maxwell and Boltzmann applied *fairly vague and somewhat timid probabilistic arguments*, that in their work

The notions of the theory of probability do not appear in a precise form and are not free from a certain amount of confusion which often discredits the mathematical arguments by making them either void of any content or even definitely incorrect. The limit theorems [...] do not find any applications [...].

His statement seems too harsh. First, I believe that it was partly occasioned by Boltzmann's verbose style of writing. Second, physicists certainly applied the LLN indirectly. Third, Khinchin said nothing about positive results achieved in physics (formulation of the ergodic hypothesis, use of infinite general populations, Maxwell's indirect reasoning about randomness). My first remark is indeed essential; here is an extract from Maxwell's letter of 1873 (Knott 1911, p. 114):

By the study of Boltzmann I have been unable to understand him. He could not understand me on account of my shortness, and his length was and is an equal stumbling block to me.

And Boltzmann (1868/1909, p. 49) indeed owned that it was difficult to understand Maxwell's *Deduktion* (1867) *because of its extreme brevity.*

2) Clausius. He (1857/1867, pp. 238 and 248) asserted that molecules moved with essentially differing velocities. Even Boscovich (1758, § 481) stated something similar but perhaps presumed that the differences between these velocities were not large: The *points* [atoms] of a *particle* [of light, as in § 477, or of any body, as in § 478] move *together with practically the same velocity*, and the entire particle will *move as a whole with the single motion that is induced by the sum* [the mean] *of the inequalities pertaining to all its points.* Clausius used a single mean velocity such as to make the entire kinetic energy of a gas equal to its actual value. Later he (1862/1867, p. 320) maintained that the velocities of molecules randomly differed one from another.

And he (1858/1867, p. 268) studied the length of the free path of a molecule. Denote the probability of a unit free path by a , then

$$W = a^x = (e^{-x})^\alpha, \alpha > 0$$

will be the probability of its being equal to x ; here, α is derived from the molecular constants of the substance. Similar considerations are in other works of Clausius (1862/1867, § 29; 1889 – 1891, pp. 70 – 71 and 119). He (1889 – 1891, pp. 70 – 71) also calculated the mean free path of a molecule. Actually, without writing it out, he considered free paths of random length ξ and calculated the expected free path as an integral over all of its possible values from 0 to ∞ .

Suppose now that

$$\xi = \xi_1 + \xi_2 + \dots + \xi_m$$

where m is an arbitrary natural number. Then, according to Clausius' assumptions, ξ_k , $k = 1, 2, \dots, m$, will not depend on $(\xi_1 + \xi_2 + \dots + \xi_{k-1})$ and the characteristic function for ξ_k will be equal to the product of these functions for the previous ξ 's. In this instance, all these functions are identical, and $F(s)$, the integral distribution function of ξ , is therefore infinitely divisible. Clausius' achievements were interesting, but he did not attempt to construct the kinetic theory of gases on a stochastic basis.

3) Maxwell (1860) established his celebrated distribution of the velocities of monatomic molecules

$$\varphi(x) = \frac{1}{\alpha\sqrt{\pi}} \exp(-x^2 / \alpha^2).$$

He tacitly assumed that the components of the velocity were independent; later this restriction was weakened (Kac 1939; Linnik 1952). He then maintained that the average number of particles with velocities within the interval $[v; v + dv]$ was proportional to

$$f(x) = \frac{4}{\alpha^3 \sqrt{\pi}} v^2 \exp(-v^2 / \alpha^2) dv.$$

This can be justified by noting that the probability of such velocities can also be represented as

$$\int_0^{2\pi} d\varphi \int_0^{\pi} \sin \theta d\theta \int_v^{v+dv} t^2 \exp(-t^2 / \alpha^2) dt.$$

It is presumed here that the components of the velocity in each of the three dimensions have the same distribution.

Maxwell left interesting statements about the statistical method in general, and here is one of them (1873b/1890, p. 374):

We meet with a new kind of regularity, the regularity of averages, which we can depend upon quite sufficiently for all practical purposes, but which can make no claim to that character of absolute precision which belongs to the laws of abstract dynamics.

The drafts of the source just mentioned (Maxwell 1990 – 2002, 1995, pp. 922 – 933) include a previously unpublished and very interesting statement (p. 930): abandoning the *strict dynamical method* and adopting instead the statistical method *is a step the philosophical importance of which cannot be overestimated.*

And here is his definition (not quite formal) of the statistical method which heuristically resembles the formulation provided by Kolmogorov & Prokhorov (§ 0.2): it consisted in *estimating the average condition of a group of atoms* (1871/1890, p. 253), in studying *the probable* [not the average!] *number of bodies in each group* under investigation (1877, p. 242).

Maxwell gave indirect thought to randomness. Here is his first pronouncement (Maxwell 1859/1890, pp. 295 – 296) which was contained in his manuscript of 1856 (1990 – 2002, 1990, p. 445), and certainly describes his opinion about that phenomenon:

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 derangement of the motion [...].

Maxwell (1873a, p. 13) later noted that in some cases *a small initial variation may produce a very great change* [...]. Elsewhere he (report read 1873, see Campbell & Garnett 1882/1969, p. 440) explained that in such instances the condition of the system was unstable and prediction of future events becomes impossible. He (Ibidem, p. 442) provided an example of instability of a ray within a biaxial crystal and prophetically stated (p. 444) that in future physicists will study *singularities and instabilities*. I note that in 1873 – 1882 Engels (1925/1971, p. 213) urged scientists to study both necessity and chance.

In a manuscript of the same year (1873) Maxwell (Campbell & Garnett, p. 360), remarked that

The form and dimensions of the orbits of the planets [...] are not determined by any law of nature, but depend upon a particular collocation of matter. The same is the case with respect to the size of the earth.

This was an example illustrating Poincaré's statement concerning randomness and necessity (§ 1.1), but it was not sufficiently specific; the eccentricities of planetary orbits depend on the velocities of the planets, cf. end of § 7.3.

And here is Maxwell's position (1875/1890, p. 436) concerning randomness in the atomic world:

The peculiarity of the motion of heat is that it is perfectly irregular; [...] the direction and magnitude of the velocity of a molecule at a given time cannot be expressed as depending on the present position of the molecule and the time.

At the very end of his life Maxwell (1879/1890, pp. 715 and 721) introduced a definition for the probability of a certain state of a system of material particles:

I have found it convenient, instead of considering one system of [...] particles, to consider a large number of systems similar to each other [...]. In the statistical investigation of the motion, we confine our attention to the number of these systems which at a given time are in a phase such that the variables which define it lie within given limits.

Boltzmann (1868, § 3) defines the probability of the system being in a phase [...] as the ratio of the aggregate time during which it is in that phase to the whole time of the motion.

4) If the classical definition of probability is included here, we can say that Boltzmann used three formulations. Maxwell (item 2 above) mentioned one of them, and another reference can be added: Boltzmann (1895 – 1899, 1895, Bd. 1, p. 50). Yet another one was that applied by Maxwell (see same subsection) although sometimes Boltzmann (1878/1909, p. 252) did not indicate which one he was employing. He (1872/1909, p. 317) apparently thought that these posterior probabilities were equivalent.

In other words, with respect to separate molecules Boltzmann introduced the time average probability, – and maintained that it was equivalent to the usual phase average probability. When studying polyatomic gases, Boltzmann (1871) defined the probability of its state as a product such as $f d\omega$ where f was some function, varying in time, of the coordinates and velocities of the separate molecules and $d\omega$, the product of the differentials of those parameters. For stochastic processes, such functions determine the distribution of a system of random variables at the appropriate moment. Zermelo (1900, p. 318) and then Langevin (1913/1914, p. 3) independently stressed the demand to provide a *definition correcte et claire de la probabilité* (Langevin). Like Maxwell, Boltzmann (1887/1909, p. 264; 1895

– 1899, 1899, Bd. 2, p. 144) used the concepts of fictitious physical systems and infinite general population.

From 1871 onward Boltzmann had been connecting the proof of the second law of thermodynamics with stochastic considerations; however, he (1886/1905, p. 28) then indicated that the 19th century will be the age of *mechanical perception of nature, the age of Darwin*, and (1904a/1905, p. 368) that the theory of evolution was understandable in mechanical terms, that (1904b, p. 136) it will perhaps become possible to describe electricity and heat mechanically. The possible reason for his viewpoint was that he did not recognize objective randomness. Another reason valid for any scholar was of course the wish to keep to *abstract dynamics*, see Maxwell's statement on the *new kind of regularity* (item 3 above) and the opinion of Hertz (1894, Vorwort): *Physicists are unanimous in that the aim of physics is to reduce the phenomena of nature to the simple laws of mechanics*. And here is a lucid description of this point as far as Boltzmann was considered (Rubanovsky 1934, p. 6): in his works

Randomness [...] struggles with mechanics. Mechanical philosophy is still able [...] to overcome randomness and wins a Pyrrhic victory over it but recedes undergoing a complete ideological retreat.

10.9. Natural scientists

10.9.1. Ivory. In 1825 – 1830 Ivory published 11 papers devoted to the derivation of the flattening of the Earth's ellipsoid of rotation by means of pendulum observations. In a letter of 1827 to Olbers, Gauss (W/Erg-4, Tl. 2, pp. 475 – 476) called Ivory an *acute* mathematician, but indicated that the *spirit* of the MLSq was alien to him.

Ivory was ignorant of the MLSq, called it not good enough but applied it, perhaps not even realizing it at once and had not applied the variance. Then, having at his disposal 5 – 7 observations, only one of which was made at a southern station, he (1826a, p. 9) combined it with each of the others (to have pairs with a large latitudinal difference) and calculated the flattening from the thus obtained pairs. The weight of the equatorial observation became absurdly great and its error corrupted all the pairs in the same way. Then, before an adjustment, stations having almost the same latitude can be combined to form a single mean station, which Ivory did not do.

Because of local anomalies of gravity Ivory (1826b, p. 242) rejected up to 31% of the available observations. His final result (1828, p. 242) was, however, sufficiently close to the flattening of the Krasovsky ellipsoid (§ 6.3). Ivory actually wished to solve two problems at once: to find out whether the observations were consistent with an ellipsoidal Earth, and to adjust them. The minimax method (§ 6.3.2-4) is best for solving the first problem.

10.9.2. Fechner. He (1860) was the founder of psychophysics and became one of the first to introduce the statistical method, although not in the crucial direction, into physics. He (1860, Bd. 1, p. 8, see also 1877, p. 213) defined that discipline as an *exact doctrine on the functional correspondence or interdependence of body and soul*. According to modern understanding, it is a study of quantitative relations between sensations and the stimuli that produce them.

Fechner (1855 and 1864) did not comment on the developing kinetic theory of gases. His mathematical tools and approach were primitive and almost everything he achieved had to be repeated at a higher level. Ebbinghaus (1908, p. 11) called Fechner *a philosopher full of fantasies but a most strict physicist* who had *put [...] together psychophysics as a new branch of knowledge*. Being the co-author of the logarithmic Weber – Fechner law connecting stimuli with sensations, Fechner extended the range of its application by experiments (1860; 1887). He studied the methods of experimentation and the modern method of paired comparisons (H. A. David 1963) owes much to him.

In the theory of errors Fechner attempted, sometimes unsuccessfully, to introduce innovations, or to repeat unknown to him previous findings and he somewhat furthered that theory. His main innovation was the collective, – the set of observed values of a random variable. He (1897) proposed to study them by applying several mean values, their mutual arrangement, and their deviations (including absolute and normed deviations) from the observations. He paid attention to asymmetric collectives and attempted to discover a universal asymmetric distribution for errors in natural sciences.

Fechner (1897, pp. 365 – 366) also studied the interdependence of the successive daily air temperatures by comparing their course with the arrangement of winning (numbered) tickets of a reputed lottery and achieved an interesting result pertaining to the runs up and down (cf. § 10.2-4). He even introduced a measure of dependence varying from 0 to 1, but describing only *positive* dependences. His contribution appeared posthumously, after the Galton correlation theory had emerged.

Mises (1928/1972, pp. 26 and 99) highly appraised Fechner's efforts and stated (p. 99) that Fechner's *constructions prompted, at least me [Mises], to adopt a new viewpoint*. K. Pearson (1905, p. 189) called him a leading statistician and Freud (1925/1963, p. 86) *followed that thinker upon many important points*.

10.9.3. Mendeleev. From 1893 to 1907 Mendeleev was Director of Russia's Main Board of Measures and Weights and processed observations both as a chemist and a metrologist. He (1872b/1951, p. 101) distrusted data *obtained under differing conditions, by different methods and observers* as compared with those *achieved by precise methods and experienced persons*. Then (1887/1934, p. 82), *disadvantageous* data ought to be rejected, otherwise *a realistic result* is impossible to get.

No wonder that he (1872a/1939, p. 144) preferred *to make a few but precise and repeated measurements* and objected to amassing observations; true, this attitude was partly due to his wish to avoid calculations, cf. Boyle's statement in § 1.7. Mendeleev (1875b/1950, p. 209) thought that an observational series should be *harmonious* so that its median should coincide with its arithmetic mean, or that the mean of its middlemost third should coincide with the mean of the means of its extreme thirds. In the first case, he mistakenly added that the coincidence meant that the appropriate distribution was normal. He had not said how to treat observations which did not obey his wish.

The deviation of the arithmetic mean from the median, normed in a certain way, is nowadays recognized as a measure of asymmetry of the appropriate distribution (Yule & Kendall 1937/1958, p. 161). Mendeleev had not mentioned the second Gaussian justification of the MLSq and made

a few mistakes in his theoretical considerations. One of them was an excessive belief in the arithmetic mean (1856/1937, p. 181; 1877/1949, p. 156; 1895/1950, p. 159).

11. Bertrand and Poincaré

Bertrand criticized everything, often mistakenly, but he turned attention to probability, and especially to the concept of uniform randomness. Poincaré achieved interesting results in geometric probability and in interpreting randomness. He referred almost exclusively to Bertrand, never to Chebyshev or Markov, and expressed strange ideas about the MLSq.

Key words: random chord, randomness in nature, geometric probability, justification of the MLSq

11.1. Bertrand: General Information

In 1855 Bertrand had translated Gauss' works on the MLSq into French, but his own work on probability began in essence in 1887 – 1888 when he published 25 notes and his treatise (1888a), written in great haste and carelessly and lacking a systematic description of its subject but in a very good literary style. Gauss died the same year (1855) and was only able to send *observations about details* (Bertrand, *C. r. Acad. Sci. Paris*, t. 40, 1855, p. 1190). Gauss is known to have refused to publish in French, – but apparently did not object to being translated into that language.

1) Statistical probability and the Bayesian approach. Heads appeared $m = 500,391$ times in $n = 10^6$ tosses of a coin (p. 276; here and below I only provide the page number of the treatise). Nonsense followed: the unreliable statistical probability of that event is $p = 0.500391$, 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 (cf. the problem about the sunrise in § 5.1). 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. § 5.2.

2) Statistics of population. Bertrand indicated that there existed a dependence between trials (or their series) and that the probabilities of the studied events could change. He referred only to Dormoy (§ 14.1.1) and had not provided any concrete examples, but he (p. 312) noted that, when studying the sex ratio at birth, both Laplace and Poisson had assumed without justification that the probability of a male birth was constant in time and space. Yes, but their mistake was only methodological since they could not have failed to understand this circumstance, cf. § 7.1-5.

3) Bertrand paid much attention to the mathematical treatment of observations, but his reasoning was amateurish and sometimes wrong. Thus, he (pp. 281 – 282) attempted to prove that the sample variance (9.6b) might be replaced by another estimator of precision having a smaller variance but failed to notice that, unlike the Gauss statistic, his new estimator was

biased. 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.

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

a) He derived the most probable composition of the urn (pp. 152 – 153) filled with balls of two colours given a sample of extracted balls.

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). For large values of s and n Bertrand obtained an elegant formula

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

He (1887) published this formula earlier without justification and noted that that variable probability was *en quelque sorte un régulateur*.

c) An urn contains m balls favourable for candidate A, and n balls favouring B ($m > n$). The balls are extracted one by one without replacement. Then, the probability P that A was always ahead of B (p. 18) was equal to $P = (m - n)/(m + n)$. This *ballot problem* has many applications. Takácz (1967, pp. 2 – 3; 1982/2006) traced its history back to De Moivre (§ 4.1-5); he himself, in 1960, had generalized it.

d) I select one out of the few problems on the gambler's ruin (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 solved this problem by applying his previous result. Calculate the probability that A loses $(n + m)/2$ games and wins $(n - m)/2$ times; then multiply it by the probability that during that time A will never have more than m counters, that is, by m/n .

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

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

Bertrand's treatise is impregnated with its non-constructive negative (and often unjustified) attitude towards the theory of probability and treatment of observations and wrong statements. Thus, he pp. (325 – 326) alleged that Cournot (cf. § 10.3-5) had supposed that judges decided their cases independently one from another. Nevertheless, he exerted a strong influence upon Poincaré (a too strong influence!), and, its spirit and inattention to Laplace and Bienaymé notwithstanding, on the revival of the interest of French scientists in probability (Bru & Jongmans 2001).

11.2. Bertrand: The Random Chord

By several examples Bertrand proved that the expression *at random*, or even *uniformly random*, was not definite enough. Thus, he maintained that the Michell problem (§ 6.1.6) should have been generalized: remarkable was not only a small distance between stars, but some other features of their mutual arrangement as well. One of his examples (p. 4) became classical. Determine the probability, Bertrand asked, that a randomly drawn chord of a given circle was longer than the side of an equilateral triangle inscribed in the circle. He listed three possible answers:

- a) One endpoint of the chord is fixed; $p = 1/3$.
- b) The chord's direction is fixed; $p = 1/2$.
- c) The location of the centre of the chord in any point of the circle is equally probable; $p = 1/4$.

A curious statement about this problem is due to Darboux (1902/1912, p. 50):

In accord with considerations which seem equally plausible, he [Bertrand] derived two different values for the probability sought, 1/2 and 1/3. He investigated this question and found its solution, but left its discovery to the readers.

In failing to mention the third solution he possibly followed Poincaré, see below.

Poincaré (1896, p. 97; 1912, p. 118) considered the Bertrand problem. Choosing two differing pairs of parameters (call them ω , α and ρ , θ), each defining the random chord, he noted that the integrals of $d\omega d\alpha$ and $d\rho d\theta$ over the given circle were not equal to each other, which as Poincaré stated, explained the paradoxical nature of the problem.

Czuber (1903/1968, pp. 107 – 108) discovered three more natural solutions of the Bertrand problem, one of them coinciding with Bertrand's first version. The other two were $p = 1/3 + \sqrt{3}/2\pi \approx 0.609$ and $1/3 + 3\sqrt{3}/4\pi \approx 0.746$.

The Bertrand problem has an uncountable set of answers (De Montessus 1903). Suppose that a certain diameter of the given unit circumference with centre O is the x -axis and mark points D and C on its positive half, – its intersections with concentric circumferences with common centre in point O and radii $OD = 1/2$ and $OC = 1$. Each point from D to infinity can indeed belong to a chord (or its extension defining the chord) satisfying the condition of the problem. De Montessus also noted that the mean value of the probability was $1/2$.

Schmidt (1926) issued from Poincaré's considerations and indicated in addition that the probability sought should persist under translation and rotation of the coordinate system (invariance under change of scale is also needed). Accordingly, he proved that this condition is only satisfied for a certain polar coordinate system and when transforming it into another one (with the appropriate Jacobian certainly allowed for).

He also showed that the proper solution corresponded to choosing that system of coordinates with origin at the centre of the circle and fixing the chord by the coordinates of its centre. The probability was then $p = 1/2$, cf. De Montessus' study, a value gradually accepted by later commentators, which can be understood as complete ignorance (§ 8.1.4)! See Poisson's calculation of the probability of the unknown composition of an urn and

especially my example concerning the unknown probability of the outcomes of a coin toss in the same subsection.

I add a few words about geometric probability in the 19th century before Bertrand. Cournot (1843, § 74) applied it for deriving the distribution of a function of several random arguments. Here is one of his examples. The arguments of the function $u = |x - y|$ are uniformly distributed on segment $[0; 1]$. After calculating the areas of the appropriate figures, he concluded that

$$P(u \geq a) = (1 - a^2), 0 \leq a \leq 1.$$

The determination of the probability of the contrary event would have led him to the once popular encounter problem (Laurent 1873, pp. 67 – 69): two persons are to meet at a definite spot during a specified time interval, their arrivals are independent and occur *at random*. The first one to arrive waits for a certain time and then leaves. Determine the probability of the encounter.

Most eminent natural scientists of the 19th century tacitly applied geometric probability, for example Boltzmann (§ 10.8.5) and Darwin (1881/1945, pp. 52 – 55) who found out that earthworms did not seize by chance any point of the perimeter of paper triangles when carrying them off to their burrows.

Seneta et al (2001) described the investigations of geometric probability by Sylvester, Crofton and Barbier which led to the appearance of integral geometry. I mention Sylvester's remarkable problem: To determine the probability that four points taken *at random* within a finite convex domain will form a convex quadrilateral. See Czuber (1903/1968, pp. 99 – 102) for a few particular cases of that problem.

For a modern viewpoint on geometric probability see M. G. Kendall & Moran (1963). Then, Ambartzumian (1999) indicated that geometric probability and integral geometry were connected with stochastic geometry.

11.3. Poincaré

Poincaré (1896/1912) had passed over in silence not only the Russian mathematicians, but even Laplace and Poisson, and 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). Also see § 11.2 (Poincaré's contribution to the celebrated Bertrand problem).

Commenting on the first edition of his treatise, Bortkiewicz (Bortkevich & Chuprov 2005, Letter 19 of 1897) noted:

The excessively respectful attitude towards [...] Bertrand is surprising. No traces of a special acquaintance with the literature on probability are seen. The course is written in such a way as though Laplace and Poisson, especially the latter, never lived.

Several times Poincaré applied the formula

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

where $\Phi(x)$ was a restricted positive function, x_0 , 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 its proof and some restrictions should perhaps be added. To place Poincaré's trick in the proper perspective, see Erdélyi (1956, pp. 56 – 57). I discuss now some 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 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 and that the calculus of probability will evidently ruin them (*les entrainerait évidemment dans sa ruine*). He concluded that the calculus was only of practical importance. Then he (1896/1912, p. 34) apparently maintained that a mathematician was unable to understand why forecasts concerning mortality come true.

In a letter of ca. 1899 connected with the notorious Dreyfus case (*Le procès* 1900, t. 3, p. 325; Sheynin 1991a, pp. 166 – 167) Poincaré followed Mill (§ 8.3) and even generalized him 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 its author.

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 could account for all the known facts. Later 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) The Bayesian approach: estimating the total number (N) of the asteroids. Poincaré (pp. 163 – 168) assumed that only M of them were known and that, during a certain year, n minor planets were observed, m of which were known before. Introducing a constant probability $p = n/N$ of

observing an asteroid during a year and applying the Bayesian approach, he obtained $EN \approx n/p$. He was not satisfied with this pseudo-answer and assumed now that p was unknown. Again applying the Bayesian approach and supposing that p took with equal probability all values within the interval $[0; 1]$, he derived instead $EN = (M/m)n$.

He could have written this formula at once; in addition, it was possible to recall the Laplace problem of estimating the population of France by sample data (§ 7.1-5). It is nevertheless interesting that Poincaré considered the unknown number of the minor planets as a random variable.

5) Without mentioning Gauss (1816, § 5), he (pp. 192 – 194) derived the moments of the normal distribution and proved that the density function whose moments coincided with the respective moments of the normal law was normal. This proposition was due to Chebyshev (1887a), see also Bernstein (1945/1964, p. 420).

Poincaré applied his investigation to the theory of errors and non-rigorously proved the CLT: for errors of *sensiblement* the same order and constituting *une faible part* of the total error, the resulting error followed *sensiblement* the Gauss law (p. 206).

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

$$f(\alpha) = \sum p_x e^{\alpha x}, f(\alpha) = \int \varphi(x) e^{\alpha x} dx. \text{ He noted that } f(\alpha) = 1 + \alpha Ex/1! + \alpha^2 Ex^2/2! + \dots (1; 2)$$

6) Homogeneous Markov chains. Poincaré provided interesting examples that might be interpreted in the language of these chains and their ergodic properties.

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

$$Ee^{iml} = \iint \varphi(a; b) e^{im(at+b)} da db$$

(which is the appropriate characteristic function in the modern sense), Poincaré not very clearly proved his proposition that resembled the celebrated Weyl theorem (beginning of § 10.8.4). 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 might 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 having made 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 $[2\pi; A]$ and having 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 substantiation on the method of arbitrary functions (Khinchin 1961/2004, pp. 421 – 422; von Plato 1983) and himself sketched its essence.

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. See § 7.1-6.

7) 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. In his treatise he (pp. 169 – 173) derived the normal distribution of observational errors mainly following Gauss; then, like Bertrand, 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, \dots, 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.

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 $n \rightarrow \infty$). Nothing, however, followed since other linear means had the same property, as Markov (1899a/1951, p. 250) stated on another occasion. Poincaré himself (1896/1912, 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 (1) and (2) and passed on to the characteristic function of the arithmetic mean. He noted that, if that function could not be represented as (2), 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 (which directly contradicted Gauss' mature approach). In the same context Poincaré (p. 171) argued that everyone believed that the normal law was universal: experimentalists thought that that was a mathematical fact and mathematicians believed that it was experimental.

8) Randomness. Poincaré discussed randomness both in his treatise and in his scientific-popular booklets. In § 1.1 I noted his statement about the link between randomness and necessity. There also, is a description of chaotic processes, and two of his explanations of chance. Maxwell (§ 10.8.5-3) anticipated one of these, but did not mention chance.

I would argue that Poincaré initiated modern studies of randomness. For him, the theory of probability remained an accessory subject, and his almost total failure to refer to his predecessors except Bertrand testifies that he was

not duly acquainted with their work. However, his treatise had for about 20 years 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: at the time, Markov was apparently the only one who did master probability.

12. Chebyshev

Chebyshev proved his version of the LLN, almost rigorously justified the CLT and greatly influenced Russian scholars. His failure to recognize contemporary Western developments hampered Russian mathematicians.

