6

Reprinted from:

RECENT DEVELOPMENTS IN STATISTICS

Proceedings of the European Meeting of Statisticians
Grenoble, 6-11 September, 1976

Edited by

J. R. BARRA
F. BRODEAU
G ROMIER
B. VAN CUTSEM
University of Grencble

1977

NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM • NEW YORK • OXFORD

Reprinted from:

RECENT DEVELOPMENTS IN STATISTICS

Proceedings of the European Meeting of Statisticians
Grenoble, 6-11 September, 1976

Edited by

J. R. BARRA
F. BRODEAU
G. ROMIER
B. VAN CUTSEM
University of Grenoble

1977

NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM · NEW YORK · OXFORD

Recent Developments in Statistics
J.R.Barra et al., editors

INFERENCES ABOUT THE RATIO OF NORMAL MEANS: A BAYESIAN APPROACH TO THE FIELLER-CREASY PROBLEM

José M. Bernardo
Departamento de Estadística
Facultad de Ciencias
Universidad de Valencia
Valencia, Spain

In biological assay work the problem arises of estimating the ratio of the means of two normal populations. The problem gave rise to a controversy among fiducialists and confidence-intervalists and it is usually known as the Fieller-Creasy problem.

A Bayesian solution to such a problem is presented here making use of a new method, developed by the author, which provides appropriate reference 'non-informative' prior distributions.

THE PROBLEM

In biological assay work one is often interested in the relative power of two treatments or drugs, and the following problem suggests itself. Suppose that two samples $x=\{x_1,x_2,\ldots,x_n\}$ and $y=\{y_1,y_2,\ldots,y_m\}$ are available from two independent normal populations with unknown means μ,η and common unknown variance σ^2. The problem is to make inferences about the value of $\psi=\mu/\eta$, the ratio of the means.

This problem was discussed in a Symposium on Interval Estimation held by the Royal Statistical Society back in 1954. E.C. Fieller and M.A. Creasy presented there two different solutions that both claimed to be fiducial. Fieller's solution is difficult to accept for it can lead, for instance, to a 'confidence' interval consisting in the *whole* real line. Kappenman, Geisser and Antle (1970) showed that Creasy's solution may be reproduced from a Bayesian point of view by using the usual 'non-informative' prior $p(\mu,\eta,\sigma)\propto\sigma^{-1}$.

However, the uncritical use of standard so-called 'non-informative' priors has been increasingly questioned. Although Jeffreys' (1939/67) prior is often accepted in the one-dimensional continuous case, no similarly acceptable results seem to exist in the case of several parameters. A key reference is Dawid, Stone and Zidek (1973) and ensuing discussion.

Our view, developed in Bernardo (1976) is that in *each* particular inference problem a *reference* posterior distribution may be produced which only uses information provided by the data and the model specification. Such a reference posterior may be used as an *origin* with respect to which posteriors obtained from informative priors may be compared, in order to assess the relative importance of the initial opinions in the final inference. In this paper, we obtain the reference posterior distribution which corresponds to the Fieller-Creasy problem.

DEFINITION OF A REFERENCE POSTERIOR

Consider a vector θ of parameters and an experiment E providing some data z which give information about θ. Suppose that we are interested in the value of some *real* function $\psi=\psi(\theta)$ of the parameters. Without loss of generality, one may assume that ψ is the first component of θ, for otherwise an appropriate transformation could be made to achieve such a situation. Then, $\theta=\{\psi,\omega\}$, $\omega=\{\omega_1,\omega_2,\ldots,\omega_k\}$.

