

A MINIMUM DISCRIMINATION INFORMATION SHRINKAGE TO INTERVAL ESTIMATOR OF THE NORMAL MEAN

Inada, Koichi
Department of Mathematics, Kagoshima University

Kim, Hea-Jung
Department of Statistics, Dongguk University

<https://doi.org/10.5109/13423>

出版情報 : Bulletin of informatics and cybernetics. 25 (1/2), pp.61-69, 1992-03. Research
Association of Statistical Sciences

バージョン :

権利関係 :

A MINIMUM DISCRIMINATION INFORMATION SHRINKAGE TO INTERVAL ESTIMATOR OF THE NORMAL MEAN

By

Koichi INADA* and Hea-Jung KIM**

Abstract

The minimum discrimination information (MDI) procedure for lowering the mean squared error (MSE) of the minimum variance unbiased estimator (MVUE) of the normal mean is considered. The procedure is employed to shrink the MVUE toward a preliminary conjectured interval under the information measure of Kullback and Leibler (1951). MDI estimator and its mean squared error are derived. The suggested estimator compares favorably with the previously proposed estimators in terms of mean squared error efficiency.

1. Introduction

Let X_1, \dots, X_n be *i.i.d.* $N(\mu, \sigma^2)$ where μ is the parameter of interest. Then the minimum variance unbiased estimator of μ is the sample mean \bar{X}_n given by

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i. \quad (1.1)$$

However, for the problem of estimating the normal mean, if we have a preliminary conjecture that $\mu \in [a, b]$, and if that information is strong compared to other available evidence, it would be better to use an estimator $\hat{\mu}_n$ which combines the preliminary conjecture in the estimation space around which accuracy seems most crucial. So that $\hat{\mu}_n$ has smaller mean squared error (MSE) than \bar{X}_n for all μ in the conjectured interval $[a, b]$, even though its MSE is greater than that of \bar{X}_n if μ is not in the interval. Thus, if the preliminary conjectured interval $[a, b]$ contains μ , we gain in MSE efficiency by using $\hat{\mu}_n$ instead of \bar{X}_n .

For the known variance case, the literature has given considerable attention to the problem of estimating the mean of a normal population under a prior information. For the point prior information case $a = b = \mu_0$, various choices of the shrinkage type estimator have been considered. Hirano (1977) studied a special type of preliminary test estimator using Akaike's (1973) information criterion. Thompson (1968a) and Mehta and Srinivasan (1971) derived their shrinkage estimators based on the MSE

* Department of Mathematics, Kagoshima University, Kagoshima 890, Japan.

** Department of Statistics, Dongguk University, Seoul 100, Korea.

criterion. Inada (1984) derived an estimator by a minimax regret criterion. Howkins and Han (1989) proposed a class of shrinkage estimators by a minimum average risk approach. On the other hand, when prior information at hand is in a form of $\mu \in [a, b]$, Thompson (1968b) gave an estimator which shrinks to an interval centered at a point, μ_0 . Disadvantages of the estimator may be specified on two grounds. First, its loose specification gives rise to enormous numerical calculations for detecting optimal value of parameters. Secondly, its MSE sometimes takes larger value than that of \bar{X}_n , even though μ takes some value in $[a, b]$.

The object of the present study is to propose and study yet another estimator which shrinks toward an interval, and to eliminate those problems attached to Thompson's estimator. Estimation of the normal mean with the interval prior information is treated as a constrained optimization of the Kullback-Leibler discrimination information function. This is called "minimum discrimination information" (MDI) procedure (cf. Gokhale and Kullback, 1978), so that we shall call an estimator of μ obtained from this procedure as a minimum discrimination information estimator (MDIE). In section 2, we obtain the MDI estimator for the problem, and derive its MSE. In section 3, some shrinkage estimators that already exist in the literature, including preliminary test estimator, are reviewed and their MSE efficiencies over \bar{X}_n are compared with respect to the MDI estimator. Section 4 contains some conclusions and further research topics of interest related with this study.

