Some Properties of Estimates of Variance Based on Replication Methods

Robert E. Fay, U.S. Bureau of the Census

1. INTRODUCTION

This paper collects a number of results on the application of replication methods to complex sample surveys. The topics considered vary in scope somewhat between different sections, but they are related by common concepts and definitions. Section 2 summarizes interesting results presented earlier by Efron and Stein (1981) and later extended by Karlin and Rinott (1982), and discusses their implications, with some additional extensions, to the problem of complex samples for some specific replication techniques, namely half-sample and random group methods. The results employ an analysis of variance decomposition, which also motivates the introduction of notions of linear and quadratic functionals in the context of complex samples.

While the results of section 2 apply only to a few familiar replication methods under limited conditions, section 3 takes up a quite different topic: the existence of replication methods (resampling plans) to represent the variance of linear functionals in virtually any situation for which a closed-form variance estimator exists for such functionals. This existence theorem simply draws the general conclusion anticipated by the large and varied number of adaptations that researchers have made to replication methods to fit specific situations in the past. The wide class of resampling plans introduced in section 3 may include some with more desirable properties in specific applications than ones now in general use.

Section 4 addresses the issue of estimation of bias through replication methods. Any of the replication methods introduced in section 3, if based on a design-unbiased estimator of variance, may be used to produce a design-unbiased estimator of a quadratic functional evaluated on the finite population.

While sections 2 through 4 discuss properties of replication for specific finite populations, section 5 discusses asymptotic properties of replication for inference. The discussion centers on the class of asymptotically normal estimators studied earlier by Binder (1983) with respect to linearization. Basically, replication offers an asymptotically equivalent alternative, although some conditions must be placed on the occurrence of extreme deviations among replicates.

2. INEQUALITIES FOR THE EXPECTED VALUES OF REPLICATION-BASED VARIANCE ESTIMATES

A frequent observation from empirical studies has been the tendency for the jackknife estimate of variance to overestimate the variance of nonlinear statistics on average. For a statistic $S_n(X_1,\dots X_n)$, symmetric in its arguments and with finite variance, Var S_n, for independent, identically distributed X_i, and its counterpart $S_{n-1}(X_1,\dots,X_{n-1})$ for samples of size n-1, Efron and Stein (1981) reexpressed this tendency in the form of the following two inequalities:

$$E(\mathrm{Var}_J(S_n)) \geq ((n-1)/n)\ \mathrm{Var}\ S_{n-1} \qquad (2.1)$$

$$((n-1)/n)\ \mathrm{Var}\ S_{n-1} \geq \mathrm{Var}\ S_n \quad ? \qquad (2.2)$$

which would imply

$$E(\mathrm{Var}_J(S_n)) \geq \mathrm{Var}\ S_n \quad ? \qquad (2.3)$$

where

$$\mathrm{Var}_J(S_n) = ((n-1)/n) \sum_{i=1}^{n} (S_{(i)} - S_{(\cdot)})^2 \qquad (2.4)$$

with

$$S_{(i)} = S_{n-1}(X_1,\dots,X_{i-1},X_{i+1},\dots X_n) \qquad (2.5)$$

$$S_{(\cdot)} = \sum_{i=1}^{n} S_{(i)}/n \qquad (2.6)$$

Theorem 1 of their paper establishes (2.1) for any linear or non-linear statistic; this equation relates the expected value of the jackknife estimate of variance to the variance of the statistic S_{n-1} based on n-1 observations. In some sense, it is natural to relate the jackknife estimate of variance for a sample size n to the properties of statistics for samples of size n-1, since (2.4), (2.5) and (2.6) involve only S_{n-1}.

Equation (2.2) is shown with a question mark, since, as Efron and Stein note in their paper, exceptions occur. If (2.2) holds, (2.3) follows immediately from (2.1). In addition, since the inequality in (2.1) is strict except for linear functionals (to be defined later in this section), (2.3) may hold in some situations in which (2.2) fails. Efron and Stein use an analysis of variance decomposition (a method originated by Hoeffding (1948)) of the statistic S_{n-1} to prove (2.1). They write

$$\mu = E\ S_{n-1} \qquad (2.7)$$

$$A_i(x_i) = E(S_{n-1}|X_i=x_i) - \mu \qquad (2.8)$$

$$B_{ii'}(x_i,x_{i'}) = E(S_{n-1}|X_i=x_i,X_{i'}=x_{i'}) - A_i(x_i) - A_{i'}(x_{i'}) - \mu \qquad (2.9)$$

