Probability and Statistics in the 18th Century
Oscar Sheynin

This, now slightly revised text, is intended for a broader circle of readers. It appeared in Italian, although with suppressed references, as

Lo sviluppo della teoria della probabilità e della statistica in *Storia della Scienza*, t. 6. Roma, Ist. Enc. Ital., 2002, pp. 529 – 541

Contrary to the official agreement, the original English version was not published.

1. Introduction

The theory of probability can be traced back to 1654 when **Pascal** and **Fermat**, in solving the problem of points (of sharing the stakes in an uncompleted series of games of chance), indirectly introduced the notion of expected gain (of the expectation of a random variable). In 1657, **Huygens** published the first treatise on probability. There, he applied the new notion (although not its present term) for studying games of chance. His materials of 1669, which remained unknown during his lifetime, included solutions of stochastic problems in mortality. Later, in 1690, following **Descartes**, he stated that natural sciences only provided morally certain (highly probable) deductions.

Moral certainty and the application of statistical probability were discussed in in philosophical literature (Arnauld & Nicole 1662) which influenced **Jakob Bernoulli**, the future cofounder of probability theory (§2). **Petty** and **Graunt**, in the mid-17th century, created political arithmetic whose most interesting problems concerned statistics of population and its regularities. Having extremely imperfect data, the latter was nevertheless able to compile the first mortality table and to study medical statistics. In 1694 **Halley** calculated the second and much better table and laid the foundation of stochastic calculations in actuarial science. **Newton** applied stochastic reasoning to correct the chronology of ancient kingdoms, and, in a manuscript written between 1664 and 1666, invented a simple mind experiment to show that the then yet unknown geometric probability was capable of treating irrational proportions of chances.

2. The First Limit Theorem

Jakob Bernoulli blazed a new trail in probability. His *Ars Conjectandi* posthumously published in 1713 contained a reprint of **Huygens**' treatise with essential comment; a study of combinatorial analysis; solutions of problems concerning games of chance; and an unfinished part where he provided (but had not applied) a definition of theoretical probability, attempted to create a calculus of stochastic propositions, and proved his immortal theorem.

Here it is. Bernoulli considered a series of *Bernoulli trials*, of $\nu = (r + s)n$ independent trials in each of which the studied event A occurred with probability $p = r/(r + s)$. If the number of such occurrences is μ, then, as he proved,

$$P\left(\left|\frac{\mu}{\nu} - p\right| \le \frac{1}{r+s}\right) \ge \frac{c}{1+c}$$

where c was arbitrary and $\nu \geq 8226 + 5758 \lg c$. It followed that

$$\lim P(|\frac{\mu}{\nu} - p| < \varepsilon) = 1. \quad (1)$$

Bernoulli thus offered the (weak) law of large numbers and established the parity between the theoretical probability p and its statistical counterpart μ/ν. Given a large number of observations, the second provided moral certainty and was therefore not worse than the first. To paraphrase him: He strove to discover whether the limit (1) existed and whether it was indeed unity rather than a lesser positive number. The latter would have meant that induction (from the ν trials) was inferior to deduction! The application of stochastic reasoning well beyond the narrow province of games of chance, sufficiently serviced by the theoretical probability, was now justified, at least for the *Bernoulli trials.*

3. Montmort

His treatise on games of chance (1708) unquestionably influenced **De Moivre**. Unlike **Huygens**' first attempt (§1), his contribution was a lengthy book rich in solutions of many old and new problems. One of the former, which **Galileo** solved in a particular case by simple combinatorial formulas, was to determine the chances of throwing k points with n dice, each of them having f faces (alternatively: having differing number of faces). In this connection Montmort offered a statement that can now be described by the formula of inclusion and exclusion: For events $A_1, A_1, ..., A_n$,

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

where $i, j, k, \ldots = 1, 2, \ldots, n, i < j, i < j < k, \ldots$ This formula is a stochastic corollary of the appropriate general proposition about sets $A_1, A_2, \ldots, A_n$ overlapping each other in whichever way. For f = Const = 6 (say), the problem stated above is tantamount to determining the probability that the sum of n mutually independent random variables taking equally probable values 1, 2, …, 5, 6 equals k.

In 1713 Montmort also inserted his extremely important correspondence with **Niklaus Bernoulli**. One of the topics discussed by them in 1711 – 1713 was a strategic game (*her*), – a game depending both on chance and the decisions made. A theory of such games was only developed in the 20th century. For other subjects of their letters see §§6 and 10.2.

4. De Moivre

His main contribution was the *Doctrine of Chances*, where, beginning with its second edition, he incorporated his derivation of the **De Moivre – Laplace** limited theorem privately printed in 1733 but accomplished by him a *dozen years or more* earlier. And his memoir of 1712, which appeared before Jakob Bernoulli's posthumously published *Ars Conjectandi* did, can be considered as its preliminary version. It was there that he introduced the classical definition of probability, usually attributed to Laplace.

The *Doctrine* was written for non-mathematical readers. It provided solutions of many problems in games of chance but did not concentrate on scientific topics, and the proofs of many propositions were lacking. Nevertheless, this book contained extremely important findings, see below and §10.1, and both **Lagrange** and **Laplace** thought of translating it into French, see Lagrange's letter to Laplace of 30.12.1776 in t. 14 of his *Oeuvres*.

I describe now the theorem mentioned above. Desiring to determine the law underlying the ratio of the births of the two sexes (§6), De Moivre proved that for n Bernoulli trials with probability of *success* p, the number of successes μ obeyed the limiting law

$$\lim P\left(a \le \frac{\mu - np}{\sqrt{npq}} \le b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b \exp(-z^2/2)dz,\ n \to \infty \quad (2)$$

with $q = 1 - p$. Note that $np = E\mu$ and $npq = \text{var}\mu$, the expectation and variance of μ (the second notion is essentially due to Gauss).The convergence implied in (2) is uniform with respect to a and b, but, again, this is a concept introduced in the 19th century. When deriving his formula, De Moivre widely used expansions of functions into power series (sometimes into divergent series calculating the sums of several of their first terms).

Thus appeared the normal distribution. De Moivre proved (2) for the case of $p = q$ (in his notation, $a = b$) and correctly stated that his formula can easily be generalized to $p \ne q$; furthermore, the title of his study included the words *binomial* $(a + b)^n$ *expanded* … He had not however remarked that the error of applying his formula for finite values of n increased with the decrease of p (or q) from 1/2 , or, in general, had not studied the rapidity of the convergence in (2).

In following the post-Newtonian tradition, De Moivre did not use the symbol of integration; his English language was not generally known on the Continent; **Laplace** (1814) most approvingly mentioned his formula but had not provided an exact reference or even stated clearly enough his result; and Todhunter (1865), the best pertinent source of the 19th century, superficially described his finding. No wonder that for about 150 years hardly any Continental author noticed De Moivre's theorem. In 1812, Laplace proved the same proposition (hence its name introduced by **Markov**) by means of the **McLaurin – Euler** summation formula and provided a correction term which allowed for the finiteness of the number of trials.

Scientific demands led to the studying of new types of random variables whose laws of distribution did not coincide with **Jakob Bernoulli's** and De Moivre's binomial law. Nevertheless, the convergence of the sums of these variables to the normal law persisted under very general conditions and this fact is the essence of the central limit theorem of which (2) is the simplest form.

5. Bayes

His fundamental posthumous memoir of 1764 was communicated and commented on by Price. Bayes' *converse* problem, as Price called it, was to determine the unknown theoretical probability of an event given the statistical probability of its occurrence in Bernoulli trials. Here, in essence, is his reasoning. A ball falls $\alpha + \beta = n$ times on a segment AB of unit length so that its positions on AB are equally probable and c is somewhere on AB with all its positions also equally probable; α times the ball falls to the left of c (α *successes*) and β times, to the right (β *failures*; statistical probability of success, α/n). It is required to specify point c. For any $[a; b]$ belonging to AB

$$P(c \in [a; b]) = \int_a^b C_n^\alpha x^\alpha (1 - x)^\beta dx \div \int_0^1 C_n^\alpha x^\alpha (1 - x)^\beta dx. \qquad (3)$$

This is the posterior distribution of *c* given its prior uniform distribution with the latter representing our prior ignorance. The letter *x* in (3) also stands for the unknown *Ac* which takes a new value with each additional trial. At present we know that

$$P = I_b(\alpha + 1; \beta + 1) - I_a(\alpha + 1; \beta + 1)$$

where *I* is the symbol of the incomplete Beta function. The denominator of (3), as Bayes easily found out, was (the complete Beta function times the factor C_n^α) the probability

$$P \text{ (The number of successes} = \alpha \text{ irrespective of } Ac) = 1/(n + 1)$$

for any acceptable value of α. Even up to the 1930's the estimation of the numerator for large values of α and β had been extremely difficult and some commentators believe that Bayes did not publish his memoir himself because he was dissatisfied with his efforts in this direction.

