Ann. Inst. Statist. Math.
Vol. 45, No. 1, 129–136 (1993)

A MINIMUM DISCRIMINATION INFORMATION ESTIMATOR OF PRELIMINARY CONJECTURED NORMAL VARIANCE

D. V. GOKHALE[1], KOICHI INADA[2] AND HEA-JUNG KIM[3]

[1]*Department of Statistics, University of California, Riverside, CA 92521-0138, U.S.A.*
[2]*Department of Mathematics, Kagoshima University, Kagoshima 890, Japan*
[3]*Department of Statistics, Dongguk University, Seoul 100-715, Korea*

(Received July 25, 1991; revised June 10, 1992)

Abstract. For the problem of estimating the normal variance σ^2 based on random sample $X_1, \dots, X_n$ when a preliminary conjectured interval $[C_0^{-1}\sigma_0^2, C_0\sigma_0^2]$ is available, the minimum discrimination information (MDI) approach is presented. This provides a simple way of specifying the prior information, and also allows to consider a shrinkage type estimator. MDI estimator and its mean square error are derived. The estimator compares favorably with the previously proposed estimators in terms of mean square error efficiency.

Key words and phrases: Minimum discrimination information, Kullback-Leibler discrimination information measure, preliminary conjecture, preliminary test, minimax criterion, mean square error efficiency.

1. Introduction

Let $X_1, \dots, X_n$ be i.i.d. $\mathrm{N}(\mu, \sigma^2)$ where σ^2 is the parameter of interest. Then usual estimator of σ^2 is the unbiased sample variance U given by

$$U = \sum_{i=1}^{n} (X_i - \bar{X})^2/(n-1) \tag{1.1}$$

where $\bar{X}$ is a sample mean. Often we might be confronted with a situation to consider estimation of the variance by combining prior information in the form of a point or an interval in the estimation space around which accuracy seems most crucial. Such information, which is available either from the past experience of the experimenter or from some reliable sources, might be useful in a number of situations. For example, in the case of normal distribution, the empirical rule, Range $\approx 4\sigma$, can be used to detect some prior value σ_0^2, near which σ^2 is expected to lie (cf. Mendenhall ((1979), p. 55)). If that information is strong enough, we would be well advised to use an unusual estimator, $\hat{\sigma}_n^2$, whose mean square error (MSE) is less than that of sample variance U if σ^2 takes some value around σ_0^2, even though its MSE is greater than that of U for σ^2 sufficiently away from σ_0^2.

One method to utilize this kind of information is through estimation after preliminary test first introduced by Bancroft (1944). This method suggests the preliminary test estimator

$$PT = \begin{cases} \sigma_0^2 & \text{if } \chi^2_{n-1}\left(1-\frac{\alpha}{2}\right) < (n-1)U/\sigma_0^2 < \chi^2_{n-1}\left(\frac{\alpha}{2}\right), \\ U & \text{otherwise} \end{cases} \tag{1.2}$$

where $P(\chi^2_{n-1} \geq \chi^2_{n-1}(\alpha)) = \alpha$. The basic idea of this estimator is that we should use σ_0^2 instead of U if $H_0 : \sigma^2 = \sigma_0^2$ appears to be true, and just the reverse if H_0 appears to be false. As this loose specification permits great latitude in choice of α, various optimality conditions have been imposed, seeking to define a choice of optimum α (cf. Toyoda and Wallace (1975), Ohtani and Toyoda (1978) and Hirano (1978, 1980)). In a case the value σ_0^2 is assumed to be always less than or equal to σ^2, Pandey and Mishra (1991) proposed an improved estimator based on a weighted function of U and σ_0^2. From a different viewpoint, Inada (1989) proposed following type of estimator

$$T(w, C) = \begin{cases} \sigma_0^2 & \text{if } C^{-1} < U/\sigma_0^2 < C, \\ wU & \text{if } U/\sigma_0^2 \geq C, \\ w^{-1}U & \text{if } U/\sigma_0^2 \leq C^{-1} \end{cases} \tag{1.3}$$

where $w \in (0,1]$ and $C > 0$ are chosen by a minimax MSE criterion under the preliminary conjecture that the true value of σ^2 lies in the interval $[C_0^{-1}\sigma_0^2, C_0\sigma_0^2]$, where C_0 is a known positive constant. Based on MSE efficiency, he made comparisons among three estimators, U, PT and $T(w, C)$, and showed that, especially for small or intermediate sample sizes, the minimax estimator $T(w, C)$ performs better than the others. There are other improved estimators, such as pre-test estimators (cf. Gelfand and Dey (1988) and Ohtani (1991)), that incorporate prior information about μ. However, since our primary concern is to get an improved estimator using the information about σ^2, those are not illustrated here.

