The bootstrap method of inference is extended to stratified sampling, involving a large number of strata with relatively few (primary) units sampled within strata, when the parameter of interest, \( \theta \), is a nonlinear function of the population mean vector. The bootstrap estimate of bias of the estimate, \( \hat{\theta} \), of \( \theta \) and the estimate of variance of \( \hat{\theta} \) are obtained. Bootstrap confidence intervals for \( \hat{\theta} \) are also given utilizing the percentile method or the bootstrap histogram of the t-statistic. Extensions to unequal probability sampling without replacement and two-stage sampling are also obtained.

1. Introduction

Resampling methods, including the jackknife and the bootstrap, provide standard error estimates and confidence intervals for the parameters of interest. These methods are simple and straightforward but are computer-intensive, especially the bootstrap. Efron (1982) has given an excellent account of resampling methods in the case of an independent and identically distributed (i.i.d.) sample of fixed size \( n \) from an unknown distribution \( F \), and the parameter of interest \( \theta = \theta(F) \). Limited empirical evidence (see Efron, 1982, p. 18) has indicated that the bootstrap standard error estimates are likely to be more stable than those based on the jackknife and also less biased than those based on the customary delta (linearization) method. Moreover, the bootstrap histogram of t-statistic approximates the true distribution of \( t \) with a remainder term of \( O_p(n^{-1}) \) in the Edgeworth expansion (Abramovitch and Singh, 1984), unlike the usual normal approximation with a remainder term of \( O_p(n^{-2}) \). Empirical results for the ratio estimate (Hinkley and Wei, 1984) also indicate that the confidence intervals from the bootstrap histogram of \( t \) (using the linearization variance estimator) perform better than those based on the normal approximation.

The main purpose of this article is to propose an extension of the bootstrap method to stratified samples in the context of sample survey data; especially to data obtained from stratified cluster samples involving large numbers of strata, \( L \), with relatively few primary sampling units (psu’s) sampled within each stratum. For nonlinear statistics \( \theta \) that can be expressed as functions of estimated means of \( p \) (\( \geq 2 \)) variables, Krewski and Rao (1981) established the asymptotic consistency of the variance estimators from the jackknife, the delta and the balanced repeated replication (BRR) methods as \( L \to \infty \) within the context of a sequence of finite populations \( \{N_j \} \) with \( L \) strata in \( \mathbb{N}_L \). Their result is valid for any multistage design in which the psu’s are selected with replacement and in which independent subsamples are selected within those psu’s sampled more than once. Rao and Wu (1983) obtained second order asymptotic expansions of these variance estimators under the above set up and made comparisons in terms of their biases.

The proposed bootstrap method for stratified samples is described in Section 2 and the properties of the resulting variance estimator are studied. The bootstrap estimate of bias of \( \theta \) is also obtained. Section 3 provides bootstrap confidence intervals for \( \theta \). The results are extended to stratified simple random sampling without replacement in Section 4. Finally, the method is extended to unequal probability sampling without replacement in Section 5 and to two-stage cluster sampling without replacement in Section 6.

2. The Bootstrap Method

The parameter of interest \( \theta \) is a nonlinear function of the population mean vector \( \bar{Y} = (Y_1, \ldots, Y_p)^T \), say \( \theta = g(\bar{Y}) \). This form of \( \theta \) includes ratios, regression and correlation coefficients. If \( n_j (\geq 2) \) psu’s are selected with replacement with probabilities \( \phi_i \) in stratum \( h \), then Krewski and Rao (1981) have shown that the natural estimator \( \hat{\theta} = g(\bar{Y}) \) can be expressed as \( \hat{\theta} = g(\bar{Y}) \). Here \( \bar{Y} \) is a design-unbiased linear estimator of \( Y = \sum_h \bar{Y}_h \), and \( \bar{Y}_h = (Y_{h1}, \ldots, Y_{hp})^T \) are the \( h \)-th stratum weight (\( \bar{Y}_h \)) and population mean vector respectively and \( \bar{Y} \) is the mean of \( n_h \) i.i.d. random vectors \( Y_{hi} = (Y_{h1}, \ldots, Y_{hp})^T \) for each \( h \) with \( E(Y_{hi}) = \bar{Y}_h \). For \( h \neq h' \), \( Y_{hi} \) and \( Y_{h'i} \) are independent but not necessarily identically distributed.

