SIMPLIFYING COMPLEX SAMPLES WITH THE BOOTSTRAP

Bruce D. Spencer, Northwestern University and Stanford University

Numerous techniques are available for estimating population characteristics from data from simple random samples, but relatively few have been adapted for application to complex samples. Methods based on resampling will be discussed as means for estimating population characteristics, standard errors, and confidence intervals. The resampling methods considered here are various versions of the bootstrap.

I. Estimation of (possibly complicated) characteristics from a complex probability sample

We first observe that many population characteristics can be expressed as functionals of the population cumulative distribution (cdf) F.

Example 1A: Mean and Variance. If the population consists of N real numbers $X_1,\ldots,X_N$, then the population cdf is given by $F(t)=N^{-1}\Sigma_N I(X_i\leq t)$ where $I(\cdot)$ is the indicator function and Σ_N denotes summation as the index runs from 1 up to N. The population mean μ_X is given by the functional $\mu(F) = \int t\,dF(t)$ and the population variance σ^2_X is given by the functional $\sigma^2(F) = \int(t-\mu_X)^2 dF(t)$.

Example 2A: Correlation and bivariate regression. If the population consists of N pairs of real numbers $(X_1,Y_1),\ldots,(X_N,Y_N)$, then the population cdf is $F(s,t) = N^{-1}\Sigma_N I(X_i\leq s)I(Y_i\leq t)$. The population correlation coefficient ρ_{XY} is the nonlinear functional $\rho(F) = (\sigma_X\sigma_Y)^{-1}\cdot\int\int(s-\mu_X)(t-\mu_Y)dF(s,t)$ and the population regression coefficient β of Y on X is given by the functional $\beta(F) = \rho(F)\cdot\sigma_Y(F)/\sigma_X(F)$. Least squares theory shows that the population regression coefficient may also be interpreted as minimizing

$$(1)\qquad \int\int g(\gamma,z)dF(z)$$

with $\gamma = (\alpha,\beta)$, $z = (s,t)$, and $g(\gamma,z)=(t-\alpha-\beta s)^2$.

Example 3A: Latent Distributions. The following latent trait model used for studying educational attainment of a group of individuals provides a more complicated example of a population characteristic of the form (1). Each of N individuals i in the population is capable of providing a random vector X_i in response to a questionnaire, and an individual's probability distribution over alternative responses is governed by his latent trait, θ_i. The probability that individual i with latent trait θ_i would choose response x is assumed to be $\Psi(x,\theta_i)$. The latent trait of individual i has distribution G_i, and $G = N^{-1}\cdot\Sigma_N G_i$ denotes the cdf of the latent trait in the population. Although estimation of Ψ is often of interest, here we will assume Ψ to be known.

Imagine the set of all responses x_j, $j=1,\ldots,MN$ that would occur if each individual answered the questionnaire independently and repeatedly a large number M of times. Ignoring "labeling," i.e., knowledge of which individual provides which responses, we see that the (marginal) probability of any response vector x is $\int\Psi(x,\theta)dG(\theta)$ and the log-likelihood of the MN responses $x_1,\ldots,x_{MN}$ is $\Sigma_{MN}\log\int\Psi(x_j,\theta)dG(\theta)$ or equivalently

$$(2)\qquad \int\log\{\int\Psi(x,\theta)dG(\theta)\}dF(x)$$

where F is the ecdf of the MN responses. Now let G^* be distribution G maximizing (2); since G^* need not be unique it may be necessary to assume that G and G^* lie in some family of distributions (Lord, 1969). Consistency results for maximum likelihood estimators imply that under some conditions on Ψ and G, G^* converges to the true cdf G. Thus, we can view the model-based concept of a latent distribution as a population characteristic if we are willing to (i) imagine a "superpopulation" of M responses from each individual, and (ii) to ignore labels (Rao, 1971).

