BAYESIAN ESTIMATORS OF THE LOCATION PARAMETER OF THE NORMAL DISTRIBUTION WITH UNKNOWN VARIANCE

Janet van Niekerk*[1] and Andriette Bekker[1]

[1]Department of Statistics, University of Pretoria, 0002, Pretoria, South Africa.

ABSTRACT

The estimation of a location parameter of the normal distribution has been widely discussed and applied in various situations. The Bayes estimators under Linear Exponential (LINEX) loss are functions of the moment generating function of the Student's t-distribution and are therefore unknown. In this paper the explicit Bayes estimators of the location parameter of the normal model for two different loss functions, the re[1]flected normal loss and the LINEX loss functions are proposed and evaluated. The performances of these estimators are evaluated using Monte Carlo simulation.

1. INTRODUCTION

In this paper the exact Bayesian estimators under the reflected normal loss function and the LINEX loss function will be derived under the assumption that we have a simple random sample from the normal population with unknown variance. In literature there has been a great amount of discussion and investigation of the normal model, see amongst others Giles(2002), DeGroot(1970) and Zellner(1971).

The quadratic loss function is the most well-known loss function and has been traditionally used by economists and decision-theoretic statisticians, but this loss function does have a couple of shortcomings. Several authors (eg. Leon and Wu, 1992) have suggested that the traditional quadratic loss function is insufficient in assessing quality and its improvement. There was thus a need for a new loss function that could uphold these standards and specifications. Taguchi (1986) used a modified form of the traditional quadratic loss function to show that there is a need to consider the proximity to the target while assessing quality and it is defined as,

$$£(\theta,\tau) = £_{RL} \equiv K\left(1 - e^{\frac{-(\theta-\tau)^2}{2\gamma^2}}\right) \tag{1}$$

where $\gamma \in R, K \geq 0$.

The LINEX loss function is an unbounded asymmetric loss function and hence the loss has no upper limit. The most commonly used quadratic loss function cannot differentiate between overestimation and underestimation due to the symmetric nature of this particular loss function.

*Janet van Niekerk is the corresponding author.

Varian (1975) introduced the LINEX loss function in an effort to overcome this problem. The parameters can be specified to accommodate the seriousness of overestimation or underestimation. The LINEX loss function is defined as

$$£(\theta,\tau) = £_L \equiv be^{a(\theta-\tau)} - a(\theta-\tau) - 1 \qquad (2)$$

where $-\infty < a < \infty$ *and* $b > 0$. If $a > 0$ then overestimation is regarded as more serious than underestimation, and vice versa. If the costs of a building project is to be estimated then certainly an underestimation of costs would have severe implications seeing that the contractor will have to carry the costs on himself and hence the LINEX loss function with $a < 0$ would be used.
Further properties of this loss function are discussed in Varian (1975) and Zellner (1986) amongst others.

In section 2 the Bayesian estimators for the location parameter of the normal model with unknown variance are proposed for the reflected normal loss and the LINEX loss respectively. The proposed estimators are then evaluated with Monte Carlo estimation in section 3.

2. BAYESIAN ESTIMATORS

2.1. The reflected normal loss function

If σ^2 is unknown, then the conjugate prior for the location parameter is assumed to be a normal distribution and for the variance it is assumed to be the inverse or inverted gamma distribution. The inverted gamma distribution is chosen as the prior distribution for the population variance since it produces mathematically favourable results (Raiffa and Schlaifer, 1961). Conjugate priors have a very powerful purpose when updating prior belief is the main focus. At each stage of collecting new data, the posterior of previous analysis is used as prior for the next stage of the Bayesian analysis.
Let $\mu \mid \sigma^2 \backsim N(\mu^1, \frac{\sigma^2}{n_0})$ and $\sigma^2 \backsim IG(\frac{v_0}{2}, \frac{S_0}{2})$.
The joint prior density function is:

$$f(\mu,\sigma^2) \propto (\sigma^2)^{-\frac{v_0+1}{2}-1} e^{\frac{1}{2\sigma^2}(S_0+n_0(\mu-\mu^1)^2)} \qquad (3)$$