2. Minimum Discrimination Information Estimator

In the analysis of discrete and categorical data, the minimum discrimination information (MDI) procedure pioneered by Kullback (1959) is defined as the problem of finding the discrete frequency distribution $p^* \in P$ that minimizes the Kullback-Leibler discrimination function

$$I(p : \pi) = \sum_{\omega} p(\omega) \log \frac{p(\omega)}{\pi(\omega)} \quad (2.1)$$

subject to linear constraints on p in the form of $Cp = \theta$, where $\pi(\omega)$ denotes a reference distribution which arises from the problem of interest, and the distribution $p(\omega)$ is a member of a family P of distributions. The problem is the "external constraints problem (ECP)" when the reference distribution is $\pi = \hat{\pi}$, the vector of observed proportions (the unconstrained maximum likelihood estimates), and the constraints are provided "externally" to the data (cf. Gokhale and Kullback, 1978).

We apply this procedure to our problem in the continuous case. More specifically, we are concerned with estimating the mean μ of a continuous distribution $f(x|\mu, \sigma^2)$ with known σ^2 and the preliminary conjecture $\mu \in [a, b]$. Usually, the interval $[a, b]$ is centered at μ_0 . In that case it can be expressed as $[\mu_0 - \delta, \mu_0 + \delta]$. We assume, to begin with, that the underlying distribution is normal under external constraint $\mu \in [a, b]$.

Let $g(x|\bar{X}_n, \sigma^2)$ be the reference distribution which uses the normal distribution with mean \bar{X}_n (the unconstrained maximum likelihood estimate of μ) and known

variance σ^2 . Then the Kullback-Leibler (1951) discrimination information is given by

$$I(f: g) = \int f(x|\mu, \sigma^2) \log \frac{f(x|\mu, \sigma^2)}{g(x|\bar{X}_n, \sigma^2)} dx \quad (2.2)$$

where $I(f: g)$ is a random variable which uses the statistics \bar{X}_n .

Motivation of choosing this disparity measure can be found in Shore and Johnson (1980). Our aim is to find a normal density $f^*(x|\mu, \sigma^2)$ from the family of normal distributions $f(x|\mu, \sigma^2)$ which is as close as possible to the reference distribution $g(x|\bar{X}_n, \sigma^2)$ subject to the preliminary conjecture $\mu \in [a, b]$ that is a constraint provided externally to the data. Since σ^2 is known, this optimization is the same as to find the mean of the normal density $f^*(x|\mu, \sigma^2)$. That is, we want to find the mean of $f^*(x|\mu, \sigma^2)$ such that $f^*(x|\mu, \sigma^2)$ minimizes $I(f: g)$ subject to the external constraint that true of μ lies in an interval $[a, b]$;

Thus we

$$\begin{aligned} & \text{minimize } I(f: g) \\ & \text{subject to } \mu \in [a, b]. \end{aligned} \quad (2.3)$$

This problem bears analogy to the “external constraints problem (ECP)” in the MDI procedure.

THEOREM 1. *The minimum discrimination information estimator of μ and its mean squared error are given by*

$$T_{MDI} = \begin{cases} b & \text{if } \bar{X}_n > b, \\ \bar{X}_n & \text{if } \bar{X}_n \in [a, b], \\ a & \text{if } \bar{X}_n < a, \end{cases} \quad (2.4)$$

and

$$\begin{aligned} MSE[T_{MDI}] = & \sigma_{\bar{X}_n}^2 [(\theta_a - \theta)\phi(\theta_a - \theta) - (\theta_b - \theta)\phi(\theta_b - \theta) \\ & + \Phi(\theta_b - \theta) - \Phi(\theta_a - \theta) + (\theta_a - \theta)^2 \Phi(\theta_a - \theta) \\ & + (\theta_b - \theta)^2 \{1 - \Phi(\theta_b - \theta)\}] \end{aligned} \quad (2.5)$$

where $\theta_a = a/\sigma_{\bar{X}_n}$, $\theta_b = b/\sigma_{\bar{X}_n}$, $\theta = \mu/\sigma_{\bar{X}_n}$, and $\phi(\cdot)$ and $\Phi(\cdot)$ respectively denote p.d.f. and c.d.f. of the standard normal distribution.