Anyway, it seems that he had not rested content with limiting relations since they were not directly applicable to the case of finite values of *n* (at least Price said so with regard to the work of **De Moivre**). However, Timerding, in his translation of the Bayes memoir into German (1908), proved that the latter's calculations could have led to

$$\lim P\left(a \le \frac{x - \alpha/n}{\sqrt{\alpha\beta / n^{3/2}}} \le b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b \exp(-z^2/2)dz, \; n \to \infty$$

where, as I myself note, $\alpha/n = Ex$ and $\alpha\beta/n^3 = \text{var } x$.

It is remarkable that Bayes, who (just like De Moivre) certainly had not known anything about variances, was apparently able to perceive that an elementary and formal transformation of the left side of (2) leading to

$$P\left(a \le \frac{\mu/n - p}{\sqrt{pq/n}} \le b\right)$$

would not have provided the proper answer to his problem. Both Jakob Bernoulli, and De Moivre mistakenly thought that they had solved the inverse problem as well just by solving the direct problem.

Only Bayes correctly perceived the proper relation between the statistical and theoretical probabilities and thus completed the first version of the theory of probability. Mises, who postulated that the theoretical probability of an event is the limit of the statistical probability of its occurrence, could have referred to Bayes; moreover, in various applications of probability this Mises conception is inevitably made use of, but the references could be and even should be made to Bayes as well!

On another level, Bayes' main result was, that, given a random variable with a superficially known distribution, it is possible to specify it by means of observation. Thus, all possible positions of c on AB were thought to be equally possible, but the n trials led to distribution (3).

Price provided an example which presumed complete previous ignorance: Sunrise had been observed a million times in succession; how probable becomes the next sunrise? According to formula (3) with $a = 1/2$, $b = 1$, $\alpha = 10^6$ and $\beta = 0$, he found that the odds of success were as the millionth power of 2 to one.

Just as it was with **De Moivre** (§4), Continental mathematicians were hindered from studying the Bayes memoir by his English language and his failure to interpret his subtle reasoning, see Gillies (1987), who discusses the recent debates (and reasonably describes Price's own contribution).

Let incompatible events $A_1, A_2, \ldots, A_n$, have probabilities $P(A_i)$ before an event B happens; suppose also that B occurs with one, and only one of the A_i's, after which these events acquire new probabilities. Then

$$P(A_i/B) = P(B/A_i)P(A_i) \div \sum_{j=1}^{n} [P(B/A_j)\, P(A_j)].$$

This is the so-called Bayes formula, see **Cournot** (1843, §88), nevertheless lacking in the Bayes memoir. However, in the discrete case it also describes the transition from prior probabilities to posterior. It was **Laplace** (1774) who had expressed it (in words only) and proved it later (1781, p. 414).

Laplace (1786) also extended the Bayes method by treating non-uniform prior distributions. And, without mentioning Bayes, he solved several problems leading to formulas of the type of (3). Best known is his calculation of the probability of the next sunrise already observed α times in succession. He (1814, p. 11 of the translation) stated, but did not prove, that this probability was $(\alpha + 1)/(\alpha + 2)$ but the explanation is in one of his earlier memoirs (1781). In 1774 he began to consider relevant urn problems, and in 1781 he went on to study the sex ratio at birth (also see §6).

An urn contains an infinite number of white and black balls. Drawings without replacement produced p white balls and q black ones; determine the probability that a white ball will be extracted next. Denote the unknown ratio of the number of white balls to all of them by x, then the obtained sample has probability
$x^p(1-x)^q$, and, since all values of x should be regarded as equally probable, the probability sought will be

$$P = \int_0^1 x \cdot x^p(1-x)^q dx \div \int_0^1 x^p(1-x)^q dx = \frac{p+1}{p+q+2}.$$

Hence (if $p = \alpha$ and $q = 0$) the conclusion above. Note that the result obtained coincides with the expectation of a random variable with density

$$\varphi(x) = Cx^p(1-x)^q,\ C = 1 \div \int_0^1 x^p(1-x)^q dx.$$

Determine now the probability of drawing m white balls and n black ones in the next $(m + n)$ extractions if these numbers are small as compared with p and q. This time making use of approximate calculations, Laplace got

$$P = \frac{p^m q^n}{(p+q)^{m+n}}$$

and noticed that this was in agreement (as it should have been) with assuming that $x \approx p/(p + q)$.

Finally, also in 1774, Laplace proved that for an arbitrary $\alpha > 0$

$$\lim P\left(\frac{p}{p+q} - \alpha \le x \le \frac{p}{p+q} + \alpha\right) = 1,\ p,\ q \to \infty.$$

In 1781 he applied this result to state that, when issuing from extensive statistical data, the sex ratio at birth could be calculated as precisely as desired [provided that it remained constant!]. See §11 for still another related problem studied by Laplace.

The difference between the statistical and the theoretical values of such magnitudes as $p/(p + q)$ could have also been estimated by means of the **De Moivre – Laplace** theorem; indeed, for $p,\ q \to \infty$ the probabilities of extracting balls of the two colours remain constant even when they are not returned back into the urn.

6. Population Statistics

The fathers of political arithmetic (§1) had good grounds to doubt, as they really did, whether quantitative studies of population were necessary for anyone excepting the highest officials. Indeed, social programmes began appearing in the 1880's (in Germany); before that, governments had only been interested in counting taxpayers and men able to carry arms.

A new study belonging to population statistics, the calculation of the sex ratio at birth, owed its origin and development to the general problem of isolating randomness from Divine design. **Kepler** and **Newton** achieved this aim with respect to inanimate nature, and scientists were quick to begin searching for the laws governing the movement of population.

In 1712 Arbuthnot put on record that during 82 years (1629 – 1710) more boys had been yearly christened in London than girls. Had the probability of a male birth been 1/2, he continued, the probability of the observed fact would have been 2^{-82}, i.e., infinitesimal. He concluded that the predominance of male births was a Divine law which *repaired* the comparatively higher mortality of men.

Even now the divide between random and non-random sequences remains more than subtle, but at least Arbuthnot's series *m, m, m*, … could not have been attributed to chance.

Nevertheless, his reasoning was feeble. Baptisms were not identical with births; Christians were perhaps somehow different from others, and London could have differed from the rest of the world; and, finally, the comparative mortality of the two sexes was unknown. A special point is that Arbuthnot only understood randomness in the sense of equal chances of a male and female birth whereas the supposed Divine law could have well been expressed by a general binomial distribution with $p > 1/2$.

De Moivre (§4) and **Niklaus Bernoulli** had developed Arbuthnot's arguments. Here is the latter's result which he formulated in a letter to **Montmort** of 1713. Denote the ratio of registered male births to those of females by m/f, the total yearly number of births by n, the corresponding number of boys by μ and set
$n/(m+f) = r$, $m/(m+f) = p$, $f/(m+f) = q$, $p + q = 1$ and let $s = 0(\sqrt{n})$. Then Bernoulli's derivation (Montmort 1708, pp. 388 – 394 in 1713) can be presented as follows:

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

The last formula means that Bernoulli indirectly, since he had not written it down, introduced the normal law as the limit of the binomial distribution much earlier than De Moivre (directly) did. However, his finding does not lead to an integral limit theorem since s should remain small as compared with n (see above), and neither is it a local theorem.

In the mid-18th century **Achenwall** created the Göttingen school of Staatswissenschaft (statecraft) which strove to describe the climate, geographical position, political structure and economics of given states and to estimate their population by means of data on births and deaths. In this context, the gulf between political arithmetic and statecraft was not therefore as wide as it is usually supposed to have been, and **Leibniz**' manuscripts written in the 1680's indeed testify that he was both a political arithmetician and an early advocate of tabular description (with or without the use of numbers) of a given state. By the 19th century statecraft broke down because of the heterogeneity of its subject, whereas statistics, as we now know it, properly issued from political arithmetic.