Subsequently, the likelihood function and posterior density are given, respectively, by

$$L(\mu,\sigma^2 \mid x_1, x_2, \dots, x_n) \propto (\sigma^2)^{-\frac{n}{2}} e^{\frac{1}{2\sigma^2}(n(\mu-\bar{x})^2+S)} \qquad (4)$$

and

$$f(\mu,\sigma^2 \mid x_1, x_2, \dots, x_n) \propto (\sigma^2)^{-\frac{v_0+1+n}{2}-1} e^{\frac{1}{2\sigma^2}Q(\mu)} \qquad (5)$$

where $Q(\mu) = S + S_0 + n(\mu-\bar{x})^2 + n_0(\mu-\mu^1)^2, S = \sum_{i=1}^{n} x_i{}^2 - n\bar{x}\ .^2$

The marginal posterior density of μ follows as,

$$p(\mu \mid x_1, x_2, \ldots, x_n) \propto [\{(\frac{\mu - \frac{n\bar{x}+n_0\mu^1}{n+n_0}}{\sqrt{\frac{\frac{S_1}{n+n_0}}{\nu_0+n}}})^2 \div (\nu_0+n)\} + 1]^{\frac{(\nu_0+n)+1}{2}} \quad (6)$$

where $S_1 = S + S_0 + \frac{nx^2+n_0(\mu^1)^2}{n+n_0} - (\frac{nx+n_0\mu^1}{n+n_0})^2$.

Now let $\frac{\mu - \frac{n\bar{x}+n_0\mu^1}{n+n_0}}{\sqrt{\frac{\frac{S_1}{n+n_0}}{\nu_0+n}}} = x$ then from (6),

$$p(\mu \mid x_1, x_2, \ldots, x_n) \propto [\frac{x^2}{\nu_0+n} + 1]^{\frac{(\nu_0+n)+1}{2}} \quad (7)$$

From (7) it can be seen that the marginal posterior distribution of μ is the noncentral Student's t-distribution with $\nu_0 + n$ degrees of freedom, location parameter of $\frac{n\bar{x}+n_0\mu^1}{n+n_0}$ and scale parameter $\frac{\frac{S_1}{n+n_0}}{\nu_0+n}$, hence $\frac{\mu - \frac{n\bar{x}+n_0\mu^1}{n+n_0}}{\sqrt{\frac{\frac{S_1}{n+n_0}}{\nu_0+n}}} \mid x_1, x_2, \ldots, x_n$ has a central t-distribution.

Then the Bayesian estimator obtained is,

$$\hat{\mu}_B = \mu \quad (8)$$

However, this estimator given in (8) in unsatisfactory since μ is the unknown parameter. A feasible estimator for μ is $\hat{\mu}_{FB} = \mu^1$ which states that the Bayesian estimator is equal to the mean of the prior distribution of $\mu \mid \sigma^2$. An improvement on this feasible estimator can be made by incorporating the sample information into the prior information as opposed to applying only the prior information. An improved feasible estimator is then given by the $\hat{\mu}_{FB}^1$ which minimizes the posterior density function and hence minimizes the product of the likelihood function (see (4)) and the prior density function (see (3)) i.e. solve for the $\hat{\mu}_{FB}^1$ which minimizes

$$L(\hat{\mu}_{FB}^1, \sigma^2 \mid x_1, x_2, \ldots, x_n) \times f(\hat{\mu}_{FB}^1 \mid \sigma^2 = r^2) = e^{-\frac{n^0}{2\sigma^2}(\hat{\mu}_{FB}^1 - \mu^1)^2 - \frac{1}{2\sigma^2}\{n(\hat{\mu}_{FB}^1 - \bar{x})^2 + S\}}$$

Hence,

$$\hat{\mu}_{FB}^1 = \frac{n_0\mu^1 + n\bar{x}}{n_0+n} \quad (9)$$

This feasible estimator of μ is a weighted average between the mean of the prior distribution and the sample mean. Note that $\hat{\mu}_{FB}$ and $\hat{\mu}_{FB}^1$ are unbiased estimators of μ.