PROOF. Integrating the right hand side of (2.2) with respect to x ,

$$I(f: g) = \frac{1}{2\sigma^2} (\mu - \bar{X}_n)^2. \quad (2.6)$$

This is a convex function of μ with global minimum at \bar{X}_n . Subject to the external constraint $\mu \in [a, b]$, $I(f: g)$ minimized by $\mu = b$ if $\bar{X}_n > b$, $\mu = \bar{X}_n$ if $\bar{X}_n \in [a, b]$ and $\mu = a$ if $\bar{X}_n < a$. Thus we get the minimum discrimination information estimator T_{MDI} given by (2.4). Using the p.d.f. of \bar{X}_n , $h(\cdot)$, we have the mean squared error of T_{MDI} given by

$$\begin{aligned} \text{MSE}[T_{MDI}] &= \int_b^\infty (b - \mu)^2 h(y) dy + \int_a^b (\bar{X}_n - \mu)^2 h(y) dy \\ &+ \int_{-\infty}^a (a - \mu)^2 h(y) dy, \end{aligned} \quad (2.7)$$

which yields (2.5).

Further, we consider the estimation problem subject to the following condition

(A): the estimators have the form

$$\alpha \bar{X}_n + (1 - \alpha) \mu_0 \quad (0 \leq \alpha \leq 1), \quad (2.8)$$

and belong to the conjectured interval $[\mu_0 - \delta, \mu_0 + \delta]$.

Usually the estimator having the above form is said shrinkage estimator.

THEOREM 2. *Subject to (A), the minimum discrimination information estimator of μ given by*

$$T_{MDI}^* = \begin{cases} \mu_0 + \delta & \text{if } \bar{X}_n > \mu_0 + \delta \\ \bar{X}_n & \text{if } \bar{X}_n \in [\mu_0 - \delta, \mu_0 + \delta], \\ \mu_0 - \delta & \text{if } \bar{X}_n < \mu_0 - \delta. \end{cases} \quad (2.9)$$

PROOF. The problem is reduced to finding an optimal α ($0 \leq \alpha \leq 1$) which

$$\begin{aligned} &\text{minimizes } I(f: g) \\ &\text{subject to } \alpha \bar{X}_n + (1 - \alpha) \mu_0 \in [\mu_0 - \delta, \mu_0 + \delta] \quad (0 \leq \alpha \leq 1). \end{aligned} \quad (2.10)$$

$I(f: g)$ is found to be

$$I(f: g) = (\bar{X}_n - \mu_0)^2 (\alpha - 1)^2. \quad (2.11)$$

Thus subject to $\alpha \bar{X}_n + (1 - \alpha) \mu_0 \in [\mu_0 - \delta, \mu_0 + \delta]$ ($0 \leq \alpha \leq 1$) we get estimator T_{MDI}^* which is given by (2.9).

If we let $a = \mu_0 - \delta$ and $b = \mu_0 + \delta$, then the estimator T_{MDI} is equal to T_{MDI}^* , which has exactly the same expression as (2.4) with $\mu_0 - \delta, \mu_0 + \delta$ instead of a, b , respectively.

3. Comparison of MDIE with Other Estimators

The estimator T_{MDI} claims only to minimize the Kullback-Leibler information measure that have externally constrained parameter space, $\mu \in [a, b]$. Therefore, under the same preliminary conjecture, we now compare its MSE efficiency with that of some other estimators in the introduction. When considering the normal distribution which is translation invariant, we can clearly confine the comparison without loss of generality to the case of $\mu_0 = 0$. So that, in our comparison, we shall take $[-\delta, \delta]$ as the preliminary conjectured interval and the origin as the center of the interval of critical accuracy.

For comparison with the MDI estimator given by (2.5) with replacements $b = \delta$ and $a = -\delta$, we have chosen four estimators as described below.