The basic idea underlying our construction of a reference posterior distribution for ψ is as follows. First, generalizing earlier work by Lindley (1956), the expected information about ψ to be provided by E when the prior density of θ is $p(\theta)$ is defined to be

$$I^{\psi}\{E,p(\theta)\} = \iint p(z,\theta)\log\frac{p(\psi|z)}{p(\psi)}d\theta\, dz$$

whenever the integral exists. Similarly, the *residual* expected information about each of the ω_i's when ψ is known may be defined as

$$I^{\omega_i|\psi}\{E,p(\theta)\} = \iint p(z,\theta)\log\frac{p(\omega_i|\psi,z)}{p(\omega_i|\psi)}d\theta\, dz$$

Now, let $E(n)$ be the experiment which consists of n independent replications of E and consider $I^{\psi}\{E(n),p(\theta)\}$, the expected information about ψ to be provided by $E(n)$ when the prior density of $\theta=\{\psi,\omega\}$ is $p(\theta)=p(\psi)p(\omega|\psi)$. By performing infinite replications of E one could expect to know θ, and therefore ψ. Thus, the number $I^{\psi}\{E(\infty),p(\theta)\}$, if it exists, measures the amount of *missing information* about ψ that one could obtain by repeating E when the prior is $p(\theta)$. It seems natural to define *diffuse* opinions about ψ relative to E as those described by that density $\pi(\psi)$ maximizing the missing information about ψ, $I^{\psi}\{E(\infty),p(\theta)\}$, for any fixed $p(\omega|\psi)$, providing such a density exists.

All prior densities of the form $\pi(\psi)p(\omega|\psi)$ will be called *ψ-diffuse relative to E*; they differ in the opinions they describe about the value of ω given ψ. We propose to select as reference prior the more 'ω-diffuse' among them that we define to be the one maximizing *each* of the missing residual informations about the ω_i's given ψ, $I^{\omega_i|\psi}\{E(\infty),\pi(\psi)p(\omega|\psi)\}$, when the marginal prior of ψ is $\pi(\psi)$, providing such a density $\pi(\omega|\psi)$ exists. The reference posterior distribution of ψ will be that obtained by the formal use of Bayes theorem with the reference prior $\pi(\theta)=\pi(\psi)\pi(\omega|\psi)$ just described.

Only a slight generalization of earlier work by Stone (1958) is necessary to establish that if, as is usual, the posterior distribution of ψ is asymptotically normal with precision $nh_0(\hat\theta)$ where $\hat\theta$ is the maximum likelihood estimator of θ then for any positive prior $p(\theta)$, i.e. such that $p(\theta)>0$ for any θ, and $n\to\infty$,

$$I^{\psi}\{E(n),p(\theta)\} = \frac{1}{2}\log\frac{n}{2\pi e} + \int p(\theta)\log\frac{h_0(\theta)^{1/2}}{p(\psi)}d\theta + o(1)$$

It follows (Bernardo, 1976) that if $h_0(\theta)$ may be decomposed such that

$$h(\theta)^{1/2} = \pi(\psi)f(\omega)$$

then the missing information about ψ is maximized, for any fixed $p(\omega|\psi)$, when $p(\psi)=\pi(\psi)$.

Similarly, if the posterior distribution of ω_i given ψ is asymptotically normal with precision $nh_i(\hat\theta)$ and

$$h_i(\theta)^{1/2} = \pi(\omega_i)g(\theta_i^*)$$

where θ_i^* contains all the components of θ except ω_i, then the missing information about ω_i given ψ is maximized when $p(\omega_i|\psi)=\pi(\omega_i)$ for any fixed $p(\omega_j|\psi)$, $j\neq i$, and for any fixed $\pi(\theta)$. Therefore, in such a case the ψ-reference prior is

$$\pi(\psi)\prod_i\{\pi(\omega_i)\}$$

and the reference posterior for ψ is that obtained from such a prior.

