Comment

K. Brenda MacGibbon and Glenn E. Shorrock

Professors Maätta and Casella have written an excellent account of the theory of variance estimation using the approach originally conceived by Stein (1964). We especially appreciate the way in which they follow the historical development of the problem, and their discussion of the conditional properties of frequentist procedures gives new insight into the nature of these procedures. Moreover, their presentation is not purely expository: the prior given in (4.12) is, we believe, new material and is clearly a necessary step in the further derivation of the decision theoretic properties of these intervals.

Invariance seems to play a small role in this paper, while the original work in the field, including that of Stein (1964) and Brewster and Zidek (1974), gave great importance to the invariant group structure of the problem. It's not clear to us whether invariance is just a way of restricting the class of possible estimators to a more manageable subclass or whether it is essential to the problem. Our own experience in the application of this approach to other problems has been mixed; in some cases invariance has been crucial, in others not.

In the estimation of the generalized variance, the invariant structure plays an important role. This is perhaps due to the complicated multivariate structure of the problem. Starting with a multivariate normal linear model in canonical form, a minimal sufficient statistic is (X, S), where X is a normally distributed $p \times k$ matrix with independent columns $X_i \sim N(\xi, \Sigma)$, S is a $p \times p$ Wishart matrix with n degrees of freedom such that $ES = n\Sigma$, X and S are independent, and Σ is positive definite. We seek a point estimate of the determinant $|\Sigma|$ of Σ with the quadratic loss function

$$L\{\varphi(X, S); \Sigma, \xi\} = |\Sigma|^{-2}\{\varphi(X, S) - \Sigma\}^2.$$

This problem is invariant under the transformations

$$X \to AX + B \quad S \to ASA',$$
$$\xi \to A\xi + B \quad \Sigma \to A\Sigma A',$$

where A is any nonsingular $p \times p$ matrix and B is any $p \times k$ matrix. Estimators that are equivariant under

K. Brenda MacGibbon and Glenn E. Shorrock are Professors, Department of Mathematics and Computer Science, University of Quebec at Montreal, C.P. 8888, Succursale "A", Montreal, Quebec, Canada H3C 3P8.

this group satisfy

$$\varphi(AX + B, ASA') = |A|^2 \varphi(X, S)$$

and have the form $\varphi(S) = c|S|$ where c is a constant. Such estimators have constant risk (expected loss), which is minimized by taking $c = (n - p + 2)!/(n + 2)!$ (Selliah, 1964). Thus, in this problem,

$$\varphi_0(S) = \frac{(n - p + 2)!}{(n + 2)!}|S|,$$

as an estimator of $|\Sigma|$, plays a role analogous to $s^2/(n + 1)$ as an estimator of σ^2.

Using the ideas of Stein (1964) on variance estimation, exploiting zonal polynomials, and searching in a larger class than that of the affine equivariant estimators, Shorrock and Zidek (1976) showed that

$$\varphi(X, S)$$
$$= \min\left\{\frac{(n - p + 2)!}{(n + 2)!}|S|, \frac{(n + k + 2 - p)!}{(n + k + 2)!}|S + XX'|\right\}$$

has uniformly smaller risk than $[(n - p + 2)!/(n + 2)!]|S|$. The estimator $\varphi(X, S)$ is the analogue of Stein's estimator of σ^2. Shorrock and Zidek's proof depends heavily on invariance, as does that of Sinha (1976), who extended their results.

A problem in which invariance seems to play a minor role is the estimation of the parameter λ of an Inverse Gaussian distribution. The setup is as follows: Let $X_1, X_2, \cdots, X_n$ be a random sample from an Inverse Gaussian distribution with parameters μ and λ (denoted by $IG(\mu, \lambda)$) and density function

$$f(x; \mu, \lambda) = \sqrt{\lambda/2\pi x^3} \exp\left\{\frac{-\lambda}{2\mu^2 x}(x - \mu)^2\right\}, \quad x, \mu, \lambda > 0.$$

The mean of this distribution is μ and the variance is μ^3/λ. Writing $\bar{X} = (1/n)\sum_{i=1}^{n} X_i$ and $U = \sum_{i=1}^{n}(X_i^{-1} - \bar{X}^{-1})$, it is well known that $\bar{X} \sim IG(\mu, n\lambda)$, $U \sim \lambda^{-1}\chi_{n-1}^2$, and $\bar{X}$ and U are independent (Tweedie, 1957). The statistic $V = n\lambda(\bar{X} - 1)^2/\bar{X}$ has a χ_1^2 distribution when $\mu = 1$ (Shuster, 1968), which suggests that the ratio V/U can play a similar role in this problem to that of $Z = \sqrt{n}\bar{X}/S$ in the normal problem; there, a small value of Z indicates that μ is close to 0. In the inverse Gaussian case, a small value of V/U indicates that μ is close to 1. For the loss function $L(\delta, \lambda^{-1}) = (\delta - \lambda^{-1})^2$, the best estimator of λ^{-1}

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

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

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

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