

# Comment

James O. Berger

The article is an extremely lucid discussion of the variance estimation problem, highlighting the major intertwining theoretical developments. Numerous interesting conceptual issues are raised in the article but, before turning to these, the issue of importance to practitioners needs further elaboration.

The authors finish Section 2 by saying “the maximum relevant improvement of Rukhin’s estimators . . . is only 4%, suggesting that there would be very little practical benefit associated with these improved estimators . . . . However, as we shall see in Section 5, there are interesting cases where substantial improvement is possible.” Turning to Section 5, we see (in Table 1) only 2%, 3%, at most 5.3%, improvement in confidence interval length. In what way are these improvements more “substantial” than Rukhin’s, especially since they are achieved only for the less realistic generalized linear model with common unknown  $\sigma^2$ ?

The author’s replies will likely be that it is more difficult to obtain improvement in the “length of confidence set” problem than in the “point estimation” problem, and to reiterate that any guaranteed improvement is useful, no matter how small. I would agree with these points, but feel it is important to acknowledge that only small gains are available here. Careful study of these developments by practitioners is not very cost effective as compared to, say, careful study of the more familiar “shrinkage estimation of means” literature, in which practical improvements in excess of 50% are common. Practitioners have the time to access only a small fraction of the theory and methodology that is produced, and I have always thus felt it to be imperative to carefully indicate the likely practical potential of new developments.

Conceptually, there are many fascinating developments in the paper. For instance, starting with the “hard to take seriously” problem of estimating  $\sigma^2$  with quadratic loss (which I would say is hard to take seriously because decision-theoretic estimation of  $\sigma^2$  is very sensitive to the loss, and there are many different reasonable losses), the authors push the developments toward the easy-to-take-seriously problem of obtaining reduced length confidence sets. Insisting,

---

*James O. Berger is the Richard M. Brumfield Distinguished Professor of Statistics. His address is Department of Statistics, Purdue University, West Lafayette, Indiana 47907.*

at the same time, on good conditional performance of the resulting sets is a further bold step toward realism. All in all, the development is an exemplary illustration of how to turn an originally abstract decision-theoretic benefit into a clearly concrete inferential gain.

As a Bayesian reading the paper, I naturally got excited that the most useful confidence sets were actually generalized Bayes credible sets, and that the posterior probabilities of the sets appeared plausible. For instance, a Bayesian might well be very happy with using  $\langle I_S, \gamma_0(\bar{X}, S^2) \rangle$  or  $\langle I_T, \gamma_T(\bar{X}, S^2) \rangle$  (see the discussion following (4.14) and (4.15)). A frequentist might also be very happy with these reports, especially if  $E_{\mu, \sigma^2} \gamma(\bar{X}, S^2) < \text{Coverage Probability}$ , for all  $\mu, \sigma^2$ , since then  $\gamma$  has complete frequentist justification as a report (cf. Berger, 1988). It would be interesting to know if this is so.

Let me finish by raising a few old pet peeves:

- (i) Following (3.4): “The choice of interval should depend on more than just ease of calculation, which is the only favorable factor associated with  $I_{ET}$ .” To the contrary, I would argue that  $I_{ET}$  (equal-tailed interval) is often most favored because (a) it conveys the uncertainty present in each tail separately, and (b) it is invariant under all monotonic transformations.
- (ii) The “betting” approach to evaluating conditional behavior: I tend to view this approach with suspicion (cf. Berger, 1985a), feeling that confidence sets are constructed to communicate something, not to provide a basis for betting. If one really takes betting on confidence sets seriously, one is quickly led, via usual axiomatics, to Bayesianism (which is the only thing that makes me at all sympathetic to the betting framework).
- (iii) “Positive bias” in conditional coverage is, in several places, stated to be of lesser concern. I am uncomfortable with this. I can think of many problems in which it is much more serious, in the practical context, to underestimate coverage than to overestimate coverage. We are all taught in our statistical education that it is okay to err on the conservative side, but I am not sure that this is indeed generally true in practical decision contexts where statistics is used.