TRABAJOS DE ESTADISTICA Y
DE INVESTIGACION OPERATIVA
Vol. 29, Cuad. 3, Madrid, 1978

UNACCEPTABLE IMPLICATIONS OF THE LEFT HAAR MEASURE IN A STANDARD NORMAL THEORY INFERENCE PROBLEM

José M. Bernardo
Departamento de Estadística
Universidad de Valencia

SUMMARY

For a very common statistical problem, inference about the mean of a normal random variable, some inadmissible consequences of the left Haar invariant prior measure, which is that recommended as a suitable prior by Jeffreys' multivariate rule and by the methods of Villegas and Kashyap, are uncovered and investigated.

Some key words: Improper prior distributions; Inference about the normal means; Multivariate Jeffreys' rule; Non informative priors.

1. Introduction.

The difficulties in finding a consistent scheme for the selection of 'objective', usually improper priors, for Bayesian inference with 'no initial information' are well known. Thus, although Jeffreys' (1939/67 ch.3) prior is often accepted in the one dimensional continuous case, no similarly acceptable results seem to exist in the case of several paramenters. Key references are Dawid, Stone and Zidek (1973) and Stone (1976) and ensuing discussions.

However, in some situations, improper priors may be used to produce suitable approximations to proper posterior distributions (see e.g. Lindley, 1965, §5.2) and, as a matter of fact, this is sistematically done in most Bayesian textbooks, like Box and Tiao (1973) and Zellner (1971). Therefore, it seems important to explore critically the implications of the different improper priors that have been proposed.

Perhaps the more common statistical problem is that of making inferences about the mean μ of a random variable x distributed as $N(\mu, \sigma^2)$. Several so-called 'non-informative' prior measures have been proposed in this case, usually within the class of relatively invariant measures $\{d\mu d\sigma/\sigma^{\lambda+1}, -\infty < \lambda < \infty\}$. With $\lambda = 1$ one obtains $d\eta d\sigma/\sigma^2$, the left Haar measure, which is the one recommended by Jeffreys' multivariate rule, and by Villegas (1971) and Kashyap (1971) methods. We present here a consequence of the use of such a prior which, in our opinion, casts serious doubts on it reasonableness and, consequently, on the methods which produce it.

2. The example.

Let x_1, x_2 be two observations of a random variable x which is distributed as $N(\mu, \sigma^2)$, and consider the probability of the event A that $x_1 < \mu < x_2$ after x_1 and x_2 have been observed, i.e. $p(A|x_1, x_2)$, where for convenience, it is assumed that $x_1 < x_2$. If the prior measure is $d\mu d\sigma/\sigma^{\lambda+1}$, then

$$p(\mu, \sigma | x_1, x_1) \propto p(x_1, x_2 | \mu, \sigma)\, p(\mu, \sigma)$$

$$\propto \sigma^{-2} \exp\left\{-\frac{1}{2\sigma^2}[(x_1-\mu)^2+(x_2-\mu)^2]\right\} \sigma^{-\lambda-1} =$$

$$= \sigma^{-\lambda-3} \exp\left\{-\frac{1}{\sigma^2}[(\bar{x}-\mu)^2 + d^2/4]\right\}$$

where $\bar{x} = (x_1 + x_1)/2$ and $d^2 = (x_1 - x_2)^2$. Therefore,

$$p(\mu | x_1, x_2) \propto \int_0^\infty \sigma^{-\lambda-3} \exp\left[-\frac{1}{\sigma^2}\{(\bar{x}-\mu)^2 + d^2/4\}\right] d\sigma \propto$$

$$\propto \{(\mu-\bar{x})^2 + d^2/4\}^{-(\lambda+2)/2} \propto$$

$$\propto \left\{1 + \frac{4}{d^2}(\mu-\bar{x})^2\right\}^{-(\lambda+2)/2}$$

so that the posterior distribution of μ is

$$p(\mu | x_1, x_2) = \frac{2}{d} \frac{\Gamma\left(\frac{\lambda+2}{2}\right)}{\Gamma\left(\frac{\lambda+1}{2}\right)\Gamma\left(\frac{\lambda}{2}\right)} \left\{1 + \frac{4}{d^2}(\mu-\bar{x})^2\right\}^{-(\lambda+2)/2}$$

a Student t with $\lambda + 1$ degrees of freedom.

Now, $p(A|x_1, x_2)$ may be computed by direct integration; for,

$$p(A|x_1, x_2) = \int_{x_1}^{x_2} k \left\{1 + \frac{4}{d^2}(\mu-\bar{x})^2\right\}^{-(\lambda+2)/2} d\mu =$$

making $t = 2(\mu - \bar{x})/d$, $dt = 2d\mu/d$,

$$= \int_0^1 kd\, \{1 + t^2\}^{-(\lambda+2)/2} dt \qquad (1)$$

where $kd = 2\Gamma\left(\frac{\lambda+2}{2}\right) \Gamma\left(\frac{\lambda+1}{2}\right)\Gamma\left(\frac{\lambda}{2}\right)$.

