Journ@l Electronique d'Histoire des Probabilités et de la Statistique

Electronic Journ@l for History of Probability and Statistics

Vol 8; Décembre/December 2012

www.jehps.net

Laplace - a pioneer of statistical inference

VESA KUUSELA[1]

Abstract

It is generally held that R.A. Fisher, Jerzy Neyman and Egon Pearson laid the foundations of statistical inference in the 1920s and 1930s. Recent research concerning the history of statistical ideas has discovered that already at the end of the 18th century Laplace sketched a method, called the principle of inverse probability, which involved characteristics features of statistical inference. Laplace's estimation of the size of population in France, applying inverse probability principle, is the first scientifically ambitious partial investigation, or sample survey. Laplace's investigation embraced many of the phases that modern sample surveys have. The inverse probability principle and his later contributions, such as the Central Limit Theorem, eventually laid foundations for the later development of statistical inference. Laplacian paradigm was the dominant model in statistical science up to the 1920s but since then it has fallen into oblivion – apparently, because it has been regarded as 'Bayesian'. Recent research has shown that Laplace's and Fisher's ideas actually were close to each other even though they were based on different inference model.

1 Introduction

In the contemporary literature on statistical science, statistical inference denotes a procedure by which it is possible to draw probabilistic conclusions about characteristics of a population in the presence of random variation. The source of random variation may be the randomized drawing of a sample from a population, or the randomization in an experiment, or the fact that measurements involve random errors. Statistical inference assumes slightly different forms when applied in the finite population framework or in the infinite population framework. In inference for finite or fixed populations the emphasis is on estimation of population parameters and confidence intervals for estimates. In inference for infinite or hypothetical populations emphasis is on significance or hypothesis testing.

A widespread conception has been that the fundamental ideas of statistical inference were first established by R. A. Fisher in 1920s [Fisher, 1922, 1925a, 1925b, 1930]. Later, Jerzy Neyman and Egon Pearson developed, basing on Fisher's ideas, a modified version of statistical inference for infinite populations [Neyman and Pearson, 1933]. In 1934, Neyman published the famous paper in which he presented, applying again Fisher's ideas, an inference method for finite populations or for sample surveys [Neyman, 1934]. The contemporary literature on statistical inference does not recognize that mathematical methods for inference had existed before Fisher. Obviously, in the modern sense the theory

[1] University of Helsinki. Contact email: vesamatti.kuusela@elisanet.fi

did not exist. However, quite recently Hald [2007] has brought up the fact that many ideas of statistical inference can be traced back to the end of the 18th century.

Modern statistical inference can be seen as an outgrowth of the idea of inverse probability that was developed at the end of the 18th century mainly as a creation of Pierre Simon Laplace. In inverse probability the question is, given the outcome of an experiment or the observations, what can be concluded about the underlying causes of outcomes and their stochastic characteristics? Clearly, inverse probability addresses the same question as statistical inference.

The application of inverse probability, or statistical inference, in real world problems involves two different but intimately related tasks: the derivations of direct probabilities and probability distributions, which is a mathematical exercise and hence deductive inference; and the application of probability distributions to make inferences about the unobserved part of the population, which is inherently inductive.

Direct probabilities are abstract models that in a general form cannot be applied as such because the situations in the real world are much too diverse and complex. Application of direct probability model to a real world phenomenon requires a link in the form of an intermediate model – an inference model – that attaches the abstract probability model to the observed phenomenon. An inference model is a thought model or thought experiment, i.e. a procedure or a conceptual model through which an abstract probability model is linked to real world events. Typical to a thought experiment is that it is composed of such a setup that can or could be tested experimentally if necessary.

In the current forms of statistical inference, the inference model is composed of repeated drawings of samples from the same population. These repeatedly drawn samples yield the sampling distribution of an estimate. This inference model was introduced by Fisher in the 1920s. In practical work, repeated sampling is never carried out, though. Samples are concretely drawn repeatedly only occasionally in theoretical developments or for teaching purposes. Before Fisher, the prevalent inference model was based on urn model or Bernoulli experiments. Laplace's inverse probability theory was based on the urn model. It was the model that mathematicians and statisticians subsequently applied throughout the 20th century.

Laplace's derivation of inverse probability principle is described in Chapter 3. In Chapter 4 is described how Laplace planned to estimate the size of population of France using this principle. In Chapter 5, it is explained how he eventually put his plan in practice by conducting a partial investigation, or sample survey, and estimated the size of population. In Chapter 7, comprises of a brief description of Laplace's other contributions related to statistical inference. The progressions of Laplace's ideas in course of the 20th century are sketchily described in Chapter 8.

2 Probability theory in 18th century France

In the second half of the 18th century, France was rich in mathematicians dealing with problems of probability theory and topics around probability calculus: such as Jean le Rond D'Alembert, Joseph Louis Lagrange, Marie Caritat de Condorcet, Antoine Lavoisier, Andrien-Marie Legrendre, and Pierre Simon Laplace. They had connections with Daniel Bernoulli, Leonhard Euler, and Heinrich Lambert, who mainly worked in Switzerland. Evidently, the French mathematicians were also aware of the contributions of the British mathematicians. Simpson's as well as De Moivre's works were known and often cited.

However, Bayes' Essay obviously was not known – at least not immediately (see also [Stigler, 1986] and [Hald, 1998]).

Stigler called the era as "the golden age of French science". A line of research was theory of probability, including the methods related to inductive inference. In doing this, the French mathematicians laid the foundations of inverse probability. Both Pearson [1978] and Hacking [1975] share the opinion that Laplace's and Condorcet developed together central parts of its theory. Condorcet's writing style was somewhat perplexing and his papers never gained a wider audience nor importance[2]. Laplace's contributions, on the other hand, were path breaking, not only for the development probability theory in the 19th century, but also for statistical science in general.

Euler's contributions in mathematical analysis and series expansions paved the ground for the developments in probability theory. Especially, Laplace's mathematical analysis leaned heavily on his works[3]. Apparently, inspired by Euler's mathematical innovations Laplace developed an analytic approach to inverse probability. Laplace's thought model was an urn trial (Bernoulli trial) and mathematical analysis typically involved (Euler's) series expansions, solving large factorials with Stirling's formula[4], and omitting terms that were negligible in large samples. Laplace's approach was inherently different from that of Thomas Bayes' that was based on Newtonian geometric reasoning.

According to Pearson [1978], Laplace wrote some eighteen memoirs dealing with the theory of the probability. The most significant were written in the years 1772 – 1783[5]. Especially, two of them, "*Mémoire sur la probabilités des causes par les évènemens*"[6] (PCE) in 1774 and "*Mémoire sur les probalités*"[7] (MOP, written in 1778 but published in 1781) became very influential. It is noteworthy that Laplace was only 25 years old when he published PCE. When referring to Laplace's contributions, it is customary to refer to Laplace's later publications, "*Essai Philosophique sur les Probabilités*" and "*Théorie Analytique des Probabilités*". They were both first published in 1812 (Laplace 1812a, 1812b], and several editions were published in the first half of the 19th century. Although these two are among the best-known contributions to probability theory of Laplace, they were partly composed of the results already published in the earlier memoirs.

In the PCE, Laplace presented his idea of inverse probability. Compared to some of Laplace's other works, this memoir reads fairly clearly and effortlessly and, as Stigler [1986] points out, even after more than two centuries, it seems almost like a contemporary work. The second memoir, MOP, deals with more topics than the first one and is more elaborate. In it, Laplace completed his application of probability calculus to the analyses of

---

[2] Dale [1999] gives a comprehensive account and analysis about Condorcet's papers.

[3] Laplace gave three advices to his students how to learn mathematics: "Read Euler! Read Euler! Read Euler!"

[4] Strirling's formula for $\ln(n!)$:

$$\ln(n!) = \frac{1}{2}\ln(2\pi) + (n+\frac{1}{2})\ln(n) - n + \frac{1}{12}n^{-1} - \frac{1}{360}n^{-3} + \frac{1}{1260}n^{-5}...$$

[5] Laplace's mathematical developments and the contributions on probability theory, especially on inverse probability, are comprehensively and minutely explored and analysed by Stigler [1986], Dale [1999], and Hald [1998, 2007].

[6] "Memoir on the Probability of the Causes of Events" (PCE). Stephan Stigler has translated the memoir and it was published in *Statistical Science* [Stigler, 1986b].

[7] "Memoir on Probabilities" (MOP)

errors in observations, which was left unfinished in the previous memoir. He also introduced the principle of probabilistic inferences in a more explicit manner.

