The Annals of Statistics
1988, Vol. 16, No. 4, 1684–1695

A BAYESIAN BOOTSTRAP FOR A FINITE POPULATION[1]

BY ALBERT Y. LO

State University of New York at Buffalo

A Bayesian bootstrap for a finite population is introduced; its small-sample distributional properties are discussed and compared with those of the frequentist bootstrap for a finite population. It is also shown that the two are first-order asymptotically equivalent.

1. Introduction. The bootstrap method for a finite population (FPB, i.e., finite population bootstrap) was suggested by Gross (1980). Bickel and Freedman (1984) and Chao and Lo (1986) gave a first-order asymptotic justification for the FPB mean. The FPB is derived from a frequentist viewpoint and, for a large population size, the FPB becomes the frequentist bootstrap of Efron (1979).

The frequentist bootstrap of Efron (1979) has a Bayesian analogue called the Bayesian bootstrap [Rubin (1981)]. The operational and small-sample similarities between the bootstrap and the Bayesian bootstrap are well known [Rubin (1981) and Efron (1982)]. Recently, Lo (1987) showed that the Bayesian bootstrap and the frequentist bootstrap are first-order asymptotically equivalent.

The question is then: Is there a Bayesian analogue (FPBB, i.e., finite population Bayesian bootstrap) to the frequentist FPB? This paper is the result of an investigation into this question. The answer is affirmative. A FPBB can be defined in terms of Pólya's urn scheme [Feller (1971)], by simulating a posterior distribution with respect to a "flat" Dirichlet-multinomial prior [Lo (1986)]; this definition of a FPBB is analogous to Rubin's definition of the infinite Bayesian bootstrap in terms of gaps of uniform random variables that are used to simulate a posterior distribution with respect to a "flat" Dirichlet process prior [Ferguson (1973)]. We also find that the FPBB and the FPB share similar operational characteristics and small-sample properties, and that for a large population size the FPBB reduces to the Bayesian bootstrap. In addition, a first-order asymptotic justification of the FPBB and a first-order asymptotic equivalence of the FPBB and FPB are provided.

In Section 2, we define the FPBB in terms of the urn of observations. The "Bayesianess" of the FPBB is illustrated in Section 3. In the same section we also show that the FPBB reduces to the Bayesian bootstrap for a large population size. The FPBB mean and variance are then computed and compared with the FPB mean and variance. Section 4 shows that, given a sample of size n from the population, the FPBB distribution of the standardized unknown empirical distribution of the finite population converges weakly to a Brownian bridge as the sample size tends to ∞, as long as the sample empirical distribution

Received February 1987; revised November 1987.
[1]This work was supported in part by NSF Grant MCS-81-02523-01.
AMS 1980 *subject classification*. Primary 62A15.
Key words and phrases. Finite population, Bayesian bootstrap, bootstrap, Pólya's urn.

converges to a distribution function. The convergence of the FPBB distribution of the population mean (and/or any continuous functional of the population empirical distribution) follows as a corollary. Section 5 discusses the effect of stratification.

2. The FPB and the FPBB. Let $\mathbf{x} = \{x_1, \dots, x_n\}$ be a given sample from a finite population $\{y_1, \dots, y_N\}$, $n \le N - 1$. (Here we interpret a sample as data obtained either from sampling with replacement, or as a realization of the initial sequence of an exchangeable sequence $\{y_1, \dots, y_N\}$; the latter case implies that given the y's, $\mathbf{x}$ is a simple random sample from the y's [Scott (1971)].) Denote the sample empirical distribution function, putting mass $1/n$ on each of the x's, by F_n, and the population empirical distribution, putting mass $1/N$ on each of the y's, by H_N. The y's are elements of a q-dimensional Euclidean space R; denote the Borel σ-field on R by $\mathscr{B}$. Consider a functional $\theta(\mu, \nu)$ of interest; μ and ν are probabilities defined on $\mathscr{B}$. A frequentist is interested in the sampling distribution $\angle\{\theta(F_n, H_N)|H_N\}$. The frequentist FPB method of Gross is to create a FPB population of N objects based on duplicating the x's, to sample without replacement from this bootstrap population to obtain $x_1^*, \dots, x_n^*$ (called a FPB sample; denote the empirical distribution of the x^*'s by F_n^*), and to replace F_n and H_N in $\angle\{\theta(F_n, H_N)|H_N\}$ by F_n^* and F_n, respectively, to obtain a FPB distribution $\angle\{\theta(F_n^*, F_n)|F_n\}$, which approximates the sampling distribution $\angle\{\theta(F_n, H_N)|H_N\}$.