Key words: LLN, CLT

12.1. His Contributions

1) The Poisson LLN (Chebyshev 1846); see Prokhorov (1986) for a detailed exposition. Chebyshev solved the following problem. In n [independent] trials the probability of success was p_1, p_2, \dots, p_n . Determine the probability that the total number of successes was not less than μ . By clever reasoning he obtained the formula

$$P(\mu \geq m) \leq \frac{1}{2\sqrt{n}} \frac{\sqrt{m(n-m)}}{m-ns} \left(\frac{ns}{m}\right)^m \left(\frac{n(1-s)}{n-m}\right)^{n-m+1}$$

where $m > ns + 1$ and s was the mean probability of success.

His proof was rigorous (although he had not indicated that the trials were independent) and he (p. 259) had the right to reproach Poisson whose method of derivation did not provide the limits of the error of his approximate analysis. Later Chebyshev (1879 – 1880/1936, pp. 162 – 163) explicated one of his intermediate transformations more clearly, also see Bernstein (1945/1964, p. 412). Chebyshev also became able to prove the Poisson LLN, cf. § 8.1.5, in the form

$$\lim P(|(\mu/n) - s| < \varepsilon) = 1, n \rightarrow \infty.$$

Then Chebyshev (1867) generalized this formula and proved the *Chebyshev form of the LLN*: for random variables ξ_i having $E\xi_i \leq C_1$ and $E\xi_i^2 \leq C_2$

$$\lim P[(1/n)|\sum(\xi_i - E\xi_i)| < \varepsilon] = 1, n \rightarrow \infty.$$

2) The Bienaymé – Chebyshev inequality (cf. § 10.2-4). In his lectures Chebyshev (1879 – 1880/1936, pp. 166 – 167) specified it for coinciding random variables and obtained a most important and very simple corollary: the arithmetic mean was a consistent estimator of the expectation of a random variable. He again assumed that the expectations and variances of the appropriate variables were uniformly restricted.

Unlike Heyde & Seneta (§ 10.2-4), I believe that Chebyshev derived this inequality in about the same way as Bienaymé did, only in much more detail. True, he restricted his attention to discrete variables whereas Bienaymé, without elaborating, apparently had in mind the continuous case; his memoir was devoted to the mathematical treatment of observations. Modern authors, whom I mentioned in § 10.2-4, repeat the derivation for the latter instance; actually, already Sleshinsky (1893) had done it.

3) The CLT. Chebyshev (1887b) noted that that theorem led to the MLSq (in accordance with the Laplacean approach). He issued from his inequalities (1874) published without proof for an integral of a non-negative function whose moments up to some order coincided with the same moments of the appropriate, in a definite sense, normal distribution. Markov (1884) and then Stieltjes substantiated them but later he (1885) expressed his regrets at having missed Markov's contribution. Chebyshev justified his inequalities afterwards but without mentioning his predecessors, see Krein (1951).

Chebyshev considered random variables u_1, u_2, \dots, u_n having densities $\varphi_i(x)$ and uniformly bounded moments. He had not expressly assumed independence and did not indicate the restriction

$$\lim [uu]/n \neq 0, i = 1, 2, \dots, n, n \rightarrow \infty. \quad (1)$$

It was not necessary for the moments to be uniformly bounded, but Liapunov (1901b, p. 57) explained that demand by Chebyshev's peculiar turn of speech.

Chebyshev noted that the density $f(x)$ of the fraction

$$x = \Sigma u_i / \sqrt{n} \quad (2)$$

can be determined by means of the multiple integral

$$f(x)dx = \int \int \int \varphi_1(u_1)\varphi_2(u_2) \dots \varphi_n(u_n)du_1du_2 \dots du_n \quad (3)$$

extended over the values of the variables at which the fraction above is situated within the interval $[x; x + dx]$. He multiplied both parts of (3) by e^{sx} where s was some constant and integrated them over $(-\infty; +\infty)$ so that the right side became separated into a product of n integrals with the same limits of integration. Chebyshev then developed both parts in powers of s (the right side, after taking its logarithm) and equated the coefficients of the same powers of that magnitude to each other. Thus the integrals

$$\int f(x)dx, \int xf(x)dx, \int x^2 f(x)dx, \dots$$

or the moments of magnitude (2), were determined up to some order $(2m - 1)$. It occurred that, as $n \rightarrow \infty$, again with the same limits of integration,

$$\int e^{sx} f(x)dx = \exp(s^2/2q^2) \quad (4)$$

where $1/q^2$ was the arithmetic mean of the second moments of u_i and it is here that the condition (1) was needed. Applying his previously mentioned

estimates of the integral of a non-negative function, Chebyshev now completed his proof:

$$\lim P\left(\alpha \leq \frac{\sum u_i}{\sqrt{2\sum E[uu]}} \leq \beta\right) = \frac{1}{\sqrt{\pi}} \int_{\alpha}^{\beta} \exp(-x^2) dx, n \rightarrow \infty. \quad (5)$$

For finite values of n the same probability, as Chebyshev indicated without a rigorous demonstration, was determined by a development in polynomials now called after Chebyshev and Hermite.

Markov (1898/1951, p. 268), when proving the Chebyshev theorem anew, without explaining the situation had eliminated a defect by introducing instead of (1) additional restrictions

$$Eu_i = 0, E u_i^m < \infty, \lim E[uu] \neq 0, n \rightarrow \infty. \quad (6a, b, c)$$

Sleshinsky (1892) issued from Cauchy's findings (§ 10.1) and apparently proved the CLT rigorously even before Markov did, although only for a linear function of observational errors having an even density.

12.2. His Lectures

From 1860 to 1882 Chebyshev delivered lectures on the theory of probability at Petersburg University. In 1936, A. N. Krylov published those read in 1879/1880 as recorded by Liapunov and I refer to his publication by mentioning only the page numbers of this source. I translated this book correcting perhaps a hundred (I repeat: a hundred) mathematical misprints. Ermolaeva (1987) briefly described a more detailed record of Chebyshev's lectures read during September 1876 – March 1878, discovered by herself but still unpublished. She had not indicated whether the newly found text essentially differed from the published version.

The lectures were devoted to definite integrals, the theory of finite differences and the theory of probability. Chebyshev attempted to apply the simplest methods; for example, he used summing, and, if necessary, went on to integration only at the last moment; he introduced characteristic functions only in the discrete case; he did not specify that he considered independent events or variables; he was not interested in the philosophical aspect of probability (Prudnikov 1964, p. 91); and, among the applications of the theory of probability, he almost exclusively discussed (not quite properly) the mathematical treatment of observations.

1) The main notions. Chebyshev (p. 148) declared that the aim of the theory of probability was

to determine the chances of the occurrence of a certain event, and that the word 'event' means anything whose probability is being determined, and probability serves to denote some magnitude that is to be measured.

Boole (1851/1952, p. 251) expressed similar ideas:

The object of the theory of probabilities may be thus stated: Given the separate probabilities of any propositions to find the probability of another proposition.

According to Prokhorov & Sevastianov (1999, p. 77), the theory of probability studies mathematical models of random events and,

Given the probabilities of some random events, makes it possible to determine the probabilities of other random events somehow connected with the first ones.

Tacitly following Laplace (§ 7.1-3), Chebyshev (p. 165) indicated that the concept of limit in probability theory differed from that in analysis, but I am still unable to understand such equalities (or are they misprints?) as on pp. 167, 183, 204/156, 171, 190

$$\lim m/n = p. \quad (7)$$

2) The limit theorem for Poisson trials (pp. 167 and 201ff). Determine the probability

$P_{n,m}$ that in n trials an event having probabilities $p_i, i = 1, 2, \dots, n$, respectively, occurred m times. Applying a little known formula from the first section of his *Lectures* Chebyshev obtained

$$P_{n,m} = \frac{1}{2\pi} \int_{-\pi}^{\pi} [p_1 e^{i\phi} + q_1] [p_2 e^{i\phi} + q_2] \dots [p_n e^{i\phi} + q_n] e^{-m\phi} d\phi, \quad q_i = 1 - p_i.$$

After some transformations and considering only small values of ϕ it occurred that

$$P_{n,m} = \frac{1}{\pi} \int_0^{\pi} \exp(-nQ\phi^2/2) \cos [(np - m)\phi] d\phi$$

where p was the mean probability of success and $Q = [pq]/n$. Assuming for large values of n an infinite upper limit in the obtained integral, Chebyshev finally got

$$P[|m/n - p| < t\sqrt{2Q/n}] = \frac{2}{\sqrt{\pi}} \int_0^t \exp(-z^2) dz$$

(without the sign of limit!) and noted that formula (7), or, as he concluded, the Poisson LLN, followed from it. He naturally did not here admonish his predecessor.

3) The CLT (pp. 219 – 224). At the time, Chebyshev had not yet known its rigorous proof. I only note his pronouncement (p. 224): the formula that he obtained was not derived

in a rigorous way [...]. We have made various assumptions but did not determine the boundary of the ensuing error. In its present state,

mathematical analysis cannot derive this boundary in any satisfactory fashion.

4) Statistical inferences. Chebyshev solved two problems which, however, were considered before him. In the first of these he (pp. 187 – 192) derived the Bayes limit theorem (§ 5.2) but did not cite anyone, and in the second he (pp. 193 – 201) studied the probability of a subsequent result in Bernoulli trials. An event occurred m times in n trials; determine the probability that it will happen r times in k new trials. Guiding himself mostly by the Stirling theorem, Chebyshev non-rigorously derived an integral limit theorem similar to that obtained by Laplace (§ 7.1-5).

5) Mathematical treatment of observations (pp. 224 – 252). Chebyshev (p. 227) proved that the arithmetic mean was a consistent estimator of the unknown constant. Unlike Poincaré (§ 11.3-7), he (pp. 228 – 231) justified its optimality by noting that, among linear estimators, the mean ensured the shortest probable interval for the ensuing error. The variance of the arithmetic mean was also minimal (Ibidem); although Chebyshev had not paid special attention to that estimator of precision, it occurred that he, in principle, based his reasoning on the definitive Gaussian substantiation of the MLSq (§ 9.1.3).

At the same time Chebyshev (pp. 231 – 236) derived the normal distribution as the universal law of error in about the same way as Gauss did in 1809. *The Gauss method*, Chebyshev (p. 250) maintained, bearing in mind exactly that attempt later abandoned by Gauss, was based on the doubtful *law of hypotheses*, – on the *Bayes theorem* with equal prior probabilities. Chebyshev several times censured that *law* when discussing the Bayesian approach in his lectures and he (p. 249) wrongly thought that the Gauss formula (9.6b) had only appeared *recently* and that it assumed a large number of observations. He did not mention that the Gauss formula provided an unbiased estimation. It might be concluded that the treatment of observations hardly interested him.

6) Cancellation of a fraction (pp. 152 – 154). Determine the probability P that a *random* fraction A/B cannot be cancelled. Markov remarked that Kronecker (1894, Lecture 24) had solved the same problem and indicated Dirichlet's priority. Kronecker had not supplied an exact reference and I was unable to check his statement; he added that Dirichlet had determined the probability sought *if it existed at all*. Anyway, Bernstein (1928/1964, p. 219) refuted Chebyshev's solution and indicated (p. 220), that the theory of numbers dealt with regular number sequences whose limiting or asymptotic frequencies of numbers of some class, unlike probabilities, *which we will never determine experimentally*, might be studied. See Postnikov (1974) on the same problem and on the stochastic theory of numbers.

12.3. Some General Considerations

And so, Chebyshev argued that the propositions of the theory of probability ought to be rigorously demonstrated and its limit theorems should be supplemented by estimation of the errors of *pre-limiting* relations (Kolmogorov 1947, p. 56). He himself essentially developed the LLN and, somewhat imperfectly, proved for the first time the CLT; on the study of these two issues depended the *destiny* of the theory of probability (Bernstein

1945/1964, p. 411). His students also contributed to the theory (§§ 13.1, 13.2, 13.4).

Kolmogorov continued: Chebyshev was the first to appreciate clearly and use *the full power* of the concepts of random variable and [its] expectation. However, Chebyshev had not made use of Poisson's heuristic definition of random variable (§ 8.1), had not applied this term and did not study densities or generating functions as mathematical objects. Then, the entire development of the theory of probability from Chebyshev onward might be described as an ever fuller use of the power of the abovementioned concepts; thus, it had since begun to study dependent random variables, their systems and chains.

Here also is Bernstein's conclusion (1945/1964, p. 432):

The genius of Chebyshev and his associates, who, in this field [theory of probability], have left mathematicians of Western Europe far behind, have surmounted the crisis of the theory of probability that had brought its development to a stop a hundred years ago.

However, Novikov (2002, p. 330) stated that *in spite of his splendid analytical talent, Chebyshev was a pathological conservative*. He corroborated it by referring to V. F. Kagan (1869 – 1953), an eminent geometrician. The latter, *when being a young Privat-Dozent*, had listened to Chebyshev's scornful statement on the *trendy disciplines like the Riemann geometry and complex-variable analysis*. Even Liapunov (1895/1946, pp. 19 – 20) called Riemann's ideas *extremely abstract*; his investigations, *pseudo-geometric* and sometimes, again, too abstract and having nothing in common with Lobachevsky's *deep geometric studies*. Liapunov did not recall Klein, who had in 1871 presented a unified picture of the non-Euclidean geometry in which the findings of Lobachevsky and Riemann appeared as particular cases. On the other hand, Tikhomandritsky (1898, p. IV) testified that in 1887 Chebyshev had *stated that [...] it is necessary to transform the entire theory of probability*. It is difficult to say what exactly did he mean.

13. Markov, Liapunov, Nekrasov

I consider here the work of three outstanding scholars. Markov completed the proof of the CLT and opened up a new direction in probability. Liapunov proved the CLT by following latest mathematical developments. Nekrasov attacked the CLT purely analytically and was the first to consider the CLT in case of large deviations but got entangled. Then, he hopelessly linked probability with religion and shallow philosophy.

Key words: CLT, case of large deviations, Markov chains

13.1. Markov: Personal Traits

For his biography see Markov Jr (1951), a noted mathematician in his own right, and Grodzensky (1987). They describe his principled stand on burning social and political issues whereas Grodzensky also published many of his pertinent newspaper letters, some of them for the first time; apparently, the newspapers did not always accept them. Markov struggled against anti-Semitism and denounced the Russian Orthodox Church, see also Sheynin (1989, pp. 340 – 341; 2007b). The Press used to call him *Militant academician* (Nekrasov 1916, p. 9) and *Andrew the Furious* (Neyman 1978).

In 1901 Tolstoy was excommunicated from the Church. During his last days, the Most Holy Synod discussed whether he should be *admitted to the bosom of the Church* and decided against it (Anonymous 1910), so that in 1912 Tolstoy's excommunication was likely well remembered. Then, in 1912 Markov submitted a request to the Synod for excommunication mentioning his doubts about events *allegedly having occurred in bygone times* and adding that he did not *sympathise with religions which, like Orthodoxy, are supported by, and in turn lend their support to fire and sword*. The Synod resolved that Markov *had seceded from God's church* (Emeliakh 1954, pp. 400 – 401 and 408). In a letter of 1915 (Sheynin 1993a, p. 200) Markov maintained that graduates of Russian Orthodox seminaries

are getting accustomed [...] to a special kind of reasoning. They must subordinate their minds to the indications of the Holy fathers and replace their minds by the texts from the Scripture.

In 1921 (Grodzensky 1987, p. 137) 15 professors of the Petrograd University declared that applicants ought to be chosen according to their knowledge rather than to class or political considerations; Markov was the first to sign their unsuccessful statement.

Markov's attitude towards other scholars had often been wrong. Just one example, Andreev's letter of 1915 to Nekrasov, see Sheynin (1994e, p. 132): Markov

Remains to this day an old and hardened sinner in provoking debate. I had understood this long ago, and I believe that the only way to save myself from the trouble of swallowing the provocateur's bait is a refusal to respond to any of his attacks...

In his own scientific work, Markov had been too rigid, see §§ 13.2 – 13.3, which negatively influenced his work. During his last years, in spite of extremely difficult conditions of life in Russia and his worsened health, he completed the last posthumous 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 even formulated an absolutely wrong statement (end of § 13.2.5). 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).

13.2. Markov: General Scientific Issues

1) History of the theory of probability. Markov investigated the Bernoulli LLN (§ 3.2.3); in 1913 he initiated a jubilee meeting of the Petersburg Academy of Sciences celebrating the bicentenary of that law, commented on the history of the Bienaymé – Chebyshev inequality and the method of moments (§ 10.2-2) and stressed De Moivre's part in establishing the *Stirling formula*. The last edition of his *Treatise* includes many interesting historical remarks.

2) Insurance of life. Markov collaborated with pension funds (Sheynin 1997c) and in 1906 he destructively criticized a proposed scheme for insuring children (reprinted in same article).

3) Calculations. I mention his table of the normal distribution (1888) 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 1962). Markov (1899b, p. 30) indirectly expressed his attitude toward calculations:

Many mathematicians apparently believe that going beyond the field of abstract reasoning into the sphere of effective calculations would be humiliating.

4) Correlation theory. In a letter of 1912 to him Slutsky (Sheynin 1990a/2011, p. 64) stated that *the shortcomings of Pearson's exposition are temporary* and will be overcome. Markov, however, continued largely to ignore him. Thus, he (1916/1951, p. 533) reasonably criticized the correlation theory, actually since it was still imperfect, but did not mention its possible worth:

Its positive side is not significant enough and consists in a simple usage of the method of least squares to discover linear dependences. However, not being satisfied with approximately determining various coefficients, the theory also indicates their probable errors and enters here the region of fantasy, hypnosis and faith in such mathematical formulas that, in actual fact, have no sound scientific justification.

Now, discovering dependences, even if only linear, is indeed important; and the estimation of plausibility of the results obtained is an essential part of any investigation.

5) Principles of the theory of probability. Markov (1911c/1981, pp. 149 – 150) thought that their discussion was meaningless and even declared

(1900/1924, c. 2) that *various concepts are defined not by words [...] but rather by [our] attitude towards them ascertained little by little*. The axiomatic approach had been necessary, but Markov, like a student of Chebyshev, underrated both it and the complex analysis (A. A. Youshkevich 1974, p. 125). He (1900/1924, pp. 10, 13 – 19 and 24) even claimed to have proved the addition and multiplication theorems and thus to transfer the calculus of probability to the realm of pure mathematics, – in spite of its failure to study densities or characteristic functions as mathematical objects, cf. § 7.3. P. Lévi (1925) was apparently the first to take this step.

Markov did not define probability anew either, but this seems to be impossible (and axiomatization did not help practitioners). In geometry, the situation is better since such notions as area of figure are indirectly defined by the appropriate integrals; on the other hand, the straight line remained undefined which prompted the appearance of the non-Euclidean geometry.

6) Mathematical statistics. By the end of his life Markov, mostly under the influence of Chuprov (Sheynin 1990a/2011, p. 76, his letter of ca. 1924 to another statistician), somewhat softened his attitude to Pearson:

Markov regarded Pearson, I may say, with contempt. Markov's temper was no better than Pearson's, he could not stand even slightest contradictions either. You can imagine how he took my persistent indications to the considerable scientific importance of Pearson's works. My efforts thus directed were not to no avail as proved by [Markov 1924]. After all, something [Pearsonian] was included in the field of Markov's scientific interests.

Chuprov (1925b) also published a review of the mentioned edition of Markov's *Treatise*. Here, I only cite his reasonable criticism of Markov's treatment of correlation theory:

The choice of questions on which attention is concentrated is fortuitous, their treatment within the bounds of the chapter on the method of least squares is incomplete, the connection made between the theory of correlation and the theory of probability is inadequate...

Yes, 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), but paid no attention either to the chi-squared test (§ 13.3-1) or to the Pearsonian curves.

7) Teaching probability theory in school. In 1914 Nekrasov made an attempt to introduce probability into the school curriculum. Markov (1915a) protested against the proposed school programme, but did not object to the very principle. He became a member of an ad hoc academic Commission which voiced an extremely negative opinion (*Report* 1916) about Nekrasov's programme and his understanding of the main concepts of mathematical analysis, see § 13.5.

8) Methodological issues. Many authors praised the methodological value of Markov's contributions, see however § 13.3-1 (in particular, his own letter) and Idelson (1947, p. 101) who voiced a negative opinion. Then, Markov refused to apply the term *random magnitude* (as it has been called

in Russia) and the expressions *normal law* and *coefficient of correlation* were likewise absent in his works. And, not wishing to leave his field (§13.3-1, letter to Chuprov), he never mentioned applications of his chains to natural sciences. The structure of his *Treatise* became ever more complicated with each new edition.

13.3. Markov: Main Investigations

1) Mathematical treatment of observations. In spite of several commentators, I deny Markov's accomplishments here. Neyman (1934, p. 595) invented a non-existing Gauss – Markov theorem and F. N. David & Neyman (1938) repeated this mistake but finally Neyman (1938/1952, p. 228) admitted it.

In his *Treatise* (1900) Markov combined the treatment of observations with the study of correlation, statistical series and interpolation, but his innovation was methodically doubtful. While discussing statistical series, Markov did not mention Chuprov's relevant papers (1916; 1918 – 1919). When considering Weldon's experiment with 26,306 throws of 12 dice (K. Pearson 1900), Markov (*Treatise* 1924, pp. 349 – 353) decided, after applying the CLT and the Bayes theorem with transition to the normal law, that the probability of a 5 or a 6 was higher than 1/3. Unlike Pearson, he had not used the chi-squared test and apparently left an impression that (although suitable for a small number of trials as well) it was not needed at all. Markov possibly followed here his own rigid principle (Ondar 1977/1981, Letter 44 to Chuprov of 1910): *I shall not go a step out of that region where my competence is beyond any doubt.*

The explication of the MLSq proper was involved; in a letter of 1910 to Chuprov Markov (Ondar 1977/1981, p. 21) wrote: *I have often heard that my presentation is not sufficiently clear.* In 1893, his former student, Koialovitch (Sheynin 2006a, pp. 81 and 85), writing to Markov, formulated some puzzling questions about his university lectures.

2) The LLN. Markov (1906/1951, p. 341) noted that the condition

$$\lim E\{[E\sum \xi_i - \sum E\xi_i]^2/n^2\} = 0, n \rightarrow \infty \quad (1)$$

was sufficient for the sequence $\xi_1, \xi_2, \dots, \xi_n, \dots$ of random variables to obey the LLN; or to comply with the condition

$$\lim P\{(1/n)|(\sum \xi_i - \sum E\xi_i)| < \varepsilon\} = 1, n \rightarrow \infty.$$

Then Markov (Ibidem, pp. 342 – 344; *Treatise*, 1913, pp. 116 – 129) derived a few sufficient conditions for sequences of independent, and, especially, dependent random variables (1906/1951, p. 351; *Treatise* 1913, p. 119; 1924, p. 174), provided examples of sequences not obeying the law, and (*Treatise*, 1913, p. 129), proved that independent variables obeyed the LLN if, for every i , there existed the moments

$$E\xi_i = a_i, E|\xi_i - a_i|^{1+\delta} < C, 0 < \delta < 1.$$

Again, Markov (*Treatise*, 1900; p. 86 in the edition of 1924) had proved that, for a positive random variable ξ ,

$$P(\xi \leq t^2 E\xi) > 1 - 1/t^2$$

and Bortkiewicz (1917, p. 36) and Romanovsky (1925a; 1925b) called this inequality after Markov.

3) The CLT. As I mentioned at the end of § 12.1-3, Markov specified the conditions of theorem (12.2) proved by Chebyshev. He (1898/1951, p. 268) considered independent random variables u_i with zero expectations and introduced conditions (12.3) but he returned several times to the CLT.

a) He (1899a/1951, p. 240) additionally introduced two restrictions: as $n \rightarrow \infty$,

$$\lim E[(u_1 + u_2 + \dots + u_n)^2] = \infty, \lim [E(u_1 + u_2 + \dots + u_n)^2/n] \neq \varepsilon. \quad (2; 3)$$

b) Markov (1907, p. 708) again proved the CLT. Referring to his papers (1898; 1899a), he now introduced conditions (12.3b) for finite values of i and (3) but did not restrict the values of u_i . On his next page Markov abandoned condition (3) *if only*

$$\lim E u_n^2 = \infty, n \rightarrow \infty \quad (4)$$

and the values of u_i remained finite. Restrictions (2) and (4) certainly coincided.

c) Markov (1908a) essentially extended the applicability of the method of moments by replacing his conditions by Liapunov's single restriction (1901a/1954, p. 159)

$$\lim \frac{\sum E u_i^{2+\delta}}{(\sum \text{var } u_i^2)^{1+\delta/2}} = 0, \delta > 0, n \rightarrow \infty.$$

In 1913 Markov included a modified version of his last-mentioned study in his *Treatise*; it is also reprinted (Markov 1900/1924; 1951, pp. 319 – 338).

Markov (1899b, p. 42) mentioned the example provided by Poisson (1824, § 10) who proved that the limiting distribution of the linear form

$$L = \varepsilon_1 + 1/3\varepsilon_2 + 1/5\varepsilon_3 + \dots$$

of random variables ε_i with density $e^{-2|x|}$ was

$$\lim P(|L| \leq c) = 1 - (4/\pi) \text{arc tan } e^{-2c}, n \rightarrow \infty.$$

In this example $\lim \text{var} [\varepsilon_n/(2n - 1)] = 0, n \rightarrow \infty$.

Markov himself (1899a/1951, pp. 242 – 246) also provided an example in which the condition (2) did not hold and the CLT did not take place.

The appearance of condition (3) remains, however, unclear. Nekrasov (1900 – 1902, 1902, pp. 292 and 293) introduced it for independent variables instead of restriction (4). Liapunov (1901a/1954, p. 175) maintained that it was not sufficient. Seneta (1984, p. 39) indicated, however, that Markov's published papers had not contained such examples

and that condition (4) was necessary and sufficient for the CLT in the case of uniformly restricted variables.

4) Markov chains. This term is due to Bernstein (1926, §16); Markov himself (1906/1951, p. 354) called them simply *chains*. He issued from a paper by Bruns of the same year, but the prehistory of Markov chains is much richer. Here are the main relevant issues.

a) The Daniel Bernoulli – Laplace urn problem, the predecessor of the Ehrenfests' model (§ 7.1-3);

b) The study of the Brownian movement (Brush 1968);

c) The problem of the extinction of families (§ 10.2-4);

d) The problem of random walks (Dutka 1985);

e) Some of Poincaré's findings;

f) The work of Bachelier (1900) on financial speculations, also see Courtault et al (2000) and Taqqi (2001).

