BROWN, L. D. (1978). A contribution to Kiefer's theory of conditional confidence procedures. *Ann. Statist.* **6** 59–71.
BROWN, L. D. and SACKROWITZ, H. (1984). An alternative to Student's t test for problems with indifference zones. *Ann. Statist.* **12** 451–469.
CASELLA, G., HWANG, J. T. and ROBERT, C. (1989). Loss functions for set estimation. Technical Report, Cornell Statistics Center.
DEY, D. K. and SRINIVASAN, C. (1985). Estimation of a covariance matrix under Stein's loss. *Ann. Statist.* **13** 1581–1591.
KIEFER, J. (1977). Conditional confidence statements and confidence estimators (with discussion). *J. Amer. Statist. Assoc.* **72** 789–827.
KIEFER, J. and SCHWARTZ, R. (1965). Admissible Bayes character of T^2-, R^2-, and other fully invariant tests for classical multivariate normal problems. *Ann. Math. Statist.* **36** 747–770.
LEHMANN, E. L. and STEIN, C. (1948). Most powerful tests of composite hypotheses. *Ann. Math. Statist.* **19** 495–516.

Comment

Arthur Cohen

A historical perspective on one of the more fascinating and intriguing theoretical results of statistics is most welcome. I have some comments that are concerned with rounding out the story and generalizations.

If one requires an estimator which is both location invariant and scale equivariant, then the best equivariant estimator of σ^2 with respect to squared error loss is $S^2/(n+1)$, and of course it is admissible within its class. For the confidence interval problem, if the vector loss L_1 = (0–1 for correct coverage or not, length) is replaced by the vector loss L_2 = (0–1 for correct coverage or not, 1–0 for covering false values or not), then the "usual" confidence interval is admissible. This latter fact follows from the duality of hypothesis testing and confidence intervals. In this problem and in several other interesting problems, the following pattern holds for the "usual" procedure: admissible as a test and hence admissible as a confidence interval for the vector loss L_2; inadmissible as a confidence interval for the vector loss L_1 and inadmissible as a point estimator for squared error (or other) loss. Table 1 indicates some problems where this pattern holds. A stands for "admissible" and I for "inadmissible." For the problem of estimating the normal mean vector see Stein (1956) and Brown (1966). For the common mean problem see Brown and Cohen (1974) and Cohen and Sackrowitz (1977). For the normal quantile problem see Zidek (1971), and for the Poisson problem see Clevenson and Zidek (1975).

There is a substantial amount of work in decision theory on estimating a normal covariance matrix or generalized variance. Again the "usual" estimators are inadmissible for reasonable loss functions. In some cases, the sample mean can be used as in the univariate case to get "help." This is the situation in papers by Sinha and Ghosh (1987) and Sarkar (1989). In other cases there is a "Stein" or dimensional effect and improvements can be made even without using information from the sample mean (see the survey paper of Lin and Perlman, 1985).

The statement in the paper that Stein knew his estimator was not admissible is a bit confusing. Stein may have speculated that the generalized Bayes estimators form a complete class as is the case of some one parameter exponential family models (see Sacks, 1963). The basis of the conjecture then is that, since it cannot be generalized Bayes because it lacks smoothness properties, it is inadmissible. As it turns out Stein's estimator is easily beaten and that is why it is inadmissible. It is not known whether the class of generalized Bayes estimators for problems with unbounded nuisance parameters is a complete class except in isolated examples where it is not true.

Although Brown (1968) is already referenced, it is important to note that his paper contains many results

TABLE 1
Admissibility status of "usual" procedure

Problem	Type of inference			
	Confidence set			Point estimation squared error loss
	Testing	Loss L_2	Loss L_1	
Normal variance	A	A	I	I
Normal mean vector of dimension 3 or more	A	A	I	I
Common mean of two independent normals	A	A	I	I
Normal quantile	A	A		I
Independent Poisson parameters of dimension 2(3)	A	A		I

Arthur Cohen is Professor, Department of Statistics, Hill Center, Busch Campus, Rutgers University, New Brunswick, New Jersey 08903.

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.