UNIVERSIDADE DE SÃO PAULO

Bayesian estimation of a normal mean parameter using linex loss function and robustness considerations

JOSEMAR RODRIGUES

Nº 06

NOTAS

Instituto de Ciências Matemáticas de São Carlos

Instituto de Ciências Matemáticas de São Carlos

ISSN - 0103-2577

Bayesian estimation of a normal mean parameter using linex loss function and robustness considerations

JOSEMAR RODRIGUES

Nº 06

N O T A S D O I C M S C
Série Estatística

São Carlos
set. / 1993

Bayesian Estimation of a Normal Mean Parameter Using Linex Loss Function and Robustness Considerations

Josemar Rodrigues
ICMSC - USP - 13560 - 970 - São Carlos - SP - Brasil

Abstract

In this paper, an explicit analytical form of the Bayes estimator for the normal location parameter using Linex loss function with a general class of prior distributions is derived. Exact and approximate results based on Pericchi and Smith's paper (1992) are given where the prior is double-exponential or Student t respectively. The results of this paper provide a link between the robust Bayesian analysis for the normal location parameter using the Linex loss function and the squared loss function.

Keywords: Bayesian estimation; Double-exponential prior; Linex loss function; Normal distribution; Robustness; Student t- prior.

1 Introduction

For a normal likelihood with known variance, Zellner (1986) derived an explicit analytical form of the Bayes estimator for the mean using Linex loss function and a normal prior. However, a Bayesian estimator is not readily available when other prior densities are utilized. Pericchi and Smith (1992) derived, under squared error loss, explicit expressions for the posterior mean and variance of the mean θ of a normal distribution with known variance and an arbitrary prior for θ. The purpose of this paper is to follow the same ideas of Pericchi and Smith's paper, but under the Linex loss function. In Bayesian estimation, the loss function and prior distribution play important roles. Symmetric loss functions have been used by several authors including Berger (1980) and Sinha and Kale (1980). Researchers such as Ferguson (1967), Zellner and Geisel (1968), Aitchison and Dunsmore (1975), and Varian (1975) have been pointed out that in some estimation and prediction problems use of symmetric loss functions may be inappropriate. That is, a given positive error may be more serious than a given negative error or vice-versa, e.g. in dam construction, underestimation of a peak water level is much more serious than overestimation. Varian (1975) introduced, in his applied study of real estate assessment, a useful asymmetric loss function known as Linex loss function that rises exponentially on one side of zero and almost linearly on the other side of zero. The Linex loss function is given by

$$L(\Delta) = b\epsilon^{a\Delta} - c\Delta - b, \qquad a, c \neq 0, b > 0. \tag{1}$$

where, $\Delta = \hat{\theta} - \theta$. Also, for a minimum to exist at $\Delta = 0$, we must have $ab = c$, and thus $L(\Delta)$ can be written as

$$L(\Delta) = b[\epsilon^{a\Delta} - a\Delta - 1], \qquad a \neq 0, b > 0. \tag{2}$$

This loss function reduces to squared error loss function for small values of $\mid a \mid$. Zellner (1986) used this loss function for estimating the normal location parameter with normal prior and known variance and for other problems. In this paper, we consider the double-exponential and Student t-prior for a normal location parameter with known variance and some aspects of robust Bayesian analysis in the sense of Pericchi and Smith (1992). Let $\pi(\theta \mid D)$ denotes the posterior density function of θ with respect to $\pi(\theta)$, where D denotes the sample and $\pi(\theta)$ the prior density. Let $E(\cdot \mid D)$ denotes the posterior expectation with respect to $\pi(\theta \mid D)$. The posterior expectation of the Linex loss function in (2) is

$$R_L(\hat{\theta}) = E(L(\Delta) \mid D) = b[e^{a\hat{\theta}}E(\epsilon^{-a\theta} \mid D) - a(\hat{\theta} - E(\theta \mid D)) - 1], \tag{3}$$

and the value of $\hat{\theta}$ that minimizes (3) is

$$\hat{\theta}_B = -(\frac{1}{a})\log(E(\epsilon^{-a\theta} \mid D)), \tag{4}$$

provided, of course, that $E(\epsilon^{-a\theta} \mid D)$ exists and is finite. This involves evaluation of moment generating function for posterior density. The main purpose of this paper is to present a robust Bayesian analysis similar to that of Pericchi and Smith (1992), but applicable to the Linex loss function. The basic results and the Bayes estimator of the normal location parameter θ, taking a normal and diffuse prior for θ using Linex loss function (2), are introduced in Section 2. Section 3 is devoted to the case of a double-exponential prior and the case of a Student t- prior is examined in detail in Section 4. In this paper, the influence of the prior on the Bayes estimator $\hat{\theta}_B$ is discussed in detail when the Bayes estimator under the diffuse prior and the prior location parameter are in conflict. The results of this paper provide a Bayesian analysis for a normal location parameter using the Linex loss function and non-normal priors and state that the influence of these priors is not affected, in the sense of Pericchi and Smith (1992), when replacing the Linex loss function by the squared loss function.