Markov (1906/1951, pp. 345 and 354) considered simple homogeneous chains of random events and discrete random variables and proved that the LLN was applicable both to the number of successes and to the sequences of these variables. Later he (1910/1951, p. 476) extended the first of these findings to simple non-homogeneous chains.

Markov proved the CLT for his chains. He considered simple homogeneous chains of events (1906) and of random variables (1908b); and complex homogeneous (1911a; 1911b) chains of random variables; simple homogeneous chains of indirectly observed events (1912a). While studying the chains, Markov established important ergodic theorems but had not paid them any special attention; in this connection, I mentioned one of his solved problems in § 7.1-3.

Markov widely applied the method of moments, and only he who repeats some of his investigations will be able to appreciate the obstacles which he overcame. Bernstein (1945/1964, p. 427), however, contrasted Markov and Liapunov. The latter had applied the classical transcendental analysis as developed by that time whereas the method of moments, Bernstein maintained, *did not facilitate the problem* [of proving the CLT] *but rather transferred all its difficulties elsewhere.*

13.4 Liapunov

The theory of probability remained an episode in his scientific work. He (1900; 1901a) proved the CLT assuming a single condition (5). I briefly repeat (Bernstein 1945/1964, pp. 427ff) that a characteristic function determines the sought law of distribution independently from the existence of the relevant moments. Liapunov proved that under his condition the characteristic function of a centred and normed sum of random variables tended to the characteristic function of a normed normal law. I also mention Lindeberg (1922b, p. 211) whose proof of the CLT was simpler and became better known. He referred to his previous paper (1922a) and continued:

I see now that already Liapunov had explicated general findings which not only surpass the results achieved by Mises [...] but which make it possible to derive most of what I have established. [...] The study of Liapunov's work prompted me to check anew the method that I have applied.

Chebyshev thought that the limits of integration, α and β , in formula (12.2) describing that theorem, were *any*. Nekrasov (1911, p. 449) arbitrarily interpreted that expression as *variable*. I discuss Nekrasov in § 13.5; he could have well indicated that, on the contrary, he had generalized the Chebyshev theorem. In his previous polemic paper Liapunov (1901b, p. 61) declared that he had assumed that these limits were given beforehand and that otherwise the probability, written down in the left side of formula of the CLT, could have no limit at all, – but nevertheless be asymptotically expressed by the normal law of distribution.

13.5. Nekrasov

His life and work (Sheynin 2003a) are separated into two stages. From 1885 and until about 1900 he had time to publish remarkable memoirs not connected with probability both in Russia and Germany and to become Professor and Rector of Moscow University. In 1898 he sketched the proof of the CLT for sums of lattice random variables. Then, however, his personality changed. His writings became unimaginably verbose, sometimes obscure and confusing, and inseparably linked with ethical, political and religious considerations. Here is a comparatively mild example (1906, p. 9):

*mathematics accumulated
psychological discipline as well as political and social arithmetic or the
mathematical law of the political and social development of forces
depending on mental and physiological principles.*

Furthermore, Nekrasov's work began to abound with elementary mathematical mistakes and senseless statements. For example (1901, p. 237): it is possible to assume roughly, that x^n , $n > 0$, is the limit of $\sin x$ as $|x| \rightarrow 0$, and *the conclusions made by [Chebyshev, Markov and Liapunov] never differ much from such an understanding of limit*. And here is his astounding declaration (Archive, Russian Academy of Sciences, fond 173, inventory 1, 55, No. 5) from his letter of 1913 to Markov:

I distinguish the viewpoints of Gauss and Laplace [on the MLSq] by the moment with regard to the experiment. The first one is posterior and the second one is prior. It is more opportune to judge à posteriori because more data are available, but this approach is delaying, it lags behind, drags after the event.

At least the attendant reasons for such a change were Nekrasov's religious upbringing (before entering Moscow University he graduated from a Russian Orthodox seminary), his work from 1898 onward as a high official at the Ministry of People's Education, and his reactionary views. In his letter of 1916 to the religious philosopher P. A. Florensky (Sheynin 1993a, p. 196) Nekrasov stated that *the German – Jewish culture and literature pushed us to the crossroads*. World War I was then going on which only partly exonerates Nekrasov. I shall now dwell on some concrete issues.

1) Teaching the theory of probability. In § 13.2-7 I mentioned Nekrasov's proposal for teaching probability in school and the rejection of the curriculum drawn up by him.

2) The MLSq. Nekrasov (1912 – 1914) mistakenly attributed to Legendre an interpolation-like application of the method and (1914) acknowledged his failure to notice, in 1912, the relevant work of Yarochenko (1893a; 1893b).

3) The CLT. It was Nekrasov who had considered the CLT for large deviations, – for the case that began to be studied only 50 years later. He (1898) formulated six theorems and proved them later (1900 – 1902). Neither Markov, nor Liapunov had sufficiently studied them; indeed, it was hardly possible to understand him and Soloviev (1997/2008, p. 359) reasonably stated:

I am firmly convinced that no contemporary mathematician or later historian of mathematic had (has) ever studied it [the memoir (1900 – 1902)] in any detail.

He himself only suggested that Nekrasov had indeed proved his theorems and he reminded his readers that Markov had indicated some mistakes made by Nekrasov. Furthermore, Soloviev (pp. 356 – 357) remarked that Nekrasov had wrongly understood the notion of lattice variables. He (p. 362) also stated that it was generally impossible to check some of Nekrasov's restrictions. Both he and Seneta (1984, §6), agree in that Nekrasov's findings had not influenced the development of the theory of probability which was certainly caused both by Nekrasov's inability to express himself intelligibly and by the unwieldiness of his purely analytical rather than stochastic approach (Soloviev, p. 363).

14. The Birth of Mathematical Statistics

By the end of the 19th century, Lexis and his followers began to study the behaviour of the probability of the studied event in observational series. Their work led to general achievements, but it was overshadowed by the less rigorous tireless efforts of K. Pearson and his school. Fisher, who originated mathematical statistics, indirectly owed much to Pearson.

Key words: Continental direction of statistics, Biometric school, biometry

14.1. The Stability of Statistical Series

By the end of the 19th, and in the beginning of the 20th century, statistical investigations on the Continent were mostly 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 (§ 8.1) examined the significance of statistical discrepancies for a large number of observations without providing examples. Cournot (§ 10.3-5) 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 (§ 10.5) investigated the connections between causes and effects in society, attempted to standardize statistical data worldwide and, following Süssmilch, 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 (§§ 10.7-7, 10.7-8).

14.1.1. Lexis. He (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, $i = 1, 2, \dots, m$, 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)

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

(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 (9.6b); he also knew formula (1), see 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 random deviations from unity considered admissible); 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.

But how could the probability vary? Lexis (1876, pp. 220 – 221 and 238) thought that the variations followed a normal law, but then he (1877, § 23) admitted less restrictive conditions (evenness of the appropriate density function) and noted that more specific restrictions were impossible. I am not sure that Lexis had broken off with previous traditions, see § 10.7-7. He (1879) discussed this issue once more, and even mentioned *irregular waves* (§ 22), but it is difficult to follow him. He interrupted himself by providing statistical examples and never gave precise formulations.

Lexis had not calculated either the expectation, or the variance of his coefficient (which was difficult). His main achievement was perhaps an attempt to check statistically some stochastic model. He (1879, § 1) also qualitatively separated statistical series into several types and made a forgotten attempt to define stationarity and trend.

A French actuary Dormoy (1874; 1878) preceded Lexis, but even French statisticians (who barely participated in the contemporary development of statistics) had not noticed his theory. It was Lexis who first discovered Dormoy (Chuprov 1909/1959, p. 236) and Chuprov (1926, p. 198/1960, p. 228) argued that the Lexian theory of dispersion ought to be called after Dormoy and Lexis. Bortkiewicz (1930), however, later ranked Dormoy far below Lexis and, be that as it may, later statisticians had only paid attention to Lexis.

14.1.2. Bortkiewicz. See also § 10.7-4. Of Polish descent, Vladislav Iosifovich Bortkevich was a lawyer by education. He was born and studied in Petersburg, but at the end of the 19th century continued his education in Germany (he was Lexis' student). In 1901 he secured a professorship in Berlin and remained there all his life as Ladislaus von Bortkiewicz. For further detail see Sheynin (1990a/2011, § 7.3).

He (1903) sharply criticized Nekrasov (1902) for the latter's statements that the theory of probability can soften *the cruel relations* between capital and labour (p. 215) and for attempts (p. 219) to exonerate the principles of firm rule and autocracy, for Nekrasov's *sickening oily tone* (p. 215) and *reactionary longings* (p. 216). Although Bortkiewicz was not initially acquainted with mathematics, he achieved interesting findings. Woytinsky (1961, pp. 452 – 453) stated that he was called *the statistical Pope* whereas Schumacher (1931, p. 573) explained Bortkiewicz' attitude towards science by a quotation from Exodus 20:3: *You shall have no other gods before me.*

Chuprov's student and the last representative of the Continental direction, Anderson (1932, p. 243/1963, Bd. 2, p. 531), described Bortkiewicz' achievements:

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 [later, Anderson added Chuprov] have managed to discover.

Bortkiewicz' work is insufficiently known mostly because of his pedestrian style and excessive attention to details, but also since German

statisticians and economists of the time (Bortkiewicz was also a celebrated economist) had been avoiding mathematics. He did not pay attention to improving his style and refused to mend his ways. Winkler (1931, p. 1030) quoted a letter from Bortkiewicz (date not given) who was glad to find in him one of the five expected readers of his work! Here is Anderson's appraisal (1932, p. 245/1963, Bd. 2, p. 533):

Bortkiewicz did not write for a wide circle of readers [...] and was not at all a good exponent of his own ideas. In addition, he made very high demands on the readers' schooling and intellect. With stubbornness partly caused by his reclusive life, [...] he refused to follow the advice of [...] Chuprov...

Bortkiewicz had determined EQ and EQ^2 and Markov (1911c/1981, p. 153), see also Ondar (1977/1981, Letter 47 of 1912), positively mentioned his work. Then, Bortkiewicz introduced his ill-fated law of small numbers (§ 8.1.2) for studying the stability of statistical series and did not listen to Chuprov's mild criticism. However, he was the main author who picked up Poisson's law and for a long time his contribution (1898a) had remained the talk of the town.

14.1.3. Markov and Chuprov. In his letters of 1910 to Chuprov, Markov (Ondar 1977) proved that Lexis' considerations were wrong. It occurred that the dispersion could also be normal when the observations were dependent. Also in 1910, Chuprov, in a letter to Markov, provided examples of dependences leading to super- and sub-normality of dispersion; in 1914 he decided that the coefficient of dispersion should be *shelved* to which Bortkiewicz strongly objected (Sheynin 1990a/2011, p. 140). Then, in 1916 both Markov and Chuprov proved that $EQ^2 = 1$ (p. 141). Finally, Chuprov (Ibidem, p. 142), definitively refuted the applicability of the coefficient of dispersion, but his conclusion is hardly known even now.

Chuprov (1918 – 1919, p. 205) proved, in an elementary way, a general formula

$$(1/n)E\left(\sum_{i=1}^n (x_i - \sum_{i=1}^n E\xi_i)^2\right) =$$

$$(1/n^2) \sum_{i=1}^n E(\xi_i - E\xi_i)^2 + (1/n^2) \sum_{i=1}^n \sum_{j \leq i} [E(x_i x_j) - E\xi_i E\xi_j].$$

Included here were n random variables ξ_i anyhow dependent on each other and the results of a single observation x_i of each of them.

While studying the stability of statistical series, Chuprov achieved really interesting results, see Seneta (1987), but, since he considered problems of the most general nature, he inevitably derived awkward formulas. Romanovsky (1930, p. 216) noted that Chuprov's formulas, although *being of considerable theoretical interest*, were *almost useless* due to complicated calculations involved.

14.2. The Biometric School

The first issue of *Biometrika* appeared in 1902. Its editors were Weldon (a biologist who died in 1906), Pearson and Davenport *in consultation* with Galton. The editorial there contained the following passage:

The problem of evolution is a problem in statistics [...] [Darwin established] the theory of descent without mathematical conceptions [...] [but] every idea of Darwin – variation, natural selection [...] – seems at once to fit itself to mathematical definition and to demand statistical analysis. [...] The biologist, the mathematician and the statistician have hitherto had widely differentiated fields of work. [...] The day will come [...] when we shall find mathematicians who are competent biologists, and biologists who are competent mathematicians ...

In 1920, Pearson (E. S. Pearson 1936 – 1937, vol. 29, p. 164) defined the aim of the Biometric school as making statistics a branch of applied mathematics and providing various disciplines applying it with a new and stronger technique. The success of the new school was partly caused by the efforts of Edgeworth who was excessively original, had an odd style and was unable to influence strongly his contemporaries.

Pearson's *Grammar of science* (1892) earned him the brand of a *conscientious and honest enemy of materialism* and *one of the most consistent and lucid Machians* (Lenin in 1909, in his *Materialism and Empiriocriticism*); the latter term is tantamount to Mach's philosophy and Mach (1897, Introduction) had most positively mentioned Pearson's *Grammar*. Newcomb had highly regarded Pearson (§ 10.8.4).

After Todhunter (1865), Pearson (1978) was apparently the first considerable work in its field but it is more important to mention Pearson's fundamental biography of Galton (1914 – 1930), perhaps the most immense book from among all works of such kind, wherever and whenever published. Many of his contributions are reprinted in Pearson (1948); see the bibliography of his works in Morant et al (1939) and Merrington et al (1983).

The work of Fisher began in 1911 but he was only able to publish a single paper in *Biometrika* (in 1915). However, at the end of the day he surpassed Pearson. It was he rather than his predecessor with whom the birth of the real mathematical statistics is much more closely connected.

Pearson's main merits include the compilation of numerous statistical tables, development of the principles of the correlation theory and contingency, the introduction of the *Pearsonian curves* for describing empirical distributions (1896 with additions in 1901 and 1916), rather than for replacing the normal law by another universal density, and the χ^2 test (1900) which he had been applying for checking the goodness of fit; independence in contingency tables; and homogeneity. Pearson constructed the system of those curves in accordance with practical considerations and defined it as the solution of the differential equation with four parameters. He attempted, often successfully, to apply the statistical method, and especially correlation theory, in many other branches of science (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 was difficult to correlate Mendelism and biometry: the former studied discrete magnitudes while the latter investigated continuous quantitative variations. Later developments threw a different light on this subject (Johannsen 1922).

I (2010) collected pronouncements of celebrated scientists about Pearson, both positive (Kolmogorov, Bernstein, Mahalanobis, Newcomb) and negative (Fisher). Here, I only quote two authors.

Fisher (1937, p. 306) objected to Pearson's view of maximum likelihood, stating that his

Plea of comparability [between the methods of moments and maximum likelihood] is [...] only an excuse for falsifying the comparison [...].

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.

Of special interest is the testimony of Camp (1933) who worked under Pearson at the Galton laboratory. Although patently prettifying Pearson, he put forward facts and impressions hardly available elsewhere.

It is also necessary to mention W. S. Gosset (pen-name Student). Not a member of the Biometric school, he was *one of the pioneers in the development of modern statistical method and its application to the design and analysis of experiments* (Irwin 1978, p. 409). Specifically, best known is his work on treating small samples and the *t*-test. Fisher aptly called him the *Faraday of statistics* (Ibidem, p. 410) since, in a sense, his intuitive feeling was better than his mathematics. It was perhaps this circumstance that Karl Pearson had in mind when, in a letter of ca. 1914 to Chuprov's follower, Anderson, he called Student *kein Fachmann* (Sheynin 1990a/2011, p. 153).

E. S. Pearson & Wishart (1943) published Student's collected papers and E. S. Pearson (1990) is a most informative source about Student and his contemporaries. It does not, however, include the bibliography of his works nor contain a concise description of his findings.

14.3. The Merging of the Two Streams?

I (§ 14.1-4) noted that the Continental statisticians were not recognizing Pearson. Many of his colleagues, Chuprov wrote, *like Markov, shelve the English investigations without reading them*. The cause of that attitude was the empiricism of the Biometric school (Chuprov 1918 – 1919, t. 2, pp. 132 – 133):

The reluctance, characteristic of English researchers, to deal with the notions of probability and expectation led to much trouble. It greatly damaged clearness [...] and even directed them to a wrong track. [...] However, after casting away that clothing [...] and supplementing the

neglected, [the kinship between Lexis and Pearson] will become obvious. [...] Not Lexis against Pearson, but Pearson refined by Lexis, and Lexis enriched by Pearson should be the slogan of those who are dissatisfied with the heartless empiricism.

So, did the two statistical streams merge, as Chuprov would have it? In 1923 he had become Honorary Fellow of the Royal Statistical Society and in 1926, after his death, the Society passed a resolution of condolence (Sheynin 1990a/2011, p. 156) which stated that his

Contributions to science were admired by all [...]. They did much to harmonise the methods of statistical research developed by continental and British workers.

Bauer (1955, p. 26) reported that he had investigated how both schools 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. More details about Bauer's study are contained in Heyde & Seneta (1977, pp. 57 – 58) where it also correctly indicated that, unlike the Biometric school, the Continental direction had concentrated on nonparametric statistics.

I myself (Gnedenko & Sheynin 1978/2001, p. 275) 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 the 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.

Supplement: Axiomatization

I present a bibliographic survey of some important points.

The main essays are Barone & Novikoff (1978) and Hochkirchen (1999) and among the lesser known authors is Bernstein (1917). After Hilbert (1901), Kolmogorov (1933) made the decisive step and Freudenthal & Steiner (1966, p. 190) commented: *he came with the Columbus' egg*. As the legend goes, Columbus cracked an egg which enabled it to stand firmly on his table. Among the new sources I list Hausdorff (2006) who left an important unpublished contribution, see Girlich (1996), Shafer & Vovk (2001) and Kregel (2011) who stressed the role of Bohlmann. Vovk & Shafer (2003, p. 27) characterized their book:

We show how the classical core of probability theory can be based directly on game-theoretic martingales, with no appeal to measure theory. Probability again becomes [a] secondary concept but is now defined in terms of martingales.

In concluding, I quote Boole (1854/1952, p. 288):

The claim to rank among the pure sciences must rest upon the degree in which it [the theory of probability] satisfies the following conditions: 1° That the principles upon which its methods are founded should be of an axiomatic nature.

Boole formulated two more conditions of a general scientific essence.

Bibliography

Barone J., Novikoff A. (1978), History of the axiomatic formulation of probability from Borel to Kolmogorov. *Arch. Hist. Ex. Sci.*, vol. 18, pp. 123 – 190.

Bernstein S. N. (1917, in Russian), An essay on the axiomatic justification of the theory of probability. *Sobranie Sochineniy* (Works), vol. 4. N. p., 1964, pp. 10 – 60.

Boole G. (1854), On the conditions by which the solution of questions in the theory of probability are limited. In author's *Studies in Logic and Probability*, vol. 1. London, 1952, pp. 280 – 288.

Freudenthal H., Steiner H.-G. (1966), *Die Anfänge der Wahrscheinlichkeitsrechnung*. In *Grundzüge der Mathematik*, Bd. 4. Hrsg H. Behnke et al. Göttingen, pp. 149 – 195.

Girlich H.-J. (1996), Hausdorffs Beiträge zur Wahrscheinlichkeitstheorie. In *F. Hausdorff zum Gedächtnis*, Bd. 1. Hrsg E. Brieskorn. Braunschweig – Wiesbaden, pp. 31 – 70.

Hausdorff F. (2006), *Ges. Werke*, Bd. 5. Berlin.

Hilbert D. (1901), Mathematische Probleme. *Ges. Abh.*, Bd. 3. Berlin, 1970, pp. 290 – 329.

Hochkirchen T. (1999), *Die Axiomatisierung der Wahrscheinlichkeitsrechnung*. Göttingen.

Kolmogorov A. N. (1933, in German), *Foundations of the Theory of Probability*. New York, 1950, 1956.

Krengel U. (June 2011), On the contributions of G. Bohlmann to probability theory. *J. Electron. Hist. Prob. Stat.*, www.jehps.net

Shafer G., Vovk V. (2001), *Probability and Finance*. New York.

Vovk V. G., Shafer G. R. (2003), Kolmogorov's contributions to the foundations of probability. *Problems of Information Transmission*, vol. 39, pp. 21 – 31.

Bibliography

The Supplement has its own Bibliography.

My translations of many Russian sources were privately printed in separate books or collections and are available in a few libraries and at www.sheynin.de

Abbreviation: AHES = *Arch. Hist. Ex. Sci., Hist. Scient. = Historia Scientiarum* (Tokyo),
ISR = *Intern. Stat. Rev.*, JNÖS = *Jahrbücher f. Nationalökonomie u. Statistik*, MS =
Matematich. Sbornik, OC = *Oeuvr. Compl.*

- Aaboe, A., De Solla Price, D. J.** (1964), Qualitative measurements in antiquity. In *L'aventure de la science* (Melanges A. Koyre, t. 1). Paris, pp. 1 – 20.
- Abbe, E.** (1863), Über die Gesetzmäßigkeit in der Verteilung der Fehler bei Beobachtungsreihen. *Ges. Abh.*, Bd. 2, 1989, pp. 55 – 81.
- Achenwall, G.** (1752), *Staatsverfassung der europäischen Reiche im Grundrisse*. Göttingen, this being the second edition of *Abriß der neuesten Staatswissenschaft*, etc. Göttingen, 1749. Many later editions up to 1798 but in 1768 the title changed once more.
--- (1763), *Staatsklugheit und ihren Grundsätzen*. Göttingen. Fourth edition, 1779.
- Al-Beruni (Al-Biruni)** (1887), *India*, vols 1 – 2. Delhi, 1964.
--- (1934), *The Book of Instruction in the Elements of the Art of Astrology*. London.
--- (1967), *Determination of Coordinates of Cities*. Beirut.
- Ambartzumian, R. V.** (1999, in Russian), Stochastic geometry. In Prokhorov (1999b, 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. Also in Anderson (1963, Bd. 2, pp. 530 – 538).
--- (1963), *Ausgewählte Schriften*, Bde 1 – 2. Tübingen.
- Anonymous** (1735), Géométrie. *Hist. Acad. Roy. Sci. avec les Mém. Math. et Phys.*, pp. 43 – 45 of the *Histoire*.
- Anonymous** (1839), Introduction. *J. Stat. Soc. London*, vol. 1, pp. 1 – 5.
- Anonymous** (1910, in Russian), The Holy Synod and Tolstoy. Newspaper *Rech*, 8 Nov. 1910, p. 3.
- Arbuthnot, J.** (1712), An argument for divine Providence taken from the constant regularity observed in the birth of both sexes. In M. G. Kendall & Plackett (1977, pp. 30 – 34).
- Aristotle** (1908 – 1930, 1954), *Works*, vols 1 – 12. London. I am referring to many treatises from that source although the authorship of two of them (*Problemata* and *Magna Moralia*) is doubtful. There is also a new edition of Aristotle (Princeton, 1984, in two volumes).
- Arnould, A., Nicole, P.** (1662), *L'art de penser*. Published anonymously. Paris, 1992. English translation: Edinburgh – London, 1850.
- Babbage, C.** (1857), On tables of the constants of nature and art. *Annual Rept Smithsonian Instn* for 1856, pp. 289 – 302. Abstract published in 1834.
- Bachelier, L.** (1900), *Théorie de la spéculation*. Paris, 1995. Engl. transl.: Princeton, 2006.
- Baer, K.** (1860 – 1875), *Issledovania o Sostoianii Rybolovstva v Rossii* (Investigations on the State of Fishing in Russia), vols 1 – 9. Petersburg.
- Baily, Fr.** (1835), *An Account of the Revd John Flamsteed*. London.
- Bauer, R. K.** (1955), Die Lexische Dispersionstheorie in ihren Beziehungen zur modernen statistischen Methodenlehre etc. *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 and 296 – 325. Communicated and commented on by R. Price. German transl.: Leipzig, 1908. Reprint of pt. 1: *Biometrika*, vol. 45, 1958, pp. 293 – 315, and E. S. Pearson & Kendall (1970, pp. 131 – 153).
- Bellhouse, D. R.** (1989), A manuscript on chance written by J. Arbuthnot. *ISR*, vol. 57, pp. 249 – 259.
--- (2000), De Vetula, a medieval manuscript containing probability calculations. *ISR*, vol. 68, pp. 123 – 136.
--- (2005), Decoding Cardano's Liber de Ludo alea. *Hist. Math.*, vol. 32, pp. 180 – 202.
- Belvalkar, S. K., Ranade, R. D.** (1927), *History of Indian Philosophy*, vol. 2. Poona.
- Benjamin, M.** (1910), Newcomb. In *Leading American Men of Science*. Ed., D. S. Jordan. New York, pp. 363 – 389.