The class of estimators we will consider are based on the simple idea of estimating F by $\hat{F}$, say, and then a functional $\theta(F)$ by $\theta(\hat{F})$. For example, if the data were based on equal probability sampling, we might estimate F by $\hat{F}$, the empirical cdf (ecdf), and hence estimate the moments of F by the moments of $\hat{F}$. Thus, we would estimate the mean by the sample mean, the variance by the average squared deviation about the sample mean (i.e., biased estimator of variance), correlation by the sample correlation, etc.

Example 1B. If $x_1,\ldots,x_n$ denote a simple random sample of size n from the population of example 1, then the ecdf is given by $\hat{F}(t) = n^{-1}\Sigma_n I(x_i\leq t)$. Thus, a functional estimator of the mean is $\bar{x} = \mu(\hat{F}) = \int t\,d\hat{F}(t) = n^{-1}\Sigma_n x_i$ and the functional estimator of the variance is

$$(3)\qquad \hat{\sigma}^2_X = \int(t-\bar{x})^2 d\hat{F}(t) = n^{-1}\Sigma_n(x_i-\bar{x})^2.$$

Example 2B. Let $(x_1,y_1),\ldots,(x_n,y_n)$ denote a random sample of size n from the population of example 2A. The ecdf is given by $\hat{F}(s,t) = n^{-1}\cdot\Sigma_n I(x_i\leq s)\,I(y_i\leq t)$ and a functional estimator of the correlation coefficient is $\hat{\rho} = \rho(\hat{F}) = (\hat{\sigma}_x\hat{\sigma}_y)^{-1}\cdot\int\int(s-\bar{x})(t-\bar{y})d\hat{F}(s,t)$, which is the usual sample correlation. Similarly, a functional estimator of the population regression coefficient is $\beta(\hat{F})$, which is the usual least-squares estimator. One way to calculate $\beta(\hat{F})$ is as $\hat{\rho}\hat{\sigma}_y/\hat{\sigma}_x$. An alternative is to substitute $\hat{F}$ for F in (1).

Example 3B. Suppose a simple random sample of n individuals is taken from a population, and from individual i a response vector x_i is obtained, $i=1,\ldots,n$. To estimate the population distribution of latent traits that maximizes (2), we substitute the ecdf $\hat{F}$ for F in (2) and choose G to maximize.

$$(4)\qquad \int\{\log\int\Psi(x,\theta)dG(\theta)\}d\hat{F}(x).$$

Bock and Aitkin (1981) and Mislevy (1983) provide techniques based on the EM-algorithm for performing this maximization.

Alternative estimators of F, such as kernel estimators or smoothed versions of $\hat{F}$, lead to alternative estimators of population characteristics. Many common estimators cannot be expressed as functional estimators in which the function is a proper probability distribution. Thus (recall example 1B) to get the usual unbiased estimator of variance we would have to replace $\hat{F}$ in (3) by $n(n-1)^{-1}\hat{F}$, which is not a proper cdf.

Now we consider unequal probability sampling with replacement. Suppose real values $x_1,\ldots,x_n$ are observed with known draw-by-draw selection probabilities $\pi_1,\ldots,\pi_n$ from a population of size

N. Then we might consider estimating F by the unbiased estimator $\hat{F}_U(t) = N^{-1}\Sigma_n I(x_i \le t)\cdot\pi_i^{-1}$. However, $\hat{F}_U$ need not be a proper cdf. For example, if $n=1$ and $\pi_1 = (cN)^{-1}$ with $c>1$ then $\hat{F}_U(t) = c>1$ for $t \ge x_1$.

An alternative to $\hat{F}_U$ is the ratio estimator

$$(5) \qquad \hat{F}_R(t) = N\hat{F}_U(t)/\Sigma_n \pi_i^{-1} = \Sigma_n I(x_i \le t)w_i$$

where $w_i = \pi_i^{-1}/\Sigma_n \pi_j^{-1}$.