The success of the FPB method is based on the creation of the FPB population from which the FPB sample is dawn. If the population size N is an integer multiple of the sample size n (i.e., $N = nk$), the FPB population can be easily constructed by duplicating the sample $\{x_1, \dots, x_n\}$ k times. In general, there are two integer multiples of n closest to N, i.e., k and $k + 1$, with $N = nk + r$, where $0 \le r \le n$. Each integer multiple generates a preliminary FPB population as before. The FPB sample is obtained by sampling from a mixture of the two with mixing weights β and $1 - \beta$, respectively, where

$$\beta = [1 - r/n][1 - r/(N - 1)]. \tag{2.1}$$

Whereas a frequentist is interested in $\angle\{\theta(F_n, H_N)|H_N\}$, a Bayesian constructs an exchangeable prior on the population variables y's and is interested in the posterior distribution $\angle\{\theta(H_N, F_n)|F_n\}$ [Ericson (1969)]. The FPBB method is the simulation of a random distribution function H_{mn}^*, to replace H_N in $\angle\{\theta(H_N, F_n)|F_n\}$ by H_{mn}^* to obtain a FPBB distribution $\angle\{\theta(H_{mn}^*, F_n)|F_n\}$, which approximates the posterior distribution $\angle\{\theta(H_N, F_n)|F_n\}$.

The FPBB method is based on a generalization of Pólya's urn scheme [Feller (1971), page 210]. Suppose an urn has a finite number of items. Sample from the urn successively according to the following rule: Select an element from the urn at random; it is replaced and, moreover, another element of the same kind is added to the urn. Call a sample of size m obtained from this experiment a Pólya sample of size m. Denote an urn containing $z_1, \dots, z_n$ by urn $\{z_1, \dots, z_n\}$. The

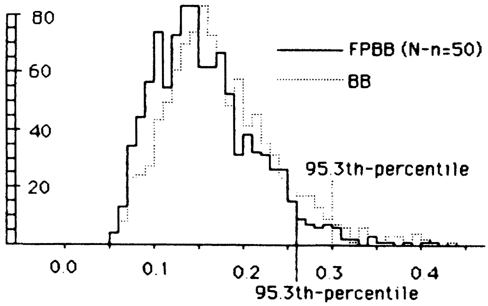

FIG. 1.

FPBB method can be described as follows:

(2.2)

Step 1: Take a Pólya sample of size $N - n$ $(= m \geq 1)$, denoted by $y_1^*, y_2^*, \ldots, y_m^*$, from the urn $\{x_1, \ldots, x_n\}$.

Step 2: Let $H_{mn}^* = [mG_m^* + nF_n]/N$, where G_m^* is the empirical distribution of $y_1^*, y_2^*, \ldots, y_m^*$. Evaluate $\theta = \theta(H_{mn}^*, F_n)$.

Step 3: Repeat the previous steps a large number of times, say B times, to obtain $G_{m1}^*, \ldots, G_{mB}^*$ and hence also $\theta_1, \ldots, \theta_B$.

Step 4: The empirical distribution of $\theta_1, \ldots, \theta_B$ approximates the posterior distribution $\angle\{\theta(H_N, F_n)|F_n\}$ for a large B.

The following example illustrates the algorithm (2.2).

EXAMPLE 2.1. To obtain a 95% probability band for the finite population empirical distribution function H_N, consider the absolute deviation functional

$$\theta(H_{mn}^*, F_n) = \sup_x |H_{mn}^*(x) - F_n(x)|. \tag{2.3}$$

The observations are the fifteen GPAs of the law school data in Efron (1982). The result of 1000 FPBB replications of the preceding functional for $m = 50$ is displayed in Figure 1 (solid line); for comparison, the result of 1000 BB replications of the same data set is also displayed (dotted line). The figure indicates that a 95.3th-percentile point is 0.26, and hence a 95.3% FPBB probability band of the finite population empirical distribution is given by $F_n \pm 0.26$. (Figure 1 also shows that a 95.3% BB probability band of the infinite population empirical distribution is given by $F_n \pm 0.30$.)

Operationally, the FPBB is simply a Pólya sampling from the data urn (Step 1), whereas the FPB is a simple random sampling from a mixture of two preliminary FPB populations. The FPBB is simpler than the FPB in the sense that no randomized population is involved. On the other hand, the Pólya sample size m $(= N - n)$ could be much larger than the FPB sample size n. The difference then is a randomized FPB population against a perhaps larger Pólya sample size.

3. The small-sample distribution of the FPBB. The paramount distribution in the FPB method is the multivariate hypergeometric distribution. The following proposition indicates that the analogue distribution in the FPBB case is the Dirichlet-multinomial distribution. Denote the distinct observations in the urn $\{x_1, \dots, x_n\}$ by $x_{(1)} < \cdots < x_{(k)}$. For $j = 1, \dots, k$, let m_j^* be the number of y^*'s equal to $x_{(j)}$, n_j be the number of x's equal to $x_{(j)}$, i.e., $m_j^* = mG_m^*\{x_{(j)}\}$ and $n_j = nF_n\{x_{(j)}\}$.

