Asymptotic Estimation of Variance

by

S. N. Joshi and Andrew L. Rukhin*
Purdue University

Technical Report #87-9

Department of Statistics
Purdue University

March 1987
Revised May 1988

* Research supported by NSF Grants DMS-8401996 and DMS-8803259.
AMS 1980 subject classifications: Primary 62F12, Secondary 62F10, 62F15, 62C15.
Currently visiting the Department of Mathematics, University of Massachusetts, Amherst

◇

ASYMPTOTIC ESTIMATION OF VARIANCE

This paper considers the asymptotic estimation problem of unknown variance from a location-scale parameter family under quadratic loss. The inadmissibility of the traditional unbiased estimator is demonstrated and a necessary and sufficient condition for the asymptotic admissibility of the corrected version of this estimator is obtained. It turns out that the latter is inadmissible if the kurtosis of the underlying distribution exceeds two.

Key words: Variance, unbiased estimator, quadratic loss, asymptotic admissibility, Brewster-Zidek estimator, quadratic polynomial of the normal mean, generalized Bayes estimators.

1. INTRODUCTION AND SUMMARY

Let $\underset{\sim}{x} = (x_1, \ldots, x_n)$ be a random sample from a location-scale parameter family with a distribution function of the form $F((x-\mu)/\sigma)$ where both μ and σ are unknown. It is convenient (after an appropriate linear transformation) to have

$$\int x dF(x) = 0, \quad \int x^2 dF(x) = 1,$$

so that μ is the mean, and σ^2 is the variance of each observation.

We study here the problem of variance estimation. For practical motivation of this problem and the use of various variance estimators in large sample surveys see the monograph of Wolter (1985).

The traditional estimator

$$\delta_U(\underset{\sim}{x}) = \sum_1^n (x_j - \overline{x})^2/(n-1) = S^2/(n-1)$$

motivated by normal distribution, is known to be unbiased (cf. Lehmann (1983) p. 102). However this estimator is always inadmissible under quadratic loss. Indeed let us consider estimators of the form cS^2 with a real constant c. The rescaled quadratic risk of these estimators $E_{\mu\sigma}\,(cS^2 - \sigma^2)^2\sigma^{-4}$ does not depend on unknown parameters, so that there is an optimal choice of c

$$c = c_0 = E_{01}S^2/E_{01}S^4 = n(n^2 - 2n + 3 + (n-1)\kappa)^{-1}$$

where $\kappa = E_{\mu\sigma}(X_1 - \mu)^4\sigma^{-4}$ is the kurtosis of the underlying distribution. See Moors (1986) and references there for the discussion of the meaning of this characteristic as a measure of peakedness or bimodality.

It is easy to see that for $n > 2$

$$c_0 \le n(n^2 - n + 2)^{-1} < (n-1)^{-1}.$$

In the case of the normal distribution function F, $\kappa = 3$ and $c_0 = (n+1)^{-1}$.

In this paper we give a necessary and sufficient condition for asymptotic admissibility in the class of scale equivariant rules of the estimator

$$\delta_0(\underset{\sim}{x}) = c_0 S^2$$

(which can be quite different from δ_U if κ is large). For this purpose in Section 2 we perform asymptotic study of the quadratic risk of properly normalized scale equivariant estimators which have been used for improved estimation of normal variance for a fixed sample size by Stein (1964), Brown (1968), Brewster and Zidek (1974) and Strawderman (1974). All known improvements over traditional estimators belong to this class, and it has been conjectured that admissibility in it implies admissibility within all statistical rules. The limiting behavior of Brewster-Zidek estimator studied by Rukhin (1987) suggests the correct normalization for the considered estimators. It turns out that the rescaled risk converges to the risk of an estimator of a random quadratic polynomial of the normal mean with known variance. This reduced problem is investigated in Section 3 where we show that the limiting version of δ_0 is admissible if and only if $\kappa \leq 2$. In the "normal" case, i.e. when $\kappa = 3$, we exhibit an improvement over δ_0 which is analogous to Brewster-Zidek variance estimator. This improvement is an admissible generalized Bayes procedure.

