Statistical Science
1986, Vol. 1, No. 3, 359–378

Laplace's 1774 Memoir on Inverse Probability

Stephen M. Stigler

Abstract. Laplace's first major article on mathematical statistics was published in 1774. It is arguably the most influential article in this field to appear before 1800, being the first widely read presentation of inverse probability and its application to both binomial and location parameter estimation. After a brief introduction, an English translation of this epochal memoir is given.

Key words and phrases: History, Bayesian, posterior distribution, predictive distribution, optimum estimation, nuisance parameter, double exponential distribution.

INTRODUCTION

The history of mathematical statistics before 1800 boasts several landmark treatises, such as Jacob Bernoulli's *Ars Conjectandi* (1713) and Abraham De Moivre's *Doctrine of Chances* (1718, second edition 1738), and a smaller number of important articles in the periodical literature. Of these latter, perhaps the single most influential was Pierre Simon Laplace's "Mémoire sur la probabilité des causes par les évènemens," published in 1774. Laplace was just 25 years old when this appeared, and it was his first substantial work in mathematical statistics. At the time he began this work in 1772, he was in a period of intensely creative scientific exploration, simultaneously making major advances in mathematics and in mathematical astronomy (Stigler, 1978), and the memoir is an explosion of ideas that left an indelible imprint on statistics. In this one article, we can recognize the roots of modern decision theory, Bayesian inference with nuisance parameters, and the asymptotic approximation of posterior distributions.

A full evaluation of Laplace's contributions in this memoir in an historical context is beyond the scope of this introduction. Some aspects of the memoir have been discussed by Todhunter (1865, pages 465–473), Gillispie (1981), Sheynin (1977), and Dale (1982), and further elucidation will be found in my book (Stigler, 1986, Chapter 3). Nonetheless, and despite the fact that the memoir reads clearly and smoothly (even after two centuries it seems like a contemporary work), a brief outline in modern terminology will be useful for most readers.

After some opening references in Section I to work of his own (Laplace, 1774a, 1776) and of Lagrange (1759) on the solution of difference equations (particularly such as arise in probability theory), Laplace goes on in Section II to treat the inference problems that are the central focus of the article. He clearly announces his first goal as that of determining an unknown binomial probability, given the outcome of $p + q$ trials, of which p result in white tickets, q in black. He states a Principle which we would now recognize as equivalent to Bayes's theorem with all causes being a priori equally likely. If F is Laplace's "event" and $\theta_1, \theta_2, \ldots, \theta_n$ the n causes, then his axiomatic "Principle" is:

$$\frac{P(\theta_i \mid F)}{P(\theta_j \mid F)} = \frac{P(F \mid \theta_i)}{P(F \mid \theta_j)}$$

and

$$P(\theta_i \mid F) = \frac{P(F \mid \theta_i)}{\sum_{j=1}^{n} P(F \mid \theta_j)}.$$

Some reasons why we can be reasonably certain Laplace was unaware of Bayes's earlier work can be found in Stigler (1978).

Indeed, one of the more compelling reasons is the very different approach taken: where Bayes gives a cogent argument why an a priori uniform distribution might be acceptable (Stigler, 1982), Laplace assumes the conclusion as an intuitively obvious axiom. Laplace gives one simple application of his Principle for the case of hypergeometric sampling from one of two

Stephen M. Stigler is a Professor in the Department of Statistics, University of Chicago, 5734 University Avenue, Chicago, Illinois 60637.

possible urns, A and B, evaluating

$$K = P(\text{draw } f \text{ white, } h \text{ black} \mid \text{urn } A),$$

and finding

$$P(\text{urn } A \mid \text{draw } f \text{ white, } h \text{ black}) = K/(K + K')$$

(where $K' = P(\text{draw } f \text{ white, } h \text{ black} \mid \text{urn } B)$), the correct result if A and B are a priori equally likely.