3 Laplace's inverse probability[8]

In the beginning of MOP, Laplace identifies three different types of probabilities:

> "In the analysis of chance, we intend to know the probability of composite events, following any law, of simple events of which the possibilities are given; these are able to be determined in these three ways: 1° *a priori*, when, by the like nature of the events, we see that they are possible in a given ratio; it is in the same way, in the game of *heads* and *tails*, if the piece that we cast into the air is homogeneous and if its two faces are entirely similar, we judge *heads* and *tails* equally possible; 2° *a posteriori*, by repeating a great number of times the experience which can bring about the event of which there is question, and by examining how many times it has happened; 3° finally, by the consideration of the grounds which can resolve for us to say on the existence of this event; if, for example, the respective skills of two players A and B are unknown, as we have no reason to suppose A more strong that B, we conclude from it that the probability of A to win a game is ½. The first of these ways gives the absolute probability of the events; the second makes it known very nearly as we will just see in the following, and the third gives only their possibility relative to the state of our knowledge.
>
> Each event being determined by virtue of the general laws of this universe, it is probable only relatively to us, and, for this reason, the distinction of its absolute possibility and of its relative possibility can seem imaginary; but we must observe that, among the circumstances which compete in the production of the events, there are some variables at each instant, such as the movement that the hand imprints on the dice, and it is the reunion of these circumstances which we name: it is of others which are constant, *chance* such as the ability of the players, the inclination of the dice to fall on one of their faces rather than on the others, etc.; these form the *absolute possibility* of the events, and their knowledge more or less extensive forms their *relative possibility*; alone, they do not suffice to produce them; it is more necessary that they be joined to the variable circumstances of which I speak; they serve thus only to augment the probability of the events, without determining necessarily their existence." [Laplace 1778]

Following the contemporary classification of probabilities, Laplace's first definition corresponds to the definition of classical probability, and the second to the definition of frequentist probability. They both are definitions of objective probability. The third definition of Laplace's corresponds to the modern meaning of subjective probability. Unlike in the modern statistical texts, for Laplace, *a priori* probability meant objective probability. In modern writings, *a priori* probability is obtained usually by subjective judgements. There is a danger of mixing up the modern definitions of probability with those of Laplace's.

3.1 The Principle of Inverse Probability

Laplace begins his presentation of inverse probability by first defining the difference between direct and indirect (or inverse) probability. An example of direct probability is an urn that is known to contain only white and black tickets in a given ratio, and one seeks the probability that a ticket drawn by chance will be white. Laplace says that in this case, the event is uncertain but the cause on which the probability of occurrence depends is known. He continues by defining the *Problem* (of inverse probability) for which he later gives solutions in different formats:

[8] Laplace did not use the term *inverse probability*. It was first used by the English scientist, August de Morgan in the 1830s (see [Dale, 1999], p. 4).

"An urn is supposed to contain a given number of white and black tickets in an unknown ratio; if one draws a ticket and finds it white, determine the probability that the ratio of white and black tickets is that of p to q. The event is known and the cause is unknown." [Laplace, 1774]

In order to solve the *Problem*, he defined the *Principle* (of inverse probability):

"If an event can be produced by a number n of different causes, the probabilities of these causes given the event are to each other as the probabilities of the event given the causes, and the probability of the existence of each of these is equal to the probability of the event given that cause, divided by the sum of all the probabilities of the event given each of these causes." [Laplace, ibid.]

In modern mathematical form[9], Laplace's *Principle* can be written as follows: If the "event" is denoted by E and $a_1, a_2, \ldots, a_n$ the n potential causes, then

$$\frac{P(a_i \mid E)}{P(a_j \mid E)} = \frac{P(E \mid a_i)}{P(E \mid a_j)}$$

and further

$$P(a_i \mid E) = \frac{P(E \mid a_i)}{\sum_j P(E \mid a_j)} \qquad (3.1)$$

It is easy to notice that Laplace's *Principle* has the same form as the Bayes' formula with equal *a priori* probabilities. Hald [1998] argued that Laplace did not realize that his *Principle* could be derived as a conditional probability under the assumption that the causes are uniformly distributed as Bayes had done. At first, *a priori* probabilities did not have any role in the *Principle*. Only in the second edition of *Théorie Analytique des Probabilités* (published in 1814), did Laplace prove the general version of his *Principle*, which is the same as Bayes' theorem.

In effect, the formula (3.1) says that the probability of a_i being the cause of the observed event E is proportional to the direct probability of the event being the cause, that is $P(a_i \mid E) \propto P(E \mid a_i)$. Later, Laplace applied this idea in solving his problems where the observations, say $x_1, \ldots, x_n$, are obtained for a given value of a continuously varying parameter o , i.e.

$$P(\theta \mid x_1, \ldots, x_n) \propto P(x_1, \ldots, x_n \mid \theta) \qquad (3.2)$$

Laplace did not try to prove the *Principle*. He presented it more like a self-evident fact - or like a postulate. Neither did he provide many clues how he had come up with it. Hald [ibid.] concluded that the intuitive background of the *Principle* may have been the same reasoning that had led Lambert [1760] to the Maximum Likelihood principle: If the probability of the observed event for a given cause is large relative to the other probabilities, then it is more likely that the event has been produced by this cause than by the other causes.

Laplace gave the following example to illustrate the *Principle*: There are two urns, A and B. The first contains p white tickets and q black tickets, and the second contains p' white tickets and q' black tickets. One draws f white and h black tickets from either of these urns, but not knowing which of the urns. It is required to determine what is the probability that the urn where the tickets were drawn was A or B?

---

[9] Laplace never used the discrete form, though. All his mathematical treatments were composed of continuous functions. The discrete forms of the formula appeared much later.

Laplace presented the following solution: Assuming that the urn was *A*, the probability of getting *f* white and *h* black tickets from it is

$$K = P(f,h \mid A) = \frac{\dfrac{(f+h)!(p+q-f-h)!}{f!h!(p-f)!(q-h)!}}{\dfrac{(p+q)!}{p!q!}} \tag{3.3}$$

The probability that the urn was *B*, i.e., *K' = P(f , h|B)*, can be obtained in a similar way replacing *p* and *q* by *p'* and *q',* respectively. Applying the *Principle*, Laplace concluded that the probability that the urn was *A* is *P(A) = K / (K + K')* and the probability that it was *B* is *P(B) = K' / (K + K')*. [See Laplace 1774]

In essence, this may be regarded as an example of testing a simple hypothesis against a simple alternative if "causes" are replaced by "hypotheses".

3.2 The Principle of Insufficient Reason

A priori probability was a significant element in Bayes' and later in Laplace's inverse probability. There is little evidence that its use had been notably challenged before R.A. Fisher in the first quarter of the 20^{th} century. One reason for the use of *a priori* probability obviously was the implicit idea of a superpopulation[10], which was involved in all inferential developments. This obviously was a consequence of the Newtonian worldview, which assumed that observations of the real world were realizations of an unknown cause system, which governed the world in the background. The parameters of the superpopulation ("states of the nature") were assumed changing constantly but unknown and therefore, they were considered as random variables, which had a probability distribution. This distribution was usually unknown and therefore, according to the *Principle of Indifference*, the rectangular prior distribution was necessary.

> **The principle of indifference**: if on the background information *B* the hypotheses (H_1, H_2, … H_N) are mutually exclusive and exhaustive and B does not favour any one of them over any other, then
>
> $$P(H_i \mid B) = \frac{1}{N}, 1 \le i \le N$$

The principle of indifference or the *principle of insufficient reason* is often dedicated to Laplace. Stigler [1986] argues that the application of the principle of indifference was not a metaphysical assumption concerning the unknown structure of the world (equally likely causes). Rather, it was an implicit assumption that for ease of analysis specified in such a manner that this principle of insufficient reason was reasonable.

3.3 Applications of the Principle[11]

Laplace demonstrated the use of his *Principle* by solving three different problems. The first was defined in following way:

---

[10] The concept of superpopulation was not used nor even recognized by Laplace or any of his contemporaries or followers. Only R.A. Fisher introduced it [Fisher, 1922].

[11] In this chapter, Laplace's original style of writing about mathematical topics is used, to demonstrate his reasoning and how he derived formulas. In general, his style was complicated and difficult to read for a modern reader.