2.1. The Naive bootstrap. In the case of an i.i.d. sample \( \{Y_{i1} \} \) with \( E(Y_{i1}) = \bar{Y} \), the bootstrap method is as follows: (i) Draw a simple random sample \( \{Y_{ib} \}^n \) with replacement from the observed values \( Y_1, Y_2, \ldots, Y_n \) and calculate \( \hat{\theta}^* = g(\bar{Y}^*) \) where \( \bar{Y}^* = EY_{ib}/n \), (ii) Independently replicate step (i) a large number, \( B \), of times and calculate the corresponding estimates \( \hat{\theta}^*_1, \ldots, \hat{\theta}^*_B \). (iii) The bootstrap variance estimator of \( \hat{\theta} = g(\bar{Y}) \) is given by

\[ v_b(a) = \frac{1}{B-1} \sum_{b=1}^{B} (\hat{\theta}^*_b - \hat{\theta}^*)^2, \]

where \( \hat{\theta}^* = E(\hat{\theta}^*) = E(\bar{Y}^*) \). The Monte-Carlo estimator \( v_b(a) \) is an approximation to

\[ v_b = \text{var}(\hat{\theta}^*) = E_b(\hat{\theta}^* - \bar{\theta}^*)^2, \]

where \( E_b \) denotes the expectation with respect to bootstrap sampling from a given sample \( Y_1, \ldots, Y_n \). No closed-form expression for \( \text{var}(\hat{\theta}^*) \) generally exists in the nonlinear case, but in the linear case with \( p = 1 \), \( \hat{\theta}^* = \bar{Y}^* \) and \( v_b \) reduces to

\[ \text{var}(\bar{Y}^*) = s^2/n \]

where \( (n-1)s^2/n = \text{var}(\bar{Y}) \) and \( \text{var}(\bar{Y}) = s^2/n \).
is the unbiased estimator of variance of $\hat{y}$. The modified variance estimator $[n/(n-1)]\text{var}_a(\hat{g}^*)$ exactly equals $\text{var}(\hat{y})$ in the linear case, but Efron (1982) found no advantage in this modification. In any case, $n/(n-1)=1$ in most applications and $\text{var}_a(\hat{g}^*)$ is a consistent estimator of the variance of $\hat{g}$, as $n \to \infty$ (Bickel and Freedman, 1981). The bootstrap histogram of $\hat{g}^1,...,\hat{g}^B$ may be used to find confidence intervals for $\hat{g}$. This method (Efron, 1982) is called the percentile method.

Noting the i.i.d. property of the $y_{hi}$'s within each stratum, a straightforward extension of the usual bootstrap method to stratified samples (referred to as the "naive bootstrap") is as follows: (i) Take a simple random sample $\{y_{hi}\}_{i=1}^{n_h}$ with replacement from the given sample $\{y_{hi}\}_{i=1}^{n_h}$ in stratum $h$, independently for each stratum. Calculate $\hat{y}_h = \frac{1}{n_h} \sum y_{hi}$, $\bar{y} = \frac{1}{\sum n_h} \sum \hat{y}_h$, and $\hat{g}^* = \hat{g}(\bar{y})$. (ii) Independently replicate step (i) a large number, $B$, of times and calculate the corresponding estimates $\hat{g}_1^*,...,\hat{g}_B^*$. (iii) The bootstrap variance estimator of $\hat{g} = \hat{g}(\bar{y})$ is given by

$$v_B(a) = \frac{1}{B-1} \sum_{b=1}^{B} (\hat{g}_b^* - \hat{g}^*)^2,$$

where $\bar{y} = \sum n_h \bar{y}_h$ and $\hat{g}_b = \hat{g}(\bar{y}_b)$. The Monte-Carlo estimator $\hat{v}_B(a)$ is an approximation to $v_B(a)$:

$$\hat{v}_B(a) = E_\bar{y}(\hat{g}_b^* - E_\bar{y}(\hat{g}_b^*))^2,$$

where $E_\bar{y}$ denotes the expectation with respect to bootstrap sampling. In the linear case with $p = 1$, $\hat{g}^* = \sum n_h \bar{y}_h / \hat{y}^*$ and $v_B$ reduces to

$$\text{var}_a(\hat{g}^*) = \frac{1}{\sum n_h} \sum (n_h - 1) \hat{y}_h^2,$$

where $(n_h - 1) \hat{y}_h^2 = \sum y_{hi}^2 - \bar{y}^2$. Comparing (2.6) with the unbiased estimator of variance of $\bar{y}$,

$$\text{var}(\bar{y}) = \sum n_h \hat{y}_h^2 / \hat{y}^2,$$

it immediately follows that $\text{var}_a(\hat{g}^*) / \text{var}(\bar{y})$ does not converge to 1 in probability, unless $L$ is fixed and $n_h \to \infty$ for each $h$. Hence, $\text{var}_a(\hat{g}^*)$ is not a consistent estimator of the variance of $\bar{y}$. It also follows that $v_B$ is not a consistent estimator of the variance of $\bar{y}$ or mean square error of a general nonlinear statistic. There does not seem to be an obvious way to correct this scaling problem except when $n_h = k$ for all $h$ in which case $k(k-1)-1 \text{var}_a(\hat{g}^*)$ will be consistent.