Notice that similar asymptotic admissibility problems have been considered in one-parameter case by Levit (1980), Ghosh and Sinha (1981) and in some multiparameter situations by Levit (1982), (1985).

2. ASYMPTOTIC ANALYSIS OF RISK OF SCALE EQUIVARIANT ESTIMATORS

We consider here scale equivariant estimators of σ^2 of the form

$$\delta(\underset{\sim}{x}) = c_0 S^2(1 - n^{-1}g(t)) \tag{2.1}$$

where $t = \sum_1^n x_j \, (\sum_1^n x_j^2)^{-1/2}$ and g is a continuously differentiable function with bounded derivative. We also suppose that for some positive ϵ

$$|t|^{1+\epsilon} g(t) \to 0, \quad |t| \to \infty. \tag{2.2}$$

Let X be a random variable with distribution function F. Other assumptions concern moment condition on X and Cramer's condition for (X, X^2). Assume that

$$EX = 0,$$

$$EX^2 = 1,$$

$$EX^3 = 0, \tag{2.3}$$

$$EX^6 < \infty \tag{2.4}$$

and

$$\limsup_{||s|| \to \infty} |E \exp(i\, s_1 X + i\, s_2 X^2)| < 1 \tag{2.5}$$

The following notation will be needed:

$$m_i = EX^i \quad 1 \leq i \leq 6,$$

$$a = m_4 - 1 = \kappa - 1 \quad (\text{as } \kappa = m_4),$$

$$Z_1 = n^{1/2}(\overline{x} - \mu),$$

$$Z_2 = n^{-1/2} \sum_1^n [(x_j - \mu)^2 - 1],$$

$$Q_n(z_1,\ z_2) = P(Z_1 \leq z_1, Z_2 \leq z_2)$$

and $\Phi(z_1, z_2)$ denotes the distribution function of the bivariate normal distribution with zero means and diagonal covariance matrix with diagonal elements 1 and a.

Note that if $\sigma = 1$ then under (2.3) and (2.4) $\Phi(z_1, z_2)$ is the limiting distribution function of the random vector (Z_1, Z_2) i.e. $Q_n \xrightarrow{w} \Phi$.

The scaled quadratic risk of estimators (2.1), $E_{\mu,\sigma}(\delta(\underset{\sim}{x}) - \sigma^2)^2\sigma^{-4}$, depends only on $\xi_n = n^{1/2}\mu\sigma^{-1}$, so that we assume henceforth that $\sigma = 1$. It can be proven that if $|\mu\sigma^{-1}|$ is bounded away from zero, then

$$\Delta_n(\xi_n) = n^2 E_{\xi_n}[(\delta_0(\underset{\sim}{x}) - 1)^2 - (\delta(\underset{\sim}{x}) - 1)^2]$$

tends to zero as n increases. Therefore we assume that $\xi_n \to \xi$ i.e. $\mu\sigma^{-1} \sim \xi n^{-1/2}$ as $n \to \infty$.

Theorem 1. Under assumption (2.2)–(2.5) one has

$$\Delta_n(\xi) \to -E[g^2(Z+\xi) + (a-2)(Z^2-1)g(Z+\xi) + a(Z+\xi)g'(Z+\xi)]$$

where Z is a standard normal random variable.

Proof: One has

$$c_0 = n^{-1} + (2 - m_4)n^{-2} + o(n^{-2}),$$
$$c_0 S^2 = 1 + n^{-1/2}Z_2 + (2 - m_4 - Z_1^2)n^{-1} + o_p(n^{-1}).$$

Therefore

$$\begin{aligned}\Delta_n(\xi) &= E_\xi\{2nc_0S^2(c_0S^2 - 1)g(t) - c_0^2S^4g^2(t)\} \\ &= E_\xi\{2n^{1/2}Z_2g(t) + 2(2 + Z_2^2 - m_4 - Z_1^2)g(t) \\ &\quad - g^2(t)\} + o(1).\end{aligned}$$

Notice that the largest order term in $\Delta_n(\xi)$ is $n^{1/2}E_\xi Z_2 g(t)$. To study its asymptotic structure we use the Edgeworth expansion, for the distribution of (Z_1, Z_2), up to order $n^{-1/2}$.