"If an urn contains an infinity of white and black tickets in an unknown ratio, and we draw $p + q$ tickets from it, of which p are white and q are black, then we require the probability that when we draw a new ticket from the urn, it will be white." [Laplace, 1774]

Implicitly Laplace assumed that the unknown ratio, x, of white tickets to all tickets in the urn has a continuous range of probabilities, $0 \le x \le 1$, all values equally possible. The probability of drawing p white tickets and q black tickets from the urn in a single drawing is $x^p(1- x)^q$. It should be noted that Laplace assumed that the urn contains an infinite number of tickets, and therefore subsequent drawings can be regarded independent. Laplace concluded by applying the *Principle*[12] that the probability that x is the true ratio is:

$$P(x \mid p,q) = \frac{x^p(1-x)^q\,dx}{\int_0^1 x^p(1-x)^q\,dx} \tag{3.4}$$

Assuming that x is the true ratio of white tickets to all tickets, the probability of drawing a white ticket from the urn is x. The probability of drawing a white ticket from the urn with true ratio x is obtained by multiplying (3.4) "by the probability of the supposition" (probability of drawing a white ticket from an urn with true ratio $x \times$ probability that x is the true ratio):

$$x \times P(x \mid p,q) = \frac{x^{p+1}(1-x)^q\,dx}{\int_0^1 x^p(1-x)^q\,dx} \tag{3.5}$$

And the total probability of drawing a white ticket from the urn, E, Laplace shows to be

$$P(E) = \frac{\int_0^1 x^{p+1}(1-x)^q\,dx}{\int_0^1 x^p(1-x)^q\,dx} \tag{3.6}$$

The expression for the probability of E, i.e., the ticket in a new drawing being white, reduces after repeated integration by parts to

$$P(E) = \frac{p+1}{p+q+2} \tag{3.7}$$

Later, this result was called Laplace's *Rule of Succession*. Laplace demonstrated the rule by posing a question: how certain can we be that the sun will rise tomorrow, given that we know that it has risen every day for the past 5000 years (1,825,000 days). One can be pretty sure that it will rise, but we can't be absolutely sure. According to the *Rule of Succession* the answer is, given that there has been a success in all earlier trials (p = 1,825,000) and no failures ($q = 0$), we can be 99.999945% sure that the sun will rise tomorrow.

Using his *Principle* and "Euler's series", Laplace continues to show that if $m + n$ new tickets are drawn from the urn, the probability of getting m white tickets and n black tickets is

$$P(E) = \frac{p^m q^n}{(p+q)^{m+n}} \tag{3.8}$$

---

[12] In the 18th century writings on probability, only continuous distributions were analysed.

Due to approximations, Laplace expected this to hold “without fearing any appreciable error” when *p* and *q* are very large, and *m* and *n* very small in comparison to *p* and *q*.

Laplace concluded that the solution to this problem provided a direct method to determine the probability of future events after those that have already occurred. This principle later came to be known as the *Principle of Learning from Experience*.

Laplace continues by saying that because it is a broad subject he gives only a “rather singular proof” of the following theorem:

> “One can suppose that the numbers *p* and *q* are so large that it becomes as close to certainty as one wishes that the ratio of the number of white tickets to the total number of tickets contained in the urn is included between the two limits *p/(p+q) – w* and *p/(p+q) + w*, one can suppose *w* to be less than any given quantity.”

Using the results of the preceding examples, Laplace concludes that the probability of ratio *x* being between the given limits is

$$P\left(p/(p+q)-w\le x\le p/(p+q)+w\right)=\frac{\int x^{p}(1-x)^{q}\,dx}{\int_0^1 x^{p}(1-x)^{q}\,dx} \qquad (3.9)$$

if the integral in the numerator of (3.9) is taken from *p/(p+q) – w* to *p/(p+q) + w*. Marking the ratio *x = p/(p+q) + z*, Laplace shows, using approximations, that the probability (3.9) is approximately

$$\frac{(p+q)^{3/2}}{\sqrt{2\pi pq}}\int_0^w 2e^{-(p+q)^3 z^2/2pq}\,dz \qquad (3.10)$$

Then using “M. Euler’s integral calculus”, he continues to show that this integral is approximately 1.

The second problem that Laplace treated in PCE dealt with the division on wins in a game of chance, which had to broke before an orderly end. It is not strictly related to the current theme and is therefore passed over.

The third and obviously most famous problem of inverse probability in PCE was “to determine the mean that one should take among three given observation of the same phenomenon”. The motivation to take up this problem originated from a writing of Daniel Bernoulli, who in a footnote stated that this was an important problem, which he had not solved. Lagrange had also touched on the problem, but had not solved it. Laplace’s treatment of this topic was important because he was able to show how the *Principle* could be applied to nontrivial practical problems.

Laplace phrased the problem as follows (see Figure 3):

> “Given three observations *a, b, c* of a phenomenon along a time axis *AB*. The time interval between *a* and *b* is *p* seconds and between *b* and *c*, *q* seconds. We wish to find the point *V* on the line where we should fix the mean that we should take between the three observations. It is supposed to represent the “true time” of the phenomenon.” [Laplace, 1774]

He assumed that any observation differing from *V* by a factor *x* would lead to a probability which could be represented by a curve $y=\varphi(x)$. He stated three conditions for the error curve $\varphi(x)$, which should help to determine its true form:

(1) $\varphi(x)$ must be symmetrical about *V*, since errors occur in both directions equally likely;

(2) $\varphi(x)$ must decrease asymptotically to ordinate as $|V - x|$ gets greater, because "the probability that the observations differs from truth by an infinite distance is evidently zero";

(3) $\int \varphi(x)dx = 1$ since it is certain that the observation will fall on a point under the curve.

Laplace concluded that the probability that three observations deviate from point *V* by distances *Va*, *Vb*, and *Vc* is $\varphi(x)\cdot\varphi(p-x)\cdot\varphi(p+q-x)$. If it is assumed that the true instance is *V'* and that *V'a* = *x'* then the probability would be $\varphi(x')\cdot\varphi(p-x')\cdot\varphi(p+q-x')$. Applying the "fundamental Principle", Laplace concluded that the probabilities that the true instance is at the point *V* or *V'* are related to each other as

$$\frac{\varphi(x)\cdot\varphi(p-x)\cdot\varphi(p+q-x)}{\varphi(x')\cdot\varphi(p-x')\cdot\varphi(p+q-x')}$$

Next, Laplace noted that in seeking the mean to be chosen, there are two things that may occur: it is equally probable that the true instant of the phenomenon falls before or after it. Laplace called this the mean of probability. The second is the instant that minimizes the sum of the "errors to be feared"[13] multiplied by their probabilities. Laplace called this the mean of error, or the astrological mean.

Laplace continues by a similar approach by which Simpson [1755] had derived his error distribution: Given the true instant *V*, the posterior probability of the three observations *a*, *b*, and *c* can be expressed as:

$$\varphi(x|p,q) = \varphi(x)\varphi(p-x)\varphi(p+q-x)$$

The point, *V,* has disappeared from the formula and it is involved only in the first error term, *x*, which was an unobserved and random quantity. The error term, *x*, became the object of Laplace's investigation, the "*cause*" to be found from the observable "*events*" *p* and *q*. The distribution of "*events*" *p* and *q*, given the cause, $x, \varphi(p,q|\, \mathbf{x})$ was proportional to $\varphi(x,p,q)$. Therefore, by his *Principle*, the distribution of *x* was

$$\varphi(x|p,q) \propto \varphi(x,p,q)$$

Laplace had an error in his formula, and therefore he ended up with an error distribution that had a wrong "shape". In his later writings, Laplace ended up with the correct form for the distribution.

4 Plan to estimate the size of the population of France

Laplace presented the idea for estimating the size of the population of France in a memoir already in 1783[14], nearly 20 years before the survey actually was undertaken. A motive to undertake the survey in 1802 may have been the fact that the French government in 1800

[13] Laplace's "error to be feared" ("*errour á craindre*") is conceptually close to the modern notion of standard error.

[14] *"On the births, the marriages and the deaths at Paris, from 1771 to 1784; & at the whole extent of France, during the years 1781 & 1782"* or *"Sur les Naissances, les Mariages et les Morts*" [Laplace, 1783]

had established a Central Statistical Office and prescribed a general enumeration of the population in 1801. It took more than two years before the census returns were received and processed, and the resulting figure of 27,349,0003 inhabitants was considered unreliable.

Laplace's plan was based on the fact that in France all births were registered in parishes and published during the last quarter of the 18th century. His idea was to take a sample of the departments, count the total population in them for a single day, and then estimate the population of the whole country, using that information combined with information on registered births in France. Based on earlier demographic studies, he assumed that the ratio of population to births during a year was relatively stable. Another essential assumption was that the proportion of women of childbearing age in the population remained stable. Bru [1988] gives a comprehensive analysis about the background of the Laplace's plan.