Bickel and Freedman (1984) also noticed the scaling problem, but they were mainly interested in bootstrap confidence intervals in the linear case ($p = 1$). They have established the asymptotic $N(0,1)$ property of the distribution of $t = (\bar{y} - \bar{y}) / (\text{var}(\bar{y}))^{1/2}$ and of the conditional distribution of $(\bar{y} - \bar{y}) / (\text{var}(\bar{y}))^{1/2}$ in stratified simple random sampling with replacement, and also proved that $(\sum n_h \hat{y}_h^2 / \hat{y}^2) / \text{var}(\bar{y})$ converges to 1 in probability as $n = \sum n_h \to \infty$,

where $(n_h - 1) \hat{y}_h^2 = \sum (y_{hi}^2 - \bar{y}_h^2)$. Their result implies that one could use the bootstrap histogram of $\hat{t}_1,...,\hat{t}_B$ to find confidence intervals for $\bar{y}$, where $\hat{t}_b = (y_{b}^2 - \bar{y}_b^2) / (\sum n_h \hat{y}_h^2 / n_h)^{1/2}$, where $\hat{y}_b$ is the value of $\hat{y}_b$ for the $b$-th bootstrap sample ($b = 1,...,B$). In the nonlinear case, there does not seem to be a simple way to construct $t^*_b$-values similar to those of Bickel and Freedman since $v_B$ has no closed form. Moreover, the straightforward extension of the bootstrap (hereafter called the naive bootstrap) does not permit the use of the percentile method based on the bootstrap histogram of $\hat{g}_1^*,...,\hat{g}_B^*$.

Although $t^*_b$ is asymptotically $N(0,1)$ in the linear case, it is not likely to provide as good an approximation to the distribution of $t$ as a statistic whose denominator and numerator are both adequate approximations to their counterparts in $t$. Such statistics will be proposed in Section 3.2. These are also applicable to the nonlinear case.

Recognizing the scaling problem in a different context, Efron (1982) suggested to draw a bootstrap sample of size $n_h - 1$ instead of $n_h$ from stratum $h$ ($h = 1,...,L$). In Section 2.2, we will instead propose a different method which includes his suggestion as a special case.

2.2 The proposed method. Our method is as follows: (i) Draw a simple random sample $\{y_{hi}\}_{i=1}^{m_h}$ of size $m_h$ with replacement from $\{y_{hi}\}_{i=1}^{n_h}$ in stratum $h$, independently for each stratum. Calculate $\bar{y}_h = \frac{1}{m_h} \sum y_{hi}$, $\bar{y} = \frac{1}{\sum m_h} \sum \bar{y}_h$, and $\hat{g}^* = \hat{g}(\bar{y})$. (ii) Independently replicate step (i) a large number, $B$, of times and calculate the corresponding estimates $\hat{g}_1^*,...,\hat{g}_B^*$. (iii) The bootstrap estimator $\hat{E}_\bar{y}(\hat{g})$ of $\hat{g}$ can be approximated by $\hat{g}_B = \frac{1}{B} \sum \hat{g}_b$. The bootstrap variance estimator of $\hat{g}$ is given by

$$\hat{v}_B = \frac{1}{B-1} \sum_{b=1}^{B} (\hat{g}_b^* - \hat{g}^*)^2,$$

where $\hat{g}_b = \hat{g}(\bar{y}_b)$ and $\hat{g}_B = \hat{g}(\bar{y}_B)$. The bootstrap variance estimator of $\hat{g}$ is given by

$$\hat{v}_B = \frac{1}{B-1} \sum_{b=1}^{B} (\hat{g}_b^* - \hat{g}^*)^2.$$

(2.9)

One can replace $E_\bar{y}(\hat{g})$ in (2.8) by $\hat{g}$.

2.3. Justification of the method. In the linear case, $\hat{g} = \bar{y}$, $v_B$ reduces to the customary unbiased variance estimator $\text{var}(\bar{y}) = \text{var}$. 

for any choice $m_h$. In the nonlinear case, it can be shown that

$$V_b = V_L + O_P(n^{-2}) \quad (2.10)$$

for any $m_h$, where $V_L$ is the linearization variance estimator (Rao and Wu, 1983). In the linear case ($p=1$), $V_b$ reduces to $\text{var}(\hat{y})$. Under reasonable regularity conditions, $V_b$ is a consistent estimator of variance of $\hat{\theta}$, $\text{Var}$(\hat{\theta}). Hence, it follows from (2.10) that $V_b$ is also consistent for $\text{Var}(\hat{\theta})$.

The asymptotic $N(0,1)$ property of the conditional distribution of $(\hat{\theta} - \hat{\theta})/\sqrt{V_b}$ can be established. Assuming that $0 < \delta_1 < m_h/(n_h-1) < \delta_2 < \infty$ for all $n_h$, i.e. the bootstrap sample size $m_h$ should be comparable to the original size $n_h$ in each stratum.