In this paper, we propose and study yet another estimator under the conjecture that σ^2 lies in the interval $[C_0^{-1}\sigma_0^2, C_0\sigma_0^2]$. It is the minimum discrimination information estimator (MDIE) which is based on the idea of a constrained optimization of the Kullback-Leibler discrimination information function. The form of the interval $[C_0^{-1}\sigma_0^2, C_0\sigma_0^2]$ is chosen to make comparisons with $T(w, C)$ possible. It is not necessary for derivation of the MDIE. Advantages of this estimator over PT and $T(w, C)$ are argued on three grounds. First, it has a closed form. Second, its derivation is straightforward. Third, it has more sensible MSE efficient interval than the other estimators. Here the MSE efficient interval defines an interval of σ^2 in which an estimator has smaller mean square error than U.

Section 2 derives the MDIE and places it in the context of an alternative approach. MSE of this estimator is also derived. Section 3 compares the MDIE with the other existing estimators based on the MSE criterion, and analyzes the results, showing reasonability of the proposed estimator. Section 4 contains some conclusions and further research topics of interest related with this study.

2. Minimum discrimination information estimator

We are concerned with the estimator of normal variance σ^2 using information obtained from both data and a preliminary conjecture about σ^2. We consider the information measure of Kullback and Leibler (1951), defined by

$$I(f : g) = \int f(x|\theta) \log \frac{f(x|\theta)}{g(x|\hat{\theta})} dx \tag{2.1}$$

where $I(f : g)$ is a random variable which uses the statistics $\hat{\theta} = \{\bar{X}, S^2\}$ with $S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/n$.

Motivation for choosing this disparity measure can be found in Shore and Johnson (1980). Our aim is to find a normal density $f(x|\theta)$ as close as possible to reference distribution $g(x|\hat{\theta})$ subject to the preliminary conjecture about σ^2 in the form of a constraint provided externally to the data. This problem bears analogy to the "external constraints problem (ECP)" in the minimum discrimination information (MDI) procedure (cf. Gokhale and Kullback (1978)). We shall call the estimator of σ^2 obtained from this procedure as minimum discrimination information estimator (MDIE). We want to find estimates of μ and σ^2 such that $I(f : g)$ is minimized, subject to the external constraint that true value of σ^2 lies in an interval $[C_0^{-1}\sigma_0^2, C_0\sigma_0^2]$, $C_0 \geq 1$; thus we

$$\begin{aligned} &\text{minimize } I(f : g), \\ &\text{subject to } \sigma^2 \in [C_0^{-1}\sigma_0^2, C_0\sigma_0^2]. \end{aligned} \tag{2.2}$$

If we write $\bar{X}$ and S^2 for the maximum likelihood estimator of μ and σ^2 respectively, then

$$I(f : g) = \frac{1}{2S^2}\{(\sigma^2 + \mu^2) - 2\bar{X}\mu + \bar{X}^2\} + \log S - \log \sigma - \frac{1}{2}. \tag{2.3}$$

This is a convex function of $\{\mu, \sigma^2\}$ with global minimum at $\{\bar{X}, S^2\}$. However, introducing external constraint $\sigma^2 \in [C_0^{-1}\sigma_0^2, C_0\sigma_0^2]$ to $I(f : g)$, we get MDIE of σ^2;

$$T_{MDI} = \begin{cases} C_0^{-1}\sigma_0^2 & \text{if } S^2 < C_0^{-1}\sigma_0^2, \\ S^2 & \text{if } S^2 \in [C_0^{-1}\sigma_0^2, C - 0\sigma_0^2], \\ C_0\sigma_0^2 & \text{if } S^2 > C_0\sigma_0^2. \end{cases} \tag{2.4}$$

