PETER HALL

Australian National University

Pressure of time prevents me from making the detailed comments I would like on this extraordinarily stimulating paper. I shall make just three points. The first is rather trite, but in view of some of the comments in the paper I feel it is worth airing.

(i) In comparisons of various "unweighted" techniques in the presence of nonexchangeability, surely the striking thing is not that certain of them perform poorly, but that some of them perform alright. I was not surprised at the poor performance of the bootstrap in non-i.i.d. contexts.

(ii) Professor Wu's remarks on the predictions of asymptotic theory of coverage probability (e.g., point 3, Section 10) deserve comment. I note that Professor Wu is constructing two-sided confidence intervals. Edgeworth expansions predict that even the simple-minded standard normal approximation to $(\hat{\theta} - \theta)/\hat{\sigma}$ will have an error in coverage probability of only $O(n^{-1})$, since *the skewness terms of order* $n^{-1/2}$ *cancel* from the coverage probability of any "reasonable" two-sided interval. I do not know the exact form of the Edgeworth expansions for all the statistics in, say, Table 3, but it would be a safe bet that each results in a zero term of order $n^{-1/2}$ and a nonzero term of order n^{-1} in the expansion of coverage probability of a two-sided interval, when variances are equal. From this point of view, theory is entirely equivocal on the matter of predicting relative performance of two-sided confidence intervals constructed by the various methods. It certainly does not favour the bootstrap. One would have to carefully dissect the $O(n^{-1})$ term to reach a more definitive conclusion. Experience suggests that such a dissection of a high-order term is likely to be misleading for samples as small as 12, particularly in complex cases such as this one.

(iii) Judging from formula (2.10) in Section 2, and the discussion preceding that formula, Professor Wu is using the percentile method without employing the bootstrap variance estimate to standardize for scale. In the case of *one*-sided confidence intervals, this procedure results in an unwanted skewness term of order $n^{-1/2}$ in the Edgeworth expansion of coverage probability. Intuition suggests that such a procedure is not the best thing to use when constructing *two*-sided intervals.

In more detail, suppose $\hat{\theta}$ and $\hat{\theta}^*$ are a standard parameter estimate and a bootstrap parameter estimate, and $\hat{\sigma}^2$ and $\hat{\sigma}^{*2}$ are a standard and a bootstrap estimate of the variance of $\hat{\theta}$, respectively. Let θ and σ^2 be the true parameter value and true variance of $\hat{\theta}$, respectively. It may be shown in a variety of ways that

(1) conditional distribution of $(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}$ approximates distribution of $(\hat{\theta} - \theta)/\sigma$

and

(2) conditional distribution of $(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^*$ approximates distribution of $(\hat{\theta} - \theta)/\hat{\sigma}$.

A little reflection, bearing in mind the conditioning argument, makes these statements seem almost tautological. In each case conditioning is on the sample, and the approximations are good up to terms of *smaller* order than $n^{-1/2}$. It is not true that the conditional distribution of $(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}$ is a good approximation to the distribution of $(\hat{\theta} - \theta)/\hat{\sigma}$. In this case the Edgeworth expansions of coverage probability for one-sided confidence intervals differ by terms of order $n^{-1/2}$. The same conclusion may be reached intuitively, noting that the statistic $(\hat{\theta} - \theta)/\sigma$ is not pivotal if σ is unknown. Work in Singh (1981), for example, concerns the approximation in (1) although I know of some authors who have tried to use it to promote an approximation of the distribution of $(\hat{\theta} - \theta)/\hat{\sigma}$ by that of $(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}$.

I should make one final remark to tie these comments to those made by Professor Wu prior to his formula (2.10). Since the conditioning in (1) and (2) is on the sample, then $\hat{\theta}$ and $\hat{\sigma}$ are effectively constant, and so the conditional distribution of $(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}$ is just a location and scale change of that of $\hat{\theta}^*$.

REFERENCE

SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195.

DEPARTMENT OF STATISTICS
AUSTRALIAN NATIONAL UNIVERSITY
GPO BOX 4
CANBERRA ACT 2601
AUSTRALIA

DAVID HINKLEY[1]

University of Texas at Austin

Professor Wu is to be congratulated for making a significant advance in jackknife methodology. The general use of information measures to determine weights in subsampling schemes is surely correct, and the implementation here for regression is most interesting.

The one somewhat negative conclusion of the paper concerns the comparatively poor performance of the bootstrap. It is to this that I shall address my remarks, because the bootstrap approach has, quite innocently, been misapplied. Good results *can* be obtained with bootstrap methods, as I hope to explain with the help of relatively simple examples.

The first point has to do with conditional probability, which in the regression context arises from conditioning on the experimental vector $\mathbf{x}$ of explanatory variables. The key issue can be seen most easily in the simple linear regression

[1]Research supported by National Science Foundation grant DMS-8505769.