pertaining to the robustness with respect to loss functions and distributions, of the results on estimation in the present paper.

ADDITIONAL REFERENCES

BROWN, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.* **37** 1087–1136.

BROWN, L. D. and COHEN, A. (1974). Point and confidence estimation of a common mean and recovery of interblock information. *Ann. Statist.* **2** 963–976.

CLEVENSON, M. and ZIDEK, J. V. (1975). Simultaneous estimation of the mean of independent Poisson laws. *J. Amer. Statist. Assoc.* **70** 698–705.

COHEN, A. and SACKROWITZ, H. B. (1977). Hypothesis testing for the common mean and for balanced incomplete blocks designs. *Ann. Statist.* **5** 1195–1211.

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.

SACKS, J. (1963). Generalized Bayes solutions in estimation problems. *Ann. Math. Statist.* **34** 751–768.

SARKAR, S. K. (1989). Shortest length confidence interval for the generalized variance and its inadmissibility. *J. Multivariate Anal.* To appear.

SINHA, B. K. and GHOSH, M. (1987). Inadmissibility of the best equivariant estimators of the variance-covariance matrix, the precision matrix and the generalized variance under entropy loss. *Statist. Decisions* **5** 201–217.

STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 197–206. Univ. California Press.

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

Comment

Edward I. George

I would like to begin by congratulating Maatta and Casella for an extraordinarily lucid and thought-provoking account of developments in decision-theoretic variance estimation. By systematically organizing so many related results, they have successfully exposed the main thread of ideas running through these developments. Effectively, this paper will serve as a springboard for further research ideas. To emphasize this point, my comments will focus on two new directions along which such ideas might proceed. The first concerns multiple shrinkage generalizations, and the second concerns further improvements to shrinkage estimators of the mean.

Let me mention before going on that, although my comments are limited to suggestions for future developments in point estimation, I am optimistic that these may also lead to analogous developments in interval estimation. I say this in light of the close connections between developments in these two areas which is brought out so clearly by Maatta and Casella.

1. MULTIPLE SHRINKAGE GENERALIZATIONS

A key idea behind the improved variance estimators described by Maatta and Casella is that of adaptively pooling possibly related information. In the single sample setting $X_1, \cdots, X_n \sim$ iid $N(\mu, \sigma^2)$, the estimators of Stein, Brown and Brewster and Zidek each improve on the "straw man" estimator $S^2/(n+1)$, $(S^2 = \sum (X_i - \bar{X})^2)$, by exploiting the possibility that $\mu/\sigma \approx 0$. The improved estimators are of the form $\phi(Z)S^2$, $(Z = \sqrt{n}\bar{X}/S)$, where $\phi(Z)$ is bounded above by $1/(n+1)$ and decreases as Z^2 decreases. When Z^2 is small, which is likely when μ^2/σ^2 is small, these estimators "shrink" $S^2/(n+1)$, effectively regaining the lost degree of freedom used in estimating μ. Indeed, Stein's estimator replaces $S^2/(n+1)$ by $\sum X_i^2/(n+2)$, an appropriate estimator when it is known that $\mu = 0$.

At first glance, this phenomenon may seem to be only a mathematical curiosity. After all, one degree of freedom will usually be a minor practical gain. This is precisely the point of the 4% bound on relative improvement described by Rukhin (1987a). However, it is straightforward to generalize these results to the general linear model case, as Maatta and Casella indicate in Section 5, where there are many more degrees of freedom and important gains may be realized. Indeed, the seminal results of Stein (1964) are obtained in such a case, although he states that "even in this case . . . the improvement is likely to be slight."

Unfortunately, there may be good reason to agree with Stein's pessimism. This can be seen in the canonical context of Section 5 where we observe independent normal variables $X_1, \cdots, X_\nu, X_{\nu+1}, \cdots, X_{\nu+p}$,

Edward I. George is Associate Professor of Statistics, Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, Illinois 60637.

with

$$X_i \sim N(0, \sigma^2) \quad \text{for } i = 1, \cdots, \nu,$$

(1.1) and

$$X_{\nu+1} \sim N(\mu_i, \sigma^2) \quad \text{for } i = 1, \cdots, p.$$

The means $\mu_1, \cdots, \mu_p$ are unknown, and the problem is to estimate σ^2. Letting

$$(1.2) \quad S^2 = \sum_1^\nu X_i^2, \quad Z = (Z_1, \cdots, Z_p)', \quad Z_i = \frac{X_{\nu+i}}{S},$$

the Stein estimator given in (5.1) of Maatta and Casella may be expressed as

$$(1.3) \quad \min\left(\frac{S^2}{\nu+2}, \frac{S^2+\sum_1^p X_{\nu+i}^2}{\nu+p+2}\right).$$

