

HOW MUCH BETTER ARE BETTER
ESTIMATORS OF A NORMAL VARIANCE

by

Andrew L. Rukhin
Purdue University

Technical Report #86-9

Department of Statistics
Purdue University

April 1986

Andrew L. Rukhin is Professor, Department of Statistics, Purdue University, West Lafayette, IN 47907. The research for this article was supported by NSF Grant DMS-8401996.

Thanks are due to H. Proskein who kindly supplied a copy of his unpublished thesis. The author is also very grateful to John Deely for many helpful remarks.

Key words: Normal variance, quadratic risk, Brewster-Zidek estimator, efficiency comparison, shrinkage estimators.

HOW MUCH BETTER ARE BETTER
ESTIMATORS OF A NORMAL VARIANCE

by

Andrew L. Rukhin
Purdue University

ABSTRACT

This paper considers the estimation problem of a normal variance σ^2 on the basis of a random sample x_1, \dots, x_n under quadratic loss when the mean ξ is unknown. The best equivariant estimator $\sum (x_j - \bar{x})^2 / (n + 1)$ is known to be inadmissible but the extent to which this inadmissibility phenomenon is serious has not previously been considered. Herein, the mean square error of the minimax and admissible estimator due to Brewster and Zidek (1974) is evaluated and an explicit formula for that estimator is given. It is shown that this risk function has maximum at $\xi = 0$ which is the mode of the corresponding generalized prior density. Locally optimal shrinkage scale-equivariant estimators are introduced and their risks are calculated for several values of the sample size n . It is observed that Brewster-Zidek estimator has risk function close to that of locally optimal minimax shrinkage estimators, but that the latter cannot give more than 4% relative improvement upon the traditional procedure.

1. Introduction

Let x_1, \dots, x_n , $n \geq 2$, be a random normal sample with unknown mean ξ and unknown variance σ^2 . We consider here the estimation problem of the variance σ^2 under quadratic loss.

Let $X = \sum_{j=1}^n x_j/n$ and $S^2 = \sum_{j=1}^n (x_j - X)^2$ be a version of the sufficient statistic. The best unbiased estimator of σ^2 is $\delta_U(X, S) = S^2/(n - 1)$. This estimator can be easily improved upon by $\delta_0(X, S) = S^2/(n + 1)$, which is optimal in the class of all procedures of the form cS^2 with some positive c . Some other choices of the constant c have been considered in the literature: Lindley (1953) suggested taking $c = 1/(n - 2)$, and Goodman (1960) proposed $c = 1/(n - 0.5)$. Notice that the estimator δ_0 is more optimistic than δ_U about the value of the unknown variance in the sense that $\delta_0 < \delta_U$. The shrinkage estimators considered in this paper are even more optimistic than δ_0 .

The inadmissibility of δ_0 has been established by Stein (1964), and a considerable amount of research has been directed to related problems after this discovery. Brown (1968) extended the inadmissibility result to more general loss functions, Strawderman (1974) obtained a class of minimax estimators some of which are generalized Bayes. Olkin and Selliah (1976) prove the admissibility of δ_0 within the class of all estimators whose risk is a multiplicative or additive function of ξ and σ . Cohen (1972) obtained improved confidence intervals for σ^2 , and Brewster and Zidek (1974) derived a minimax estimator which is admissible within the class of all scale-equivariant procedures. Proskin (1985) proved absolute admissibility of Brewster-Zidek estimator by formalizing the heuristic argument in Brown (1979).

In this paper we study the possible improvements over δ_0 in terms of quadratic risk. In particular in Section 2 we evaluate the risk of the Brewster-Zidek estimator. This risk function is shown to have a peculiar form with surprising maximum at $\xi = 0$ (which is the mode of the corresponding generalized prior density). In Section 3 the largest possible improvement among all scale-equivariant minimax shrinkage estimators for small samples is calculated as is shown to be no more than 4%. The estimator attaining this improvement is explicitly given.

2. Brewster-Zidek Estimator

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

$$\delta(X, S) = S^2 (1 - \phi(U))/(n + 1) \quad (2.1)$$

where $U = n^{1/2}|X|(nX^2 + S^2)^{-1/2} = n^{1/2}|X|/(\sum_{j=1}^n X_j^2)^{1/2}$, and ϕ is a positive measurable function.

