Studies in the history of probability and statistics:

The Bayesian contributions of Edmond Lhoste

LYLE D. BROEMELING

Department of Biostatistics, The University of Texas M D Anderson Cancer Center

1515 Holcombe Boulevard, Houston, TX 77030

Summary

The contributions of Edmond Lhoste are largely unknown outside of France, and even within that country are not well known. His main contributions were in two areas: (1) the development of distributions that represent little or no prior knowledge, and (2) a sophisticated posterior analysis for normal and binomial populations. His results are similar to those of Haldane (1948) and Jeffreys (1939), but they appeared much earlier in 1923. They represent a significant and unique contribution to Bayesian ideas in the early part of the 20th century.

1. INTRODUCTION

Edmond Lhoste was born in 1880 and died in 1948. He studied engineering at the Ecole Polytechnique, and in 1923, while serving as a captain in the French army, he wrote ***Le Calcul des probabilités appliqué a l' artillerie, lois de probabilité a priori***. This was a collection of four articles comprising about 92 pages that appeared respectively in the May, June, July, and, August issues of the *Revue d' Artillerie.* For this review, Lhoste's results from the June and July issues are emphasized, because they contain his original ideas. The May issue is an introduction and motivation for the study, and the August issue focuses on applications to artillery.

In the second article, his contributions to Bayesian statistics were substantial and include the development of vague prior distributions that represent little or no knowledge about the mean and variance of a normal distribution and about the probability of success for the binomial distribution. His results were similar to those of Jeffreys (1961) for the normal distribution, and to those of Haldane (1948) for the binomial distribution.
For example, he used a constant prior for the mean of a normal, while for the standard deviation he used a prior that was the inverse of the standard deviation, namely

$f(\sigma) = 1/\sigma$ for $\sigma > 0$.

The prior for the binomial parameter θ was

$f(\theta) \propto 1/[\theta(1-\theta)]$, which was also recommended, but much later, by Haldane (1948).

In addition to his work with prior information, he developed sophisticated Bayesian inferential procedures for the normal and binomial distributions and applied them to problems in artillery. Many of his contributions were important and novel and have been commonly regarded to be discovered by other authors.

2. THE DEVELOPMENT OF VAGUE PRIOR INFORMATION

In the June article, Lhoste begins with the declaration “ Je n’ai pas l’intention d’exposer dans ces quelques articles la théorie classique du caclcul des probabilités. Une telle étude serait superflue.” He has declared independence from the usual approach to Bayesian presentations and then announces he will use the method of Bayes Rule (regle de Bayes) to develop Bayesian inferences for the mean and standard deviation of the normal distribution. Assuming the mean of the normal is unknown and variance known, he states Bayes’ theorem, namely that the posterior density of the mean is equal to a constant times the prior density of the mean times the likelihood function for the mean, based on *n* i.i.d. observations. I have taken the liberty of using modern terminology. For example, Lhoste would use the phrase, “la courbe de dispersion” (the curve of dispersion), for what we now call the density function, and he would employ “les causes” for prior information about a parameter. Using such terms as cause and effect for prior information and observable events, respectively, was the standard terminology in statistics in discussing inverse probability at that time. Another quaint phrase employed by Lhoste was “une grandeur éventuelle” to designate a parameter as a random variable. Poincaré (1912) and Keynes (1920) employed similar language in their influential textbooks. Beginning with Haldane (1932) and Jeffreys (1939) there is a shift toward the use of more modern phrases.

Lhoste’s June article consists of developing vague prior distributions for: (1) the mean μ of a normal, (2) the standard deviation σ of a normal, and (3) the binomial parameter θ. For each case, the prior distribution for the relevant parameter is developed, using techniques based on little or no prior knowledge about the parameter.

For the mean of a normal population, Lhoste proposed a constant prior density, based on something resembling the principle of precise measurement. See Barnett (1999, page 279) for a discussion of this principle. The likelihood function and the prior density are considered density functions of the unknown mean μ, and the ratio of the variance of the likelihood function to the variance of the prior density is considered to be small (because the variance of the prior density is considered to be large relative to that of the likelihood function). This, in turn, implies that the graph of the prior density is flat relative to that of the likelihood function, or in the words of Lhoste, “En d’autres termes, si l’on admet que le rapport est tres petit’, la forme de la fonction n’a pas d’influence sur le resultat final et on peut pratiquement reamplacer cette function par une constante.” (Translation: “In other words, if one admits that the ratio is small, the form of the prior function doesn’t have influence on the final result and we can practically replace the prior by a constant.”)