Example 1C. Suppose $x_1,\ldots,x_n$ is an unequal probability sample with replacement and draw-by-draw selection probabilities $\pi_1,\ldots,\pi_n$. The functional estimator of the mean is, using $\hat{F}_U$, the unbiased estimator $\bar{x}_U = \mu(\hat{F}_U) = \int t d\hat{F}_U(t) = (nN)^{-1}\Sigma x_i/\pi_i$. Using $\hat{F}_R$, we obtain the ratio estimator of the mean, $\bar{x}_R = \mu(\hat{F}_R) = \Sigma_n x_i w_i$, which is biased.

To estimate the population variance we could use the functional estimator $\hat{\sigma}^2_{RX} = \sigma^2(\hat{F}_R) = \Sigma_n (x_i-\bar{x}_R)^2 w_i = \Sigma_n \Sigma_n (x_i-x_j)^2 w_i w_j$. This estimator is biased, but the bias can be removed by multiplying $\hat{\sigma}^2_{RX}$ by $(\Sigma_n \pi_i^{-1})^2 N^{-2} n^{-1}(n-1)^{-1}$ or equivalently by calculating $\sigma^2(n(n-1)^{-1}\hat{F}_U)$. Kish (1964, 1965), developed a different unbiased estimator of variance. Kish's approach is based on unbiasedness rather than functional estimation to develop estimators of other population characteristics.

Example 2C. Suppose $(x_1,y_1),\ldots,(y_n,y_n)$ is an unequal probability sample selected with replacement and draw-by-draw selection probabilities $\pi_1,\ldots,\pi_n$. Then the ratio estimate of the population cdf is $\hat{F}_R(s,t) = (\Sigma_n \pi_i^{-1})^{-1}\cdot\Sigma_n I(x_i \le s) I(y_i \le t)\pi_i^{-1}$. A functional estimate of ρ is the probability-inverse weighted correlation coefficient $\hat{\rho}_R = \rho(\hat{F}_R) = (\hat{\sigma}_{RX}\hat{\sigma}_{RY})^{-1}\int\int(s-\bar{x}_R)(t-\bar{y}_R)d\hat{F}_R(s,t) = \Sigma (x_i-\bar{x}_R)(y_i-\bar{y}_R)\pi_i^{-1}/(\hat{\sigma}_{RX}\hat{\sigma}_{RY})$. Similarly, a functional estimate of the population regression coefficient is

$$(6) \qquad \hat{\beta}_R = \beta(\hat{F}_R) = \hat{\rho}_R\hat{\sigma}_{RY}/\hat{\sigma}_{RX} = \Sigma_n(x_i-\bar{x}_R)(y_i-\bar{y}_R)\pi_i^{-1}/\Sigma_n (x_i-\bar{x}_R)^2\pi_i^{-1}$$

which is the "weighted least squares" estimate. Observe that the weighting in (6) has nothing to do with heteroscedasticity. Rather, the weights enter as they do because the population regression coefficient is a special kind of functional, viz., β is the solution to (1). To estimate β with a functional statistic we minimize (1) with our estimate $\hat{F}$ substituted for F, i.e., minimize $\int\int g(\gamma,z)dF_R(z) = \Sigma (y_i-\alpha-\beta x_i)^2 w_i$.

Example 3C. Suppose an unequal probability sample of n individuals is taken from a population, and from each individual i a response vector x_i and a draw-by-draw selection probability π_i are observed, $i=1,\ldots,n$. Instead of choosing G to maximize (4), we choose G to maximize $\int \log\{\int\Psi(x,\theta)dG(\theta)\}d\hat{F}_R(x)$ with $\hat{F}_R(x)$ the weighted ecdf of response vectors.

Suppose now that we want to estimate the population cdf F from an unequal probability sample of size n drawn without replacement with overall selection probabilities P_i. Thus, in a population of size N we have $n = \Sigma_N P_i$. The estimator (5) can still be applied if we set $\pi_i = P_i/n$.