Furthermore, it can be easily shown that this is the unique estimator which optimizes the criterion (2.2) under a class of shrinkage estimators of the form;

$$\hat{\sigma}_n^2(\alpha) = \alpha S^2 + (1 - \alpha)\sigma_0^2, \qquad 0 \leq \alpha \leq 1. \tag{2.5}$$

The MDIE, obtained by (2.2), may be formally viewed as a counterpart of restricted or bounded estimators of a normal mean (cf. Casella and Strawderman (1981) and Bickel (1981)) and of regression coefficients (cf. Klemm and Sposito

(1980) and Escobar and Skarpness (1987)). In other words, those estimators and the MDIE are formally similar in that they were driven by an optimization of their respective object functions under each interval constraint of parameter concerned.

Mean square error associated with the estimator (2.4) is

$$
\begin{aligned}
(2.6)\quad MSE(T_{MDI}) = {} & \frac{\sigma^4}{n^2}(n-1)(n+1)\{\Gamma(\beta_1;b)-\Gamma(\beta_1;a)\} \\
& - \frac{2\sigma^4}{n}(n-1)\{\Gamma(\beta_2;b)-\Gamma(\beta_2;a)\} \\
& + \sigma^4\{\Gamma(\beta_3;b)-\Gamma(\beta_3;a)\} \\
& + (C_0\sigma_0^2-\sigma^2)^2\{1-\Gamma(\beta_3;b)\} + (C_0^{-1}\sigma_0^2-\sigma^2)^2\Gamma(\beta_3;a)
\end{aligned}
$$

where $a = nC_0^{-1}\sigma_0^2/(2\sigma^2)$, $b = nC_0\sigma_0^2/(2\sigma^2)$ and $\beta_i = (n+5-2i)/2$, $i = 1,2,3$. Here Γ denotes the incomplete gamma function defined by

$$
(2.7)\qquad \Gamma(\beta;t) = \Gamma(\beta)^{-1}\int_0^t y^{\beta-1}e^{-y}dy, \qquad t \geq 0.
$$

The MSE of the estimator can thus vary over the range $\sigma^2 \geq 0$. Numerical evaluations, done for various values of C_0, σ_0^2 and n, indicate that the MSE of T_{MDI} is significantly small when true value of σ^2 is located in or near the conjectured interval $[C_0^{-1}\sigma_0^2, C_0\sigma_0^2]$ and gets larger as it goes further away from the interval (cf. Figs. 1, 2, 3 and 4). Although this provides an explanation of how the suggested estimator behaves in accordance with our goal of this study, they do nothing to resolve the question of which estimator is to be preferred among T_{MDI} and other estimators in the introduction. Preliminary test estimator PT in (1.2) and minimax estimator $T(w, C)$ in (1.3) can clearly be criticized on the ground that they have loose specifications. They could not be expressed as closed formulas, so it may be hard to set the best estimator for them in a real estimation situation. Since derivations of those estimators take subjective approaches, it is not possible to compare them with MDIE theoretically except with respect to their measurements of goodness. For the comparison, we will measure the goodness of an estimator by its mean square error.

3. Comparison of MDIE with other estimators

The estimator T_{MDI} claims only to minimize the Kullback-Leibler information measure that has the external constraint $\sigma^2 \in [C_0^{-1}\sigma_0^2, C_0\sigma_0^2]$. Therefore, under the same preliminary conjecture, we now compare its MSE efficiency with that of other estimators mentioned in the introduction. We shall denote the efficiencies of (1.2), (1.3) and (2.4) relative to (1.1) by $e(PT)$, $e(T(w, C))$ and $e(T_{MDI})$, respectively. That is,

$$
(3.1)\qquad
\begin{aligned}
e(PT) &= MSE(U)/MSE(PT), \\
e(T(w,C)) &= MSE(U)/MSE(T(w,C)), \\
e(T_{MDI}) &= MSE(U)/MSE(T_{MDI}).
\end{aligned}
$$