THE REFERENCE POSTERIOR OF THE RATIO OF TWO MEANS

Consider again the Fieller-Creasy problem. Here, $\theta=\{\mu,\eta,\sigma\}$ and $\psi=\mu/\eta$. The problem may be reparametrized in terms of $\zeta=\{\psi,\eta,\sigma\}$. It is known (Walker, 1968) that the posterior distribution of ψ is asymptotically normal with precision matrix $nF(\hat\zeta)$ where $\hat\zeta$ is the maximum likelihood estimator of ζ and $F(\zeta)$ is Fisher's information matrix of typical element

$$-\int p(z|\zeta)\frac{\partial^2}{\partial\zeta_i\partial\zeta_j}\log p(z|\zeta)dz$$

In this case, the matrix is easily found to be

$$F\begin{pmatrix}\psi\\ \eta\\ \sigma\end{pmatrix} = \frac{1}{\sigma^2}\begin{pmatrix}\eta^2 & \psi\eta & 0\\ \psi\eta & 1+\psi^2 & 0\\ 0 & 0 & 4\end{pmatrix}$$

It follows that the asymptotic posterior distribution of ψ and those of η and σ given ψ are normal with respective precisions (see e.g. Graybill, 1961, ch. 3) $nh_0(\hat\zeta)$, $nh_1(\hat\zeta)$, $nh_2(\hat\zeta)$, where

$$h_0(\zeta) = \eta^2(1+\psi^2)^{-1}\sigma^{-2}$$
$$h_1(\zeta) = (1+\psi^2)\sigma^{-2}$$
$$h_2(\zeta) = 4\sigma^{-2}$$

Thus, the ψ-diffuse densities are those of the form

$$p(\psi,\eta,\sigma) \propto (1+\psi^2)^{-1/2}p(\eta,\sigma|\psi)$$

and the ψ-reference prior is $\pi(\psi,\eta,\sigma)=(1+\psi^2)^{-1/2}\sigma^{-1}$ or, in terms of the original parametrization

$$\pi(\mu,\eta,\sigma) = (\mu^2+\eta^2)^{-1/2}\sigma^{-1} \tag{1}$$

The reference posterior distribution of $\psi=\mu/\eta$ after the samples $x=\{x_1,x_2,\ldots,x_n\}$ and $y=\{y_1,y_2,\ldots,y_m\}$ have been observed may now be produced via Bayes Theorem. Our reference prior combines nicely with the likelihood function so that, unlike the posterior density for ψ obtained from the usual prior $p(\mu,\eta,\sigma)=\sigma^{-1}$, the reference posterior distribution for ψ may be obtained in closed form.

Indeed, the joint density of the two samples is

$$p(x,y|\mu,\eta,\sigma) = (2\pi)^{-(n+m)/2}\sigma^{-(n+m)}\exp\left[\frac{1}{2\sigma^2}\{S^2+n(\bar x-\mu)^2+m(\bar y-\eta)^2\}\right] \tag{2}$$

where $\bar x=\Sigma x_i/n$, $\bar y=\Sigma y_i/m$, and $S^2=\Sigma(x_i-\bar x)^2+\Sigma(y_i-\bar y)^2$. Combining (1) and (2), the joint posterior density of $\theta=\{\mu,\eta,\sigma\}$ is

$$p(\mu,\eta,\sigma|x,y) \propto (\mu^2+\eta^2)^{-1/2}\sigma^{-(n+m+1)}\exp\left[\frac{1}{2\sigma^2}\{S^2+n(\bar x-\mu)^2+m(\bar y-\eta)^2\}\right]$$

so that, since the Jacobian of the one-to-one transformation from $\theta=\{\mu,\eta,\sigma\}$ to $\zeta=\{\psi,\eta,\sigma\}$ is $1/|\eta|$, the joint posterior density of $\zeta=\{\psi,\eta,\sigma\}$ is

$$p(\psi,\eta,\sigma|x,y) \propto (1+\psi^2)^{-1/2}\sigma^{-(n+m+1)}\exp\left[\frac{1}{2\sigma^2}\{S^2+n(\bar x-\psi\eta)^2+m(\bar y-\eta)^2\}\right] \tag{3}$$