In Section III, Laplace applies the Principle to the case of binomial sampling, effectively assuming a uniform prior distribution for the parameter. In Problem I, he derives the beta posterior of the binomial parameter x and also the predictive distribution (Laplace's E) for $m + n$ future trials given $p + q$ trials have occurred, first for $m = 1$ and $n = 0$, then for the general case. He applies Stirling's formula in a form he obtained from a text of Leonhard Euler's to derive a large sample approximation to the latter. He does his asymptotics in two different ways. First, he supposes $p + q$ is large (while $m + n$ is not), and he concludes that the predictive distribution is approximately equal to the one found by simply taking x equal to the sample fraction, $p/(p + q)$. He then shows that a very different conclusion holds if $m + n$ is large also, say $m + n = p + q$. The remainder of the section is taken up with an analysis of the posterior distribution for large samples. He states (in a surprisingly modern δ-ε form) a theorem that claims the posterior consistency of the relative frequency of successes, but his proof is a tour de force that does more: it both introduces Laplace's own method for the asymptotic approximation of integrals, by expanding the integrand about its maximum at $x = p/(p + q)$, and it effectively demonstrates the asymptotic normality of the beta posterior distribution, as he approximates the posterior probability

$$E = P(\,|\,x - p/(p + q)\,| < w \mid p \text{ white tickets, } q \text{ black})$$

by a normal integral, and then takes the limit as $p + q \to \infty$, with $w = (p + q)^{-1/n}$, $2 < n < 3$. Since the proof includes what seems to be the earliest evaluation of the definite integral

$$2 \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{z^2}{2\sigma^2}\right) dz = 1,$$

with $\sigma^2 = pq/(p + q)^3$ here, Laplace may have been the first to integrate the normal density. In 1809, Gauss was to refer to this evaluation as "an elegant theorem of Laplace." The section ends with an intricate attempt to approximate the error made in taking $E = 1$.

Section IV is given over to Problem II, an analysis of the classical "problem of points" from a Bayesian perspective. We might now describe this as finding the expectation of the exact predictive distribution for binomial sampling, with a uniform prior distribution.

In Section V, Laplace moves on to the estimation of a location parameter for the "simple" case of three observations. He was evidently spurred on by reading a brief footnote in a 1772 review by Jean Bernoulli III: "The problem of finding the true mean among several observations, which is rarely the arithmetic mean, is of considerable interest to astronomers." Bernoulli cited works by Boscovich and Lambert that had been published, and by Lagrange and Daniel Bernoulli that were only to appear in 1776 and 1778. (The entire footnote is reproduced, with references, in Stigler (1978, page 248).) Laplace begins his treatment of his problem ("Problem III") by speculating upon the nature of the error distribution and finding the general expression for the posterior distribution of what we would now call the location parameter. Figure 1 shows the three observations at a, b, and c, and Figure 2 shows the error distribution. V represents the true value of the location parameter, and p and q the gaps between a and b and b and c, respectively. We might now write $a = X^{(1)}$, $b = X^{(2)}$, $c = X^{(3)}$, $p = X^{(2)} - X^{(1)}$, $q = X^{(3)} - X^{(2)}$, and note that p and q are ancillary statistics. Laplace lets x be the distance from V to a, and Figure 1 gives a stylized version of the posterior distribution of x given p and q, the curve *HOL*. Laplace suggests both the posterior median ("the mean of probability") and the value that minimizes the posterior expected loss ("the mean of error") as a posterior estimate of the location parameter, and proves (with the aid of Figure 3) that these are always the same. This result (characterizing the posterior median as optimal for a certain loss function) is surely one of the earliest we can recognize as truly belonging to mathematical statistics, rather than probability theory. He then returns to the question of specifying the error distribution and presents an argument for the double exponential density,

$$\phi(x) = \frac{m}{2} e^{-m|x|}$$

He finds an explicit expression for the posterior median in this case, assuming the scale parameter m is known, and he shows it differs from the commonly used arithmetic mean, as Jean Bernoulli III had said.

