Journal of the Korean
Data & Information Science Society
2006, Vol. 17, No. 4, pp. 1365 ~ 1374

An Objective Bayesian Inference for the Difference between Two Normal Means

Eun Jin Jang[1] · Dal Ho Kim[2] · Kyeong Eun Lee[3]

Abstract

In this paper, we consider a decision-theoretic oriented, objective Bayesian inference for the difference between two normal means with known variances. We derive the Bayesian reference criterion as well as the intrinsic estimator and the credible region which correspond to the intrinsic discrepancy loss and the reference prior. We show the similarity between derived two-sample results and the results for the one-sample case in Bernardo(1999).

Keywords : Bayesian reference criterion, Credible region, Intrinsic estimator, Intrinsic expected loss, Reference prior

1. Introduction

Statistical inference problems might concerned with point estimation, interval estimation, and precise hypothesis testing. Conventional frequentist solutions to those problems typically need ad hoc modifications when dealing with non-regular models, and may prove to be misleading. For example, the asymptotic normality of maximum likelihood estimators(MLE) is not necessarily hold when the sampling distribution is not regular, and frequenntly ad-hoc modifications are needed, reducing the generality of the procedure. Moreover, a number of drawbacks have been exposed about the use of p-values in sharp hypothesis testing (see e.g.

1) Doctoral Course, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea
E-mail: ejjang@knu.ac.kr

2) Corresponding Author : Professor, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea
E-mail: dalkim@knu.ac.kr

3) Full-time Lecturer, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea
E-mail: artlee@knu.ac.kr

Berger and Selke, 1987; Selke *et al.*, 2001).

From a decision-theoretic Bayesian standpoint, any of these problems may be posed as the sampling model for observable data, a prior distribution describing the decision maker's prior beliefs about parameter, the space of possible actions, and a loss function that measures the consequences of deciding to act according to choice of the action. Bayes decision rule which minimizes the posterior expected loss is then optimal for the specific problem at hand. But it is well known that the solution may dramatically depend both on the choice of the loss function and on the choice of the prior distribution.

From an objective Bayesian perspective, it is well known that the irrespective use of flat priors may be misleading. Also Jeffrey' prior may not be defined when the assumed model is non-regular. Furthermore, the decision oriented objective Bayesian approach to point estimation using conventional loss functions (e.g., quadratic, zero-one and linear) produces non-invariant solutions. Regarding sharp hypothesis testing, the use of improper priors may lead to indeterminate answers. The commonly used tool to overcome this problem, conventional Bayes factors (Jeffreys, 1961) suffer from Jeffrey-Lindley-Bartlett paradox (Bartlett, 1957; Lindley, 1957). Various attempts to overcome this difficulties have been made, including intrinsic Bayes factors (Berger and Pericchi, 2001), fractional Bayes factors (O'Hagan, 1997) and neutral Bayes factors (Robert and Caron, 1996). However, the resulting factors do not necessarily correspond to any prior, thus being open to criticism.

Recently Bernardo(1999) and Bernardo and Rueda(2002) introduced a new model selection criterion, called the Bayesian Reference Criterion(BRC). Bernardo takes a decision theoretic approach to developing an objective Bayes solution to test for nested hypotheses. Furthermore, Bernardo and Juarez(2003) addressed intrinsic point estimation and Bernardo(2005) defined intrinsic credible region based on the information-theory based loss function as well as reference prior from an objective Bayesian viewpoint. The procedures merge the use of the reference algorithm (Berger and Bernardo, 1992; Bernardo, 1979) to derive non-informative priors, with the intrinsic discrepancy (Bernardo and Rueda, 2002; Bernardo and Juarez, 2003) as loss function to obtain an objective answer for statistical problem.

The normal distribution is probably the most commonly used parametric distribution among other special distributions. The two-sample comparison under the normal setup is the most well known methodology in data analysis. In this paper, we consider objective Bayesian answer to the problems of sharp hypothesis testing as well as both point and region estimation for the difference between two normal means with known variances.