There are two natural ways to estimate the population cdf F from a stratified sample: the "separate" ratio estimator and the "combined" ratio estimator. Assume there are L strata, N_h sampling units in stratum h with draw-by-draw selection probabilities π_{hi} such that $1 = \Sigma_{N_h}\pi_{hi}$, and n_h observations from stratum h drawn with replacement. Defining the stratum weights as $W_h = N_h/\Sigma_L N_k$, the "estimated" weights as $\hat{W}_h = \Sigma_{n_h}\pi_{hi}^{-1}/\Sigma_L \Sigma_{n_h}\pi_{kj}^{-1}$, and $\hat{F}_{Rh}$ as the estimator (5) computed for stratum h, we may estimate F by the separate ratio estimator $\Sigma_L W_h\hat{F}_{Rh}$ or by the combined ratio estimator $\Sigma_L\hat{W}_h F_{Rh}$. If L is large and the n_h are small then the combined ratio estimator may be better.

To estimate F from a multistage sample we may simply apply the estimator (5) to the ultimate sampling units. If the sample is a stratified multistage sample, we may apply the separate ratio estimator or the combined ratio estimator to the ultimate sampling units within each stratum.

II. Motivation for $\hat{F}_R$ for Samples Drawn with Replacement

To develop estimators for the population cdf when the data come from a complex sample, we will be guided by the ecdf for simple random samples. To generalize the ecdf to complex samples we will use a simple idea: rather than make the estimation procedure more complex, we will simplify the sample and then use the ecdf.

To "simplify" a sample we modify the original sampling procedure so that no clustering and stratification effects are present and so that every element in the population has equal chance of being in the sample. Although we usually cannot completely simplify a complex sample drawn with replacement, we can come close if we resample (subsample) the given sample so that

(1) stratification and clustering are ignored, and
(2) an element's probability of appearing in the simplified sample (subsample) at any draw is inversely proportional to the original draw-by-draw selection probability.

Note that the simplified sample is a subsample of the original sample, and it is drawn with replacement. Thus, the subsampling may lose some information. If we have data to spare, this inefficiency may be tolerable, as when we pull a small subsample from the large data set to perform preliminary analyses. Since in most cases we will not want to lose information, two alternative methods of preserving information are presented in section III. One is based on the "expected subsample" and the other is based on the expected value of the functional estimator, with expectation referring to the subsampling when the original sample is held fixed. For linear functionals the methods give essentially the same results.

Let $x_1,\ldots,x_n$ be an unequal probability sample drawn with replacement and draw-by-draw selection probabilities $\pi_1,\ldots,\pi_n$, and define $w_i = \pi_i^{-1}/\Sigma_n\pi_j^{-1}$ as before. Then a simplified sample of size m, say $x_1^S,\ldots,x_m^S$, is constructed by sampling m times with replacement from $x_1,\ldots,x_n$

using selection probabilities $w_1,\ldots,w_n$. The ecdf of the simplified sample is $\hat{F}^S(t) = m^{-1} \cdot \Sigma_m I(x_i^S \le t)$. However, a better estimator than F^S is the conditional expectation of $\hat{F}^S$ given the original sample: this estimator is equal to $\hat{F}_R$.

III. Computational Considerations

In many cases (including stratified cluster samples with unequal seleciton probabilities) we may use $\hat{F}_R$ to estimate the population cdf F, and to estimate a population characteristic $\theta(F)$ we may use $\theta(\hat{F}_R)$. We will assume that we can readily calculate the estimate $\theta(\hat{F})$ for $\hat{F}$ the ecdf of a simple random sample. However, we might not be able to so readily compute $\theta(\hat{F}_R)$. For example, we might have software for computing ordinary least squares regression estimates (example 2B) but not weighted least squares estimates (example 2C).

If we choose a simplified sample and calculate the ecdf $\hat{F}^S$ then we can compute $\theta(\hat{F}^S)$, but some information may be lost. Two ways around this problem will now be presented.