2 Estimation of mean: Basic results

Consider n observations $x_i, i = 1, \ldots, n$ drawn independently from a normal distribution with unknown mean θ and known variance σ^2. It is known that $\overline{X} = n^{-1}(X_1 + \ldots + X_n)$ is a sufficient statistic for θ, so, we replace our original sample D by $\overline{X}$. We shall denote the density corresponding to $Y = \overline{X} \sim N(\theta, \sigma^2/n)$ by $p(y - \theta)$. All the integrals in the following theorem are assumed to converge over the range $(-\infty, \infty)$.

Theorem 1. Define

$$m(\cdot) = \int p(\cdot - \theta)\pi(\theta)d\theta$$

for any $\pi(\theta)$. Further define

$$s(y) = -\frac{\partial\{\log \ m(y)\}}{\partial y}, \quad H(y) = \frac{1}{a}\log(\frac{m(y - a\sigma^2/n)}{m(y)}),$$

$$\overline{S}(y) = -H(y) + \frac{\sigma^2}{n}s(y).$$

Then

$$\text{(a)} \quad \hat{\theta}_B = y - \frac{a\sigma^2}{2n} - H(y) \tag{5}$$

$$\text{(b)} \quad R_L(\hat{\theta}_B) = bz - ab\overline{S}(y) \tag{6}$$

$$\text{(c)} \quad H(y) = \frac{1}{a}\int_{y-a\sigma^2/n}^{y} s(t)dt \tag{7}$$

where $z = a^2\sigma^2/2n$.

Proof.

(a)

It is not hard to see that

$$\begin{aligned} m(y)E(\epsilon^{-a\theta} \mid y) &= \int \epsilon^{-a\theta} p(y-\theta)\pi(\theta)d\theta \\ &= \epsilon^{\frac{a^2\sigma^2}{2n} - ay} m(y - \frac{a\sigma^2}{n}). \end{aligned}$$

Hence, from (4) we have that the Bayes estimator is given by

$$\hat{\theta}_B = -\frac{1}{a}\log E(\epsilon^{-a\theta}) = y - \frac{a\sigma^2}{2n} - \frac{1}{a}\log(\frac{m(y - a\sigma^2/n)}{m(y)}).$$

(b)

From the theorem given in Pericchi and Smith (1992) we have that

$$E(\theta \mid y) = y - \frac{\sigma^2}{n}s(y). \tag{8}$$

It is easily seen that

$$E(\epsilon^{-a\theta} \mid y) = \exp\{-a(y - a\sigma^2/2n)\}\frac{m(y - a\sigma^2/n)}{m(y)},$$

so that, by (3) and (8) the result easily follows.

(c)

The definitions of $H(y)$ and $s(y)$ straightforwardly yield:

$$H(y) = -\frac{1}{a}[log(m(y)) - log(m(y - a\sigma^2/n))] = \frac{1}{a}\int_{y-a\sigma^2/n}^{y} s(t)dt.$$

If a diffuse prior ,$\pi(\theta) \propto$ const., is employed, the optimal estimator in (5) and the posterior expectation of the Linex loss function in (6) are given by:

$$\hat{\theta}_C = y - \frac{a\sigma^2}{2n} \quad (9)$$

$$R_L(\hat{\theta}_C) = bz, \quad (10)$$

as shown in Zellner (1986). Then, as in Pericchi and Smith (1992), the representation of $\hat{\theta}_B$ in (5) provides insight into the influence of the prior distribution on the optimal estimator $\hat{\theta}_B$. That is, if the prior $\pi(\theta)$ is such that $H(y)$ is bounded as a function of $\hat{\theta}_C - \mu$, where μ is the prior mean and y is fixed, then $\hat{\theta}_B - \hat{\theta}_C$ is bounded. We note that this is is not true for the the conjugate prior $\pi(\theta \mid \mu, \tau^2) = \phi(\theta \mid \mu, \tau^2)$, for which $H(y) = \frac{\sigma^2}{n(\tau^2+\sigma^2/n)}(\hat{\theta}_C - \mu)$ and discrepancies between $\hat{\theta}_C$ and prior mean μ can lead to unbounded departures of $\hat{\theta}_B$ from $\hat{\theta}_C$. For this normal prior, the optimal estimator of θ can be rewritten as