With the 'usual' prior $d\mu d\sigma/\sigma$, obtained for $\lambda = 0$, the integral (1) reduces to

$$\frac{2}{\pi}\int_0^1 \{1 + t^2\}^{-1} dt = \frac{2}{\pi} \arctan t \Big|_0^1 = \frac{2}{\pi}\frac{\pi}{4} = \frac{1}{2} \qquad (2)$$

while with the left Haar measure $d\mu d\sigma/\sigma^2$, obtained for $\lambda = 1$, the integral (1) becomes

$$\int_0^1 \{1 + t^2\}^{-3/2} dt = \frac{t}{\sqrt{1+t^2}}\Big|_0^1 = \frac{1}{\sqrt{2}} \simeq 0.7071 \qquad (3)$$

Let us now consider the problem conditionally to σ. In such case, asuming σ known and the uniform prior measure $d\mu$ for μ,

$$p(\mu|x_1, x_2) \propto p(x_1, x_2 | \mu, \sigma)\, p(\mu) \propto \exp\left\{-\frac{1}{\sigma^2}(\bar{x}-\mu)^2\right\}$$

so that the posterior density of μ is $N(\bar{x}, \sigma^2/2)$. Again, the posterior probability of $A \equiv \{x_1 < \mu < x_2\}$ given x_1 and x_2 may be computed by direct integration, for

$$p(A|x_1, x_2) = \int_{x_1}^{x_2} \frac{1}{\sigma\sqrt{\pi}} \exp\left\{-\frac{1}{\sigma^2}(\mu-\bar{x})^2\right\} d\mu =$$

making $t = (\mu - \bar{x})\sqrt{2}/\sigma$, $dt = d\mu\sqrt{2}/\sigma$,

$$= 2\int_0^{t_1} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{t^2}{2}\right\} dt = 2\left\{\Phi(t_1) - \frac{1}{2}\right\} = 2\Phi(t_1) - 1 \qquad (4)$$

where $t_1 = |x_1 - x_2|/\sigma\sqrt{2}$.

As one would expect, (4) depends on the data x_1 and x_2. However, one may obtain an upper limit for the expected value of (4). Indeed, $2\Phi(t_1) - 1$ is a concave function of t_1, for $\Phi(t)$ is the standarized normal distribution function and $t_1 \geqslant 0$. Thus, taking expectations with respect to the random variable t_1,

$$E\{2\Phi(t_1) - 1\} \leqslant 2\Phi\{E(t_1)\} - 1$$

But $d = x_1 - x_2$ is distributed as $N(0, 2\sigma^2)$ so that the probability density function of $\delta = |d| = |x_1 - x_2|$ given σ is

$$p(\delta|\sigma) = \frac{1}{\sigma\sqrt{\pi}} \exp\{-\delta^2/4\sigma^2\}$$

whose expected value is easily seen to be $2\sigma/\sqrt{\pi}$, so that

$$E(t_1|\sigma) = \frac{1}{\sigma\sqrt{2}} \frac{2\sigma}{\sqrt{\pi}} = \sqrt{(2/\pi)}$$

independently of σ. Therefore, we have an upper bound for the expected valve of $p(A|x_1, x_2)$ given by

$$E\; p(A|x_1, x_2) \leqslant 2\Phi\{\sqrt{(2/\pi)}\} - 1 \simeq 0.575$$

This is certainly compatible with the result (2) obtained with the prior measure $d\mu d\sigma/\sigma$ but it makes unacceptable the result (3) obtained with the left Haar measure $d\mu d\sigma/\sigma^2$ as a prior. We believe this is a serious objection to the use of such a prior and, consequently, to the methods which recommend it.

3. Discussion.

The argument in the preceeding Section suggests that, *in the absence* of other sources of information, one should have probability 1/2 that the mean of a normal random variable of unknown variance lies between the first two observations.

One may try to investigate whether the natural extension of the measure $d\mu d\sigma/\sigma$ to the multinormal homocedastic model, i.e. $d\mu_1, d\mu_2, \ldots, d\mu_k d\sigma/\sigma$ is consistent with this result. For simplicity, we shall concentrate in the case $k = 2$. Thus, let us consider the bivariate random variable $z = (x, y)$ which is normally distributed with mean (μ_1, μ_2) and covariance matrix $\sigma^2 I$; let $z_1 = (x_1, y_1)$ and $z_2 = (x_2, y_2)$ be two observations from z and let us compute $p(A|x_1, x_2)$ where A is the event that μ_1 lies between x_1 and x_2. Clearly,

$$p(A|x_1, x_2) = \int_{x_1}^{x_2} p(\mu_1|x_1, x_2)\, d\mu_1 =$$

$$= \int_{x_1}^{x_2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\mu_1|x_1, x_2, y_1, y_2)\, p(y_1, y_2|x_1, x_2)\, dy_1\, dy_2\, d\mu_1$$

an expression which, *if all densities involved were proper*, would reduce to that computed in Section 2, and we would obtain again the value 1/2. However, the 'predictive' density $p(y_1, y_2|x_1, x_2)$ of the y's given the x's is *not* proper and thus one cannot be sure that the argument goes through.