PROPOSITION 3.1.

(3.1) $(m_1^*, \dots, m_k^*)|F_n$ *is a Dirichlet-multinomial* $(m; n_1, \dots, n_k)$ *vector*;

(3.2) $\lim_{m \to \infty} \angle\{(m^{-1}m_1^*, \dots, m^{-1}m_k^*)|F_n\}$ *is a Dirichlet* $(n_1, \dots, n_k)$ *vector*.

PROOF. The first statement (3.1) follows from the definition of the m^*'s and the fact that the y^*'s form a Pólya sample from the urn $\{x_1, \dots, x_n\}$. Putting $n = m$ and $\alpha = nF_n$ in Proposition 2.1 in Lo (1986) results in (3.2). □

In general, the number of distinct x's (i.e., k) increases as the sample size n grows. Eventually we let n tend to ∞ and it is more convenient to present a functional form of Proposition 3.1 which is independent of k. To do this, we need a slightly different, yet equivalent, definition of a Dirichlet-multinomial process [Lo (1986)].

DEFINITION 3.1. For a positive integer m and a finite measure α on $(R, \mathscr{B})$, there is a point process $N(A)$, $A \in \mathscr{B}$ such that for any partition $A_1, \dots, A_k$ of R, $(N(A_1), \dots, N(A_k))$ is a Dirichlet-multinomial $(m; \alpha(A_1), \dots, \alpha(A_k))$ vector [compound multinomial in Mosimann (1962)]; $N(\cdot)$ is called a Dirichlet-multinomial point process with parameters (m, α), denoted concisely by $\angle\{N(\cdot)\} = DM(m; \alpha)$.

Routine calculation using conditional expectation [or invoking Theorem 4 in Ferguson (1973)] shows that, $\angle\{N(\cdot)\} = DM(m; \alpha)$ implies

$$(3.3) \quad E\left[\int g(x)N(dx)\right] = m\int g(x)\alpha(dx)/\alpha(R) \quad \text{if } \int |g(x)|\alpha(dx) < \infty,$$

and if $\int g^2(x)\alpha(dx) < \infty$ and $\alpha' = \alpha/\alpha(R)$,

$$
\begin{aligned}
&E\left[\int g(x)N(dx)\right]^2 \\
&\quad = m\left[\int g^2(x)\alpha'(dx)\right] \\
&\qquad + m(m-1)\left\{\alpha(R)\left[\int g(x)\alpha'(dx)\right]^2 \right. \\
&\qquad\qquad \left. + \int g^2(x)\alpha'(dx)\right\} \Big/ (\alpha(R)+1).
\end{aligned}
\tag{3.4}
$$

THEOREM 3.1.

$$\angle\{mG_m^*(\cdot)|F_n\} = DM(m; nF_n), \tag{3.5}$$

$$\lim_{m\to\infty} \angle\{G_m^*(\cdot)|F_n\} = \textit{a Dirichlet process with shape measure } nF_n. \tag{3.6}$$

PROOF. (3.5) follows from (3.1) and the Kolmogorov consistent property [Theorem 5.5 in Hoadley (1969)] of a Dirichlet-multinomial vector. Statement (3.6) follows from (3.2). □

The statement (3.5) and a result in Lo (1986), Theorem 2.1, which states that for an exchangeable prior on the sequence $\{y_1, \dots, y_N\}$ such that $\angle\{NH_N(\cdot)\} = DM(N; \alpha)$, $\angle\{m(\cdot)|F_n\} = DM(m; \alpha + nF_n)$, where $m^{-1}m(\cdot)$ is the empirical distribution function of the unobserved population variables, indicate that one can view $\angle\{mG_m^*(\cdot)|F_n\}$ as a posterior distribution of $m(\cdot)$ with respect to a "flat" Dirichlet-multinomial process prior.

REMARK 3.1. A FPBB which allows the user input of prior information can be obtained similarly. Summarize the prior information of the finite population by p initial data, say, $\{x_1^\#, \dots, x_p^\#\}$ (allow ties among the $x^\#$'s). Denote the empirical distribution function of the $x^\#$'s by $F_p^\#$. Replace urn $\{x_1, \dots, x_n\}$ by urn $\{x_1, \dots, x_n, x_1^\#, \dots, x_p^\#\}$ in Step 1 of the algorithm (2.2). Then $\angle\{mG_m^*(\cdot)|F_n\} = DM(m; pF_p^\# + nF_n)$, which can be used to approximate $DM(m; \alpha + nF_n) = \angle\{m(\cdot)|F_n\}$. [Note that in this setting, $\{x_1^\#, \dots, x_p^\#\}$ reflects the knowledge of the prior parameter α.]