Instead of using the simplified sample, we can use the "expected value" of the simplified sample, i.e., the "blowup" sample; Gross (1980), Rao and Wú (1984). Thus, if $x_1,\ldots,x_n$ is the original with-replacement sample with draw-by-draw selection probabilities $\pi_1,\ldots,\pi_n$, and $w_i = \pi_i^{-1} / \Sigma_n \pi_j^{-1}$, then we choose m so that $k_i = mw_i$ is an integer (or approximately an integer; Tukey (1948)), and let the blowup sample consist of k_i copies of x_i, $i=1,\ldots,n$. Observe that the ecdf of the blowup sample, $\hat{F}^S$, equals $\hat{F}_R$. Thus, the estimate $\theta(\hat{F}^S)$ which we can calculate, equals $\theta(\hat{F}_R)$. The possible drawback to this approach is that m may need to be quite large.

An alternative approach is a bootstrap method for performing a stochastic numerical integration. We repeatedly and independently simplify the original sample a number B of times, yielding estimates $\hat{\theta}_b = \theta(\hat{F}^{Sb})$ for $\hat{F}^{Sb}$ the ecdf of the b^{th} simplified sample, $b=1,\ldots,B$. Then we choose the average or the median of $\theta_1,\ldots,\theta_B$, say $\tilde{\theta}$, to estimate $\theta(F)$. If we use the average, then $\tilde{\theta}$ is a random approximation to $E(\theta(\hat{F}^S) \mid \text{original sample}) = \theta(\hat{F}_R) + O(m^{-1})$. If the cost of computing $\theta(\hat{F}^S)$ increases rapidly with sample size m, then this bootstrap/stochastic numerical integration method may be less expensive than using the blowup sample.

Example 1D. Suppose $x_1,\ldots,x_n$ is an unequal probability sample drawn with replacement with draw-by-draw selection probabilities $\pi_1,\ldots,\pi_n$. Then application of $\bar{x}$ and σ^2 to the blowup sample produces the estimators $\bar{x}_R$ and σ^2_{RX}. Application of $\bar{x}$ and $\hat{\sigma}^2$ to a simplified subsample leads to statistics $\bar{x}_{RS}$ and $\hat{\sigma}^2_{RXS}$ having the properties that their expected values (over repeated subsamplings) are $\bar{x}_R$ and $\hat{\sigma}^2_{RX}$. As the number B of subsamples used to compute $\bar{x}_{RS}$ and $\hat{\sigma}^2_{RXS}$ increases, or as the size m of the subsamples increase, then $\bar{x}_{RS}$ and $\hat{\sigma}^2_{RXS}$ converge to $\bar{x}_R$ and $\hat{\sigma}^2_{RX}$.

Example 2D. Suppose $(x_1,y_1),\ldots,(x_n,y_n)$ is an unequal probability sample drawn with replacement with draw-by-draw selection probabilities $\pi_1,\ldots,\pi_n$. Applying the ordinary least squares (OLS) estimator of the regression coefficient to the blowup sample yields the probability-inverse weighted regression estimator $\hat{\beta}_R$. To see that the stochastic numerical integration approach not yield $\hat{\beta}_R$, suppose all π_i are equal and consider estimating the slope of the population regression by OLS. If we choose subsamples of size m=2, calculate the OLS estimate of slope from each, and take the median of these over a large number B of subsamples, we get Theil's (1950) "nonparametric" estimator of slope.