4.1 Determining the required sample size

Laplace's plan was to use *a ratio estimator* but he did not invent the ratio estimation. Graunt had estimated the population of London in 1662 by a similar method [see Graunt 1662]. However, Laplace extended the survey to the whole of France while Graunt investigated only London and in a more intuitive manner[15]. The true importance of Laplace's plan lies on the fact that he also wanted to estimate the accuracy of the estimate. In that context, he developed a method to calculate how large a sample was needed to obtain the required accuracy.

In the memoir of 1783, Laplace shows how the needed sample size could be estimated.

> "The ratio of the population to the births … can never be rigorously exact: by supposing in it even a rigorous precision, there would remain still on the population of France, the incertitude which is born of the action of the variable causes. The population of France, drawn from the annual births, is therefore only a probable result, & consequently susceptible to errors. It is to the analysis of chances to determine the probability of these errors, & to what point we must carry the denumeration, in order that it be very probable that they are contained within narrow limits. These researches depend on a new & yet little known theory, that of the probability of future events takes from observed events; they lead to some formulas of which the numerical calculation is impractical, because of the great numbers which we consider: but having given in this Volume & in the preceding, the principles necessary to resolve this kind of questions, & a general method to have in highly convergent series, the functions of great numbers; I have made application of it to the theory of the population deduced from births. The denumerations already made in France, & compared to the births, give very nearly 26 for the ratio of the population to the annual births; now if we take a mean among the births of the years 1781 & 1782, we have 973054½ for the number of annual births in the whole extent of the Realm, containing in it Corsica; by multiplying therefore this number by 26, the population of the whole of France, will be 25299417 inhabitants. Now I find by my analysis, that in order to have a probability of a thousand to one, of not being deceived by a half-million in this evaluation of the population of France, it would be necessary that the denumeration which has served to determine the factor of 26 had been of 771469 inhabitants. If we would take 26½ the ratio of the population to the births, the number of the inhabitants of France will be 25785944; & in order to have the same probability of not being deceived by a half-million on this result, the factor 26½ must be determined after a denumeration of 817219 inhabitants. It follows thence that if we wish to have for this object the precision which its importance requires, it is necessary to carry this denumeration to a million or twelve hundred thousand inhabitants." [Laplace, 1783]

It was an outstanding vision of Laplace that an estimate of the accuracy of the estimate was needed and that the accuracy depended on the size of the sample. He does not explain

[15] Graunt also planned to estimate the population of England but he never put his plan into action (see also [Hald, 1998]).

which analysis led to this conclusion. Twenty years later, Laplace published the *Central Limit Theorem*, which might have alluded to this.

Laplace continues explaining his inference model for applying the *Principle* to calculate the size of the sample:

> "We consider an urn which contains an infinity of white & black tickets in an unknown ratio, & and we suppose that in a first drawing we have extracted p white tickets & and q black tickets; we suppose next that in a second drawing we have extracted q' black tickets, but we are ignorant of the number of white tickets brought forth in this drawing; the mean which naturally presents itself in order to know this number in an approximate manner, is to suppose it with q' in the ratio of p to q, that gives pq'/q for this number. We determine presently the probability that the true unknown number will be contained in the limits
>
> $$\frac{pq'}{q}(1-\omega), \frac{pq'}{q}(1+\omega)$$
>
> or that which returns to the same, that the error of the result $\frac{pq'}{q}$ will not surpass $\frac{pq'\omega}{q}$." [Laplace 1783]

The ratio estimate of population is the number of white tickets, that is, $p' = pq'/q$. Karl Pearson [1978] pointed out that his model was unsatisfactory in two respects: the births are not regarded as part of the population, and the sample is not considered as part of the finite population.

In order to find out the needed sample size, it was necessary to derive the (sampling) distribution of p'. This Laplace solved using his *Principle* in a manner that does not open readily (see [Dale 1999], p. 218). If the unknown ratio of white tickets to the total number of ticket is x, and the unknown number of white tickets in the second drawing is p', the probability is

$$P(p', q' \mid x) = \frac{(p'+q')!}{p'!q'!} x^{p'}(1-x)^{q'} \qquad (4.1)$$

Laplace assumed that p' may obtain all values from $p'=0$ to $p'=\infty$ and "these values are more or less probable, according as they render the second drawing more or less probable."

Earlier he had shown that $P(q' \mid x) = \sum_{p'=0}^{\infty} \frac{(p'+q')!}{p'!q'!} x^{p'}(1-x)^{q'} = \frac{1}{1-x}$

It follows that:

$$P(p' \mid q', x) = \frac{P(p', q' \mid x)}{P(q' \mid x)} = \frac{(p'+q')!}{p'!q'!} x^{p'}(1-x)^{q'+1}, p' = 0,1,2,... \qquad (4.2)$$

Laplace assumed that all values of the ratio are equally probable in the range from $x = 0$ to $x = 1$, i.e., that the *a priori* distribution of x is uniform.

Next, Laplace observes that the "entire probability" of p' is

$$P(p' \mid p, q, q') = \frac{(p'+q')!}{p'!q'!} \frac{\int_{x=0}^{1} x^{p+p'}(1-x)^{q+q'+1}\, dx}{\int_{x=0}^{1} x^{p}(1-x)^{q}\, dx} \qquad (4.3)$$

Obviously, at the time of this memoir, Laplace had not found the asymptotic expansion solutions, which were so typical in his later works. Therefore, he had to attack the problem in a different – and complicated – manner (for details, see [Laplace, 1783]).

Based on his earlier results, Laplace concludes that if s is 'less than and hardly different' from pq'/q, then

$$P(0 \leq p' \leq s \mid p,q,q') = \frac{1}{\sqrt{\pi}} \int_T^{\infty} e^{-t^2}\, dt \qquad (4.4)$$

where $T^2 = \dfrac{\left(\dfrac{p}{p+q} - \dfrac{s}{s+q'}\right)^2 \cdot (p+q)^3 (s+q')^3}{2sq'(p+q)^3 + 2pq(s+q')^3}$

The right hand side of formula (4.4) is approximately the normal distribution with zero mean and unit variance, i.e. *N(0,1)*. The approximation holds 'near the most probable values' of p'. If s is greater than pq'/q but close to it, the probability becomes

$$P(0 \leq p' \leq s \mid p,q,q') = 1 - \frac{1}{\sqrt{\pi}} \int_T^{\infty} e^{-t^2}\, dt \qquad (4.5)$$

Hence, it follows that

$$P(s \leq \frac{pq'}{p} \leq s') = 1 - \frac{1}{\sqrt{\pi}} \int_T^{\infty} e^{-t^2}\, dt - \frac{1}{\sqrt{\pi}} \int_{T'}^{\infty} e^{-t^2}\, dt$$

where T' is defined as T replacing s with s'.

If one sets $s = \dfrac{pq'}{q}(1-\omega), s' = \dfrac{pq'}{q}(1+\omega)$

and disregards terms of order ω^3 and two values of T^2 and T'^2 and takes $V^2 = \dfrac{pqq'\omega^2}{2(p+q)(q+q')}$ then

$$P(\frac{pq'}{q}(1-\omega) \leq \frac{pq'}{p} \leq \frac{pq'}{q}(1+\omega)) = 1 - \frac{2}{\sqrt{\pi}} \int_V^{\infty} e^{-t^2}\, dt \qquad (4.6)$$

In order to determine p (the "size of sample") Laplace denotes that proportion of white tickets to black by $i = p/q$ and the accepted error by $a = \dfrac{pq'}{q}\upsilon$ and hence $\upsilon = \dfrac{a}{iq'}$. The previous expression of V^2 yields $p = \dfrac{2i^2(i+1)q'^2 V^2}{a^2 - 2i(i+1)q'V^2}$. The value of a depends on the limits between which the error of the estimate is supposed to be.

Laplace gives an example by first letting a = 500000. The value of q' was the number of annual births in France, which was known to be q' = 973054.5 (the decimal expression resulted from a three-year average). The value of V depends on the probability, P, that the population size would be enclosed within the limits

$$P\left(\frac{pq'}{q} - a \leq p' \leq \frac{pq'}{q} + a\right)$$

Laplace set the probability to "a thousand to one", that is P =1000/1001;

$$\frac{2\int e^{-t^2}dt}{\sqrt{\pi}} = \frac{1}{1001}, or \int e^{-t^2}dt = \frac{\sqrt{\pi}}{2002}$$

where the integral is taken from $t = V$ to $t = \infty$. Laplace concludes that V^2 = 5.415. Based on the earlier enumerations, it was known "very nearly" that i = 26. Laplace also carries out similar calculations for i = 25.5 and i = 26.5. Apparently, there was some uncertainty in his mind concerning the value of i.