With regard to his choice of the prior for the standard deviation σ, Lhoste reasons that the prior density for σ should be the same as that for its reciprocal, and he lets

$$f(\sigma) = 1/\sigma \text{ for } \sigma > 0, \tag{2.1}$$

be the prior density for the standard deviation. Thus, our lack of knowledge about σ should be the same as our lack of knowledge about $1/\sigma$. This is similar to Jeffreys' (1961, page 119) invariance principle that states that prior information about σ should be the same as that for any power of σ.

To determine the prior density for the binomial parameter θ, Lhoste used the prior induced by the prior density for the odds $\gamma = \theta/(1-\theta)$, and since the odds and σ for the normal distribution have the same domain, the positive numbers, the same prior density for the odds is assigned as that for σ of a normal distribution (2.3), which leads to

$$f(\theta) \propto 1/[\theta(1-\theta)] \tag{2.2}$$

as the prior density for θ.

As mentioned earlier, these determinations of prior information resemble those used by Jeffreys (1945, 1961) for the mean and variance of a normal population, and by Haldane (1948) for the binomial parameter. Later we will discuss and compare the manner in which vague prior information was justified and assigned to the parameters by Lhoste, Jeffreys, and Haldane.

3. BAYESIAN INFERENCES FOR NORMAL AND BINOMIAL POPULATIONS

In his July article, Lhoste develops Bayesian inferences for the parameters of normal and binomial populations, using the vague priors he developed in the first article. Several cases are considered: (1) a normal population with known standard deviation σ (l'erreur moyenne quadratique), but unknown mean μ (la valeur probable), (2) the mean known, but standard deviation unknown, (3) both mean and variance unknown, and (4) the unknown binomial parameter θ.

3.1 Mean unknown but standard deviation known. Using the constant prior density, the mean has posterior density

$$f(\mu \mid data) \propto \exp\text{-}[n(\mu - \bar{x})^2 / (2\sigma^2)], \qquad (3.1)$$

thus, by inspection, the posterior distribution of the mean is normal with mean the sample mean and variance σ^2 / n.

3.2 Mean is known but the standard deviation is unknown. With a vague prior density for σ given above (2.1), the posterior density of the standard deviation is f(σ |data) $\propto$ $\sigma^{-(n+1)}$ exp- A/ 2 σ^2, (3.2)

where A= $\sum_{i=1}^{i=n}(x_i-\mu)^2$.

Thus, the posterior distribution of σ^2 is an inverse gamma with alpha = n/2 and beta = A/2. See Gelman et.al. (1995) for the properties of the inverse gamma. This would give as the posterior mean

E(σ^2 |data) = A/(n-2).

Actually, Lhoste works with the posterior distribution of a precision parameter h = 1/$\sqrt{2}$ σ and derives its mean as

E(h|data) = $\sqrt{(n-1)S^2/(n-3/2)}$,

by using the value of the normalizing constant given for the posterior distribution of σ above.

3.3. For the third case above when both parameters are unknown, the marginal posterior distributions of the mean μ and standard deviation σ are derived employing as the joint prior density

$$f(\mu, \sigma \,|\, \text{data}) \propto 1/\sigma. \quad (3.4)$$

He writes that the joint density of the two parameters is a constant times the prior density of the mean, times the prior density of the standard deviation, times the likelihood function for both parameters based on *n* i.i.d. observations. He does this without stating that in effect one is assuming, *a priori*, that the mean and standard deviation are independent, with a constant prior for the mean and the vague prior (2.1) for the standard deviation. Nevertheless, he derives as the marginal posterior distribution for the mean as a *t*-distribution with *n*-1 degrees of freedom, the mean is the sample mean, and the precision parameter is n/S^2, where S^2 is the sample variance. See DeGroot (1970) for this parameterization of the *t*-distribution with density

$$f(\mu \,|\, \text{data}) \propto [1+n(\mu-\bar{x})^2/(n-1)S^2]^{-n/2}.$$

The posterior moments of μ are

$E(\mu \,|\, \text{data}) = \bar{x}$, and

$\text{Var}(\mu \,|\, \text{data}) = S^2(n-1)/n(n-3)$.