for $i \neq i'$, with higher-order terms defined in a similar manner, in the same fashion as the analysis of variance. (The next order term in this decomposition starts with $E(S_{n-1}|X_i=x_i,X_{i'}=x_{i'}, X_{i''}=x_{i''})$ and subtracts three linear terms (2.8), three quadratic terms (2.9), and (2.7). The highest order term begins with $S_{n-1}(x_1,\dots x_{n-1})$, and subtracts all terms of lower order.) Consequently, S_{n-1} can be completely decomposed as the sum of such terms. Except for μ, all terms in this decomposition have mean zero and are uncorrelated. The orthogonality of terms permits useful decomposition of the variance of expressions such as $S_{(i)} - S_{(i')}$, which in turn leads to a relatively simple, and certainly elegant, proof of (2.1).

Efron and Stein (1981) and Efron (1982) discuss examples where (2.2) fails; another illustration, more similar to typical problems encountered in survey sampling, is to consider a positive random variable X, and a random variable Y with a conditional distribution given X=x having expected value rx and variance $\sigma^2 x^2$. In this situation, the ratio estimate based on n independent selections from this population

$$r_n = (\sum_i^n Y_i)/(\sum_i^n X_i) \qquad (2.10)$$

is an unbiased estimator for the population ratio R. If the distribution of X is non-degenerate, both (2.2) and (2.3) fail for n=2. For larger n, both may fail depending upon circumstances, or (2.2) may fail while (2.3) holds; the appendix discusses this example in more detail.

Although (2.3) may fail, another inequality relating the expected value of the jackknife variance estimate to an estimate based on all n observations always holds, namely

$$E(Var_J(S_n)) \geq Var\ S_{(\cdot)} \qquad (2.11)$$

(Theorem 2, Efron and Stein 1981). Comparing (2.3) with (2.11), an interesting conclusion may be drawn: (2.3) fails only in applications in which the variance of $S_{(\cdot)}$ is less than that of S_n. Thus, it is only in cases in which S_n is, in the sense of variance, an inferior estimate to $S_{(\cdot)}$ that (2.3) is in question.

Equations (2.1) and (2.2) break the question of the performance of the jackknife estimate of variance into its properties, (2.1), relative to statistics computed on the same sample size as the jackknife subsets, and the effect on the variance of the change in sample size, (2.2). A similar approach may be taken to the random group method, based upon splitting n = rk observations into r groups of k each. For the case of iid random variables

$$E(Var_{RG}(S_n)) = (1/r)\ Var\ S_k \qquad (2.12)$$

$$(1/r)\ Var\ S_k \geq Var\ S_{rk} \quad ? \qquad (2.13)$$

Here, the inequality in (2.1) may be replaced by equality in (2.12), but the effect of sample size modification is perhaps more in question in (2.13). These relationships hold even if the original sample design selects r sample cases, with replacement, from k strata, and the r random groups are formed by selecting one sample case without replacement from each of the k strata.

Half-sample or balanced-repeated replication is also a frequent choice for complex sample surveys. Assume two sample observations, X_{i1} and X_{i2}, are drawn, with replacement, from each of k = n/2 strata. One variance estimator in this situation is based on computation of the statistic S_k both for an original half-sample of n/2 observations and for the complementary sample of the remaining n/2 observations, and computing their squared difference; this procedure is a special case of the random group method described by (2.12) and (2.13) with r=2. Of course, in application this procedure is repeated for random or balanced half-samples and the resulting variance estimates (each a single degree of freedom) averaged; this averaging does not affect (2.12) or (2.13).

For other half-sample methods, however, the analysis of variance decomposition provides (2.12) as an inequality. For each j = 1,...,J, let h(·,j) be a mapping from {1,...,k} to {1,2}, so that $S_k(X_{1h(1,j)},\ldots,X_{kh(k,j)})$ represents the statistic computed on the j-th half-sample. Let

$$S_{(\cdot)} = \sum_{j=1}^{J} S_k(X_{1h(1,j)},\ldots,X_{kh(k,j)})/J \qquad (2.14)$$

be the average of J half-sample estimates. If the half-samples are constructed by independent random selection, with equal probability from the set of all 2^k possible half-samples, then the variance estimator

