DISCUSSION

C. F. J. Wu, University of Wisconsin-Madison

Let me first congratulate the speakers for their interesting and novel works. Their presentations cover a broad range of topics in variance estimation from complex surveys. Folsom develops a theory of U statistics for complex survey data. Rust examines the theoretical aspect of variance estimation using reduced replication. Valliant studies model-based inference for stratified samples. These topics complement each other very well. It is interesting to notice that all the speakers are relatively recent Ph.D.'s and, prior to their doctoral studies, had working experience in government agencies. They are both practitioners and theoreticians. Dr. C. Dippo, the session's organizer, deserves special thanks for putting together such an interesting session.

Folsom. His work is, in my opinion, significant and fundamental. Prior to his work, the only theory available for variance estimation from complex survey samples is for the linear statistics or their smooth functions (Kish and Frankel, 1974; Rao and Wu, 1985, 1986). As an example, his extension of the Yates-Grundy-Sen estimator to U-statistic is significant, although its use in practice may be limited since it requires the knowledge of higher-order inclusion probabilities. Empirical comparisons of his proposal and other estimators are needed. For example, it was found in a recent study (Rao and Wu, 1986) that the bootstrap never performs spectacularly. More work is needed to bridge the gap between asymptotic results and finite sample performance.

Next I would like to relate his generalized BRR to a "bootstrap" proposed in Rao and Wu (1984, 1986) for probability samples. Our bootstrap method consists of repeated computations of

$$\tilde{Y} = \hat{Y}_{HT} + \sum_{(i,j)=1}^{m} \sqrt{w_{ij}}\,(z_i - z_j)\ ,$$

$$z_i = \frac{y_i}{\pi_i},\quad w_{ij} = \frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\ , \qquad (1)$$

and the pairs (i,j) are selected randomly with replacement from the $n(n-1)$ pairs. Since the pairs (i,j) and (j,i), $i \neq j$, are selected with equal probability in (1), we may consider a modification of (1),

$$\tilde{Y}^* = \hat{Y}_{HT} + \frac{1}{\binom{n}{2}} \sum_{i<j} \delta^*_{ij}\sqrt{w_{ij}}\,(\pi_i - \pi_j)\ , \qquad (2)$$

where the summation is over all the ordered pairs $i < j$ and δ^*_{ij} are independent with $\text{Prob}(\delta^*_{ij} = \pm 1) = \frac{1}{2}$. This method may be called a generalized random repeated replication (RRR). If $\delta^*_{ij} = \pm 1$ in (2) are chosen more systematically such as from a Hadamard matrix, it reduces to Folsom's generalized BRR. Details of this connection may be found in Rao and Wu (1986).

My appreciation of his work may be best served by urging him to make his work available soon.

Rust. Grouping units and/or strata to save computation is a common practice even with today's computing capacity. Rust's paper addresses the important question of how grouping should be done. His main technical tool is the use of degrees of freedom as a measure of goodness of the t-approximation. Such an approach has enabled him to obtain guide on the choice of grouping procedures. The utility of his results depend on the adequacy of degrees of freedom as a measure of the t-approximation. Other measures may include the bias of v and the correlation between v and $\hat{\theta}$. They seem to be more difficult to manipulate.

The author argues in favor of small G (i.e. grouping of more strata) especially for domain estimation (Section 5). Consider the extreme case of $G = 1$ and $n_h = 2$, each replicate becomes a half-sample. Will balanced half-samples be a better method? Even for small $G > 1$, certain degree of balancing in the choice of replicates may improve the efficiency. Is this aspect covered by his theory? My next technical comment is on the method of combining strata for BRR (Section 7). The author's proposal of equalizing $\sum_{h\in g} W_h^2\sigma_h^2$ over the combined strata is very interesting. This is in contrast with Valliant's proposal of α-balance. It would be interesting to compare the various approaches.

My last comment, perhaps a more philosophical one, is on the choice between methods based on resampling such as JRR and BRR and the linearization method. If the formula for the latter method is available, which is often the case, would it be necessary to consider JRR and/or BRR, which are more computer-intensive and may be difficult to develop for complex designs?

Valliant. Valliant's work on model-based inference for stratified samples is a natural extension of work by Royall and coauthors to a more interesting and realistic setting. This is especially welcome since the impact of Royall's theoretical work could be greater if more realistic problems were addressed.

My main complaint is on the author's choice of v_L for empirical and theoretical comparison. Take the ratio estimator as an example. He uses

$$v_L = \sum_h N_h^2(1 - f_h) \sum_{s_h} r_{1hi}^2/[n_h(n_h - 1)]$$

as *the* linearization variance estimator, while another linearization variance estimator

$$v_L' = \sum_h N_h^2(1 - f_h)\left(\frac{\bar{X}_h}{\bar{x}_{hs}}\right)^2 \sum_{s_h} r_{1hi}^2/[n_h(n_h - 1)]$$

is ignored. In the case of simple random sampling, both v_L and v_L' can be derived from linearization (Cochran, 1977). Empirical and theoretical comparison of v_L and v_L' has been given in a series of papers (Wu, 1982, 1985; Wu and Deng, 1983). In general v_L' performs better than v_L. This is also supported by the model-based argument of Royall and Cumberland (1981). The omission of v_L' in the study, which is apparently inherited from Royall, makes his case less convincing. Some of the conclusions drawn would be less dramatic if v_L' were included in the comparison.

My next question is on the generality of conclusions from his empirical study. The natural populations chosen are well suited for the theory. What would happen if some key assumptions such as $\alpha_h = 0$, $\mathrm{var}(y) \propto x$, break down? It would be useful to include in the study other populations which incorporate such features. I am somewhat disappointed with the inability of the model-based theory to distinguish v_B, v_J and v_L from each other. One exception is the concept of α-balance, $\bar{x}_{h1} = \bar{x}_{h2}$. Note that its derivation is based on a model with a linear term in x_{hi}. If the model also includes a quadratic term in x_{hi}, quadratic balance $s_{h1}^2 = s_{h2}^2$ may also be required. Fortunately the "basket method" described in Section 5, by its very nature, will achieve both mean- and variance-balance.

My last comment is on the choice of variance estimator as suggested by the empirical study. The jackknife v_J appears to have the best overall performance. Its main disadvantage relative to the others under comparison is the computational burden. A method that retains the efficiency of v_J but requires less computation is desired. One such method that is related to but different from $v_B(\mathrm{bas})$ is proposed as follows: (i) take a half-sample in each stratum randomly subject to mean-balance (and, if possible, variance-balance); (ii) repeat (i) L times to get L replicates. Note that, in $v_B(\mathrm{bas})$, grouping to achieve mean-balance is done once and then replicates are formed systematically based on an orthogonal matrix, whereas, in the proposed method, grouping subject to mean-balance is done repeatedly and independently.

Acknowledgement. The research is supported by the National Science Foundation under Grant No. DMS-8502303.

Additional References

Rao, J. N. K. and Wu, C. F. J. (1984). Bootstrap inference for sample surveys. Proceedings of the American Statistical Association, Section on Survey Research Methods, 106-112.

Rao, J. N. K. and Wu, C. F. J. (1986). Resampling inference with complex survey data, Technical Report, University of Wisconsin-Madison.

Wu, C. F. J. (1982). Estimation of variance of the ratio estimator. Biometrika, 69, 183-189.

Wu, C. F. J. and Deng, L. Y. (1983). Estimation of variance of the ratio estimator: an empirical study. Scientific Inference, Data Analysis and Robustness (G. E. P. Box et al., eds.), 245-277, Academic Press.