Remark 1. Jeffreys' prior is used as a non-informative prior (Lee,1989),

$$f(\mu, \sigma^2) \propto \frac{1}{\sigma^2} \quad (10)$$

The feasible Bayesian estimator of μ for the non-informative prior is

$$\hat{\mu}_{FB}^1 = \bar{x} \quad (11)$$

2.2. LINEX loss function

From equation (6) it can be seen that the function of μ has the Student's t-distribution with $\nu_0 + n$ degrees of freedom hence

$$\frac{\mu - \frac{n\bar{x}+n_0\mu^1}{n+n_0}}{\sqrt{\frac{\frac{S_1}{n+n_0}}{\nu_0+n}}} \mid x_1, x_2, \ldots, x_n \sim t(\nu_0 + n)$$

The Bayesian estimator arrived at is,

$$\hat{\mu}_B = -\frac{1}{a} ln[b\sqrt{\frac{S_1}{n+n_0(\nu_0+n)}} M_t\left(-a\sqrt{\frac{S_1}{n+n_0(\nu_0+n)}}\right)] + \frac{n\bar{x}+n_0\mu^1}{n+n_0} \quad (12)$$

The Bayesian estimator from (12) is unknown since $M_t\left(-a\sqrt{\frac{S_1}{n+n_0(\nu_0+n)}}\right)$ is the moment generating function of the Student's t-distribution in the point $-a\sqrt{\frac{S_1}{n+n_0(\nu_0+n)}}$, which is unknown.

Remark 2. The Bayesian estimator for the non-informative prior is

$$\hat{\mu}_B = -\frac{1}{a} ln[b\sqrt{\frac{S}{n(n-1)}} M_t\left(-a\sqrt{\frac{S}{n(n-1)}}\right)] + \bar{x} \quad (13)$$

from (10).

Note that the moment generating function in (12) can be written as:

$$M_t\left(-a\sqrt{\frac{S_1}{n+n_0(\nu_0+n)}}\right) = E[e^{-ta\sqrt{\frac{S_1}{n+n_0(\nu_0+n)}}}] \quad (14)$$

where $t \sim t(n-1)$.

Now consider the sample at hand $x_1, x_2, \ldots, x_n$. The sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \frac{\sigma^2}{n})$.

Therefore $(X_i - \bar{X}) \sim N(0, {\sigma_1}^2)$. Hence,

$$\begin{aligned}
{\sigma_1}^2 &= Var(X_i - \bar{X}) \\
&= VarX_i - 2Cov(X_i, \bar{X}) + Var\bar{X} \\
&= VarX_i - 2Cov(X_i, \frac{1}{n}\sum_{j=1}^{n} X_j) + Var\bar{X} \\
&= VarX_i - 2(1/n)Cov(X_i, X_i) + Var\bar{X} \\
&= VarX_i - 2(1/n)VarX_i + Var\bar{X} \\
&= \sigma^2 - 2\frac{\sigma^2}{n} + \frac{\sigma^2}{n} \\
&= \frac{\sigma^2(n-1)}{n}
\end{aligned}$$

and therefore $\frac{X_i - \bar{X}}{\sqrt{\frac{\sigma^2(n-1)}{n}}} \sim N(0,1)$. Note that if $Z \sim N(0,1)$ independently of $W \sim \chi^2(\nu)$ then $B = \frac{Z}{\sqrt{\frac{W}{\nu}}} \sim t(\nu)$. It is known that the sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is a bounded complete sufficient statistic for σ^2 and also $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$.

