The Annals of Statistics
1993, Vol. 21, No. 4, 2138–2148

A BAYESIAN METHOD FOR WEIGHTED SAMPLING[1]

BY ALBERT Y. LO

SUNY, Buffalo

Bayesian statistical inference for sampling from weighted distribution models is studied. Small-sample Bayesian bootstrap clone (BBC) approximations to the posterior distribution are discussed. A second-order property for the BBC in unweighted i.i.d. sampling is given. A consequence is that BBC approximations to a posterior distribution of the mean and to the sampling distribution of the sample average, can be made asymptotically accurate by a proper choice of the random variables that generate the clones. It also follows from this result that in weighted sampling models, BBC approximations to a posterior distribution of the reciprocal of the weighted mean are asymptotically accurate; BBC approximations to a sampling distribution of the reciprocal of the empirical weighted mean are also asymptotically accurate.

1. Introduction. The weighted distribution model is one where the probability of including an observation in the sample is proportional to a weighting function. This model can be defined by

$$X_1, \dots, X_n | G \quad \text{are i.i.d. from a distribution } F(\cdot | G), \tag{1.1}$$

where

$$F(ds|G) = \omega(s)G(ds) \Big/ \int \omega(s)G(ds), \tag{1.2}$$

$\omega(s)$ is a known weighting function with $0 < \omega(s) < \infty$ on the support of G; G is the unknown parameter. This model arises naturally in several applied areas: in sampling fiber length [Palmer (1948) and Cox (1969)] where the X_i's are univariate, in cell sampling [Takahashi (1966) and Zelen (1974)] where the X_i's (and the s's) are bivariate, and in aerial survey in traffic and ecology problems [Brown (1972) and Cook and Martin (1974)]. Rao (1965) gives a unified formulation for this model, and Patil and Rao (1977) is an excellent source of references.

Cox (1969) proposes an estimate for a freely varying G for a length-biased model which corresponds to $\omega(s) = s$; he also discusses statistical inference about G based on this estimate. Recently, Vardi (1985) and Gill, Vardi and Wellner (1988) study the k-sample problem of (1.1) using the MLE method.

In this paper, we provide a Bayesian solution to the model (1.1) as an alternative to Cox's classical approach. The Bayesian method is important

---

Received November 1989; revised November 1992.

[1]Supported in part by NSF Grants DMS-89-06382 and DMS-91-06460.

AMS 1991 *subject classifications*. Primary 62G09; secondary 62G20, 62G99.

Key words and phrases. Weighted distribution models, weighted gamma process priors, bootstrap approximations, Bayesian bootstrap clone approximations, asymptotic accuracy.

since it provides an alternative approach to statistical thinking, and the posterior mean as an estimate allows user input of prior information and thus is useful for Bayesians and frequentists alike. Furthermore, it is well known that a Bayesian solution is of great importance from a decision theorist's viewpoint.

Section 2 shows that normalized weighted gamma process priors for G are conjugate priors. Section 3 develops a large-sample approximation to a posterior distribution of G. It is shown that the posterior distribution of G can be approximated by a Gaussian process discussed by Cox (1969).

Section 4 discusses Bayesian bootstrap clone approximations [Lo (1991)] in the weighted sampling models. The accuracy of the BBC approximation for the reciprocal of the weighted mean is discussed based on techniques developed in Section 5. Section 5 deals with some accuracy questions of BBC approximations: a class of BBC, including Rubin's (1981) Bayesian bootstrap, is found to provide an asymptotically accurate approximation to the posterior distribution of the population mean with respect to Dirichlet type priors; another class of BBC's is found to provide an asymptotically accurate approximation to the sampling distribution of the sample average.

2. The prior to posterior analysis. The key to the Bayesian solution for these types of problems is the choice of a nontrivial, yet manageable, prior; a conjugate one is even more desirable since it allows for an easy description of the posterior distribution. For sampling from a weighted distribution, normalized weighted gamma processes are found to be conjugate priors. Such priors are defined naturally through weighted gamma processes [see Dykstra and Laud (1981), Lo (1982) and Lo and Weng (1989)] just as Dirichlet processes [Ferguson (1973)] can be defined in terms of gamma processes. Let $\gamma(\cdot)$ be a gamma process with shape $\alpha(\cdot)$, [i.e., $\gamma(\cdot)$ is an "independent-increment" process, and for each t, $\gamma(t)$ is a gamma $(\alpha(t); 1)$ random variable]. For a fixed $\beta \geq 0$, define $v(t) = \int I_{\{s \leq t\}} \beta(s) \gamma(ds)$ [the inequality is coordinate-wise if s is k-variate]. Then $v(\cdot)$ is called a weighted gamma$(\alpha; \beta)$ process and the random distribution function $v(\cdot)/v(\infty)$ a normalized weighted gamma process (with shape α and weight β), that is, a $D(\alpha; \beta)$ process. Note that a $D(\alpha; 1)$ process is a Dirichlet process with shape measure α. Let $\mathbf{X} = (X_1, \ldots, X_n)$, and let $\mathbf{x} = (x_1, \ldots, x_n)$.