The minimax estimator δ_1 obtained by Brewster and Zidek (1974) has this form. It is generalized Bayes against the prior density

$$\lambda(\xi, \sigma) = \int_0^{\infty} \exp \{-nt\xi^2/(2\sigma^2)\} t^{-1/2} (1 + t)^{-1} dt/\sigma$$

with respect to the invariant measure $d\xi d\sigma/\sigma$ when scaled quadratic loss $(\delta/\sigma^2 - 1)^2$ is used. We shall need the following representation of δ_1 :

$$\begin{aligned}
\delta_1(x,s) &= \frac{\int_{-\infty}^{\infty} \int_0^{\infty} \sigma^{-n-3} \exp\{[n(x - \xi)^2 + s^2]/(2\sigma^2)\} \lambda(\xi, \sigma) \, d\xi d\sigma}{\int_{-\infty}^{\infty} \int_0^{\infty} \sigma^{-n-5} \exp\{[n(x - \xi)^2 + s^2]/(2\sigma^2)\} \lambda(\xi, \sigma) \, d\xi d\sigma} \\
&= \frac{\int_{-0}^{\infty} \int_0^{\infty} \sigma^{-n-3} \exp\{-[nx^2(1+t^{-1})^{-1} + s^2]/(2\sigma^2)\} (1+t)^{-\frac{3}{2}} t \, dt d\sigma}{\int_0^{\infty} \int_0^{\infty} \sigma^{-n-5} \exp\{-[nx^2(1+t^{-1})^{-1} + s^2]/(2\sigma^2)\} (1+t)^{-\frac{3}{2}} t^{-\frac{1}{2}} \, dt d\sigma} \\
&= \frac{s^2 \int_0^u (1-v^2)^{(n-1)/2} \, dv}{(n+2) \int_0^u (1-v^2)^{(n+1)/2} \, dv} \\
&= \frac{s^2}{n+1} \left(1 - \frac{u(1-u^2)^{(n+1)/2}}{(n+2) \int_0^u (1-v^2)^{(n+1)/2} \, dv} \right) \\
&= s^2 (1 - \phi_1(u))/(n+1). \tag{2.2}
\end{aligned}$$

Formula (2.2) allows easy calculation of δ_1 . Indeed if $n = 2p - 1$ where p is positive then

$$\phi_1(u) = (1-u^2)^p / \sum_{k=0}^p \binom{p}{k} \frac{(-1)^k (2p+1)}{2k+1} u^{2k}.$$

and for even values of the sample size n , $n = 2p$,

$$\phi_1(u) = (1 - u^2)^{p + \frac{1}{2}}$$

$$/ \left[\sum_{k=0}^p \frac{(2p+1) \dots (2p - 2k + 3)}{2^k} (1 - u^2)^{p-k+\frac{1}{2}} + \frac{(2p+1)!!}{(2p)!!} \frac{\arcsin u}{u} \right]$$

These formulae are convenient for numerical evaluation of quadratic risk of δ_1 for small sample sizes.

It is easy to see that the mean squared error of δ_1 as well as any other estimator (2.1) depends only on $\eta = n^{\frac{1}{2}}|\xi|/\sigma$, and for any estimator δ of this form because of (4.3)

$$\Delta(\eta) = E_{\eta} \{ [S^2/(n+1) - 1]^2 - [S^2(1 - \phi)/(n+1) - 1]^2 \}$$

$$= 4 e^{-\eta^2/2} (n+1)^{-2} (d_{n-2})^{-1}$$

$$\left[\int_0^1 \phi(u) (1 - \phi(u)/2) (1 - u^2)^{(n+1)/2} \pi_{n+3}(u) du \right.$$

$$\left. - (n+1) \int_0^1 \phi(u) (1 - u^2)^{(n-1)/2} \pi_{n+1}(u) du \right], \quad (2.3)$$

where

$$d_m = \int_0^{\infty} s^m e^{-s^2/2} ds = 2^{(m-1)/2} \Gamma(0.5(m+1))$$

and

$$\pi_m(u) = (2\pi)^{-1/2} \int_0^{\infty} s^m e^{-s^2/2} \cosh su ds. \quad (2.4)$$