The paper is organised as follows. Section 2 gives a review of the reference -intrinsic methodology to derive a decision rule for sharp hypothesis testing, and to provide a point estimator and credible region. In Section 3, we derive the Bayesian reference criterion, the intrinsic estimator and the credible region for the difference between two normal means with known variances and give simulation results. Some concluding remarks are given in Section 4.

2. Preliminaries

Suppose that available data $\boldsymbol{x}$ consist of a random sample $\boldsymbol{x} = x_1, \cdots, x_n$ from the family $M \equiv p(\mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\lambda}), \mathbf{x} \in X, \boldsymbol{\theta} \in \Theta, \boldsymbol{\lambda} \in \Lambda$, where $\boldsymbol{\theta}$ is some vector of interest and $\boldsymbol{\lambda}$ is some vector of nuisance parameters.

The intrinsic discrepancy loss $\delta_{\boldsymbol{x}} \tilde{\boldsymbol{\theta}}, (\boldsymbol{\theta}, \boldsymbol{\lambda})$, introduced by Bernardo and Rueda (2002), is basically used to measure the "distance" between the probability density $p(\boldsymbol{x} | \boldsymbol{\theta}, \boldsymbol{\lambda})$ and the family of probability densities $p(\mathbf{x} | \tilde{\boldsymbol{\theta}}, \boldsymbol{\lambda}), \boldsymbol{\lambda} \in \Lambda$, defined as

$$\delta_{\boldsymbol{x}}\{\tilde{\Theta}, (\Theta, \boldsymbol{\lambda})\} = \inf_{\tilde{\boldsymbol{\lambda}} \in \Lambda} \delta\{p(\boldsymbol{x} | \tilde{\Theta}, \tilde{\boldsymbol{\lambda}}), p(\boldsymbol{x} | \Theta, \boldsymbol{\lambda})\} \tag{2.1}$$

where $\delta\{p(\boldsymbol{x} | \tilde{\Theta}, \tilde{\boldsymbol{\lambda}}), p(\boldsymbol{x} | \Theta, \boldsymbol{\lambda})\} = \min\{k(\tilde{\Theta}, \tilde{\boldsymbol{\lambda}} | \Theta, \boldsymbol{\lambda}), k(\Theta, \boldsymbol{\lambda} | \tilde{\Theta}, \tilde{\boldsymbol{\lambda}})\}$ and $k\{\Theta, \boldsymbol{\lambda} | \tilde{\Theta}, \tilde{\boldsymbol{\lambda}}\} = \int_X p(\boldsymbol{x} | \tilde{\Theta}, \tilde{\boldsymbol{\lambda}}) \log \frac{p(\boldsymbol{x} | \tilde{\Theta}, \tilde{\boldsymbol{\lambda}})}{p(\boldsymbol{x} | \Theta, \boldsymbol{\lambda})} d\boldsymbol{x}$, that is the Kullback-Leibler divergences. Computation of intrinsic loss functions in regular models may be simplified by

$$\delta_{\boldsymbol{x}}\{\tilde{\Theta}, (\Theta, \boldsymbol{\lambda})\} = \min\{\inf_{\tilde{\boldsymbol{\lambda}} \in \Lambda} k(\tilde{\Theta}, \tilde{\boldsymbol{\lambda}} | \Theta, \boldsymbol{\lambda}), \inf_{\tilde{\boldsymbol{\lambda}} \in \Lambda} k(\Theta, \boldsymbol{\lambda} | \tilde{\Theta}, \tilde{\boldsymbol{\lambda}})\}. \tag{2.2}$$