The father of population statistics was **Süssmilch**. He collected vast data on the movement of population and attempted to prove Divine providence as manifested in every field of vital statistics. He treated his materials faultily; thus, he combined towns and villages without taking weighted means, and he had not tried to allow for the difference in the age structures of the populations involved. Nevertheless, his life tables remained in use well into the 19th century.

Euler actively participated in preparing the second edition (1765) of Süssmilch's main work, the *Göttliche Ordnung*, and one of its chapters was partly reprinted in his *Opera omnia*. Later on **Malthus**, without any references, adopted their indirect conclusion that population increased in a geometric progression. Euler left several contributions on population statistics, now collected in his *Opera omnia*. With no censuses (as we understand them now) at his disposal, he was unable to recognize the importance of some demographic factors, but he introduced such concepts as increase in population, and the period of its doubling. He worked out the mathematical theory of mortality and formulated rules for establishing life insurance in all its forms, cf. §7 where I mention several previous scholars whom Euler had not cited.

During 1766 – 1771 **Daniel Bernoulli** contributed three memoirs to population statistics. In the first of these he examined the benefits of inoculation, – of communicating a mild form of smallpox from one person to another one, – which had been the only preventive measure against that deadly disease. The Jennerian vaccination became known at the turn of the 18^{th} century, whereas inoculation had been practised in Europe from the 1720's. This procedure was not safe: a small fraction of those inoculated were dying, and, in addition, all of them spread the disease among the population.

Bernoulli's memoir was the first serious attempt to study it, but even he failed to allow properly for the second danger. He formulated (necessarily crude) statistical hypotheses on smallpox epidemics and calculated the increase in the mean duration of life caused by inoculation. Concluding that this treatment prolonged life by two years, he came out in its favour. In 1761, even before Bernoulli's memoir had appeared, D'Alembert voiced reasonable objections. Not everyone, he argued, will agree to expose himself to a low risk of immediate death in exchange for a prospect of living two remote years longer. And there also existed the moral problem of inoculating children. In essence, he supported inoculation, but regarded its analysis impossible.

In his second memoir Bernoulli studied the duration of marriages, a problem directly connected with the insurance of *joint lives*. He based his reasoning on an appropriate problem of extracting strips of two different colours from an urn which he solved in the same year (in 1768).

Bernoulli devoted his third memoir of 1770 – 1771 to studying the sex ratio at birth. Supposing that male and female births were equally probable, he calculated the probability that out of $2N$ newly-born m were boys:

$$P = [1\cdot3\cdot5\cdot\ldots\cdot(2N-1)] \div [2\cdot4\cdot6\cdot\ldots\cdot2N] = q(N).$$

He calculated this fraction not by the **Wallis** formula or the local **De Moivre – Laplace** theorem, but by means of differential equations. After deriving
$q(N-1)$ and $q(N+1)$ and the two appropriate values of Δq, he obtained

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

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

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

Application of differential equations was Bernoulli's usual method in probability.

Bernoulli also determined the probability of the birth of approximately m boys:

$$P(m = N \pm \mu) = q \exp(-\mu^2/N) \text{ with } \mu = 0(\sqrt{N}). \qquad (4)$$

He then generalized his account to differing probabilities of the births of both sexes, and, issuing from some statistical data, compared two possible values of the sex ratio but had not made a definite choice.

A special feature of this memoir is that Bernoulli determined such a value of μ that the total probability (4) from $\mu = 0$ to this value ($\mu = 47$) was 1/2. He calculated this total by summing rather than by integration and thus failed to obtain directly the De Moivre – Laplace theorem (2).

In 1772 **Lambert** followed Daniel Bernoulli in studying population statistics. He offered an empirical law of mortality, examined the number of children in families and somewhat extended Bernoulli's memoir on smallpox by considering children's mortality from this disease. Before treating the second-mentioned subject, Lambert increased the number of children by 1/2 thus apparently allowing for stillbirths and infant mortality. This rate of increase was arbitrary, but at least he attempted to get rid of a gross systematic mistake. Along with Bernoulli and Euler he created the methodology of mathematical demography.

7. Civil Life; Moral and Economic Issues

Jakob Bernoulli thought of applying probability to civil life and moral and economic affairs, but he did not have time to accomplish much in this direction. One aspect of civil life, i. e., games of chance, had indeed promoted the origin of the theory of probability (§1) and offered meaningful problems whose solutions became applicable in natural sciences and led to the creation of new mathematical tools used also in probability (§10.1). I shall now discuss other pertinent points.

In 1709, **Niklaus Bernoulli** published a dissertation on applying the *art of conjecturing* to jurisprudence, and, it ought to be added, he plagiarized Jakob Bernoulli by borrowing from his as yet unpublished classical book of 1713 and even from his *Meditationes* (Diary) never meant for publication. Niklaus repeatedly mentioned his late uncle, which does not exonerate him.

Niklaus recommended the use of mean longevity and mean gain (or loss) in calculations concerning annuities, marine insurance, lotteries And in deciding whether an absent person ought to be declared dead both he and Jakob were prepared to weigh the appropriate probabilities against each other. Mentality really changed since the time when **Kepler** correctly, but in a restricted way, had simply refused to say whether the absent man was alive or dead.

In connection with a problem in mortality (and, therefore, life insurance) Niklaus effectively introduced the continuous uniform distribution which was the first continuous law to appear in probability. Important theoretical work inspired by life insurance was going on from 1724 (**De Moivre**) onward (**Thomas Simpson**). Actually, insurance societies date back to the beginning of the 18th century, but more or less honest business, based on statistics of mortality, hardly superseded downright cheating before the second half of the 19th century. And, although governments sold annuities even in the 17th century, their price had then been largely independent from statistical data.

Stochastic studies of judicial decisions, of the voting procedures adopted by assemblies and at general elections, had begun in the late 18th century, but many later scientists denied any possibility of numerically examining these subjects. Thus, probability, misapplied to jurisprudence, had become "the real opprobrium of mathematics" (Mill 1886, p. 353); or, in law courts people act like the "moutons de Panurge" (**Poincaré** 1912, p. 20).

So, is it possible to determine the optimal number of jurors, or the optimal majority of their votes (when a wrong decision becomes hardly possible)? To determine the probability of an extraordinary fact observed by witnesses? **Condorcet** studied these and similar problems although hardly successfully. First, it was difficult to follow his exposition, and, second, he had not made clear that his attempt was only tentative, that he only meant to show what could be expected in the ideal case of independent decisions being made. But at least he emphasized that *les hommes* should be educated and unprejudiced.

Laplace followed suit declaring that the representation of the nation should be the *élite* of men of exact and educated minds. Later he (1816, p. 523) remarked, although only once and in passing, that his studies were based on the assumption that the jurors acted independently one from another.

One of Condorcet's simple formulas (which can be traced to **Jakob Bernoulli's** study of stochastic arguments in his *Ars Conjectandi* and which Laplace also applied in 1812) pertained to extraordinary events (above). If the probabilities of the event in itself and of the trustworthiness of the report are p_1 and p_2, then the event acquires probability

$$P = \frac{p_1 p_2}{[p_1 p_2 + (1-p_1)(1-p_2)]}.$$

This formula is however hardly applicable. Indeed, for $p_1 = 1/10,000$ and $p_2 = 0.99$, $P \approx 0.01$ so that the event will not be acknowledged by a law court, and a second trustworthy witness will have to be found.

Moral applications of probability at least emphasized the importance of criminal statistics and assisted in evaluating possible changes in the established order of legal proceedings. As **Gauss** correctly remarked in 1841, the appropriate studies were unable to help in individual cases, but could have offered a clue to the lawgiver for determining the number of witnesses and jurors.

Applications of probability to economics began in 1738 with **Daniel Bernoulli**. In attempting to solve the Petersburg paradox (§10.2), he assumed that the advantage (y) of a gambler was connected with his gain (x) by a differential equation (likely the first such equation in probability theory)

$$y = f(x) = c \ln(x/a)$$

where a was the initial fortune of the gambler. Bernoulli then suggested that the *moral expectation* of gain, be chosen instead of its usual expectation,

$$\sum p_i f(x_i)/\sum p_i \text{ instead of } \sum p_i x_i / \sum p_i;$$

the p_i's were the probabilities of the respective possible gains.

The distinction made between gain and advantage enabled Bernoulli to replace the infinite expectation (10) appearing in a paradoxical situation by a new expression which was finite and thus to get rid of the paradox, see §10.2. Neither did he fail to notice that, according to his innovation, a fair game of chance became detrimental to both gamblers.