Notice that

$$\phi_1(u) = \frac{u}{(n+2)} \frac{d}{du} \log \int_0^u (1-v^2)^{(n+1)/2} dv$$

so that

$$\phi_1^2(u) = -u\phi_1'(u)/(n+2) + [n+2 - (n+1)u^2/(1-u^2)]\phi_1(u).$$

Integration by parts now gives

$$\Delta(n) = 4\pi e^{-n^2/2} (n+1)^{-2} (n+2)^{-1} (d_{n-2})^{-1}$$

$$\int_0^1 \phi_1(u) u (1-u^2)^{(n-1)/2} [(n+1) \pi_{n+1}'(un) - 0.5 (1-u^2) \pi_{n+3}'(un)] du. \quad (2.5)$$

It follows that $\Delta(0) = 0$, i.e. at $\xi = 0$ the risk of the Brewster-Zidek estimator takes its largest value. This is to be contrasted with the fact that $\xi = 0$ is the mode of the generalized prior density.

Using formulas for ϕ_1 and formulas (4.4) - (4.10) from the Appendix one can evaluate the integral (2.5), for instance, by Romberg adaptive extrapolation rule. Notice that for odd sample sizes ϕ_1 is a rational function so that (2.5) can be calculated in closed form.

Table 1 contains results of such numerical integration for the relative risk improvement $\Delta_{rel}(\eta) = (n+1)\Delta(\eta)/2$ for $3 \leq n \leq 8$ and $0 < \eta \leq 3$. These results show that δ_1 cannot give more than 3% relative improvement over δ_0 .

Asymptotical formula when n is large can be derived if one observes that

$$(n+2)\phi_1(u(n+1)^{-1/2}) \sim u e^{-u^2/2} / \int_0^u e^{-t^2/2} dt$$

and because of (4.11)

$$\Delta_{rel}(\eta) \sim \eta e^{-\eta^2/2} (n+1)^{-1} \times \int_0^\infty v^2 e^{-v^2} \sinh v\eta dv / \int_0^\infty e^{-t^2/2} dt. \quad (2.6)$$

Approximations for $\Delta(\eta)$ when $|\eta|$ is large can be obtained from (4.12).

There is asymptotical series for Δ in powers η^{-2} :

$$\Delta_{\text{rel}}(\eta) = A\eta^{-n-1} [1 - (n+2)/(2\eta^2) + \dots] \quad (2.7)$$

where

$$A = 2^{(n+3)/2} \Gamma(n+1) \Gamma(0.5(n+2)) / [\Gamma(0.5) \Gamma(0.5(n-1)) \Gamma(0.5(n+3))]$$

The formula (2.6) gives a good approximation for Δ_{rel} if $n \geq 10$, and (2.7) provides an accurate answer if $\eta > 6$.

3. Locally Optimal Shrinkage Estimators

It follows from Section 2 that the Brewster-Zidek estimator does not lead to substantial improvement over δ_0 . Here we address the question of how much improvement can be made upon δ_0 in the class of shrinkage estimators (2.1).

Because of (2.3) the nonnegative function ϕ which minimizes $\Delta(\eta_0)$ has the form

$$\begin{aligned} \phi(u) &= \max \{0, 1 - (n+1) E_{\eta_0}(S^2/U) / E_{\eta_0}(S^4/U)\} \\ &= \max \{0, 1 - (n+1) \pi_{n+1}(u\eta_0) / [(1-u^2)\pi_{n+3}(u\eta_0)]\}. \end{aligned} \quad (3.1)$$

where π_m is defined by (2.4).