The remainder of Section V is concerned with the case where the scale parameter m is unknown but a priori uniformly distributed. His analysis here is intricate, the more so because he makes a subtle error in finding an explicit expression for the posterior median. The crucial passages are in the two consecutive paragraphs that follow his correct derivation of the posterior distribution of m, the two paragraphs that begin "Next, if we denote by y ..." Since the nature and importance of this error seem to have escaped other commentators, it is worth detailed comment. To help clarify the argument, I shall employ the symbol f as generic notation for density, so for example $f(x, m \mid p, q)$ is the conditional joint density of x and m given p and q. Recall that Laplace has framed his problem in terms of the correction, x, to be made to the first observation, a, and he wishes the posterior median of x (which I shall denote x_0). Now, earlier, when he supposed m known, Laplace had correctly given the equation that would determine the posterior median x_0 as, essentially,

$$\int_{-\infty}^{x_0} f(x, p, q \mid m)\, dx = \frac{1}{2} \int_{-\infty}^{\infty} f(x, p, q \mid m)\, dx,$$

which, since this joint density is proportional to the conditional density, $f(x \mid p, q, m)$, with constant of proportionality $[\int f(x, p, q \mid m)\, dx]^{-1}$, is equivalent to

$$\int_{-\infty}^{x_0} f(x \mid p, q, m)\, dx = \frac{1}{2} \int_{-\infty}^{\infty} f(x \mid p, q, m)\, dx.$$

Here, when m is unknown, he has derived $f(m \mid p, q)$ under the assumption that m is a priori uniformly distributed on $(0, \infty)$, as

$$f(m \mid p, q) \propto m^2 e^{-m(p+q)}(1 - \tfrac{1}{3}e^{-mp} - \tfrac{1}{3}e^{-mq}).$$

Now his $y = f(x \mid p, q, m)$, and Laplace "evidently" wants to proceed as follows:

We want x_0 so that

$$\int_{-\infty}^{x_0} f(x \mid p, q)\, dx = \frac{1}{2} \int_{-\infty}^{\infty} f(x \mid p, q)\, dx,$$

or equivalently,

$$\int_0^{\infty} \int_{-\infty}^{x_0} f(x, m \mid p, q)\, dx\, dm = \frac{1}{2} \int_0^{\infty} \int_{-\infty}^{\infty} f(x, m \mid p, q)\, dx\, dm,$$

or equivalently,

$$\int_0^{\infty} \int_{-\infty}^{x_0} f(x \mid m, p, q) f(m \mid p, q)\, dx\, dm = \frac{1}{2} \int_0^{\infty} \int_{-\infty}^{\infty} f(x \mid m, p, q) f(m \mid p, q)\, dx\, dm.$$

Now if this last equation were equivalent to

$$\int_0^{\infty} \int_{-\infty}^{x_0} f(x, p, q \mid m) f(m \mid p, q)\, dx\, dm = \frac{1}{2} \int_0^{\infty} \int_{-\infty}^{\infty} f(x, p, q \mid m) f(m \mid p, q)\, dx\, dm,$$

we would indeed have Laplace's solution. This would be true if

$$f(x \mid m, p, q) \propto f(x, p, q \mid m)$$

and the constant of proportionality did not depend on m, but here it does! In fact,

$$\begin{aligned} f(x \mid m, p, q) &= f(x, p, q \mid m)/f(p, q \mid m) \\ &\propto f(x, p, q \mid m)/f(m \mid p, q), \end{aligned}$$

for Laplace's prior, so the correct solution would have been

$$\int_0^{\infty} \int_{-\infty}^{x_0} f(x, p, q \mid m)\, dx\, dm = \frac{1}{2} \int_0^{\infty} \int_{-\infty}^{\infty} f(x, p, q \mid m)\, dx\, dm,$$

an even simpler pair of integrals than Laplace considered. Because of his error, Laplace is led to finding the root of a 15th degree equation, which he does iteratively, and presents the solution in the form of a table. No wonder he only considered the case of three observations! Laplace's error is important, as it sheds light on how he conceived of conditional distributions as only defined up to proportionality, and helps explain how his axiomatic "Principle" could have such intuitive appeal to him (Stigler, 1986, Chapter 3).

Finally, Section VI explores the effect of allowing a probability to have an a priori distribution in two classical settings: coin tossing and the rolling of a die. This work departs from the earlier sections of the paper in allowing a nonuniform prior; Laplace effectively "unsharpens" the sharp null hypothesis of a fair coin by first permitting the probability to be $\tfrac{1}{2}(1 - \pi)$ rather than $\tfrac{1}{2}$, and then allowing π a uniform distribution over $[0, q^{-1}]$. He notes that the results (the expectation of the predictive distribution, in particular) can be dramatically different for composite events from those where a fair coin is assumed. His analysis for the die is similar, although more complex. It is interesting that he calls unbalanced dice "English Dice"; presumably they had a different nickname in London. He ends with some explicit derivations for the case of a three-sided die, concluding what must be the earliest application of Bayesian ideas to a multinomial setting.