Given data $\boldsymbol{x}$ generated by $p(\boldsymbol{x} | \boldsymbol{\theta}, \boldsymbol{\lambda})$, a situation with no prior information about the value of $\boldsymbol{\theta}$ is formally described by the reference prior $\pi(\boldsymbol{\theta}, \boldsymbol{\lambda})$ when $\boldsymbol{\theta}$ is the quantity of interest and in any decision problem, the quantity of interest is that function of the parameters which enters the loss function. Consequently, the intrinsic discrepancy reference expected loss, or intrinsic expected loss, defined as

$$d(\tilde{\boldsymbol{\theta}} | \boldsymbol{x}) = \int_{\Theta} \int_{\Lambda} \delta_{\boldsymbol{x}}\{\tilde{\boldsymbol{\theta}}, (\boldsymbol{\theta}, \boldsymbol{\lambda})\} \pi(\boldsymbol{\theta}, \boldsymbol{\lambda} | \mathbf{x}) d\boldsymbol{\theta} d\boldsymbol{\lambda} \tag{2.3}$$

where $\pi(\boldsymbol{\theta}, \boldsymbol{\lambda} | \boldsymbol{x}) \propto p(\boldsymbol{x} | \boldsymbol{\theta}, \boldsymbol{\lambda}) \pi(\boldsymbol{\theta}, \boldsymbol{\lambda})$, and $\pi(\boldsymbol{\theta}, \boldsymbol{\lambda})$ is the joint reference prior when $\boldsymbol{\theta}$ is the quantity of interest.

The intrinsic statistic $d(\boldsymbol{\theta}_0 | \boldsymbol{x})$ is a measure of the evidence against the simplified model $p(\boldsymbol{x} | \boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{\lambda})$ provided by the data $\boldsymbol{x}$ The hypothesis $H_0 \equiv \{\boldsymbol{\theta} = \boldsymbol{\theta}_0\}$ should be rejected if(and only if) the posterior expected loss is sufficiently large. To decide whether or not the precise value $\boldsymbol{\theta}_0$ may be used as a proxy for the unknown value of $\boldsymbol{\theta}$, the Bayesian reference criterion(BRC) might be used as follows:

$$\text{Reject } H_0 \text{ iff } d(\boldsymbol{\theta}_0|\boldsymbol{x}) = \int_{\Theta}\int_{\Lambda} \delta_{\boldsymbol{x}}\{\boldsymbol{\theta}_0, (\boldsymbol{\theta},\boldsymbol{\lambda})\}\pi(\boldsymbol{\theta},\boldsymbol{\lambda}|\mathbf{x})d\boldsymbol{\theta}d\boldsymbol{\lambda} > d^* \tag{2.4}$$

The values of d^* around 2.5 would imply a ratio of $e^{2.5} \approx 12$, providing mild evidence against the null; while values around $5(e^5 \approx 150)$ can be regarded as strong evidence against H_0; values of $d^* \geqq 7.5(e^{7.5} \approx 1800)$ can be safely used to reject the null.

Bayes estimates are those which minimize the expected posterior loss. The intrinsic estimate is the Bayes estimate which corresponds to the intrinsic discrepancy loss and the reference posterior distribution. Introduced by Bernardo and Juárez(2003), this is a completely general objective Bayesian estimator which is invariant under reparametrization. The intrinsic estimate of $\boldsymbol{\theta}$

$$\tilde{\boldsymbol{\theta}}_{int}(\boldsymbol{x}) = \operatorname{argmin}_{\tilde{\boldsymbol{\theta}} \in \Theta} d(\tilde{\boldsymbol{\theta}}|\mathbf{x}) \tag{2.5}$$

is that parameter value which minimizes the reference posterior expected intrinsic loss (2.3).

Bayesian region estimation is typically based on posterior credible regions, *i.e.*, sets of $\boldsymbol{\theta}$ values with pre-specified posterior probabilities. The p-credible intrinsic region is the lowest posterior loss p-credible region which corresponds to the intrinsic discrepancy loss and the reference prior. An intrinsic p-credible region is a subset $R_p^{int} = R_p^{int}(\boldsymbol{x}, \Theta)$ on the parameter space Θ such that,

$$\int_{R_p^{int}} p(\boldsymbol{\theta}|\boldsymbol{x})d\boldsymbol{\theta} = p \ , \forall\, \tilde{\boldsymbol{\theta}}_i \in R_p^{int}, \forall\, \tilde{\boldsymbol{\theta}}_j \notin R_p^{int}, d(\tilde{\boldsymbol{\theta}}_i|\boldsymbol{x}) \le d(\tilde{\boldsymbol{\theta}}_j|\boldsymbol{x}), \tag{2.6}$$

where $d(\tilde{\boldsymbol{\theta}}_i|\boldsymbol{x})$ is the intrinsic expected loss (2.3).