- Bernoulli, Daniel** (1735), Recherches physiques et astronomiques ... Quelle est la cause physique de l'inclinaison des plans des planètes... In Bernoulli, D. (1987, pp. 303 – 326). [All the other memoirs are reprinted in Bernoulli, D. (1982). When possible, I provide their titles in translation from Latin.]
- (1738; 1982, pp. 223 – 234, in Latin), Exposition of a new theory on the measurement of risk. *Econometrica*, vol. 22, 1954, pp. 23 – 36.
- (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.
- (1768a; 1982, pp. 276 – 287), De usu algorithmi infinitesimalis in arte coniectandi specimen.
- (1768b; 1982, pp. 290 – 303), De duratione media matrimoniorum, etc. [On the mean duration of marriages for any ages of husband and wife and on other bordering issues]. Russian transl. in Ptukha (1955, pp. 453 – 464).
- (1769, in Latin), Manuscript, same title as in 1778. In English, 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; 1982, pp. 361 – 375, in Latin), The most probable choice between several discrepant observations and the formation therefrom of the most likely induction. *Biometrika*, vol. 48, 1961, pp. 3 – 13, with translation of Euler (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** (manuscript; 1975, partial publ.), Meditationes. In Bernoulli, J. (1975, pp. 21 – 90).
- (1713), *Ars coniectandi*. Reprint: Bernoulli J. (1975, pp. 107 – 259). German transl.: *Wahrscheinlichkeitsrechnung*. Leipzig, 1899. Its reprint: Frankfurt/Main, 1999. References in text to German translation.
- (1975), *Werke*, Bd. 3. Basel. Includes reprints of several memoirs of other authors and commentaries.
- (1986), *O Zakone Bolshikh Chisel* (On the Law of Large Numbers). Moscow. Ed., Yu. V. Prokhorov. Contains commentaries.
- (2005), *On the Law of Large Numbers*. Berlin, this being a translation of pt. 4 of the *Ars Coniectandi*.
- Bernoulli, Johann III** (1785), Milieu à prendre entre les observations. *Enc. Méthodique*. Mathématiques, t. 2. Paris, pp. 404 – 409.
- Bernoulli, Niklaus** (1709), *De Usu Artis Coniectandi in Jure*. Reprint: Bernoulli J. (1975, pp. 289 – 326).
- Bernstein, S. N.** (1926), Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Math. Annalen*, Bd. 97, pp. 1 – 59.
- (1928, in Russian), The present state of the theory of probability and its applications. In Bernstein (1964, pp. 217 – 232).
- (1945, in Russian), On Chebyshev's work on the theory of probability. *Ibidem*, pp. 409 – 433.
- (1964), *Sobranie Sochineniy* (Coll. Works), vol. 4. Moscow.
- Bertrand, J.** (1887), Sur les épreuves répétées. *C. r. Acad. Sci. Paris*, t. 105, pp. 1201 – 1203.
- (1888a), *Calcul des probabilités*. 2nd edition, 1907. Reprints: New York, 1970, 1972. Second edition practically coincides with the first one.
- (1888b), Sur l'évaluation a posteriori de la confiance méritée par la moyenne d'une série de mesures. *C. r. Acad. Sci. Paris*, t. 106, pp. 887 – 891.
- (1888c), Sur l'erreur à craindre dans l'évaluation des trois angles d'un triangle. *Ibidem*, pp. 967 – 970.
- Bessel, F. W.** (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 (1876, Bd. 3, pp. 300 – 304).
- (read 1832), Über ein gegenwärtigen Standpunkt der Astronomie. In author's book (1848), *Populäre Vorlesungen über wissenschaftliche Gegenstände*. Hamburg, pp. 1 – 33.

- (1838a), Untersuchung über die Wahrscheinlichkeit der Beobachtungsfehler. In Bessel (1876, Bd. 2, pp. 372 – 391).
- (1838b), *Gradmessung in Ostpreussen*. Berlin.
- (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: *L'Institut*, t. 7, No. 286, pp. 187 – 189.
- (1840a), Principe nouveau du calcul des probabilités avec ses applications aux sciences d'observation. *Soc. Philomat. Paris, Extraits*, sér. 5, pp. 37 – 43. Also: *L'Institut*, t. 8, No. 333, pp. 167 – 169.
- (1840b), Quelques propriétés des moyens arithmétiques de puissances de quantités positives. *Soc. Philomat. Paris, Extraits*, sér. 5, pp. 67 – 68. Also: *L'Institut*, t. 8, No. 342, pp. 216 – 217.
- (1845), De la loi de multiplication et de la durée des familles. Reprint: D. G. Kendall (1975, pp. 251 – 253).
- (1852), Sur la probabilité des erreurs d'après la méthode des moindres carrés. *J. Math. Pures Appl.*, sér. 1, t. 17, pp. 33 – 78. Also: *Mém. pres. Acad. Sci. Inst. France*, sér. 2, t. 15, 1858, pp. 615 – 663.
- (1853), Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés. *C. r. Acad. Sci. Paris*, t. 37, pp. 309 – 324. Also: *J. Math. Pures Appl.*, sér. 2, t. 12, 1867, pp. 158 – 176.
- (1855), Sur un principe que M. Poisson avait cru découvrir et qu'il avait appelé Loi des grands nombres. *C. r. Acad. Sci. Morales et Politiques*, ser. 3, t. 11, pp. 379 – 389. Also: *J. Soc. Stat. Paris*, 1876, pp. 199 – 204.
- (1874), Sur une question de probabilités. *Bull. Soc. Math. France*, t. 2, pp. 153 – 154.
- (1875), Application d'un théorème nouveau du calcul des probabilités. *C. r. Acad. Sci. Paris*, t. 81, pp. 417 – 423. Also: *Bull. Math. Astr.*, t. 9, pp. 219 – 225.
- Biermann, K.-R.** (1955), Über eine Studie von Leibniz zu Fragen der Wahrscheinlichkeitsrechnung. *Forschungen und Fortschritte*, Bd. 29, No. 4, pp. 110 – 113.
- (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. 2nd edition.
- (1855), Sur les observatoires météorologiques permanents que l'on propose d'établir en divers points de l'Algérie. *C. r. Acad. Sci. Paris*, t. 41, pp. 1177 – 1190.
- Birg, S.**, Editor (1986), *Ursprünge der Demographie in Deutschland. Leben und Werke J. P. Süßmilch's*. [Coll. Papers.] Frankfurt/Main.
- Black, W.** (1788), *Comparative View of the Mortality of the Human Species*. London.
- (1782, in English), *Esquisse d'une histoire de la médecine*. Paris, 1798.
- Block, M.** (1878), *Traité théorique et pratique de statistique*. Paris, 1886.
- Boltzmann, L.** (1868), Studien über das Gleichgewicht der lebenden Kraft. In Boltzmann (1909, Bd. 1, pp. 49 – 96).
- (1871), Analytischer Beweis des zweiten Hauptsatzes. *Ibidem*, pp. 288 – 308.
- (1872), Weitere Studien über das Wärmegleichgewicht. *Ibidem*, pp. 316 – 402.
- (1878), Weitere Bemerkungen über einige Probleme. *Ibidem*, Bd. 2, pp. 250 – 288.
- (1886), Der zweite Hauptsatz der mechanischen Wärmetheorie. In Boltzmann (1905, pp. 25 – 50).
- (1887), Über die mechanischen Analogien des zweiten Hauptsatzes. In Boltzmann (1909, Bd. 3, pp. 258 – 271).
- (1895), On certain questions of the theory of gases. *Ibidem*, pp. 535 – 544.
- (1895 – 1899), *Vorlesungen über Gastheorie*, Bde 1 – 2. Leipzig.
- (1896), Entgegnung auf die [...] Betrachtungen des Hrn Zermelo. In Boltzmann (1909, Bd. 3, pp. 567 – 578).
- (1904a), Entgegnung auf einen von [...] Ostwald [...] gehaltenen Vortrag. In Boltzmann (1905, pp. 364 – 378).
- (1904b), *Vorlesungen über die Prinzipie der Mechanik*, Tl. 2. Leipzig.
- (1905), *Populäre Schriften*. Leipzig, 1925. Later edition, 1979.
- (1909), *Wissenschaftliche Abhandlungen*, Bde 1 – 3. Leipzig.
- Boole, G.** (1851), On the theory of probabilities. In Boole (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.

- Bortkiewicz, L. von (Bortkevich, V. I.)** (1894 – 1896), Kritische Betrachtungen zur theoretischen Statistik. JNÖS, 3. Folge, Bde 8, 10, 11, pp. 641 – 680, 321 – 360, 701 – 705.
 --- (1898a), *Das Gesetz der kleinen Zahlen*. Leipzig.
 --- (1898b), Das Problem der Russischen Sterblichkeit. *Allg. stat. Archiv*, Bd. 5, pp. 175 – 190, 381 – 382.
 --- (1903, Russian), The theory of probability and the struggle against sedition. *Osvobozhdenie*. Stuttgart, book 1, pp. 212 – 219. Signed “B”. Only contained in some copies of that periodical.
 --- (1904), Anwendungen der Wahrscheinlichkeitsrechnung auf Statistik. *Enc. math. Wiss.*, Bd. 1, pp. 821 – 851.
 --- (1917), *Die Iterationen*. Berlin.
 --- (1930), Lexis und Dormoy. *Nordic Stat. J.*, vol. 2, pp. 37 – 54.
- Boscovich, R.** (1758). *Philosophiae Naturalis Theoria*. Latin – English edition: Chicago – London, 1922. English translation from the edition of 1763: *Theory of Natural Philosophy*. Cambridge, Mass., 1966.
- Bouillaud, J.** (1836), *Essai sur la philosophie médicale*. Paris.
- Boyle, R.** (posthumous, 1772), A physico-chymical essay. *Works*, vol. 1. Sterling, Virginia, 1999, pp. 359 – 376.
- Brendel, M.** (1924), Über die astronomische Arbeiten von Gauss. In Gauss, W-11, Tl. 2, Abt. 3. Separate paging.
- Brownlee, J.** (1915), Historical note on Farr’s theory of epidemic. *Brit. Med. J.*, vol. 2, pp. 250 – 252.
- Bru, B.** (1981), Poisson, le calcul des probabilités et l’instruction publique. In Métivier et al (1981, pp. 51 – 94).
 --- (1988), Laplace et la critique probabiliste des mesures géodésiques. In Lacombe, H., Costabel, P. (1988, pp. 223 – 244).
 --- (1991), A la recherche de la démonstration perdue de Bienaymé. *Math. Inf. Sci. Hum.*, 29^e année, No. 114, pp. 5 – 17.
- Bru, B., Bru, M.-F., Bienaymé, O.** (1997), La statistique critiquée par le calcul des probabilités. *Rev. Hist. Math.*, t. 3, pp. 137 – 239.
- Bru, B., Jongmans, F.** (2001), Bertrand. In Heyde et al (2001, pp. 185 – 189).
- Brush, S. G.** (1968), Brownian movement from Brown to Perrin. AHES, vol. 5, pp. 1 – 36. Reprint: M. G. Kendall et al (1977, pp. 347 – 382).
- Budd, W.** (1849), *Malignant Cholera*. London.
- Bühler, G.**, Editor (1886), *Laws of Manu*. Oxford, 1967.
- Buffon, G. L. L.** (1777), Essai d’arithmétique morale. In Buffon (1954, pp. 456 – 488).
 --- (1954), *Ceuvres philosophiques*. Paris. Editors, J. Piveteau, M. Fréchet, C. Bruneau.
- Bull J. P.** (1959), Historical development of clinical therapeutic trials. *J. Chronic Diseases*, vol. 10, pp. 218 – 248.
- Buniakovsky, V. Ya.**, all publications in Russian (1836 – 1837), On the application of the analysis of probabilities to determining the approximate values of transcendental numbers. *Mém. Imp. Acad. Sci. St. Péters.*, Sci. math., phys. et nature, t. 3, No. 4, pp. 457 – 467 and No. 5, pp. 517 – 526.
 --- (1846), *Osnovania Matematicheskoi Teorii Veroiatnostei* (Principles of the Math. Theory of Probability). Petersburg. Abbreviated “Detailed Contents” and chapter “History of Probability” reprinted in Prokhorov (1999b, pp. 867 – 869 and 863 – 866).
 --- (1866a), Essay on the law of mortality in Russia and on the distribution of the Orthodox believers by ages. *Zapiski Imp. Akad. Nauk St. Petersb.*, vol. 8, No. 6. Separate paging.
 --- (1866b), Tables of mortality and of population for Russia. *Mesiatseslov* (Calendar) for 1867. Petersburg. Supplement, pp. 3 – 53.
 --- (1874), Anthropological [Anthropometric] studies. *Zapiski Imp. Akad. Nauk St. Petersb.*, vol. 23, No. 5. Separate paging.
 --- (1875a), On a problem concerning the partition of numbers. *Ibidem*, vol. 25, No. 1. Separate paging.
 --- (1875b), On the probable number of men in the contingents of the Russian army in 1883, 1884 and 1885. *Ibidem*, No. 7. Separate paging.
 --- (1880), On maximal quantities in issues concerning moral benefit. *Ibidem*, vol. 36, No. 1. Separate paging.
- Burov, V. G., Viatkin, R. V., Titarenko, M. A.**, Editors (1972 – 1973), *Drevnekitaiskaia Filosofia* (Ancient Chinese Philosophy), vols 1 – 2. Moscow.
- Bursill-Hall, P., Editor** (1993), *R. J. Boscovich. His Life and Scientific Work*. Rome.
- Buys Ballot C. H. D.** (1847), *Les changements périodiques de temperature*. Utrecht.

- (1850), Die periodischen [...] Änderungen der Temperatur. *Fortschritte Phys.*, Bd. 3 für 1847, pp. 623 – 629.
- Camp, B. H.** (1933), Karl Pearson and mathematical statistics. *J. Amer. Stat. Assoc.*, vol. 28, pp. 395 – 401.
- Campbell, L., Garnett, W.** (1882), *Life of Maxwell*. London, 1884, New York – London, 1969.
- Cardano, G.** (ca. 1564, publ. 1663, in Latin), *The Book on Games of Chance*, 1953. New York, 1961.
- (1575, in Latin), *The Book of My Life*, 1931. New York, 1962.
- Cauchy, A. L.** (1853a), Sur la nouvelle méthode d'interpolation comparée à la méthode des moindres carrés. OC, t. 12. Paris, 1900, pp. 68 – 79.
- (1853b), Sur les résultats moyens d'observations de même nature, et sur les résultats les plus probables. Ibidem, pp. 94 – 104.
- (1853c), Sur la probabilité des erreurs qui affectent des résultats moyens d'observations de même nature. Ibidem, pp. 104 – 114.
- (1853d), Sur les résultats moyens d'un très grand nombre d'observations. Ibidem, pp. 125 – 130.
- Celsus, A. C.** (1st century, 1935), *De Medicina*, vol. 1. London. In English.
- Chadwick, E.** (1842), *Report on the Sanitary Condition of the Labouring Population*. Edinburgh, 1965.
- Chapman, S.** (1941), *Halley As a Physical Geographer*. London.
- Chebyshev, P. L.** (1846), Démonstration élémentaire d'une proposition générale de la théorie des probabilités. *J. reine u. angew. Math.*, Bd. 33, pp. 259 – 267.
- (1867), Des valeurs moyennes. *J. Math. Pures et Appl.*, t. 12, pp. 177 – 184.
- (1874), Sur les valeurs limites des intégrales. Ibidem, t. 19, pp. 157 – 160.
- (Lectures 1879/1880), *Teoria Veroiatnostei* (Theory of Probability). Moscow – Leningrad, 1936.
- (1887a, in Russian), Sur les résidus intégraux qui donnent des valeurs approchées des intégrales. *Acta Math.*, t. 12, 1888 – 1889, pp. 287 – 322.
- (1887b, in Russian), Sur deux théorèmes relatifs aux probabilités. Ibidem, t. 14, 1890 – 1891, pp. 305 – 315.
- (1899 – 1907), *Oeuvres*, tt. 1 – 2. Pétersbourg. Reprint: New York, 1962.
- (1944 – 1951), *Polnoe Sobranie Sochineniy* (Complete Works), vols 1 – 5. Moscow – Leningrad.
- Chirikov, M. V., Sheynin, O. B.** (1994), see Sheynin (1994e).
- Chuprov, A. A.** (1905), Die Aufgaben der Theorie der Statistik. *Schmollers Jahrb. f. Gesetzgebung, Verwaltung u. Volkswirtschaft in Dtsch. Reiche*, Bd. 29, pp. 421 – 480. Quoted from its Russian translation (Chuprov 1960, pp. 43 – 90).
- (1909), *Ocherki po Teorii Statistiki* (Essays on the Theory of Statistics). Second edition, 1910. Moscow, 1959.
- (1916, in Russian), On the expectation of the coefficient of dispersion. *Izvestia Imp. Akad. Nauk*, vol. 10, pp. 1789 – 1798.
- (1918 – 1919), Zur Theorie der Stabilität statistischer Reihen. *Skand. Aktuarietidskr.*, t. 1 – 2, pp. 199 – 256, 80 – 133.
- (1925a, in German), *Principles of the Mathematical Theory of Correlation*. London, 1939. Russian version: 1926.
- (1925b), Review of Markov (1924). In Ondar (1977a/1981, pp. 154 – 157).
- (1926), Teorien för statistiska räckors stabilitet. *Nord. Stat. Tidskrift*, Bd. 5, pp. 195 – 212. Russian translation: Chuprov (1960, pp. 224 – 239).
- (1960), *Voprosy Statistiki* (Issues in Statistics). Reprints and/or translations of papers. Moscow, 1960.
- Cicero(n), M. T.** (1991), *Über die Wahrsagung*. München – Zurich.
- (1997), *Nature of the Gods*. Oxford.
- Clausius, R.** (1849), Über die Natur derjenigen Bestandteile der Erdatmosphäre. *Annalen Phys. Chem.*, Bd. 76 (152), pp. 161– 188.
- (1857), Über die Art der Bewegung welche wir Wärme nennen. In Clausius (1867, pp. 229 – 259).
- (1858), Über die mittlere Länge der Wege ... bei der Molekularbewegung. Ibidem, pp. 260 – 276.
- (1862), Über die Wärmeleitung gasförmiger Körper. Ibidem, pp. 277 – 326.
- (1867), *Abhandlungen über die mechanische Wärmetheorie*, Abt. 2. Braunschweig.
- (1889 – 1891), *Die kinetische Theorie der Gase*. Braunschweig.

- Cohen J., Chesnick E. I.** (1970), The doctrine of psychological chances. *Brit. J. Psychol.*, vol. 61, pp. 323 – 334.
- Cohen J., Chesnick E. I., Haran D.** (1971), Evaluation of compound probabilities in sequential choice. *Nature*, vol. 232, pp. 414 – 416.
- Commelin, C.** (1693), *Beschryvinge der Stadt Amsterdam*. Amsterdam.
- Comte A.** (1830 – 1842), *Cours de philosophie positive*, tt. 1 – 6. Paris, 1893, t. 3.
- Condamine, C. M. de la** (1751), *Mesure des trois premiers degrés du méridien*. Paris.
- (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.** (1784), Sur le calcul des probabilités. *Hist. Acad. Roy. Sci. Paris 1781 avec Mém. Math. et Phys. pour la même année*. Paris, 1784, pp. 707 – 728.
- (1785), Eloge de M. [D.] Bernoulli, *Hist. Acad. Roy. Sci. Paris 1782 avec Mém. Math. et Phys.*, pp. 82 – 107 of the *Histoire*. Reprint: *Oeuvres*, t. 2. Paris, 1847, pp. 545 – 580.
- (1795), *Esquisse d'un tableau historique des progrès de l'esprit humain siuvi de Fragment sur l'Atlantide*. Paris, 1988.
- Coolidge, J. L.** (1926), Adrain and the beginnings of American mathematics. *Amer. Math. Monthly*, vol. 33, No. 2, pp. 61 – 76.
- Cotes, R.** (1722), Aestimatio errorum in mixta mathesis per variationes partium trianguli plani et sphaerici. *Opera Misc.* London, 1768, pp. 10 – 58.
- Cournot, A. A.** (1838), Sur l'applications du calcul des chances à la statistique judiciaire. *J. Math. Pures et Appl.*, sér. 1, t. 3, pp. 257 – 334.
- (1843). *Exposition de la théorie des chances et des probabilités*. Paris, 1984. Editor B. Bru.
- (1851), *Essai sur les fondements de nos connaissances et sur les caractères de la critique philosophique*. Paris, 1975.
- (1861), *Traité de l'enchaînement des idées fondamentales dans les sciences et dans l'histoire*. Paris, 1982.
- (1875), *Matérialisme, vitalisme, rationalisme*. Paris, 1979.
- Courtault, J.-M. et al** (2000), Bachelier. On the centenary of “Théorie de la spéculation”. *Math. Finance*, vol. 10, pp. 341 – 353.
- Cramér, H.** (1946), *Mathematical Methods of Statistics*. Princeton. 13th printing, 1974.
- Cubranic, N.** (1961), *Geodetski rad R. Boskovicica*. Zagreb.
- Czuber, E.** (1891), Zur Kritik einer Gauss'schen Formel. *Monatshefte Math. Phys.*, Bd. 2, pp. 459 – 464.
- (1903), *Wahrscheinlichkeitsrechnung und ihre Anwendung*, Bd. 1. New York, 1968.
- (1921), *Die statistischen Forschungsmethoden*. Wien.
- D'Alembert, J. Le Rond** (1754), Croix ou pile. *Enc. ou dict. raisonné des sciences, des arts et des métiers*, t. 4. Stuttgart, 1966, pp. 512 – 513.
- (1759), Essai sur les éléments de philosophie. OC, t. 1, pt. 1. Paris, 1821, pp. 116 – 348.
- (1761), Sur l'application du calcul des probabilités à l'inoculation de la petite vérole. *Opusc. math.*, t. 2. Paris, pp. 26 – 95.
- (1768a), Doutes et questions sur le calcul des probabilités. In author's book *Mélanges de littérature, d'histoire et de philosophie*, t. 5. Amsterdam, pp. 239 – 264.
- (1768b), Sur la durée de la vie. *Opusc. Math.*, t. 4. Paris, pp. 92 – 98.
- (1768c), Sur un mémoire de M. Bernoulli concertant l'inoculation. Ibidem, pp. 98 – 105.
- (1768d), Sur le calcul des probabilités. Ibidem, pp. 283 – 310.
- D'Amador R.** (1837), *Mémoire sur le calcul des probabilités appliqué à la médecine*. Paris.
- Darboux G.** (read 1901; 1902), Eloge historique de Bertrand. In author's *Eloges académiques et discours*. Paris, 1912, pp. 1 – 60.
- Darwin, C.** (1859), *Origin of Species*. London – New York, 1958. [Manchester, 1995.]
- (1868), *The Variation of Animals and Plants under Domestication*, vols 1 – 2. London, 1885. [London, 1988.]
- (1876), *The Effects of Cross and Self-Fertilisation in the Vegetable Kingdom*. London, 1878. [London, 1989.]
- (1881), *The Formation of Vegetable Mould*. London, 1945. [London, 1989.]
- (1887), *Life and Letters*. New York – London, 1897, vols 1 – 2. [New York, 1969.]
- (1903), *More Letters*, vol. 1. London.
- David, F. N.** (1962), *Games, Gods and Gambling*. London.

- David, F. N., Neyman, J.** (1938), Extension of the Markoff theorem on least squares. *Stat. Res. Memoirs*, vol. 2, pp. 105 – 117.
- David, H. A.** (1963), *The Method of Paired Comparisons*. London – New York, 1988.
- David, H. A., Edwards, A. W. F.** (2001), *Annotated Readings in the History of Statistics*. New York.
- Davidov, A. Yu.** (1854?), *Lektsii Matematicheskoi Teorii Veroiatnostei* (Lectures on Math. Theory of Probability). N. p., n. d.
- (1857), Theory of means. *Rechii i Otchet Proiznesennye v Torzestvenom Sobranii Mosk. Univ.* (Orations and Report Made at the Grand Meeting of Moscow Univ.). Moscow, first paging.
- [1885], *Teoria Veroiatnostei, 1884 – 1885* (Theory of Probability). N. p., n. d.
- Daw, R. H.** (1980), J. H. Lambert, 1727 – 1777. *J. Inst. Actuaries*, vol. 107, pp. 345 – 350.
- DeCandolle, Aug. P.** (1832), *Physiologie végétale*, tt. 1 – 3. Paris.
- 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. *ISR*, vol. 52, 1984, pp. 236 – 262. Commentary (A. Hald): *Ibidem*, pp. 229 – 236.
- (1718), *Doctrine of Chances*. Later editions: 1738, 1756. References in text to reprint of last edition: New York, 1967.
- (1725), *Treatise on Annuities on lives*. London. Two later editions: 1743 and 1756, incorporated in the last edition of his *Doctrine*, pp. 261 – 328. On p. xi of the same source (*Doctrine* 1756), the anonymous Editor stated that the appended *Treatise* was its improved edition [of 1743]. German transl.: Wien, 1906.
- (1730), *Miscellanea Analytica de Seriebus et Quadraturis*. London. French transl.: Paris, 2009.
- (1733, in Latin), Transl. by author: A method of approximating the sum of the terms of the binomial $(a + b)^n$ expanded into a series from whence are deduced some practical rules to estimate the degree of assent which is to be given to experiments. Incorporated in subsequent editions of the *Doctrine* (in 1756, an extended version, pp. 243 – 254).
- (1756), This being the last edition of the *Doctrine*.
- De Montessus, R.** (1903), Un paradoxe du calcul des probabilités. *Nouv. Annales Math.*, sér. 4, t. 3, pp. 21 – 31.
- 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 (the whole issue). Paris, 1978. Reprint of the edition of 1647.
- De Witt, J.** (1671, in vernacular), Value of life annuities in proportion to redeemable annuities. In Hendriks (1852, pp. 232 – 249).
- Dietz, K.** (1988), The first epidemic model: historical note on P. D. Enko. *Austr. J. Stat.*, vol. 30A, pp. 56 – 65.
- Dietz, K., Heesterbeek, J. A. P.** (2000), D. Bernoulli was ahead of modern epidemiology. *Nature*, vol. 408, pp. 513 – 514.
- (2002), D. Bernoulli's epidemiological model revisited. *Math. Biosciences*, vol. 180, pp. 1 – 21.
- Dormoy E.** (1874), Théorie mathématique des assurances sur la vie. *J. des actuaries française*, t. 3, pp. 283 – 299, 432 – 461.
- (1878), Same title, t. 1. Paris. Incorporates his paper of 1874.
- Double F. J., Dulong P. L., Larrey F. H., Poisson S. D.** (1835), Review of contribution on medical statistics. *C. r. Acad. Sci. Paris*, t. 1, pp. 167 – 177.
- Dove, H. W.** (1839), Über die nicht periodischen Änderungen der Temperaturvertheilung auf der Oberfläche der Erde. *Abh. Kgl. Preuss. Akad. Wiss. Berlin*, Phys. Abh. 1838, pp. 285 – 415.
- (1850), Über Linien gleicher Monatswärme. *Ibidem*, Phys. Abh. 1848, pp. 197 – 228.
- 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.
- Dutka, J.** (1985), On the problem of random flights. *AHES*, vol. 32, pp. 351 – 375.
- (1988), On the St. Petersburg paradox. *AHES*, vol. 39, pp. 13 – 39.

