DISCUSSION

Ralph E. Folsom, Research Triangle Institute

I will begin by congratulating the authors collectively for a very creative and thought provoking set of papers. The chairman also deserves credit for assembling a set of papers that compliment each other so well. The Dippo-Wolter and Royall-Cumberland papers aptly demonstrate just how poorly symetric intervals may perform for small to moderate size samples. The asymetric bootstrap intervals explored in the Rao and Wu paper provide an attractive alternative solution.

I particularly enjoyed studying the Rao and Wu paper. They have provided an excellent tutorial on how bootstrap methods should be applied to probability samples. Their bootstrap results for general unequal probability designs and two stage designs represent a breakthrough for replication variance technology. Previous applications of replication methodology to unequal probability designs have been wedded to the variance formula for with replacement primary unit selections. I have also recently developed some analogs of BRR and Jackknife methods that reproduce the Yates-Grundy-Sen variance estimator for linear statistics. These extensions also remove the equal stratum sample size restriction for BRR (Folsom, 1984).

There are two additional points that I would like to raise regarding the Rao and Wu results. The first relates to their assessment of the bootstrap variance estimator for combined ratios. Their asymptotic bias evaluation shows that including first and second order terms the bias is equivalent for the bootstrap and BRR variance estimators in the common paired selection design. If, for a particular application, one has reason to believe that a symetric t-interval will perform adequately, then I would argue that this second order asymptotic equivalence result provides strong support for prefering the more computationally efficient BRR variance estimator.

My second comment has to do with the proposed asymetric bootstrap t-interval. This would appear to be a promising method for dealing with the badly skewed t-distributions observed in the other two papers. I am somewhat skeptical however about the likely performance of such intervals in deeply stratified samples. The sample t-statistic can be viewed as a nonlinear function of U-statistics, with the square root of a degree 2 variance statistic in the denominator. Since the variance of the denominator has a strong influence on the shape of the associated t-distribution, one would hope that the bootstrap variance of the denominator would be reasonably unbiased. To the contrary, with many designs having 4 or fewer selections per strata, small sample U-statistics theory strongly suggests that the variance of the bootstraped t distribution may be substantially overestimated. This could tend to make the bootstraped intervals too wide. In my opionion, much more theoretical and empirical work is needed before the purported robustness of asymetric bootstrap intervals is validated.

Turning to the Dippo and Wolter paper, I would like to strongly recommend their second order analytic approach for evaluating alternative variance estimators. I am hopeful that results of this nature will begin to clarify the mixed results that have shown up in previous monte carlo studies. In combination with Rao and Wu's recent second order bias analysis, some interesting general results are emerging. Rao and Wu's discovery that the bias of the linearization and jackknife variance estimators are equivalent including second order terms for the paired selection design is a notable example.

The empirical evaluation of the Taylor Series convergence in the Dippo and Wolter paper gave a somewhat mixed picture. The fact that the second order contributions averaged less than 1% of the first two terms for the smallest sample size was comforting. On the other hand, the monte carlo results suggested that the higher order contributions might exceed 10 percent for several of the statistics in the small sample size evaluation. I would like the authors to give their reaction to the convergence issue.

To mention a few critical points, I was disappointed that the authors didn't include the linearization variance estimator in their analytical and empirical evaluation. Rao and Wu's analysis shows that in terms of minimum bias, the delta estimator is a very strong competitor for combined ratios. On the other hand, the random group methods with 2 or 4 groups are not very strong competitors. I was also surprised that the authors seemed to expect convergence to normality for the random group estimators. I would have preferred to see comparisons against the t distribution with appropriate degrees of freedom; that is, df = 1, 3, and 7.

The most striking and disturbing result in the Dippo and Wolter paper is the uniformly poor performance of symetric t-intervals for combined ratio statistics with indicator variable denominators. The high positive correlation between the sample ratio and sample variance estimator seems to be the common culprit here as well as in the Royall and Cumberland paper.

Turning to the Royall and Cumberland paper, I would like to congratulate the authors on an excellent job of presenting their thought provoking results with great clarity. I do wish that they would have taken the recommendation of John Rao and other reviewers of their earlier paper and considered the linearization variance estimator based on the delta variance approximation for the associated ratio statistic. While I realize that this paper is intended as a direct sequel to their 1981 paper, I would also have appreciated a more informative competition between the author's nonprobability sample designs for achieving balance and traditional stratified probability sampling methods. For the simple random samples explored by the authors, the distinction between robust prediction theory variance estimators and probability sampling theory methods is blurred.

I would contend that the probability sampling camp has at least as legititimate a claim on the jackknife estimator as does the prediction theorists. Otherwise, I thought the authors made a strong case for well balanced sample designs when good auxillary variables are available on the frame. Royall and Cumberland also succeed in damning the least squares variance estimator along with the standard variance approximations presented for the ratio and regression total estimators in sampling texts. Royall and Cumberland's contention that robust variance estimators should be used for regression based inference in all branches of statistical application, is an admonition that I heartily endorse. In this respect, I also believe that the linearization or delta method deserves more attention.

Again, the most dramatic aspect of the Royall and Cumberland results is the failure of symetric t-intervals for their City and 1970 county populations. These results suggest that a usefull diagnostic for symetric intervals would be a robust estimator for the correlation between the sample statistic and its estimated standard error. When this correlation is high, asymetric bootstrap intervals may provide a robust alternative. On the other hand, I would caution that the validity of bootstrap interval robustness claims has yet to be convincingly demonstrated.

Reference

Folsom, R. E. (1984). Probability Sample U-Statistics: Theory and Applications for Complex Sample Designs. Institute of Statistics Mimeo Series No. 1464, Chapel Hill, N.C.