### 2.4. Estimate of bias of $\hat{\theta}$

Our bootstrap estimate of bias is

$$\hat{b}(\hat{\theta}) = E_*(\hat{\theta}) - \hat{\theta} \quad (2.11)$$

which is approximated by $\hat{b}(\hat{\theta}) = \hat{\theta}$. It can be shown that $\hat{b}(\hat{\theta})$ is a consistent estimator of $B(\hat{\theta})$ (Rao and Wu, 1983). On the other hand, the bias estimate $B(\hat{\theta}) = E_*(\hat{\theta}) - \hat{\theta}$, based on the naive bootstrap, is not a consistent estimator of $B(\hat{\theta})$ (Rao and Wu, 1983).

### 3. Confidence Intervals

We now consider different bootstrap methods for setting confidence intervals for $\theta$.

#### 3.1. Percentile method

For ready reference, we now give a brief account of the percentile method based on the bootstrap histogram of $\hat{\theta}^1, \ldots, \hat{\theta}^B$. Define the cumulative bootstrap distribution function as

$$\hat{\text{CDF}}(t) = \#(\hat{\theta}^b \leq t; b = 1, \ldots, B)/B. \quad (3.1)$$

For $0 \leq Q \leq 0.5$, define $\hat{\theta}_\text{LOW}(Q) = \hat{\text{CDF}}^{-1}(Q)$ and $\hat{\theta}_\text{UP}(Q) = \hat{\text{CDF}}^{-1}(1-Q)$. Then the interval

$$\{\hat{\theta}_\text{LOW}(Q), \hat{\theta}_\text{UP}(Q)\} \quad (3.2)$$

is an approximate $(1-2Q)$-level confidence interval for $\theta$. It has the central $1-2Q$ portion of the bootstrap distribution (Efron, 1982, p. 70). One can also consider a bias-corrected percentile method, following Efron (1982, p. 82). This method leads to

$$\{\hat{\text{CDF}}^{-1}(\Phi(2z_0 - z_0)), \hat{\text{CDF}}^{-1}(\Phi(2z_0 + z_0))\} \quad (3.3)$$

as an approximate $(1-2Q)$-level confidence interval for $\theta$, where $\Phi$ is the cumulative distribution function of a standard normal, $z_0 = \Phi^{-1}(\hat{\text{CDF}}(\hat{\theta}))$ and $z_0 = \Phi^{-1}(1-Q)$. The advantage of the interval (3.3) over (3.2) has been demonstrated by Efron (1982) in the i.i.d. case.

#### 3.2. Bootstrap t-statistics

Instead of approximating the distribution of $\hat{\theta}$ by the bootstrap distribution of $\hat{\theta}$, we can approximate the distribution of the t-statistic $t = (\hat{\theta} - \theta)/\sqrt{V_b}$ by its bootstrap counterpart $t^* = (\hat{\theta} - \theta)/\sqrt{V_b^*}(a)$ where $V_b^*(a) = V_b^*(a)$ is the bootstrap variance estimator obtained from (2.9) by bootstrapping the particular bootstrap sample $\{Y_{hi}^b\}$ i.e. by replacing $Y_{hi}$ by $Y_{hi}^b$ in the proposed method. For the second phase bootstrapping one could choose values $(m_h^b, B^b)$ different from $(m_h, B)$. This double-bootstrap method thus leads to $B$ values $t_1^*, \ldots, t_B^*$ of $t^*$. Utilizing the bootstrap histogram of $t_1^*, \ldots, t_B^*$, we define

$$\hat{\text{CDF}}_j(t) = \#(t^* \leq t)/B, \quad \hat{\text{CDF}}_j^{-1}(Q) = \hat{\text{CDF}}_j^{-1}(1-Q),$$

and construct an approximate $(1-2Q)$-level confidence interval for $\theta$ given by

$$\{\hat{\theta} - \hat{\theta}_\text{UP}^b, \hat{\theta} - \hat{\theta}_\text{LOW}^b\} \quad (3.4)$$

We now provide an asymptotic justification for $t^*$. Noting that $\sqrt{V_b}$ is a Monte Carlo approximation to

$$\sqrt{V_b} = \frac{1}{B} \left[ \sum_{b=1}^{B} (\hat{\theta}^b - \hat{\theta})^2 \right]^{1/2}, \quad (3.5)$$

where $\hat{\theta}^b$ is the value of $\hat{\theta}$ obtained from bootstrapping the particular sample $\{Y_{hi}^b\}$ and $E^b$ is the second phase bootstrap expectation, we can write $t^* = (\hat{\theta} - \theta)/V_b$. In the linear case $\theta = \hat{\theta}$ it is easily seen that

$$E_*(t^*^2) = \hat{V}_b = \frac{1}{B} \left[ \sum_{b=1}^{B} (\hat{\theta}^b - \hat{\theta})^2 \right]^{1/2} \quad (3.6)$$