- (1996), On Gauss's priority in the discovery of the method of least squares. *AHES*, vol. 49, pp. 355 – 370.
- Ebbinghaus H.** (1908), *Abriss der Psychologie*. Leipzig.
- 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.
- Edwards, A. W. F.** (1987), *Pascal's Arithmetical Triangle*. Baltimore, 2002.
- Ehrenfest, P. & T.** (1907), Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem. In Ehrenfest, P. (1959), *Coll. Scient. Papers*. Amsterdam, pp. 146 – 149.
- 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. Reprint: Ku (1969, pp. 265 – 274).
- (1974), Pearson. *Dict. Scient. Biogr.*, vol. 10, pp. 447 – 473.
- Emeliakh L. I.** (1954, in Russian), The case of A. A. Markov's excommunication from the Church. *Voprosy Religii i Ateizma*, vol. 2, pp. 397 – 411.
- Encke, J.F.** (1851), Über die Bestimmung der elliptischen Elemente bei Planetenbahnen. *Abh. Kgl. Preuss. Akad. Wiss. Berlin* for 1849, pp. 1 – 68 of second paging.
- Eneström G.** (1896, in Swedish), Sur la méthode de J. de Witt (1671) pour le calcul de rentes viagères. *Archief voor de verzekerings-wetenschap*, t. 3, 1897, pp. 62 – 68.
- Engels, F.** (compiled 1873 – 1882, publ. 1925), *Dialektik der Natur*. Berlin, 1971. [Berlin, 1985.]
- Enko, P. D.** (1889, in Russian), On the course of epidemics of some infectious diseases. *Intern. J. Epidemiology*, vol. 18, 1989, pp. 749 – 755. Partial transl. by K. Dietz.
- Erdélyi, A.** (1956), *Asymptotic Expansions*. New York.
- Erman, A.,** Editor (1852), *Briefwechsel zwischen W. Olbers und F. W. Bessel*, Bde 1 – 2. Leipzig.
- Ermolaeva, N. S.** (1987, in Russian), On an unpublished course on the theory of probability by Chebyshev. *Voprosy Istorii Estestvozn. i Tekhniki* No. 4, pp. 106 – 112.
- Euler, L.** (1748 in Latin), *Introduction to Analysis of the Infinite*, Book 1. New York, 1988.
- (1749), Recherches sur la question des inégalités du mouvement de Saturne et de Jupiter. *Opera Omnia*, ser. 2, t. 25. Zürich, 1960, pp. 45 – 157.
- (1755), Elémens de la trigonométrie sphéroïdique tirés de la méthode des plus grands et plus petits. *Ibidem*, ser. 1, t. 27. Zürich, 1954, pp. 309 – 339.
- (1767), Recherches générales sur la mortalité et la multiplication du genre humain. *Ibidem*, ser. 1, t. 7. Leipzig, 1923, pp. 79 – 100. His manuscript Sur multiplication du genre humaine is also there, pp. 545 – 552.
- (1770), Expositio methodorum, cum pro determinanda parallaxi solis [...]. *Ibidem*, ser. 2, t. 30. Zürich, 1964, pp. 153 – 231.
- (1776), Éclaircissements sur les établissemens publics [...] avec la description d'une nouvelle espèce de Tontine [...]. *Ibidem*, ser. 1, t. 7, pp. 181 – 245.
- (1778, in Latin). [Commentary to D. Bernoulli (1778).] English transl.: 1961, and its reprint 1970, are published together with the transl. and reprint of D. Bernoulli.
- Faraday, M.** (1991 – 2008), *Correspondence*, vols 1 – 5. London.
- Farr, W.** (1885), *Vital Statistics*. London.
- Fechner, G. T.** (1855), *Über die physikalische und philosophische Atomenlehre*. Leipzig, 1864.
- (1860), *Elemente der Psychophysik*, Bde 1 – 2. Leipzig. Third edition, 1907, its reprint, 1964.
- (1877), *In Sachen der Psychophysik*. Leipzig.
- (1887), Über die Methode der richtigen und falschen Fälle. *Abh. Kgl. Sächsische Ges. Wiss.*, Bd. 13 (22), pp. 109 – 312.
- (1897), *Kollektivmasslehre*. Leipzig. Editor (and actual coauthor) G. F. Lipps.
- Fedorovitch, 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.
- Fisher, R. A.** (1922), On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc.*, vol. A222, pp. 309 – 368.
- (1936), Has Mendel's work been rediscovered? *Annals of Science*, vol. 1, pp. 115 – 137.

- (1937), Professor K. Pearson and the method of moments. *Annals Eugenics*, vol. 7, pp. 303 – 318.
- Fletcher, A., Miller, J. C. P. et al** (1962), *Index of Mathematical Tables*, vol. 1. Oxford. First edition, 1946.
- 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.
- (1829), Historical Eloge of the Marquis De Laplace. *London, Edinburgh and Dublin Phil. Mag.*, 2nd ser., vol. 6, pp. 370 – 381. The original French Eloge was published in 1831: *Mém. Acad. Roy. Sci. Inst. de France*, t. 10, pp. LXXX – CII.
- Fourier, J. B. J. et al** (1826), Rapport sur les tontines. *Oeuvr.*, t. 2. Paris, 1890, pp. 617 – 633.
- Franklin, J.** (2001), *The Science of Conjecture*. Baltimore.
- Fréchet, M.** (1949), Réhabilitation de la notion statistique de l'homme moyen. In Fréchet, *Les mathématiques et le concret*. Paris, 1955, pp. 317 – 341.
- Freud, S.** (1925), Selbstdarstellung. *Werke*, Bd. 14. Frankfurt/Main, 1963, pp. 31 – 96.
- Freudenthal, H.** (1951), Das Peterburger Problem in Hinblick auf Grenzwertsätze der Wahrscheinlichkeitsrechnung. *Math. Nachr.*, Bd. 4, pp. 184 – 192.
- (1961), 250 years of mathematical statistics. In *Quantitative Methods in Pharmacology*. Amsterdam, 1961, pp. xi – xx. Editor H. De Jonge.
- (1971), Cauchy. *Dict. Scient. Biogr.*, vol. 3, pp. 131 – 148.
- Fuss, P. N.** (1843), *Correspondance mathématique et physique de quelques célèbres géomètres du XVIII siècle*, tt. 1 – 2. New York – London, 1968.
- Galen, C.** (2nd century, 1946), *On Medical Experience*. London.
- (2nd century, 1951), *Hygiene*. Springfield, Illinois.
- Galilei, G.** (1613, in Italian), History and demonstrations concerning sunspots and their phenomena. In Galilei, *Discoveries and Opinions of Galileo*. Garden City, New York, 1957, pp. 88 – 144.
- (written ca. 1613 – 1623, publ. 1718), Sopra le scoperte dei dadi. English transl. without English title in David, F. N. (1962, pp. 192 – 195).
- (1632, in Italian), *Dialogue concerning the Two Chief World Systems*. Berkeley – Los Angeles, 1967.
- Galle, E.** (1924), Über die geodätischen Arbeiten von Gauss. In Gauss, W-11/2, Abt. 1. Separate paging.
- 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.
- (1892), *Finger Prints*. London.
- Garber, D., Zabell, S.** (1979), On the emergence of probability. *AHES*, vol. 21, pp. 33 – 53.
- Gatterer, J. C.** (1775), *Ideal einer allgemeinen Weltstatistik*. Göttingen.
- Gauss, C. F.** (1809, in Latin), *Theorie der Bewegung*, Book 2, Section 3. German transl. in Gauss (1887, pp. 92 – 117).
- (1816), Bestimmung der Genauigkeit der Beobachtungen. *Ibidem*, pp. 129 – 138.
- (1823a, in German), 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.
- (1828, in Latin), Supplement to Gauss (1823b). German transl.: *Ibidem*, pp. 54 – 91.
- (1855), *Méthode des moindres carrés*. Paris.
- (1860 – 1865), *Briefwechsel mit H. C. Schumacher*. W/Erg-5, Tle 1 – 3. Hildesheim, 1975.
- (1863 – 1930), *Werke*, Bde 1 – 12. Göttingen a.o. Reprint: Hildesheim, 1973 – 1981.
- (1887), *Abhandlungen zur Methode der kleinsten Quadrate*. Hrsg. A. Börsch & P. Simon. Latest edition: Vaduz, 1998.
- (1899), *Briefwechsel mit W. Bolyai*. W/Erg-2. Hildesheim, 1987.
- (1900 – 1909), *Briefwechsel mit Olbers*, Bd. 2. W/Erg-4, Tl. 2. Hildesheim, 1976.
- (1927), *Briefwechsel mit C. L. Gerling*. W/Erg-3. Hildesheim, 1975.
- (1995, Latin and English), *Theory of Combinations of Observations Least Subject to Error*. Includes Gauss (1823a) in German and English. Transl. with Afterword by G. W. Stewart. Philadelphia.

- Gavarret, J.** (1840), *Principes généraux de statistique médicale*. Paris.
- Gelfand, A. E., Solomon, H.** (1973), A study of Poisson's model for jury verdicts etc. *J. Amer. Stat. Assoc.*, vol. 68, pp. 271 – 278.
- Gerardy T.** (1977), Die Anfänge von Gauss's geodätische Tätigkeit. *Z. f. Vermessungswesen*, Bd. 102, pp. 1 – 20.
- 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.** (1954, in Russian). *Theory of probability*. Moscow, 1973. [Providence, RI, 2005.] First Russian edition, 1950.
- (1959), On Liapunov's work on the theory of probability. *IMI*, vol. 12, pp. 135 – 160.
- Gnedenko, B. V., Sheynin, O. B.** (1978). See Sheynin (1978a).
- Goldstein, B. R.** (1985), *The 'Astronomy' of Levi ben Gerson (1288 – 1344)*. New York.
- Gower, B.** (1993), Boscovich on probabilistic reasoning and the combination of observations. In Bursill-Hall (1993, pp. 263 – 279).
- Gowing R.** (1983), *Roger Cotes – Natural Philosopher*. Cambridge.
- 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.
- (1940), A statistical mare's nest? *J. Roy. Stat. Soc.*, vol. 103, pp. 246 – 248.
- (1941 – 1943), Medical statistics from Graunt to Farr. *Biometrika*. Reprint: Pearson & Kendall (1970, pp. 47 – 120).
- Grigeman, N. T.** (1960), Geometric probability and the number π . *Scripta Math.*, t. 25, pp. 183 – 195.
- Grodzensky, S. Ya.** (1987, in Russian), *A. A. Markov*. Moscow.
- Gumbel, E. J.** (1917), Eine Darstellung statistischer Reihe durch Euler. *Jahresber. Deutschen Mathematiker-Vereinigung*, Bd. 25, pp. 251 – 264.
- 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.
- (2002), *On the History of Series Expansions of Frequency Functions and Sampling Distributions, 1873 – 1944*. Copenhagen, this being *Roy. Danish Acad. Sci. Letters*, Mat.-fys. Meddelelser No. 49.
- Halley, E.** (1694a), *An Estimate of the Degree of Mortality of Mankind*. Baltimore, 1942.
- (1694b), Some further considerations on the Breslaw bills of mortality. *Phil. Trans. Roy. Soc.*, vol. 17 for 1693, No. 198, pp. 654 – 656. Reprint of volume: New York, 1963.
- Halperin, T.** (1988), The development of probability logic from Leibniz to MacColl. *Hist. and Phil. of Logic*, vol. 9, pp. 131 – 191.
- Hartley D.** (1749), *Observations on Man*. London. Reprint of edition of 1791: Poole, 1998.
- Harvey, W.** (1651, in Latin), *Anatomical Exercises on the Generation of Animals*. In *Great Books* (1952), vol. 28, pp. 329 – 496.
- Haushofer, D. M.** (1872), *Lehr- und Handbuch der Statistik*. Wien. Second edition, 1882.
- Hellman, C. D.** (1970), Brahe. *Dict. Scient. Biogr.*, vol. 2, pp. 401 – 416.
- Helmert, F. R.** (1868), Studien über rationelle Vermessungen im Gebiete der höhern Geodäsie. *Z. Math. Phys.*, Bd. 13, pp. 73 – 120, 163 – 186.
- (1872), *Ausgleichsrechnung nach der Methode der kleinsten Quadrate*. Leipzig. Subsequent editions: 1907 and 1924.
- (1875), Über die Berechnung des wahrscheinlichen Fehlers aus einer endlichen Anzahl wahrer Beobachtungsfehler. *Z. Math. Phys.*, Bd. 20, pp. 300 – 303.
- (1876), Über die Wahrscheinlichkeit der Potenzsummen der Beobachtungsfehler. *Ibidem*, Bd. 21, pp. 192 – 218.
- (1886), *Lotabweichungen*, Heft 1. Berlin.
- (1904), Zur Ableitung der Formel von Gauss etc. In Helmert (1993, pp. 173 – 188).
- (1905), Über die Genauigkeit der Kriterien des Zufalls bei Beobachtungsreihen. *Ibidem*, pp. 189 – 208.
- (1993), *Akademie-Vorträge*. Frankfurt am Main. Reprints of author's reports.
- 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.

- Henny, J.** (1975), Niklaus und Johann Bernoullis Forschungen auf dem Gebiet der Wahrscheinlichkeitsrechnung in ihrem Briefwechsel mit Montmort. In J. Bernoulli (1975, pp. 457 – 507).
- Henry M. Ch.** (1883), *Correspondance inédite de Condorcet et de Turgot*. Genève, 1970.
- Herschel, W.** (1783), On the proper motion of the Sun. In Herschel (1912, vol. 1, pp. 108 – 130).
- (1784), Account of some observations. *Ibidem*, pp. 157 – 166.
- (1805), On the direction and motion of the Sun. *Ibidem*, 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. [London, 2003.]
- Hertz H.** (1894), *Die Principien der Mechanik*, this being the author's *Ges. Werke*, Bd. 3. Leipzig.
- Heyde, C. C., Seneta, E.** (1977), *Bienaymé*. New York.
- , Editors (2001), *Statisticians of the Centuries*. New York.
- Hill, D., Elkin, W. L.** (1884), Heliometer-determination of stellar parallax. *Mem. Roy. Astron. Soc.*, vol. 48, pt. 1 (the whole issue).
- Hippocrates** (3rd – 4th centuries BC, 1952), Aphorisms. In *Great Books* (1952, vol. 10, pp. 131 – 144).
- Hobbes, T.** (1646, in Latin), Of liberty and necessity. *English Works*, vol. 4. London, 1840, pp. 229 – 278.
- Hogan, E. R.** (1977), R. Adrain: American mathematician. *Hist. Math.*, vol. 4, pp. 157 – 172.
- Hostinský, B.** (1932), Application du calcul des probabilités à la théorie du mouvement Brownien. *Annales Inst. H. Poincaré*, t. 3, pp. 1 – 72.
- Humboldt, A.** (1816), Sur les lois que l'on observe dans la distribution des formes végétales. *Annales Chim. Phys.*, t. 1, pp. 225 – 239.
- (1817), Des lignes isothermes. *Mém. Phys. Chim. Soc. d'Arcueil*, t. 3, pp. 462 – 602.
- (1818), De l'influence de la déclinaison du Soleil sur le commencement des pluies équatoriales. *Annales Chim. Phys.*, t. 8, pp. 179 – 190.
- (1831), *Fragmens de géologie et de climatologie asiatiques*, t. 2. Paris.
- (1845 – 1862), *Kosmos*, Bde. 1 – 5 (1845, 1847, 1850, 1858, 1862). Stuttgart. English transl. of vol. 4: New York, 1858.
- Humboldt, A., Bonpland, A. J. A.** (1815 – 1825), *Nova genera et species plantarum*, tt. 1 – 7. Paris.
- Huygens, C.** (1657), De calcul dans les jeux de hasard. In Huygens (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.
- Idelson, N. I.** (1947), *Sposob Naimenshikh Kvadratov etc* (Method of Least Squares etc). Moscow.
- Irwin, J. O.** (1978), Gosset W. S. In Kruskal & Tanur (1978, vol. 1, pp. 409 – 413).
- Ivory, J.** (1826a), On the ellipticity of the Earth as deduced from experiments with the pendulum. *London, Edinburgh and Dublin Phil. Mag.*, vol. 68, pp. 3 – 10, 92 – 101.
- (1826b), On the methods proper to be used for deducing a general formula for the length of the seconds pendulum. *Ibidem*, pp. 241 – 245.
- (1828), Letter to the Editor relating to the ellipticity of the Earth as deduced from experiments with the pendulum. *Ibidem*, New ser., vol. 3, pp. 241 – 243.
- Johansenn, W.** (1922), Biology and statistics. *Nordic Stat. J.*, vol. 1, 1929, pp. 351 – 361.
- Jorland, G.** (1987), The Saint Petersburg paradox 1713 – 1937. In *The Probabilistic Revolution*, vols 1 – 2. Cambridge (Mass.), 1987, vol. 1, pp. 157– 190. Editors, L. Kruger, G. Gigerenzer, M. S. Morgan.
- Juskevic [Youshkevitch], A. P., Winter, E., Hoffmann, P., Editors** (1959), *Die Berliner und die Petersburger Akademie der Wissenschaften in Briefwechsels Eulers*, Bd. 1. Berlin.
- Kac, M.** (1939), On a characterization of the normal distribution. In Kac, *Probability, Number Theory and Statistical Physics*. Cambridge (Mass.), 1979, pp. 77 – 79.
- Kant, I.** (1755), Allgemeine Naturgeschichte und Theorie des Himmels. *Ges. Schriften*, Bd. 1. Berlin, 1910, pp. 215 – 368.
- Kapteyn, J. C.** (1906a), *Plan of Selected Areas*. Groningen.
- (1906b), Statistical methods in stellar astronomy. [*Reports*] *Intern. Congr. Arts & Sci. St. Louis 1904*. N. p., vol. 4, 1906, pp. 396 – 425.

- (1909), Recent researches in the structure of the universe. *Annual Rept Smithsonian Instn* for 1908, pp. 301 – 319.
- 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 edition. Fifth, posthumous edition, Moscow, 1928. German edition: *Theorie und Methoden der Statistik*. Tübingen, 1913.
- Kendall, D. G.** (1975), The genealogy of genealogy: branching processes before (and after) 1873. *Bull. London Math. Soc.*, vol. 7, pp. 225 – 253.
- Kendall, M. G. (Sir Maurice)** (1956), The beginnings of a probability calculus. *Biometrika*, vol. 43, pp. 1 – 14. Reprint: E. S. Pearson & Kendall (1970, pp. 19 – 34).
- (1960), Where shall the history of statistics begin? *Biometrika*, vol. 47, pp. 447 – 449. Ibidem, pp. 45 – 46.
- (1971), The work of Ernst Abbe. *Biometrika*, vol. 58, pp. 369 – 373. Reprint: M. G. Kendall & Plackett (1977, pp. 331 – 335).
- 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. Collected reprints of papers.
- Kepler, J.** (1596, in Latin). *Mysterium cosmographicum. Ges. Werke*, Bd. 8. Munich, 1963, pp. 7 – 128, this being a reprint of second edition (1621, with additions to many chapters). English transl: New York, 1981.
- (1606, in Latin), *Über den neuen Stern im Fuß des Schlangenträger*. Würzburg, 2006.
- (1609, in Latin), *New Astronomy*. Cambridge, 1992. Double paging.
- (1610), Tertius interveniens. Das ist Warnung an etliche Theologos, Medicos und Philosophos. *Ges. Werke*, Bd. 4. München, 1941, pp. 149 – 258.
- (1618 – 1621, in Latin), *Epitome of Copernican Astronomy*, Books 4 and 5, 1620 – 1621. *Great Books* (1952, vol. 16, pp. 845 – 1004).
- (1619, in Latin), *Harmony of the World*. Philadelphia, 1997.
- Khinchin, A. Ya.** (1943, in Russian), *Mathematical Foundations of Statistical Mechanics*. New York, 1949.
- (1961, in Russian), The Mises frequency theory and modern ideas of the theory of probability. *Science in Context*, vol. 17, 2004, pp. 391 – 422.
- Kington, J. A.** (1974), The Societas meteorologica Palatina. *Weather*, vol. 29, No. 11, pp. 416 – 426.
- Knapp, G. F.** (1872), Quetelet als Theoretiker. *JNÖS*, Bd. 18, pp. 89 – 124.
- Knies, C. G. A.** (1850), *Die Statistik als selbstständige Wissenschaft*. Kassel.
- Knott C. G.** (1911), *Life and Work of P. G. Tait*. Cambridge.
- Köppen, W.** (1872), Die Aufeinanderfolge der unperiodischen Witterungserscheinungen. *Repert. Met.*, Bd. 2. Petersburg, pp. 187 – 238.
- (1874), Über die Abhängigkeit des klimatischen Characters der Winde von ihrem Ursprunge. Ibidem, Bd. 4, No. 4, pp. 1 – 15.
- (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.
- Körber H.-G.** (1959), Über Humboldts Arbeiten zur Meteorologie und Klimatologie. *Humboldt Gedenkschrift*. Berlin, pp. 289 – 335.
- Kohli, K.** (1975a), Spieldauer: von J. Bernoullis Lösung der fünfte Aufgabe von Huygens bis zu den Arbeiten von de Moivre. In J. Bernoulli (1975, pp. 403 – 455).
- (1975b), Aus dem Briefwechsel zwischen Leibniz und Jakob Bernoulli. Ibidem, pp. 509 – 513.
- (1975c), Kommentar zur Dissertation von N. Bernoulli. Ibidem, pp. 541 – 556.
- Kohli, K., van der Waerden, B. L.** (1975), Bewertung von Leibrenten. In J. Bernoulli (1975, pp. 515 – 539).
- Kolmogorov, A. N.** (1931, in Russian), The method of median in the theory of errors. In Kolmogorov (1992, pp. 115 – 117).
- (1947, in Russian), The role of Russian science in the development of the theory of probability. *Uch. Zapiski Moskovsk. Gos. Univ.* No. 91, pp. 53 – 64.
- (1954, in Russian), Law of small numbers. *Bolshaia Sov. Enc.* (Great Sov. Enc.), 2nd edition, vol. 26, p. 169. Published anonymously.
- (1985 – 1986, in Russian), *Selected Works*, vols. 1 – 2. Dordrecht, 1991 – 1992.
- Kolmogorov, A. N., Petrov, A. A., Smirnov, Yu. M.** (1947, in Russian), A formula of Gauss in the method of least squares. *Izvestia Akad. Nauk SSSR, ser. math.*, vol. 11, pp. 561 – 566. English transl. in Kolmogorov (1985 – 1986/1992, pp. 303 – 308).

- Kolmogorov, A. N., Prokhorov, Yu. V.** (1988, in Russian), Mathematical statistics. *Enc. Math.*, vol. 6, 1990, pp. 138 – 142.
- Koopman, B. O.** (1940), The bases of probability. *Bull. Amer. Math. Soc.*, vol. 46, pp. 763 – 774.
- Kotz, S., Johnson, N. L.**, Editors (1982 – 1989), *Enc. of Statistical Sciences*, second edition, vols 1 – 16 with single paging. Hoboken, New Jersey, 2006.
- Krein, M. G.** (1951, in Russian), Chebyshev's and Markov's ideas in the theory of the limiting values of integrals and their further development. *Uspekhi Matematich. Nauk*, vol. 6, No. 4, pp. 3 – 120.
- Kronecker, L.** (1894), *Vorlesungen über Mathematik*, Bd. 1. Leipzig.
 --- (1901), *Vorlesungen über Zahlentheorie*, Bd. 1. Leipzig.
- Kruskal, W.** (1978), Formulas, numbers, words: statistics in prose. In: *New Directions for Methodology of Social and Behavioral Science*. San Francisco, 1981. Editor, D. Fiske, pp. 93 – 102.
- Kruskal, W., Tanur, J. M.**, Editors (1978), *International Encyclopedia of Statistics*, vols 1 – 2. New York.
- Ku, H. H.**, Editor (1969), *Precision Measurement and Calibration*. Nat. Bureau Standards Sp. Publ. 300, vol. 1. Washington.
- Lacombe, H., Costabel, P.**, Editors (1988), *La figure de la Terre du XVIII^e siècle à l'ère spatiale*. Paris.
- Lagrange, J. L.** (1776a), Sur l'utilité de la méthode de prendre le milieu entre les résultats de plusieurs observations. *Oeuvr.*, t. 2. Paris, 1868, pp. 173 – 236.
 --- (1776b), Letter to Laplace of 30 Dec. *Oeuvr.*, t. 14. Paris, 1892, p. 66.
 --- (1777), Recherches sur les suites récurrentes. *Oeuvr.*, t. 4. Paris, 1869, pp. 151 – 251.
- Lamarck, J. B.** (1800 – 1811), *Annuaire météorologique* [, tt. 1 – 11]. Paris. Extremely rare.
 --- (1809), *Philosophie zoologique*, t. 2. Paris, 1873.
 --- (1815), *Histoire naturelle des animaux sans vertèbres*, t. 1. Paris.
- Lambert, J. H.** (1760), *Photometria*. Augsburg. Incomplete transl. into German: *Ostwald Klassiker*, NNo. 31 – 33, 1892.
 --- (1765a), Anmerkungen und Zusätze zur practischen Geometrie. In Lambert, *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.
 --- (1771), *Anlage zur Architectonic*, Bd. 1. Also *Phil. Schriften*, Bd. 3. Hildesheim, 1965.
 --- (1772), Anmerkungen über die Sterblichkeit, Todtenlisten, Geburthen und Ehen. *Beyträge*, Tl. 3. Berlin, 1772, pp. 476 – 569.
 --- (1772 – 1775), Essai de taxéométrie ou sur la mesure de l'ordre. *Nouv. Mém. Acad. Roy. Sci. et Belles-Lettres Berlin* for 1770, pp. 327 – 342; for 1773, pp. 347 – 368. Also *Phil. Schriften*, Bd. 8/1. Hildesheim, 2007, pp. 423 – 460.
 --- (1773), Observations sur l'influence de la Lune dans le poids de l'atmosphère. *Nouv. Mém. Acad. Roy. Sci. Berlin* for 1771, pp. 66 – 73.
- Lamont, J.** (1867), Über die Bedeutung arithmetischer Mittelwerthe in der Meteorologie. *Z. Öster. Ges. Met.*, Bd. 2, No. 11, pp. 241 – 247.
- Langevin P.** (1913), La physique du discontinu. In collected articles *Les progrès de la physique moléculaire*. Paris, 1914, pp. 1 – 46.
- Laplace, P. S.** (1774), Sur la probabilité des causes par les événements. *OC*, t. 8. Paris, 1891, pp. 27 – 65.
 --- (1776), Recherches sur l'intégration des équations différentielles aux différences finies. *Ibidem*, pp. 69 – 197.
 --- (1781), Sur les probabilités. *OC*, t. 9. Paris, 1893, pp. 383 – 485.
 --- (1786), Sur les approximations des formules qui sont fonctions de très-grands nombres, Suite. *OC*, t. 10. Paris, 1894, pp. 295 – 338.
 --- (1789), Sur quelques points du système du monde. *OC*, t. 11. Paris, 1895, pp. 477 – 558.
 --- (1796), *Exposition du système du monde*. *OC*, t. 6. Paris, 1884 (the whole volume, this being a reprint of the edition of 1835).
 --- (1798 – 1825), *Traité de mécanique céleste*. *OC*, tt. 1 – 5. Paris, 1878 – 1882. English transl. of vols 1 – 4 by N. Bowditch: *Celestial Mechanics* (1829 – 1839). New York, 1966.
 --- (1810a), Sur les approximations des formules qui sont fonctions de très grands nombres et sur leur application aux probabilités. *OC*, t. 12. Paris, 1898, pp. 301 – 345.
 --- (1810b), Same title, supplement. *Ibidem*, pp. 349 – 353.
 --- (1811), Sur les intégrales définies. *Ibidem*, pp. 357 – 412.