Note that,

$$\begin{aligned}
X_i - \bar{X} &= X_i - \mu + \mu - \bar{X} \\
&= (X_i - \mu) + (\mu - \bar{X}) \\
&= f(X_i - \mu)
\end{aligned}$$

and $(X_1 - \mu, X_2 - \mu, \ldots, X_n - \mu)$ are ancillary statistics for σ^2.

Therefore by Basu's theorem (Basu, 1955), $(X_1-\mu, X_2-\mu, \ldots, X_n-\mu)$ and s^2 are independent and also $X_i-\bar{X}$ and $\frac{(n-1)s^2}{\sigma^2}$ for $i=1,\ldots,n$ and then $\frac{X_i-\bar{X}}{\sqrt{\frac{\sigma^2(n-1)}{n}}}$ and $\frac{(n-1)s^2}{\sigma^2}$ are independent.

Define

$$d = \frac{\frac{X_i-\bar{X}}{\sqrt{\frac{\sigma^2(n-1)}{n}}}}{\sqrt{\frac{\frac{(n-1)s^2}{\sigma^2}}{n-1}}}$$

$$= \frac{X_i-\bar{X}}{\sqrt{\frac{s^2(n-1)}{n}}} \qquad (15)$$

Then $d \sim t(n-1)$. Therefore t in equation (14) and d in equation (15) is equal in distribution. From this equality in distribution,

$$E[e^{tc}] = E[e^{dc}]$$

see Randles and Wolfe (1955).

Now,

$$M_t\left(-a\sqrt{\frac{S_1}{n+n_0(\nu_0+n)}}\right) = E_t\left[e^{-ta\sqrt{\frac{S_1}{n+n_0(\nu_0+n)}}}\right]$$

$$= E_d\left[e^{-da\sqrt{\frac{S_1}{n+n_0(\nu_0+n)}}}\right]$$

$$= E_X[e^{-\frac{X_i-\bar{X}}{\sqrt{\frac{s^2(n-1)}{n}}}a\sqrt{\frac{S_1}{n+n_0(\nu_0+n)}}}]$$

An unbiased estimator for the expression in (14) is:

$$\frac{1}{n}\sum_{i=1}^{n} e^{-\frac{X_i-\bar{X}}{\sqrt{\frac{s^2(n-1)}{n}}}a\sqrt{\frac{S_1}{n+n_0(\nu_0+n)}}}$$

since $E_X[\frac{1}{n}\sum_{i=1}^{n} e^{-\frac{X_i-\bar{X}}{\sqrt{\frac{s^2(n-1)}{n}}}a\sqrt{\frac{S_1}{n+n_0(\nu_0+n)}}}] = E_X[e^{-\frac{X_i-\bar{X}}{\sqrt{\frac{s^2(n-1)}{n}}}a\sqrt{\frac{S_1}{n+n_0(\nu_0+n)}}}]$. An unbiased estimator for the expression in (12) can then be written as:

$$\hat{\mu}_B = -\frac{1}{a}ln[\frac{b}{n}\sqrt{\frac{S_1}{n+n_0(\nu_0+n)}}\sum_{i=1}^{n} e^{-\frac{X_i-\bar{X}}{\sqrt{\frac{s^2(n-1)}{n}}}a\sqrt{\frac{S_1}{n+n_0(\nu_0+n)}}}] + \frac{n\bar{x}+n_0\mu^1}{n+n_0}. \qquad (16)$$

Equivalently (13) can be written as:

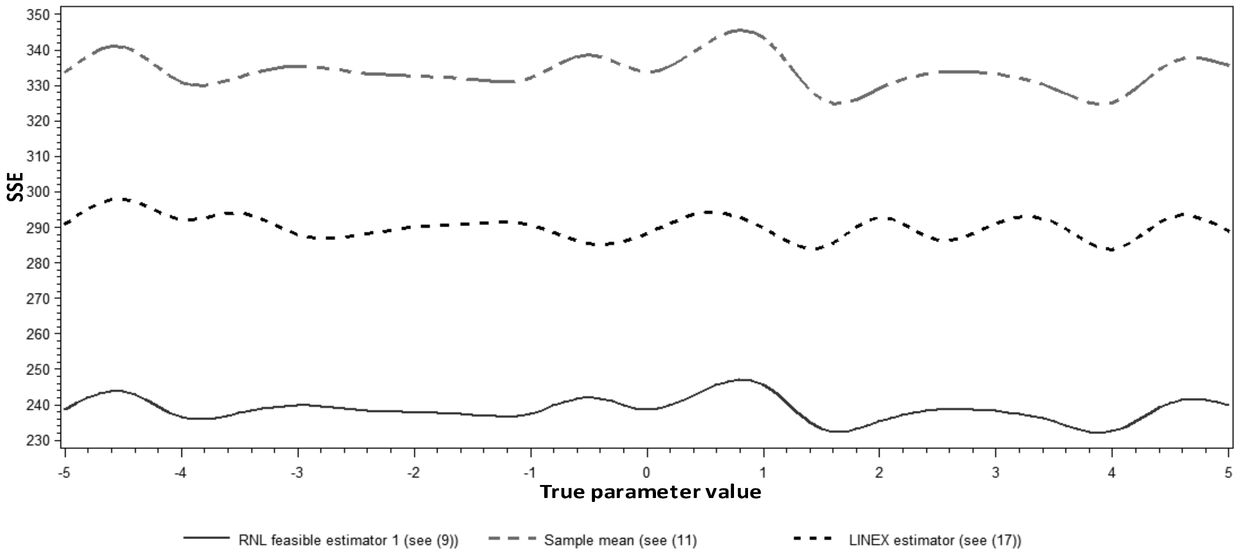
$$\hat{\mu}_B = -\frac{1}{a}ln[\frac{b}{n}\sqrt{\frac{S}{n(n-1)}}\sum_{i=1}^{n} e^{-\frac{X_i-\bar{X}}{\sqrt{\frac{s^2(n-1)}{n}}}a\sqrt{\frac{S}{n(n-1)}}}] + \bar{x} \qquad (17)$$

To evaluate (16) and (17) Monte Carlo simulation is used.

3. EVALUATION OF BAYESIAN ESTIMATORS

The estimators are evaluated by means of their individual sum of squared errors. The hyper parameters are chosen in such a way that the expected value of the prior distribution for the parameter in question is equal to the specific parameter of the population distribution. The number of simulations used is 10000. The sample size that was used was 15.

Figure 1. The sum of squared errors for 10000 simulations with n=15.



It is evident from the graph that estimator (17) has a lower sum of squared errors than the feasible estimator (11) which is the commonly known maximum likelihood estimator of μ. However, the feasible estimator (9) has the lowest sum of squared errors. There are thus two estimators with better performance than the maximum likelihood estimator (the sample mean) which is also the Bayesian estimator of μ under squared error loss.

4. CONCLUSION

In this paper we considered again the classical case of the estimation of the location parameter of the normal model with unknown variance. The exact expressions of the Bayesian estimators derived under reflected normal loss are as follows:

Table 1.

Prior distribution	**Bayesian estimator**
Conjugate prior	$\hat{\mu}_{FB} = \mu^1$ and $\hat{\mu}_{FB}^1 = \frac{n_0\mu^1+n\bar{x}}{n_0+n}$
Non-informative prior	$\hat{\mu}_{FB}^1 = \bar{x}$

The explicit Bayesian estimators obtained under LINEX loss are as follows:

Table 2.