In the nonlinear case, following Bickel and Freedman (1984), we can show that $E_*(t^*^2)/\hat{V}_b$ converges to 1 in probability as $n \to \infty$. Hence, it follows that the conditional distribution of $t^*$ is asymptotically $N(0,1)$. One could use a jackknife t-statistic $t_j = (\hat{\theta} - \hat{\theta})/\hat{V}_j$ instead of $t^*$, where $\hat{V}_j$ is a jackknife variance estimator of $\hat{\theta}$ (see Krewski and Rao, 1981). The corresponding confidence interval is then given by

$$\{\hat{\theta} - \hat{t}_\text{UP}^j, \hat{\theta} - \hat{t}_\text{LOW}^j\} \quad (3.7)$$

where $\hat{t}_\text{LOW}$ and $\hat{t}_\text{UP}$ are the lower and upper $\alpha$-points of the statistic $t_j = (\hat{\theta} - \hat{\theta})/\hat{V}_j^*$ obtained from the bootstrap histogram of $t_j^1, \ldots, t_j^B$, and $\hat{V}_j^*$ is obtained from $\hat{V}_j$ by jackknifing the particular bootstrap sample $\{Y_{hi}^b\}$. It can be shown that the confidence interval (3.7) is also asymptotically correct. A confidence interval of this type was considered by Efron (1981) in the case of an i.i.d. sample $\{Y_i\}$.

It is possible to replace $\hat{V}_j^*$ by the BRR or the linearization variance estimator and obtain a confidence interval similar to (3.7).

#### 3.3. Choice of $m_h$

The choice $m_h = n_h$ is a natural one. The choice $m_h = n_h - 1$ gives $Y_{hi}^* = \hat{Y}_{hi}$ and our method reduces to the naive bootstrap, except that in step (i) of the latter method a simple random sample of size $n_h - 1$ is
selected from \( \{y_i\}_{i=1}^{n_h} \) is stratum \( h \). For \( n_h = 2 \), \( m_h = 1 \), the method reduces to the well-known random half-sample replication and the resulting variance estimators are less stable than those obtained from BRR for the same number, \( B \), of half samples (McCarthy, 1969). It may be worth considering a bootstrap sample size \( m_h \) larger than \( n_h = 2 \), say \( m_h = 3 \) or 4.

In the linear case (\( p = 1 \)), it can be shown that the bootstrap third moment, \( \mathbb{E}(y - y)^3 \), matches the unbiased estimate of the third moment of \( y \), if \( m_h = (n_h - 2)/(n_h - 1) \), \( n_h \geq 4 \). This property is not enjoyed by the jackknife or the BRR. Similarly, it can also be shown that the same choice of \( m_h \) ensures that the bootstrap histogram of a t-statistic (see section 3.2) approximates the true distribution of \( t \) with a remainder term of \( O(n^{-1}) \). Details will be provided in a separate paper.

4. Stratified Simple Random Sampling Without Replacement

All the previous results apply to the case of stratified simple random sampling without replacement by making a slight change in the definition of \( \tilde{y}_{hi} \):

\[
\tilde{y}_{hi} = y_h + m_h(n_h - 1)^{-1}(1-f_h)^n(y_{hi} - \bar{y}_h), \tag{4.1}
\]

where \( f_h = n_h/N_h \) is the sampling fraction in stratum \( h \). It is interesting to observe that, even by choosing \( m_h = n_h - 1 \), \( \tilde{y}_{hi} \neq y_{hi} \). Hence the naive bootstrap using \( \tilde{y}_{hi} \) will still have the problem of giving a wrong scale as discussed before. In the special case of \( n_h = 2 \) for all \( h \), McCarthy (1969) used a finite population correction similar to (4.1) in the context of BRR. It can be shown that the choice

\[
m_h = \left( \frac{(n_h - 2)}{(n_h - 1)} \right) \left( \frac{1}{1 - 2f_h} \right)^2 \tag{4.2}
\]

matches the third moments, but it provides an approximation to the true distribution of the t-statistic with a remainder term of \( O(n^{-1}) \) only, as in the case of a normal approximation.

Nickel and Preisendorfer (1984) considered a different bootstrap sampling method in order to recover the finite population correction, \( 1-f \), in the variance formula. This method essentially creates populations consisting of copies of each \( y_{hi} \), \( i = 1, \ldots, n_h \) and \( h = 1, \ldots, L \) and then generates \( \{y_i^*\}_{i=1}^{m} \) as a simple random sample without replacement from the created population, independently in each stratum. This "blow-up" bootstrap was first proposed by Gross (1980) and also independently by Chao and Lo (1983). The variance estimator resulting from this method (by working directly with \( y_{hi}^* \)), however, remains inconsistent for estimating the true variance of \( \hat{\delta} \). It is possible, however, to make the variance estimator consistent by reducing the bootstrap sample size to \( n_h - 1 \), as in Section 2.

The "blow-up" bootstrap approximates the true distribution of the t-statistic with a remainder term of \( O_p(n^{-2}) \) (see Abramovitch and Singh, 1984), unlike our method.