3. Reference-Intrinsic Analysis for the Difference between Two Normal Means with Known Variance

Assuming the data $x_{ij}(i=1,2; j=1,\cdots,n_i)$ are independent and normally distributed with means μ_i and known variances σ_i^2. We derive the Bayesian reference criterion, the intrinsic estimator and intrinsic credible region when the parameter of interest is $\theta = \mu_1 - \mu_2$, the difference between two normal means.

With parametrization $\theta = \mu_1 - \mu_2, \lambda = \mu_2$, the joint probability density function of $\boldsymbol{x}$ with parameters (θ, λ) is given by

$$p(\boldsymbol{x}|\theta,\lambda) = (\sqrt{2\pi}\sigma_1)^{-n_1}(\sqrt{2\pi}\sigma_2)^{-n_2} \times \exp\left[-\frac{1}{2\sigma_1^2}\sum_{j=1}^{n_1}\{x_{1j}-(\theta+\lambda)\}^2 - \frac{1}{2\sigma_2^2}\sum_{j=1}^{n_2}(x_{2j}-\lambda)^2\right] = N(\boldsymbol{x}_1|\theta+\lambda,\sigma_1)N(\boldsymbol{x}_2|\lambda,\sigma_2) \tag{3.1}$$

where $\boldsymbol{x}$ is $\{\mathbf{x}_1,\mathbf{x}_2\} = \{x_{11},\cdots,x_{1n_1},x_{21},\cdots,x_{2n_2}\}$.

The directed divergence $k\{\tilde{\theta},\tilde{\lambda}|\theta,\lambda\}$ is

$$k\{\tilde{\theta},\tilde{\lambda}|\theta,\lambda\} = \int p(\boldsymbol{x}|\theta,\lambda)\log\frac{p(\boldsymbol{x}|\theta,\lambda)}{p(\boldsymbol{x}|\tilde{\theta},\tilde{\lambda})}d\boldsymbol{x} = \left[\frac{n_1}{2}\left\{\frac{(\theta+\lambda)-(\tilde{\theta}+\tilde{\lambda})}{\sigma_1}\right\}^2 + \frac{n_2}{2}\left(\frac{\lambda-\tilde{\lambda}}{\sigma_2}\right)^2\right]. \tag{3.2}$$

As a function of $\tilde{\lambda}$, the directed divergence $k\{\tilde{\theta},\tilde{\lambda}|\theta,\lambda\}$ is minimized when $\tilde{\lambda}$ takes the value $\lambda+\frac{\sigma_2^2/n_2}{\sigma_1^2/n_1+\sigma_2^2/n_2}(\theta-\tilde{\theta})$. Thus the minimum directed divergence is

$$\begin{aligned} &inf_{\tilde{\lambda}\in R}k\{\tilde{\theta},\tilde{\lambda}|\theta,\lambda\} \\ &= \frac{n_1}{2\sigma_1^2}\left(\frac{\sigma_1^2/n_1}{\sigma_1^2/n_1+\sigma_2^2/n_2}\right)^2(\theta-\tilde{\theta})^2 + \frac{n_2}{2\sigma_2^2}\left(\frac{\sigma_2^2/n_2}{\sigma_1^2/n_1+\sigma_2^2/n_2}\right)^2(\theta-\tilde{\theta})^2 \\ &= \frac{1}{2(\gamma_1+\gamma_2)}(\theta-\tilde{\theta})^2, \end{aligned}$$

where $\gamma_1 = \sigma_1^2/n_1$, $\gamma_2 = \sigma_2^2/n_2$.