The influence of this memoir was immense. It was from here that "Bayesian" ideas first spread through the mathematical world, as Bayes's own article (Bayes, 1764) was ignored until after 1780 and played no important role in scientific debate until the twentieth century (Stigler, 1982). It was also this article of Laplace's that introduced the mathematical techniques for the asymptotic analysis of posterior distributions that are still employed today. And it was here that the earliest example of optimum estimation can be found, the derivation and characterization of an estimator that minimized a particular measure of posterior expected loss. After more than two centuries, we mathematical statisticians cannot only recognize our roots in this masterpiece of our science, we can still learn from it.

ACKNOWLEDGMENTS

Support for this research was provided in part by National Science Foundation Grant MCS-8301459. This manuscript was prepared using computer facilities supported in part by National Science Foundation Grant MCS-8404941 to the Department of Statistics at the University of Chicago.

REFERENCES

BAYES, T. (1764). An essay towards solving a problem in the doctrine of chances. *Philos. Trans. Roy. Soc. London* **53** 370–418. (Reprinted, with an introduction by George Barnard, in 1958 in *Biometrika* **45** 293–315, and in 1970 in *Studies in the History of Statistics and Probability* (E. S. Pearson and M. G. Kendall, eds.) 131–153. Hafner, Darien, Conn.)

DALE, A. I. (1982). Bayes or Laplace? An examination of the origin and early applications of Bayes's theorem. *Arch. History Exact Sci.* **27** 23–47.

GILLISPIE, C. C. (1981). Laplace, Pierre Simon. *Dictionary of Scientific Biography* **15,** 273–403. Scribner, New York.

LAGRANGE, J. L. (1759). Sur l'integration d'une équation différentielle à différences finies, qui contient la théorie des suites récurrentes. *Miscellanea Philos. Math. Soc. Privatae Taurinenesis* **1** 33–42.

LAPLACE, P. S. (1774a). Mémoire sur les suites récurro-récurrentes et sur leurs usages dans la théorie des hasards. *Mémoires de mathématique et de physique presentés à l'Académie royale des sciences, par divers savans, & lûs dans ses assemblées* **6** 353–371. (Reprinted in Laplace's *Oeuvres complètes* **8** 5–24.)

LAPLACE, P. S. (1774b). Mémoire sur la probabilité des causes par les évènemens. *Mémoires de mathématique et de physique presentés à l'Académie royale des sciences, par divers savans, & lûs dans ses assemblées* **6** 621–656. (Reprinted in Laplace's *Oeuvres complètes* **8** 27–65.)

LAPLACE, P. S. (1775). Mémoire, sur les solutions particulières des équations différentielles, & sur les inégalités séculaires des plànetes. *Histoire de l'Académie royale des sciences, année 1772, première partie, avec les Mémoires de mathématique et de physique, pour la même année, tirés des registres de cette académie* (published 1775) 343–377. (Reprinted in Laplace's *Oeuvres complètes* **8** 325–366.)

LAPLACE, P. S. (1776). Recherches, sur l'intégration des équations différentielles aux différence finies, & sur leur usage dans la théorie des hasards. *Mémoires de mathématique et de physique presentés à l'Académie royale des sciences, par divers savans, & lûs dans ses assemblées, 1773* (published 1776) 37–162. (Reprinted in Laplace's *Oeuvres complètes* **8** 69–197.)

LAPLACE, P. S. (1886–1912). *Oeuvres complètes de Laplace.* 14 volumes. Gauthier-Villars, Paris.