Expression (3.6) states that, for a large population size (and a fixed sample size), the FPBB reduces to the Bayesian bootstrap since a Dirichlet process with a shape measure nF_n is the backbone of the Bayesian bootstrap method [Lo (1987)].

The distribution of the FPBB being known (similarly, a FPB version of Definition 3.1 and Theorem 3.1 can be stated in terms of a multivariate hypergeometric point process), we give the first two moments for the FPBB and FPB means in the case of $q = 1$. These moment expressions will be needed in

Section 5. Denote the sample variance by s^2 and the finite population correction factor (fpc) $1 - n/N$ by $1 - f$.

It follows from (3.3)–(3.5) and $H_{mn}^* = (mG_m^* + nF_n)/N$ that

$$E\left[\int x H_{mn}^*(dx) \middle| F_n\right] = n^{-1} \sum_{1 \le i \le n} x_i \tag{3.7}$$

and

$$\begin{aligned} \operatorname{Var}\left\{\int x H_{mn}^*(dx) \middle| F_n\right\} &= [(N-n)/N][(n-1)/(n+1)]s^2/n \\ &= (1-f)[(n-1)/(n+1)]s^2/n. \end{aligned} \tag{3.8}$$

In the FPB case, Bickel and Freedman (1984) and Chao and Lo (1985) noted that a β determined by (2.1) leads also to the following "right" conditional asymptotic variance for $\int x F_n^*(dx)$,

$$E\left[\int x F_n^*(dx) \middle| F_n\right] = n^{-1} \sum_{1 \le i \le n} x_i \tag{3.9}$$

and

$$\begin{aligned} \operatorname{Var}\left\{\int x F_n^*(dx) \middle| F_n\right\} &= [(N-n)/(N-1)][(n-1)/n]s^2/n \\ &= (1-f)[N/(N-1)][(n-1)/n]s^2/n \\ &\approx (1-f)[(n-1)/n]s^2/n, \quad \text{for a large } N. \end{aligned} \tag{3.10}$$

4. The first-order asymptotic equivalence of the FPBB and FPB. In this section, we show that the common limit for both $\mathcal{L}\{[n/(1-f)]^{1/2}[H_{mn}^* - F_n]|F_n\}$ and $\mathcal{L}\{[n/(1-f)]^{1/2}[F_n^* - F_n]|F_n\}$ is a Brownian bridge, subject to a change of time scale by the limit of F_n and is identical to the limit of the corresponding posterior distribution [Lo (1986)] and the sampling distribution [Bickel (1969)], respectively. The key assumption for the validity of these limit results is the convergence of the sample empirical distribution; it was the same condition imposed by Scott (1971) and Lo (1986) in order to achieve asymptotic posterior normality.

ASSUMPTION 4.1. For each t, $F_n(t) \to F(t)$, which is a distribution function.

A straightforward proof of these results can be obtained by appealing to the Markov property of the Dirichlet-multinomial process and the multivariate hypergeometric process, respectively. However, such a method requires a strenuous amount of notation to account for the jumps of the limit of F_n. In the following a simpler proof of the first limit based on an adaptation of the well-known quantile transform method in the iid case is given for $R = (-\infty, \infty)$, which we assume throughout this section. An almost identical proof of the second limit [which can also be deduced from a result in Bickel (1969)] is provided for completeness.

The quantile method relies on two lemmas. The first lemma deals with the quantile transform in the Pólya and simple random sampling schemes, and the second lemma extends the weak convergence of the uniform empirical distribution function to the Pólya and simple random sampling schemes. Denote the empirical distribution function of $\{1/n, \dots, (n-1)/n, 1\}$ by $E_n(\cdot)$. Call the urn $\{1/n, \dots, (n-1)/n, 1\}$ the standard urn (n). For any nondecreasing function K on the line, define its inverse by $K^{-1}(s) = \inf\{x\colon K(x) \ge s\}$ for $s \in (0,1)$.