Theorem 20.1 of Bhattacharya and Rao (1975 p. 56) guarantees that

$$\iint f(z_1,z_2)dQ_n(z_1,z_2) = \iint f(z_1,z_2)d\Phi(z_1,z_2) + n^{-1/2}\iint f(z_1,z_2)P(z_1,z_2)d\Phi(z_1,z_2) + o(n^{-1/2}) \tag{2.6}$$

for a large class of functions f. Here $P(z_1, z_2)$ is a polynomial in z_1 and z_2; it follows from (7.19) of Bhattacharya and Rao (1975 p. 56) that $P(z_1, z_2)$, in our situation, has the form

$$P(z_1,z_2) = (m_6 - 3m_4 + 2)(z_2^3a^{-3} - 3z_2a^{-2})/6 + (z_1^2z_2 - z_2)/2 + m_5(z_1z_2^2a^{-2} - z_1a^{-1})/2.$$

Let $A = A_n = \{(Z_1, Z_2)\colon\ |Z_i| < \log n, \quad i = 1, 2\}$. On A we have

$$\begin{aligned} t &= (Z_1 + \xi)[1 + Z_2 n^{-1/2} + \xi(\xi + 2Z_1)n^{-1}]^{-1/2} \\ &= (Z_1 + \xi)[1 - Z_2/(2n^{1/2}) + o_p(n^{-1/2})] \end{aligned}$$

and hence

$$\begin{aligned} g(t) &= g(z_1 + \xi) - (z_1 + \xi) z_2 g'(z_1 + \xi) n^{-1/2}/2 \\ &\qquad + o(n^{-1/2}) \\ &= h_n(z_1, z_2) + o(n^{-1/2}) \qquad \text{(say)}. \end{aligned}$$

It can be proved that

$$\begin{aligned} \Delta_n(\xi) &= n^{1/2} \int_A z_2 h_n(z_1, z_2) dQ_n(z_1, z_2) \\ &\quad + 2 \int_A [(2 + z_2^2 - m_4 - z_1^2) h_n(z_1, z_2) - g^2(z_1 + \xi)] dQ_n(z_1, z_2). \end{aligned}$$

Using (2.6) for the first term one obtains

$$\begin{aligned} \Delta_n(\xi) &= 2n^{1/2} \int z_2 g(z_1 + \xi) d\Phi(z_1, z_2) \\ &\quad + 2 \int z_2 g(z_1 + \xi) P(z_1, z_2) d\Phi(z_1, z_2) \\ &\quad - \int (z_1 + \xi) z_2^2 g'(z_1 + \xi) d\Phi(z_1, z_2) \\ &\quad + 2 \int (2 + z_2^2 - m_4 - z_1^2) g(z_1 + \xi) d\Phi(z_1, z_2) \\ &\quad - \int g^2(z_1 + \xi) d\Phi(z_1, z_2) + o(1) \\ &= -E\{a(Z + \xi) g'(Z + \xi) + (a - 2)(Z^2 - 1) g(Z + \xi) \\ &\quad + g^2(Z + \xi)\} + o(1) \end{aligned}$$

and Theorem 1 is proved.