THEOREM 2.1. *For the model* (1.1), $\mathscr{L}\{G(\cdot)\} = D(\alpha; 1/\omega)$ *implies*

$$\mathscr{L}\{G(\cdot)|\mathbf{X} = \mathbf{x}\} = D\Big(\alpha + \sum_{1 \leq i \leq n} \delta_{x_i}; 1/\omega\Big).$$

PROOF. Note that (1.2) and a routine computation shows that $F(\cdot)$ is a $D(\alpha; 1)$ process, that is, a Dirichlet process with shape measure α [Ferguson (1973)]. It follows from (1.1) and a result of Ferguson that the posterior distribution of $F(\cdot)$ given $\mathbf{X} = \mathbf{x}$ is a $D(\alpha + \sum_{1 \leq i \leq n} \delta_{x_i}; 1)$ process. Since $G(t) = \int I_{\{s \leq t\}} \omega(s)^{-1} F(ds) / \int \omega(s)^{-1} F(ds)$ follows from (1.2), another computation shows that $\mathscr{L}\{G(\cdot)|\mathbf{X} = \mathbf{x}\} = D(\alpha + \sum_{1 \leq i \leq n} \delta_{x_i}; 1/\omega)$. $\square$

The conclusion of Theorem 2.1 suggests the following construction of a "posterior random variable;" which simplifies the study of posterior distributions. Let γ be a gamma(α; 1) process, and let $\{Z_1, \dots, Z_n, \dots\}$ be a sequence of i.i.d. standard exponential random variables which is also independent of γ and the X_i's. For each $\mathbf{X} = \mathbf{x}$, define

$$G\gamma_n(t) = \frac{\int I_{\{s \le t\}} \omega(s)^{-1} \gamma(ds) + \sum_{1 \le i \le n} \omega(x_i)^{-1} Z_i I_{\{x_i \le t\}}}{\int \omega(s)^{-1} \gamma(ds) + \sum_{1 \le i \le n} \omega(x_i)^{-1} Z_i}. \tag{2.2}$$

Note that $\mathscr{L}\{G\gamma_n(\cdot)|\mathbf{X} = \mathbf{x}\} = \mathscr{L}\{G(\cdot)|\mathbf{X} = \mathbf{x}\}$ and hence

$$\begin{aligned} E\left\{\left[\int \omega(s) G(ds)\right]^{-1} \middle| \mathbf{x}\right\} &= E\left\{\left[\int \omega(s) G\gamma_n(ds)\right]^{-1} \middle| \mathbf{x}\right\} \\ &= \{\alpha(\infty)/[\alpha(\infty) + n]\}\left[\int \omega(s)^{-1} \alpha(ds)/\alpha(\infty)\right] \\ &\quad + \{n/[\alpha(\infty) + n]\}\left[n^{-1} \sum_i \omega(x_i)^{-1}\right]. \end{aligned} \tag{2.3}$$

The posterior mean (2.3) is a weighted average of the prior estimate $\int \omega(s)^{-1}\alpha(ds)/\alpha(\infty)$ just as the Cox estimate, $n^{-1}\sum_i \omega(x_i)^{-1}$, of $[\int \omega(s)G(ds)]^{-1}$.

Putting $\alpha = 0$ in $G\gamma_n(\cdot)$ results in

$$G_n(t) = \frac{\sum_{1 \le i \le n} \omega(x_i)^{-1} Z_i I_{\{x_i \le t\}}}{\sum_{1 \le i \le n} \omega(x_i)^{-1} Z_i} \tag{2.4}$$

and $\mathscr{L}\{G_n(\cdot)|\mathbf{X} = \mathbf{x}\}$ is the posterior distribution with respect to a "flat" prior. The Cox estimate of $[\int \omega(s)G(ds)]^{-1}$ is the posterior mean of $[\int \omega(s)G(ds)]^{-1}$ with respect to the "flat" prior,

$$E\left\{\left[\int \omega(s) G_n(ds)\right]^{-1} \middle| \mathbf{x}\right\} = n^{-1} \sum_{1 \le i \le n} \omega(x_i)^{-1}. \tag{2.5}$$

Cox (1969) also proposes

$$\hat{C}(t) = \sum_{1 \le i \le n} \omega(x_i)^{-1} I_{\{x_i \le t\}} \Big/ \sum_{1 \le i \le n} \omega(x_i)^{-1} \tag{2.6}$$

as an estimate of $G(t)$. The posterior mean of $G(t)$ with respect to the "flat" prior does not simplify to $\hat{C}(t)$.