Bernoulli next applied moral expectation to studying the shipping of freight and stated that (in accordance with common sense) it was beneficial to carry the goods on several ships. He did not prove this statement (which was done by **Laplace**).

Moral expectation became fashionable and Laplace (1812, p. 189) therefore qualified the classical expectation by the adjective *mathematical*. Nowadays, it is still used in the French and Russian literature. In 1888 **Bertrand** declared that the theory of moral expectation had become classical but remained useless. However, already then economists began developing the theory of marginal utility by issuing from Bernoulli's fruitful idea.

The term *moral expectation* is due to Gabriel Cramer who had expressed thoughts similar to those of Daniel Bernoulli and the latter published a passage from his pertinent letter of 1732 to **Niklaus Bernoulli**.

8. The Theory of Errors

8.1. The Main problem. Suppose that m unknown magnitudes $x, y, z, \ldots$ are connected by a redundant system of n physically independent equations ($m < n$)

$$a_i x + b_i y + c_i z + \ldots + s_i = 0 \qquad (5)$$

whose coefficients are given by the appropriate theory and the free terms are measured. The approximate values of $x, y, z, \ldots$ were usually known, hence the linearity of (5). The equations are linearly independent (a later notion), so that the system is inconsistent (which was perfectly well understood). Nevertheless, a solution had to be chosen, and it was done in such a way that the residual free terms (call them v_i) were small enough.

The case of direct measurements ($m = 1$) should be isolated. Given, observations $s_1, s_2, \ldots, s_n$ of an unknown constant x (here, $a_i = 1$); determine its true value. The choice of the arithmetic mean seems obvious and there is evidence that such was the general rule at least since the early 17th century. True, ancient astronomers treated their observations in an arbitrary manner and in this sense even astronomy then had not yet been a quantitative science. However, since errors of observations were large, the absence of established rules can be justified. Thus, for *bad* distributions of the errors the arithmetic mean is not stochastically better (or even worse) than a single observation.

In 1722, **Cotes'** posthumous contribution appeared. There, he stated that the arithmetic mean ought to be chosen, but he had not justified his advice, nor did he formulate it clearly enough. Then, in 1826, **Fourier** had defined the *veritable object of study* as the limit of the arithmetic mean as the number of observations increased indefinitely, and many later authors including **Mises**, independently one from another and never mentioning Fourier, introduced the same definition for the *true value*.

The classical problem that led to systems (5) was the determination of the figure of the Earth. Since **Newton** had theoretically discovered that our planet was an ellipsoid of rotation with its equatorial radius (a) larger than its polar radius (b), numerous attempts were made to prove (or disprove) this theory. In principle, two meridian arc measurements were sufficient for an experimental check (for deriving a and b), but many more had to be made because of the unavoidable errors of geodetic and astronomical observations (and local deviations from the general figure of the Earth).

At present, the adopted values are roughly a = 6,378.1 *km* and b = 6,356.8 *km*. That $2\pi \cdot 6{,}356.8 = 39{,}941$ which is close to 40,000 is no coincidence: in 1791, the meter was defined as being $1/10^7$ of a quarter of the Paris meridian. This *natural* standard of length lasted until 1872 when the meter of the Archives (called for the place it was kept in), a platinum bar, was adopted instead. From 1960, the meter is being defined in terms of the length of a light wave. The introduction of the metric system as well as purely astronomical problems had necessitated new observations so that systems (5) had to be solved time and time again, whereas physics and chemistry began presenting their own demands by the mid-19th century.

8.2. Its Solution. Since the early 19th century the usual condition for solving (5) was that of least squares

$$v_1^2 + v_2^2 + \ldots + v_n^2 = \min.$$

Until then, several other methods were employed. Thus, for $m = 2$ the system was broken up into all possible subsystems of two equations each, and the mean value of each unknown over all the subsystems was then calculated. As discovered in the 19th century, the least-squares solution of (5) was actually some weighted mean of these partial solutions.

The second important method of treating systems (5) devised by **Boscovich** consisted in applying conditions

$$v_1 + v_2 + \ldots + v_n = 0, \quad |v_1| + |v_2| + \ldots + |v_n| = \min \qquad (6a, 6b)$$

(Maire & Boscovich 1770, p. 501). Now, (6a) can be disposed of by summing up all the equations in (5) and eliminating one unknown. And, as **Gauss** noted in 1809, (6b) led exactly to m zero residuals v_i, which follows from an important theorem in the then not yet known linear programming. In other words, after allowing for restriction (6a), only $(m - 1)$ equations out of n need to be solved, but the problem of properly choosing these still remained. Boscovich himself applied his method for adjusting meridian arc measurements and he chose the proper equations by a geometric trick. Then, **Laplace** repeatedly applied the Boscovich method for the same purpose, for example, in vol. 2 of his *Mécanique céleste* (1799).

A special condition for solving systems (5) was $|v_{max}| = \min$, the minimax principle. **Kepler** might have well made his celebrated statement about being unable to fit the **Tychonian** observations to the **Ptolemaic** theory after having attempted to apply this principle (in a general setting rather than to linear algebraic equations). In 1749, **Euler** achieved some success in employing its rudiments. The principle is not supported by stochastic considerations, but it has its place in decision theory and Laplace (1789, p. 506) clearly stated that it was suited for checking hypotheses (cf. Kepler's possible attitude above) although not for adjusting observations. Indeed, if even this principle does not achieve a concordance between theory and observation, then either the observations are bad, or the theory wrong.

8.3. Simpson. I return now to the adjustment of direct observations. In 1756 Simpson proved that at least sometimes the arithmetic mean was more advantageous than a single observation. He considered the uniform, and the triangular distributions for the discrete case. After calculating the error of the mean he recommended the use of this estimator of the true value of the constant sought. Simpson thus extended stochastic considerations to a new domain and effectively introduced random observational errors, i. e. errors taking a set of values with corresponding probabilities. His mathematical tool was the generating function introduced by **De Moivre** in 1730 for calculating the chances of throwing a certain number of points with a given number of dice. De Moivre first published the solution of that problem without proof in 1712, somewhat earlier than **Montmort** (§3) who employed another method.

For that matter, no doubt following De Moivre, Simpson himself had earlier (1740) described the same calculations, and he now noted the similarity of both problems. Consider for example his triangular distribution with errors

$$- v, \ldots, -2, -1, 0, 1, 2, \ldots, v \qquad (7)$$

having probabilities proportional to

$$1, \ldots, (v-2), (v-1), v, (v-1), (v-2), \ldots, 1.$$

Simpson's (still unnamed) generating function was here

$$f(r) = r^{-v} + 2r^{-v+1} + \ldots + (v+1)\, r^0 + \ldots + 2r^{v-1} + r^v$$

and the chance that the sum of t errors equalled m was the coefficient of r^m in $f^t(r)$.

In 1757 Simpson went on to the continuous triangular distribution by introducing a change of scale: the intervals between integers (7) now tended to zero so that it became possible to regard the segment $[-v; v]$ as consisting of an infinitely large number of such intervals, and the distribution, as though given on a continuous set.

In 1776 **Lagrange** extended Simpson's memoir to other (purely academic) distributions. He introduced integral transformations, managed to apply generating functions to continuous distributions and achieved other general findings.

8.4. Lambert. Let $\varphi(x; \hat{x})$ with unknown parameter $\hat{x}$ be the density law of independent observational errors $x_1, x_2, \ldots, x_n$. Then the value of

$$\varphi(x_1; \hat{x}) \cdot \varphi(x_2; \hat{x}) \ldots \cdot \varphi(x_n; \hat{x}) \qquad (8)$$

will correspond to the probability of obtaining such observations. Hence the maximal value of (8) will provide the *best* value of $\hat{x}$. Now suppose, as it was always done in classical error theory, that the density is $\varphi(x - \hat{x})$, a curve with a single peak (mode) at point $x = \hat{x}$. The determination of the true value of the constant sought may then be replaced by calculation of the most probable value of $\hat{x}$. The derivation of the unknown parameter(s) of density laws became an important problem of statistics, and the principle of maximum likelihood (of maximizing the product (8)) provides its possible solution.

It was Lambert who first formulated this principle for unimodal densities in 1760. Actually, he studied the most important aspects of treating observations. He returned to this subject in 1765, this time attempting to determine the density of pointing a geodetic instrument by starting from the principle of insufficient reason (the term was introduces later) and to estimate numerically the precision of observations.