3. ADMISSIBILITY AND INADMISSIBILITY RESULTS

Theorem 1 shows that estimator (2.1) is asymptotically better than δ_0 if

$$\begin{aligned} \Delta_g(\xi) &= E[g^2(Z + \xi) + (a - 2)(Z^2 - 1) g(Z + \xi) + a(Z + \xi) g'(Z + \xi)] \\ &= E[g^2(Z + \xi) + (a - 2)(Z^2 - 1) g(Z + \xi) + a(Z^2 + Z\xi - 1) g(Z + \xi)] \le 0, \end{aligned} \tag{3.1}$$

The alternative form for $\Delta_g(\xi)$ in (3.1) is obtained by integration by parts. Inequality (3.1) means that

$$E[(a - 1)(Y^2 - 1) + g(Y) - \theta]^2 \le E[(a - 1)(Y^2 - 1) - \theta]^2$$

where $Y = Z + \xi$ is a normal random variable with mean ξ and unit variance, and

$$\theta = \theta(\xi, y) = [(3a-4)\xi y - (a-2)\xi^2]/2. \tag{3.2}$$

Thus the existence of an improvement on δ_0 in the original problem of asymptotic variance estimation is equivalent to inadmissibility of

$$d_0(y) = (a-1)(y^2-1) \tag{3.3}$$

as an estimator of a random quadratic polynomial θ of normal mean ξ on the basis of normal observation y with this mean and unit variance for quadratic loss $(d-\theta)^2$.

In the case of inadmissibility of d_0 in the latter problem to obtain an asymptotically better estimator in the original situation one has to replace in (2.1), $g(t)$ by $g(y)$. Indeed $t = y + 0_p(1)$.

Notice that $d_0(y)$ is an "unbiased" estimator of θ in the sense that

$$E_\xi d_0(Y) = (a-1)\xi^2 = E_\xi \theta(\xi, Y).$$

We look first at the form of Bayes estimators in this reduced problem. If Λ denotes prior distribution of ξ then the Bayes estimator of δ_B of θ has the form

$$\delta_B(y) = 0.5[(3a-4)y \int \xi \exp\{-(y-\xi)^2/2\} d\Lambda(\xi)$$
$$- (a-2) \int \xi^2 \exp\{-(y-\xi)^2/2\} d\Lambda(\xi)] / \int \exp\{-(y-\xi)^2/2\} d\Lambda(\xi).$$

Assume that Λ possesses a twice differentiable density λ and put

$$\delta_B(y) = (a-1)(y^2-1) + g_B(y).$$

Then integration by parts shows that

$$g_B(y) = 0.5 \int [-2(a-1)\lambda''(\xi) + a\xi\lambda'(\xi) + a\lambda(\xi)] \exp\{-(y-\xi)^2/2\} d\xi$$
$$/ \int \exp\{-(y-\xi)^2/2\} d\Lambda(\xi). \tag{3.4}$$

In particular if $\lambda(\xi) \equiv 1$, then $g_B(y) \equiv 0.5a$. It is easy to see that the latter estimator is inadmissible. In fact d_0 is always better.

It follows from (3.4) that $g_B(y) \equiv 0$ if and only if

$$D\lambda = \lambda''(\xi) - b\xi\lambda'(\xi) - b\lambda = 0 \tag{3.5}$$

where $b = 0.5a/(a-1)$.

Differential equation (3.5) has a solution

$$\lambda_0(\xi) = \exp\{b\xi^2/2\}$$

(In fact this is the only symmetric solution, $\lambda(-\xi) = \lambda(\xi)$.) Therefore if $b < 0$, i.e. if $a < 1$, λ_0 is a proper density and d_0 is a proper Bayes estimator which is trivially admissible.

Also if $a > 1$, then $b > 0$ and λ_0 cannot be approximated (in terms of posterior risk) by proper densities. This fact suggests the inadmissibility of d_0 for $a > 1$, which we prove later.

