that depends only on U is $U/(n+1)$. Considering estimators of the form $U\varphi(V/U)$, and using techniques analogous to those of Stein (1964), Pal and Sinha (1989) showed that the choice

$$\varphi(V/U) = \min\left\{\frac{1}{n+1}, \frac{1}{n+2}\left(1+\frac{V}{U}\right)\right\}$$

produces an estimator that dominates $U/(n+1)$. For the same loss function, and using techniques analogous to those of Brewster and Zidek (1974), we can find a smooth estimator of λ^{-1} (MacGibbon and Shorrock, 1989). Because of the scale invariance of the problem, the distribution of V/U is independent of λ and this appears to be the only place in the argument where invariance plays a role.

ADDITIONAL REFERENCES

MacGibbon, K. B. and Shorrock, G. E. (1989). Estimation of the lamda parameter of an inverse Gaussian distribution. Technical Report, Université du Québec à Montréal.

Pal, N. and Sinha, B. K. (1989). Improved estimators of dispersion of an inverse Gaussian distribution. Technical Report, Univ. Maryland, Baltimore County.

Selliah, J. B. (1964). Estimation and testing problems in a Wishart distribution. Technical Report No. 10, Dept. Statistics, Stanford Univ.

Shorrock, R. W. and Zidek, J. V. (1976). An improved estimator of the generalized variance. *Ann. Statist.* **4** 629–638.

Shuster, J. (1968). On the inverse Gaussian distribution function. *J. Amer. Statist. Assoc.* **63** 1514–1516.

Sinha, B. K. (1976). On improved estimators of the generalized variance. *J. Multivariate Anal.* **6** 617–625.

Tweedie, M. C. K. (1957). Statistical properties of inverse Gaussian distributions. I. *Ann. Math. Statist.* **28** 362–377.

Comment

Andrew L. Rukhin

Maatta and Casella start their interesting paper with an analogy between the estimation of a multivariate normal mean and that of a normal variance. Indeed, in both of these problems a surprising inadmissibility phenomenon of a traditional and intuitively reasonable estimator has been discovered. However, each of these problems has distinctive features, and I would like to start by discussing two of them and then to comment on the asymptotic variance estimator and the variational form of Bayes estimators.

1. THE PROBLEM OF ESTIMATING A MULTIVARIATE NORMAL MEAN IS EASIER IN A SENSE

Let X have multivariate normal distribution $N_k(\mu, \sigma^2 I)$ and let S^2 be a statistic which is independent of X and such that S^2/σ^2 has a chi-squared distribution with ν degrees of freedom. This setting arises in a classical linear model where X represents the least squares estimator, and S^2, the residual sum of squares.

If μ is to be estimated under, say, quadratic loss, then one can demonstrate the inadmissibility of X for $k \geq 3$ using Stein's by now popular technique of integrating by parts. Indeed, if $\delta(X, S)$ is a smooth estimator, then one can obtain an unbiased estimator $D_\delta(X, S)$ of the risk difference

$$\Delta(\mu, \sigma) = [E\|X-\mu\|^2 - E\|\delta(X, S)-\mu\|^2]\sigma^{-2},$$

i.e.,

$$ED_\delta(X, S) = \Delta(\mu, \sigma).$$

It is also possible to choose δ so that $D_\delta \geq 0$, and hence this estimator, δ, improves on X.

In the problem of variance estimation, one can derive unbiased estimates of the risk difference for quadratic loss for the best equivariant estimator $S^2/(\nu+2)$. However, there is no alternative estimator for which this estimate is nonnegative. Conditioning on $\|X\|/S$ or representing the noncentral t-distribution, that of this statistic, as a Poisson mixture of central t-distributions is crucial for the inadmissibility proof. Notice that to estimate the risk difference Strawderman (1974) had used the so-called Baranchik lemma, which implies the nonnegativity of the expected value of a product of one monotone and one which changes signs.