(a) Two types of preliminary test estimator for this problem (cf. Bancroft and Han, 1977) is given by

$$PT(i) = \begin{cases} 0 & \text{if } |\bar{X}_n|/\sigma_{\bar{X}_n} < Z(\frac{\alpha_i}{2}), \\ \bar{X}_n & \text{if } |\bar{X}_n|/\sigma_{\bar{X}_n} \geq Z(\frac{\alpha_i}{2}), \end{cases} \quad i = 1, 2, \quad (3.1)$$

and if we denote $\Phi(\cdot)$ and $\phi(\cdot)$ are respectively c.d.f. and p.d.f. of the standard normal distribution, final expressions of their mean squared error are

$$\begin{aligned} MSE[PT(i)] &= \sigma_{\bar{X}_n}^2 [\theta^2 \{ \Phi(C_i - \theta) + \Phi(C_i + \theta) - 1 \} \\ &\quad + (C_i - \theta)\phi(C_i - \theta) + (C_i + \theta)\phi(C_i + \theta) \\ &\quad + 2 - \Phi(C_i - \theta) - \Phi(C_i + \theta)], \quad i = 1, 2, \end{aligned} \quad (3.2)$$

where $C_i = Z(\frac{\alpha_i}{2})$ and $\theta = \mu/\sigma_{\bar{X}_n}$.

Here $PT(1)$ denotes the preliminary test estimator with the significant level $\alpha_1 = 0.05$ for the test $H_0: \mu = 0$ against $H_1: \mu \neq 0$, and $PT(2)$ indicates that with an optimal level of significance ($\alpha_2 = 0.1572\dots$) for the test, obtained by Akaike information criterion (cf. Hirano, 1977).

(b) Two shrinkage to an interval estimators proposed by Thompson (1968b) are given by

$$TH(j) = \bar{X}_n + \frac{k_j \sigma_{\bar{X}_n}^2}{4\delta} \ln \left[\frac{(\bar{X}_n - \delta)^2 + k_j \sigma_{\bar{X}_n}^2}{(\bar{X}_n + \delta)^2 + k_j \sigma_{\bar{X}_n}^2} \right], \quad j = 1, 2. \quad (3.3)$$

Here $TH(1)$ corresponds to the Thompson's estimator with value k_1 selected on the basis of the maximum width of the interval for which $MSE/\sigma_{\bar{X}_n}^2 < 1$, and the other estimator $TH(2)$ denotes that with k_2 chosen on the basis of the minimum value of $MSE/\sigma_{\bar{X}_n}^2$. Mean squared errors of these estimators are given by

$$\begin{aligned} MSE[TH(j)] &= \sigma_{\bar{X}_n}^2 \int_{-\infty}^{\infty} \left[x + \frac{k_j \sigma_{\bar{X}_n}^2}{4\theta_0} \ln \left\{ \frac{(x + \theta - \theta_0)^2 + k_j \sigma_{\bar{X}_n}^2}{(x + \theta + \theta_0)^2 + k_j \sigma_{\bar{X}_n}^2} \right\} \right]^2 \\ &\quad \times \phi(x) dx, \quad j = 1, 2, \end{aligned} \quad (3.4)$$

where $\theta_0 = \delta/\sigma_{\bar{X}_n}$, $\theta = \mu/\sigma_{\bar{X}_n}$ and $\phi(\cdot)$ denotes p.d.f. of standard normal distribution. Here we may need massive numerical calculation to get the values of k_1 and k_2 which correspond to a given value of δ . For our comparison we used values of δ whose corresponding k_1 and k_2 values are tabulated in Thompson (1968b).

We shall denote MSE efficiencies of T_{MDI} , $PT(i)$ and $TH(j)$ over \bar{X}_n as

$$\begin{aligned} e(T_{MDI}) &= MSE(T_{MDI})/\sigma_{\bar{X}_n}^2, \\ e(PT(i)) &= MSE(PT(i))/\sigma_{\bar{X}_n}^2, \quad i = 1, 2, \\ e(TH(j)) &= MSE(TH(j))/\sigma_{\bar{X}_n}^2, \quad j = 1, 2. \end{aligned} \quad (3.5)$$

Following figures show the graphs of (3.5) plotted against $|\theta| = |\mu/\sigma_{\bar{X}_n}|$ for 2 different

preliminary conjectured interval cases, $\theta_0 = 1, 2$, where $\theta_0 = \delta/\sigma_{\bar{x}_n}$. Since our aim is to compare the efficiencies of the shrinkage estimators in or near the preliminary conjectured interval, it would be sufficient to depict the efficiencies of them only for the region of $|\theta| \leq 5$. These figures are symmetric with respect to the origin.