At the end of the 19th century the just mentioned principle was applied to substantiate the existence of *equally possible cases* appearing in the formulation of the notion of probability and soon afterwards Poincaré managed to soften essentially this delicate issue. In actual fact, the very notion of expectation, if not understood as an abstract concept (which it really is), can hardly be justified in any other way excepting *insufficient reason*.

Lambert (1765, §321) also defined the *Theorie der Fehler* including into its province both the stochastic and the deterministic studies of errors. **Bessel** had picked up this term, *Theory of errors*, and, although neither **Laplace**, nor **Gauss** ever applied it, it came in vogue in the mid-19th century.

A classical example of the deterministic branch of the error theory is Cotes' solution (1722) of 28 problems connecting the differentials of the various elements of plane and spherical triangles with each other. He thus enabled to calculate the effect of observational errors on indirectly determined sides of the triangles.

8.5. Daniel Bernoulli. In 1778, Daniel Bernoulli denied the arithmetic mean and, without mentioning **Lambert**, advocated the principle of maximum likelihood. Taking a curve of the second degree as the density law of the observational errors, and examining the case of only three observations, he obtained an algebraic equation of the fifth degree in $\hat{x}$, the estimator of the constant sought.

In a companion commentary, **Euler** reasonably denounced the principle of maximum likelihood since in the presence of an outlying observation the product (8) becomes small, and, in addition, contrary to common sense, the decision of whether to leave or reject it becomes important. Then, nevertheless following Bernoulli but misinterpreting him, he derived a cubic equation in $\hat{x}$ and noted that it corresponded to the maximal value of the sum of the squares of the weights of the observations. If the small terms of this sum are rejected, his condition becomes

$$(\hat{x} - x_1)^2 + (\hat{x} - x_2)^2 + + \ldots + (\hat{x} - x_n)^2 = \min \qquad (9)$$

which leads to the arithmetic mean, still alive and kicking!

Heuristically, (9) resembles the condition of least squares (and, indeed, in case of $m = 1$ least squares lead to this mean). Furthermore, **Gauss**, in 1823, in his definitive formulation of this celebrated method, derived it from the principle of maximum weight which might, again heuristically, be compared with Euler's condition (9).

Finally, in 1780 Bernoulli considered pendulum observations. Drawing on his previous memoir, he applied formula (4), i. e., the normal law, for calculating the error of time-keeping accumulated during 24 hours. He then isolated random (*momentanearum*) errors, whose influence was proportional to the square root of the appropriate time interval, from systematic (*chronicarum*), almost constant mistakes. These two categories are still with us, but his definitions are not.

8.6. Laplace. Laplace's main achievements in error theory belong to the 19th century. Before that, he published two memoirs (1774; 1781) bearing on this subject and interesting from the modern point of view but 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 sum of *errors to be feared of* multiplied by their probabilities (i. e., the absolute expectation of error) should be minimal. In the 19th century, he applied the same principle for justifying the method of least squares, which was only possible for the case of normal distribution (existing on the strength of his non-rigorous proof of the central limit theorem when the number of observations was large).

Also in 1781, Laplace proposed, as a density curve,

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

His deliberations might be described by the **Dirac** delta-function. However, one of his conclusions was based on considering an integral of

$$\varphi\ [\alpha(x - x_1)] \cdot \varphi\ [\alpha(x - x_2)] \ \ldots \ \cdot \varphi\ [\alpha(x - x_n)]$$

(where the x_i's were the observations made) which has no meaning in the language of generalized functions.

From its very beginning, the theory of errors belonged to probability theory (**Simpson**), but its principles of adjusting observations (of maximal likelihood; of least absolute expectation; of least squares) had been subsequently taken over by statistics.

9. Laplace's Determinism

According to Laplace's celebrated utterance (1814/1995, p. 2), for an omniscient intelligence "nothing would be uncertain, and the future, like the past, would be open to its eyes". He did not say that initial conditions could not be known precisely and of course he did not know anything about instability of motion (**Poincaré**) or about modern ideas on the part of randomness (or chaos) in mechanics.

Already in the beginning of his career he (1776, p. 145) denied randomness ("Le hasard n'a … aucune réalité en lui-même") but remarked that "le plus grand nombre des phénomènes" could only be studied stochastically and attributed the emergence of the "science des hasards ou des probabilités" to the feebleness of the mind. The real cause for the origin of probability was rather the existence of stochastic laws determining the behaviour of sums (or other functions) of random variables; or, the dialectical interrelation between the randomness of a single event and the necessity provided by mass random phenomena.

A case in point is the *statistical determinism*. Thus, in 1819 Laplace noticed that the receipts from the Lottery of France had been stable. Elsewhere, he (1795/1812, p. 162) remarked that the same was true with regard to the yearly number of dead letters. The generally known statement about the *figures of moral statistics* (of marriages, suicides, crimes) is due to **Quetelet**. Owing to his careless formulation it is hardly known that he actually meant stability under constant social conditions.

Two additional points are worth stating. First, nobody ever claimed that Laplace's philosophy had hindered his studies in astronomy or population statistics (based on stochastic examination of observations, see §11). Moreover, he (1796, p. 504) effectively recognized randomness when discussing the eccentricities of planetary orbits and other small deviations from "une parfaite régularité".

Second, belief in determinism and actual recognition of randomness did not begin with Laplace. **Kepler** denounced chance as an abuse of God, but he had to explain the eccentricities by random causes. Laplace (and **Kant**) likely borrowed this idea from him, or from **Newton** (1718/1782, Query 31, p. 262) who actually recognized randomness as Kepler did: The "wonderful uniformity in the planetary orbits" was accompanied by "inconsiderable irregularities … which may have risen from the mutual actions of comets and planets upon one another". Finally, Laplace might have found his statement about the omniscient intelligence in earlier literature (**Maupertuis** 1756, p. 300; **Boscovich** 1758, §§384 – 385).

10. Some Remarkable Problems

10.1. The Gambler's Ruin. A series of games of chance is played by *A* and *B* until one of them is ruined. How long can the series be? What is the probability that *A* (or *B*) will be ruined not more than in *n* games? These are some questions here. In its simplest form the problem of ruin is due to Huygens.

Suppose that *A* has *a* counters, the probability of his winning a game is *p*, and the respective magnitudes for *B* are *b* and *q* ($p + q = 1$). Call P_a the probability of *A*'s loosing all his counters before winning all those belonging to *B*, let P_{an} be the probability of his ruin in not more than *n* games and denote the respective magnitudes for *B* by P_b and P_{bn}. The entire game can be imagined as a movement of a point *C* along a segment of length $(a + b)$, up to *b* units to the left and up to *a* units to the right. After each game *C jumps* to the left with probability *p* or to the right with probability *q*, and the play ends when *C* arrives at either end of the segment. Between these barriers *C* will *walk* randomly. And a random walk (which can also be imagined in a three-dimensional space) is a crude model of diffusion and Brownian motion.

Jakob Bernoulli several times treated this problem either incompletely (like **Huygens** did) or leaving the proof of his formula to his readers. It was **De Moivre**, who already in 1712 proved the same formula by an ingenious reasoning. He established that

$$\frac{P_A}{P_B} = \frac{a^q(a^p - b^p)}{b^p(a^q - b^q)},\ a \neq b.$$

He also offered rules for calculating either the probability ($P_{an} + P_{bn}$) that the play will end within *n* games or the probabilities P_{an} and P_{bn} separately, and, in addition, he considered the case of $a = \infty$. De Moivre extended his research: in 1718 he provided answers to other problems although without justifying the results obtained. The demonstrations are now reconstructed (**Hald** 1990, §20.5).

De Moivre's later findings were especially important because of the new method which he devised and applied here, the method of recurring sequences. **Laplace** discussed the problem of the gambler's ruin in several memoirs. He (1776) solved it by means of partial difference equations even for the case of three gamblers. **Lagrange** devoted the last section of his memoir of 1777 on these equations to their application in probability. There, he solved several problems which, in particular, were concerned with the gambler's ruin.