$$Var_{HS1}(S_n) = (J-1)^{-1} \sum_{j=1}^{J} (S_k(X_{1h(1,j)},\ldots,X_{kh(k,j)}) - S_{(\cdot)})^2 \qquad (2.15)$$

satisfies

$$E(Var_{HS1}(S_n)) \geq (1/2)\ Var\ (S_k) \qquad (2.16)$$

which parallels (2.1). If S_n is symmetric with respect to permutations of the order of X_{i1}, X_{i2} within each stratum i, i.e.,

$$S_n(X_{11},X_{12},\ldots,X_{i1},X_{i2},\ldots,X_{k1},X_{k2}) = S_n(X_{11},X_{12},\ldots,X_{i2},X_{i1},\ldots,X_{k1},X_{k2}) \qquad (2.17)$$

for all i, then the alternative variance estimator

$$Var_{HS2}(S_n) = J^{-1} \sum_{j=1}^{J} \{S_k(X_{1h(1,j)},\ldots,X_{kh(k,j)}) - S_n(X_{11},\ldots,X_{k2})\}^2 \qquad (2.18)$$

also satisfies (2.16) for any set of predetermined or random half-samples that do not depend upon the values of $X_{11},\ldots,X_{k2}$ observed. Balanced and partially-balanced half-sample replication satisfy these conditions, as well as independent random selection of half samples. Proofs of these results using the ANOVA decomposition ((2.7)-(2.9) etc.) are presented in the appendix.

None of the preceding results requires the observations X to be univariate random variables; indeed, these random variables may represent multivariate weighted results from additional stages of sampling. Selection with replacement is required, however, in order to give the necessary independence.

These results show that half-sample replication, using (2.15) or (2.18), generally tends to produce overestimates of variance in the sense of expectation, although exceptions to (2.13) may lead in turn to exceptions to this rule. If the inequality (2.13) is quite strong, that is, if Var S_k considerably exceeds k Var S_n, then the replication methods can be expected to produce a similarly substantial overestimate.

The analysis of variance decomposition (2.7) - (2.9) serves to introduce important concepts discussed in the literature on replication methods.

One such notion in the context of simple random sampling from an infinite population is to consider the empirical distribution function, the non-parametric estimate of the population distribution function given by the distribution function derived from the probability measure with mass 1/n at each of the n sample points. S is a functional statistic if it is solely a function of the empirical distribution function, independent of n. The sample mean and median are two examples of functional statistics.

Again in the context of simple random samples, a functional statistic is a linear functional statistic if only the linear term (2.8) in its analysis of variance decomposition is nondegenerate, i.e.

$$S_n(x_1,\dots,x_n) = \mu + \sum_i A_i(x_i) \qquad (2.19)$$

Similarly, quadratic functional statistics are those whose decomposition includes only terms through the quadratic expectations (2.9),

$$S_n(x_1,\dots,x_n) = \mu + \sum_i A_i(x_i) + \sum_{i<i'} B_{ii'}(x_i,x_{i'}) \qquad (2.20)$$

(These definitions and their applications are discussed by Efron (1982) and other general references on replication methods.)

Generalization of these concepts to the context of complex samples appears to favor a different approach, however. For example, for simple random samples, the meaning of the sample size n is unambiguous, whereas the number of units in the population is often unknown for multi-stage sample designs. Instead, one may consider for a finite population the multivariate cumulative frequency function. The Horwitz-Thompson estimator formed by placing mass equal to the inverse probability of selection for each sample case provides the finite population equivalent of the empirical distribution function. Functional statistics in this context are those that depend only on this estimated cumulative frequency function. Examples in survey estimation are numerous, including typical estimates of means, proportions, ratios, etc.

Linear functional statistics in the context of complex samples are those that are linear mappings from the space of cumulative frequency functions to R^n (to cover multivariate versions). Similarly, quadratic functional statistics in this context are linear functional statistics augmented by terms $B(x_i,x_i')$ arising from a bilinear operator B(X,X) on the space of cumulative frequency functions, using X redundantly as both arguments (e.g. Liu and Thompson 1983). Consequently, linear and quadratic functionals have essentially the same form as (2.19) and (2.20), respectively, but with terms not necessarily derived from their ANOVA decomposition based upon the complex sample design.