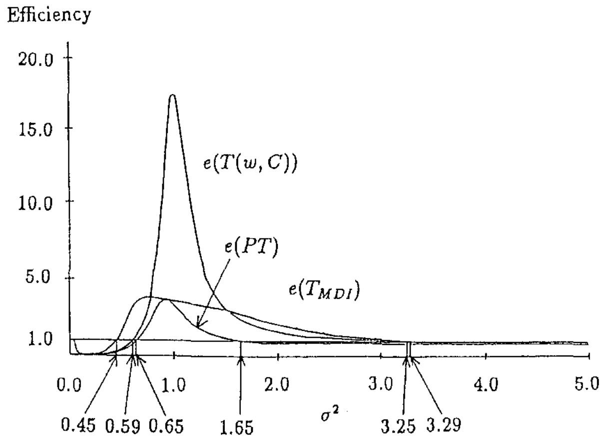

Fig. 1. Efficiencies of the three estimators under the set $\{n = 7,\ C_0 = 1.5,\ \sigma_0^2 = 1.0,\ w = 0.422,\ C = 2.37,\ \alpha = 0.05\}$.

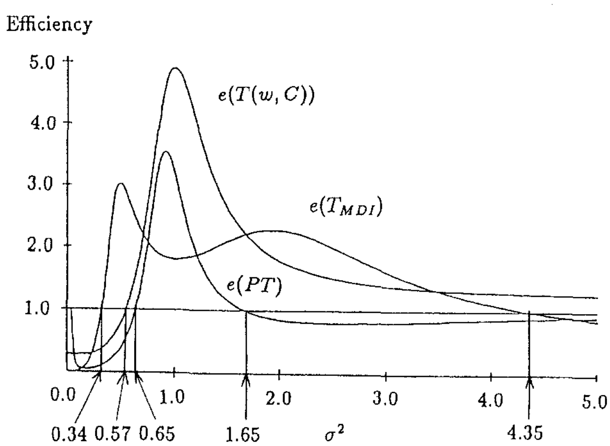

Fig. 2. Efficiencies of the three estimators under the set $\{n = 7,\ C_0 = 2.0,\ \sigma_0^2 = 1.0,\ w = 0.626,\ C = 1.64,\ \alpha = 0.05\}$.

Expressions of mean square errors of U, PT and $T(w, C)$ (cf. Inada (1989)) are

$$\begin{aligned}
&MSE(U) = 2\sigma^4/(n-1),\\
&MSE(PT) = \sigma^4[\{\Gamma(\beta_1; d) - \Gamma(\beta_1; h) + 1\}(n+1)/(n-1)\\
&\qquad\qquad - 2\{\Gamma(\beta_2; d) - \Gamma(\beta_2; h)\} - 1]\\
&\qquad\qquad + (2\sigma_0^2\sigma^2 - \sigma_0^4)\{\Gamma(\beta_3; d) - \Gamma(\beta_3; h)\},\\
&MSE(T(w, C)) = \sigma^4[\{w^{-2}\Gamma(\beta_1; l) + w^2(1 - \Gamma(\beta_1; v))\}(n+1)/(n-1)\\
&\qquad\qquad - 2\{w^{-1}\Gamma(\beta_2; l) + w(1 - \Gamma(\beta_2; v))\} + 1]\\
&\qquad\qquad + 2(\sigma_0^2\sigma^2 - \sigma_0^4)\{\Gamma(\beta_3; l) - \Gamma(\beta_3; v)\},
\end{aligned} \tag{3.2}$$

where $d = \chi^2_{n-1}(1 - \alpha/2)\sigma_0^2/(2\sigma^2)$, $h = \chi^2_{n-1}(\alpha/2)\sigma_0^2/(2\sigma^2)$, $l = (n-1)C^{-1}\sigma_0^2/$