10.2. The Petersburg Paradox. In a letter to **Montmort** of 1713 **Niklaus Bernoulli** described his invented game (Montmort 1713, p. 402). *A* gives *B* an *écu* if he throws a six at the first attempt with a common die; he also promises 2, 4, 8, … *écus* if the six first appears at the second, the third, the fourth, … throw. Required is the expectation of *B*'s gain (call it Eξ). The conditions, but not the essence of the problem soon changed with a coin replacing the die. In this new setting

$$E\xi = 1 \cdot 1/2 + 2 \cdot 1/4 + 4 \cdot 1/8 + \ldots = \infty \qquad (10)$$

whereas no reasonable man would have given much in exchange for a promised Eξ. This remarkable paradox has been discussed to this very day; here are the pertinent points.

a) It introduced a random variable with an infinite expectation.

b) It inspired scholars to emphasize that a low probability of gain (lower than some positive α) should be disregarded, i. e., that only a few terms of the infinite series be taken into account). But how large ought to be the maximal value of α? And a similar question for probabilities of loss higher than 1 – α? There is no general answer, everything depends on circumstances lying beyond the province of mathematics. The value α = 1/10,000 recommended by **Buffon** in 1777, – the probability that a healthy person aged 56 years dies within the next 24 hours, – had intuitive appeal, but it was too low and never really adopted as a universal estimate. Cf. the concept of moral certainty introduced by **Descartes** and **Huygens** (§1) and taken up by **Jakob Bernoulli**.

c) It prompted **Daniel Bernoulli** to introduce the moral expectation (§6) which enabled him to solve the paradox by getting rid of the infinity in (10). His contribution was published in a periodical of the Petersburg Academy of Sciences, hence the name of the paradox.

d) It led to an early and possibly the first large-scale statistical experiment: Buffon, in the same contribution of 1777, described his series of 2,048 Petersburg games. The average payoff per game occurred to be only 4.9 and the maximal number of tosses in a game was nine, and then only in six cases.

e) **Condorcet**, and later **Lacroix** discovered a more proper approach to the paradox: the possibly infinite game, as they maintained, presented one single experiment so that only a mean characteristic of many such games can provide a reasonable clue. **Freudenthal** (1951) studied a series of Petersburg games with the gamblers taking turns by lot in each of them.

f) A digression. **Buffon's** experiment illustrated runs (sequences) of random events with one and the same probability of *success*. **Montmort** testified that gamblers were apt to make wrong conclusions depending on the appearance (or otherwise) of a run in a series of independent games of chance. At present, runs are made use of to distinguish between chance and regularity. Suppose that a certain dimension of each machine part in a batch is a bit larger than that of a standard part; how probable is it that something went wrong?

De Moivre solved important problems connected with probabilities of number sequences in sampling. In 1767 **Euler** met with similar problems when studying lotteries and solved them by the combinatorial method. In 1793 **John Dalton** applied elementary considerations when studying the influence of auroras on the weather and in the 19th century **Quetelet** and **Köppen** described the tendency of the weather to persist by elements of the theory of runs.

10.3. The Ehrenfests' Model

Each of two urns contains an equal number *n* of balls, white and black, respectively. Determine the (expected) number of white balls in the first urn after *r* cyclic interchanges of one ball. **Daniel Bernoulli** solved this problem by the combinatorial method and, in addition, by applying differential equations. He also generalized his problem to three urns with balls of three colours and noted the existence of a limiting case, of an equal (mean) number of balls of each colour in each urn. At present, this can be proved by referring to a theorem concerning homogeneous **Markov** chains.

In 1777 **Lagrange** solved a similar problem for any finite number of urns and balls of two colours. He employed partial difference equations as did **Laplace** in 1811 when solving a similar problem. Laplace (1814/1995, p. 42) also poetically interpreted the solution of such problems:

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

Nevertheless, it is difficult to discover his *constant forces*, and a later author (**Bertrand** 1888, p. xx) put it better: "Le hazard, à tout jeu, corrige ses caprices". True, he only connected his remark with the action of the law of large numbers; in his case, the less was the relative number of white balls (say) in an urn, the less probable became their future extractions.

The future history of such urn problems as described above includes the celebrated **Ehrenfests'** model of 1907 which is usually considered as the beginning of the history of stochastic processes.

11. Mathematical Statistics

Roughly speaking, the difference between probability and statistics consists in that the former is deductive whereas the latter (excepting its own theoretical part) is inductive and has to do with making conclusions from quantitative data. Mathematical statistics emerged in the 20th century and the term itself had hardly appeared before C. G. A. Knies introduced it in 1850.

However, problems connected with inductive inference are very old: even ancient scholars and lawgivers, drawing on numerical data, strove to distinguish between causality and randomness, e. g., between deaths from an emerging epidemics and the "normal" mortality (the Talmud, see its treatise Taamit). Beginning with **Petty** and **Graunt** (§1), crude statistical probabilities were being applied for estimating populations, and **Arbuthnot's** problem concerning the births of boys and girls (§6) was also inductive. The main goal of **De Moivre's** *Doctrine of Chances*, as he himself declared, was the choice between Design and randomness.

By studying the statistical determination of the probability of a random event, **Bayes** (§5) opened up a chapter of mathematical statistics. For **Laplace**, probability became the decisive tool for discovering the laws of nature (he never mentioned Divine Design). Thus, after establishing that the existence of a certain astronomical magnitude, as indicated by observations, was highly probable, he (1812, p. 361) felt himself obliged to investigate its cause and indeed proved its reality. Several chapters of his classic *Théorie analytique* … could now be called statistical. Since he based it on his earlier memoirs, it is natural that there we find him (1774, p. 56) mentioning *un nouveau genre de problème les hasards* and even *une nouvelle branche de la théorie des probabilités* (1781, p. 383). The expression *nouvelle branche* was due to **Lagrange**, see his letter to Laplace of 13.1.1775 in t. 14 of his *Oeuvres*, who thus described the latter's estimation of a certain probability.

A remark made by Laplace in 1812 can be connected with the present-day statistical simulation. He enlarged on **Buffon** whose study was first announced in an anonymous abstract in 1735 and published in 1777. A needle of length $2r$ falls on a set of parallel lines. The probability that it intersects a line, as he had found out, was $p = 4r/\pi a$ where a was the distance between adjacent lines, and Laplace noted that from a large number of such falls the value of π can thus be estimated. Note that Buffon had made use of geometric probability.

A curious and wrong statement made by the astronomer **William Herschel** (1817, p. 579) shows that statistics was sometimes thought to be more powerful than it was (or is). He argued that the size of any star, "promiscuously chosen" out of the 14,000 stars of the first seven magnitudes, was "not likely to differ much from a certain mean size of them all". Unlike observational errors (say), stars (of differing physical nature!) could not have belonged to one and the same statistical population. Only in the former case we may estimate (by applying the later **Bienaymé – Chebyshev** inequality and issuing from data!) the deviations of the possible values of a random variable from their mean.

Sampling theory is a chapter of statistics, but the practice of sampling in England goes back at least to the 13th century when it began to be applied for assaying the new coinage (Stigler 1977). For many years, W. Herschel engaged in counting the stars in heaven. In his report of 1784 he noted that in one section of the Milky Way their multitude prevented him so that he only counted the stars in six "promiscuously chosen" fields, i. e., applied the principle of sampling. He also counted the stars in a "most vacant" field, obviously for checking the lower bound of his calculated estimate of the total number of stars in the section.

In the absence of censuses, **Laplace** (1786) employed sampling for calculating the population of France (M). He knew the population of a small (sample) part of the country (m), the yearly number of births both there and over entire France (n and N), and, assuming that the ratio of births to population was constant, he concluded that $M = Nm/n$. Laplace then applied his earlier formulas (end of §5) for estimating the possible error of this figure. In 1928 **Karl Pearson** reasonably remarked that Laplace's urn model (§5) of which he made use here was not adequate and that his relevant approximate calculations were imperfect. Still, Laplace was the first to study the error of sampling whereas his method of calculation (of the incomplete B function) was not improved for more than a century, cf. §5 on the appropriate efforts made by Bayes.

12. The Opposition

The theory of probability did not develop unopposed. **Leibniz**, in his correspondence with **Jakob Bernoulli** (Kohli 1975), denied that statistical probability should be regarded as an equal of its theoretical counterpart. The former, he argued, depended on an infinity of circumstances and could not be determined by a finite number of observations. Jakob, however, remarked that the opposite might be true for the ratio of two infinities (apparently: for the rate of success in Bernoulli trials). Later on Leibniz changed his opinion. In any case, in a letter of 1714 he even claimed, without any justification, that the late Bernoulli "a cultivé" probability "sur mes exhortations".