2. RELATIVE RISK REDUCTIONS OF VARIANCE ESTIMATORS ARE SMALLER THAN FOR MEAN ESTIMATORS

It is known that in our setup for the crude James–Stein estimator

$$\delta(X, S) = \left[1 - \frac{(k-2)S^2}{(\nu+2)\|X\|^2}\right]$$

Andrew L. Rukhin is Professor, Department of Mathematics and Statistics, University of Maryland at Baltimore County, Baltimore, Maryland 21228.

the relative risk reduction $\Delta(\mu, \sigma)/E\| X - \mu \|^2$ tends to 1 as $k \to \infty$ when $\mu = 0$, i.e., percentage risk reduction achieved by this estimator approaches 100%. As pointed out by Maatta and Casella in the end of Section 2, the univariate ($k = 1$) problem of estimating a normal variance for quadratic loss does not yield a substantial risk reduction. In the multivariate case this situation changes.

Let

$$\psi(X, S) = S^2(\nu + 2)^{-1}[1 - \varphi(S[\| X \|^2 + S^2]^{-1/2})] \tag{1}$$

be a scale-equivariant estimator of σ^2. For the Brewster–Zidek estimator

$$\varphi(v) = \varphi_{\mathrm{BZ}}(v) = v^\nu(1 - v^2)^{k/2}(\nu + k + 2)^{-1} \cdot \left[\int_v^1 t^{\nu+1}(1 - t^2)^{(k-2)/2}\, dt\right]^{-1}$$

and for the original Stein estimator

$$\varphi(v) = \varphi_s(v) = \max[0, 1 - (\nu + 2)(\nu + k + 2)^{-1}v^{-2}].$$

A study of the risks of these estimators has been done by Rukhin and Ananda (1989). They showed that for any dimension k the quadratic risk of the Brewster–Zidek estimator at $\mu = 0$ coincides with that of $S^2/(\nu + 2)$. This is a surprising fact since the generalized prior density which gives rise to the Brewster–Zidek procedure has the maximum at $\mu = 0$. Thus the traditional interpretation of prior distribution as a weight function reflecting the relative importance of different parametric values does not hold for improper priors.

Incidentally, for this reason the minimaxity of the Brewster–Zidek estimator cannot be deduced directly from Strawderman's theorem (1974). However, its slight modification can deliver the fact that the Brewster–Zidek procedure is better than $S^2/(\nu + 2)$.

It can be shown that the result quoted above holds for any smooth unimodal loss $W(\psi/\sigma^2)$, i.e., the corresponding risk of the Brewster–Zidek estimator at $\mu = 0$ coincides with that of the best equivariant estimator (which is proportional to S^2).

It is also demonstrated in Rukhin and Ananda (1989) that the Brewster–Zidek estimator provides increasingly larger percentage reductions in average loss (PRIAL in terminology of Lin and Perlman, 1985) as k increases. For instance, if $\nu = 10$ the relative risk improvement is 10.7% for $k = 9$ and 13.4% for $k = 15$. However, the norm of the vector μ where this maximum occurs, rescaled by σ^2, is also increasing (at a rate of $k^{1/2}$). Since the specification of the origin (or the point toward which shrinkage occurs) is needed for an efficient use of improved estimators (cf. Berger, 1982), the Brewster–Zidek estimator is not very easily implemented in practice for large k.

One can criticize the use of the quadratic loss function in variance estimation because it more heavily penalizes overestimation. For this reason entropy loss

$$W(t) = t - 1 - \log t$$

is often suggested. For this loss function the best equivariant estimator is the classical unbiased estimator S^2/ν. It is also inadmissible. Numerical results in Rukhin and Ananda (1989) show that its inadmissibility is even more pronounced than for quadratic loss, i.e., the relative risk improvement for the corresponding Brewster–Zidek estimator is larger.