This and the corresponding generalizations of Brown and Brewster and Zidek for this situation, are all of the form

$$(1.4) \quad \phi(Z)S^2,$$

where $\phi(Z)$ is bounded above by $1/(\nu + 2)$ and decreases as $\|Z\|^2$ $(= Z'Z)$ decreases. Analogous to the univariate case, these estimators improve on the "straw man" estimator

$$(1.5) \quad S^2/(\nu + 2)$$

by exploiting the possibility that $\sum_1^p \mu_i^2/\sigma^2 \approx 0$. When $\|Z\|^2$ is small, which is likely when $\sum_1^p \mu_i^2/\sigma^2$ is small, these estimators "shrink" (1.5), effectively regaining the p degrees of freedom used in estimating $\mu_1, \cdots, \mu_p$. The potential gains from using such improved variance estimators in this context can be very substantial, especially if p is large compared to ν. The drawback however, is that every one of $\mu_1/\sigma, \cdots, \mu_p/\sigma$ must be small to insure that $\sum_1^p \mu_i^2/\sigma^2$ will be small. This will be rare when p is large. The larger the potential gain, the more unlikely it will be achieved.

The far more common situation in most realistic general linear model settings is that some unknown subset of the means $\mu_1, \cdots, \mu_p$ is small and can be neglected. Indeed, the determination of such a subset is often a principal strategy for building parsimonious models. Thus, what is needed are improved variance estimators which exploit the possibility that for any unspecified subset $A \subset \{1, 2, \cdots, p\}$, $\sum_A \mu_i^2/\sigma^2$ is small. Such estimators should be obtainable as generalizations of the estimators of Stein, Brown and Brewster and Zidek. These generalizations would be of the form

$$(1.6) \quad \psi(Z)S^2,$$

where $\psi(Z)$ is bounded above by $1/(\nu + 2)$, and is decreasing in $\sum_A Z_i^2$ for any $A \subset \{1, 2, \cdots, p\}$. What is desired is that when $\sum_A Z_i^2$ is small for some particular A, which would be likely when $\sum_A \mu_i^2/\sigma^2$ is small, these estimators would "shrink" $S^2/(\nu + 2)$ regaining $|A|$ degrees of freedom ($|A|$ is the size of the set A).

Perhaps the most straightforward such generalization would be that of Stein's estimator (1.3) to

$$(1.7) \quad \min\{T_1, \cdots, T_N\}, \quad T_j = \frac{S^2 + \sum_{A_j} X_{\nu+i}^2}{\nu + |A_j| + 2},$$

where $A_1, \cdots, A_N$ are the $N = 2^p$ distinct subsets of the integers $\{1, 2, \cdots, p\}$. Brown's estimator can also be straightforwardly generalized to one of the form

$$(1.8) \quad c_j S^2 \quad \text{when } Z \in B_j,$$

where $B_1, \cdots, B_N$ are a partition of $\mathbb{R}^p$, and $c_1, \cdots, c_N$ are fixed prespecified constants whose selection could be based on risk calculations similar to those of Brown. Of course, any such estimators would be nonanalytic and hence inadmissible.

I believe a more promising approach would be to consider what I would call multiple shrinkage generalizations of the Brewster–Zidek estimators. As opposed to the estimators of Stein and Brown, the Brewster–Zidek estimators "shrink" $S^2/(\nu + 2)$ smoothly as a function of $\|Z\|^2$. Furthermore, the Brewster–Zidek estimator for the general linear model context (1.1) can be obtained as a generalized Bayes estimator for σ^2 using a prior on $\eta = (\mu_1/\sigma, \cdots, \mu_p/\sigma)'$. Thus, it may be possible to obtain a multiple shrinkage estimator as a Bayes estimator for a finite mixture of related priors. What is needed is for each subset $A_j \subset \{1, 2, \cdots, p\}$, a prior π_j on η which yields as the Bayes estimator, a Brewster–Zidek analog which "shrinks" $S^2/(\nu + 2)$ whenever $\sum_{A_j} Z_i^2$ is small. If such priors could be found, the multiple shrinkage generalization of the Brewster–Zidek estimator could be obtained as a Bayes rule for a mixture prior of the form

$$(1.9) \quad \pi_*(\eta) = \sum_1^N p_j \pi_j(\eta),$$

where $p_1, \cdots, p_N$, $(N = 2^p)$, are a set of prespecified prior weights. The resulting estimator would be an adaptive convex combination of different Brewster–Zidek estimators which would put most weight on the estimator "shrinking" most. My experience with related multiple shrinkage estimators in the context of estimating a multivariate normal mean (see George, 1986a, b), suggests that such estimators might offer substantial risk reduction in a much larger region of the parameter space compared to any one Brewster–Zidek estimator.