Lemma 4.1. *Denote the empirical distribution function of the real numbers $z_1, \dots, z_j$ by K_j.*

(i) *$U_1, \dots, U_m$ is a Pólya sample from the standard urn (n) implies $K_n^{-1}(U_1), \dots, K_n^{-1}(U_m)$ is a Pólya sample from the urn $\{z_1, \dots, z_n\}$.*

(ii) *$U_1, \dots, U_n$ is a simple random sample from the standard urn (N) implies $K_N^{-1}(U_1), \dots, K_N^{-1}(U_n)$ is a simple random sample from the urn $\{z_1, \dots, z_N\}$.*

Proof. According to Lo (1986), page 1227, the joint distribution of a Pólya sample of size m from the urn $\{z_1, \dots, z_n\}$ is the average of the m-fold product of Dirichlet $(nK_n(\cdot))$ random probabilities; similarly, $\angle\{U_1, \dots, U_m\}$ is the average of the m-fold product of Dirichlet $(nE_n(\cdot))$ random probabilities. Note that, if G is a Dirichlet $(\alpha(\cdot))$ random probability on $[0,1]$ and H is a given nondecreasing function, the Borel function H^{-1} maps G to a Dirichlet $(\alpha(H(\cdot)))$ random probability. Hence, $\angle\{K_n^{-1}(U_1), \dots, K_n^{-1}(U_m)\}$ is the average of the m-fold product of Dirichlet $(nE_n(K_n(\cdot)))$ [$=$ Dirichlet $(nK_n(\cdot))$] random probabilities, proving (i).

The proof of (ii) is completely elementary and is omitted. □

Next, we turn to the second lemma. For any $-\infty \le a < b \le \infty$, let $D[a, b]$ be the space of "cadlag" functions defined on $[a, b]$. Equip $D[a, b]$ with the projection σ-field; metrize this space with the uniform metric. Unless otherwise specified, weak convergence of random elements taking values in $D[a, b]$ will be studied under this setting; see Pollard (1984) for details. Denote a standard Brownian bridge in $D[0,1]$ by $B(s)$, $0 \le s \le 1$.

Lemma 4.2. (i) *Let $U_1, \dots, U_m$ be a Pólya sample from the standard urn (n) and $Y_{mn}(s) = [mn/N]^{1/2}[m^{-1}\Sigma_{1 \le j \le m} I\{U_j \le s\} - E_n(s)]$ for $0 \le s \le 1$; then*

$$\angle\{Y_{mn}(\cdot)|E_n\} \to \angle\{B(\cdot)] \qquad \text{in } D[0,1] \text{ as } m, n \to \infty.$$

(ii) *Let $U_1, \dots, U_n$ be a simple random sample from the standard urn (N) and $X_{mn}(s) = [n/(1-f)]^{1/2}[n^{-1}\Sigma_{1 \le j \le n} I\{U_j \le s\} - E_N(s)]$ for $0 \le s \le 1$; then*

$$\angle\{X_{mn}(\cdot)|E_N\} \to \angle\{B(\cdot)\} \qquad \text{in } D[0,1] \text{ as } m, n \to \infty.$$

Proof. We first prove (i). Note that $\angle\{\Sigma_{1 \le j \le m} I\{U_j \le \cdot\}|E_n\}$ is a $DM(m; nE_n)$ process and $E_n(s) \to s$ for each $s \in [0,1]$. First, denote $D[0,1]$

equipped with the Skorohod metric and the Borel σ-field by $D_s[0,1]$. It is well known that the finite-dimensional distributions of Y_{mn} converge appropriately [Scott (1971)]. A computation in Lo (1986), page 1231, yields

$$E\left[(Y_{mn}(D)Y_{mn}(C))^2 | E_n\right] \leq 13E_n(D)E_n(C), \quad \text{for all } m \text{ and } n,$$

where D and C are neighboring blocks; tightness follows from the extension to Theorem 3 in Bickel and Wichura (1971). Therefore, $\angle\{Y_{mn}(\cdot)|E_n\} \to \angle\{B(\cdot)\}$ in $D_s[0,1]$. Next, the continuous mapping theorem in Pollard (1984), page 70, can be applied to the identity map χ from $D_s[0,1]$ onto $D[0,1]$ to yield $\angle\{\chi(Y_{mn})|E_n\} \to \angle\{\chi(B)\}$, proving (i).

The proof of (ii) is similar. It suffices to show that $\angle\{X_{mn}(\cdot)|E_n\} \to \angle\{B(\cdot)\}$ in $D_s[0,1]$. According to Hartley and Rao (1968), the finite-dimensional distributions of X_{mn} converge appropriately. A computation of Bickel and Wichura (1971), page 1668, yields

$$E\left[(X_{mn}(D)X_{mn}(C))^2 | E_N\right] \leq 33E_N(D)E_N(C), \quad \text{for all } N \geq 4,$$

implying that the sequence $\{X_{mn}\}$ is tight, proving (ii). □

Theorem 4.1. *Under Assumption* 4.1,

(i) $$\angle\left\{[n/(1-f)]^{1/2}[H^*_{mn}(\cdot) - F_n(\cdot)] | F_n\right\} \to \angle\{B(F(\cdot))\}$$ *in* $D[-\infty,\infty]$ *as* $m, n \to \infty$,

and

(ii) $$\angle\left\{[n/(1-f)]^{1/2}[F^*_n(\cdot) - F_n(\cdot)] | F_n\right\} \to \angle\{B(F(\cdot))\}$$ *in* $D[-\infty,\infty]$ *as* $n \to \infty$.