His conclusion was that in order to have a "probability of one thousand against one, of not being deceived by more than one half-million in the evaluation of the population of France", it is necessary that the "sample size", p, is either 727,210 inhabitants (for the smallest value of i), or 771,469 inhabitants, or 817,219 inhabitants (for the largest value of i).

Finally, Laplace concluded that "if we wish to have for this object, the probability that its importance requires, it is necessary to carry to a million or twelve hundred thousand inhabitants, the denumeration p which must determine the factor i."

5 Estimation of the size of the population

For estimation, Laplace selected 30 departments distributed over the area of France applying two criteria. First, all types of climates were represented. In this way, the effects of climate on the birth rate were compensated. Second, Laplace selected departments that had communes with mayors he thought were capable of providing accurate information. In modern terms, Laplace applied a two-stage cluster design except that the departments were selected with a purpose. However, Laplace did not indicate how he ended up with exactly 30 departments.

When the estimation took place, Laplace applied a slightly more elaborate method in estimation than he had proposed earlier in the derivation of the required sample size. Since the first paper, he had managed to solve some mathematical problems, and he had also developed new mathematical tools for probability calculus. The inference model was based on a Bernoulli trial, i.e., a box with white and black tickets.

A black ticket still represents a birth and each white ticket represents an individual living in the country. The first drawing represents the enumeration in which it is observed y births and the number of inhabitants is x[16]. The second drawing represents the population of the whole of France and the total number of births, X, is known while the corresponding population, Y, is unknown.

From the known total number of registered births, x, during the preceding year in the selected departments and that in the whole country, X, the ratio estimate of the population of France, Y, could be calculated as:

The combined population of the sampled Departments as of September 22, 1802, was

$$\hat{Y}_R = X\frac{y}{x}$$

2,037,615. As for births, Laplace totalled the sample births for the three-year period from

[16] In this chapter, Laplace's notations are replaced by a more modern style found, e.g., in Cochran's book on sampling [Cochran, 1953].

September 22, 1799, to September 22, 1802, obtaining a value of 215,599, so that his sample x is 215,599/3. By taking the average of the number of births, he hoped to eliminate random variation. Finally, in the numerical estimate of the sample ratio, y/x, was 28.352845 and then

> "supposing that the number of annual births in France is one million, which is nearly correct, we find, -- , the population of France to be 28 352 845 persons."

Laplace assumed that an infinite urn consisted of white and black tickets representing a population of French citizens on a specified day. In modern terms, he regards the number of known births in the country, X, as a random variable from a sample of unknown size Y, the population of France.

White tickets represented registered births in the preceding year. The ratio θ (proportion of white tickets) is unknown. He regarded the ratio x/y from the sample as a binomial estimate of θ .

Given the binomial sample data from the communes (x successes out of y trials), the posterior distribution of θ is (obtained by an application of the *Principle*)

$$\frac{\theta^x(1-\theta)^{y-x}}{\int_0^1 \theta^x(1-\theta)^{y-x}d\theta} \qquad (5.1)$$

In order to analyse the error in the ratio estimate, Laplace marked $Y = \hat{Y}_R + z$ and focuses on the distribution of z. The approximate form of the distribution function $f(z)$ when x, X, and y are all large, Laplace observed to be (for details, see [Cochran, 1978], or [Hald, 1998]).

$$f(z) \cong \exp\left\{-\frac{1}{2}\frac{x^3}{X(X+x)y(y-x)}\left[z^2 + z - \frac{2X(y-x)}{x^2}z\right]\right\} \qquad (5.2)$$

That shows that in large samples, the distribution of the error z in the ratio estimate $\hat{Y}_R$ was approximately normal with a variance

$$V(z) = X(X+x)y(y-x) / x^3$$

Laplace then calculated that the "standard error", given the data, was 107.550 persons, and he concluded that it makes "the odds about 300,000 to 1 against an error of more than half a million". In modern terms the result was that with probability $p < 3.33*10^{-6}$ the population size is outside limits 27,852,845 and 28,852,845.

6 How good was Laplace's estimator

The Laplace's method to estimate the size of the population is close to the methods that are given in the modern textbooks on survey sampling. The main differences of Laplace's method compared to modern sampling theory are the purposive selection of sample and the inference model that was based on Bernoulli trial. Thatcher [1964] compared so-called binomial prediction, based on Laplace's theory, and the theory of confidence limits. The interpretations of the inference models are different, but the comparison is interesting anyway. Thatcher [ibid.] found that the confidence limits lie outside the Laplacian limits, but the difference between them is no larger than the effect of one extra observation.

Cochran [1978][17] analysed the estimation method of Laplace and notice that, seen from modern perspective, the errors in it where minor. He contended that it was the first time that an asymptotic distribution of the ratio estimate has been worked out. In addition, he argued that an infinite superpopulation to study the properties of estimators had not been applied in the same manner as Laplace did since Brewer published his paper in 1963 [Brewer, 1963].

Applying Brewer's model to Laplace's problem, results conditional on the known value of X are obtained by writing $\hat{Y}_R = X/o_y$ and $Y = X/o_Y$, where o_y and o_Y are estimates of o obtained from binomial samples of sizes y and Y, respectively. Then

$$\hat{Y}_R - Y = X(\frac{1}{\theta_y} - \frac{1}{\theta_Y}) \cong X(\theta_Y - \theta_y)/\theta^2$$

Averaging over repeated selections of both drawings (France and a sample of communes) gives nearly the same as that of Laplace's estimate if the binominal selection was replaced by the hypergeometric [Cochran, 1978]. Cochran pointed out that the difference is caused by Laplace's implicit assumption that France itself and the sample of departments were independent binomial samples from an infinite superpopulation, or from an infinite urn, with an unknown ratio of births to population. In Brewer's model, the sample of communes is a subsample drawn from France.

7 Laplace's other contributions to probability theory

Laplace was primarily interested in astronomy but also his contributions in probability theory were diversified. He had a great number of ideas but he was not able to finalise all of them, and some of them were in a hidden form. For example, Laplace's *Principle* involves a similar reasoning as the idea of maximum likelihood. Another example is his work on a topic that is currently known as *sufficiency*. Stigler [1973] notes that Laplace did similar investigations as Fisher did a century later. Stigler [ibid.] found it surprising how close Laplace came to discovering sufficiency in 1818.

7.1 Central limit theorem

From the standpoint of statistical science, obviously the most important invention of Laplace was the *Central Limit Theorem* (CLT)[18]. It is generally held as Laplace's most significant result in probability theory. However, also Gauss' influence in the deriving the CLT was significant. The CLT asserts that, under certain general conditions, the sum of a large number of independent variables is approximately normally distributed.

Before Laplace published the CLT, the distribution of the arithmetic mean had been studied for many error distributions, resulting in complicated formulas. To avoid the complications, Laplace developed a new mathematical technique that was a combination of two of his main lines of mathematical methods: the theory of generating functions and the

[17] An interesting nuance is that Cochran referred to the later versions of *Théorie analytique des probabilités* and *Essai philosophique sur les probabilités.* It seems that he was unaware of Laplace's contributions in the 18th century.

[18] The actual term "central limit theorem" (in German: "zentraler Grenzwertsatz") was first used by George Pólya in 1920 in the title of a paper. Pólya (1920), "Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem", Mathematische Zeitschrift 8: 171–181.

method of asymptotic expansions of an integral. The characteristic functions, or Fourier transforms, were an outgrowth of a technique Lagrange had employed a few years earlier.

The memoir containing the CLT was read to the Academy of France in 1810 [Laplace, 1810]. Laplace's work on the CLT did not include the regularity conditions familiar in the modern form. They and the exceptional cases were later treated by several mathematicians e.g., Poisson, Cauchy, Liapounov, Lindeberg, etc.

The importance of CLT for all statistical inference is undeniable (see also [Fischer, 2010]). Only after the CLT was discovered the development of partial investigations, or sample surveys, could be justified.

7.2 Hypothesis testing

Laplace was keen in applying probability calculus on new areas. For example, he undertook an analysis on the sex ratio at birth. He had data from a twenty-six-year series in Paris. He found the total number of births to be $y = 251{,}527$ for boys and $z = 241{,}945$ for girls. If x represents the probability that a given birth is male, he calculated in a straightforward application the posterior probability:

$$P(x \leq ½ \mid y = 251527, z = 241945) = 1.1521 * 10^{-42}$$

He therefore regarded it as certain that the probability for a male birth is $p_x > ½$. This is clearly an example of a test of a hypothesis the null hypothesis being H_0: $p_x \leq ½$.