5. Unequal Probability Sampling Without Replacement


Rao, Hartley and Cochran (1962) proposed a simple method of sampling with unequal probabilities and without replacement. The population of \( N \) units is partitioned at random into \( n \) groups \( G_1, \ldots, G_n \) of sizes \( N_1, \ldots, N_n \) respectively (\( \sum N_i = N \)) and then one unit is drawn from each of the \( n \) groups with probabilities \( p_k = \frac{N_k}{N} \) for the \( k \)-th group. Here \( p_k = x_k/X \), \( x_k = \sum_i x_{ki} \) is a measure of size of the \( t \)-th unit \( (t = 1, \ldots, N) \) that is approximately proportional to \( y_t \) (in the scalar case of \( p = 1 \)) and \( X = \sum x_i \). Their unbiased estimator of \( \bar{Y} \) is

\[
\widehat{\bar{Y}} = \frac{n}{\sum (k = 1)} \frac{x_k}{p_k} (\hat{y}_k - \hat{\bar{y}}), \tag{5.1}
\]

where \( \hat{x}_k = x_k/(Np_k) \) and \( (\hat{y}_k, p_k) \) denote the values for the unit selected from the \( k \)-th group \((G_k)\). An unbiased estimator of variance of \( \hat{\bar{Y}} \) is

\[
\text{var}(\hat{\bar{Y}}) = \lambda^2 \sum p_k (\hat{y}_k - \hat{\bar{Y}})^2, \tag{5.2}
\]

where

\[
\lambda^2 = \frac{(N^2 - N)}{(N^2 - \sum N_k^2)}. \tag{5.3}
\]

In the special case of equal group sizes, \( N_k = N/n \), \( \lambda \) reduces to \( (1-f)^3(n(n-1))^{-1} \) where \( f = n/N \).

Our bootstrap sampling for the above method is as follows: (1) Attach the probability \( p_k \) to the sampled unit from \( G_k \) and then select a sample \( \{y_i^*, p_i^*\}_{i=1}^{m} \) of size \( m \) with replacement with probabilities \( p_k \) from \( \{y_k, p_k\}_{k=1}^{n} \).

Calculate \( z_i^* = y_i^*/(p_i^*) \),

\[
\frac{\hat{z}_1}{m} = \frac{\hat{\bar{Y}} + \lambda m^{1/2}(z_1 - \hat{\bar{Y}})}{m} = \hat{z}_1 = \hat{\bar{Y}} + \lambda^{1/2}(z_1 - \hat{\bar{Y}}) \tag{5.4}
\]

Calculate \( \hat{y} = \theta N^2/(N^2 - N) \),

\[
\bar{y} = \frac{\hat{y}}{n}, \tag{5.5}
\]

In the linear case with \( p = 1, \theta = \bar{y} \), we get

\[
\text{var}(\hat{\bar{Y}}) = \text{var}(z_1) = \lambda^2 \text{var}(z_1), \tag{5.6}
\]

Hence, the bootstrap variance estimator, \( \text{var}(\hat{\bar{Y}}) \), reduces to \( \text{var}(\bar{y}) \) in the linear case. The properties of \( \text{var}(\theta) \) in the nonlinear case and the bootstrap methods of setting confidence
intervals for $\theta$, given in Section 3, are being investigated.

5.2 General Results. A general linear unbiased estimator of the population total $Y$ (in the scalar case of $p=1$) based on a sample $s$ of size $n$ with associated probability of selection $p(s)$, is given by

$$\hat{Y} = \sum_{i \in s} d_i(s)y_i,$$

where the known weights $d_i(s)$ can depend on $s$ and the unit $i (i \in s)$. Rao (1979) proved that a nonnegative unbiased quadratic estimator of variance of $\hat{Y}$ is necessarily of the form

$$\text{var}(\hat{Y}) = -\sum_{i<j} d_{ij}(s)w_i w_j (z_i - z_j)^2,$$

(5.8)

where $z_i = y_i/w_i$, the $w_i$ are known constants such that $\text{Var}(Y) = 0$ when $y_i = w_i$. Here the known coefficients $d_{ij}(s)$ satisfy the unbiasedness condition

$$Ed_{ij}(s) = \text{Cov}[d_i(s), d_j(s)],$$

(5.9)

where $d_{ij}(s) = 0$ if $i \not\in s$, and $d_{ij}(s) = 0$ if $s$ does not contain both $i$ and $j$. As an example, the well-known Horvitz-Thompson estimator $Y_{HT}$ satisfies (5.7) with $d_{i}(s) = 1/n_i$, $i \in s$ and $w_i = n_i$ and

$$-d_{ij}(s)w_i w_j = \pi_i \pi_j - \pi_{ij},$$

(5.10)

where $\pi_i = \sum_{s} p(s)$ and $\pi_{ij} = \sum_{s} p(s)$ are the first and second order inclusion probabilities. Let $b_{ij}(s) = -d_{ij}(s)/2$ and assume that $b_{ij}(s) > 0$ for all $i \in s$ and all $j \not\in s$. This is true for many well-known schemes, including the Rao-Hartley-Cochran (RHC) method.