In the case $a = 1$, (3.5) takes the form

$$\xi\lambda'(\xi) + \lambda(\xi) = 0.$$

We take a sequence of approximate solutions for $\epsilon > 0$

$$\lambda_\epsilon(\xi) = |\xi|^{\epsilon-1}, \quad |\xi| < 1; = |\xi|^{-\epsilon-1}, \quad |\xi| \geq 1.$$

The Bayes estimator g_ϵ for such a prior is

$$g_\epsilon(y) = 0.5\epsilon \left[\int_{|\xi|<1} |\xi|^{\epsilon-1} \exp\{-(y-\xi)^2/2\} d\xi - \int_{|\xi|>1} |\xi|^{-\epsilon-1} \exp\{-(y-\xi)^2/2\} d\xi\right]$$
$$/ \left[\int_{|\xi|<1} |\xi|^{\epsilon-1} \exp\{-(y-\xi)^2/2\} d\xi + \int_{|\xi|>1} |\xi|^{-\epsilon-1} \exp\{-(y-\xi)^2/2\} d\xi\right]$$

An easy calculation shows that as ϵ tends to 0

$$\int\int [g_\epsilon(y) - g_o(y)]^2 \exp\{-(y-\xi)^2/2\}\lambda_\epsilon(\xi) d\xi dy \to 0,$$

and this fact is known to imply the admissibility of g_o (or d_0).

Now we consider the case $1 < a$. One has for a symmetric function g

$$\Delta_g(\xi) = E_\xi g(Y)[g(Y) + 2(a-1)(Y^2-1) - \xi Y(3a-4) + (a-2)\xi^2]$$
$$= (2\pi)^{-1/2} \exp(-\xi^2/2)$$
$$\sum_{k=0}^{\infty} \xi^{2k}/[(2k)!] \int_{-\infty}^{\infty} \exp(-y^2/2) y^{2k} g(y)$$
$$[g(y) + 2(a-1)(y^2-1) - (3a-4)2k + (a-2)2k(2k-1)y^{-2}]dy$$

Therefore $\Delta_g(\xi) \leq 0$ for all ξ if for $k = 0, 1, \ldots$

$$\mathcal{L}_k(g) = \int_0^{\infty} \exp(-y^2/2) y^{2k} g(y)$$
$$[g(y) + 2(a-1)(y^2-1) - (3a-4)2k + (a-2)2k(2k-1)y^{-2}]dy \leq 0. \qquad (3.6)$$

To find solutions to (3.6) we put

$$g(y) = r\exp(-\alpha y^2/2)$$

so that

$$\mathcal{L}_k(g) = r\int_0^\infty \exp(-y^2/2)y^{2k}dy(1+\alpha)^{-k-1/2}$$
$$\{r[(1+\alpha)/(1+2\alpha)]^{k+1/2} - 2(a-1)\alpha/(1+\alpha)$$
$$+ 2k[2(a-1)/(1+\alpha) - 3a + 4 + (a-2)(1+\alpha)]\}$$

Inequalities (3.6) hold for $\alpha > 0$, if

$$r = 2(a-1)\alpha(1+2\alpha)^{1/2}(1+\alpha)^{-3/2}$$

and

$$2(a-1)/(1+\alpha) - 3a + 4 + (a-2)(1+\alpha) \leq 0. \tag{3.7}$$

Clearly one can find positive α such that (3.7) is met. Indeed if $a \leq 2$, any sufficiently large α satisfies (3.7). If $a > 2$, it suffices to put

$$(1+\alpha)^2 = 2(a-1)/(a-2)$$

in which case (3.7) also holds.

We summarize our results.

Theorem 2. The asymptotic admissibility of estimator (2.1) of σ^2 is equivalent to the admissibility of estimator (3.3) of random quadratic polynomial (3.2) of normal mean ξ on the basis of a normal random variable Y with mean ξ and unit variance. Estimator $d_0(Y)$ is admissible in this problem if and only if $\kappa \leq 2$.

Now we consider the case $m_4 = 3$, i.e. $a = 2$ in more detail. In this case $d_0(Y) = Y^2 - 1$ is an inadmissible estimator of $\theta = \xi Y$ and we obtain an admissible improvement over d_0.

Let

$$\lambda_1(\xi) = \int_0^\infty \exp(-t\xi^2/2)t^{-1/2}(1+t)^{-1}dt \tag{3.8}$$

be generalized prior density which is analogous to the one introduced by Brewster and Zidek (1974).