For this ϕ

$$\Delta(\eta_0) = 2 e^{-\eta^2/2} (n+1)^{-2} (d_{n-2})^{-1}$$

$$\int_0^{u_0} (1-u^2)^{(n+1)/2} \pi_{n+3}(u\eta_0) [1 - (n+1)\pi_{n+1}(u\eta_0) / [(1-u^2)\pi_{n+3}(u\eta_0)]]^2 du \quad (3.2)$$

where u_0 solves the equation

$$(n+1) \pi_{n+1}(u_0 \eta_0) = (1 - u_0^2) \pi_{n+3}(u_0 \eta_0). \quad (3.3)$$

For numerical calculation of $\Delta(\eta_0)$ put $v_0 = u_0 \eta_0$, so that

$$\eta_0^2 = v_0^2 \pi_{n+3}(v_0) / [\pi_{n+3}(v_0) - (n+1)\pi_{n+1}(v_0)].$$

This formula gives the value of η_0^2 as a function of v_0 , and

$$\Delta_{rel}(\eta_0) = e^{-\eta^2/2} (n+1)^{-1} (d_{n-2})^{-1} \eta^{-n-2}$$

$$\int_0^{v_0} (\eta_0^2 - v^2)^{(n-3)/2} [\eta_0^2 (\pi_{n+3}(v) - (n+1)\pi_{n+1}(v)) - v^2 \pi_{n+3}(v)]^2 dv / \pi_{n+3}(v).$$

The function $\Delta_{rel}(\eta_0)$ is monotonically increasing in η_0 . When $\eta_0 = 0$

$$\phi(u) = \max\{0, 1 - (n+1)/[(1-u^2)(n+2)]\},$$

which corresponds to the original Stein's (1964) estimator. As is seen from Table 2 where the function (3.2) is tabulated (see the column corresponding to $\eta = 0$) its improvement over δ_0 is small. For larger values of η_0 one obtains more noticeable improvements.

Unfortunately these improvements are unattainable in the class of minimax estimators. In fact functions $\phi(u)$ generate minimax estimators only if $\eta \leq \bar{\eta}$. This value $\bar{\eta}$ can be determined by the condition $\Delta(0) = 0$, where Δ corresponds to estimator (3.1) with $\eta_0 = \bar{\eta}$. Thus these estimators have a risk function similar in form to that of the Brewster-Zidek estimator.

Table 3 provides values $\bar{\eta}$ at which the largest gain by a scale equivariant shrinkage estimator $\bar{\delta}$ is attained. The shifted estimator $\bar{\delta}(X+c, S) - c$ gives the maximal improvement at $\eta = c + \bar{\eta}$. Therefore if prior information about the ratio $\eta = n^{1/2}|\xi|/\sigma$, say $\eta \approx \eta_0$, is available, then the best choice of c is $c = \eta_0 - \bar{\eta}$. Notice that the estimator $\bar{\delta}$ cannot be improved upon in the class of scale equivariant shrinkage estimators, but it is not smooth and hence cannot be admissible. It appears that further scale equivariant improvements are not possible and the risk of the alternative improvements will be a complicated function of both η and σ .

In Table 3 we give the values $\bar{\eta}$, solutions u_0 of (3.3) and relative risk improvements of the corresponding locally optimal shrinkage estimators for $n = 3, \dots, 8$. It can be seen that these estimators give just slightly larger

maximal improvement over δ_0 than the Brewster-Zidek estimator. Since there is no improvement larger than 3.5% the answer, perhaps nonsurprisingly, to the title question of this paper is "not much".

However this situation changes in the problem of simultaneous estimation of a number of unknown variances, of a covariance matrix or of a function of the latter. For instance, Lin and Perlman (1985) report significant improvements upon traditional estimate of a covariance matrix. If this matrix is known to be of a diagonal form, the methods of the present paper are applicable and the corresponding vector estimators provide sizeable improvement upon the analogue of δ_0 . In this situation the fact that shrinkage scale-equivariant estimators cannot give substantial improvements for small values of $|\xi|/\sigma$, but do noticeably better for larger values of this ratio, can be useful.

Appendix

We assemble here some results about the functions (2.4) which are closely related to the functions of a parabolic cylinder. These results are useful for practical evaluation of estimators and their risks in many other estimation problems involving normal parameters (see for example Rukhin (1986)).