$$\begin{aligned} \hat{\theta}_B &= \frac{y}{1+\lambda} + \frac{\lambda}{1+\lambda}\mu - \frac{a\sigma^2}{2n(1+\lambda)}, & (11) \\ &= E(\theta \mid y) - \frac{a}{2}Var(\theta \mid y), \quad \text{where} & (12) \\ \lambda &= \frac{\sigma^2}{n\tau^2}. \end{aligned}$$

The expression above were obtained by Zellner (1986) without considering the robustness problem.

The next two results provide an interesting insight into the behavior of the optimal estimator to the prior distribution using the Linex loss function.

Theorem 2.

(a) If $s(t)$ is continuos on $J = [y - a\sigma^2/n, \quad y]$, $(a > 0)$. then

$$\begin{aligned} H(y) &= \frac{\sigma^2}{n}s(\hat{\theta}_C) \quad \text{and} & (13) \\ \hat{\theta}_B &= E[\theta \mid \hat{\theta}_C]. \end{aligned}$$

(b) If $s(t)$ is increasing on J, then there exists a number c on J such that

$$H(y) = \frac{\sigma^2}{n}[w_c s(y - a\sigma^2/n) + (1 - w_c)s(y)], \quad \text{where} \quad (14)$$

$$w_c = \frac{c - y + a\sigma^2/n}{a\sigma^2/n}.$$

Proof.

(a) The results trivially follows from (7) with a straightforward application of The First Mean Value Theorem
(b) The result trivially follows from (7) with a straightforward application of the Second Mean Value Theorem.
These two results provide an important insight into the behavior of $H(y)$. The first one shows,when the prior mean is in conflict with $\hat{\theta}_C$, that the sensitivity to the prior is the same under the Linex loss function and the squared loss function. That is, the parameter a of the Linex loss function does not affect the influence of the prior. The second one shows that $H(y)$ is a weighted average of $s(y)$ and $s(y - a\sigma^2/n)$. In particular, for all c in J, we observe from (14) that

$$s(y - a\sigma^2/n) \leq H(y) \leq s(y),$$

which states that the influence of prior mean μ on $\hat{\theta}_B$ is bounded in J. Also, if $s(t)$ is bounded for any t, we see from (7) that $H(t)$ is bounded.

It will be shown in the next two sections, that the lack of robustness property of the normal prior is not shared by the double-exponential and Student t-priors. We see that (5) and (6) exhibit the modification of $\hat{\theta}_C$ and $R_L(\hat{\theta}_C)$ due to the prior input. To study the sensibility of the prior input we consider $H(y)$ or $s(y)$ as a function of $\hat{\theta}_C - \mu$ or $y - \mu$, respectively. The sensibility of the prior distribution is quantified by the behavior of $H(y)$ when using the Linex loss function and by $s(y)$ when using the squared loss function. Result (7) provides an interesting link between the sensibility of the prior input under the Linex loss function and the squared loss function. We show in the next two sections, from (13) and Pericchi and Smith's results (1992), that the sensibility of the double-exponential prior and the Student t-prior is not affected when replacing the squared loss function by the Linex loss function. In conclusion, the robustness for these non-normal priors with respect to $\hat{\theta}_B$ is not affected by the parameters a and b of the Linex loss function (2).

3 Exact representation of $\hat{\theta}_B$ with double-exponential prior

Pericchi and Smith (1992) derived, under regularity conditions, explicit expressions for the posterior mean and variance of the normal mean parameter θ with known variance σ^2 and a double-exponential prior for θ with mean and variance μ and ν^2, respectively, that is, a prior distribution for θ with density function

$$\pi(\theta) = \frac{1}{\sqrt{2}\nu}\exp\{-\frac{\sqrt{2}}{\nu} \mid \theta - \mu \mid\}.$$