3. GENERAL RESAMPLING PLANS FOR COMPLEX DESIGNS

Most variance estimators for linear functional statistics from complex samples take the form

$$\mathrm{Var}(S) = \sum_j b_j \left(\sum_k a_{jk} X_k\right)^2 \qquad (3.1)$$

where X_k represents the linear functional evaluated on the frequency distribution placing mass equal to the inverse probability of selection for sample observation k from sample s, and the factors b_j and a_{jk} may depend on s but not $\underset{\sim}{X} = \{X_k\}^T$. (The evaluation of the linear functional for the entire sample is thus $\underset{\sim}{1}^T\underset{\sim}{X}$, where $\underset{\sim}{1}$ represents a column vector of 1's.) Such variance estimators may be written

$$\mathrm{Var}(S) = \underset{\sim}{X}^T \underset{\sim}{C}_s \underset{\sim}{X} \qquad (3.2)$$

for a symmetric matrix $\underset{\sim}{C}_s$ determined by the a_{jk} and b_j.

For a given sample s and symmetric matrix $\underset{\sim}{C}_s$, a resampling plan corresponding to C_s is a family of random variables d_r* and of non-negative random variables $\underset{\sim}{p}^{(r)*} = \{p_k^{(r)*}\}$ such that, for any $\underset{\sim}{x}$,

$$E^*\{d_r^*(\underset{\sim}{1}^T\underset{\sim}{x}(\underset{\sim}{p}^{(r)*}) - \underset{\sim}{1}^T\underset{\sim}{x})^2\} = \underset{\sim}{x}^T\underset{\sim}{C}_s\underset{\sim}{x} \qquad (3.3)$$

where $\underset{\sim}{x}(\underset{\sim}{p}^{(r)*}) = \{x_k p_k^{(r)*}\}$. The expectation E* is over the probability distribution of the random variables p_k*. Efron (1982) discusses resampling plans for simple random samples, although not necessarily conforming to (3.3).

Theorem 1 For any symmetric matrix C_s, there exist corresponding resampling plans d_r^*, $\underset{\sim}{p}^{(r)*}$. Furthermore, a plan may be chosen so that $E^*(p_k^{(r)*}) = 1$ and $E^*(d_r^*(p_k^{(r)}-1)) = 0$ for all k. If $\underset{\sim}{C}_s$ is positive semi-definite, there exists a plan such that d_r* is constant.

Proof Let $\lambda_1,\dots,\lambda_M$ be an enumeration of the nonzero eigenvalues of $\underset{\sim}{C}_s$, including multiplicities, and let $\underset{\sim}{v}^{(1)},\dots,\underset{\sim}{v}^{(M)}$ be a corresponding set of orthonormal eigenvectors. For any x,

$$\underset{\sim}{x}^T\underset{\sim}{C}_s\underset{\sim}{x} = \sum_{m=1}^{M} \lambda_m (\underset{\sim}{v}^{(m)T}\underset{\sim}{x})^2 \qquad (3.4)$$

Let

$$\lambda_+ = \sum_{m=1}^{M} |\lambda_m| \qquad (3.5)$$

Define the random variables $\underset{\sim}{p}^{(r)*}$ and d_r* by

$$\underset{\sim}{p}^{(r)*} = \underset{\sim}{1} + \underset{\sim}{v}^{(m)} \qquad (3.6)$$

$$d_r^* = \lambda_+^{-1} \lambda_m/|\lambda_m| \qquad (3.7)$$

with probability $|\lambda_m|/(2\lambda_+)$ and

$$\underset{\sim}{p}^{(r)*} = \underset{\sim}{1} - \underset{\sim}{v}^{(m)} \qquad (3.8)$$

$$d_r^* = \lambda_+^{-1} \lambda_m/|\lambda_m| \qquad (3.9)$$

with probability $|\lambda_m|/(2\lambda_+)$. Then (3.3) and the other conditions of the theorem are satisfied.

Although other researchers have derived resampling plans for most practical situations, the theorem emphasizes the existence of resampling plans under all situations in which the variance estimator takes the form (3.2), establishing the generality of replication. Other approaches to the construction of resampling plans may be taken, besides the one given in the proof.

The index r in the plan implies the computation of the resampled statistics for multiple replicates, which may be generated through independent selections, but not necessarily so. The jackknife and balanced repeated replication may be viewed as resampling plans with specific dependencies across r, where the probability distribution P^* may be induced simply by permutation of r. A generalization of this notion of balancing is given by the following theorem.