Let for $m > -1$, $u > 0$,

$$\pi_m(u) = (2\pi)^{-\frac{1}{2}} \int_0^{\infty} e^{-s^2/2} s^m \cosh su \, ds.$$

Then

$$\begin{aligned} \pi_m'(u) &= (2\pi)^{-1/2} \int_0^\infty e^{-s^2/2} s^{m+1} \sinh su \, ds \\ &= [\pi_{m+2}(u) - (m+1) \pi_m(u)]/u, \end{aligned} \quad (4.1)$$

$$\pi_m''(u) = \pi_{m+2}(u). \quad (4.2)$$

The importance of functions π_m is due to the fact that for any integrable function $\phi(U)$, $U = n^{1/2} |X| (nX^2 + S^2)^{-1/2}$, and $\alpha > 1 - n$

$$E_\eta S^\alpha \phi(U) = 2e^{-\eta^2/2} \int_0^1 \phi(u) (1-u^2)^{(n+\alpha-3)/2} \pi_{n-1+\alpha}(u\eta) \, du / d_{n-2}. \quad (4.3)$$

Here $\eta = n^{1/2} |\xi|/\sigma$.

Define the following polynomials in u^2

$$P_p(u) = \sum_{k=0}^p \binom{2p}{2k} (2p - 2k - 1)!! u^{2k}, \quad (4.4)$$

$$Q_p(u) = \sum_{k=0}^p \binom{2p}{2k} \frac{2p+1}{2k+1} (2p - 2k - 1)!! u^{2k} \quad (4.5)$$

$$R_p(u) = \sum_{k=0}^p r_k^p u^{2k},$$

$$r_k^p = \sum_{j=0}^{p-k+1} \frac{(2p-2j)(2p+2k) \dots (2p+2k-2j+4)}{2^{k-1}} (2p-j+1) (2p-2k-2j+1)!! \quad (4.6).$$

Proposition. If $m = 2p$ with integer p , then

$$\pi_m^p(u) = 0.5 e^{u^2/2} P_p(u), \quad (4.7)$$

$$\pi_m^p(u) = 0.5 u e^{u^2/2} Q_p(u). \quad (4.8)$$

If $m = 2p + 1$ with integer p , then

$$\pi_m^p(u) = (2\pi)^{-1/2} R_p(u) + 0.5 u Q_p(u) e^{u^2/2} \operatorname{erf}(u 2^{-1/2}), \quad (4.9)$$

$$\pi_m^p(u) = (2\pi)^{-1/2} R_p'(u) + u Q_p(u) + 0.5 P_{p+1}(u) e^{u^2/2} \operatorname{erf}(u 2^{-1/2}), \quad (4.10)$$

where erf denotes the error function.

These formulae can be proved by induction if one uses recurrent formulas (4.1) or (4.2) from the known representations for functions of a parabolic cylinder, $D_{-m}(u)$. Indeed

$$\pi_m(u) = 0.5 e^{u^2/4} \Gamma(m+1) [D_{-m-1}(u) + D_{-m-1}(-u)] / (2\pi)^{\frac{1}{2}}.$$

$$= 0.5 \frac{d^m}{du^m} (e^{u^2/2}) \quad m - \text{even}$$

$$= 0.5 \frac{d^m}{du^m} [e^{u^2/2} \operatorname{erf}(u2^{-\frac{1}{2}})] \quad m - \text{odd}.$$

Asymptotical formulae can be obtained from Laplace's method which, for instance, gives

$$\pi_m(u2^{-\frac{1}{2}}) \sim e^{-(m+1)/2} (m+1)^{m/2} 2^{-\frac{1}{2}} \cosh u,$$

$$\pi'_m(u(m+1)^{-\frac{1}{2}}) \sim e^{-(m+2)/2} (m+2)^{(m+1)/2} 2^{-\frac{1}{2}} \sinh u, \quad (4.11)$$

and asymptotical expansions for large u are also easily derived:

$$\pi_m(u) = (2\pi)^{-\frac{1}{2}} e^{u^2/2} \left[\int_{-u}^{\infty} (t+u)^m e^{-t^2/2} dt + \int_u^{\infty} (t-u)^m e^{-t^2/2} dt \right] / 2$$