IV. Estimation of Standard Errors

Standard errors can be estimated by pseudo-replication methods. Attention here is focused on the bootstrap. Before discussing how to bootstrap complex samples, we review the idea behind the bootstrap in simple random sampling with replacement. As usual, F denotes the population cdf and $\theta(F)$ is a population characteristic. Let s denote an observed simple random sample of size n from F, let $\hat{F}$ be the ecdf of s, and let $\theta(\hat{F})$ be the estimate of $\theta(F)$. To study the sampling distribution of $\theta(\hat{F})$ over possible samples, we hold s fixed and:

1. Draw a subsample s^* of size n from $\hat{F}$,
2. Calculate the ecdf of s^*, $\hat{F}^*$, and also $\theta^* = \theta(\hat{F}^*)$,
3. Repeat steps 1 and 2 many times to generate an ecdf of θ^*. Then use the observed distribution of $\theta^* - \theta(\hat{F})$ to make inferences about the unobserved distribution of $\theta(\hat{F}) - \theta(F)$.

The subsamples s^* are called "bootstrap samples" and the ecdf of θ^* is the "bootstrap distribution."

The bootstrap procedure extends in a generally straightforward way to complex samples (except for two special considerations to be discussed below).

To show how to extend the bootstrap procedure to complex samples we first consider unequal probability sampling (with replacement). Suppose real values $x_1,\ldots,x_n$ are observed with known draw-by-draw selection probabilities $\pi_1,\ldots,\pi_n$ from a population of size N. To estimate the population cdf F use $\hat{F}_R$ given by (5). Holding the sample fixed, choose bootstrap samples from $\hat{F}_R$ but according to the original sampling scheme. The net effect is to draw the bootstrap sample by simple random sampling with replacement from the original sample! For example, suppose $x_1,\ldots,x_n$ are all distinct. Then sampling from $\hat{F}_R$ assigns x_i a selection probability of $w_i \propto \pi_i^{-1}$, and sampling according to the original sampling scheme assigns x_i a selection probability proportional to π_i. The product of these probabilities is constant for all $x_1,\ldots,x_n$, i.e., we use simple random sampling with replacement.

Example 1E. Suppose $x_1,\ldots,x_n$ is an unequal probability sample drawn with replacement with draw-by-draw selection probabilities $\pi_1,\ldots,\pi_n$. Each bootstrap sample is obtained by simple random sampling with replacement from $x_1,\ldots,x_n$. Let x_j and π_j denote the particular values of x_i and π_i corresponding to the i^{th} draw of the bootstrap sample, $j=1,\ldots,n$. Consider estimating the sampling variance of the ratio-estimator of the mean, $\bar{x}_R = \mu(\hat{F}_R)$. For each bootstrap sample form

$$\hat{F}^*_R(t) = \Sigma_n I(x^*_i \le t)\pi^{*-1}_i / \Sigma_n \pi^{*-1}_i$$

and

$$\bar{x}_R = \mu(F_R^*) = \Sigma_n x^* \pi_i^{*-1} / \Sigma_n \pi_i^{*-1} .$$

Repeating the bootstrap sampling procedure a large number, B , of times leads to an ecdf for $\bar{x}_R$. The variance of this ecdf , which is the bootstrap estimator of variance of $\bar{x}_R = \mu(\hat{F}_R)$, tends (as B increases) to

$$(7) \qquad \Sigma_n x_i^2 w_i^2 - 2\bar{x}_R \Sigma_n x_i w_i + \bar{x}_R^2 \Sigma_n w_i^2 + O(n^{-1})$$

with $w_i = \pi_i^{-1} / \Sigma_n \pi_j^{-1}$. The ratio of (7) to the true variance of $\bar{x}_R$ tends in probability to 1 for large n .

If the original sampling procedure is multistage, say primary sampling units (psu's) are selected with possibly unequal selection probabilities and with replacement, then the bootstrap sampling procedure focusses entirely on the primary selections. To bootstrap, draw a simple random sample of psu's with replacement form the original sample, and then select all ultimate sampling units within the chosen psu's. (The use of single-stage replication sampling from a multistage original sample is common to jackknifing and balanced repeated replications (BRR) also, and is justified when the original primary selections are made with replacement.)