Theorem 2 Under the conditions of Theorem 1, there exists a completely balanced resampling plan of order 2M, where M is the rank of C_s, in the sense that

$$x^T C_s x = (2M)^{-1} \sum_{r=1}^{2M} d_r^*(1^T x(p^{(r)*}) - 1^T x)^2 \qquad (3.10)$$

If C_s is positive semi-definite, there exists a plan such that d_r^* is constant.

Proof Using the same definitions as the proof of Theorem 1, let

$$\lambda_{max} = \max\{|\lambda_m|\} \qquad (3.11)$$

$$p^{(2m-1)*} = 1 + \lambda_m^{1/2}/\lambda_{max}^{1/2}\, \nu^{(m)} \qquad (3.12)$$

$$p^{(2m)*} = 1 - \lambda_m^{1/2}/\lambda_{max}^{1/2}\, \nu^{(m)} \qquad (3.13)$$

$$d_{2m-1}^* = d_{2m}^* = \lambda_{max}^{-1}\, \lambda_m/|\lambda_m| \qquad (3.14)$$

The index r may then be defined as a random permutation of the index 2m-1 or 2m in (3.12)-(3.14). This constructed resampling plan satisfies the statement of the theorem.

It should be noted that $1^T x(p^{(r)*})$ in (3.3) represents the evaluation of the linear functional for the resampled distribution based on reweighting each sample case k by the factor $p_k^{(r)*}$. For general functional statistics, S_g, an implied variance estimator is

$$\mathrm{Var}(S_g) = 1/R \sum_{r=1}^{R} d_r^*(S_g(p^{(r)*}) - S_g)^2 \qquad (3.15)$$

where $S_g(p^{(r)})$ again represents the evaluation of S_g on the resampled cumulative frequency distribution function formed by altering the weight for each case k by the factor $p_k^{(r)*}$. Naturally, properties of this approach for general functionals would depend upon the specific finite population, the specific resampling plan, and the sample design.

4. BIAS REDUCTION WITH GENERAL RESAMPLING PLANS

Bias reduction provided an initial impetus for development of replication methods, and this property has been explicitly recognized for the jackknife, half-sample, and bootstrap. The following theorem simply generalizes this aspect of replication for the resampling plans defined in the preceding section, using methods earlier incorporated in the formulation of the jackknifed chi-square test (Fay 1980, 1984).

Theorem 3 Suppose there is a C_s such that (3.2) is an unbiased estimate over the distribution of s of the design-based variance for all linear functionals, and that for each s, d_r^*, $p^{(r)*}$ is a corresponding resampling plan satisfying (3.3) with $E^*\{d_k^{(r)*}(p_k^{(r)*} - 1)\} = 0$ and $E^*(p_k^{(r)*}) = 1$ for all k. Then, if S_q is any quadratic functional,

$$S_q' = S_q - d_r(S_q(p^{(r)}) - S_q) \qquad (4.1)$$

is an unbiased estimate of S_q evaluated for the cumulative frequency function of the finite population.

The proof is given in the appendix.

Naturally, in application (4.1) would typically be averaged over a series of replicates r=1, ...,R. The problem of estimation of quadratic functionals for finite populations has recently been treated by Liu and Thompson (1983), who consider the Horwitz-Thompson estimator based on the joint inclusion probabilities. An advantage to (4.1), however, is its immediate extension to functionals that are locally approximated by quadratic functionals, where (4.1) may yield bias reduction in place of bias removal.

5. ASYMPTOTIC RESULTS FOR IMPLICITLY DEFINED ESTIMATORS

Binder (1983) considered the question of estimation of the asymptotic variance for asymptotically normal estimators of population parameters, θ_0, defined as the solution to an equation of the form

$$W(\theta) = \sum_{k=1}^{N} u(Z_k,\theta) - v(\theta) = 0 \qquad (5.1)$$

where Z_k represents the data for unit k in the population of size N. For this class, the true θ_0 representing the solution to (5.1) for the population may be estimated from a sample of size n through estimation of $W(\theta)$ by

$$\hat{W}(\theta) = \sum_{k} w_k\, u(Z_k,\theta) - v(\theta) \qquad (5.2)$$

where w_k represents the design-based weight for sample unit k, and defining $\hat{\theta}$ as the solution of

$$\hat{W}(\hat{\theta}) = 0 \qquad (5.3)$$

As examples of this class of estimators, he discussed generalized linear models, which include linear and logistic regression, and log-linear models in general. This formulation obviously covers most maximum likelihood estimators, where the derivative of the log-likelihood for the population in (5.1) is estimated through (5.2). More generally, many M-estimators are also of this form. Thus, this class encompasses most analytic statistics of interest.