Similarly, the directed divergence $k\{\theta,\lambda|\tilde{\theta},\tilde{\lambda}\}$ is also minimized when $\tilde{\lambda}$ takes the value $\lambda+\frac{\sigma_2^2/n_2}{\sigma_1^2/n_1+\sigma_2^2/n_2}(\theta-\tilde{\theta})$, to yield

$$inf_{\tilde{\lambda}\in R}k\{\theta,\lambda|\tilde{\theta},\tilde{\lambda}\} = \frac{1}{2(\gamma_1+\gamma_2)}(\tilde{\theta}-\theta)^2.$$

Hence, using (2.2), the intrinsic discrepancy loss $\delta_{\boldsymbol{x}}\{\tilde{\theta},(\theta,\lambda)\}$ from using $\tilde{\theta}$ as a proxy for θ is

$$\delta_{\boldsymbol{x}}\{\tilde{\theta},(\theta,\lambda)\} = \frac{1}{2(\gamma_1+\gamma_2)}(\tilde{\theta}-\theta)^2. \tag{3.3}$$

It follows that the intrinsic expected loss is

$$d(\tilde{\theta}|\boldsymbol{x}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta_{\boldsymbol{x}}\{\tilde{\theta},(\theta,\lambda)\}\pi(\theta,\lambda|\mathbf{x})\mathrm{d}\lambda\mathrm{d}\theta \tag{3.4}$$

where $\pi(\theta,\lambda|\boldsymbol{x})$ is the reference posterior which corresponds to the reference prior $\pi(\theta,\lambda)$. The reference prior when θ is the parameter of interest is given by

$$\pi(\theta,\lambda) \propto 1. \tag{3.5}$$

Hence, the posterior distribution of (θ,λ) is given by

$$\pi(\theta,\lambda|\boldsymbol{x}) \propto p(\boldsymbol{x}|\theta,\lambda)\pi(\theta,\lambda) = N(\boldsymbol{x_1}|\theta+\lambda,\sigma_1)N(\boldsymbol{x_2}|\lambda,\sigma_2). \tag{3.6}$$

In our case, in order to calculate the reference prior when $\delta_{\boldsymbol{x}}\{\tilde{\theta},(\theta,\lambda)\}$ is the parameter of interest, note that this is a one-to-one piecewise function of Θ Thus given that the reference posterior is invariant under this kind of transformations (Bernardo and Smith, 1994, p. 326), we can use the reference posterior (3.6) when $\delta_{\boldsymbol{x}}\{\tilde{\theta},(\theta,\lambda)\}$ is the parameter of interest. Using the intrinsic discrepancy loss (3.3) and the reference posterior (3.6), the intrinsic discrepancy reference expected loss $d(\tilde{\theta}|\boldsymbol{x})$ is given as follow:

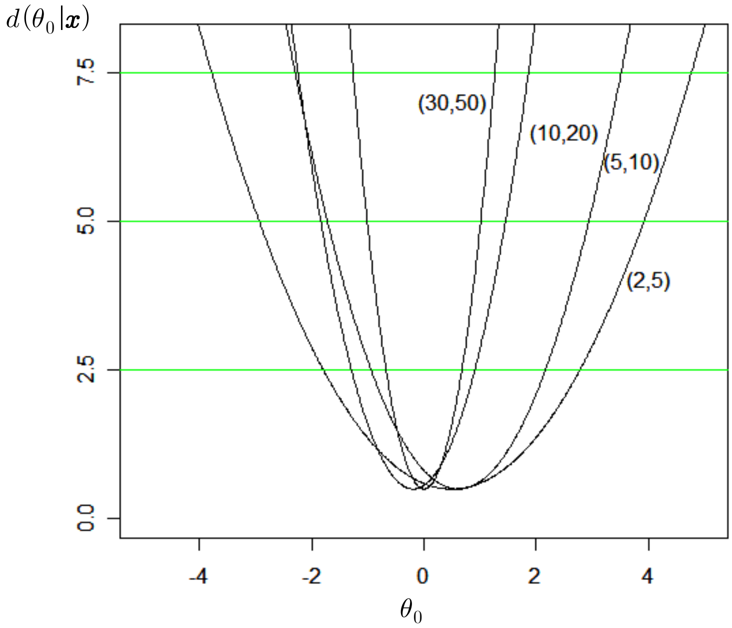
$$d(\tilde{\theta}|\boldsymbol{x}) = \frac{1}{2}\left[1+\left\{\frac{\tilde{\theta}-(\overline{x_1}-\overline{x_2})}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right\}^2\right]. \tag{3.7}$$