Below is a citation of another application that is also close to modern hypothesis testing.

> "We have observed that, in the interval of the 85 years elapsed from 1664 to 1757, there are born, in London, 737629 boys and 698958 girls, which gives around 19/18 for the ratio of births of boys to those of girls; this ratio being greater than the one of 105 to 101 which took place in Paris, and the number of births observed in London being very considerable, we would find for this city a greater probability that the births of boys are more possible than those of girls; but, when the probabilities differ likewise little from unity, they can be counted equal and confused with certitude.
>
> The preceding method gives a quite simple solution to an interesting problem, which it is perhaps very difficult to resolve by other methods: we have seen that the ratio of the births of boys to those of girls is sensibly greater in London than in Paris; this difference seems to indicate in London a greater facility for the birth of boys: the question is to determine how much this is probable.
>
> This value of P (the probability that the birth of a boy is less possible in London than in Paris = 1/410458) is a little too great; but, since taking in it only the first two terms of the series we would have a value too small, it is easy to conclude from it that the preceding can differ from the truth by the 1/142 part of its value, so that it is a strong approximation: there is therefore odds of more than four hundred thousand against one that the births of boys are more facile in London than in Paris. Thus we can regard as a very probable thing that it exists, in the first of these two cities, a cause more than in the second, which facilitates the births of boys, and which depends either on the climate or on the nourishment of the mothers." [Laplace, 1783]

8 Development of probability theory after Laplace

Laplace and Carl Friedrich Gauss were the central persons in shaping statistical science in its current form. It did not happen immediately, though. Since Laplace and Gauss, the development of probability theory subsided for a long time. In 1924, Rietz wrote:

> "The mathematical theory of statistics dates back to the first publication relating to Bernoulli's theorem in 1713. The line of development started by Bernoulli was carried forward by DeMoivre, Stirling, Maclaurin, and Euler culminating in the formulation of the Bernoulli theorem by Laplace in substantially the form in which it still holds a fundamental place in mathematical statistics.

> The *Théorie Analytique des Probabilités* of Laplace is undoubtedly the most significant publication at the basis of the development of mathematical statistics. Strangely enough, for a period of more than fifty years following the publication of the work of Laplace in 1812, little of importance was contributed to the subject. To be sure, the second law of error of Laplace was developed by Gauss and given its important place in the adjustment of observations, but there was on the whole relatively little progress…." [Rietz, 1924]

The period 1830–1890, starting after Laplace's and Gauss' most productive times, has been described as the one of clarification and consolidation of the works of Laplace and Gauss. During that period, probability theory was extended to applications from the natural sciences to the social and biological sciences. In addition, statistical reasoning emerged during that period resulting in the establishment of statistical institutes and statistical societies.

Siméon-Denis Poisson was one of the most important apprentices of Laplace, and he is said to be the first to understand the fundamental importance of Laplace's probabilistic work. He was a central person who systematized and extended Laplace's works and extended its application to vital statistics and law. Bru [2001] called him "the apostle of Laplacian science".

One of the Poisson's best known achievements in probability theory is currently known as the *Weak Law of Large Numbers* or the *Poisson law of large numbers*. This provided a rationale for applying probability to social matters. It was deemed to explain how the statistical stability of social affairs was possible (see [Hacking, 1990]). It has been credited to Poisson that Laplace's ideas spread so quickly and widely in France and other parts of Europe.

Adoplphe Quetelet was another significant apprentice of Laplace. His *Social Physics* was essentially based on Laplace's idea of social phenomena being analogous to natural ones and thus measurable. Laplace's influence can be seen in the expectancy that all random fluctuations in nature and in society could be treated correspondingly to a pattern of errors in observations.

An interesting nuance in the history of statistics is that approximately 25 years after Laplace, Quetelet wanted to try a similar estimation of the size of the population in the Low Countries as Laplace had carried out in France. After Quetelet had published his plans, baron de Keverberg objected them [De Keverberg, 1827]. De Keverberg was afraid that the sample could never reach full "representativeness" because of the fundamental heterogeneity of the population. Quetelet accepted de Keverberg's argument about the lack of homogeneity and lost interest in partial investigations.

In 1849, Quetelet published a book called "*Letters on Probability"* [Quetelet, 1849] in which he outlined the use of probability in statistical research. The probability analysis of Quetelet was based on Laplace's and Poisson's ideas. Quetelet's "*Letters on Probabilities"*, especially the law of error, later inspired Francis Galton and he started to call the law of error as the Normal Distribution.

German statistician Wilhelm Lexis was a noteworthy apprentice of Quetelet's. He is best known because of his pioneering work on dispersion. Using an urn model the same way Laplace had done Lexis derived a dispersion coefficient Q (reportedly in homage to Quetelet), which is the ratio of the empirical variance of the series to the assumed theoretical variance. Lexis' most important contribution to statistical social science was the method to assess the stability of statistical series [see e.g. Lexis, 1879] but his work also had a significant impact on the development of statistical methods.

Lexis had strong and long-lasting influence in the development of statistics in Russia. In England, Lexis had the greatest influence on Edgeworth. Lexis' analysis of dispersion has also been claimed to foreshadow the statistics of Karl Pearson and even R.A. Fisher's analysis of variance.

The development, which started from the publication of Central Limit Theorem in 1810, was concluded a century later by the contributions of Edgeworth [Edgeworth, 1906, 1908 and 1909]. In these contributions, he completed the large-sample theory of statistical inference by inverse probability. Since Laplace, Edgewortht's contributions in inverse probability theory were probably the most significant.

In 1926, Arthur Bowley published a well-thought theory for statistical inference for finite populations [Bowley, 1926]. Bowley leaned on Edgeworth's contributions, especially on his version of the Central Limit Theorem, but the theoretical framework was Laplacian. Possibly, it is the last significant contribution that was an offspring of the Laplacian paradigm. The adherence to this paradigm may also be the reason why Bowley's paper fell into oblivion – deemed erroneously as 'Bayesian'.

Jeffreys fostered and upheld the Bayes–Laplace theory – as he called it – still in the late 1930s and even later [e.g. Jeffreys, 1983], but obviously, he did not gain wider acceptance, anymore.

9 Discussion and conclusions

In 1774, Laplace presented the idea of inverse probability in a memoir titled "*Mémoire sur la probabilités des causes par les évènemens*". The influence of this piece of work was immense. It was from this memoir that the ideas, now called "Bayesian", started to spread through the mathematical world [Stigler, 1986]. Obviously, Laplace's 1774 memoir is one of the revolutionary papers in the history of statistical inference (see also [Hald, 1998]).

Laplace is also a pioneer of sample surveys: In 1802, he undertook first scientifically ambitious partial investigation, or sample survey. Its purpose was to estimate the size of population of France. For the survey, he developed the method for statistical inference applying his Principle of inverse inference. In addition, the plan included the calculation of the sample size that was needed to reach the predefined accuracy, and the method to estimate the accuracy of estimates. His theory is essentially correct for simple random sampling, even though his model did not correspond to this mode of sampling.

Apart from the method of inverse inference, Laplace developed many of the central ideas in probability theory and statistical science. Weatherford [1982] claims that the classical theory of probability reached its zenith in the work of Laplace and that he solved more problems and developed more important mathematical tools than any of his predecessors. These methods opened a new era in the development of probability theory and its application to empirical sciences. His contributions were so influential that they dominated statistical thinking nearly for a century.

9.1 Was Laplace aware of Bayes' essay?

There is little evidence available on how Laplace originally derived the general theory of inverse probability. He was only 25 when he published the PCE. The greatest motivation for his work, at least in the beginning, may have been the problem of merging discrepant observations in astronomy. Primarily, Laplace was an astronomer.

In 1764, William Price, a friend of Thomas Bayes, read an essay that Bayes had written [Bayes, 1763 or Bayes, 1958] to the Royall Society. Bayes had died few years earlier and Price had found the Essay among Bayes' papers.

The first version of Laplace's *Principle* and Bayes' method resemble each other but they are not the same. The later version of the *Principle,* published in 1814 in the second edition of *Théorie Analytique des Probabilités*, is the same as Bayes method, though. Some writers, e.g., Pearson [1920] and Fisher [1930], claim that Laplace had copied the *Principle* from Bayes. On the other hand, Stigler [1982] claims that Bayes' Essay was ignored until after 1780 and it played no important role in the scientific debate until the 20th century.