In Figure 1 we see the efficiencies of the five different estimators which shrink toward an interval constructed by setting $\theta_0 = 1$, i.e., $|\theta| \leq 1$. Several behaviors of the efficiencies may be pointed out from this figure. First, the preliminary test estimators, $PT(1)$ and $PT(2)$, are inferior to the other estimators in a sense that, comparing to the others, these estimators do not attain any single most efficient point throughout the values of $|\theta|$, and yield worst case, $e(PT(i)) > 1$, $i = 1, 2$, even in some regions of the conjectured interval $|\theta| \leq 1$ (cf. Table 1 and Table 2). Secondly, Thompson's estimators $TH(1)$ and $TH(2)$ show their smallest $e(TH(j))$, $j = 1, 2$, at the middle of the interval of shrinkage, and give the $|\theta|$ values at which $e(TH(j)) = 1$, $j = 1, 2$, i.e., the breakeven points, outside the interval of shrinkage. On the other hand, the MDI estimator achieves smallest $e(T_{MDI})$ around the edge of the interval of shrinkage. This behavior, which is not in general equated to θ , gives rise to the estimator to have the widest efficient interval, for which $e(T_{MDI}) < 1$, of all the five estimators. If we compare the MDI estimator with Thompson's estimators in terms of MSE efficiency, it appears that Thompson's estimators are better when the value of $|\theta|$ locates in the

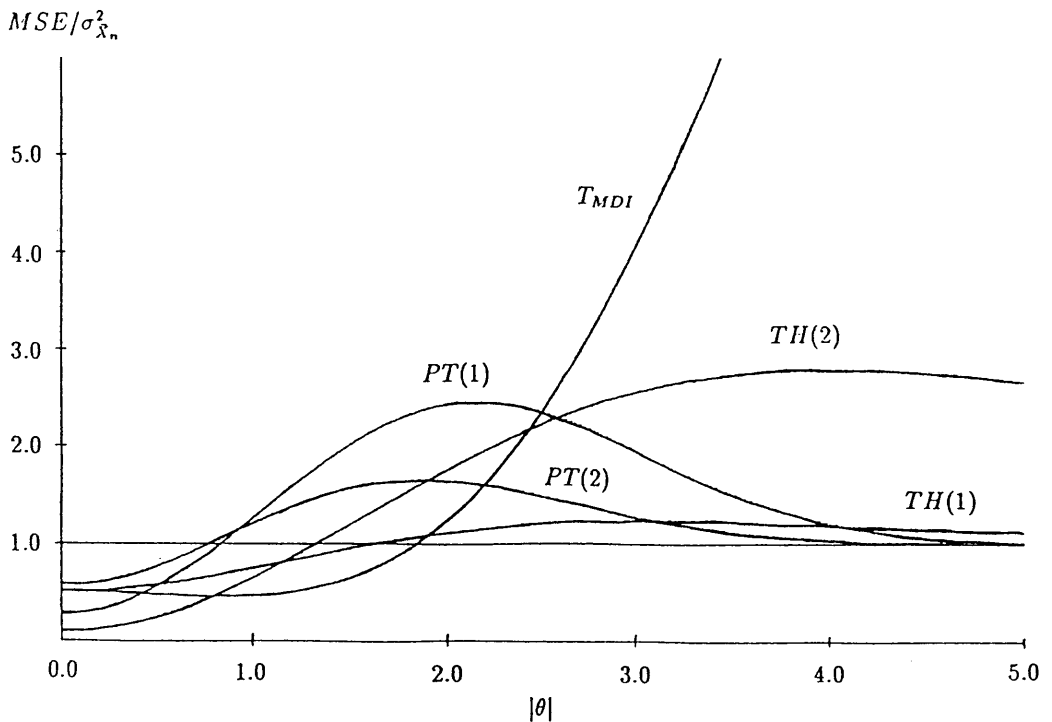


Fig. 1. Efficiencies of the five estimators in case of $\theta_0 = 1$.

middle parts of the interval of shrinkage, but get worse than the MDI estimator as that of $|\theta|$ lies around the edge of the interval (cf. Table 1).