3. ASYMPTOTICALLY THE PROBLEM OF ESTIMATING THE VARIANCE REDUCES TO THAT OF ESTIMATING A POSITIVE NORMAL MEAN

To explain the behavior of the risk of the Brewster–Zidek estimator, let us look at the version of variance estimation problem for a smooth unimodal loss function $W(\psi/\sigma^2)$. Assume that $k \to \infty$, $\nu \to \infty$ and

$$\nu^{-1/2}\varphi((v^{1/2} - 2^{1/2}z)k^{-1/2}) \to 2^{1/2}f(z).$$

This limiting behavior is suggested by the asymptotic theory of generalized Bayes estimators, in particular of the Brewster–Zidek estimator. It can be shown that if

$$\| \mu \|^2\sigma^{-2} \sim 2^{1/2}k\nu^{-1/2}\theta, \quad \theta \ge 0,$$

then

$$\frac{EW(c_0S^2/\sigma^2) - EW(c_0S^2(1 - \varphi)/\sigma^2)}{EW(c_0S^2/\sigma^2)} \to E(Z - \theta)^2 - E(Z + f(Z) - \theta)^2.$$

Here c_0S^2 is the best equivariant estimator under W, and the random variable Z is normal with nonnegative mean θ and unit variance. Thus asymptotically the variance estimation of problem reduces to the estimation of a positive normal mean. The Brewster–Zidek estimator corresponds to the generalized Bayes estimator with respect to Lebesgue measure on the positive half-line, so that

$$f(z) = f_{\mathrm{BZ}}(z) = \exp(-z^2/2) \Big/ \int_{-\infty}^{z} \exp(-t^2/2)\, dt.$$

The Stein estimator transforms into the maximum likelihood estimator,

$$f(z) = f_s(z) = \max(-z, 0).$$

Both of these estimators are minimax, i.e., their risk functions are bounded from above by 1. The risk of

$Z + f_{\mathrm{BZ}}(Z)$ is a unimodal function which takes a (minimax) value of 1 at $\theta = 0$ and $\theta = \infty$, and whose minimum, .584, is attained at $\theta = 1.08$. The risk of the maximum likelihood estimator has different form: it is a monotonically increasing function which takes value .5 at $\theta = 0$ and tends to 1 as $\theta \to \infty$. It is almost paradoxical that the inadmissible maximum likelihood estimator provides smaller minimal risk value than the admissible generalized Bayes estimator.

A long unresolved problem posed by Herbert Robbins is to find an explicit improvement over the maximum likelihood estimator and to determine its minimal risk value (which cannot exceed .5). Notice that such an improvement cannot be a shrinkage estimator, i.e., the function f *must* take both positive and negative values. This follows from the fact that the Stein variance estimator, as well as Z_+, is locally optimal at $\theta = 0$ in the class of shrinkage estimators and therefore cannot be improved on within the class (cf. Rukhin, 1987a). In view of our interpretation of the positive mean estimation problem, it is also natural to ask for the smallest minimal risk value within the class of all minimax estimators.

This is a fascinating problem not only because it is so simply formulated but also because it is a version of the asymptotic variance estimation question. In fact, this problem arises in practice when two samples before and after treatment are compared, and it is known that the treatment cannot diminish the mean. However minimal risk values considerably smaller than .5 are very unlikely to happen for minimax estimators.

Joshi and Rukhin (1989) consider another asymptotic setting involving variance estimation for an arbitrary location-scale univariate family. In this situation, under quadratic loss the unbiased estimator is always inadmissible, but the right multiple of it is asymptotically admissible if and only if the kurtosis of the underlying distribution does not exceed 2.

4. WHY BE OPTIMISTIC ABOUT THE VALUE OF AN UNKNOWN VARIANCE?

All known improvements over the best equivariant estimator of variance happen to be scale-equivariant shrinkage estimators, ψ, of the form (1), i.e., $\psi \le S^2(\nu + 2)^{-1}$. In other words, one should be rather optimistic about the value of the unknown variance, and this is true not only for quadratic loss.