Proof. We prove (i) first. Denote F_∞ by F. For each $n = 1, \dots, \infty$, define the map Q_n from $D[0,1]$ to $D[-\infty,\infty]$ by $Q_n(y) = (Q_n y)(\cdot) = y(F_n(\cdot))$ for each $y \in D[0,1]$. The Q-maps are measurable with respect to the projection σ-fields; furthermore,

$$\sup_{-\infty \leq t \leq \infty} |(Q_n x)(t) - (Q_n y)(t)| \leq \sup_{0 \leq s \leq 1} |x(s) - y(s)|,$$

implying that the Q-maps are equicontinuous. Hence, Lemma 4.2(i) can be applied to yield that $\angle\{Q_n(Y_{mn})|F_n\}$ and $\angle\{Q_n(B)|F_n\}$ have the same limit [the Y's are defined in Lemma 4.2(i)]; the latter distribution equals $\angle\{B(F_n(\cdot))|F_n\}$, which converges to $\angle\{B(F(\cdot))\}$ under Assumption 4.1.

It remains to show that

$$\angle\{Q_n(Y_{mn})|F_n\} = \angle\left\{[n/(1-f)]^{1/2}[H^*_{mn}(\cdot) - F_n(\cdot)] | F_n\right\}.$$

For each t, $E_n(F_n(t)) = F_n(t)$ and $U_j \leq F_n(t)$ if and only if $F_n^{-1}(U_j) \leq t$. Hence,

$$\begin{aligned}(Q_n Y_{mn})(t) &= Y_{mn}(F_n(t)) \\ &= [mn/N]^{1/2}\left[m^{-1} \sum_{1 \leq j \leq m} I\{F_n^{-1}(U_j) \leq t\} - F_n(t)\right],\end{aligned}$$

which, according to Lemma 4.1(i), has the same distribution as

$$[mn/N]^{1/2}\left[m^{-1}\sum_{1\le j\le m} I\{y_j^*\le t\} - F_n(t)\right]$$
$$= [n/(1-f)]^{1/2}[H_{mn}^*(t) - F_n(t)].$$

The proof of (ii) is almost identical. First, suppose N is an integer multiple of n. Then $F_n = H_N^*$, which is the empirical distribution function of the FPB population, and $n\to\infty$ implies $m = N - n\to\infty$. Hence (ii) is equivalent to

$$\text{(ii}'\text{)}\qquad \angle\left\{[n/(1-f)]^{1/2}[F_n^*(\cdot) - H_N^*(\cdot)]|H_N^*\right\} \to \angle\{B(F(\cdot))\}$$
$$\text{in } D[-\infty,\infty] \text{ as } n\to\infty,$$

which can be proved by arguing verbatim, using Lemmas 4.1(ii) and 4.2(ii) instead, as in the proof of (i) [an alternative proof of (ii′) follows from applying Theorem 3.1 in Bickel (1969)]. Next, if N is not an integer multiple of n, the first argument can be applied to the populations corresponding to each of the two integer multiples of n that is closest to N; the proof can be completed by noting that two sequences of distributions converge weakly to an identical limit implies that any mixture of the two also converges weakly to the same limit. □

COROLLARY 4.1. *Under Assumption* 4.1, *F continuous implies*

$$\lim_{m,n\to\infty} P\left\{[n/(1-f)]^{1/2}\sup_t|H_{mn}^*(t) - F_n(t)| > \lambda|F_n\right\}$$
$$(4.1)\qquad = \lim_{n\to\infty} P\left\{[n/(1-f)]^{1/2}\sup_t|F_n^*(t) - F_n(t)| > \lambda|F_n\right\}$$
$$= 2\sum_{1\le j<\infty}(-1)^{j+1}\exp(-2j^2\lambda^2);$$

F has a bounded support implies

$$\lim_{m,n\to\infty}\angle\left\{[n/(1-f)]^{1/2}s^{-1}\left\{\int xH_{mn}^*(dx) - \int xF_n(dx)\right\}\middle|F_n\right\}$$
$$(4.2)\qquad = \lim_{n\to\infty}\angle\left\{[n/(1-f)]^{1/2}s^{-1}\left\{\int xF_n^*(dx) - \int xF_n(dx)\right\}\middle|F_n\right\}$$
$$= \Phi,$$

where Φ is the distribution function of a standard normal random variable.