These results derived by Lhoste are used today in many Bayesian texts for inferences about the mean and variance of a normal. See, for example, Zellner (1971), Box and Tiao (1973), and Broemeling (1985).

I believe Lhoste was the first to employ $1/\sigma$ as the joint prior density for μ and σ, however, before Lhoste, others had derived the posterior distribution of the mean of normal population when both parameters are unknown. See Pfanzagl and Sheynin (1996) for the history of the early Bayesian work on the *t*-distribution. Luroth (1876), Edgeworth (1883), and Burnside (1923) all derived the posterior distribution of the mean as a *t*-distribution (with various degrees of freedom), all of which were related to that given by Lhoste; however, they used a constant joint prior density for their particular parameterization of the normal.

With regard to marginal posterior distribution of σ, Lhoste arrived at

$$f(\sigma|\text{data}) \propto \sigma^{-n} \exp - (n-1)S^2 / 2\sigma^2, \quad (3.5)$$

thus, the posterior distribution of σ^2 is inverse gamma with alpha =(n-1)/2 and beta = (n-1) S^2/2 and this differs by one degree of freedom from the posterior distribution (3.2) of the same parameter when the mean is known. The posterior mean of the variance is

$$E(\sigma^2|\text{data}) = (n-1)S^2/(n-3),$$

where $(n-1)S^2 = \sum_{i=1}^{i=n} (x_i - \bar{x})^2$.

I believe that Lhoste was the first to develop the posterior distribution of the standard deviation of the normal distribution.

3.4 When making inferences for the binomial parameter and using his prior density (2.2), Lhoste derives the posterior distribution as the Beta density

$$f(\theta|\text{data}) \propto \theta^{x-1}(1-\theta)^{n-x-1}, \quad (3.6)$$

with parameters *x* and *n-x* , where *x* is the number of successes in n independent Bernoulli trials with probability of success θ. His use of this prior density results in the posterior mean

$$E(\theta|\text{data}) = x/n , \quad (3.7)$$

the sample proportion, which is the same as Haldane's (1948) prior distribution. He also found the posterior variance of θ as

$V(\theta|\text{data}) = x(n-x)/n^2(n+1)$.

I believe Lhoste was the first, or certainly among the first (see Villegas, 1990), to use this prior density, resulting in the usual maximum likelihood estimator of the proportion. Of course, both Bayes and Laplace employed the uniform prior density, which would give

$$E(\theta\ |\text{data}) = (x+1)/(n+2) \quad (3.8)$$

for the posterior mean. On the other hand, with Jeffreys' (1961) prior (alpha = beta = ½), the posterior mean would be

$$E(\theta\ |\text{data}) = (x+1/2)/(n+1). \quad (3.9)$$

4. HISTORICAL SIGNIFICANCE OF LHOSTE

Lhoste is essentially unknown outside of France and even inside that county is little-known. He is mentioned by Villegas (1990), but I can find very little information about him in the English language journals. In France, Dumas (1945, 1947, 1982, 1985), an obvious protégé of Lhoste, was quite enthusiastic in his admiration, and many of his publications concerned the articles by Lhoste. He was well aware that those contributions were much earlier than those of Jeffreys and Haldane. It is interesting to note that both Dumas and Lhoste were professional engineers.

What is the historical significance of the work of Lhoste? He was aware of the Bayesian literature and references such authors as Bayes (1763), Laplace (1774), Boole (1854), Venn (1866), Bertrand (1889), de Morgan (1847) , Jevons (1874), Lotze (1874), Czuber(1921), Poincaré (1912), Keynes (1920), and Pearson (1920). He would have had to have been heavily influenced by Bertrand and Poincaré.