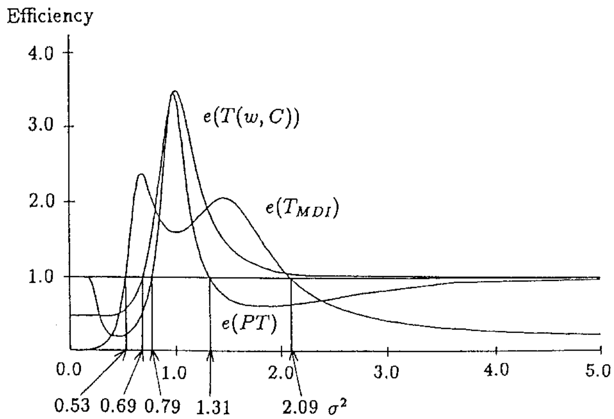

Fig. 3. Efficiencies of the three estimators under the set $\{n = 20,\ C_0 = 1.5,\ \sigma_0^2 = 1.0,\ w = 0.801,\ C = 1.25,\ \alpha = 0.05\}$.

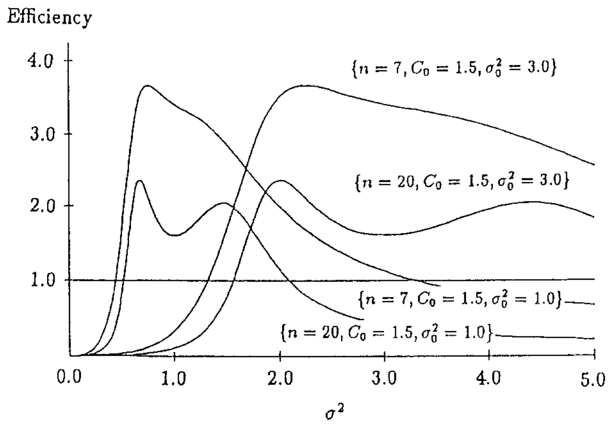

Fig. 4. Graph of $e(T_{MDI})$ under each case of the estimation situations.

$(2\sigma^2)$, $v = (n-1)C\sigma_0^2/(2\sigma^2)$, and the other notations are the same as in (1.2), (1.3), (2.4) and (2.6). When considering U which is consistent estimator of σ^2, we shall confine our comparison of the efficiencies of the estimators for the cases of small and moderate sample sizes.

Following figures show the graphs of (3.1) plotted against σ^2 for 3 numerical comparisons characterized by the set $\{n, C_0, \sigma_0^2, w, C, \alpha\}$. Here w and C denote some constants obtained from the criterion of

$$\inf_{0<w\leq 1, C>0} \sup_{\sigma^2 \in [C_0^{-1}\sigma_0^2, C_0\sigma_0^2]} MSE(T(w, C))$$

and α is a significance level to be determined for calculating $MSE(PT)$. In our comparisons, we used the optimum values of w and C tabulated in Inada (1989) and set α as 0.05. Since our aim is to compare the efficiencies of the three estimators in or near the preliminary conjectured interval, it would be sufficient to depict the efficiencies of them only for the region of $\sigma^2 \in [0, 5]$.

As seen in the figures, PT, $T(w, C)$ and T_{MDI} outperform U in most part of the conjectured interval and, in some extent, in the outside vicinity of the interval. Specifically, for various sample size, PT and $T(w, C)$ are more efficient than MDIE when true variance σ^2 lies in or very near the value σ^2. But their efficiencies drastically decrease as σ^2 goes outside that region. Particularly near the left limit of the conjectured interval, they are even worse than the usual unbiased estimator. On the other hand, T_{MDI} retains fairly good efficiency throughout and, to some extent, beyond the conjectured interval and has best efficiency of all when σ^2 lies around the conjectured interval limits. Figure 4 shows that this property of T_{MDI} remains same regardless of the sample size and the width of preliminary conjectured interval.

In practical estimation situation, we would not insure that the true value of σ^2 actually lies near σ_0^2 unless there is very strong prior belief that true variance is σ_0^2. Therefore, it would be better to use somewhat robust estimator T_{MDI} in a sense that it has the property of uniformly smaller MSE (compared to U) in or around the conjectured interval so that even in the case of a wrong preliminary conjecture, if it is not crucial, the estimator can still be able to give us an efficient estimation of σ^2.