If the original sample was drawn by stratified sampling (possibly with unequal selection probabilities within strata), then the bootstrap sample is to be drawn by stratified simple random sampling from the original sample. A special consideration arises in stratified sampling because the bootstrap estimator of variance of the sample mean (in a simple random sample of size n) is biased, and it becomes unbiased if multiplied by $n(n-1)^{-1}$. In stratified sampling with few observations per stratum and many strata these biases become important, and one way of coping with this is to decrease by 1 the size of the bootstrap samples drawn from each stratum. Rao and Wu (1984) discuss this and alternative techniques. But, aside from this consideration, the across-strata allocation of the bootstrap sample should be the same as for the original sample.

A second special consideration arises if the original sample was drawn with replacement. If we use techniques described above to develop a bootstrap estimate of variance of the mean of a simple random sample, then we need to multiply that estimate by the finite population correction factor. For stratified simple random sampling without replacement and with varying sampling fractions across strata, multiplication by a single finite population correction (fpc) will not suffice. One way of coping with this is to differentially increase the sizes of the bootstrap samples drawn from the different strata. (This technique may also be used for estimating variances of nonlinear statistics in simple random sampling.) For multistage samples we would focus on the sampling fractions for the psu's. For unequal probability sampling a sampling design need not have a unique fpc , but we could try to use the factor appropriate for simple random samples,

(8) $(N-n)/N-1)$.

This fpc performed well in limited empirical studies of unequal probability sampling (Cochran, 1977, p. 130). Alternative fpc's may be derived from superpopulation models. For example, if we draw an unequal probability sample of size n from a population of size N without replacement and with overall selection probabilities $P_1,\ldots,P_N$, the superpopulation model of Brewer (1963, p. 100) leads to the fpc $(N'-n)/(N'-1)$ with $N' = n\Sigma_N P_i /N$. This fpc is always between (8) and 1.

V. Computational Issues.

We will now consider some computational issues that arise in using the bootstrap to estimate standard errors for an estimator that is calculated by stochastic numerical integration. If B_1 subsamples are drawn each time the estimator is computed, and if B_2 bootstrap samples are used to calculate the bootstrap distribution, then a total of B_1B_2 subsamples are used. The amount of computation required could be excessive. To reduce the computation needed to compute standard errors, one may compute the bootstrap distribution using $B_1=1$ even though a larger value was used to develop the parameter estimate. The effect of doing this is to inflate the variance of the bootstrap distribution. However, it may be possible to estimate the inflation of the variance and adjust for it, based on the following rationale.

Let E_1 and V_1 denote expectation and variance computed with respect to the subsampling used in the stochastic numerical integration--recall, this subsampling "simplified" the original sample. Let s denote the original sample, s* a bootstrap sample from s , s_1 a simplified sample from s , and s_1^* a simplified sample from s* , and denote the ecdf's of these samples by $\hat{F}$, $\hat{F}^*$, $\hat{F}_1$, and $\hat{F}_1^*$ respectively. Let E and V denote expectation and variance with respect to the original sampling procedure. The estimator derived by the stochastic numerical integration has the form $E_1(\theta(\hat{F}_1)|s)$, and what we wish to estimate is $V(E_1(\theta(\hat{F}_1)|s))$. The same monte-carlo that provides an estimate of $E_1(\theta(\hat{F}_1)|s)$ can give us an estimate of $V_1(\theta(\hat{F}_1)|s)$, say v_1 (e.g., let v_1 be the observed variance of the $\theta(\hat{F}_1)$ values generated by the monte-carlo).

Observe that, by the usual formula relating conditional and unconditional variances and expectations,

$$(9) \qquad V(E_1(\theta(\hat{F}_1)|s)) = V(\theta(\hat{F}_1)) - E(V_1(\theta(\hat{F}_1)|s)) .$$

The second term on the right hand side of (9) is estimated by v_1 . The first term on the right can be estimated by a kind of "two-stage bootstrapping." First draw the bootstrap sample s* from s , and then draw a simplified sample, s_1^* , from s_* and calculate the ecdf $\hat{F}_1^*$ and then calculate $\theta(\hat{F}_1^*)$. Repeat this two-stage bootstrap sampling B_2 times, and calculate the ecdf of $\theta(F_1^*)$. We will refer to this ecdf as the "two-stage bootstrap distribution." The variance of the two-stage distribution, v_2 , serves as an estimate of $V(\theta(\hat{F}_1))$. Now subtracting v_1 from v_2 yields an estimate of the variance of the stochastic numerical integration estimator $E_1(\theta(\hat{F}_1)|s)$.