These authors, in a straightforward manipulation established that

$$\begin{aligned} m(y) &= \frac{a^*}{\nu\sqrt{2}}\{F(y) + G(y)\}, \qquad (15) \\ F(y) &= \exp\{c(y)\}\Phi\{\frac{\sqrt{n}}{\sigma}(\mu - y - b^*)\}, \\ G(y) &= \exp\{-c(y)\}\Phi\{-\frac{\sqrt{n}}{\sigma}(\mu - y + b^*)\}, \\ a^* &= \exp\{\frac{\sigma^2}{n\nu^2}\}, \quad b^* = \frac{\sigma^2\sqrt{2}}{n\nu}, \quad c(y) = \frac{\sqrt{2}}{\nu}(y - \mu), \end{aligned}$$

where $\Phi(\ )$ denotes the standard normal acumulative density function. In particular, they established that

$$-b^*\frac{n}{\sigma^2} \leq s(y) \leq b^*\frac{n}{\sigma^2}. \qquad (16)$$

From (7) or (13), a straightforward computation yields

$$\begin{aligned} -b^* \leq \ & H(y) \ \leq b^*, \quad \text{or,} \\ -b^* + \hat{\theta}_C \leq \ & \hat{\theta}_B \ \leq b^* + \hat{\theta}_C, \end{aligned}$$

which establishes that the influence of μ is bounded. Assuming a squared loss function, Pericchi and Smith (1992) obtained the same b^* bounds for $E[\theta \mid y] - y$. Using (5) and (15), the optimal estimator relative to the Linex loss function (1) and the double-exponential prior is

$$\hat{\theta}_B = \hat{\theta}_C - \frac{1}{a}log\{\frac{F(y - a\sigma^2/n) + G(y - a\sigma^2/n)}{F(y) + G(y)}\}.$$

4 Approximate representation of $\hat{\theta}_B$ with a Student t-prior

If we chose a Student t-prior for θ, with location and scale μ and τ, and degrees of freedom α, then the prior density has the form

$$\pi(\theta) \propto \{1 + \frac{(\theta - \mu)^2}{\alpha\tau^2}\}^{-\frac{\alpha+1}{2}}.$$

It will be convenient to denote this density by $t(\theta \mid \alpha, \mu, \tau^2)$. Assuming the squared loss function, Pericchi and Smith studied the model

$$\begin{aligned} y &\sim N(\theta, \sigma^2/n), \\ \theta &\sim t(\theta \mid \alpha, \mu, \tau), \end{aligned}$$

where the parameters μ,τ,α and σ^2 are supposed known. These authors developed an insightful approximation to $s(y)$ given by

$$s(y) \doteq \frac{(\alpha + 1)(y - \mu)}{\alpha\tau^2 + (y - \mu)^2}. \tag{17}$$

Thus, using (5), (7) and (17), the optimal estimator relative to the Linex function and the Student t-prior $t(\theta \mid \alpha, \mu, \tau)$ is

$$\hat{\theta}_B \doteq y - \frac{a\sigma^2}{2n} - \frac{(\alpha + 1)}{2a} log\{1 - \frac{\sigma^2}{n}\frac{2a(y - a\sigma^2/2n - \mu)}{\alpha\tau^2 + (y - \mu)^2}\}. \tag{18}$$

Also, from (13) we observe that $H(y)$ in (18) behaves like $s(\hat{\theta}_C)$ in (17) and is not affected by the parameter a. Figure 1 shows the influence of the discrepancy, computed from (18), between the prior mean μ and the optimal estimator $\hat{\theta}_C$ on the optimal estimator (18) for, $\alpha = 9$, $n = 10$, $\tau = \sigma = 1$ and different values for a. The figure shows that the qualitative form of this influence is equal to that obtained by Pericchi and Smith and that this influence measured by $H(y)$ is not affected by the value of a. In conclusion, the sensibility of this prior for the normal mean parameter is not affected when replacing the squared loss function by the Linex function. Future extensions of the results of this paper, when the likelihood has an exponential family form and any prior using the Linex loss function, will be given elsewhere.