4. Concluding remarks

We have introduced an estimator using the MDI approach which is widely applicable to a class of estimation problems where prior informations are available. The estimator compares favorably with the previously proposed estimators in a sense that unlike others, in every case of the preliminary conjectured interval, T_{MDI} dominates U uniformly in MSE when the true value of σ^2 lies in the interval, and this dominance continues to exist in the outside vicinity of the interval. Basic idea of the approach is to obtain an estimator that minimizes the Kullback-Leibler information of type under an externally given constraint formulated by the prior information. Clearly, there is much scope to supplement our estimation problem by further theoretical and practical work. In particular, the work of Section 2 might be carried out under a more general class of shrinkage estimators of σ^2 in the presence of the preliminary conjecture $[C_0^{-1}\sigma_0^2, C_0\sigma_0^2]$. For example, a class of estimators $\hat{\sigma}_n^2 = \alpha(T_n)S^2 + (1 - \alpha(T_n))\sigma_0^2$ can be considered, where $T_n = nS^2/\sigma_0^2$ and α is a weight function. It may also be pointed out that σ^2 was constrained to lie in the interval of the form $[C_0^{-1}\sigma_0^2, C_0\sigma_0^2]$ so as to be able to compare its performance with $T(w, C)$. Expression (2.4) can be easily modified under a general constraint of the form $[a\sigma_0^2, b\sigma_0^2]$, $0 < a < b < \infty$. Alternative estimation schemes, Bayesian approaches, are also of interest. Such research interests are currently being investigated.

Acknowledgements

The detailed and constructive comments of the referees and their thorough corrections to some errors were very useful in making the paper more readable. Their efforts are greatly appreciated.

REFERENCES

Bancroft, T. A. (1944). On biases in estimation due to the use of preliminary tests of significance, *Ann. Math. Statist.*, **15**, 190–204.

Bickel, P. J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted, *Ann. Statist.*, **9**, 1301–1309.

Casella, G. and Strawderman, W. E. (1981). Estimating a bounded mean, *Ann. Statist.*, **9**, 870–878.

Escobar, L. A. and Skarpness, B. (1987). Mean square error and efficiency of the least squares estimator over interval constraints, *Comm. Statist. Theory Methods*, **16**, 397–406.

Gelfand, A. E. and Dey, D. K. (1988). Improved estimation of the disturbance variance in a linear regression model, *J. Econometrics*, **39**, 387–395.

Gokhale, D. V. and Kullback, S. (1978). The minimum discrimination information approach in analyzing categorical data, *Comm. Statist. Theory Methods*, **7**, 987–1005.

Hirano, K. (1978). A note on the level of significance of the preliminary test in pooling variances, *J. Japan Statist. Soc.*, **8**, 71–75.

Hirano, K. (1980). On the estimators for estimating variance of a normal distribution, *Recent Developments in Statistical Inference and Data Analysis* (ed. K. Matusita), 105–115, North-Holland, Amsterdam.

Inada, K. (1989). Estimation of variance after preliminary conjecture, *Bull. Inform. Cybernet.*, **23**, 183–198.

Klemm, R. J. and Sposito, V. A. (1980). Least squares solutions over interval restrictions, *Comm. Statist. Simulation Comput.*, **9**, 423–425.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency, *Ann. Math. Statist.*, **22**, 79–86.

Mendenhall, W. (1979). *Introduction to Probability and Statistics*, 5th ed., Ouxbury Press, Bdmont.

Ohtani, K. (1991). Estimation of the variance in a normal population after the one-sided pre-test for the mean, *Comm. Statist. Theory Methods*, **20**, 219–234.

Ohtani, K. and Toyoda, T. (1978). Minimax regret critical values for a preliminary test in pooling variances, *J. Japan Statist. Soc.*, **8**, 15–20.

Pandey, B. N. and Mishra, G. C. (1991). Weighted estimators for a normal population variance, *Comm. Statist. Theory Methods*, **20**, 235–247.

Shore, J. E. and Johnson, R. W. (1980). Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy, *IEEE Trans. Inform. Theory*, **26**, 26–37.

Toyoda, T. and Wallace, T. D. (1975). Estimation of variance after a preliminary test of homogeneity and optimal levels of significance for the pre-test, *J. Econometrics*, **3**, 395–404.