A drawback of this method is the possibility that v_1 could exceed v_2 , giving a negative variance estimate.

In principle, the subsampling leading to s_1 and s_* can be replicated enough times to ensure

non-negative variance estimates, but if too many replications are needed the purpose of this two-stage approach--reducing computation expense--is sacrificed.

VI. Confidence Intervals

Various approaches may be used to develop confidence intervals from the bootstrap distribution, such as the percentile method and the modified percentile method. Since these methods are described elsewhere, we won't discuss them here; see Efron (1982, 1981), and Tibshirani (1984). If the two-stage bootstrap distribution is used, the intervals will be too wide because the variance is too large by the amount $E(V_1(\theta(\hat{F}_1)|s))$. One untested approach is to rescale the two-stage bootstrap distribution to reduce its variance by a proportion v_1/v_2. To accomplish this, we could try to rescale the intervals, or we could recalculate the two two-stage bootstrap distribution, with all bootstrap sample sizes multiplied by $(1-v_1/v_1)^{\frac{1}{2}}$.

The author is grateful* for advice from Marty Frankel, Lincoln Moses, Tim Hesteberg, and Rob Tibshirani, who in no way should be blamed for any errors that may be in this paper. This research was supported by grant NIE-G-83-044 from the National Institute of Education.

References

Bock, R. D., and Aitkin, M. (1981) "Marginal Maximum Likelihood Estimation of Item Parameters: An Application of an EM Algorithm." Psychometrika, 46: 443-359.

Cochran, W. G. (1977) Sampling Techniques, 3rd ed. New York: Wiley.

Efron, B. (1982) The Jackknife, the Bootstrap, and Other Resampling Plans. SIAM, monograph #38, CBMS-NSF. Philadelphia.

_____ (1981) "Nonparametric Confidence Intervals and Standard Errors" Canadian Journal of Statistics. 9, 139-172.

Gross, S. T. (1980) "Median Estimation in Sample Surveys." Proceedings of the American Statistical Association, Section on Survey Research Methods: 181-184.

Kish, L. (1964) "Generalizatons for Complex Probability Sampling." Proceedings of the American Statistical Association, Social Statistics Section: 63-67.

Kish, L. (1965) Survey Sampling. New York: Wiley.

Lord, F. (1967) "Estimating True-Score Distributions in Psychological Testing (An Empirical Bayes Estimation Problem), Psychometrika, 34: 259-299.

Mislevy, R. J. (1983) "Characterizations of latent distributions." Unpublished manuscript.

Rao, C. R. (1971) "Some Aspects of Statistical Inference in Problems of Sampling from Finite Populations." Pp. 177-202 in V. P. Godambe and D. A. Sprott, Foundations of Statistical Inference, Canada: Holt, Rinehard, and Winston.

Rao, J. N. K. and Wu, C. F. J. (1984) "Bootstrap Inference with Stratified Samples." Mathematics Research Center, University of Wisconsin, Madison. Technical report #2629.

Theil, H. (1950) "A Rank-Invariant Method of Linear and Polynomial Regression Analysis, I." Proc. Kon. Ned. Akad. v. Wetensch. A., 53: 386-392.

Tibshirani, R. J. (1984) "Bootstrap Confidence Intervals." Division of Biostatistics, Stanford University. Technical report #91.

Tukey, J. W. (1948) "Approximate Weights," Annals of Mathematical Statistics, 19: 91-92.