3.1 The Bayesian Reference Criterion

The intrinsic statistic $d(\theta_0|\boldsymbol{x})$ is the reference posterior expectation of the amount of information $\delta_{\boldsymbol{x}}\{\theta_0,(\theta,\lambda)\}$ which could be lost if the null model were used. Thus the Bayesian reference criterion(BRC) is

$$\text{Reject } \theta=\theta_0 \quad \text{iff } d(\theta_0|\boldsymbol{x}) > d^*. \tag{3.8}$$

Figure 1 represents the intrinsic statistic calculated from simulated data with $\theta=0$ and for several sample sizes of (n_1,n_2). We can see that for small (n_1,n_2), it is not possible to reject almost any value of the parameter and that the criterion becomes more discriminating as the sample size increases.



<Figure 1> The intrinsic statistic for the difference between two normal means θ_0, fixing $\theta = \mu_1 - \mu_2 = 0$.

<Table 1> Correspondence between the threshold value d^* of the test statistic $d(\theta_0|\boldsymbol{x})$ and type 1 error probabilities, $P[d > d^* | \theta = \theta_0]$, for several sample sizes (n_1, n_2).

d^*	$P[d > d^* \vert \theta = \theta_0]$	d^*	$P[d > d^* \vert \theta = \theta_0]$
1.8528	0.0100	1.00	0.3173
2.4207	0.0500	2.00	0.0833
3.8175	0.0100	3.00	0.0254
4.4398	0.0050	4.00	0.0082
5.9138	0.0010	5.00	0.0027
6.5578	0.0005	6.00	0.0009
8.0684	0.0001	7.00	0.0003

The frequentist behavior of the proposed test under the null is easily found. Indeed, if $\theta = \theta_0$, then the sampling distribution of $\overline{x}_1 - \overline{x}_2$ is $N(\theta_0, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$. Since $z^2 \sim \chi_1^2$ under H_0, the 'type 1' error probabilities $P[d > d^* | \theta = \theta_0]$ are given by $\Pr[\chi_1^2 > 2d^* - 1]$ as a function of the threshold value d^*. Table 1 gives a one-to-one correspondence between d^*-values and frequentist significance levels. However this exact correspondence is generally not to be expected.

3.2 The Intrinsic Estimation

The intrinsic expected loss $d(\tilde{\theta}|\boldsymbol{x})$ is minimized when $\tilde{\theta}=\overline{x}_1-\overline{x}_2$. Thus the intrinsic estimator of θ is

$$\tilde{\theta}_{int}(\boldsymbol{x})=\mathrm{argmin}_{\tilde{\theta}\in \mathrm{R}}\mathrm{d}(\tilde{\theta}|\mathbf{x})=\overline{\mathrm{x}}_1-\overline{\mathrm{x}}_2. \tag{3.9}$$

Moreover, $d(\tilde{\theta}|\boldsymbol{x})$ is symmetric around $\overline{x}_1-\overline{x}_2$, it follows that p-credible regions for Θ must be centered at $\overline{x}_1-\overline{x}_2$, and from the reference posterior (3.6), the marginal posterior density of Θ is $N(\theta|\overline{x_1}-\overline{x_2},\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)^{1/2})$.

Consequently, the intrinsic p-credible interval for the difference between two normal means θ is

$$R_p^{int}(\boldsymbol{x},R)=(\overline{x}_1-\overline{x}_2)\pm z_p\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)^{\frac{1}{2}} \tag{3.10}$$

where z_p is the $(p+1)/2$ quantile of a standard normal distribution. This credible interval is also the HPD interval, and the frequentist confidence interval of level $1-p$.