To give a heuristic motivation for shrinkage estimators, let us examine the form of generalized Bayes estimators; after all, they form a complete class for this estimation problem. Denote by $\lambda(\mu, \sigma)$ a smooth prior density with respect to measure $d\mu d\sigma/\sigma$. It can be shown (cf. Rukhin, 1987b) that the corresponding generalized Bayes estimator δ_λ has the representation

$$\begin{aligned} \delta_\lambda(X,S) - S^2/\nu &= \nu^{-1} \iint [\mathscr{D}\lambda]\sigma^{1-\nu-k} \\ &\qquad \cdot \exp\{-[\| X-\mu \|^2 + S^2]/2\sigma^2\}\, d\mu\, d\sigma \\ &\quad \div \iint \lambda\sigma^{-\nu-k-3} \\ &\qquad \cdot \exp\{-[\| X-\mu \|^2 + S^2]/2\sigma^2\}\, d\mu\, d\sigma, \end{aligned} \tag{2}$$

where

$$\mathscr{D}\lambda = \sum \frac{\partial^2}{\partial \mu_i^2}\lambda + \frac{\partial}{\sigma\partial\sigma}\lambda.$$

It immediately follows from (2) that S^2/ν is a generalized Bayes estimator against the "noninformative" prior $\lambda \equiv 1$, which of course is well known. Also it is evident that $\delta_\lambda = S^2/\nu$ if and only if λ solves the parabolic differential equation

$$\mathscr{D}\lambda = 0,$$

which is closely related to the adjoint heat equation. Its typical solutions are of the form

$$\sigma^{-k}\exp\{\| \mu \|^2/(2\sigma^2)\}$$

or convolutions of these kernels with initial values $\lambda(\mu, 0)$. Therefore, these solutions look even less like probability densities than $\lambda \equiv 1$, and they do not admit a good approximation by proper densities. Since such an approximation is responsible for the admissibility of the corresponding generalized Bayes estimator, the inadmissibility of S^2/ν can be inferred from the properties of the adjoint heat operator. For the situations when the corresponding differential operator sometimes has solutions approximable by proper densities (see Rukhin, 1987c).

Also if

$$\lambda(\mu, \sigma) = \sigma^{-k}\lambda(\| \mu \|/\sigma) \tag{3}$$

then with $t = \| \mu \|/\sigma$,

$$\mathscr{D}\lambda = \sigma^{-k-2}[\lambda''(t) - t\lambda'(t) - k\lambda(t)],$$

and integration by parts shows that

$$\begin{aligned} &\iint [\mathscr{D}\lambda]\sigma^{1-\nu-k}\exp\left\{-\frac{\| \mu \|^2 + S^2}{2\sigma^2}\right\} d\mu\, d\sigma \\ &\quad = -k \iint \lambda\sigma^{-1-\nu-k}\exp\left\{\frac{\| \mu \|^2 + S^2}{2\sigma^2}\right\} d\mu\, d\sigma < 0, \end{aligned}$$

i.e.,

$$\delta_\lambda(0, S) < S^2/\nu.$$

In other terms, all Bayes estimators with respect to priors in (3) shrink at $X = 0$. Observe that priors of the form displayed in (3) lead to admissible Bayes estimators if λ has a sufficient number of moments. In particular, in the formula of prior densities (4.10) in Maatta and Casella, one can just as well put $a = .5$.

The utility of the variational representation of Bayes estimators in (2) is not exhausted by the facts quoted above. An easy calculation shows that if

$$\tilde{\lambda}(\mu, \sigma) = \int_0^\infty \exp\left\{-\frac{u\|\mu\|^2}{2\sigma^2}\right\} u^{k/2-1}(1+u)^{-1}\, du/\sigma^k$$

is the prior density of Brewster and Zidek, then $\mathscr{D}\tilde{\lambda}$ as a function of μ is proportional to Dirac's delta function so that $\tilde{\lambda}$ is a fundamental solution of the equation $\mathscr{D}\lambda = 0$. This fact, which shows a special role of Brewster–Zidek prior densities, may have significant implications in other estimation problems.

Also, it is worth noticing that the variational form of the Bayes risk has been used by Haff (1984) in estimating an unknown covariance matrix. This problem is an interesting generalization of that of variance estimation with its idiosyncratic results, and a survey revealing, among other things, unpublished work by Charles Stein in this area would be desirable.

One of the important applications of ideas and techniques displayed by Maatta and Casella arises in the theory of linear models when a covariance matrix is represented as a combination of given matrices with unknown weights. The estimation of these variance components has been considered from decision-theoretical point of view in Klotz, Milton and Zacks (1969).