Figure 2 depicts the efficiencies of the five shrinkage estimators with the interval of shrinkage $|\theta| \leq 2$. The analyses of this figure are same as those of Figure 1. But it is important to notice that, compared to \bar{X}_n , unlike the MDI estimator both Thompson's estimators do not achieve uniformly smaller MSE over the interval of shrinkage (cf. Table 1). For the purpose of overall comparison of the five estimators, we show in Table 1 and Table 2 two key indications of MSE efficiency of each estimator for $\theta_0 = 1, 2, 3, 5$.

These tables show that unlike others, over the every case of shrinkage interval, the MDI estimator produces uniformly lower MSE than \bar{X}_n , and tends to sustain this property in the outside vicinity of the interval of shrinkage (or preliminary conjectured interval).

4. Concluding Remarks

We have proposed and studied yet another estimator for the mean of a univariate normal distribution by shrinkage toward a preliminary conjectured interval. The suggested estimator is obtained from the minimum discrimination information (MDI) procedure using the reference distribution as the maximum entropy distribution with

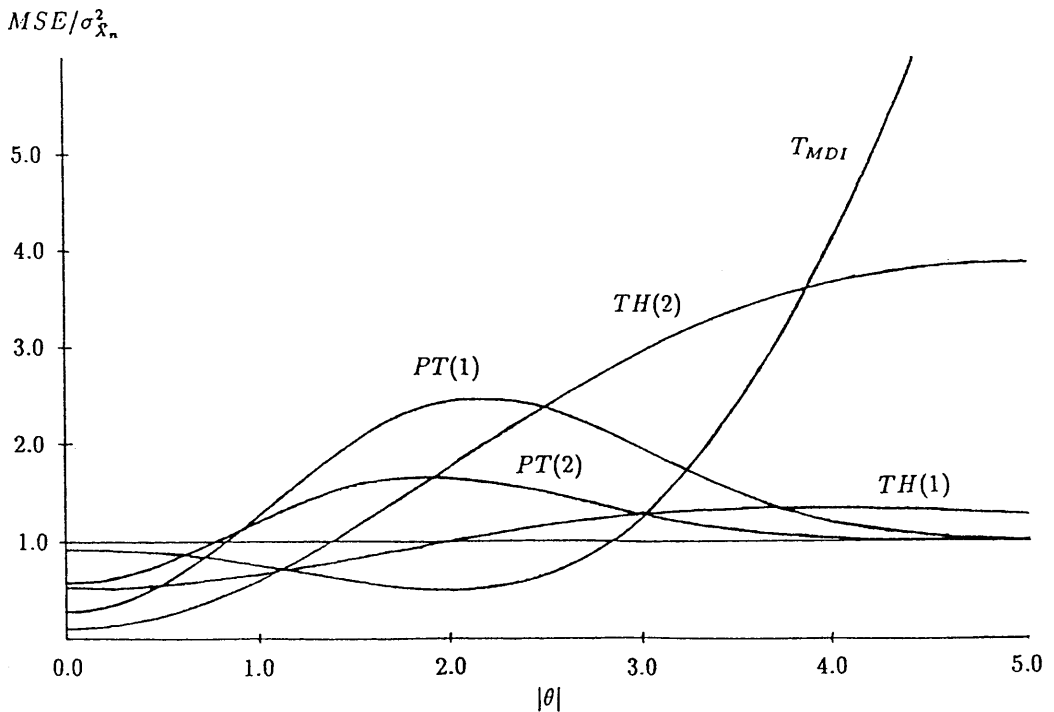


Fig. 2. Efficiencies of the five estimators in case of $\theta_0 = 2$.

Table 1. Breakeven points (BP); value of $|\theta|$ on which $e(PT(i))$, $e(TH(j))$, and $e(T_{MDI}) = 1$.

θ_0	$PT(1)$	$PT(2)$	$TH(1)$	$TH(2)$	T_{MDI}
1.0	0.83	0.77	1.63	1.33	1.84
2.0	0.83	0.77	2.01	1.37	2.83
3.0	0.83	0.77	2.22	1.39	3.83
5.0	0.83	0.77	2.40	1.40	5.83

Table 2. Most efficient intervals; intervals of $|\theta|$ on which certain estimator achieves the minimum MSE efficiency of all estimators.