Nevertheless, we next prove that indeed, with the prior $d\mu_1 d\mu_2 d\sigma/\sigma$, and only with that prior, one obtains again $p(A|x_1, x_2) = 1/2$.

Omitting the details for brevety, the posterior distribution of μ_1 given x_1, x_2, y_1 and y_2 is the Student t with two degrees of freedom

$$p(\mu_1|x_1, x_2, y_1, y_2) = \frac{\sqrt{2}}{2s}\left\{1 + \frac{2}{s^2}(\bar{x} - \mu_1)^2\right\}^{-3/2}$$

where $s^2 = (d_1^2 + d_2^2)/2$ with $d_1 = x_1 - x_2$ and $d_2 = y_1 - y_2$. Thus,

$$p(A|x_1, x_2, y_1, y_2) = \int_{x_1}^{x_2} p(\mu_1|x_1, x_2, y_1, y_2)\, d\mu_1 =$$

$$= \int_{x_1}^{x_2} \frac{\sqrt{2}}{s}\left\{1 + \frac{2}{s^2}(\bar{x} - \mu_1)^2\right\}^{-3/2} d\mu_1 =$$

with $t = (\bar{x} - \mu_1)\sqrt{2}/s$, $dt = d\mu_1\sqrt{2}/s$,

$$= \int_0^{t_1} (1 + t^2)^{-3/2}\, dt = \frac{t_1}{\sqrt{1 + t_1^2}} \qquad (5)$$

where $t_1 = |d_1|/(d_1^2 + d_2^2)^{1/2}$. The value of (5) changes from $\sqrt{2}/2$ to 0 as $|d_2| = |y_1 - y_2|$ increases.

Moreover, contitionally to σ, $r = d_2^2$ is gamma distributed with parameters 1/2 and 1/4 σ^2 i.e.

$$p(r|\mu_2, \sigma) = (2\sigma\sqrt{\pi})^{-1} r^{-1/2} \exp\left\{-\frac{r}{4\sigma^2}\right\}$$

and the posterior distribution of σ after x_1 and x_2 have been observed is

$$p(\sigma|x_1, x_2) = \frac{d_1}{\sqrt{\pi}} \sigma^{-2} \exp\left\{-\frac{d_1^2}{4\sigma^2}\right\}$$

Hence, the posterior density of $r = d_2^2$, after x_1 and x_2 have been observed is the inverted beta

$$p(r|x_1, x_2) = \int_0^\infty p(r|\sigma)\, p(\sigma|x_1, x_2)\, d\sigma = \frac{d_1}{\pi} r^{-1/2} (r + d_1^2)^{-1}$$

so that the posterior density of $t_1 = |d_1|/(d_1^2 + r)^{1/2}$ is

$$p(t_1|x_1, x_2) = p(r|x_1, x_2) \Big/ \left|\frac{dt_1}{dr}\right| = \frac{2}{\pi} \frac{1}{\sqrt{1 - t_1^2}} \qquad (6)$$

which is independent of x_1 and x_2. Combining (5) and (6),

$$p(A|x_1, x_2) = \int_0^1 \frac{t}{\sqrt{1+t^2}} \frac{2}{\pi} \frac{1}{\sqrt{1-t^2}}\, dt =$$

$$= \frac{2}{\pi} \int_0^\infty \frac{t\,dt}{\sqrt{1-t^4}} = \frac{1}{\pi} \int_0^1 \frac{dx}{\sqrt{1-x^2}} = \frac{1}{\pi} \arcsin x \Big|_0^1 = \frac{1}{\pi} \frac{\pi}{2} = \frac{1}{2}$$

as we expected. It may be verified that this result cannot be obtained with any other prior of the relatively invariant class.

The content of this paper may be interpreted as an argument for the use of $d\mu_1, d\mu_2, \ldots, d\mu_k\, d\sigma/\sigma$ as a formal prior measure if one tries to make inferences about *one* of the means of a multinormal homocedastic model without making use of any information other than that provided by the data.

REFERENCES.

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

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

Jeffreys, H. (1939/67): *Theory of Probability (3rd ed.).* Oxford: Clarendon Press.

Kashyap, R.L. (1971): Prior Probability and Uncertainty. *IEEE Trans. Information Theory IT-14*, 641-650.

Lindley, D.V. (1965): *Introduction to Probability and Statistics from a Bayesian Viewpoint. Vol 2: Inference.* Cambridge: University Press.

Stone, M. (1976): Strong inconsistency from uniform priors. *J. Amer. Statist. Assoc. 71*, 114-125.

Villegas, C. (1971): On Haar priors. In *Foundations of Statistical Inference*, ed. V.P. Godambe and D.A. Sprott, pp. 409-414. Toronto, Montreal: Holt, Rinehart and Winston.

Zellner, A. (1971): *An Introduction to Bayesian Inference in Econometrics.* New York: Wiley.