Binder's results include conditions under which the estimator $\hat{\theta}$ from (5.3) is asymptotically normal and the asymptotic variance may be estimated by a Taylor series (linearization) method appropriate to the form of (5.2). He considers a sequence of populations of size N_t, t=1,2,3,..., where the population value from (5.1), $W_t(\theta)$, is defined for all $\theta \varepsilon \Theta$, the parameter space. Except for slight notational changes and generalization, his conditions are:

Condition 1.

$$\lim_{t\to\infty} N_t^{-1} W_t(\theta) = \Omega(\theta) \tag{5.4}$$

exists for all $\theta \varepsilon \Theta$.

Condition 2. $\Omega(\theta)$ is a one-to-one function, so that $\Omega^{-1}(.)$ exists.

Conditions 3 and 4. There exists a θ_0, an interior point of Ω, such that

$$\Omega(\theta_0) = 0. \tag{5.5}$$

Condition 5. (A generalization of Binder's formulation) There exists a sequence h_t, t=1,2,3,... with $h_{t\to\infty}$ and

$$h_t N_t^{-1} (\hat{W}(\theta) - W_t(\theta)) \xrightarrow{D} N(0, \Phi(\theta)) \tag{5.6}$$

for positive-definite covariance matrix $\Phi(\theta)$ for all θ in a neighborhood of θ_0.

Condition 6. $\hat{W}(\theta)$ is totally differentiable in a neighborhood of θ_0.

Condition 7.

$$\lim_{\theta\to\theta_0, t\to\infty} N_t^{-1} \left(\frac{\partial \hat{W}(\theta)}{\partial \theta}\right) = \left(\frac{\partial \Omega(\theta)}{\partial \theta}\right)\Bigg|_{\theta=\theta_0} = J(\theta_0) \tag{5.7}$$

in probability.

Condition 8. $J(\theta_0)\Phi(\theta_0)J(\theta_0)^T$ is of full rank.

Condition 9. $\Phi(\theta)$ is a continuous function of θ.

Condition 10. $\partial\Omega(\theta)/\partial\theta$ is a continuous function of θ.

Condition 11. $\hat{\Phi}(\theta)$, assumed to exist under the sample design, gives a consistent estimator for $\Phi(\theta)$.

Lemma 1 (Binder 1983) The distribution of $h_t(\hat{\theta}-\theta_0)$ is asymptotically equivalent to the distribution of

$$D_t = -J^{-1}(\theta_0)\, h_t\, N_t^{-1}\, W(\theta_0) \tag{5.8}$$

Corrollary 1 (Binder 1983) The asymptotic distribution of $h_t(\hat{\theta}-\theta_0)$ is the normal law with mean 0 and variance matrix $J^{-1}(\theta_0)\Phi(\theta_0)J^{-1}(\theta_0)^T$.

Corrollary 2 (Binder 1983) Let F_t be the distribution function for $h_t(\hat{\theta}-\theta_0)$, based on the t-th sample and let G_t be the distribution function of a multivariate normal distribution with mean zero and variance matrix

$$h_t^2\, \hat{Var}(\hat{\theta}) = \hat{J}^{-1}(\hat{\theta})\hat{\Phi}(\hat{\theta})\hat{J}^{-1}(\hat{\theta})^T \tag{5.9}$$

where $\hat{J}(\hat{\theta}) = N_t^{-1}\, \partial\hat{W}(\hat{\theta})/\partial\theta$. Then, by virtue of conditions 9 to 11

$$\lim_{t\to\infty} \sup |F_t - G_t| = 0 \tag{5.10}$$

These results establish the asymptotic inferential validity of using (5.9) for confidence statements about θ_0.

The stringent nature of the global nature of conditions 1 and 2, as opposed to a simpler requirement that they apply in a neighborhood of θ_0, serves to guarantee consistency of $\hat{\theta}$. The global nature of these conditions could be removed by adding conditions assuring the consistency of $\hat{\theta}$.

The following theorem states conditions under which a generalized replication approach gives the same outcome as Corrollary 2.