De Moivre (1756, p. 254) stated that

There are Writers, of a Class indeed very different from that of James Bernoulli, who insinuate as if the Doctrine of Probabilities could have no place in any serious Enquiry … [that its study was] *trivial and easy* [and] *rather disqualifies a man from reasoning on every other subject.*

Simpson (1756, p. 82) defined the aim of his memoir on the arithmetic mean (§8.3) as refuting

Some persons, of considerable note, who … even publickly maintained that one single observation taken with due care, was as much to be relied on as the mean of a great number of them …

Indeed, natural scientists might have persisted in **Robert Boyle's** belief (1772, p. 376) that "experiments ought to be estimated by their value, not their number". However, the two approaches should be complementary rather than contradictory.

The main culprit was however **D'Alembert** (who nevertheless did not check the advance of probability). In 1754 and again in 1765 he claimed that the probability of throwing two heads consecutively was 1/3 rather than 1/4. He also believed that after several heads in succession tails will become more likely and he aggravated this nonsense by an appeal to determine probabilities statistically (which would have proved him wrong). Then, in 1768, he was unable to understand why the mean and the probable duration of life did not coincide.

Euler (Juskevic et al 1959, p. 221), in a letter of 27 May/7 June 1763, mentioned Dalembert's "unbearable arrogance" and argued that he had tried "most shamelessly to defend all his mistakes" [possibly not only in probability]. Witness also Dalembert's invasion (1759, p. 167) of an alien field of knowledge: "The physician most worthy of being consulted is the one who least believes in medicine".

True, Dalembert also put forward some reasonable ideas. He remarked, after Buffon, that low probabilities of gain ought to be discarded and he noted that the benefits of inoculation (§6) should be reassessed. In general, some of his criticisms were ahead of the time since they implied that the theory of probability ought to be built up more rigorously.

13. On the Threshold of the Next Century

The new century began with the appearance, in 1812, of **Laplace's** *Théorie*(which I had to mention above). There, he brought together all his pertinent memoirs (including those of 1809 – 1811), but failed to merge them into a coherent whole. True, he applied the **De Moivre – Laplace** limit theorem wherever possible, but he did not introduce, even on a heuristic level, the notion of a random variable, did not therefore study densities or characteristic functions per se; his theory of probability still belonged to applied mathematics and did not admit of development.

But what was achieved up to 1801? The first limit theorems were proved; generating functions and difference equations were introduced and applied; and integrals were approximated by new and complicated methods. The study of games of chance originated important topics with future applications in natural sciences and economics. Probability became widely applied to population statistics and treatment of observations (and jurisprudence), but natural sciences did not yet yield to this new discipline. Problems really belonging to mathematical statistics were being solved again and again and the time became ripe for **Gauss** to develop the method of least squares.

Bibliography

1. Sources

Arbuthnot, J. (1712), An argument for Divine Providence taken from the constant regularity observed in the births of both sexes. In Kendall & Plackett (1977, pp. 30 – 34).
Arnauld, A., Nicole, P. (1662), *L'art de penser*. Paris, 1992.
Bayes, T. (1764), An essay towards solving a problem in the doctrine of chances, with commentary by R. Price. Reprinted: *Biometrika*, vol. 45, 1958, pp. 293 – 315 and in E. S. Pearson & Kendall (1970, pp. 131 – 153). German transl.: Leipzig, 1908.
Bernoulli, D. (1738, in Latin), Exposition of a new theory on the measurement of risk. *Econometrica*, vol. 22, 1954, pp. 23 – 36.
--- (1766), Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir. *Werke*, Bd. 2. Basel, 1982, pp. 235 – 267.
--- (1768a), De usu algorithmi infinitesimalis in arte coniectandi specimen. Ibidem, pp. 276 – 287.
--- (1768b), De duratione media matrimoniorum. Ibidem, pp. 290 – 303.
--- (1770), Disquisitiones analyticae de nouo problemate coniecturale. Ibidem, pp. 306 – 324.
--- (1770 – 1771), Mensura sortis ad fortuitam successionem rerum naturaliter contingentium applicata. Ibidem, pp. 326 – 360.
--- (1778, in Latin), The most probable choice between several discrepant observations and the formation therefrom of the most likely induction. *Biometrika*, vol. 48, 1961, pp. 1 – 18. Reprinted in E. S. Pearson & Kendall (1970, pp. 155 – 172).
--- (1780), Specimen philosophicum de compensationibus horologicis. *Werke*, Bd. 2, pp. 376 – 390.
Bernoulli, J. (1713), *Ars Conjectandi*. *Werke*, Bd. 3. Basel, 1975, pp. 107 – 259. Translated into German, and the most important pt 4 was also translated into Russian and French.
Bernoulli, N. (1709), De usu artis conjectandi in iure. In Bernoulli, J. (1975, pp. 287 – 326).
Bertrand, J. (1888), *Calcul des probabilités*. Second ed., 1907. Reprinted: New York, 1970, 1972.
Boscovich, R. G. (1758, in Latin), *Theory of Natural Philosophy.* Cambridge (Mass.) – London, 1966. Translated from the edition of 1763.
Boyle, R. (1772), A Physico-Chymical Essay. *Works*, vol. 1. Sterling, Virginia, 1999, pp. 359 – 376.
Buffon, G. L. L. (1777), Essai d'arithmétique morale. *Oeuvr. Phil.* Paris, 1954, pp. 456 – 488.
Condorcet, M. A. N. Caritat de (1986), *Sur les éléctions et autres textes*. Paris. Contains Discourse préliminaire de l'essai sur l'application de l'analyse a la probabilité des voix (1785), pp. 7 – 177 and Elements du calcul des probabilités (1805), pp. 483 – 623. The entire Essai (not just the Discourse) is reprinted separately: New York, 1972.
--- (1994), *Arithmétique politique*. Paris. Contains reprints of Sur le calcul des probabilités (1784 – 1787), of his articles from the *Enc. Méthodique* and previously unpublished or partly published MSS.
Cotes, R. (1722), *Aestimatio errorum in mixta mathesi per variationes partium trianguli plani et sphaerici*. In *Opera misc.* London, 1768, pp. 10 – 58.
Cournot, A. A. (1843), *Exposition de la théorie des chances et des probabilités.* Paris, 1984.
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 elemens de philosophie. The passage quoted in text appeared in 1821 (OC, t. 1, pt 1. Paris, pp. 116 – 348).
--- (1761a), Réflexions sur le calcul des probabilités. *Opuscules math.*, t. 2. Paris, pp. 1 – 25.
--- (1761b), Sur l'application du calcul des probabilités à l'inoculation de la petite vérole. Ibidem, pp. 26 – 95.
--- (1768), Sur la durée de la vie. Ibidem, t. 4, pp. 92 – 98.
De Moivre, A. (1712, in Latin), De mensura sortis, or, On the measurement of chance. *Intern. Stat. Rev.*, vol. 52, 1984, pp. 237 – 262 with comment by A. Hald (pp. 229 – 236).
--- (1718), *Doctrine of Chances*. London, 1738 and 1756. Reprint of third ed.: New York, 1967. The two last editions include the author's translation of his Method of approximating the sum of the terms of the binomial … (1733, in Latin). The third edition also carries a reprint of the Dedication of the first edition to Newton (p. 329).
--- (1724), *Treatise of Annuities on Lives*. In De Moivre (1756, pp. 261 – 328).