Some other developments inspired by Stein's result can be found in Olkin and Selliah (1977) and Gelfand and Dey (1988). Another relevant problem is the estimation of generalized variance (see Sinha, 1976; Tsui, Weerahandi and Zidek, 1980).

To conclude, I would like to mention a classical problem where a standard deviation σ is to be estimated and in which therefore one can expect an inadmissible traditional procedure. (Incidentally, admissibility results for any power of σ are the same as for σ^2.) This is a confidence interval estimation problem for a mean μ. Even in the univariate case, the Student confidence interval $(X - aS, X + aS)$ probably is inadmissible under a risk function combining the probability of coverage and the expected width of the interval. Some inadmissibility results for other losses are given by Brown and Sackrowitz (1984) and in Rukhin (1988). Clearly these results are closely related to the property of the sets $\{|X| < \alpha S\}$ being positively biased as discussed by Maatta and Casella in Section 4.

One of the difficulties presented by this problem and not in the earlier ones considered is that it is not possible to restrict attention to procedures which depend only on $|X|$. The same feature occurs in the quantile estimation problem (Zidek, 1971) or in the estimation of a linear function of the mean and variance (Rukhin, 1987d). In these problems, no final results (which probably should mean admissible improvements) have been obtained.

ACKNOWLEDGMENTS

This work and most of the author's research mentioned here were supported by National Science Foundation Grant DMS-88-03259. The author is grateful to Jim Zidek for careful reading of the original version and many helpful remarks.

ADDITIONAL REFERENCES

BERGER, J. (1982). Selecting a minimax estimator of a multivariate normal mean. *Ann. Statist.* **10** 81–92.

BROWN, L. D. and SACKROWITZ, H. (1984). An alternative to Student's t-test for problems with indifference zones. *Ann. Statist.* **12** 451–469.

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

HAFF, L. R. (1984). Solutions of the Euler-Lagrange equations for certain multivariate normal estimation problems. Unpublished manuscript.

JOSHI, S. R. and RUKHIN, A. L. (1989). Asymptotic estimation of variance. Unpublished manuscript.

KLOTZ, J. H., MILTON, R. C. and ZACKS, S. (1969). Mean square efficiency of estimators of variance components. *J. Amer. Statist. Assoc.* **64** 1383–1402.

LIN, S. P. and PERLMAN, M. D. (1985). A Monte Carlo comparison of four estimators of a covariance matrix. In *Multivariate Analysis VI* (P. R. Krishnaiah, ed.) 411–429. North-Holland, Amsterdam.

OLKIN, I. and SELLIAH, J. B. (1977). Estimating covariances in a multivariate distribution. In *Statistical Decision Theory and Related Topics II* (S. S. Gupta and D. S. Moore, eds.) 313–326. Academic, New York.

RUKHIN, A. L. (1987b). Bayes estimators in log-normal regression model. In *Advances in Multivariate Statistical Analysis* (A. K. Gupta, ed.) 315–325. Kluwer, Norwell, Mass.

RUKHIN, A. L. (1987c). Quadratic estimators of quadratic functions of normal parameters. *J. Statist. Plann. Inference* **15** 301–310.

RUKHIN, A. L. (1987d). Estimating a linear function of the normal mean and variance. *Sankhyā Ser. A* **49** 72–77.

RUKHIN, A. L. (1988). Estimated loss and admissible loss estimators. In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds.) 409–420. Springer, New York.

RUKHIN, A. L. and ANANDA, M. M. A. (1989). Risk behavior of variance estimators in multivariate normal distributions. Technical Report, Univ. Maryland, Baltimore County.

SINHA, B. K. (1976). On improved estimators of the generalized variance. *J. Multivariate Anal.* **6** 617–625.

TSUI, K.-W., WEERAHANDI, S. and ZIDEK, J. V. (1980). Inadmissibility of the best fully equivariant estimator of the generalized residual variance. *Ann. Statist.* **8** 1156–1159.

ZIDEK, J. V. (1971). Inadmissibility of a class of estimators of a normal quantile. *Ann. Math. Statist.* **42** 1444–1447.