Our bootstrap sampling is as follows: (1) consider all the $n(n-1)$ pairs $(i,j)$, $i \not\in j$, and select $m$ pairs $(i^*, j^*)$ with replacement with probabilities $\pi_{i^* j^*} (= \pi_{j^* i^*})$ to be specified. Calculate

$$\tilde{Y} = \hat{Y} + \sum_{(i^*,j^*) \in s^*} k_{i^* j^*}(z_{i^*} - z_{j^*}),$$

(5.11)

$$\tilde{\theta} = g(\tilde{Y}),$$

where $k_{i^* j^*} = k_{ij}$ for $i < j < s$ to be specified, and $s^*$ denotes the set of $m$ bootstrap pairs. (2) Implement steps (ii) and (iii) of Section 2.2 using $\theta$ obtained from (5.11).

In the linear case with $p=1$, $\theta = \gamma$, we get

$$E_s(\tilde{\gamma}) = \hat{\gamma} + \sum_{i<j \not\in s} k_{ij} z_i z_j (z_i - z_j).$$

(5.12)

Letting $\Sigma k_{ij}^2 = c_i$ and noting that $j(\hat{d}_i)$, we get

$$E_s(\hat{Y}) = \hat{\gamma} + \sum_{i \in s} c_i (z_i - c_i) = \hat{\gamma}.$$

Similarly,

$$\text{var}_s(\tilde{\gamma}) = \frac{1}{m} E_s((k_{i^* j^*}(z_{i^*} - z_{j^*}))^2)$$

(5.13)

$$= \frac{1}{m} \sum_{i \not\in j} k_{ij}^2 (z_i - z_j)^2.$$ (5.14)

We now choose $k_{ij}^2 = \pi_{ij}/d_{ij}(s) w_i w_j$. Thus

$$\text{var}_s(\tilde{\gamma}) = \frac{1}{m} \sum_{i < j} k_{ij}^2 \pi_{ij}.$$ (5.15)

The choice of $k_{ij}$ and $\lambda_{ij}$ satisfying (5.14) is not unique. One simple choice is $\lambda_{ij} = 1/(n(n-1))$, i.e., equal probabilities for each of the $n(n-1)$ pairs of $i \not\in j$. Another choice is

$$\lambda_{ij} = \pi_{ij}/d_{ij}(s),$$

(5.16)

using (5.10) in (5.14). If $m = n(n-1)$ and $\lambda_{ij} = 1/(n(n-1))$, then (5.16) reduces to

$$\lambda_{ij} = \pi_{ij}/d_{ij}(s).$$

(5.17)

Hence, $\tilde{\gamma}_{HT}$ reduces to

$$\tilde{\gamma}_{HT} = \gamma_{HT} + \sum_{(i^*,j^*) \in s^*} \pi_{i^* j^*} (z_{i^*} - z_{j^*}).$$ (5.18)

It follows from (5.13) that $\text{var}_s(\tilde{\gamma}_{HT})$ reduces to the well-known Yates-Grundy variance estimator:

$$\text{Var}(\tilde{\gamma}_{HT}) = \sum_{i < j} \pi_{ij} (z_i - z_j)^2.$$ (5.19)

The properties of $\text{var}_s(\tilde{\gamma})$ in the nonlinear case and the bootstrap confidence intervals on $\theta$ are being investigated.

Remark. In the case of RHC method, we now have two different bootstrap methods: (1) The method of Section 6.1 based on selecting a bootstrap
sample with probabilities $P_k$ and with replacement from $\{y_k, P_k\}_{k=1}^n$. (2) The present method of selecting a bootstrap sample of pairs with replacement from the $n(n-1)$ pairs $(i, j) \in \mathcal{S}$, $i \neq j$, with probabilities $\lambda_{i,j}$. We are investigating their relative merits, but intuitively method (1) is more appealing.

6. Two-stage Cluster Sampling Without Replacement

Suppose that the population is comprised of $N$ clusters with $M_t$ elements (subunits) in the $t$-th cluster ($t = 1, \ldots, N$). The population size $M_0$ ($= \sum M_t$) is unknown in many applications. A simple random sample of $n$ clusters is selected without replacement, and $M_t$ elements are chosen, again by simple random sampling without replacement, from the $M_t$ elements in $t$-th cluster if the latter is selected. The customary unbiased estimator of the population total $Y$ is

$$\hat{Y} = \frac{N}{n} \sum_{i=1}^{n} M_i \bar{Y}_i = \frac{N}{n} \sum_{i=1}^{n} \frac{M_i Y_{ij}}{M_0}, \quad (6.1)$$

where $\bar{Y}_i$ is the sample mean for $i$-th sample cluster. The corresponding estimator of $\theta$ is written as $\hat{\theta} = g(\hat{Y})$, where