Prior distribution	**Bayesian estimator**
Conjugate prior	$\hat{\mu}_B = -\frac{1}{a} ln[\frac{b}{n}\sqrt{\frac{S_1}{n+n_0(v_0+n)}}\sum_{i=1}^{n} e^{-\frac{X_i-\bar{X}}{\sqrt{\frac{s^2(n-1)}{n}}}a\sqrt{\frac{S_1}{n+n_0(v_0+n)}}}] + \frac{n\bar{x} \quad + n_0\mu^1}{n+n_0}$
Non-informative prior	$\hat{\mu}_B = -\frac{1}{a} ln[\frac{b}{n}\sqrt{\frac{S}{n(n-1)}}\sum_{i=1}^{n} e^{-\frac{X_i-\bar{X}}{\sqrt{\frac{s^2(n-1)}{n}}}a\sqrt{\frac{S}{n(n-1)}}}] + \bar{x}$

with $S_1 = S + S_0 + \frac{n\bar{x}^2+n_0(\mu^1)^2}{n+n_0} - (\frac{nx+n_0\mu^1}{n+n_0})^2$ and $S = \sum_{i=1}^{n} x_i^{\ 2} - n\bar{x}\ .^2$

The Bayesian estimators in the case of unknown variance under LINEX loss from both the conjugate prior and the non-informative prior are functions of the moment-generating function of the Student's t-distribution, and were therefore unknown previously.

From the simulation study it is evident that the estimator

$$\hat{\mu}_B = -\frac{1}{a} ln[\frac{b}{n}\sqrt{\frac{S_1}{n+n_0(v_0+n)}}\sum_{i=1}^{n} e^{-\frac{X_i-\bar{X}}{\sqrt{\frac{s^2(n-1)}{n}}}a\sqrt{\frac{S_1}{n+n_0(v_0+n)}}}] + \frac{n\bar{x}+n_0\mu^1}{n+n_0}$$

under LINEX loss and the estimator

$$\hat{\mu}_{FB}^1 = \frac{n_0\mu^1+n\bar{x}}{n_0+n}$$

under reflected normal loss has a lower sum of squared errors than the sample mean, and are thus preferred. There is therefore an estimator for μ proposed for each loss function that is preferred above the widely used sample mean.

5. ACKNOWLEDGEMENTS

The authors would like to thank Proff. Arashi, Balakrishnan and Roux for their discussion on this topic. This work is based upon research supported by the National Research Foundation, Incentive programme.

6. REFERENCES

Akdeniz, F. (2004). New biased estimators under LINEX loss function. *Statistical Papers*, 45, 175-190.

Basu, D. (1955). On statistics independent of a complete sufficient statistic, *Sankhyā*, 15, 377-380.

DeGroot, M.H. (1970). *Optimal statistical decisions*, McGraw-Hill Inc., United States of America.

Giles, D.E.A. (2002). Preliminary-Test and Bayesian Estimation of a Location Parameter Under 'reflected normal' loss, *Statistics textbooks and monographs*, 165, 287-304.

Lee, P.M. (1989). *Bayesian statistics:An introduction,* Oxford University Press, New York.

Leon, R.V. and Wu, C.F.J. (1992). A theory of performance measures in parameter design. *Statist. Sinica*, 2(2), 335-357.
Parsian, A. (1990). Bayesian estimation using a LINEX loss function. *J.Sci.I.R.Iran*, 1, 305-307.
Raiffa, H. and Schlaifer, R. (1961). *Applied statistical decision theory*, Harvard University Press, Cambridge, MA.
Randles, R.H. and Wolfe, D.A. (1979). *Introduction to The Theory of Nonparametric Statistics*. John Wiley and Sons, New York.
Spiring, F.A. (1993). The reflected normal loss function, *The Canadian Journal of Statistics*, 21(3), 321-330.
Taguchi, G. (1986). *Introduction to Quality Engineering: Designing Quality into Products and Processes*. Kraus, White Plains, New York.
Van Niekerk, J. (2011). Preliminary-test and Bayesian estimation of a location parameter of the normal distribution under different loss functions. Honours. South Africa: University of Pretoria.
Varian, H.R. (1975). A Bayesian approach to real estate assessment. In: Feinberg, Zellner, A., eds. *Studies in Bayesian Econometrics in honor of L.J. Savage*. Amsterdam: North Holland, 195-208.
Zellner, A. (1971), *An Introduction to Bayesian Inference in* Econometrics, Wiley, New York.