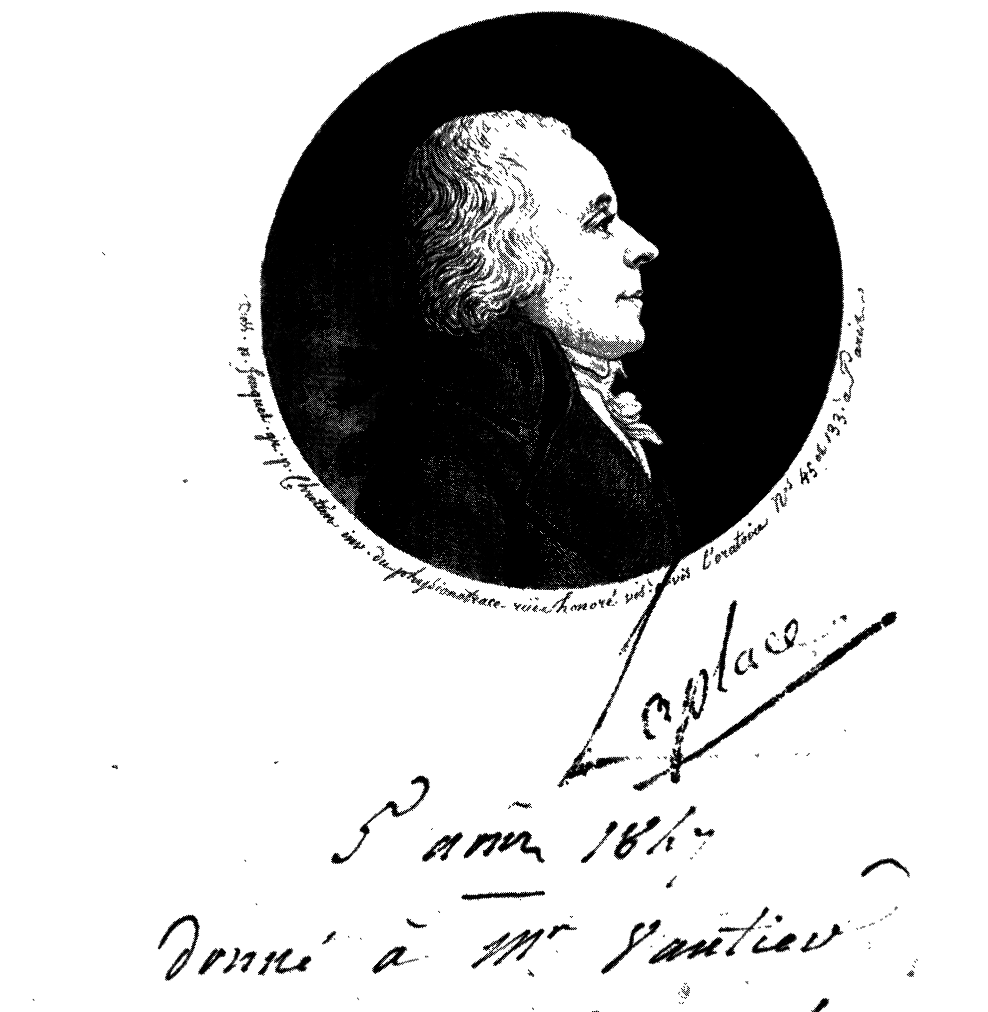
SHEYNIN, O. B. (1977). Laplace's theory of errors. *Arch. History Exact Sci.* **17** 1–61.

STIGLER, S. M. (1978). Laplace's early work: Chronology and citations. *Isis* **61** 234–254.

STIGLER, S. M. (1982). Thomas Bayes's Bayesian inference. *J. Roy. Statist. Soc. Ser. A* **143** 250–258.

STIGLER, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty Before 1900.* Harvard Univ. Press, Cambridge, Mass.

TODHUNTER, I. (1865). *A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace.* Macmillan, London. (Reprinted by Chelsea, New York, in 1949, 1965.)



This lithograph of Laplace bears his printed signature and an 1847 *autograph inscription by his widow. Laplace died* 20 *years earlier, in* 1827, *and the inscription ("Given to Mr. Vautier by Madame La Marquise de Laplace") suggests that it was given in response to a request for a remembrance. The face it shows is quite different from that shown in the stern formal pose that is usually reproduced, but the association of this picture with his widow suggests that it shows a likeness of Laplace recognizable to those who knew him well. The lithograph had been printed in* 1799, *and it thus shows Laplace at an age no more than* 50.

Laplace

5 avril 1817

donné à Mr Dantier

par Mde La Mise de Laplace