2. FURTHER IMPROVEMENTS TO SHRINKAGE ESTIMATORS OF THE MEAN

The second research direction I would like to discuss concerns bringing the improved variance estimators discussed by Maatta and Casella to bear on a different problem. I would like to consider using these variance estimators to further improve shrinkage estimators of a multivariate normal mean in the following context. Suppose we have n multivariate normal observations,

$$Y_1, \cdots, Y_n \sim \text{iid } N_p(\mu, \sigma^2 I), \tag{2.1}$$

where the $p \times 1$ vector of means μ and the scalar σ^2 are unknown. The problem here is to find an estimator δ of μ which yields small risk

$$R(\mu, \delta) = E_\mu \| \mu - \delta \|^2. \tag{2.2}$$

By sufficiency, this problem can be reduced to estimating μ on the basis of

$$\bar{Y} = \frac{1}{n} \sum_1^n Y_j \quad \text{and} \quad S^2 = \sum_1^n \| Y_j - \bar{Y} \|^2, \tag{2.3}$$

where $\bar{Y}$ and S are independently distributed as

$$\bar{Y} \sim N_p(\mu, (\sigma^2/n)I) \quad \text{and} \quad S^2 \sim \sigma^2 \chi^2_{p(n-1)}. \tag{2.4}$$

The traditional "straw man" for this problem has been the MLE and worst minimax estimator, namely $\delta^{\text{MLE}} = \bar{Y}$. When σ^2 is known, δ^{MLE} is uniformly dominated in risk by shrinkage estimators such as the James–Stein estimator

$$\delta^S = \left(1 - \frac{(p-2)\sigma^2/n}{\| \bar{Y} \|^2}\right) \bar{Y}. \tag{2.5}$$

When σ^2 is unknown, the recommended substitute for (2.5) which also dominates δ^{MLE}, is

$$\hat{\delta}^S = \left(1 - \frac{(p-2)\hat{\sigma}^2/n}{\| \bar{Y} \|^2}\right) \bar{Y}, \tag{2.6}$$

where

$$\hat{\sigma}^2 = S^2/(p(n-1)+2) \tag{2.7}$$

(see Stein, 1966). One can think of $\hat{\delta}^S$ as an estimate of δ^S based on estimating σ^2 by $\hat{\sigma}^2$.

Notice however that the context (2.4) is (after suitable transformation), just the canonical context (1.1). In fact, the estimator $\hat{\sigma}^2$ is precisely the "straw man" $S^2/(\nu + 2)$ in (1.5) which is dominated by the estimators of Stein, Brown, and Brewster and Zidek. I would like to suggest that it may be possible to further improve (2.6) by using one of these estimators in place of $\hat{\sigma}^2$. More precisely, letting $Z = \sqrt{n}\,\bar{Y}/S$, it would be interesting to consider shrinkage estimators of the form

$$\hat{\delta}^* = \left(1 - \frac{(p-2)\hat{\sigma}_*^2/n}{\| \bar{Y} \|^2}\right) \bar{Y}, \tag{2.8}$$

where

$$\hat{\sigma}_*^2 = \phi(Z) S^2 \tag{2.9}$$

is one of the improved estimators of Stein, Brown, or Brewster and Zidek. A reason to suspect that $\hat{\delta}^*$ may be successful is that both δ^S and $\hat{\sigma}_*^2$ operate by exploiting the possibility that $\| \mu \| / \sigma$ is small. Indeed, both of these estimators shrink the traditional estimates when $\| Z \|^2$ is small. Thus, $\hat{\delta}^*$ will shrink less abruptly than $\hat{\delta}^S$. Note that $\hat{\delta}^S$ and $\hat{\delta}^*$ will differ most when n is small and p is large.

Of course, the proof of the pudding is whether or not an estimator of the form $\hat{\delta}^*$ will yield desirable risk reduction over δ^{MLE}. An advantage of considering estimators like $\hat{\delta}^S$ is that $\bar{Y}$ and $\hat{\sigma}^2$ are independent, so that the same risk calculations used to evaluate δ^S may be carried out conditionally. Unfortunately, $\bar{Y}$ and $\hat{\sigma}_*^2$ are not independent so that more formidable risk calculations will probably be required. Nonetheless, shrinkage domination can be shown in situations with dependent variance estimates (see George, 1988).

3. CONCLUSION

In closing, I hope my comments have underscored the potential of the contribution by Maatta and Casella for stimulating new ideas. There are now many new directions to explore.

ADDITIONAL REFERENCES

George, E. I. (1986a). Minimax multiple shrinkage estimation. *Ann. Statist.* **14** 188–205.

George, E. I. (1986b). Combining minimax shrinkage estimators. *J. Amer. Statist. Assoc.* **81** 437–445.

George, E. I. (1988). Shrinkage domination in the multivariate common mean problem. Technical Report 72, Statistics Research Center, Graduate School of Business, Univ. Chicago.

Stein, C. (1966). An approach to the recovery of inter-block information in balanced incomplete block designs. In *Research Papers in Statistics, Festschrift for J. Neyman* (F. N. David, ed.) 351–366. Wiley, New York.