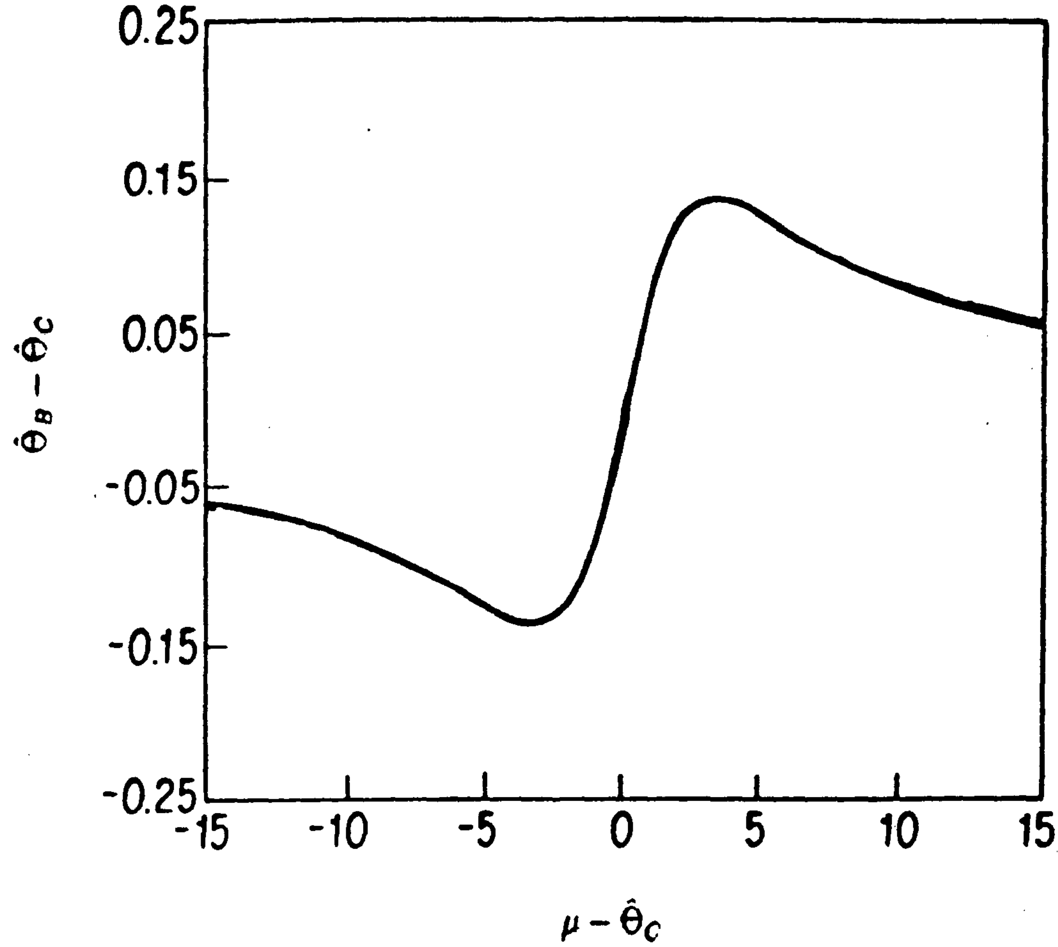

Figure 1: Student t-prior: $n = 10$, $\alpha = 9$, $\sigma = \tau = 1$, $a = 4, -4, 2, -2$.

REFERENCES

Aitchison, J. and Dunsmore, I. R. (1975), Statistical Predicton Analysis. London: Cambridge University Press.

Berger, James, O. (1980), Statistical Decision Theory. Foundations, Concepts and Methods, New York: Springer - Verlag.

Pericchi, L.R. and Smith, A.F.M. (1992), Exact and approximate posterir moments for a normal location parameter. J. R. Statist. Soc. B, 54, 3, 793-804.

Sinha, S.K. and Kale, B.K. (1980), Life Testing and Reliability Estimation, Wiley Easten Limited, New Delhi.

Ferguson, Thomas, S. (1967), Mathematical Statistics: A Decision Theoretic Approach, New York: Academic Press.

Varian, H. R. (1975), A Bayesian Approach to Real Estate Assessment. In Studies in Bayesian Econometrics and Statistics in Honor of L.J. Savage, Eds. S.E. Feinberg and A. Zellner, Amsterdam: North Holland, 195-208.

Zellner, A. (1986), Bayes estimation and prediction using asymmetric loss functions. JASA, 81, 446-451.

Zellner, A. and Geisel, Martin S. (1968), Sensitivity of control to uncertainty and form of the criterion function, in The Future of Statistics, ed. Donald G. Watts, New York : Academic Press, 269-289.

N O T A S D O I C M S C

Serie Estatistica

Nº 005/93 - ACHCAR, J.A.; FOGO, J.C. - Some useful reparametrizations for the reliability function considering the exponential distribution and censored lifetime data

Nº 004/93 - LEME, E.C. - A m.v.u. estimator vs a m.l. estimator: a way of comparing their precisions by using the Pareto Distribution

Nº 003/93 - ACHCAR, J.A. - Some practical aspects of approximate Bayesian inference

Nº 002/93 - RODRIGUES, J.; BABA, M.Y. - Bayesian estimation of a simple regression model with measurement errors

Nº 001/93 - ACHCAR, J.A. - Some aspects of reparametrization for statistical model