For a continuous F, one can use (4.1) to obtain a large-sample uniform FPBB band for the population empirical distribution given by

$$(4.3)\qquad F_n \pm \lambda[(1-f)/n]^{1/2},$$

where λ is the $(1-\alpha)$-percentile point of $\sup_t|B(t)|$ defined by $2\Sigma_{1\le j<\infty}(-1)^{j+1}\exp(-2j^2\lambda^2) = \alpha$ (note that the FPB leads to the same band). For example, $\alpha = 95.3\%$ corresponds to $\lambda = 1.37$, and if the sample size is 15 and the population size is 65, the fpc $(1 - f)$ equals 0.769231; a 95.3% large-sample FPBB probability band for the finite population empirical distribution is given by $F_n \pm 0.354(1 - f)^{1/2} = F_n \pm 0.354(0.877) = F_n \pm 0.31$, which compares favorably with the exact FPBB band obtained in Example 2.1.

5. Stratification. The extension of the previous results to several strata does not present any great difficulty. The idea is to perform the FPBB and/or the FPB in each stratum independently. Suppose the population is broken into J strata and sampling is performed in each stratum independently to obtain $\mathbf{x} = \{x_{ji};\ i = 1,\dots, n_j \text{ and } j = 1,\dots, J\}$, where n_j is the sample size for the jth stratum. Let F_{n_j} be the jth stratum sample empirical distribution function, $x_{j\cdot}$ the jth stratum sample average $n_j^{-1}\Sigma_{1 \le i \le n_j} x_{ji}$, and s_j^2 the jth stratum sample variance $(n_j - 1)^{-1}\Sigma_{1 \le i \le n_j}(x_{ji} - x_{j\cdot})^2$. Similarly, given $\mathbf{x}$, for $j = 1,\dots, J$, the Pólya sample from the jth urn [Step 1 in (2.2)] is denoted by $\{y_{ji}^*;\ i = 1,\dots, m_j\}$, where m_j is the Pólya sample size for the jth urn, and given $\mathbf{x}$, sampling from different urns is assumed to be independent. Let $H_{mn_j}^*$ be the empirical distribution function of the Pólya sample in the jth urn, and $y_{j\cdot}^*$ $(= m_j^{-1}\Sigma_{1 \le i \le m_j} y_{ji}^*)$ be the average of the Pólya sample within the jth urn. Note that $n_j + m_j = N_j$, which is the population size of the jth stratum. Suppose the functional of interest is the weighted average

$$\sum_{1 \le j \le J} w_{n_j}\left\{\theta\left(H_{N_j} - F_{n_j}\right)\right\}, \tag{5.1}$$

where H_{N_j} is the jth stratum population empirical distribution function; the w_{n_j}'s are weights which may depend on the F_{n_j}'s (but not the H_{N_j}'s).

Our discussion will be restricted to the case of a linear combination of strata means, i.e., θ is a mean functional given by

$$\theta\left(H_{N_j}, F_{n_j}\right) = \int x H_{N_j}(dx) - \int x F_{n_j}(dx). \tag{5.2}$$

The next theorem states a first-order asymptotic result for the FPBB and FPB subject to stratification.

THEOREM 5.1. *Suppose that the population is numerical and bounded and that* $F_{n_j}(t) \to F_j(t)$ *for each* j *and* t. *Let* $w_{\mathbf{n}} = [\Sigma_{1 \le j \le J} w_{n_j}^2 s_j^2 (1 - f_j)/n_j]^{1/2}$.

(i) $\min\{n_1,\dots, n_J\} \to \infty$ *and* $\min\{m_1,\dots, m_J\} \to \infty$ *imply*

$$\begin{aligned} &\angle\left\{w_{\mathbf{n}}^{-1} \sum_{1 \le j \le J} w_{n_j}\left[\int x H_{mn_j}^*(dx) - \int x F_{n_j}(dx)\right]\middle|\mathbf{x}\right\} \\ &\quad = \angle\left\{w_{\mathbf{n}}^{-1} \sum_{1 \le j \le J} w_{n_j}\left(m_j/N_j\right)\left(y_{j\cdot}^* - x_{j\cdot}\right)\middle|\mathbf{x}\right\} \to \Phi; \end{aligned} \tag{5.3}$$

(ii) $\min\{n_1,\dots, n_J\} \to \infty$ *implies*

(5.4) *(5.3) holds with* $H_{mn_j}^*$ *replaced by* $F_{n_j}^*$ *(i.e.,* $y_{j\cdot}^*$ *replaced by* $x_{j\cdot}^*$*), where* $F_{n_j}^*$ *is the jth stratum FPB sample empirical distribution function and* $x_{j\cdot}^*$ *is the jth stratum FPB sample average.*

PROOF. First, note that according to (3.8),

$$\begin{aligned}(5.5)\qquad &\operatorname{Var}\Big\{\big[n_j/(1-f_j)\big]^{1/2}s_j^{-1}(m_j/N_j)(y_{j\cdot}^{*}-x_{j\cdot})\big|F_{n_j}\Big\}\\ &\quad=(n_j-1)/(n_j+1)\to 1,\quad \text{for } j=1,\dots,J,\end{aligned}$$

and that expression (4.2) in Corollary 4.1 yields for each j, as $n_j, m_j \to \infty$,

$$(5.6)\quad \angle(j, m_j, n_j) = \angle\Big\{\big[n_j/(1-f_j)\big]^{1/2}s_j^{-1}(m_j/N_j)(y_{j\cdot}^{*}-x_{j\cdot})\big|F_{n_j}\Big\} \to \Phi,$$

independently (in j).

