ADDITIONAL REFERENCES

BERGER, J. O. (1985a). In defense of the Likelihood Principle: Axiomatics and Coherency. In *Bayesian Statistics 2* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 33–66. North-Holland, Amsterdam.

BERGER, J. O. (1988). An alternative: The estimated confidence approach. Discussion of "Conditionally acceptable frequentist solutions" by G. Casella. In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds.) **1** 85–90. Springer, New York.

Comment

Lawrence D. Brown

It is a pleasure and an embarrassment to read a historical story in which one plays an integral role. From my perspective the story has been accurately related, but I do have some miscellaneous comments to make which are related to the general topic.

THE LOSS FUNCTION

The point estimation segment of this article deals exclusively with the loss function (1.5)—i.e., $L(\delta, \sigma^2) = ((\delta/\sigma^2) - 1)^2$. Although this loss is relatively easy to handle analytically it seems somewhat inappropriate for a broad range of applications. Let me repeat informally some thoughts I tried to convey formally in Brown (1968).

Strictly from a qualitative point of view, the loss (1.5) is very skewed. Note that $\lim_{\delta\to\infty} L(\delta, \sigma^2) = \infty$ but $\lim_{\delta\to 0} L(\delta, \sigma^2) = 1$. Hence overestimation of σ^2 is much more severely penalized than underestimation. Furthermore, the best invariant estimator for this loss is $S^2/(\nu + 2)$, which is smaller than the maximum likelihood value of $S^2/(\nu + 1)$, or the intuitively appealing and best unbiased estimator, which is S^2/ν. One rationalization for this discrepancy could be that the intuition supporting use of S^2/ν is in error; that (1.5) is the actual loss and that therefore $S^2/(\nu + 2)$ is to be preferred to S^2/ν (and Brewster and Zidek's (2.20) is then to be preferred to $S^2/(\nu + 2)$).

However, another interpretation is possible. Note that historically use of S^2/ν was proposed and, presumably, found generally satisfactory without elicitation of or reference to a specific loss function. If indeed S^2/ν is satisfactory among invariant procedures perhaps that is because it matches the actual (but subconscious) loss function measuring the experimenters' preferences. Thus one asks, "For what loss is the best unbiased estimator, S^2/ν, also best invariant?".

Stein (1964) found a loss function for which S^2/ν is best invariant. It is

(1) $$L_S(\delta, \sigma^2) = \delta/\sigma^2 - \ln(\delta/\sigma^2) - 1.$$

Note that $L_S(\delta, \sigma^2) \ge 0$ and attains the value 0 uniquely at $\delta = \sigma^2$. Also, $L_S(\delta, \sigma^2)$ is strictly convex in δ and $\lim_{\delta\to 0} L_S(\delta, \sigma^2) = \lim_{\delta\to\infty}(L_S(\delta, \sigma^2) = \infty$. Thus L_S has a number of pleasing qualitative properties.

In Brown (1968) more was established about L_S. It was shown that for virtually any problem of estimating a single scale parameter the best unbiased estimator is also best invariant for this loss, and the loss function L_S is *the only loss function possessing this global property* (up to affine transformations, which do not affect admissibility). Thus a belief in the suitability among invariant estimators of the best unbiased estimator is *equivalent to* a belief in the suitability of L_S. In summary, my own feeling is that the loss L_S is the most appropriate for *general* studies of estimation of scale parameters. (Of course, other loss functions may be appropriate in *specific* applications.)

The story related for loss (1.5) by Maatta and Casella applies equally to the loss L_S. The analog of Stein's estimator, (2.4), is $\tilde{\delta}(\bar{X}, S^2) = \tilde{\phi}(Z^2)S^2$, where

(2) $$\tilde{\phi}(Z^2) = \min\left(\frac{1}{\nu}, \frac{1 + Z^2}{\nu + 1}\right).$$

Under L_S this estimator dominates the usual S^2/ν.

Loss L_S is explicitly considered in Brown (1968) where it is shown (as in (2.16)) that the choice

(3) $$\tilde{\phi}^*(Z^2) = \begin{cases} \tilde{c}_{0,1}(r^2,) & \text{if } Z^2 \le r^2, \\ 1/\nu & \text{if } Z^2 > r^2 \end{cases}$$

yields an estimator better than S^2/ν when

$$\tilde{c}_{0,1}(r^2) = 1/E_{0,1}(S^2 \mid Z^2 \le r^2).$$

The algorithm of Brewster and Zidek then applies, and shows the estimator with

(4) $$\tilde{\phi}^{**}(Z^2) = \tilde{c}_{0,1}(Z^2)$$

Lawrence D. Brown is Professor, Department of Mathematics, and Director of the Cornell Statistics Center, Cornell University, Ithaca, New York 14853-7901.