3. A large-sample theory for posterior distributions. The posterior distribution $\mathscr{L}\{\theta(G(\cdot))|\mathbf{x}\}$ is the Bayesian solution to the problem of estimating $\theta(G(\cdot))$. Unfortunately, a simple description of this problem is not yet available. This section provides a large-sample approximation to the posterior distribution of $G(\cdot)$. Here the approximation is stated in terms of a functional central limit theorem. Random functions are regarded as elements of the $D[-\infty, \infty]^k$ space, which is equipped with the uniform metric and the projection

σ-field [see Pollard (1984)]; that is, the observations are values in $[-\infty, \infty]^k$. Let G_0 be the "true" parameter G. Define the function $\omega^{-1}G_0(t)$ by

$$\omega^{-1}G_0(t) = \int I_{\{s \le t\}}\omega(s)^{-1}G_0(ds) \quad \text{for each } t \text{ in } [-\infty, \infty]^k \tag{3.1}$$

and denote a standard Brownian motion process on $[0, \infty)$ by $\{W(s)\colon 0 \le s < \infty\}$.

THEOREM 3.1. *If* $\int \omega(s)G_0(ds)$ *and* $\int \omega(s)^{-1}G_0(ds)$ *are both finite, then*

(i)
$$\mathscr{L}\left\{n^{1/2}\left[G_n(\cdot) - \hat{C}(\cdot)\right]\middle|\mathbf{x}\right\} \to \left[\int \omega(s)G_0(ds)\right]^{1/2} \mathscr{L}\{W(\omega^{-1}G_0(\cdot)) - G_0(\cdot)W(\omega^{-1}G_0(\infty))\} \quad a.s.\ F(\cdot|G_0);$$

(ii) *if in addition,* $\int \omega(s)^{-1}\alpha(ds)$ *is finite,* (i) *remains valid with* $G(\cdot)$ *replacing* $G_n(\cdot)$.

PROOF. Note that $G_n(t) = Y_n(t)/Y_n(\infty)$ where

$$Y_n(t) = \sum_{1 \le i \le n} \omega(x_i)^{-1} Z_i I_{\{x_1 \le t\}} \quad \text{for all } t \text{ in } [-\infty, \infty]^k.$$

Next,

$$\begin{aligned} &n^{1/2}\{Y_n(t)/Y_n(\infty) - E[Y_n(t)|\mathbf{x}]/E[Y_n(\infty)|\mathbf{x}]\} \\ &\quad = [n/Y_n(\infty)]n^{-1/2}\{Y_n(t) - E[Y_n(t)|\mathbf{x}]\} \\ &\qquad - [n/Y_n(\infty)](E[Y_n(t)|\mathbf{x}]/E[Y_n(\infty)|\mathbf{x}]) \\ &\qquad \times n^{-1/2}\{Y_n(\infty) - E[Y_n(\infty)|\mathbf{x}]\}. \end{aligned} \tag{3.2}$$

It suffices to show that a.s. $F(\cdot|G_0)$,

$$\mathscr{L}\{n^{-1/2}\{Y_n(\cdot) - E[Y_n(\cdot)|\mathbf{x}]|\mathbf{x}\} \to \mathscr{L}\left\{W\left(\omega^{-1}G_0(\cdot)\middle/\int \omega(s)G_0(ds)\right)\right\}; \tag{3.3}$$

the proof is then completed by applying the continuous mapping theorem. Assume that G_0 is continuous. Note that (conditional on $\mathbf{X} = \mathbf{x}$) the $Y_n(\cdot)$ is an "independent increment" process. Hence, the finite dimensional distribution convergence is equivalent to the following: for each t in $[-\infty, \infty]^k$, $\mathscr{L}\{n^{-1/2}\{Y_n(t) - E[Y_n(t)|\mathbf{x}]\}|\mathbf{x}\}$ has an appropriate normal limit a.s. $F(\cdot|G_0)$. The last statement follows from arguments similar to that of Theorem 4.1 in Lo (1987). Next, we turn to tightness. Tightness [a.s. $F(\cdot|G_0)$] follows from a result in Bickel and Wichura (1971) and the strong law of large numbers. An argument in Lo (1993) can be applied to extend the limit result to a discontinuous G_0.