Euler, L. (1778, in Latin), Commentary on Bernoulli D. (1778). Translation into English published together with Bernoulli's memoir. Euler's memoirs on probability, statistics and treatment of observations reprinted in his *Opera omnia*, ser. 1, t. 7. Leipzig – Berlin, 1923.
Herschel, W. (1784), Account of some observations. *Scient. Papers*, vol. 1. London, 1912, pp. 157 – 166.
--- (1817), Astronomical observations and experiments tending to investigate the local arrangement of celestial bodies in space. Ibidem, vol. 2, pp. 575 – 591.
Juskevic, A. P. et al, Editors (1959), *Die Berliner und die Petersburger Akademie der Wissenschaften in Briefwechsel L. Eulers*, Bd. 1. Berlin.
Kendall, M. G., Plackett, R. L., Editors (1977), *Studies in the History of Statistics and Probability*, vol. 2. London. Coll. reprints.
Lagrange, J. L. (1867 – 1892), *Oeuvres*, tt. 1 – 14. Paris.
In t. 2 (1868): Sur l'utilité de la méthode de prendre le milieu entre les résultats de plusieurs observations (1776), pp. 173 – 234.
In t. 4 (1869): Recherches sur les suites récurrentes (1777), pp. 151 – 251.
In t. 13 (1882): his correspondence with Dalembert.
In t. 14 (1892): his correspondence with other scientists.
Lambert, J. H. (1760), *Photometria.* Augsburg.
--- (1765 – 1772), *Beyträge zum Gebrauch der Mathematik und deren Anwendung*, Tl. 1 – 3. Berlin. The first part (1765) contains *Anmerkungen und Zusätze zur practischen Geometrie* (pp. 1 – 313) and *Theorie der Zuverlässigkeit der Beobachtungen und Versuche* (pp. 424 – 488). The third part (1772) contains *Anmerkungen über die Sterblichkeit, Todtenlisten, Geburthen und Ehen* (pp. 476 – 569).
Laplace, P. S. (1798 – 1825), *Traité de mécanique céleste*, tt. 1 – 5. Paris. See below his *Oeuvr. Compl.*
--- (1878 – 1912), *Oeuvres complètes*, tt. 1 – 14. Paris.
TT. 1 – 5 (1878 – 1882) this being a reprint of the *Méc. Cél.* English transl.: *Celestial Mechanics* (1832), vols 1 – 4. New York, 1966.
T. 6 (1884) is a reprint of the 1835 edition of *Exposition du système du monde* (1796).
T. 7 (1886) is the *Théorie analytique des probabilités* (1812) with its preface, *Essai philosophique sur les probabilités* (1814) and four Supplements (1816 – ca. 1819). Transl. of the *Essai*: *Philosophical Essay on Probabilities*. New York, 1995.
T. 8 (1891) contains Sur les suites récurro-récurrentes (1774), pp. 5 – 24; Sur la probabilité des causes par les événements (1774), pp. 27 – 65) and Recherches sur l'intégration des équations différentielles aux différences finies (1776), pp. 69 – 197.
T. 9 (1893) contains Sur les probabilités (1781), pp. 383 – 485.
T. 10 (1894) contains Sur les approximations des formules qui sont fonctions de très-grands nombres (1785 – 1786), pp. 209 – 338.
T. 11 (1895) contains Sur les naissances, les mariages et les morts (1786), pp. 35 – 46, and Sur quelques points du système du monde (1789), pp. 477 – 558.
T. 14 (1812) contains Leçons de mathématiques données à l'École normale en 1795 (1812), pp. 10 – 177.
Maire [, C.], Boscovich [, R. G.] (1770), *Voyage astronomique et géographique dans l'Etat de l'Eglise*. Paris. The adjustment of observations is treated in Livre 5 written by Boscovich.
Maupertuis, P. L. M. (1756), Lettres. *Oeuvres*, t. 2. Lyon, 1756, pp. 185 – 340.
Montmort, P. R. (1708), *Essay d'analyse sur les jeux de hazard*. Paris, 1713. Reprinted: New York, 1980.
Newton, I. (1704), *Optics*. *Opera quae extant omnia*, vol. 4. London, 1782, pp. 1 – 264. Reprinted from edition of 1718.
Poincaré, H. (1896), *Calcul des probabilités*. Paris, 1912. Reprinted: Paris, 1923 and 1987.
Schneider, I., Editor (1988), *Die Entwicklung der Wahrscheinlichkeits-theorie von den Anfangen bis 1933*. Darmstadt. Collection of reprints and translations, mostly in English.
Simpson, T. (1740), *Nature and Laws of Chance*. London.
--- (1756), On the advantage of taking the mean of a number of observations in practical astronomy. *Phil. Trans. Roy. Soc.*, vol. 64, pp. 82 – 93.
--- (1757), Revised version of same in author's *Misc. Tracts on Some Curious … Subjects in Mechanics* … London, pp. 64 – 75.
Süssmilch, J. P. (1741), *Die Göttliche Ordnung*. Several later editions. Reprint of the edition of 1765 with Bd. 3 of 1776: Göttingen – Augsburg, 1988.
2. Studies
Eisenhart, C. (1989), Laws of error. In Kotz et al (1982 – 1989, vol. 4, pp. 530 – 566).

Farebrother, R. W. (1993), Boscovich's method for correcting discordant observations. In P. Bursill-Hall, Editor, *Boscovich. Vita e attività scientifica. His Life and Scientific Work*. Roma, pp. 255 – 261.
Fieller, E. C. (1931), The duration of play. *Biometrika*, vol. 22, pp. 377 – 404.
Freudenthal, H. (1951), Das Petersburger Problem in Hinblick auf Grenzwertsätze der Wahrscheinlichkeitsrechnung. *Math. Nachr.*, Bd. 4, pp. 184 – 192.
Freudenthal, H., Steiner, H.- G. (1966), Aus der Geschichte der Wahrscheinlichkeitstheorie und der mathematischen Statistik. In Behnke, H. et al, Editors, *Grundzüge der Mathematik*, Bd. 4. Göttingen, pp. 149 – 195.
Gillies, D. A. (1987), Was Bayes a Bayesian? *Hist. Math.*, vol. 14, pp. 325 – 346.
Hald, A. (1990), *History of Probability and Statistics and Their Applications before 1750*. New York.
--- (1998), *History of Mathematical Statistics from 1750 to 1930*. New York.
Henny, J. (1975), Niklaus und Johann Bernoullis Forschungen auf dem Gebiet der Wahrscheinlichkeitsrechnung. In J. Bernoulli (1975, pp. 457 – 507).
Heyde, C. C., Seneta, E., Editors (2001), *Statisticians of the Centuries*. New York.
Johnson, N. L., Kotz, S., Editors (1997), *Leading Personalities in Statistical Sciences*. New York.
Jorland, G. (1987), The St.-Petersburg paradox, 1713 – 1937. In Krüger, L. et al, Editors, *Probabilistic Revolution*, vol. 1. Cambridge (Mass.), pp. 157 – 190.
Kohli, K. (1975), Spieldauer. In J. Bernoulli (1975, pp. 403 – 455).
--- (1975), Aus de Briefwechsel zwischen Leibniz und J. Bernoulli. Ibidem, pp. 557 – 567.
Kotz, S., Johnson, N. L., Editors (1982 – 1989), *Encyclopedia of Statistical Sciences*, vols 1 – 9. Update vols 1 – 3, 1997 – 1999. New York.
Paty, M. (1988), Dalembert et les probabilités. In Roshdi, R., Editor, *Les sciences à l'époque de la Révolution Française*. Paris, pp. 203 – 265.
Pearson, E. S., Plackett, R. L., Editors (1970), *Studies in the History of Statistics and Probability*. London. Coll. reprints.
Pearson, K. (1924), Historical note on the origin of the normal curve of errors. *Biometrika*, vol. 16, pp. 402 – 404.
--- (1925), James Bernoulli's theorem. Ibidem, vol. 17, pp. 201 – 210.
--- (1978), *History of Statistics in the 17th and 18th Centuries* etc (Lectures of 1921 – 1933). London.
Schneider, I. (1968), Der Mathematiker A. De Moivre. *Arch. Hist. Ex. Sci*., vol. 5, pp. 177 – 317.
Seal, H. L. (1949), Historical development of the use of generating functions in probability theory. *Bull. Assoc. Actuaires Suisses*, t. 49, pp. 209 – 229. Reprinted: Kendall & Plackett (1977, pp. 67 – 86).
Sheynin, O. Many contributions; see at www.sheynin.de
Shoesmith, D. (1987), The Continental controversy over Arbuthnot's argument etc. *Hist. Math.*, vol. 14, pp. 133 – 146.
Stigler, S. M. (1977), Eight centuries of sampling inspection. The trial of the pyx. *J. Amer. Stat. Assoc.*, vol. 72, pp. 493 – 500.
--- (1986), *History of Statistics*. Cambridge (Mass.) – London. Contains slandering statements concerning Euler and Gauss.
Takacs, L. (1969), On the classical ruin problem. *J. Amer. Stat. Assoc.*, vol. 64, pp. 889 – 906.
Thatcher, A. R. (1957), Note on the early solutions of the problem of the duration of play. *Biometrika*, vol. 44, pp. 515 – 518. Reprinted: E. S. Pearson & Kendall (1970, pp. 127 – 130).
Todhunter, I. (1865), *History of the Mathematical Theory of Probability*. New York, 1949, 1965.
Walker, Helen M. (1929), *Studies in the History of the Statistical Method*. New York, 1975.
Westergaard, H. L. (1932), *Contributions to the History of Statistics*. New York, 1968.
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.
Zabell, Sandy L. (1988), The probabilistic analysis of testimony. *J. Stat. Planning and Inference*, vol. 20, pp. 327 – 354.
--- (1989), The rule of succession. *Erkenntnis*, Bd. 31, pp. 283 – 321.