Using gamma-related integral results, σ is easily integrated out from (3) to get

$$p(\psi,\eta|x,y) \propto (1+\psi^2)^{-1/2}\left[S^2+\frac{nm(\bar x-\psi\bar y)^2}{m+\psi^2 n}\right]^{-(n+m)/2}$$
$$\cdot\left[1+\left\{\frac{m+\psi^2 n}{S^2+\frac{nm(\bar x-\psi\bar y)^2}{m+\psi^2 n}}\right\}\left\{\eta-\frac{m\bar y+\psi n\bar x}{m+\psi^2 n}\right\}^2\right]^{-(n+m)/2} \tag{4}$$

Finally, using Student-t related integral results, the parameter η may be integrated out from (4) to obtain

$$p(\psi|x,y) = C\,(1+\psi^2)^{-1/2}(m+\psi^2 n)^{-1/2}\{S^2+\frac{nm(\bar{x}-\psi\bar{y})^2}{m+\psi^2 n}\}^{-(n+m-1)/2} \quad (5)$$

where C is a normalizing constant. The density (5) is our reference posterior density for the ratio of the means.

This could be compared with the posterior density of ψ that one obtains from the usual so-called 'non-informative' prior

$$p(\mu,\eta,\sigma) = \sigma^{-1} \quad (6)$$

Using a similar procedure to that followed above, (6) and (2) may be combined to obtain the posterior density of $\theta=\{\mu,\eta,\sigma\}$ which is

$$p(\mu,\eta,\sigma|x,y) \propto \sigma^{-(m+n+1)}\exp\left[\frac{1}{2\sigma^2}\{S^2+n(\bar{x}-\mu)^2+m(\bar{y}-\eta)^2\}\right]$$

so that the corresponding posterior density of $\zeta=\{\psi,\eta,\sigma\}$ is

$$p(\psi,\eta,\sigma|x,y) \propto |\eta|\,\sigma^{-(m+n+1)}\exp\left[\frac{1}{2\sigma^2}\{S^2+n(\bar{x}-\psi\eta)^2+m(\bar{y}-\eta)^2\}\right] \quad (7)$$

As before, σ is easily integrated out of (7) to obtain an expression for the posterior distribution of ψ and η which is the same as (4) except that the factor $(1+\psi^2)^{-1/2}$ is substituted by $|\eta|$. To obtain the posterior distribution of the parameter of interest ψ one would now integrate out η. However, because of the factor $|\eta|$, this integral can no longer be evaluated in terms of Student-t intergral results and the posterior density of ψ cannot be expressed in closed form as above.

Our reference posterior density for the ratio of the means, given by (5), is always proper provided $m\geq 1$ and $n\geq 1$. It can have one or two modes which, as the constant C may be determined numerically. Clearly, it is symmetric about the origin when either $\bar{x}=0$ or $\bar{y}=0$. This was to be expected since, in either case, there is no information to decide on the sign of ψ. This feature is not obtained with the usual prior (6). When $n=m=1$ and $x=y=0$, the reference density (5) reduces to a Cauchy density function.

DISCUSSION

The reference density (5) has been studied using Montecarlo methods. For instance, random samples of size 5 were independently generated from Normal distributions with means $\mu=3$ and $\eta=1$ and variance $\sigma^2=1$. The results of three of those and the corresponding values for the statistics $\bar{x}$, $\bar{y}$ and S^2 are described below.

1.	$x=\{3.31, 3.90, 3.22, 2.00, 2.88\}$	$\bar{x} = 3.062$	$S^2 = 3.1608$
	$y=\{0.49, 0.64, 1.58, 1.53, 0.57\}$	$\bar{y} = 0.962$	
2.	$x=\{3.01, 3.16, 4.31, 2.62, 3.38\}$	$\bar{x} = 3.296$	$S^2 = 3.7728$
	$y=\{1.37, 0.17, 0.18, 0.74, 1.42\}$	$\bar{y} = 0.776$	
3.	$x=\{1.55, 3.33, 3.87, 1.10, 3.69\}$	$\bar{x} = 2.708$	$S^2 = 7.7297$
	$y=\{1.18, 1.30, 1.04, 1.67, 0.68\}$	$\bar{y} = 1.174$	

In Figure 1, the corresponding posterior distributions of ψ, calculated from (5) estimating numerically the corresponding proportionality constants, are plotted together. All of them seem to assess a reasonable high posterior density to the true value $\psi=3$.