To prove (ii), it suffices to note that

$$\sup_t n|G_n(t) - G\gamma_n(t)| \le 2\int \omega(s)^{-1}\gamma(ds) \Big/ \Big[n^{-1} \sum_{1\le i\le n} \omega(x_i)^{-1} Z_i \Big], \tag{3.4}$$

implying that, a.s. $F(\cdot|G_0)$,

$$\limsup_n \sup_t n|G_n(t) - G\gamma_n(t)| \le 2\int \omega(s)^{-1}\gamma(ds) \int \omega(s) G_0(ds) \quad \text{a.s. } P\{\cdot|\mathbf{x}\}.$$

□

If the sample is unweighted, that is, $\omega(s) = 1$, the limit in Theorem 3.1 reduces to the usual Brownian bridge $B(G_0(\cdot)) = W(G_0(\cdot)) - G_0(\cdot)W(1)$; see Lo (1987).

Next, we turn to the study of the unknown mean of G. Let $\mathscr{L}\{Y\} = G_0$, and define

$$\begin{aligned} \kappa^2[G_0] = [E\omega(Y)]\big\{E[Y^2/\omega(Y)] - 2(EY)E[Y/\omega(Y)] \\ + (EY)^2 E[1/\omega(Y)]\big\}. \end{aligned} \tag{3.5}$$

PROPOSITION 3.2. *Assume that* $\int s^k \omega(s)^{-1} G_0(ds)$ *is finite for* $k = 0, 1, 2$.

(i) $\mathscr{L}\Big\{n^{1/2}\Big[\int sG_n(ds) - \int s\hat{C}(ds)\Big]\Big|\mathbf{x}\Big\} \to N(0, \kappa^2[G_0])$ *a.s.* $F(\cdot|G_0)$;

(ii) *if, in addition* $\int s^k \omega(s)^{-1}\alpha(ds)$ *is finite for* $k = 0, 1$, *(i) holds with* G_n *replaced by* G.

PROOF. The proof of (i) essentially follows the finite dimensional convergence arguments in Theorem 3.1. To prove (ii), use an inequality similar to (3.4) to get

$$\begin{aligned} &\limsup_n n\left|\int sG\gamma_n(ds) - \int sG_n(ds)\right| \\ &\quad \le E\omega(Y)\left\{\int s\omega(s)^{-1}\gamma(ds) + E(Y)\int \omega(s)^{-1}\gamma(ds)\right\}. \end{aligned} \tag{3.6}$$

□

A $(1-\alpha)$ posterior interval estimate for $\int sG_0(ds)$ is then

$$\int s\hat{C}(ds) \pm z_{\alpha/2}\kappa[\hat{C}]n^{-1/2}, \tag{3.7}$$

where $z_{\alpha/2}$ is the $1 - \alpha/2$ percentile point of a standard normal; $\kappa^2[\hat{C}]$ is an estimate of $\kappa^2[G_0]$ obtained by substituting G_0 by $\hat{C}$.

In the length-biased model, $\kappa^2[G_0]$ reduces to $(EY)^2\{(EY)E(1/Y) - 1\}$, which is the frequentist asymptotic variance derived by Cox [see (5.4) in Cox (1969)]. If the sample is unweighted, that is, $\omega(s) = 1$, then $\kappa^2[G_0]$ reduces to the variance of G_0.

REMARK 3.1. (i) A referee pointed out that (3.3) can also be derived from a result of Alexander (1987).

(ii) One can also argue as in the proof of Theorem 3.1(i) to obtain a frequentist functional central limit theorem for Cox's estimate $\hat{C}(\cdot)$ [see Cox (1969), page 517]. A frequentist analogy to Proposition 3.2 appeared in Cox [(1969), page 516].

4. BBC approximations for weighted sampling. Rubin's (1981) Bayesian bootstrap, adapted to the weighted sampling models, suggests the following: given the data $\mathbf{x}$, the posterior distribution $\mathscr{L}\{\theta(G - \hat{C})|\mathbf{x}\}$ can be approximated by $\mathscr{L}\{\theta(G_n - \hat{C})|\mathbf{x}\}$; the latter can in turn be approximated by simulating i.i.d. exponential Z_i's in the construction of G_n. Bayesian bootstrap clone approximations [Lo (1991)] use i.i.d. positive Z_i's to play the role of the exponentials: Suppose the data $\mathbf{x}$ are given. Let $\rho = \rho(Z_1) = E(Z_1)/\sigma(Z_1)$.