On the other hand, I do not think he knew about Luroth (1876), Edgeworth, (1883), or Burnside (1923), all of whom derived the posterior *t*-distribution of the mean of a normal distribution, with variance unknown. These were *t*-distributions all based on uniform priors for some parameterization of the normal distribution. For example, Edgeworth (1883) used a uniform prior distribution for the mean and precision (the inverse of the variance). These authors employed their priors without comment and without a rational justification; however Lhoste goes to great length to justify his choice of vague prior information. With regard to the mean of the normal, and as we have seen, he used a constant, based on assuming the prior density has a variance that is large relative to the 'variance' of the likelihood function. This implied to him that the prior can be chosen as a constant.

Lastly, for the binomial population, Lhoste developed the Haldane (1948) prior, so-called by Jeffreys (1961), some twenty five years before Haldane. It is interesting to note that Haldane chose his prior so that the mean of the posterior distribution of θ was the sample proportion x/n, because the sample proportion has optimal sampling properties. Lhoste's approach was quite different and based on a consistency argument linking the prior of the odds and the prior he had developed for standard deviation of a normal population. The same approach was also examined by Jeffreys (1961, page 123).

Lhoste was an innovator with his detailed and informative posterior analysis of the standard deviation of a normal population. He devotes 6 pages of Appendix II of the July article to the properties of the posterior distribution of the standard deviation and to the precision $h = 1/2\sigma$, and, using the Euler integral of the second kind (the integral of the gamma function), derives the mean, variance, and coefficient of variation of the posterior distribution of h. One would expect Luroth, Edgeworth, or Burnside to have derived the posterior distribution of the standard deviation, because they were analyzing a normal population with both parameters unknown. Apparently they did not, however, Edgeworth (1883) did calculate and estimate σ by maximizing the joint posterior distribution of μ and σ.

Jeffreys (1961, page 139, equation 7) gave a similar analysis, but unlike Lhoste, did not go further in using it to make inferences about the normal population. For example, Jeffreys did not derive the posterior mean and variance of σ, as did Lhoste.

To summarize his work, Lhoste's contributions fall into two main areas: (1) a justification for the choice of prior distributions that convey 'little' information about the parameters of the model, and (2) a detailed and modern posterior analysis for the normal distribution, including a unique focus on the standard deviation of a normal population.

Our study of Lhoste gives us a better understanding of the history of Bayesian ideas in the early 20th century. Dale's (1991) history of inverse probability can now be augmented by adding Lhoste's name to the list of innovative Bayesian researchers. Also, we can now complete the gap between what was provided by Poincaré, Pearson, and Keynes during the period of 1900 to 1920, and what Jeffreys gave us beginning in 1933. We now have a better understanding of the significance and originality of Lhoste's impact on the field.

Acknowledgement

I wish to thank Monsieur Maurice Dumas who was Ingenieur en chef de l'Artillerie navale and President de la Societé de statistique de Paris. He initiated this study by contacting me some years ago in making me aware of the many contributions of Edmond Lhoste. I am indebted for his efforts, and this review is dedicated to his memory.

The author is indebted to Ana Langevin Broemeling for the English translation of ***Le Calcul des probabilités applique a l' artillerie, lois de probabilité a priori*** by Edmond Lhoste.

Steve Stigler was very kind in providing me with the Pfanzagl–Sheynin (1996) article. I was unaware of this information and it helped me to better appreciate the historical significance of Lhoste's work and how it related to earlier Bayesian work on the *t*-distribution.

Also, I would like to thank Christian Robert, who provided me with valuable information about Maurice Dumas, and his articles that referenced Lhoste.

References

Barnett, V. (1999) *Comparative Statistical Inference* John Wiley & Sons Ltd.

Bayes, R.T. (1763) An essay toward solving a problem in the doctrine of chances. *Phil. Trans. Royal Soc*. 53, 370-418.

Bertrand, J.L.F. (1889) *Calcul des Probabilités*: Paris; 2-nd edition 1907, Reprinted as 3-rd edition, 1972, New York, Chelsea.

Boole, G. (1854) *An Investigation of the Laws of Thought, on Which are Founded the Mathematical Theories of Logic and Probabilities*. London, Macmillan. Reprinted in 1958, Dover, New York.