$$\hat{\theta} = \frac{\hat{Y}}{M_0} = \frac{1}{n} \sum_{i=1}^{n} \frac{M_i Y_{ij}}{M_0}, \quad (6.2)$$

where $\bar{M}_0 = M_0/N$. For instance, if $\theta = Y/M_0$ where $M_0$ is unknown, then $\hat{\theta} = \hat{\theta}/M_0$ and $\hat{\theta}_i = (N/n) \sum M_i \bar{Y}_i$ with $\bar{Y}_i = 1$ for all elements $j$ in any cluster $i$. An unbiased estimator of variance of $\hat{\theta}$ is given by (Cochran, 1977, p.303)

$$\text{var} (\hat{\theta}) = \frac{(1-f_1) n (\bar{Y} - \hat{\theta})^2}{n(n-1)} \sum_{i=1}^{n} \left( \frac{M_i}{M_0} - \frac{1}{M_0} \right)^2 + \frac{f_1}{n^2} \sum_{i=1}^{n} \frac{(1-f_21) M_i (\bar{Y}_{ij} - \hat{\theta}_i)^2}{M_{ij} (M_0 - \bar{Y}_0)^2}, \quad (6.3)$$

where $Y_{ij}$ is the $y$-value for $j$-th sample element in the $i$-th sample cluster, $f_1 = n/N$ and $f_21 = M_1/M_i$.

We employ two-stage bootstrap sampling to obtain $\{y_{ij}\}$ from $\{Y_{ij}\}$ as follows: (1) Select a simple random sample of $n$ clusters with replacement from the $n$ sample clusters and then draw a simple random sample of $M_t$ elements with replacement from the $M_t$ elements in $t$-th sample cluster if the latter is chosen. (Independent bootstrap subsampling for the same cluster chosen more than once.) We use the following notation: $\hat{Y}_i = \hat{Y}$-value of the $i$-th bootstrap cluster; $\hat{Y}_i = \hat{Y}_i$-value of the $i$-th bootstrap cluster (similarly $M_i$), and

$$\hat{Y}_i = \hat{Y}_i$$

Calculate

$$\hat{Y}_{ij} = \hat{Y} + \lambda_1 \left( \hat{Y}_i - \hat{Y}_j \right) + \lambda_2 \left( \frac{M_i Y_{ij} - M_j Y_{ij}}{M_0} \right), \quad (6.4)$$

and

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} \frac{M_i}{m_{ij}} \hat{Y}_{ij}, \quad (6.5)$$

where $\lambda_{i,j} = M_{i,j}/M_0$, $\hat{\theta}_i = \hat{\theta}_i/\bar{M}_0$, and

$$\lambda_1 = \frac{n-1}{n-1} \frac{1}{m_{ij}} \hat{Y}_{ij}, \quad \lambda_2 = \frac{1}{n(n-1)} \frac{1}{m_{ij}} M_{i,j}/M_0. \quad (6.6)$$

(2) Implement steps (ii) and (iii) of Section 2.2 with $\hat{\theta} = g(\hat{Y})$.

Let $E^*_1$ and $\text{var}^*_1$ respectively denote the conditional bootstrap expectation and variance, for a given bootstrap sample of clusters. Similarly, $E^*_2$ and $\text{var}^*_2$ denote the bootstrap expectation and variance respectively for the sample clusters. Then

$$E^*_1 (\hat{\theta}) = \hat{\theta} + \frac{1}{n} \sum_{i=1}^{n} \frac{Y_{ij} - \hat{Y}_i}{M_0} \quad (6.7)$$

Similarly,

$$\text{var}^*_1 (\hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{Y_{ij} - \hat{\theta}_i}{M_0} \right)^2 \quad (6.8)$$

Also

$$\text{var}^*_2 (\hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \frac{M_i \hat{Y}_{ij} - M_0 \hat{Y}_i}{M_0} \quad (6.9)$$

where $\lambda_2 = \frac{1}{n(n-1)} \frac{1}{m_{ij}} M_{i,j}/(M_0-\hat{\theta}_0)$. Hence, combining (6.8) and (6.9) we get

$$\text{var}^*_2 (\hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{Y_{ij} - \hat{\theta}_i}{M_0} \right)^2 \quad (6.10)$$

Hence, the bootstrap variance estimator, $\text{var}^*_1 (\hat{\theta})$, reduces to $\text{var}(\hat{\theta})$ in the linear case. Thus the variability between bootstrap estimates correctly accounts for the between-cluster and within-cluster components of the variance without the necessity of estimating them separately as in the case of BRR or the jackknife.

Acknowledgement.

This research was supported by a grant from
the Natural Sciences and Engineering Research Council of Canada. C.F.J. Wu is partially supported by NSF grant No. MCS-83-00140. Our thanks are due to David Binder for useful discussions on the bootstrap for unequal probability sampling without replacement.

References