<Table 2> The mean value and standard deviation of the intrinsic estimators, the coverage probability and expected length of the intrinsic 0.95-credible intervals for several sample sizes (n_1,n_2).

(n_1,n_2)	Intrinsic estimators		Coverage Probability	Expected Length
	Mean	SD		
(2,5)	-0.0267	2.6324	0.9493	10.3713
(5,10)	-0.0189	1.8062	0.9526	7.1210
(10,20)	0.0182	1.2881	0.9481	5.0353
(30,50)	-0.0054	0.7869	0.9547	3.1196
(50,100)	-0.0048	0.5714	0.9514	2.2519

As numerical illustration, we simulated ten thousand data sets for several sample sizes (n_1,n_2) from a normal distributions $N(x_1|1,1)$ and $N(x_2|1,2)$, and calculated the mean value and standard deviation of the intrinsic estimators as well as both the coverage probability and expected length of the intrinsic 0.95 -credible intervals. The results are summarized in Table 2.

Table 2 provides that the intrinsic statistic appears to be consistently closer to the true value of the parameter, the frequentist coverage of reference p -credible regions is indeed approximately equal to p for all sample sizes and the expected length decrease according to the increase of sample sizes.

4. Concluding Remarks

The reference-intrinsic approach described provides a powerful alternative to point estimation and sharp hypothesis testing, with a clear interpretation in terms of information units. This study considers the reference-intrinsic approach for the difference between two normal means with known variances. We derived the Bayesian reference criterion, the intrinsic estimator and the credible region which corresponds to the intrinsic discrepancy loss and the reference prior in two sample case, and showed the possible extension of reference-intrinsic analysis in more complex setup.

References

1. Bartlett, M. (1957). A comment on D.V. Lindley's statistical paradox, *Biometrika*, **44**, 533-534.
2. Berger, J. O. and Bernardo, J. M. (1992). On the development of reference priors. *Bayesian Statistics 4* (J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, eds.), Oxford: University Press, 35-60 (with discussion).
3. Berger, J. O. and Pericchi, L. R. (2001). Objective Bayesian methods for model selection: Introduction and comparison, *Model Selection, Lecture Notes*, vol. 38 (P. Lahiri, ed.), Institute of Mathematical Statistics, pp. 135-207. (with discussion).
4. Berger, J. O. and Selke, T. (1987). Testing a point null hypothesis: The irreconcilability of *p*-values and evidence. *Journal of the American Statistical Association*, 82, 112-139.
5. Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society B*, 41, 113-147.
6. Bernardo, J. M. (1999). Nested hypothesis testing: The Bayesian reference criterion. *Bayesian Statistics 6* (J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, eds.), Oxford: University Press, 101-130.
7. Bernardo, J. M. (2005). Intrinsic credible regions: An objective Bayesian approach to interval estimation. *Test*, 14, 317-384.

8. Bernardo, J. M. and Juárez, M. A. (2003). Intrinsic Estimation. *Bayesian Statistics 7* (J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith and M. West, eds.), Oxford: University Press, 465-476.

9. Bernardo, J. M. and Rueda, R. (2002). Bayesian hypothesis testing: A reference approach. *International Statistical Review* **70**, 351-372.

10. Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*, Chichester: John Wiley.

11. Jeffreys, H. (1961). *Theory of Probability*, 3rd ed., Oxford: University Press.

12. Lindley, D. V. (1957). A statistical paradox, *Biometrika*, **44**, 187-192.

13. O'Hagan, A. (1997). Properties of intrinsic and fractional Bayes factors, *Test*, **6**, 101-118.

14. Robert, C. P. and Caron, N. (1996). Noninformative Bayesian testing and neutral Bayes factor, *Test*, **5**, 411-437.

15. Selke, T., Bayarri, M. J. and Berger, J. O. (2001). Calibration of p-values for testing precise null hypotheses. *The American Statistician*, 55, 62-71.

[received date : Aug. 2006, accepted date : Oct. 2006]