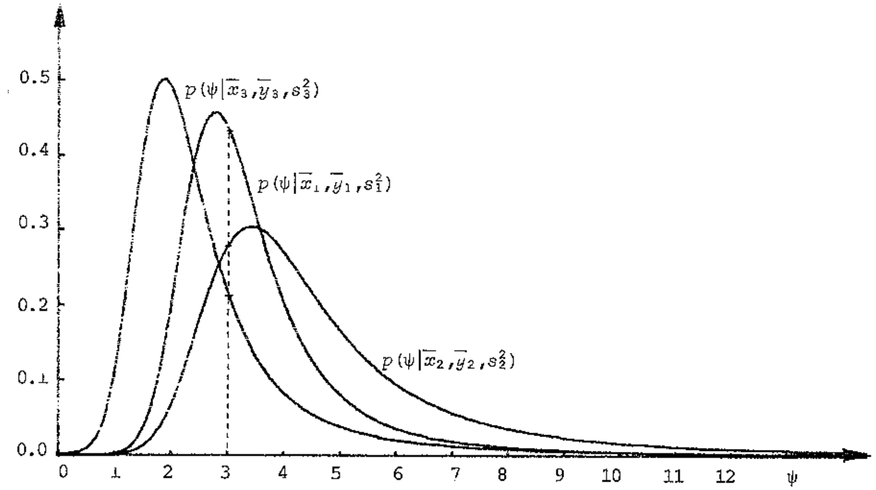

Figure 1. Example of posterior distributions of ψ.

No claim of 'objectivity' (whatever that may mean) is made for the set of inferences about ψ that could be produced from (5). It is only argued that (5) is a useful *reference* density to look at for it gives a feeling of the sort of values of ψ that the data $z=\{x,y\}$ are supporting. One should compare (5) with the posterior density of ψ obtained from an informative prior which describes one's initial information on the particular problem dealt with. From this comparison, one would obtain information about the relative importance of data and prior information in the final conclusions.

ACKNOWLEDGEMENTS

I am grateful to Professor D.V. Lindley for his useful comments and to Dr. J. Basulto and Mr. I. Zabala for their help with the computations.

REFERENCES

Bernardo, J.M. (1976). *The use of information in the design and analysis of scientific experimentation*. Ph.D. Thesis, University of London.

Creasy, M.A. (1954). Limits for the ratio of the means. *J. Roy. Statist. Soc. Ser. B 16*, 186-194 (with discussion).

Dawid, A.P., Stone, M. and Zidek, J.V. (1973). Marginalization Paradoxes in Bayesian and Structural Inference. *J. Roy. Statis. Soc. Ser. B 35*, 189-233 (with discussion).

Fieller, E.C. (1954). Some problems in interval estimation. *J. Roy. Statist. Soc. Ser. B 16*, 175-185 (with discussion).

Graybill, F.A. (1961). *An Introduction to Linear Statistical Models*. (McGraw-Hill, New York)

Jeffreys, H. (1939/67). *Theory of Probability*. (Clarendon Press, Oxford).

Kappenman, R.F., Geisser, S. and Antle, C.E. (1970). Bayesian and fiducial solutions to the Fieller-Creasy problem. *Sankhya B 32*, 331-340.

Lindley, D.V. (1956). On a measure of the Information provided by an Experiment. *Ann. Math. Statist. 27*, 986-1005.

Stone, M. (1958). *Studies with a measure of information*. Ph.D. Thesis, University of Cambridge.

Walker, A.M. (1969). On the asymptotic behaviour of posterior distributions. *J. Roy. Statist. Soc. Ser. B 31*, 80-88.