In the introduction to the memoir, Laplace says:

> "I propose to determine the probability of the causes of events, a question which has not received due consideration before, but which deserves even more to be studied, for it is principally from this point of view that the science of chances can be useful in civil life." [Laplace, 1774]

An obvious conclusion from this citation is that Laplace was not aware of Bayes' Essay. Stigler [1986] claims that there are many reasons why it is reasonably certain that Laplace was unaware of Bayes' earlier work. Comparing Bayes' Essay and Laplace's PCE it is easy to agree with Hald [1998] who argued that Bayes' and Laplace's theories are conceptually and mathematically so different that they cannot be related.

In deriving his method, Bayes applied Newtonian geometric reasoning and there are only few formulas in the Essay. The famous Bayes' formula cannot be found in it. The modern forms of the Bayes' Theorem have been given long time after his death. According to Fienberg (2006) the modern formula was introduces in the beginning of 20th century. In addition, Bayes' inference model was not a Bernoulli trial. Instead, the model was based on rolling balls on a (billiard) table. Bayes' method was an extension conditional probability.

Todhunter [1865] is the first comprehensive account on the history of probability. There he says:

> "This memoir [Laplace 1774] is remarkable in the history of the subject, as being the first which distinctly enunciated the principle of estimating the probabilities of the causes by which an observed event may be produced." [Todhunter, 1865]

Todhunter considers this memoir as the first contribution on inverse probability, and not that of Bayes'.

In the 19th century, Bayes' Essay had only minor importance. It is briefly mentioned only in few statistical texts, whereas references to Laplace's works are frequent. Typically, Todhunter [ibid.] only briefly and superficially mentions Bayes (5 pages) whereas the part which deals with Laplace's contributions, covers nearly one quarter of the book (c. 150 pages).

Todhunter [ibid.] concludes "…on the whole the Theory of probability is more indebted to him [Laplace] than to any other mathematician." Indirectly, the citations of Todhunter indicate that Bayes' Essay was nearly unknown to the 19th century scientific world. In the statistical literature Bayes appeared only in the 20th century [see Fienberg 2006].

An apparent conclusion is that Laplace's two early memoirs [Laplace, 1774 and 1781] were original and the first significant contributions on inverse probability.

9.3 Revolutions in inference theory

Anders Hald begins his book about the history of parametric statistical inference [Hald, 2007] by saying: "The three revolutions in parametric statistical inference are due to Laplace, Laplace and Gauss (1809-1811), and Fisher." Hald observed that the first revolution in statistical inference, due to Laplace, took place between 1774 and 1786 when Laplace turned his attention from direct probability and derived his Principle of Inverse Probability.

The second revolution that Hald [ibid.] indentified took place in 1809-1828 when Gauss and Laplace, with the help of the Principle of Inverse Probability, discovered the Central Limit Theorem and the method of least squares. Hald [ibid.] concluded that the second revolution was concluded by the contributions of Edgeworth [Edgeworth, 1906, 1908 and 1909] in which he completed the large-sample theory of statistical inference by inverse probability initiated by Laplace and generalises the Central Limit Theorem.

In the 1920s, R. A. Fisher initiated the third revolution in statistical inference that Hald [2007] recognises. Fisher pioneered a recasting of statistics, moving away from the reliance on large sample approximations and on inverse probability. Fisher focused on the use of small samples and finding causes by designing experiments.

Before Fisher's contributions, the prevailing idea of statistical inference was Laplace's inverse probability principle, and its validity was not questioned. Indirectly, this can be concluded from the texts of the leading statisticians in the beginning of the 20th century. Apparently, they all worked from the Laplacian paradigm – even though they usually are (erroneously) depicted "Bayesians".

In 1920, Karl Pearson described the "*Fundamental problem of practical statistics*":

> "…The problem I refer to is that of 'inverse probabilities' and in practical statistics it takes the following form:
>
> An 'event' has occurred p times out of $p + q = n$ trials, where we have no *a priori* knowledge of the frequency of the event in the total population of occurrences. What is the probability of its occurring r times in a further $r + s = m$ trials?
>
> In statistical language, a first sample of n shows p marked individuals, and we require a measure of the accordance which future samples are likely to give with this result. For example, a medical treatment is found to be successful in p out of n cases, we require some measure of the probable stability of this ratio. It is on this stability and its limits according to the size of the first sample that the whole practical doctrine of statistics, which is the theory of sampling, actually depends. We usually state the "probable errors" of results without visualising the strength or weakness of the logic behind them, and without generally realising that if the views of some authors be correct our superstructure is built, if not on a quicksand, at least in uncomfortable nearness to an abyss." [Pearson, 1920]

This problem is close to the one that Laplace presented in his *Memoir on the Probability of the Causes of Events*. Inference is seen as a method that is based on learning from experience and a key element was the assumption of the stability of statistical ratios.

A few years later, Arthur Bowley presented a related line of reasoning while analysing inferences in the context of survey sampling [Bowley, 1923].

Obviously still in the beginning of the 20th century, inverse probability had been a part of mathematicians' training. Maybe the most assuring indication is what R.A. Fisher wrote in 1936:

> "…In the latter half of the nineteenth century the theory of inverse probability was rejected more decisively [than Boole] by Venn and by Chrystal, but so retentive is the tradition of mathematical

> teaching that I may myself say that I learned it at school as an integral part of the subject, and for some years saw no reason to question its validity." [Fisher, 1936]

Erich Lehmann, a friend and colleague of Neyman's at the University of California at Berkeley, wrote in 2008 that the years 1925-26 were difficult for Neyman and Egon Pearson. They began to realize that Fisher's work required rethinking the current philosophy of inference. According to Lehmann, this was exceptionally difficult for Egon Pearson because his father, Karl Pearson, "was not able or never saw the need to" make such a shift [Lehmann, 2008].

9.4 Emergence of the new inference theory

In 1922, R. A. Fisher published a general theory of estimation in the paper titled *On the Mathematical Foundations of Theoretical Statistics* [Fisher, 1922] that can be regarded as a watershed in statistical science. In this paper, Fisher introduced a totally new set of concepts and new thinking in statistical science. In addition, he created a new terminology for statistical science that gained recognition quickly when the new generation of statisticians adopted it in late the 1920s. It has been dominant ever since. Fisher also introduced a new inference model: repeated sampling from the same distribution that yielding a sampling distribution for a statistics. This was the first time in the history of statistical science when a framework for frequency-based theory of parametric statistical inference was clearly formulated.

Nearly from the very beginning of his career, Fisher had attacked the principle of inverse probability and said that it was the greatest flaw in modern science. In addition, all his comments about Laplace were critical, belittling, or sarcastic. In 1930s, Fisher presented his famous *fiducial argument* to replace the inverse probability principle [Fisher, 1930]. In this context, he also defined a new mode of statistical inference, which he called inductive reasoning.

Jerzy Neyman is another main architect of the frequency-based statistical inference. However, he developed inference theory for both hypothetical and finite populations. Neyman adopted Fisher's inference model of drawing repeated samples, and applied it also in sampling from finite populations. An interesting detail is that Neyman he was trained within the Laplacian paradigm as can be concluded from his early contributions [see Spława-Neyman, 1923 and 1925]. In the late 1920s, while visiting the UK, he began to realize that the work of Fisher required a rethinking of his philosophy of inference [Lehmann, 2008]. As a result, Neyman adopted the new Fisherian approach to statistical inference (see also [Kuusela, 2011]) and eventually applied it in inference both for infinite and finite populations. Only after Neyman's 1934 paper, the two modes of statistical inference were separated. Currently, Fisher's and Neyman's inference model is the corner stone of modern inference theory for both hypothetical and finite populations.

Fisher's contributions revolutionized almost every part of statistical science, especially the theory of estimation and statistical inference. It replaced the Laplacian paradigm in a relatively short time. At first sight, Fisher's revolution seems to put statistical inference into a totally new form. However, as Hald [2007] points out: ". . . many of Fisher's asymptotic results are identical to those of Laplace from a mathematical point of view, only a new interpretation is required."

References

Bayes, T. (1763),”An essay towards solving a problem in the doctrine of chances,” *Philosophical Transactions of the Royal Society*, 53, 370-418.

Bayes, T. (1958),”An Essay Towards Solving a Problem in the Doctrine of Chances (with biographical note by G.A. Barnard),” *Biometrika*, 45, 293-315.

Bowley, A. L. (1923),”The precision of measurements estimated from samples,” *Metron*, 2, 494-500.

Bowley, A. L. (1926),”Measurement of the precision attained in sampling,” *Bulletin of the International Statistical Institute*, 22, 6-62.

Brewer, K. R. W. (1963),”Ratio estimation and finite populations: Some results deducible from the assumption of an underlying stochastic process,” *Australian Journal of Statistics*, 5, 93-105.