Theorem 4 Under conditions 1-11, suppose that in a neighborhood $N(\theta_0)$ of θ_0, for each t=1,2,... there exists a resampling plan, $d_r{*}^{(t)}$, $p^{(r,t)}{*}$, r=1,2,...R_t, giving reweighted $\hat{W}(\theta)(p^{(r,t)}{*})$ for all θ in $N(\theta_0)$, and that, as $t\to\infty$,

$$\sup_{\substack{r<R_t \\ \theta\varepsilon N(\theta_0)}} N_t^{-1} \|\hat{W}(\theta)(p^{(r,t)*}) - \hat{W}(\theta)\| \xrightarrow{P} 0 \tag{5.11}$$

Suppose also that $Var_r\ \hat{W}(\theta)$ is the estimated variance for $\hat{W}(\theta)$ of the form (3.15) and that, in $N(\theta_0)$, $h_t^2 N_t^{-2} Var_r\ \hat{W}(\theta)$ consistently estimates $\Phi(\theta)$. If $\hat{\theta}(p^{(r,t)})$ are replicate estimates based on $\hat{W}(\theta)(p^{(r,t)})$, and if $Var_r\ \hat{\theta}$ is of the form (3.15), then $h_t^2\ Var_r\ \hat{\theta}$ may replace (5.9) in the statement of Corrollary 2 (Binder 1983).

Comment on proof Condition 7 implies, for any $\delta > 0$, there exists an $\varepsilon > 0$, and T, such that

$$P\{\|N_t^{-1}\frac{\partial W(\theta)}{\partial \theta} - J(\theta_0)\| > \delta\} < \delta \tag{5.12}$$

for all θ with $\|\theta - \theta_0\| < \varepsilon$, t>T. With the exception of the set described by (5.12) and an additional set of arbitrarily small probability, the expression, (3.15), for $Var_r\ \theta$ may be expanded with the methods used in the proof of Lemma 1; (5.11) bounds the contribution of all terms except for another set of arbitrarily small probability.

The conditions of the theorem are often easier to check in practice than they might at first seem. Note that the scaling factor h_t does not appear in (5.11).

For fixed θ, (5.1) is a linear functional modified by $v(\theta)$, which is constant for fixed θ. Thus, although $\hat{\theta}$ itself is typically non-linear, resampling plans for $W(\theta)$ in $N(\theta_0)$ generally would exist.

APPENDIX

A.1 Example of Ratio Estimation from Section 2

Section 2 introduces an example with $E(Y|X=x) = rx$, $Var(Y|X=x) = \sigma^2x^2$ as a counterexample to (2.3) when n=2 and an illustration when (2.2) may fail. For r_n defined by (2.10),

$$Var(r_n) = n\, E(X_1^2/(\sum_{i=1}^{n} X_i)^2) \tag{A.1}$$

$$E(Var_J(r_n)) = ((n-1)/n)\, Var(r_{n-1}) +$$

$$((n-1)^2/2n)\ E\ \frac{(X_n - X_{n-1})^2 \sum_{i=1}^{n-2} X_i^2}{(\sum_{i=1}^{n-1} X_i)^2\ (\sum_{i=1}^{n-2} X_i + X_n)^2} \quad (A.2)$$

where no second term in (A.2) appears for n=2. The form of (A.2) is consistent with (2.1), but the author has not derived general simplifying conditions under which (2.3) fails except through direct computation of (A.1) and (A.2).

A.2 Proof of Inequality (2.16)

The variance estimator (2.15) based upon independent half-sample replications may be written

$$Var_{HS1} = (J(J-1))^{-1} \sum_{j=1}^{J} \sum_{j<j'} \{S_k(X_{1h(1,j)},\ldots,X_{kh(k,j)}) - S_k(X_{1h(1,j')},\ldots,X_{kh(k,j')})\}^2 \quad (A.3)$$

There are $J(J-1)/2$ terms in the summation (A.3), each with the same expectation. For each i, $h(i,j) = h(i,j')$ with probability 1/2 for $j = j'$; therefore, the contribution to the total (A.3) of terms $(A_i(X_{ih(i,j)}) - A_i(X_{ih(i,j')}))^2$ is 1/2 $Var\ (A_i(X_i))$. Similarly, for $i \neq i'$, the probability of $\{h(i,j) = h(i',j')$ and $h(i',j) = h(i',j')\}$ is 1/4, so that the contribution to (A.3) of terms $\{B_{ii'}(X_{ih(i,j)},X_{i'h(i',j)}) - B_{ii'}(X_{ih(i,j')},X_{i'(i',j')})\}^2$ is 3/4 $Var\ B_{i'}(X_i,X_{i'})$. In general, for the m-th order terms in the ANOVA decomposition, the contribution to (A.3) will be $1 - (1/2)^m$ times the variance contribution to the decomposition of S_k. Thus, (2.16) is established, with strict inequality whenever any terms in the decomposition of S_k above the first order are non-degenerate.