Box, G.E.P. and Tiao, G.C. (1973) *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Reading, Mass.

Broemeling, L.D. (1985) *Bayesian Analysis of Linear Models,* Marcel-Dekker, New York.

Burnside, W. (1923) On errors of observation. *Proc. Camb. Phil. Soc.* 21, 482-487.

Czuber, E. (1921) *Die Statistischen Forschungmethod*. Vienna, L.W. Seidel.

Dale A.I. (1991) *A History of Inverse Probability, from Thomas Bayes to Karl Pearson.* Springer–Verlag, New York.

de Morgan. (1847) *Formal Logic*. London, Taylor & Watson.

Dumas, M. (1945) Sur les lois de probabilités divergentes en la formule de Fisher. Pages 127-130 of *Congrés de la Vicloire, Association Francaise pour l'avancement de sciences, 64-th Session, Paris* .

Dumas, M. (1947) Sur une loi de probabilité a priori conduisant aux arguments fiducaires de Fisher, *Le Revue Scientifique* , 3-18.

Dumas, M. (1982) Lois de probabilité a priori de Lhoste pages 3-17 in *Sciences et techniques de l'armement, Memorial de l'artillerie francaise.*

Dumas, M. (1985) Sénéque les évidences a posterior. Journal de la Societé de statistique de Paris, Vol.126, No. 1, 136-138.

Edgeworth, F.Y. (1883) The method of least squares. *The London, Edinburgh, and Dublin Phil. Mag*. Ser. 5, 16, 360-375.

Degroot, M.H. (1970) *Optimal Statistical Decisions*, McGraw-Hill Book Company.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (1995) *Bayesian Data Analysis.* Chapman & Hall/CRC.

Haldane, J.B.S. (1932) A note on inverse probability. *Proc. Cam.Phil. Soc*. 28, 55-61.

Haldane, J.B.S. (1948) The precision of observed values of small frequencies. Biometrika, 297-300.

Jeffreys, H. (1939, 1945, and 1961) First, second, and third editions of *Theory of Probability*. Oxford U. Press.

Jevons, W.S. (1874) *The Principles of Science. A treatise on logic and scientific method*, 2 vols. (Reprinted in one volume in 1877), London, Macmillan.

Keynes, J.M. (1920) *A Treatise on Probability.* Macmillan & Co. Ltd.

Laplace, P.S. (1774) Mémoire sur les suites récurro-récurrentes et sur leurs usages dans la théorie des hasards. *Mémoires des l'Academie royale des Sciences de Paris (Savants* é*trangers)* 6: 353-371. Reprinted in *Euvres completes de Laplace* 8:5-24.

Lhoste, E. (1923) Le Calcul des probabilités appliqué l’artillerie, lois de probabilité a priori. *Revue d’artillerie*, Mai-Août, Berger-Levrault, Paris.

Lotze, R.H. (1874) *Logik, drei Bucher, vom Denken, vom Untersuchen und vom Erkennen*. Leipzig:. Hirzel. 2-nd edition 1880, 3-rd edition 1912. (Translated by B. Bosanquet 1884 as *Logic in Three Books of Thought, of Investigation, and of Knowledge*. Oxford Clarendon Press).

Luroth, J. (1876) Vergleichung von zwei werten der wahrscheinlechen fehler. *Astron*. *Nach*. 87, 209-220.

Pearson, K. (1920) The fundamental problem of practical statistics, Biometrika 13, pp 1-16.

Pfanzagl, J. and Sheynin, O. (1996) Studies in the history of probability and statistics XVIV, A forerunner of the t-distribution. Biometrika 83, 4, pp 891-898.

Poincaré, H. (1912) *Calcul des probabilités* 2-nd edition, Paris.

Venn, J. (1866) *The Logic of Chance*. London, Macmillan (2-nd edition 1876, 3-rd edition 1888).

Villegas, C. (1990) Bayesian inference in models with Euclidean structures. J. Am. Stat. Assoc. 85, No. 412, 1159-1164.

Zellner. A. (1971) *An Introduction to Bayesian Inference in Econometrics*. John Wiley & Sons Inc., New York.