Denote the conditional distribution defined in (5.3) by G_{mn} and the Wasserstein metric defined in Mallows (1972) by ρ. According to the Mallows inequality (1972), Lemma 3,

$$\begin{aligned}(5.7)\qquad \rho^2(G_{mn},\Phi) &\le w_{\mathbf{n}}^{-2}\Bigg[\sum_{1\le j\le J} w_{n_j}^2 s_j^2(1-f_j)/n_j\Bigg]\rho^2\big(\angle(j,m_j,n_j),\Phi\big)\\ &\le \max_{1\le j\le J}\rho^2\big(\angle(j,m_j,n_j),\Phi\big).\end{aligned}$$

By (5.5) and (5.6), the last expression goes to 0 as both n_j and $m_j \to \infty$, proving the theorem. □

The choice of $w_{n_j} = N_j/N$ in Theorem 5.1 yields a large-sample $(1-\alpha)$ FPBB or FPB interval for the population average in a finite population subject to stratification given by

$$(5.8)\qquad \sum_{1\le j\le J} N_j N^{-1} x_{j\cdot} \pm \lambda w_{\mathbf{n}},$$

where λ is the $(1-\alpha/2)$-percentile point of Φ. This interval agrees with the classical one based on the stratified simple random sampling [Cochran (1977), page 95]. For the case of one stratum, i.e., $J = 1$, (5.8) reduces to the classical large-sample $(1-\alpha)$ interval estimate for a finite population mean based on the simple random sampling [Cochran (1977), page 27].

REMARK 5.1. Bickel and Freedman (1984) proved a version of Theorem 5.1(ii) under a weaker Lindeberg-type condition on the population variables; moreover, their theory also allows for the possibility of "many small strata," whereas our theory treats the case of a fixed number of strata only. However, the "in-probability" conclusion of their result dictates that their result does not imply Theorem 5.1(ii).

REFERENCES

BICKEL, P. J. (1969). A distribution free version of the Smirnov two sample test in the p-variate case. *Ann. Math. Statist.* **40** 1–23.

BICKEL, P. J. and FREEDMAN, D. A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *Ann. Statist.* **12** 470–482.

BICKEL, P. J. and WICHURA, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42** 1656–1670.

CHAO, M.-T. and LO, S.-H. (1985). A bootstrap method for finite population. *Sankhyā Ser. A* **47** 399–405.

COCHRAN, W. G. (1977). *Sampling Techniques*, 3rd ed. Wiley, New York.

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26.

EFRON, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. SIAM, Philadelphia.

ERICSON, W. A. (1969). Subjective Bayesian models in sampling finite populations (with discussion). *J. Roy. Statist. Soc. Ser. B* **31** 195–233.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York.

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230.

GROSS, S. (1980). Median estimation in sample surveys. Presented at 1980 Amer. Statist. Assoc. meeting.

HARTLEY, H. O. and RAO, J. N. K. (1968). A new estimation theory for sample surveys. *Biometrika* **55** 547–557.

HOADLEY, B. (1969). The compound multinomial distribution and Bayesian analysis of categorical data from finite populations. *J. Amer. Statist. Assoc.* **64** 216–229.

LO, A. Y. (1986). Bayesian statistical inference for sampling a finite population. *Ann. Statist.* **14** 1226–1233.

LO, A. Y. (1987). A large sample study for the Bayesian bootstrap. *Ann. Statist.* **15** 360–375.

MALLOWS, C. L. (1972). A note on asymptotic joint normality. *Ann. Math. Statist.* **43** 508–515.

MOSIMANN, J. E. (1962). On the compound multinomial distribution, the multivariate β-distribution, and correlations among proportions. *Biometrika* **49** 65–77.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

RUBIN, D. B. (1981). The Bayesian bootstrap. *Ann. Statist.* **9** 130–134.

SCOTT, A. (1971). Large-sample posterior distributions for finite populations. *Ann. Math. Statist.* **42** 1113–1117.

DEPARTMENT OF STATISTICS
STATE UNIVERSITY OF NEW YORK
AT BUFFALO
AMHERST, NEW YORK 14261