To show (2.16) for the variance estimator (2.18), which includes application to balanced repeated replication, note that the expected value of the right-hand side of (2.18) is minimized, over all symmetric statistics in the sense of (2.17), by replacing $S_n(X_{11},\ldots,X_{k2})$ by

$$S_{(\cdot)}(X_{11},\ldots,X_{k2}) = 2^{-k} \sum_h S_k(X_{1h(1,j)},\ldots,X_{kh(k,j)}) \quad (A.4)$$

where summation in (A.4) is over all 2^k possible half-samples. Thus,

$$E(Var_{HS2}(S_n) \geq E(\ J^{-1} \sum_{j=1}^{J} \{S_k(X_{1h(1,j)},\ldots X_{kh(k,j)} - S_{(\cdot)}(X_{11},\ldots,X_{k2})\}^2) \quad (A.5)$$

The ANOVA decomposition may be used to reexpress (A.4); for example, the first-order terms of the decomposition of (A.4) are 1/2 $A_i(X_{ij})$, the second-order terms 1/4 $B_{ii'}(X_{ij},X_{i'j'})$ for $(i,j) \neq (i',j')$, etc. With this decomposition, the right-hand side of (A.5) may be easily shown to have the same expectation as (A.3), establishing result (2.16) and providing a concidental demonstration of

$$E(Var_{HS2}(S_n)) \geq E(Var_{HS1}(S_n)) \quad (A.6)$$

for any predetermined half-sample plan on the left-hand side of (A.6) and implementation of random half-samples on the right-hand side.

A.3 Proof of Theorem 3

It is sufficient to prove the result for a quadratic functional of the form $B_{ii'} = b_{ii'}c_ic_{i'}$, where $b_{ii'}$ is a constant and c_i, $c_{i'}$ represent counts for cells or characteristics i and i', not necessarily distinct, in the finite population.

Let $X^{(i)}$ and $X^{(i')}$ denote weighted indicator variables for characteristics i and i' based on sample s. Over the sampling distribution of s,

$$E(b_{ii'}(1^TX^{(i)})(1^TX^{(i')})) = b_{ii'}\ (c_ic_{i'} + Cov(1^TX^{(i)},1^TX^{(i')})) \quad (A.7)$$

For a specific sample s,

$$E^*\{d_r^*\{b_{ii'}(1^TX^{(i)}(p^{(r)}))(1^TX^{(i')}(p^{(r)*}))) - b_{ii'}(1^TX^{(i)})(1^TX^{(i')})\}\} = X^{(i)T}C_sX^{(i')} \quad (A.8)$$

using the specific assumptions in the statement of the theorem. The unbiasedness of the variance estimator for linear functionals implies unbiasedness in the corresponding covariance estimator on the right-hand side of (A.8), and the theorem follows by taking the expectation of (A.8) over s.

References

Binder, D.A. (1983), "On the Variances of Asymptotically Normal Estimators from Complex Samples," International Statistical Review, 51, 279-292.

Efron, B. (1982), The Jackknife, the Bootstrap and Other Resampling Plans, Society for Industrial and Applied Mathematics, Philadelphia, PA.

Efron, B. and Stein, C. (1981), "The Jackknife Estimate of Variance," Annals of Statistics, 9, 586-596.

Fay, R.E. (1980), "On Jackknifing Chi-Square Test Statistics - Part II: Asymptotic Theory," unpublished manuscript.

______ (1984), "A Jackknifed Test for Complex Samples," to appear in the Journal of the American Statistical Association.

Hoeffding, W. (1948), "A Class of Statistics with Asymptotically Normal Distributions," Annals of Mathematical Statistics, 19, 293-325.

Liu, T.P. and Thompson, M.E. (1983), "Properties of Estimates of Quadratic Finite Population Functions: The Batch Approach," Annals of Statistics, 11, 275-285.

Karlin, S. and Rinott, Y. (1982), "Applications of ANOVA Type Decompositions for Comparisons of Conditional Variance Statistics Including Jackknife Estimates," Annals of Statistics, 10, 485-501.