Direct calculation shows that for the corresponding Bayes estimator d_1

$$g_1(y) = y\exp(-y^2/2)/\int_0^y \exp(-t^2/2)dt = y\frac{d}{dy}\log\int_0^y \exp(-t^2/2)dt. \tag{3.9}$$

We show that d_1 improves on d_0 by verifying (3.6). Notice first that

$$g_1^2(y) = g_1(y)(1-y^2) - yg_1'(y). \tag{3.10}$$

It follows from (3.10) that

$$\int_0^\infty \exp(-y^2/2)y^{2k}g_1^2(y)dy = 2\int_0^\infty \exp(-y^2/2)y^{2k}(k+1-y^2)g_1(y)dy.$$

Therefore

$$\mathcal{L}_k(g_1) = -2k\int_0^\infty \exp(-y^2/2)y^{2k}g_1(y)dy \le 0$$

and d_1 is indeed better than d_0. Standard admissibility argument, as before, in which λ_1 is approximated by proper densities

$$\int_0^\infty \exp(-t\xi^2/2)t^{\epsilon-1/2}(1+t)^{-1}dt, \quad \epsilon > 0$$

proves admissibility of d_1.

Thus we have proved

Theorem 3. In the case $\kappa = 3$, generalized Bayes estimator (3.9) against prior density (3.8) improves on d_0. This estimator is admissible.

The second statement of Theorem 3 also can be obtained from an extension of admissibility criterion for estimators of normal mean due to Brown (1971) Theorems 5.5.1 and 6.2.1. This extension shows that in our case any admissible estimator d has the form

$$d(y) = y^2 + yh'(y)/h(y) \tag{3.11}$$

where $h(y) = \int \exp(-(y-\xi)^2/2)d\Lambda(\xi)$ with a nonnegative measure Λ. Estimator (3.11) is admissible if and only if

$$\int_1^\infty y^{-2}h^{-1}(y)dy = \int_{-\infty}^{-1} y^{-2}h^{-1}(y)dy = \infty. \tag{3.12}$$

Notice that for $d_1(y)$

$$h(y) = \int_0^1 \exp(-t^2y^2/2)dt \le (0.5\pi)^{1/2}/|y|$$

so that (3.12) is satisfied.

References

Bhattacharya, R. N. and Ranga Rao, R. (1976). Normal Approximation and Asymptotic Expansions, Wiley, New York.

Brewster, J. F. and Zidek, J. V. (1974). Improving on equivariant estimators. *Ann. Statist.* **2**, 21–38.

Brown, L. D. (1968). Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters. *Ann. Math. Statist.* **39**, 29–48.

Brown, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Statist.* **42**, 855–903.

Ghosh, J. K. and Sinha, B. K. (1981). A necessary and sufficient condition for second order admissibility with applications to Berkson's bioassay problem. *Ann. Statist.* **9**, 1334–1338.

Lehmann, E. L. (1983). Theory of Point Estimation, Wiley, New York.

Levit, B. Ya (1980). On asymptotic minimax estimators of the second order. *Theory Probab. Appl.* **25**, 552–568.

Levit, B. Ya (1982). Minimax estimation and positive solutions of elliptic equations. *Theory Probab. Appl.* **27**, 563–586.

Levit, B. Ya. (1985). Second-order asymptotic optimability and positive solutions of the Schrödinger equation. *Theory Probab. Appl.* **30**, 333–363.

Moors, J. J. A. (1986). The meaning of kurtosis: Darlington reexamined. *Amer. Statist.* **40**, 283–284.

Rukhin, A. L. (1987). How much better are better estimators of a normal variance, *Journ. Amer. Statist. Assoc.* **82**, 925–928.

Stein, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.* **16**, 155–160.

Strawderman, W. E. (1974). Minimax estimation of powers of the variance of a normal population under squared error loss. *Ann. Statist.* **2** 190–198.

Wolter, K. M. (1985). Introduction to Variance Estimation, Springer-Verlag, New York.