- (1812), *Théorie analytique des probabilités*. OC, t. 7, No. 1 – 2. Paris, 1886. Consists of two parts, an Introduction (1814) and supplements, see below. Theory of probability proper is treated in pt. 2.
- (1814), *Essai philosophique sur les probabilités*. In Laplace (1812/1886, No. 1, separate paging). English transl: *Philosophical Essay on Probabilities*. New York, 1995. Translator and editor A.I. Dale.
- (1816), *Théor. anal. prob., Supplément 1*. OC, t. 7, No. 2, pp. 497 – 530.
- (1818), *Théor. anal. prob., Supplément 2*. Ibidem, pp. 531 – 580.
- (ca. 1919), *Théor. anal. prob., Supplément 3*. Ibidem, pp. 581 – 616.
- Laurent, H.** (1873), *Traité du calcul des probabilités*. Paris.
- Le Cam, L.** (1986), The central limit theorem around 1935. *Stat. Sci.*, vol. 1, pp. 78 – 96.
- Legendre, A. M.** (1805), *Nouvelles méthodes pour la détermination des orbites des comètes*. Paris.
- Leibniz, G. W.** (1680?, manuscript, publ. 1872), Öffentliche Assecuranzen. In author's book (1986, pp. 421 – 432).
- (1686, manuscript), Allgemeine Untersuchungen über die Analyse der Begriffe und wahren Sätze. In author's book *Fragmente zur Logik*. Berlin, 1960, pp. 241 – 303.
- (1704), *Neue Abhandlungen über menschlichen Verstand*. Hamburg, 1996.
- (1986), *Sämmtl. Schriften und Briefe, Reihe 4, Bd. 3*. Berlin.
- (2000), *Hauptschriften zur Versicherungs- und Finanzmathematik*. Editors, E. Knobloch et al. Berlin.
- Le procès** (1900), *Le procès Dreyfus devant le Conseil de guerre de Rennes*, tt. 1 – 3. Paris.
- Levi ben Gerson** (14th century, publ. 1999), *The Wars of the Lord*, vol. 3. New York.
- Lévy, M.** (1844), *Traité d'hygiène*. Paris, 1862.
- Lévy, P.** (1925), *Calcul des probabilités*. Paris.
- Lexis, W.** (1874, report), Naturwissenschaft und Sozialwissenschaft. In Lexis (1903, pp. 233 – 251).
- (1876), Das Geschlechtsverhältnis der Geborenen und die Wahrscheinlichkeitsrechnung. JNÖS, Bd. 27, pp. 209 – 245. Reprinted in Lexis (1903, pp. 130 – 169).
- (1877), *Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft*. Freiburg i/Bayern.
- (1879), Über die Theorie der Stabilität statistischer Reihen. JNÖS, Bd. 32, pp. 60 – 98. Reprinted in Lexis (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.
- Liapunov, A. M.** (1895), Chebyshev. In Chebyshev, P. L. *Izbrannye Matematich. Trudy* (Sel. Math. Works). Moscow – Leningrad, 1946, pp. 9 – 21.
- (1900), Sur une proposition de la théorie des probabilités. *Izvestia Imp. Acad. Sci. St. Pétersb.*, sér. 5, t. 13, pp. 359 – 386.
- (1901a), Nouvelle forme du théorème sur la limite de probabilité. *Mém. Imp. Acad. Sci. St. Pétersb.*, sér. 8, Cl. phys.-math., t. 12, No. 5, separate paging. Also in Liapunov (1954, pp. 157 – 178).
- (1901b), An answer to Nekrasov. *Zapiski Kharkov Univ.*, vol. 3, pp. 51 – 63.
- (1954), *Sobranie Sochineniy* (Coll. Works), vol. 1. Moscow.
- Libri – Carrucci, G. B. I. T., Lacroix, S. F., Poisson, S. D.** (1834), Report on Bienaymé's manuscript [Bienaymé (1838)]. *Procès verbaux des séances, Acad. Sci. Paris*, t. 10, pp. 533 – 535.
- Liebermeister, C.** (ca. 1877), Über Wahrscheinlichkeitsrechnung in Anwendung auf therapeutische Statistik. *Sammlung klinischer Vorträge* No. 110 (Innere Medizin No. 39). Leipzig, pp. 935 – 961.
- Lindeberg, J. W.** (1922a), Über das Exponentialgesetz in der Wahrscheinlichkeitsrechnung. *Annales Acad. Scient. Fennicae*, t. A16 für 1920, No. 1, pp. 1 – 23.
- (1922b), Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.*, Bd. 15, pp. 211 – 225.
- Linnik, Yu. V.** (1951), Commentary on Markov (1916a). In Markov (1951, pp. 668 – 670).
- (1952, in Russian), Remarks concerning the classical derivation of the Maxwell law. *Doklady Akad. Nauk SSSR*, vol. 85, pp. 1251 – 1254.

- (1958, in Russian), *Method of Least Squares etc.* Oxford, 1968. Reference in text to English edition of 1961.
- Lloyd G. E. R.** (1982), Observational error in later Greek science. In *Science and Speculation*. Eds, J. Barnes et al. Cambridge, pp. 128 – 164.
- Louis, P. C. A.** (1825), *Recherches anatomico-pathologiques sur la phthisie*. 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.
- Maimonides, M.** (12th century, 1975), *Introduction to the Talmud*. New York.
- (12th century, 1977), *Letters*. New York.
- Maire, [C.], Boscovich, [R. J.]** (1770), *Voyage astronomique et géographique dans l'État de l'Église*. Paris.
- Makovskiy, [O.]** (1914), *Dosokratiki* (Presocratics), pt. 1. Kazan.
- Malthus T. R.** (1798), *Essay on the Principle of Population*. Works, vol. 1. London, 1986.
- Maltsev, A. I.** (1947, in Russian), Remark on the work of Kolmogorov et al (1947). *Izvestia Akad. Nauk SSSR, Ser. math.*, vol. 11, pp. 567 – 578.
- Markov, A. A. [Jr.]** (1951, in Russian), Biography of A. A. Markov [Sr.]. In Markov (1951, pp. 599 – 613).
- Markov, A. A. [Sr.]** (1884, in Russian), Proof of some of Chebyshev's inequalities. In author's book *Izbrannye Trudy po Teorii Nepreryvnykh Drobei* (Sel. Works on Theory of Continued Fractions). Moscow – Leningrad, 1948, pp. 15 – 24.
- (1888), *Table des valeurs de l'intégrale ...* St. Pétersbourg.
- (1898), Sur les racines de l'équation ... In Markov (1951, pp. 255 – 269).
- (1899a), The law of large numbers and the method of least squares. *Ibidem*, pp. 231 – 251.
- (1899b, in Russian), Application of continued fractions to calculation of probabilities. *Izvestia Fiz.-Mat. Obshchestvo Kazan Univ.*, ser. 2, vol. 9, No. 2, pp. 29 – 34.
- (1899c, in Russian), Answer [to Nekrasov]. *Ibidem*, No. 3, pp. 41 – 43.
- (1900), [*Treatise*] *Ischislenie Veroiatnostei* (Calculus of Probabilities). Subsequent editions: 1908, 1913, and (posthumous) 1924. German transl.: Leipzig – Berlin, 1912.
- (1906), Extension of the law of large numbers to magnitudes dependent on one another. In Markov (1951, pp. 339 – 361).
- (1907, in Russian), On some cases of the theorems on the limit of expectation and on the limit of probability. *Izvestia Imp. Acad. Sci. St. Pétersb.*, ser. 6, t. 1, pp. 707 – 714.
- (1908a, in Russian), On some cases of the theorem on the limit of probability. *Ibidem*, t. 2, pp. 483 – 496.
- (1908b), Extension of the limit theorems of the calculus of probability to a sum of magnitudes connected into a chain. In Markov (1951, pp. 363 – 397).
- (1910), Investigation of the general case of trials connected into a chain. *Ibidem*, pp. 465 – 507.
- (1911a), On connected magnitudes not forming a real chain. *Ibidem*, pp. 399 – 416.
- (1911b), On a case of trials connected into a complex chain. *Ibidem*, pp. 417 – 436.
- (1911c, in Russian), On the basic principles of the calculus of probability and on the law of large numbers. In Ondar (1977/1981, pp. 149 – 153).
- (1912a), On trials connected into a chain by unobserved events. In Markov (1951, pp. 437 – 463).
- (1912b, in Russian), A rebuke to Nekrasov. MS, vol. 28, pp. 215 – 227.
- (1914, in Russian), Bicentennial of the law of large numbers. In Ondar (1977a/1981, pp. 158 – 163).
- (1915a, in Russian), On Florov's and Nekrasov's scheme of teaching the theory of probability in school. *Zhurnal Ministerstva Narodn. Prosv.*, New ser., vol. 57, May, pp. 26 – 34 of section on Modern chronicle.
- (1915b), On a problem by Laplace. In Markov (1951, pp. 549 – 571).
- (1916), On the coefficient of dispersion. *Ibidem*, pp. 523 – 535.
- (1951), *Izbrannye Trudy* (Sel. Works). N. p.
- Marsden, B. G.** (1995), 18- and 19-th century developments in the theory and practice of orbit determination. In Taton & Wilson (1995, pp. 181 – 190).
- Materialy** (1917), *Materialy dlia Biograficheskogo Slovaria Deistvitelnykh Chlenov Akademii Nauk* (Materials for a Biographical Dictionary of the Full Members of the Acad. Sci.). Petrograd.

- Maupertuis, P. L. M.** (1738), Relation du voyage fait par ordre du Roi au cercle polaire etc. *Œuvres*, t. 3, pp. 68 – 175.
- (1745), Venus physique. *Oeuvres*, t. 2, pp. 1 – 133.
- (1756a), Sur le divination. *Ibidem*, pp. 298 – 306.
- (1756b), Opérations pour déterminer la figure de la Terre et les variations de la pesanteur. *Oeuvres*, t. 4, pp. 285 – 346.
- (1756c), *Ouvres*, tt. 1 – 4. Lyon.
- Maxwell, J. C.** (1859), On the stability of the motion of Saturn's rings. In Maxwell (1890, vol. 1, pp. 288 – 376).
- (1860), Illustrations of the dynamical theory of gases. *Ibidem*, pp. 377 – 410.
- (1867), On the dynamical theory of gases. *Ibidem*, vol. 2, pp. 26 – 78.
- (1871), Introductory lecture on experimental physics. *Ibidem*, pp. 241 – 255.
- (1873a), *Matter and Motion*. London.
- (1873b), Molecules. In Maxwell (1890, vol. 2, pp. 361 – 378).
- (1873, report), Does the progress of physical science tend to give any advantage to the opinion of necessity over that of the contingency of events. In Campbell & Garnett (1884/1969, pp. 434 – 444).
- (manuscript 1873), Discourse on molecules. *Ibidem*, pp. 358 – 361. Possibly an excerpt.
- (1875), On the dynamical evidence of the molecular constitution of bodies. In Maxwell (1890, vol. 2, pp. 418 – 438).
- (1877), Review of H. W. Watson (1876), *Treatise on the Kinetic Theory of Gases*. Oxford. *Nature*, vol. 16, pp. 242 – 246.
- (1879), On Boltzmann's theorem. In Maxwell (1890, vol. 2, pp. 713 – 741).
- (1890), *Scientific Papers*, vols 1 – 2. Cambridge. Reprints: Paris, 1927, New York, 1965.
- (1990 – 2002), *Scientific Letters and Papers*, vols 1 – 3. Cambridge.
- May, K. O.** (1972). Gauss. *Dict. Scient. Biogr.*, vol. 5, pp. 298 – 315.
- Mayer, T.** (1750), Abhandlung über die Umwälzung des Mondes um seine Axe. *Kosmograph. Nachr. u. Samml. für 1748*, pp. 52 – 183.
- Meadowcroft, L. V.** (1920), On Laplace's theorem on simultaneous errors. *Messenger Math.*, vol. 48, pp. 40 – 48.
- Mendel, J. G.** (1866, in German), Experiments in plant hybridization. In Bateson, W. (1909), *Mendel's Principles of Heredity*. Cambridge, 1913, pp. 317 – 361.
- (1905, in German), Letters to C. Naegeli, 1866 – 1873. *Genetics*, vol. 35, No. 5, pt. 2, pp. 1 – 28.
- Mendeleev, D. I.**, all publications in Russian (1856), Specific volumes. In Mendeleev (1934 – 1952, vol. 1, 1937, pp. 139 – 311).
- (1872a), On the compressibility of gases. *Ibidem*, vol. 6, 1939, pp. 128 – 171.
- (1872b), Report on the experiments of 1867 and 1869. *Ibidem*, vol. 16, 1951, pp. 99 – 113.
- (1875), Progress of work on the restoration of the prototypes of measures of length and weight. *Ibidem*, vol. 22, 1950, pp. 175 – 213.
- (1876), On the temperatures of the atmospheric layers. *Ibidem*, vol. 7, 1946, pp. 241 – 269.
- (1877), The oil industry of Pennsylvania and in the Caucasus. *Ibidem*, vol. 10, 1949, pp. 17 – 244.
- (1885), Note on the scientific work of A. I. Voeikov. *Ibidem*, vol. 25, 1952, pp. 526 – 531.
- (1887), Investigation of the specific weight of aqueous solutions. *Ibidem*, vol. 3, 1934, pp. 3 – 468.
- (1895), On the weight of a definite volume of water. *Ibidem*, vol. 22, 1950, pp. 105 – 171.
- (1934 – 1952), *Sochinenia (Works)*. Moscow – Leningrad.
- Merrington, M. et al** (1983), *List of the Papers and Correspondence of Karl Pearson*. London.
- Métivier, M., Costabel, P., Dugac, P.**, Editors (1981), *Poisson et la science de son temps*. Paris. New edition to appear.
- Meyer, Hugo** (1891), *Anleitung zur Bearbeitung meteorologischer Beobachtungen*. Berlin.
- 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.
- 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. Partly reprinted in Mises (1964a, pp. 35 – 56).
- (1928), *Wahrscheinlichkeit, Statistik und Wahrheit*. Wien. Subsequent editions, for example, Wien, 1972. English transl.: New York, 1981.
- (1964a), *Selected Papers*, vol. 2. Providence, Rhode Island.
- (1964b), *Mathematical Theory of Probability and Statistics*. New York. Editor, H. Geiringer.
- Molina, E. C.** (1930), The theory of probability: some comments on Laplace's Théorie analytique. *Bull. Amer. Math. Soc.*, vol. 36, pp. 369 – 392.
- (1936), A Laplacean expansion for Hermitian – Laplace functions of high order. *Bell Syst. Techn. J.*, vol. 15, pp. 355 – 362.
- Montmort, P. R.** (1708), *Essay d'analyse sur les jeux de hazard*. Paris, 1713. Published anonymously. References in text to reprint: New York, 1980.
- Morant, G. M. et al** (1939), *Bibliography of the Statistical and Other Writings of Karl Pearson*. London.
- Muncke, G. W.** (1837), Meteorologie. *Gehler's Phys. Wörterbuch*, Bd. 6/3. Leipzig, pp. 1817 – 2083.
- Nekrasov, P. A.**, all publications in Russian (1898), General properties of mass independent phenomena in connection with approximate calculation of functions of very large numbers. MS, vol. 20, pp. 431 – 442.
- (1900 – 1902), New principles of the doctrine of probabilities of sums and means. MS, vols 21 – 23, pp. 579 – 763; 1 – 142, 323 – 498; 41 – 455.
- (1901), Concerning a simplest theorem on probabilities of sums and means. MS, vol. 22, pp. 225 – 238.
- (1902), *Filosofia i Logika Nauki* etc. (Philosophy and Logic of Science). Moscow.
- (1906), *Osnovy Obshchestvennykh i Estestvennykh Nauk v Srednei Shkole*. (Principles of Social and Natural Sciences in High School). Petersburg.
- (1911), On the principles of the law of large numbers, the method of least squares and statistics. MS, vol. 27, pp. 433 – 451.
- (1912 – 1914), The Laplacean theory of the method of least squares simplified by a theorem of Chebyshev. MS, vols 28 – 29, pp. 228 – 234 and 190 – 191.
- (1916), *Srednia Shkola* [...] (Secondary School [...]). Petrograd.
- Neugebauer, O.** (1948), Mathematical models in ancient astronomy. In author's book (1983, pp. 99 – 127).
- (1975), *History of Ancient Mathematical Astronomy*, pts 1 – 3. Berlin. [Berlin, 2004.]
- (1983), *Astronomy and History. Sel. Essays*. New York
- Newcomb, S.** (1859 – 1861), Notes on the theory of probability. *Math. Monthly*, vol. 2, 1860, pp. 134 – 140, 272 – 275.
- (1861a), On the secular variations and mutual relations of the orbits of the asteroids. *Mem. Amer. Acad. Arts and Sciences*, New ser., vol. 8, pt. 1, pp. 123 – 152.
- (1861b), Solution of problem. *Math. Monthly*, vol. 3, pp. 68 – 69.
- (1862), Determination of the law of distribution of the nodes and perihelia of the small planets. *Astron. Nachr.*, Bd. 58, pp. 210 – 220.
- (1867), Investigation of the orbit of Neptune. *Smithsonian Contr. to Knowledge*, vol. 15. Separate paging.
- (1872), *On the Right Ascensions of the Equatorial Fundamental Stars*. Washington.
- (1874), Investigation of the orbit of Uranus. *Smithsonian Contr. to Knowledge*, vol. 19. Separate paging.
- (1881), Note on the frequency of use of the different digits in natural numbers. *Amer. J. Math.*, vol. 4, pp. 39 – 40.
- (1886), A generalized theory of the combinations of observations. *Ibidem*, vol. 8, pp. 343 – 366.
- (1892), On the law and the period of the variation of terrestrial latitudes. *Astron. Nachr.*, Bd. 130, pp. 1 – 6.
- (1895), *The Elements of the Four Inner Planets and the Fundamental Constants of Astronomy*. Washington.
- (1896), On the solar motion as a gauge of stellar distances. *Astron. J.*, vol. 17, pp. 41 – 44.
- (1897a), A new determination of the precessional constant. *Astron. Papers*, vol. 8, pp. 1 – 76.
- (1897b), A new determination of the precessional motion. *Astron. J.*, vol. 17, pp. 161 – 167.

- (1901), On the period of the Solar spots. *Astrophys. J.*, vol. 13, pp. 1 – 14.
- (1902), On the statistical relations among the parallaxes and the proper motions of the stars. *Astron. J.*, vol. 22, pp. 165 – 169.
- (1904a), *On the Position of the Galactic*. Carnegie Instn of Wash., *Publ. No.* 10.
- (1904b), *Statistical Inquiry into the Probability of Causes of Sex in Human Offspring*. Ibidem, *Publ. No.* 11.
- Newcomb, S., Holden, E. S.** (1874), On the possible periodic changes of the Sun's apparent diameter. *Amer. J. Sci.*, ser. 3, vol. 8 (108), pp. 268 – 277.
- Newton, I.** (1669 – 1671), *Lectiones opticae. Opera quae Extant Omnia*, vol. 3. London, 1782, pp. 250 – 437. [London, 1931.]
- (1704), *Optics*. Ibidem, vol. 4, pp. 1 – 264. [London, 1931.]
- (1728), *Chronology of Ancient Kingdoms Amended*. London. [London, 1770.]
- (1729), *Mathematical Principles of Natural Philosophy*. Third edition. Berkeley, 1960.
- (1958), *Papers and Letters on Natural Philosophy*. Cambridge.
- (1967), *Mathematical Papers*, vol. 1. Cambridge.
- Neyman, J.** (1934), On two different aspects of the representative method. *J. Roy. Stat. Soc.*, vol. 97, pp. 558 – 625. Reprinted in Neyman (1967, pp. 98 – 141).
- (1938), L'estimation statistique traitée comme un problème classique de probabilité. Reprinted Ibidem, pp. 332 – 353.
- (1938), *Lectures and Conferences on Mathematical Statistics and Probability*. Washington, 1952.
- (1967), *Selection of Early Statistical Papers*. Berkeley.
- (1978), Review of Ondar (1977). *Hist. Math.*, vol. 5, pp. 485 – 486.
- Nicomachus of Gerasa** (1st century, 1952), *Introduction to Arithmetic*. In *Great Books* (1952, vol. 11, pp. 811 – 848).
- Novikov, S. P.** (2002, in Russian), The second half of the 20th century and its result: the crisis of the physical and mathematical association in Russia and in the West. *IMI*, vol. 7 (42), pp. 326 – 356.
- Novoselsky, S. A.** (1916), *Smertnost i Prodolzhitelnost Zhizni v Rossii* (Mortality and Duration of Life in Russia). Petrograd.
- Ondar, Kh. O.** (1971, in Russian), On the work of A. Yu. Davidov in the theory of probability etc. *Istoria i Metodologia Estestvennykh Nauk*, No. 11, pp. 98 – 109.
- Editor (1977, in Russian), *Correspondence between Markov and Chuprov on the Theory of Probability and Mathematical Statistics*. New York, 1981.
- Ore, O.** (1953), *Cardano, the Gambling Scholar*. Princeton, 1963; New York, 1965.
- Oresme, N.** (14th century, publ. 1966, Latin and English), *De Proportionibus Proportionum* and *Ad Pauca Respicientes*. Editor E. Grant. Madison..
- Paevsky, V. V.** (1935, in Russian), Euler's work in population statistics. In *Eiler* (Euler. Memorial volume). Moscow – Leningrad, pp. 103 – 110.
- Pannekouk A.** (1961), *History of Astronomy*. New York, 1989.
- Pascal, B.** (1654a), Correspondence with P. Fermat. OC, t. 1, pp. 145 – 166.
- (1654b), A la très illustre Académie Parisienne de Mathématique (Latin and French). Ibidem, pp. 169 – 173.
- (1665), *Traité du triangle arithmétique*. Ibidem, pp. 282 – 327.
- (1669), *Pensées*. Collection of fragments. OC, t. 2, pp. 543 – 1046.
- (1998 – 2000), *Oeuvres complètes*, tt. 1 – 2. Paris.
- Paty, M.** (1988), D'Alembert et les probabilités. In Rashed R., Editor, *Sciences à l'époque de la Révolution française*. Paris, pp. 203 – 265.
- 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. Gosset*. Editors, R. L. Plackett, 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. Collected reprints of papers.
- Pearson E. S., Wishart J.**, Editors (1943), *Student's Collected Papers*. London.
- Pearson, K.** (1892), *Grammar of Science*. [Bristol, 1991; Tokyo, 1991.]
- (1894), On the dissection of asymmetrical frequency curves. *Phil. Trans. Roy. Soc.*, vol. A185, pp. 71 – 110.
- (1896), Skew variation in homogeneous material. Ibidem, vol. A186, pp. 343 – 414.
- (1898), Cloudiness. *Proc. Roy. Soc.*, vol. 62, pp. 287 – 290.

- (1900), On a criterion that a given system of deviations ... can be reasonably supposed to have arisen from random sampling. *London, Edinburgh and Dublin Phil. Mag.*, ser. 5, vol. 50, pp. 157 – 175.
- (1905), Das Fehlergesetz und seine Verallgemeinerung durch Fechner und Pearson: a rejoinder. *Biometrika*, vol. 4, pp. 169 – 212.
- (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.
- (1919), Peccavimus! *Biometrika*, vol. 12, pp. 259 – 281.
- (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.
- (1948), *Early Statistical Papers*. Cambridge.
- (1978), *History of Statistics in the 17th and 18th Centuries against the Changing Background of Intellectual, Scientific and Religious Thought*. Lectures 1921 – 1933. Editor E. S. Pearson. London.
- Pettenkofer, M.** (1865), Über die Verbreitungsart der Cholera. *Z. Biologie*, Bd. 1, pp. 322 – 374.
- (1873), *Über den Werth der Gesundheit für eine Stadt*. Braunschweig. English transl.: *Bull. Hist. Med.*, vol. 10, No. 3 and 4, 1941.
- (1886 – 1887), Zum gegenwärtigen Stand der Cholerafrage. *Arch. f. Hyg.*, Bd. 4, pp. 249 – 354, 397 – 546; Bd. 5, pp. 353 – 445; Bd. 6, pp. 1 – 84, 129 – 233, 303 – 358, 373 – 441; Bd. 7, pp. 1 – 81.
- 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.** (1662), *A Treatise of Taxes and Contributions*. In Petty (1899, vol. 1, pp. 1 – 97).
- (1674), *Discourse Read before the Royal Society*. London.
- (1690), *Political Arithmetic*. In Petty (1899, vol. 2, pp. 239 – 313).
- (1899), *Economic Writings*, vols 1 – 2. Ed., C. H. Hull. Cambridge. Reprint: London, 1997.
- (1927), *Papers*, vols 1 – 2. London. Reprint: London, 1997.
- Pfanzagl, J., Sheynin, O.** (1997), Süsmilch. In second edition of Kotz & Johnson (2006, vol. 13, pp. 8489 – 8491) but somehow appeared there anonymously.
- Picard, J.** (1693), Observations astronomiques faites en diverse endroits du royaume. *Mém. Acad. Roy. Sci. 1666 – 1699*, t. 7. Paris, 1729, pp. 329 – 347.
- Pirogov, N. I.** (1849, in Russian), On the achievements of surgery during the last five years. *Zapiski po Chasti Vrachebn. 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).
- (1864), *Grundzüge der allgemeinen Kriegschirurgie*. Leipzig. Russian version: 1865 – 1866.
- (1871), *Bericht über die Besichtigung der Militär-Sanitäts-Anstalten in Deutschland, Lothringen und Elsass 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. Kendall & Plackett (1977, pp. 279 – 291).
- Plato, J. von** (1983), The method of arbitrary functions. *Brit. J. Phil. Sci.*, vol. 34, pp. 37 – 47.
- Poincaré, H.** (1892a), *Thermodynamique*. Paris.
- (1892b), Réponse à P. G. Tait. *Œuvres*, t. 10. Paris, 1954, pp. 236 – 237.
- (1894), Sur la théorie cinétique des gaz. *Ibidem*, pp. 246 – 263.
- (1896), *Calcul des probabilités*. Paris, 1912, reprinted 1923 and 1987.
- (1902), *La science et l'hypothèse*. Paris, 1923.
- (1905), *La valeur de la science*. Paris, 1970.
- (1921), Résumé analytique [of own works]. In *Mathematical Heritage of H. Poincaré*. Providence, Rhode Island, 1983. Editor F. E. Browder, pp. 257 – 357.
- Poisson, S.-D.** (1824 – 1829), Sur la probabilité des résultats moyens des observations. *Conn. des Temps*, pour 1827, pp. 273 – 302; pour 1832, pp. 3 – 22.