θ_0	$PT(1)$	$PT(2)$	$TH(1)$	$TH(2)$	T_{MDI}
1.0	—	—	—	(0, 0.79)	(0.79, 1.84)
2.0	—	—	(1.09, 1.15)	(0, 1.09)	(1.15, 2.83)
3.0	—	—	(1.04, 1.75)	(0, 1.04)	(1.75, 3.83)
5.0	—	—	(1.03, 2.38)	(0, 1.03)	(2.38, 5.83)

unconstrained sample estimates of the corresponding parameters.

For various choices of the interval of shrinkage, we compared MSE efficiency over \bar{X}_n and the most efficient interval (MEI) with the other shrinkage estimators. In particular, some choices of the interval of shrinkage, $\theta_0 = 2, 3, 5$, unlike the others, the suggested estimator dominates \bar{X}_n uniformly in MSE when the true value of θ lies in the intervals, and this dominance continues to exist for the regions of somewhat beyond the intervals. These regions are not, in general, expressible as a function of θ . For a practical view point, the MDI estimator is unique, making the shrinkage estimation of the normal mean simple, while the other estimators have loose specifications, permitting great latitude in choices of the unspecified parameters. Unlike other estimation situations (cf. Gokhale, Inada and Kim, 1991), it is interesting to see that our MDI estimator coincides with that obtained by the constrained maximum likelihood method under the constraint $\mu \in [a, b]$.

Finally, the principle of constructing the MDI estimator used in this paper can be extended easily to the derivation of a shrinkage estimator of the normal mean for the unknown variance case, and this estimator is now under investigation.

References

- [1] AKAIKE, H.: *Information theory and an extension of the maximum likelihood principle*, 2nd International Symposium on Information Theory (eds. B.N. Petrov and F. Csaki), Akademiai Kiado, Budapest (1973), 267–281.
- [2] BANCROFT, T.A. and HAN, Chien-Pai: *Inference based on conditional specification: a note and a bibliography*, International Statistical Review, **45** (1977), 117–127.
- [3] GOKHALE, D.V., INADA, K. and KIM H.J.: *A minimum discrimination information estimator of*

- preliminary conjectured normal variance*. Submitted to *Ann. Inst. Statist. Math.*, (1991)
- [4] GOKHALE, D.V. and KULLBACK, S.: *The minimum discrimination information approach in analyzing categorical data*. *Communications in Statistics* **A7**, No.10 (1978), 987–1005.
 - [5] HAWKINS, D.R. and HAN, Chien-Pai: *A minimum average risk approach to shrinkage estimators of the normal mean*. *Ann. Inst. Statist. Math.*, **41** (1989), 347–363.
 - [6] HIRANO, K.: *Estimation procedures based on preliminary test, shrinkage technique and information criterion*. *Ann. Inst. Statist. Math.*, **29** (1977), 21–34.
 - [7] INADA, K.: *A minimax regret estimator of a normal mean after preliminary test*. *Ann. Inst. Statist. Math.*, **36** (1984), 207–215.
 - [8] KULLBACK, S. and LEIBLER, R.A.: *On information and sufficiency*. *Annals of Mathematical Statistics*, **22** (1951), 79–86.
 - [9] KULLBACK, S.: *Information Theory and Statistics*, New York, John Wiley, (1959).
 - [10] MEHTA, J.S. and SRINIVASAN, R.: *Estimation of the mean by shrinkage to a point*. *J. Amer. Statist. Assoc.*, **66** (1971), 86–90.
 - [11] SHORE, J.E. and JOHNSON, R.W.: *Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy*. *IEEE Transactions on Information Theory*, Vol. **IT-26**, No. 1 (1980), 26–37.
 - [12] THOMPSON, J.R.: *Some shrinkage techniques for estimating the mean*. *J. Amer. Statist. Assoc.*, **63** (1968a), 113–122.
 - [13] THOMPSON, J.R.: *Accuracy borrowing in the estimation of the mean by shrinkage to an interval*. *J. Amer. Statist. Assoc.*, **63** (1968b), 953–963.

Received August 1, 1991

Revised September 2, 1991

Communicated by H. Yamato