Bru, B. (1988), ”Estimations Laplaciennes. Un exemple: la recherche de la population d’un grand Empire, 1785-1812 ,” in *Estimation et sondages. Cinq contributions à l’histoire de la statistique*, ed. Mairesse, J., Paris: Economica, pp. 7-46.

Bru, B. (2001), ”Siméon-Denis Poisson,” in *Statisticians of the Centuries*, eds. C. C. Heyde and E. Seneta: Springer, pp. 123-126.

Cochran, W.G. (1953), *Sampling Techniques*, New York, John Wiley & Sons.

Cochran, W. G. (1978), ”Laplace’s Ratio Estimator,” in *Contributions to Survey Sampling and Applied Statistics*, ed. H. A. David, New York: Academic Press, pp. 3-10.

Dale, A. I. (1999), *The History of Inverse Probability (2nd ed.)*, New York: Springer-Verlag.

Egdeworth, F. Y. (1906),”The Generalized Law of Error, or Law of Great Numbers,” *Journal of the Royal Statistical Society*, 69, 497-539.

Egdeworth, F. Y. (1908),”On the Probable Errors of Frequency-Constants,” *Journal of the Royal Statistical Society*, 71, 381-397.

Egdeworth, F. Y. (1909),”Addendum on”Probable Errors of Frequency-Constants”,” *Journal of the Royal Statistical Society*, 72, 81-90.

Fienberg, S. E. (2006), ”When Did Bayesian Inference Become ”Bayesian”?,” *Bayesian Analysis*, 2006, 1-40.

Fischer, H. (2010), *A History of the Central Limit Theorem; From Laplace to Donsker*, Heidelberg: Springer.

Fisher, R. A. (1922),”On the Mathematical Foundations of Theoretical Statistics,” *Philosophical Transactions of the Royal Society, A*, 222, 309-368.

Fisher, R. A. (1925a),”Theory of Statistical Estimation,” *Proceedings of Cambridge Philosophical Society*, 22, 700-725.

Fisher, R. A. (1925b), *Statistical Methods for Research Workers*, Edinburgh: Oliver and Boyd.

Fisher, R. A. (1930),”Inverse Inference,” *Proceedings of Cambridge Philosophical Society*, 26, 528-535.

Fisher, R. A. (1935),"The Logic of Inductive Inference," *Journal of the Royal Statistical Society*, 98, 39-82.

Graunt, J. (1662), *Natural and Political Observations upon the Bills of Mortality*, London: John Martyn.

Hacking, I. (1975), *The Emergence of Probability*, Cambridge: Cambridge University Press.

Hacking, I. (1990), *The Taming of Chance*, Cambridge: Cambridge University Press.

Hald, A. (1998), *A History of Mathematical Statistics from 1750 to 1930*: John Wiley & Sons.

Hald, A. (1999),"On the History of Maximum Likelihood in Relation to Inverse Probability and Least Squares," *Statistical Science*, 14, 214-222.

Hald, A. (2007), *A History of Parametric Statistical Inference from Bernoulli to Fisher, 1713-1935*, New York: Springer.

Hume, D. (1748), *An Enquiry Concerning Human Understanding*, London. Reprinted by Clarendon Press, Oxford by comments Selby-Bigge, L. A. (2nd edition 1902). Published in 1748 under the title *Philosophical Essay Concerning Human Understanding*.

Jeffreys, H. (1983), *Theory of Probability,* 3rd edition (1st edition 1939), London: Clarendon Press.

Keverberg, Baron de (1827), "Notes sur Quetelet," *Nouveaux Mémoires de l'Academie royal des sciences et belles-lettres de Bruxelles,* 4, 175-192

Kuusela, V. (2011), *Paradigms in Statistical Inference for Finite Populations ; up to the 1950s*, Helsinki : Statistics Finland Research Reports 257. (Also available at the University of Helsinki electronic archives https://helda.helsinki.fi/handle/10138/27416?locale-attribute=en).

Lambert, J. (1760), *Photometria sive de Mensura et Gradibus Luminis, Colorum et Umrae,* Ausburg : Detleffsen.

Laplace, P. S. (1774), "Mémoire sur la probabilité des causes par les événements," *Mémoires de l'Académie des Sciences de Paris*, 6, 621-656.

Laplace, P. S. (1778), "Mémoire sur les probabilités," in *Mémoires de l'Académie Royale des Sciences de Paris,* Paris, pp. 1-98.

Laplace, P. S. (1783), "Sur les naissances, les mariages et les morts," *Mémoires de l'Académie Royale des Sciences de Paris,* 693-702.

Laplace, P. S. (1810), "Mémoire sur les approximations des formules qui sont fonctions de très grands nombres et sur leur application aux probabilités," *Mémoires de l'Académie Royale des Sciences de Paris,* 301-345.

Laplace, P. S. (1812a), *Théorie analytique des probabilités*, Paris: Gauthier-Villar.

Laplace, P. S. (1812b), *Essai philosophique sur les probabilités*, Paris: Bougois.

Lehmann, E. L. (2008), *Reminiscences of a Statistician*, New York: Springer.

Lexis, W. (1877), *Zur Theorie der Massenerscheinungen in der meschlichen Geselschaft*, Freiburg: Fr. Wagner'sche Buchhandlung.

Lexis, W. (1879), "Über die Theorie der Stabilität statistischer Reihen," in *Jahrbücher für Nationalökonomie und Statistik* (Vol. 32), pp. 60-98.

Neyman, J., and Pearson, E. (1933), ”On the Problem of the most Efficient Test of Statistical Hypotheses,” *Philosophical Transactions of the Royal Society, A*, 289-337.

Neyman, J. (1934), ”On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection (with discussion),” *Journal of the Royal Statistical Society*, 97, 558-606.

Pearson, K. (1920), ”The Fundamental Problem of Practical Statistics,” *Biometrika*, 13.

Pearson, K. (1928), ”On a Method Ascertaining Limits to the Actual Number of Marked Members in a Population of Given Size from a Sample,” *Biometrika*, 20A, 149-174.

Pearson, K. (1978), ”The history of statistics in the 17th and 18th centuries against the changing background of intellectual, scientific and religious though,” in *Lectures by Karl Pearson given at the University College London during the academic sessions 1921 - 1933*, ed. E. Pearson, London: Charles Griffin & Co.

Pólya, G. (1920), ” Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem,” *Mathematische Zeitschrift*, 8, 171-181.

Quetelet, A. (1849), *Lettres à S.A.R. le duc régnant de Saxe-Coburg et Gotha, sur la théorie des probabilités, appliquée aux sciences morales et politiques s*, Bruxelles: Reprinted by Arno Company *(Letters addressed to H.R.H. the Grand Duke of Saxe Coburg and Gotha, On the theory of probabilities, as applied to the moral and political science)*, New York 1981.

Rietz, H. L. (1924), ”On Certain Topics in the Mathematical Theory of Statistics,” *Bulletin of the American Mathematical Society*, 417-453.

Simpson, T. (1755), ”On the Advantage of Taking the Mean of a Number of Observations in practical Astronomy,” *Philosophical Transactions of the Royal Society*, 49, 82-93.

Spława-Neyman, J. (1925), ”Contributions of the theory of small samples drawn from a finite population,” *Biometrika*, 17, 472-479.

Spława-Neyman, J. (1923), ”On the application of probability theory to agricultural experiments: Essay on principles. (Translated and published in English 1990 in Statistical Science),” *Statistical Science*, 5, 465-472.

Stigler, S. (1973), ”Studies in the History of Probability and Statistics. XXXII; Laplace, Fisher, and the discovery of the concept of sufficiency,” *Biometrika*, 60, 439-445.

Stigler, S. (1982), ”Thomas Bayes´s Bayesian Inference,” *Journal of the Royal Statistical Society, Ser. A*, 145, 250-258.

Stigler, S. (1986), *The History of statistics; the Measurement of Uncertainty before 1900*, Cambridge: The Belknap Press of Harvard University Press.

Stigler, S. (1986), ”Laplace's 1774 memoir on inverse probability,” *Statistical Science*, 1, 359-378.

Thatcher, A. R. (1964), ”Relationship between Bayesian and confidence limits for prediction,” *Journal of the Royal Statistical Society, Ser. B*, 26, 176-192.

Todhunter, I. (1865), *A History of the Mathematical Theory of Probability; From the time of Pascal to that of Laplace*, Cambridge and London: Macmillan and Co. (Reprinted by Chelsea, New York, 1965.)

Weatherford, T. (1982), *Philosophical Foundations of Probability Theory*, London: Routledge & Kegan Paul.