- (1825 – 1826), Sur l'avantage du banquier au jeu de trente-et-quarante. *Annales Math. Pures et Appl.*, t. 16, pp. 173 – 208.
- (1833), *Traité de mécanique*, t. 1. Paris. Second edition. First edition, 1811.
- (1835), Recherches sur la probabilité des jugements etc. *C. r. Acad. Sci. Paris*, t. 1, pp. 473 – 494.
- (1836), Note sur la loi des grands nombres. *Ibidem*, t. 2, pp. 377 – 382.
- (1837a), *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*. Paris. [Paris, 2003.]
- (1837b), Sur la probabilité du tir a la cible. *Mémorial de l'Artillerie*, No. 4, pp. 59 – 94.
- (1837c), *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.
- (1954), *Mathematics and Plausible Reasoning*. Princeton.
- Postnikov, A. G.** (1974), *Veroiatnostnaia Teoria Chisel* (Stochastic Number Theory). Moscow.
- Prevost, P., Lhuillier, S. A. J.** (1799), Sur l'art d'estimer la probabilité des causes par les effets. *Mém. Acad. Roy. Sci. et Belles-Lettres Berlin avec l'Histoire 1796*, pp. 3 – 24 of the second paging.
- Proctor, R. A.** (1872), On star-grouping. *Proc. Roy. Instn Gr. Brit.*, vol. 6, pp. 143 – 152.
- (1874), *The Universe*. London.
- Prokhorov, Yu. V.** (1986, in Russian), The law of large numbers and the estimation of the probabilities of large deviations. In Bernoulli, J. (1986, pp. 116 – 150).
- , Editor (1999), *Veroiatnost i Matematicheskaia Statistika. Enziklopedia* (Probability and Math. Statistics. Enc.). Moscow.
- Prokhorov, Yu. V., Sevastianov, B. A.** (1999, in Russian), Probability theory. In Prokhorov (1999b, pp. 77 – 81).
- Prudnikov, V. E.** (1964, in Russian), *Chebyshev*. [Leningrad, 1976.]
- Ptolemy** (2nd century, 1956), *Tetrabiblos*. In Greek and English. London.
- (2nd century, 1984), *Almagest*. London.
- Ptukha, M. V.** (1955), *Ocherki po Istorii Statistiki v SSSR* (Essays on the History of Statistics in the Soviet Union), vol. 1. Moscow.
- Quetelet, A.** (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*. Paris.
- (1848b), Sur la statistique morale. *Mém. Acad. Roy. Sci., Lettres et Beaux-Arts Belg.*, t. 21. Separate paging.
- , publie par (1849 – 1857), *Sur le climat de la Belgique*, tt. 1 – 2. Bruxelles.
- (1852), Sur quelques propriétés curieuses qui présentent les résultats d'une série d'observations. *Bull. Acad. Roy. Sci., Lettres, Beaux-Arts Belg.*, t. 19, pt. 2, pp. 303 – 317.
- (1853), *Théorie des probabilités*. Bruxelles.
- (1869), *Physique sociale*, tt. 1 – 2. Bruxelles, this being a new edition of Quetelet (1836). Reprint 1897.
- (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.
- Rabinovitch, N. L.** (1973), *Probability and Statistical Inference in Ancient and Medieval Jewish Literature*. Toronto.
- Radelet de Grave P., Scheuber, V.** (1979), *Correspondance entre D. Bernoulli et J.-H. Lambert*. Paris.
- Raikher, V. K.** (1947), *Obshchestvenno-Istoricheskie Tipy Strakhovania* (Social Types of Insurance in History). Moscow – Leningrad.
- Raimi, R. A.** (1976), The first digit problem. *Amer. Math. Monthly*, vol. 83, pp. 521 – 538.

- Raju, C. K.** (2010), Probability in ancient India. *Handbook of the Philosophy of Science*, vol. 7. N. p., pp. 1 – 24.
- Réaumur, R. A.** (1738), Observations du thermomètre etc. *Hist. Acad. Roy. Sci. Paris avec Mém. math.-phys.* 1735, pp. 545 – 576.
- Rehnsch, E.** (1876), Zur Orientierung über die Untersuchungen und Ergebnisse der Moralstatistik. *Z. Philos. u. philos. Kritik*, Bd. 69, pp. 43 – 115, this being pt 2 of his article. Part 3 had not appeared.
- Report** (1916, in Russian), Report of the [ad hoc] Commission to discuss some issues concerning the teaching of mathematics in high school. *Izvestia Imp. Akad. Nauk*, ser. 6, vol. 10, No. 2, pp. 66 – 80.
- Rigaud, S. P.** (1832), *Miscellaneous Works and Correspondence of J. Bradley*. Oxford. [Oxford, 1972.]
- (1841), *Correspondence of Scientific Men of the 17th Century*, vol. 2. Oxford.
- Romanovsky, V. I.** (1912), *Zakon Bolshikh Chisel i Teorema Bernoulli* (The Law of Large Numbers and the Bernoulli Theorem). Warsaw. Also in *Protokoly Zasedaniy Obshchestva Estestvoispytatelei Univ. Varshava* for 1911, No. 4, pp. 39 – 63.
- (1924, in Russian), Theory of probability and statistics according to some newest works of Western scholars. *Vestnik Statistiki*, No. 4 – 6; No. 7 – 9, pp. 1 – 38 and 5 – 34.
- (1925a, in Russian), Generalization of the Markov inequality and its application in correlation theory. *Bull. Sredneaziatsk. Gos. Univ.*, vol. 8, pp. 107 – 111.
- (1925b), Généralisation d'une inégalité de Markoff. *C. r. Acad. Sci. Paris*, t. 180, pp. 1468 – 1470.
- (1930), *Matematicheskaiia Statistika* (Math. Statistics). Moscow – Leningrad.
- (1961), Same title, book 1. Tashkent. Editor, T. A. Sarymsakov.
- Rubanovsky L. M.** (1934), *Metody Fizicheskoi Statistiki* (Methods of Physical Statistics). Leningrad – Moscow.
- Sakatov, P. S.** (1950, in Russian), *Lehrbuch der höheren Geodäsie*. Berlin, 1957. Other Russian editions, 1953, 1964.
- Schlözer, A. L.** (1804), *Theorie der Statistik*. Göttingen.
- Schmidt, O. Yu.** (1926, in Russian), On the Bertrand paradox in the theory of probability. *MS*, vol. 33, pp. 33 – 40.
- Schneider, I., Editor** (1988), *Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfängen bis 1933*. Darmstadt. Collection of fragments almost exclusively in German with some comment.
- Schumacher H.** (1931), Ladislaus von Bortkiewicz. *Allg. stat. Arch.*, Bd. 21, pp. 573 – 576.
- 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.** (1983), Modern probabilistic concepts in the work of E. Abbe and A. De Moivre. *Math. Scientist*, vol. 8, pp. 75 – 80.
- (1984), The central limit theorem and linear least squares in pre-revolutionary Russia: the background. *Ibidem*, vol. 9, pp. 37 – 77.
- (1987), Chuprov on finite exchangeability, expectation of ratios and measures of association. *Hist. Math.*, vol. 14, pp. 243 – 257.
- (1994), Carl Liebermeister's hypergeometric tails. *Ibidem*, vol. 21, pp. 453 – 462.
- (1998), Bienaymé (1798 – 1878): criticality, inequality and internationalization. *ISR*, vol. 66, pp. 291 – 301.
- Seneta, E., Parshall, K. H., Jongmans, F.** (2001), Nineteenth-century developments in geometric probability. *AHES*, vol. 55, pp. 501 – 524.
- Shafer, G.** (1978), Non-additive probabilities in the work of Bernoulli and Lambert. *AHES*, vol. 19, pp. 309 – 370.
- 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.
- (1968), On the early history of the law of large numbers. *Biometrika*, vol. 55, pp. 459 – 467. Reprinted in E. S. Plackett & Kendall (1970), pp. 231 – 239).
- (1970), D. Bernoulli on the normal law. *Biometrika*, vol. 57, pp. 199 – 202. Reprinted in M. G. Kendall & Plackett (1977), pp. 101 – 104).
- (1971a), Newton and the classical theory of probability. *AHES*, vol. 7, pp. 217 – 243.
- (1971b), Lambert's work in probability. *Ibidem*, pp. 244 – 256.

- (1971c), On the history of some statistical laws of distribution. *Biometrika*, vol. 58, pp. 234 – 236. Reprinted in M. G. Kendall & Plackett (1977, pp. 328 – 330).
- (1972), D. Bernoulli's work on probability. *RETE. Strukturgeschichte der Naturwissenschaften*, Bd. 1, pp. 273 – 300. Reprinted in M. G. Kendall & Plackett (1977, pp. 105 – 132).
- (1973a), Finite random sums. Historical essay. AHES, vol. 9, pp. 275 – 305.
- (1973b), Boscovich's work on probability. AHES, pp. 306 – 324.
- (1973c), Mathematical treatment of astronomical observations. Historical essay. AHES, vol. 11, pp. 97 – 126.
- (1974), On the prehistory of the theory of probability. AHES, vol. 12, pp. 97 – 141.
- (1976), Laplace's work on probability. AHES, vol. 16, pp. 137 – 187.
- (1977a), Laplace's theory of errors. AHES, vol. 17, pp. 1 – 61.
- (1977b), Early history of the theory of probability. AHES, pp. 201 – 259.
- (1978a, in Russian), Theory of probability. Coauthor, B. V. Gnedenko. Chapter in *Mathematics of the 19th century* [vol. 1]. Editors, A. N. Kolmogorov, A. P. Youshkevitch. Basel, 1992 and 2001, pp. 211 – 288. Gnedenko died in 1995; I did not know about the reprint and the information there is dated.
- (1978b), Poisson's work in probability. AHES, vol. 18, pp. 245 – 300.
- (1979), Gauss and the theory of errors. AHES, vol. 20, pp. 21 – 72.
- (1980), On the history of the statistical method in biology. AHES, vol. 22, pp. 323 – 371.
- (1982), On the history of medical statistics. AHES, vol. 26, pp. 241 – 286.
- (1983), Corrections and short notes on my papers. AHES, vol. 28, pp. 171 – 195.
- (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 – 93.
- (1985), On the history of the statistical method in physics. AHES, vol. 33, pp. 351 – 382.
- (1986a), Quetelet as a statistician. AHES, vol. 36, pp. 281 – 325.
- (1986b, in Russian), J. Bernoulli and the beginnings of probability theory. In Bernoulli J. (1986, pp. 83 – 115).
- (1989), Markov's work on probability. AHES, vol. 39, pp. 337 – 377; vol. 40, p. 387.
- (1990a, in Russian), *Chuprov: Life, Work, Correspondence*. Göttingen, 1996 and 2011.
- (1990b, in Russian), Markov's review of a paper by B. B. Golitzin. IMI, vol. 32 – 33, pp. 451 – 455.
- (1991a), Poincaré's work in probability. AHES, vol. 42, pp. 137 – 172.
- (1991b), On the works of Buniakovsky in the theory of probability. AHES, vol. 43, pp. 199 – 223.
- (1991c), On the notion of randomness from Aristotle to Poincaré. *Math., Inform. et Sciences Humaines*, No. 114, pp. 41 – 55.
- (1992), Al-Biruni and the mathematical treatment of observations. *Arabic Sciences and Phil.*, vol. 2, pp. 299 – 306.
- (1993a, in Russian), Markov's letters in newspaper *Den* in 1914 – 1915. IMI, vol. 34, pp. 194 – 206.
- (1993b), On the history of the principle of least squares. AHES, vol. 46, pp. 39 – 54.
- (1993c), Treatment of observations in early astronomy. AHES, pp. 153 – 192.
- (1994a), Gauss and geodetic observations. AHES, vol. 46, pp. 253 – 283.
- (1994b), Chebyshev's lectures on the theory of probability. AHES, pp. 321 – 340.
- (1994c), Bertrand's work on probability. AHES, vol. 48, pp. 155 – 199.
- (1994d), Ivory's treatment of pendulum observations. *Hist. Math.*, vol. 21, pp. 174 – 184.
- (1994e, in Russian), Correspondence between P. A. Nekrasov and K. A. Andreev. IMI, vol. 35, pp. 124 – 147. Co-author, M. V. Chirikov.
- (1995a), Helmholtz's work in the theory of errors. AHES, vol. 49, pp. 73 – 104.
- (1995b), Density curves in the theory of errors. AHES, pp. 163 – 196.
- (1995c), Introduction of statistical reasoning into astronomy. In Taton & Wilson (1995, pp. 191 – 197).
- (1996), Selection and treatment of observations by Mendeleev. *Hist. Math.*, vol. 23, pp. 54 – 67.
- (1997a), Achenwall. In second edition of Kotz & Johnson (2006, vol. 1, pp. 26 – 27).
- (1997b), Süssmilch. See Pfanzagl & Sheynin.

- (1997c, in Russian), Markov and insurance of life. *Math. Scientist*, vol. 30, 2005, pp. 5 – 12.
- (1999a), Statistics, definition of. In second edition of Kotz & Johnson (2006, vol. 12, pp. 8128 – 8135).
- (1999b), Discovery of the principle of least squares. *Hist. Scient.*, vol. 8, pp. 249 – 264.
- (1999c, in Russian), Slutsky: fifty years after his death. IMI, vol. 3 (38), pp. 128 – 137.
- (1999d), Gauss and the method of least squares. JNÖS, Bd. 219, pp. 458 – 467.
- (2000), Bessel: some remarks on his work. *Hist. Scient.*, vol. 10, pp. 77 – 83.
- (2001a), Pirogov as a statistician. Ibidem, pp. 213 – 225.
- (2001b), Gauss, Bessel and the adjustment of triangulation. Ibidem, vol. 11, pp. 168 – 175.
- (2002a, in Russian), On Cournot's heritage in the theory of probability. IMI, vol. 7 (42), pp. 301 – 316.
- (2002b), Newcomb as a statistician. *Hist. Scient.*, vol. 12, pp. 142 – 167.
- (2002c), Sampling without replacement: history and applications. *NTM, Intern. Z. f. Geschichte u. Ethik Naturwiss., Technik, Med.*, Bd. 10, pp. 181 – 187.
- (2003a), Nekrasov's work on probability: the background. AHES, vol. 57, pp. 337 – 353.
- (2003b), On the history of Bayes's theorem. *Math. Scientist*, vol. 28, pp. 37 – 42.
- (2003c), Geometric probability and the Bertrand paradox. *Hist. Scient.*, vol. 13, pp. 42 – 53.
- (2004), Fechner as a statistician. *Brit. J. Math., Stat. Psychology*, vol. 57, pp. 53 – 72.
- (2006a), Markov's work on the treatment of observations. *Hist. Scient.*, vol. 16, pp. 80 – 95.
- (2006b), *Theory of Probability and Statistics As Exemplified in Short Dictums*. Berlin, 2009.
- (2007a), The true value of a measured constant and the theory of errors. *Hist. Scient.*, vol. 17, pp. 38 – 48.
- (2007b), Markov: integrity is just as important as scientific merits. *NTM, Intern. Z. f. Geschichte u. Ethik Naturwiss., Technik, Med.*, Bd. 15, pp. 289 – 294.
- (2007c), Euler's work in probability and statistics. In *Euler Reconsidered. Tercentenary Essays*. Ed., R. Baker. Heber City, Uta, pp. 281 – 316.
- (2008), Bortkiewicz' alleged discovery: the law of small numbers. *Hist. Scient.*, vol. 18, pp. 36 – 48.
- (2009), *Theory of Probability. Historical Essay*. Berlin.
- (2010), Karl Pearson a century and a half after his birth. *Math. Scientist*, vol. 35, pp. 1 – 9.
- (2011), Randomness and determinism etc. Unpublished.
- (2012), New exposition of Gauss's final introduction of least squares. *Math. Scientist*, to appear.
- Shoesmith, E.** (1985a), N. Bernoulli and the argument for Divine Providence. *ISR*, vol. 53, pp. 255 – 259.
- (1985b), 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.
- (1869 – 1870), Hospitalism. Ibidem, pp. 289 – 405.
- Simpson, T.** (1740), *Nature and Laws of Chance*. London.
- (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.
- Sleshinsky, I. V.** (1892, in Russian), On the theory of the method of least squares. *Zapiski Novoross. Obshchestva Estestvoispytatelei*, vol. 14, pp. 201 – 264.
- (1893, in Russian), On Chebyshev's theorem. *Zapiski Novoross. Univ.*, vol. 59, pp. 503 – 506.
- Slutsky, E. E.** (1912), *Teoria Korreliatsii (Correlation theory)*. Kiev.

- Snow, J.** (1855), On the mode of communication of cholera. In *Snow on Cholera*. New York, 1965, pp. 1 – 139.
- Sofonea, T.** (1957a), Leibniz und sein Projekt zur Errichtung staatlicher Versicherungsanstalten. *Schweiz. Versicherungs-Zeitschrift*, Jg. 25, pp. 144 – 148.
- (1957b), E. Halley (1656 – 1742) und seine Sterbetafel 300 Jahre nach seiner Geburt. *Het Verzerkerings Archief*, t. 34, pp. 31* – 42*.
- Soloviev, A. D.** (1997, in Russian), Nekrasov and the central limit theorem of the theory of probability. *Archives Intern. d'Hist. Sci.*, vol. 58, 2008, No. 160 – 161, pp. 353 – 364.
- Spieß, O.** (1975), Zur Vorgeschichte des Peterburger Problems. In J. Bernoulli (1975, pp. 557 – 567).
- Steklov, V. A.** (1915, in Russian), On a problem of Laplace. *Izvestia Imp. Akad. Nauk*, Ser. 6, t. 9, pp. 1515 – 1537.
- (1924, in Russian), A. A. Markov. *Izvestia Ross. Akad. Nauk*, ser. 6, vol. 16 for 1922, pp. 169 – 184.
- Stigler, S. M.**, Editor (1980), *American Contributions to Mathematical Statistics in the 19th Century*, vols 1 – 2. New York. Reprints of papers of early American authors. No single paging.
- (1986), *History of Statistics*. Cambridge (Mass.)
- (1999), *Statistics on the Table*. Cambridge (Mass.). Collected revised papers.
- Stiltjes, T. J.** (1885), Note a l'occasion de la reclamation de M. Markoff. OC, t. 2. Berlin, 1993, pp. 430 – 431/518 – 519.
- Strecker, H., Strecker, R.** (2001), Anderson. In Heyde & Seneta (2001, pp. 377 – 381).
- Struve, P. B.** (1918, in Russian), Who was the first to indicate that statistics can be applied to philological studies? *Izvestia Ross. Akad. Nauk*, Ser. 6, t. 12, pp. 1317 – 1318.
- Süssmilch, J. P.** (1741), *Die Göttliche Ordnung in den Veränderungen des menschlichen Geschlechts, aus der Geburt, dem Tode und der Fortpflanzung desselben*. Berlin, 1765. Several subsequent editions.
- (1758), Gedanken von dem epidemischen Krankheiten, etc. In Wilke, J., Editor, *Die Königliche Residenz Berlin und die Mark Brandenburg im 18. Jahrhundert*. Berlin, 1994, pp. 69 – 116.
- Tait, P. G.** (1892), Poincaré's Thermodynamics. *Nature*, vol. 45, pp. 245 – 246.
- Takács (Takács), L.** (1967), *Combinatorial Methods in the Theory of Stochastic Process*. New York.
- (1969), On the classical ruin problems. *Ibidem*, vol. 64, pp. 889 – 906.
- (1982), Ballot problems. In second edition of Kotz & Johnson (2006, vol. 1, pp. 183 – 188).
- (1994), The problem of points. *Math. Scientist*, vol. 19, pp. 119 – 139.
- Taqqi, M. S.** (2001), Bachelier and his times; a conversation with B. Bru. *Finance Stoch.*, vol. 5, pp. 3 – 32.
- Taton, R., Wilson, C.**, Editors (1995), *General History of Astronomy*, vol. 2B. Cambridge.
- Thatcher, A. R.** (1957), A note on the early solutions of the problem of the duration of play. *Biometrika*, vol 44, pp. 515 – 518. Reprinted in E. S. Pearson & Kendall (1970, pp. 127 – 130).
- Tikhomandritsky, M. A.** (1898), *Kurs Teorii Veroiatnostei* (Course in Theory of Probability). Kharkov.
- Tikhomirov, E. I.** (1932, in Russian), Directions to Russian meteorological stations in the 18th century. *Izvestia Glavnoi Geofisich. Obs.*, No. 1 – 2, pp. 3 – 12.
- Timerding, H. E.** (1915), *Analyse des Zufalls*. Braunschweig.
- Toaldo, J.** (1775, in Italian), *Witterungslehre für den Feldbau*. Berlin, 1777.
- (1777), Essai de météorologie. *J. Phys.*, t. 10, pp. 249 – 279, 333 – 367.
- 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.** (1975), Early kinetic theories of gases. AHES, vol. 15, pp. 1 – 66.
- (1984), *An Idiot's Fugitive Essays on Science*. New York. Collected reprints of author's introductions and reviews on history and philosophy of natural sciences.
- Tsinger, V. Ya.** (1862), *Sposob Naimenshikh Kvadratov* (Method of Least Squares). Dissertation. Moscow.
- Tutubalin, V. N.** (1972), *Teoria Veroiatnostei v Estestvoznanii* (Theory of probability in Natural Sciences.). Moscow.
- Waerden van der, B. L.** (1968), Mendel's experiments. *Centaurus*, vol. 12, pp. 275 – 288.

- 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.
- Wilks, S. S.** (1962), *Mathematical Statistics*. New York.
- Wilson, C.** (1980), Perturbations and solar tables from Lacaille to Delambre. *AHES*, vol. 22, pp. 53 – 304.
- Winkler W.** (1931), Ladislaus von Bortkiewicz. *Schmollers Jahrbuch f. Gesetzgebung, Verwaltung u. Volkswirtschaft im Deutschen Reich*, 55. Jg, pp. 1025 – 1033.
- Wolf, Abr.** (1935), *History of Science, Technology and Philosophy in the 16th and 17th Centuries*. London, 1950.
- Wittstein, Ph.** (1867), *Mathematische Statistik*. Hannover.
- Woolhouse, W. S. B.** (1873), On the philosophy of statistics. *J. Inst. Actuaries*, vol. 17, pp. 37 – 56.
- Woytinsky W. S.** (1961), *Stormy Passage*. New York.
- Yamazaki, E.** (1971), D’Alembert et Condorcet: quelques aspects de l’histoire du calcul des probabilités. *Jap. Studies Hist. Sci.*, vol. 10, pp. 60 – 93.
- Yarochenko, S. P.** (1893a, in Russian), On the theory of the method of least squares. *Zapiski Novoross. Univ.*, vol. 58, pp. 193 – 208 of second paging.
- (1893b), Sur la méthode des moindres carrés. *Bull. Sciences Math.*, sér. 2, t. 17, pp. 113 – 125.
- You Poh Seng** (1951), Historical survey of the development of sampling theories and practice. *J. Roy. Stat. Soc.*, vol. A114, pp. 214 – 231. Reprinted: M. G. Kendall & Plackett (1977, pp. 440 – 458).
- Youshkevitch, A. A.** (1974), Markov. *Dict. Scient. Biogr.*, vol. 9, pp. 124 – 130.
- Youshkevitch, A. P.** (1986, in Russian), N. Bernoulli and the publication of J. Bernoulli’s *Ars Conjectandi. Theory of Probability and Its Applications*, vol. 31, 1987, pp. 286 – 303.
- Yule, G. U.** (1900), On the association of attributes in statistics. *Phil. Trans. Roy. Soc.*, vol. A194, pp. 257 – 319. Also in Yule, *Statistical Papers*. London, 1971, pp. 7 – 69.
- Yule, U., Kendall, M. G.** (1937), *Introduction to Theory of Statistics*. London, 1958.
- Zach, F. X. von** (1813), Sur le degré du méridien. *Mém. Acad. Imp. Sci., Littérature, Beaux-Arts Turin* pour 1811 – 1812. *Sci. math. et phys.*, pp. 81 – 216.
- Zermelo E.** (1900), Über die Anwendung der Wahrscheinlichkeitsrechnung auf dynamische Systeme. *Phys. Z.*, Bd. 1, pp. 317 – 320.

Index of Names

The Index does not cover either the Introduction, Supplement or Bibliography. The numbers refer to subsections rather than to pages. Bold type signifies chapter or (sub)section devoted to the person in question.