$$= 0.5 e^{u^2/2} u^m [1 + 0.5m(m-1)u^{-2} + \dots]. \quad (4.12)$$

REFERENCES

- Brewster, J. F. and Zidek, J. V., (1974), "Improving on Equivariant Estimators," The Annals of Statistics, 2, 21-38.
- Brown, L. (1968), "Inadmissibility of the Usual Estimators of Scale Parameters in Problems with Unknown Location and Scale Parameters," The Annals of Mathematical Statistics, 39, 29-48.
- (1979), "A Heuristic Method for Determining Admissibility of Estimators--with Applications," Annals of Statistics, 7, 960-994.
- Cohen, A. (1972). "Improved Confidence Intervals for the Variance of a Normal Distribution," Journal of the American Statistical Association, 76, 382-387.
- Goodman, L. A. (1960). "A Note on the Estimation of Variance," Sankhyā, 22, 221-228.
- Lin, S. P., and Perlman, M. D. (1985), "A Monte Carlo Comparison of Four Estimators of a Covariance Matrix," in Multivariate Analysis - VI, ed. P. R. Krishnaiah, Amsterdam: Elsevier Science Publishers, 411-429.
- Lindley, D. V., (1953), "Statistical Inference" (with discussion). Journal Royal Statistical Society, ser B, vol. 15, 30-76.
- Olkin, I., and Selliah, J. B. (1976), "Estimating Covariances in a Multivariate Normal Distribution", in Statistical Decision Theory and Related Topics II (1977 ed), ed. S. Gupta and D. Moore, New York: Academic Press, 313-326.
- Proskin, H. M. (1985), "An Admissibility Theorem with Applications to the Estimation of the Variance of the Normal Distribution," unpublished Ph.D. dissertation, Rutgers University, Department of Statistics.
- Rukhin, A. L. (1986), "Estimating Normal Tail Probabilities," Naval Research Logistics Quarterly, 33, 91-99.
- Stein, Ch., (1964), "Inadmissibility of the Usual Estimator for the Variance of a Normal Distribution with Unknown Mean," Annals of the Institute of Statistical Mathematics, 16, 155-160.
- Strawderman, W. E. (1974), "Minimax Estimation of Powers of the Variance of a Normal Population Under Squared Error Loss," Annals of Statistics, 2, 190-198.

Table 1. The Relative Risk Improvements of Brewster-Zidek Estimator
(in Percents)

n	η							
	.1	.5	.75	1.00	1.50	2.00	2.50	3.00
3	.01	.38	.70	1.15	2.04	2.41	2.84	2.69
4	.01	.36	.78	1.25	2.20	2.83	2.99	2.77
5	.01	.35	.59	1.26	2.22	2.82	2.94	2.68
6	.01	.38	.65	1.20	2.13	2.71	2.82	2.54
7	.01	.34	.72	1.18	2.05	2.60	2.68	2.38
8	.01	.29	.69	1.11	1.98	2.57	2.52	2.21

Table 2. The Largest Possible Improvements Within the Class of Shrinkage Estimators (in Percents)

n	η						
	.00	.5	1.0	2.0	3.0	4.0	6.0
3	1.79	1.80	2.07	4.30	6.78	8.68	10.95
4	1.83	1.84	2.18	4.61	7.60	10.23	13.24
5	1.76	1.79	2.09	4.53	7.85	10.78	14.66
6	1.67	1.68	2.05	4.42	7.80	11.11	15.32
7	1.57	1.57	2.01	4.30	7.75	12.07	15.91
8	1.47	1.47	1.96	4.27	7.69	10.59	16.50

Table 3. The Relative Risk Improvements of Locally Optimal Estimators
(in percents)

n	\bar{n}	u_0	η					
			.5	1.0	1.5	2.0	2.5	3.0
3	1.38	.68	.53	1.76	2.88	3.25	2.83	1.98
4	1.34	.63	.54	1.85	3.00	3.51	2.85	1.92
5	1.31	.59	.56	1.83	2.93	3.21	2.67	1.75
6	1.29	.56	.54	1.76	2.81	3.03	2.52	1.67
7	1.28	.53	.51	1.67	2.66	2.88	2.33	1.48
8	1.27	.51	.48	1.55	2.50	2.72	2.17	1.28