Aaboe, A. 1.5
Abbe, E. 9.2
Achenwall, G. 6.2.1
Adrain, R. 9.1.2
Afanasieva – Ehrenfest, T. A. 6.1.1, 7.1-3
Al-Biruni (973 – 1048) 1.4 – 1.6
Ambartsumian, R. V. 11.2
Anchersen, J. P. 6.2.1, 10.8
Anderson, O. 14.1.2, 14.2
Andreev, K. A. 13.1
Arbuthnot, J. **2.2.4**
Archimedes (ca. 287 – 212 BC) 1.5, 7.3
Aristarchus (end of 4th c. – first half of 3rd c. BC) 1.5
Aristotle (384 – 322 BC) 1.1 – 1.4, 2.1.1
Arnauld, A. 2.1.1, 2.1.4, 2.2.1, 3.1.2, 3.3.1
Austin, E. 10.8.3
Babbage, C. 10.5, 10.8.2
Bachelier, L. 13.3
Baer, K. 10.8.2
Baily, F. 1.7
Barbier, E. 11.2
Bauer, R. K. 14.3
Bayes, T. 2.2.3, **5**, 6.1.6, 7.1-1, 7.1.-3, 7.1-5, 7.2-5, 8.1.5, 11.1, 12.2
Bellhouse D. R. 2.1.1, 2.2.4
Belvalkar, S. K. 1.1
Benjamin M. 10.8.4
Bentley, R. 2.2.3
Bernoulli, Daniel, 3.3.4, **6.1.1**, 6.1.4 – 6.1.6, 6.2.3, 6.2.4, 6.3.1, 7.1-3, 8.5, 10.2.2, 10.4.2, 10.7, 10.8.4, 13.3
Bernoulli, Jakob 1.4, 2.1.1, 2.1.2, 2.1.4, 2.2.2, 2.2.4, **3**, 4.3, 5.2, 7.1-8, 8.1.2, 8.1.3, 8.1.5, 8.3, 10.2, 11.3, 13.2
Bernoulli, Johann I 3.1.2
Bernoulli, Johann III 6.3.1
Bernoulli, Niklaus 3.1.2, **3.3.2**, 3.3.3, 3.3.4, 6.1.1, 8.5, 12.1
Bernstein, S. N. 11.3, 12.2, 12.3, 13.3, 13.4, 14.2
Bertrand, J. 2.1.1, 6.1.1, 6.1.2, 6.1.6, 8.1.5, 9.1.3, 10.5, 11, **11.1**, **11.2**, 11.3
Bessel, F. W. 6.3.1, 8.1.5, 9.1.3, 9.1.5, **9.3**
Bienaymé, I. J. 7.2, 8.1.1, 8.1.5, **10.2**, 10.3, 11.1, 12.1, 13.2, 14.1
Biermann, K.-R. 2.1.1, 9.1.2
Biot, J. B. 6.3.2, 10.8.3
Birg, S. 6.2.2
Black, W. 6.2.3
Block, M. 10.7
Boltzmann, L. 10.8.5, 11.2
Bolyai, J. 9.1.5
Bolyai, W. 9.1.5
Bonpland, A. J. A., 10.8.2
Boole, G. 8.1.1, 10.8.4, 12.2
Bortkiewicz, L. von 3.2.3, 8.1.2, 8.1.5, 10.4.6, 10.7, 11.3, 13.3, 14.1.1, **14.1.2**
Boscovich, R. J. 3.2.3, 6.3, 6.3.2, 7.2-5, 7.3, 10.8.4, 10.8.5
Bouillaud, J. 10.8
Bowditch N. 7.2-5
Boyle, R. 1.7, 10.9.3
Bradley, J. 9.3
Brahe, T. 1.6, 1.7, 6.3, 10.7
Bredikhin, F. A. 10.8.4

Brendel, M. 9.1.5
 Brownlee, J. 10.8.1
 Bru, B. 6.3.2, 10.2, 10.2.5, 11.1
 Bruns, H. 13.3
 Brush, S. G. 13.3
 Budd, W. 6.2.2
 Bühler, G. 1.3
 Buffon, G. L. L. 3.3.4, **6.1.4**, 6.1.6, 7.1-4, 10.4.3
 Bull, J. P. 8.5
 Buniakovsky, V. Ya. 8.1.5, **10.4**
 Burov, V. G. 1.4, 3.2.2
 Buys Ballot, C. H. D. 10.8.3
Camp, B. H. 14.2
 Campbell, L. 10.8.5-3
 Cardano, G. 2.1.1, 3.2.3
 Cauchy, A. L. 8.1.5, 9.3, **10.1**, 11.3, 12.1
 Chadwick E. 10.8.1
 Chapman, S. 2.1.4
 Charlier, C. V. L. 10.2.5
 Chebyshev, P. L. 5.1, 7.1, 7.1-3, 8.1.5, 10.2-2, 10.2-3, 10.4-7, 11.3, **12**, 13.2 – 13.5
 Chuprov, A. A. 6.1.3, 8.1.5, 10.3, 10.7, 11.3, 13.2, 13.3, 14.1.1, 14.1.2, **14.1.3**, 14.2, 14.3
 Cicero M. T. (106 – 43 BC) 3.1.2, 3.2.1, 3.2.2
 Clausius, R. 10.8.4, 10.8.5
 Cohen, J. 1.4
 Columbus Chr. 2.1.4
 Commelin, C. 2.1.3
 Comte, A. 10.8
 Condamine, C. M. de la 6.2.3, 6.3.1
 Condorcet, M. J. A. N. 2.2.3, 3.3.4, **6.1.5**, 6.2.3, 6.2.4, 6.3.1, 10.8.1, 11.1, 11.3
 Confucius (ca. 551 – 479 BC) 1.4
 Coolidge, J. L. 9.1.2
 Copernicus, N. 1.6
 Cotes, R. 6.3.1
 Cournot, A. A. 1.1, 5.1, 6.1.3, 8.1.2, 8.1.5, **10.3**, 10.7, 10.8.4, 11.1, 11.2, 14.1
 Courtaut, J.-M. 13.3
 Cramer, G. 3.3.4, 6.1.1
 Cramér, H. 9.1.3
 Crofton, M. W. 11.2
 Cubranic, N. 6.3.2
 Czuber, E. 5.1, 9.2, 10.7, 11.2
D'Alembert, J. Le Rond 5.1, **6.1.2**, 6.1.3, 6.1.5, 6.2.3, 6.3.1, 9.2, 10.7, 10.8, 11.2
 D'Amador, R. 10.8
 Darboux, G. 11.2
 Darwin, C. 10.5, 10.6, 10.8.2, 10.8.5-4, 11.2, 14.2
 Davenport, C. B. 14.2
 David, F. N. 2.1.1, 13.3
 David, H. A. 2.2.4, 3.3.3, 5.1, 10.9.2
 Davidov, 8.1.1, 8.1.5
 Daw, R. H. 6.2.2
 DeCandolle, Aug. P. 10.8.2
 Delambre, J. B. J. 10.7
 De Moivre, A. 1.3, 2.2.2, 2.2.3, 3.2.3, 3.3.3, 3.3.4, **4**, 5.2, 5.3, 6.1.1, 6.1.6, 6.3.1, 7.1-1, 7.1-3, 8.1.2, 8.1.5, 11.1, 13.2, 14.1
 De Montessus, R. 11.2
 De Morgan, A. 4.4
 De Morgan, Sophia 4.4
 De Solla Price, D. J. 1.5
 Derham, W. 2.2.3
 Descartes, R. 2.1.2, 7.3
 De Witt, J. 2.1.3, 3.2.3
 Dietz, K. 6.2.3, 10.8.1
 Dirac, P. A. M. 7.2-1, 8.1.1

Dirichlet, P. G. L. 7.1-2, 12.2
 Doppler, C. 10.8.4
 Dormoy, E. 11.1, 14.1.1
 Double, F. J. 7.2-1, 8.1.1, 8.5, 10.8.1
 Dove, H. W. 10.8.3
 Dreyfus, A. 8.3, 11.3
 Drinkwater-Bethune, J. E. 5.1
 Dufau, P. A. 10.7
 Du Pasquier, L. G. 2.1.3
 Dutka, J. 3.3.4, 9.1.5, 13.3
Ebbinghaus, H. 10.9.2
 Eddington, A. S. 9.2, 10.8.4
 Edgeworth, F. Y. 14.2
 Edwards, A. W. F. 2.2.1, 2.2.4, 3.3.3, 5.1
 Ehrenfest, P. 6.1.1, 7.1-3
 Ehrenfest, T., see Afanasieva-Ehrenfest T.
 Eisenhart, C. 9.1.3, 9.1.4, 14.3
 Elkin, W. L. 10.8.4
 Emeliakh, L. I. 13.1
 Encke, J. F. 9.1.5
 Eneström G. 2.1.3
 Engels, F. 10.8.5-3
 Enko, P. D. 10.8.1
 Erathosthenes (ca. 276 – 194 BC) 1.5
 Erdélyi, A. 11.3
 Erman, A. 9.3
 Ermolaeva, N. S. 12.2
 Euler, L. 2.1.3, 2.1.4, 6.1.2, 6.2.2, 6.3.1, 6.3.2, 7.1-3, 7.2-1, 9.1.5, 10.4-7
Faraday, M. 10.8.3, 14.2
 Farr, W. 10.8.1
 Fechner, G. T. 6.1.1, **10.9.2**
 Fedorovitch, L. V. 2.1.4
 Feller, W. 7.1-6
 Fermat, P. 2.2.1, 2.2.2
 Fisher, R. A. 5.1, 6.1.6, 9.2, 10.8.2, 10.8.4, 14.2
 Flamsteed, J. 1.7
 Fletcher, A. 13.2
 Florensky, P. A. 13.5
 Fourier, J. B. J. 7.3, 8.1.1, 9.1.4, 10.5, 10.8, 11.3
 Fournival, R. de 2.1.1
 Franklin, J. 2.2.1, 3.2.1
 Fréchet, M. 10.5
 Freud, S. 10.9.2
 Freudenthal, H. 2.2.4, 3.3.4, 10.1
 Fuss, P. N. 6.1.1
Galen, C. (129 – 201?) 1.1, 1.3, 1.4
 Galilei, G. 1.3, 2.1.1, 6.3.2, 7.3
 Galle, A. 9.1.5
 Galton, F. 10.2, 10.5, 10.2-6, **10.6**, 10.9.2, 14.2
 Garber, D. 3.2.1
 Garnett, W. 10.8.5-3
 Gatterer, J. C. 10.7
 Gauss, C. F. 3.1.2, 6.3.1, 6.3.2, 7.2-2, 7.3, 8.2, **9.1**, 9.3, 10.1, 10.9.1, 11.1, 11.3, 12.2, 13.5, 14.1.1
 Gavarrat, J. 8.5, 10.7, 10.8
 Gerardy, T. 9.1.5
 Gillispie, C. 8.4
 Gnedenko, B. V. 10.2-3, 14.3
 Goldstein, B. R. 1.5
 Gosset, W. S. (Student) 14.2
 Gower, B. 3.2.3
 Gowing, R. 6.3.1

Gram, J. P. 10.2-5
 Graunt, J. 2.1.4, 2.2.2 – 2.2.4, 3.1.1, 3.2.2, 6.2.2, 8.5, 10.7
 Greenwood, M. 2.1.3, 2.1.4, 10.8
 Gridgeman, N. T. 7.1-4
 Grodsensky, S. Ya. 13.1
 Guerry, A. M. 6.2.2
 Gumbel, E. J. 6.2.2
Hald, A. 2.1.1, 2.1.3, 2.1.4, 2.2.2, 3.3.3, 3.3.4, 4.1 – 4.3, 6.3.2, 7.1-3, 7.1-5, 8.1.5, 8.5, 9.1.2, 9.1.4, 9.2, 10.2-3, 10.2-5, 10.7, 10.8.3, 10.8.4, 14.2, 14.3
 Halley, E. 2.1.4, 3.2.3, 4.2
 Halperin, T. 3.2.1
 Hartley, D. 5.3
 Haushofer, D. M. 3.2.3
 Hellman, C. D. 1.6
 Helmert, F. R. 3.2.3, **9.2**
 Hendriks, F. 2.1.3
 Henny, J. 3.3.3
 Henry, M. Ch. 6.1.5
 Hermite, C. 7.1-3, 12.1
 Herschel, W. 6.3.2, 10.8.4
 Hertz, H. 10.8.5-4
 Heuschling, X. 10.5
 Heyde, C. C. 8.1.5, 8.3, 10.1, 10.2, 12.1, 14.3
 Hill, D. 10.8.4
 Hipparchus (180 or 190 – 125 BC) 1.5
 Hippocrates (460 – 377 or 356 BC) 1.3
 Hobbes, T. 1.1
 Hogan, E. R. 9.1.2
 Holden, E. S. 10.8.4
 Hostinský, B. 7.1-3
 Humboldt, A. 1.3, 9.3, 10.8.2, 10.8.3
 Huygens, C. 2.1.1, 2.1.3, 2.1.4, **2.2.2**, 3.1.2, 4.3, 6.1.2
Idelson, N. I. 13.2
 Irwin, J. O. 14.2
 Ivory, J. **10.9.1**
Jenner, E. 6.2.3
 Jevons, W. S. 8.1.1
 Johansenn, W. 14.2
 John of Salisbury (1115 or 1120 – 1180) 3.2.1
 Jorland, G. 3.3.4
 Juskevic, A. P. see Youshkevitch, A. P.
Kac, M. 10.8.5-3
 Kagan, V. F. 12.3
 Kant, I. 1.1
 Kapteyn, J. C. 10.8.4
 Karn, M. Noel, 6.2.3
 Kaufman, A. A. 10.7
 Kendall, D. G. 10.2-6
 Kendall, M. G. (Sir Maurice) 2.1.1, 2.1.4, 9.2, 10.9.3, 11.2
 Kepler, J. 1.1, 1.6, 1.7, 2.1.2, 2.1.4, 3.2.3, 3.2.4, 6.3.2, 10.7
 Khinchin, A. Ya. 10.8.5, 11.3
 Kiaer, A. N. 10.8.4
 Kington, J. A. 6.2.4
 Klein, F. 12.3
 Knapp, G. F. 10.5, 10.7
 Knies, C. G. A. 6.2.1
 Knott, C. G. 10.8.5
 Köppen, W. 10.5, 10.8.3
 Körber, H.-G. 10.8.3
 Kohli, K. 2.1.3, 2.2.2, 3.1.2, 3.3.2, 3.3.4
 Koialovitch, B. M. 13.3
 Kolmogorov, A. N. 7.2-5, 8.1.2, 9.2, 10.8.5-3, 12.3, 14.2

Koopman, B. O. 3.2.1
 Kowalski, M. A. 10.8.4
 Krasovsky, F. N. 6.3, 10.9.1
 Krein, M. G. 12.1
 Kronecker, L. 9.1.3, 12.2
 Kruskal, W. 6.3.1
 Krylov, A. N. 12.2
Lagrange, J. L. 4.3, 6.1.1, 6.3.1, 7.1-3
 Lamarck, J. B. 1.1, 10.8.3
 Lambert, J. H. 3.2.1, **6.1.3**, 6.2.2, 6.2.4, 6.3.1, 6.3.2, 8.1.1, 8.5, 10.3-4
 Lamont, J. 10.8.3
 Langevin, P. 10.8.5-3
 Laplace, P. S. 1.1, 2.1.1, 2.2.3, 2.2.4, 4.1, 4.3, 4.4, 5.1, 5.2, 6.1.1, 6.1.5, 6.1.6, 6.2.3, 6.3.1, 6.3.2, **7**, 8.1.2, 8.1.3, 8.2, 8.3, 9.1.3, 10.3-1, 10.4-5, 10.4-7, 10.7, 10.8.4, 10.8.5, 11.1, 11.3, 12.2, 13.3, 13.5
 Laurent, H. 10.4-7, 11.2
 Le Cam, L. 11.3
 Legendre, A. M. 6.3.2, 8.2, 9.3
 Leibniz, G. W. 2.1.1 – 2.1.4, 3.1.2, 6.1.3, 6.2.3, 6.2.4, 7.3, 8.3, 8.5, 10.8, 10.8.1
 Lenin, V. I. 14.2
 Levi ben Gerson (1288 – 1344) 1.5, 3.2.4
 Lévy, M. 10.8.1
 Lévy, P. 13.2-5
 Lexis, W. 10.2-2, 10.7, 14.1, **14.1.1**
 Liapunov, A. M. 10.2-1, 12.1 – 12.3, 13.3 – 13.5, 14.1, 14.1.2, 14.3
 Libri-Carrucci, G. B. I. T. 8.1.1
 Liebermeister, C. 8.5
 Lindeberg, J. W. 13.4
 Linnik, Yu. V. 10.1, 10.8.5-3
 Lloyd, G. E. R. 1.5
 Lhuillier, S. A. 3.2.1
 Lobachevsky, N. I. 12.3
 Louis, P. C. A. 8.5, 10.8
 Lubbock, I. W. 5.1
 Lueder, A. F. 10.7
Mach, E. 14.2
 Maciejewski, C. 3.2.3
 Maclaurin, C. 7.1-3
 Mahalanobis, P. C. 14.2, 14.3
 Maimonides M. (1135 – 1204) 1.2, 1.4, 1.6
 Maire, C. 6.3, 6.3.2
 Makovelsky, A. O. 1.4
 Malfatti, F. 7.1-3
 Malthus, T. R. 6.2.2
 Maltsev, A. I. 9.2
 Mann, W. 10.8.2
 Mariotte, E. 1.7.
 Markov, A. A., Jr 13.1
 Markov, A. A., Sr 3.2.3, 4.4, 6.1.1, 7.1-3, 7.1-6, 8.1.5, 9.1.3, 10.2-3, 10.4, 10.4-3, 11.3, 12.1, 12.2, **13.1 – 13.3**, 13.5, 14.1.2, 14.1.3, 14.3
 Marsden B. G. 9.1.5
 Maupertuis, P. L. M. 1.2, 6.3.2, 7.3
 Maxwell, J. C. 9.1.3, 10.8.5, 11.3
 May K. O. 9.1.2, 9.1.3
 Mayer, T. 6.3.2
 Mayr, G. von 10.7
 Meadowcroft, L. V. 7.2-4
 Mendel, J. G. 10.8.2
 Mendeleev, D. I. 10.8.3, 10.8.4, **10.9.3**
 Merrington, M. 14.2
 Meyer, H. 10.8.3
 Michell, J. 6.1.6, 10.8.4, 11.2

Mill, J. S. 3.1.2, 8.3, 11.3
 Mises, R. von 5.1, 5.2, 9.1.4, 10.9.2
 Molina, E. C. 7.1-3, 7.1-8
 Montmort, P. R. 2.1.1, **3.3.3**, 3.3.4, 4.3, 6.1.2, 6.3.1
 Moran, P. A. P. 11.2
 Morant, G. M. 14.2
 Muncke, G. W. 6.2.4
Nekrasov, P. A. 10.2-3, 13.1 – 13.4, **13.5**, 14.1.2
 Neugebauer, O. 1.5
 Newcomb, S. 6.1.6, 10.8.4, 14.2
 Newton, I. 1.1, 2.1.4, **2.2.3**, 3.2.3, 4.3, 6.1.6, 6.3, 6.3.2, 7.3
 Neyman, J. 5.1, 13.1, 13.3
 Nicole, P. 2.1.1, 2.1.4, 2.2.1, 3.1.2, 3.3.1
 Nicomachus of Gerasa (ca. 100 BC) 1.4
 Nieuwentit, B. (1654 – 1718), 2.2.3
 Nightingale, F. 2.2.3, 10.8.1
 Novikov, S. P. 12.3
 Novoselsky, S. A. 10.4-6
Olbers H. W. 9.1.5, 9.3, 10.9.1
 Ondar, Kh. O. 8.1.1, 13.3, 14.1.2, 14.1.3
 Ore, O. 2.1.1, 3.2.3
 Oresme, N. (ca. 1323 – 1382) 3.2.4
Paevsky, V. V. 6.2.2
 Pannekouk, A. 1.5
 Pascal, B. 2.1.1, 2.2.1, 2.2.2, 4.1
 Paty, M. 6.1.2
 Pearson, E. S. 2.2.3, 14.2, 14.3
 Pearson, K. 2.1.3, 2.1.4, 2.2.3, 3.2.3, 5.1, 6.1.4, 6.2.2, 6.2.3, 6.3.1, 7.1-4, 7.1-5, 10.5, 10.6, 10.8.3, 10.8.4, 10.9.2, 13.2, 13.3, 14.2, 14.3
 Petrov, V. V. 9.1.3
 Pettenkofer, M. 10.8.1
 Petty, W. 2.1.1, 2.1.4
 Pfanzagl, J. 6.2.2
 Picard, J. 6.3.1
 Pirogov, N. I. 10.8.1
 Plackett, R. L. 9.1.3, 9.1.5
 Plato, J. von 11.3
 Poincaré, H. 1.1, 8.3, 10.8.4, 10.8.5-3, 11.1, 11.2, **11.3**, 12.2, 13.3
 Poinsot, L. 8.3
 Poisson, S.-D. 2.1.1, 2.2.4, 3.2.3, 4.1, 5.2, 6.1.5, 7.1-8, 8, 10.1, 10.2-2, 10.5, 10.7, 10.8.4, 10.8.5, 11.1 – 11.3, 12.1 – 12.3, 13.3, 14.1, 14.1.2
 Polya, G. 4.4, 5.1
 Postnikov, A. G. 12.2
 Prevost, P. 3.2.1
 Price, R. 2.2.3, 5.1 – 5.3
 Proctor, R. A. 6.1.6, 10.8.4
 Prokhorov, Yu. V. 5.1, 10.8.5-3, 12.1, 12.2
 Prudnikov, V. E. 12.2
 Ptolemy C. (Ilc.) 1.5, 1.6, 3.2.3, 6.3.2
Quetelet, A. 2.2.3, 6.1.4, 6.2.2, 9.1.3, **10.5**, 10.7, 10.8.1 – 10.8.3, 14.1
Rabbi Schlomo ben Adret, 1235 – 1310, 1.2
 Rabinovitch, N. L. 1.2, 1.4
 Radelet de Grave, P. 6.2.4
 Raikher, V. K. 1.4
 Raimi, R. A. 10.8.4
 Raju, C. K. 2.1.1
 Réaumur, R. A. 10.8.2
 Rehnisch, E. 10.5
 Riemann, B. 12.3
 Rigaud, S. P. 1.7
 Robinson, G. 6.3.2
 Romanovsky, V. I. 3.2.3, 13.3, 14.1.3

Rubanovsky, C. M. 10.8.5-4
Sakatov, P. S. 6.3
 Saunderson, N. 5.3
 Schindler, A. 10.8.2
 Schlözer, A. L. 6.2.1
 Schmidt, O. Ju. 11.2
 Schumacher, H. 14.1.2
 Schumacher H. C. 9.1.2, 9.1.5
 Seidel, L. 10.8.1
 Seneta, E. 4.1, 8.1.5, 8.3, 8.5, 10.1, 10.2, 11.2, 12.1, 13.3, 13.5, 14.1.3, 14.3
 Sevastianov, B. A. 12.2
 Shafer, G. 3.2.1
 Shaw, N. 10.8.3
 Shoesmith, E. 2.2.2, 2.2.4, 6.3.1
 Short, J. 6.3.1
 Simon, J. 6.2.3
 Simpson, J. Y. 10.8.1
 Simpson, T. 4.2, 6.1.6, 6.3.1
 Sleshinsky, I. V. 12.1
 Slutsky, E. E. 13.2
 Snow, J. 10.8.1
 Sofonea, T. 2.1.4
 Soloviev, A. D. 13.5
 Spieß, O. 3.3.4
 Steklov, V. A. 7.1-3, 10.4, 10.4-7
 Stieltjes, T. J. 12.1
 Stigler, S. M. 5.3, 6.3.2, 9.1.5
 Stirling, J. 4.4, 5.2, 12.2, 13.2
 Struve, P. B. 10.4
 Student (Gosset, W.S.) 9.2, 13.1
 Süßmilch, J. P. 2.1.3, 2.1.4, 2.2.3, 6.2.2, 14.1
 Sylvester, J. J. 11.2
Tait, P. G. 11.3
 Takács (Takács), L. 2.2.1, 3.3.4, 4.1, 11.1
 Taqqi, M. S. 13.3
 Thatcher, A. R. 3.3.4
 Tikhomandritsky, M. A. 12.3
 Tikhomirov, E. I. 6.2.4
 Timerding, H. E. 5.2, 9.1.4
 Toaldo, J. 6.2.4
 Todhunter, I. 3.3.2, 4.3, 6.1.1, 6.1.2, 6.1.5, 6.2.3, 6.3.1, 7.1-3, 7.1-7, 14.2
 Tolstoy, L. N. 13.1
 Tonti, L. 2.1.3
 Toomer, 1.5
 Truesdell, C. 7.2-1, 10.8.5
 Tsinger, V. Ya. 9.1.3
 Turgot, A. R. 6.1.5
 Tutubalin, V. N. 1.4
Ulpianus D. (ca. 170 – 228) 2.1.3
Venn, J. 8.1.1
 Voltaire (Arouet, M. E.) 2.2.3
Waerden, B. L. van der, 2.1.3, 2.2.2, 10.8.2
 Wallis, J. 6.1.1
 Waterston G. 10.8.5
 Weber, E. H. 6.1.1, 10.9.2
 Weber W. E. 9.1.5
 Weldon, W. F. R. 13.3, 14.2
 Weyl, H. 10.8.4, 11.3
 White, A. D. 6.2.3
 Whiteside, D. T. 2.2.3
 Whittaker, E. T. 6.3.2
 Wittstein, Th. 10.7

Wilks, S. S. 7.1-2
William III 2.1.3
Wilson, C. 6.3.2
Winkler, W. 14.1.2
Wishart, J. 14.2
Wolf, Abr. 6.2.4
Woolhouse, W. S. B. 10.7
Woytinsky W. S. 14.1.2
Yamazaki, E. 6.1.2, 6.1.5
Yarochenko, S. P. 13.5
You Poh Seng 10.8.4
Youshkevitch, A. A. 13.2
Youshkevitch, A. P. 3.3.4, 6.1.2
Yule, G. U. 10.5, 10.9.3, 13.1
Zabel, S. L. 3.2.1
Zach, F. X. von 9.1.5
Zermelo, E. 10.8.5-3