(4.1) Construct a $G_n(\cdot)$ based on simulating i.i.d. random variables $Z_1, \dots, Z_n$,

thereby producing a realization of $\theta^* = \theta(\rho \times (G_n - \hat{C}))$. Repeat this simulation B times to obtain $G_{1n}(\cdot), \dots, G_{Bn}(\cdot)$, and the corresponding $\theta_1^*\ \theta_2^*, \dots, \theta_B^*$. The empirical distribution of $\theta_1^*, \theta_2^*, \dots, \theta_B^*$ approximates the posterior distribution $\mathscr{L}\{\theta(G - \hat{C})|\mathbf{x}\}$ with respect to smooth priors.

Next is a first-order theory for BBC approximations. The proof of this result is almost identical to that of Theorem 3.1 and is omitted.

THEOREM 4.1. *If EZ_1^2, $\int\omega(s)G_0(ds)$, and $\int\omega(s)^{-1}G_0(ds)$ are finite, then*

(i)
$$\mathscr{L}\left\{\rho n^{1/2}\left[G_n(\cdot) - \hat{C}(\cdot)\right]|\mathbf{x}\right\} \to \left[E\omega(Y)\right]^{1/2}\mathscr{L}\left\{W\left(\omega^{-1}G_0(\cdot)\right) - G_0(\cdot)W\left(\omega^{-1}G_0(b)\right)\right\} \quad a.s.\ F(\cdot|G_0);$$

(ii) *if $\int\omega(s)^{-1}\alpha(ds)$ is also finite,* (i) *remains valid with* $G(\cdot)$ *replacing* $G_n(\cdot)$.

While Theorem 4.1 states that the BBC is asymptotically correct for both Bayesians and frequentists, a more interesting question is whether the BBC are asymptotically accurate. An answer to this question is available if the parameter of interest is the reciprocal of the weighted mean $(\int\omega(s)G(ds))^{-1}$. According to (1.2),

$$\left(\int\omega(s)G(ds)\right)^{-1} = \int\omega(x)^{-1}F(dx), \tag{4.2}$$

which is a linear functional of the unknown distribution F. Hence the problem reduces to that of bootstrapping a linear functional of the population distribution F, given a sample from F. If the Z_i's are exponentials, Weng's (1989) result on the accuracy of Bayesian bootstrap mean (extended to the case of a linear functional) applies. The next section discusses the accuracy question when using nonexponential Z_i's.

5. Accurate BBC approximations for the mean of F. This section discusses the asymptotic accuracy of the BBC approximations in the case of the mean functional (or a linear functional of F) under i.i.d. unweighted sampling. A BBC cumulative distribution of the mean is

$$F^*(y|\mathbf{x}) = P\left\{\rho n^{1/2}\left(\sum_{1\le i\le n} x_i Z_i/S_n - \bar{x}\right)\Big/\hat{\sigma} \le y\mathbf{x}\right\}, \tag{5.1}$$

where $\bar{x}$ and $\hat{\sigma}$ are the mean and the standard deviation of the sample empirical distribution $\hat{F}$, respectively. (Define $F^*(y|\mathbf{x}) = 0$ if $Z_1 + \cdots + Z_n = 0$). The BBC suggests the use of $F^*(y|\mathbf{x})$ as the basis of approximating the posterior distribution of the unknown mean and the sampling distribution of the sample average. The following asymptotic expansion is critical for assessing the asymptotic accuracy of these approximations. Let

$$\tau(H) = \tau(Y) = E\{[Y - E(Y)]/\sigma(Y)\}^3 \tag{5.2}$$

be the coefficient of skewness of a random variable Y having a distribution H and a finite third moment. Let Φ and ϕ be the $N(0,1)$ distribution and density, respectively.

THEOREM 5.1. *Assume that the product Z_1X_1 is nonlattice. Then $\int |x^3| F_0(dx) < \infty$ implies, for each y,*

$$F^*(y|\mathbf{x}) = \Phi(y) + 6^{-1}n^{-1/2}\tau(Z_1)\tau(F_0)(1-y^2)\phi(y) + o(n^{-1/2}) \quad a.s.\ F_0. \tag{5.3}$$

PROOF. Let

$$\xi_n(y) = n^{1/2}y/\left[n + (y/\rho)^2\right]^{1/2}. \tag{5.4}$$

Rearrange terms in (5.1) to get

$$F^*(y|\mathbf{x}) = P\Big\{\left[\sum (Z_i - EZ_i)\left((x_i - \bar{x})/\hat{\sigma} - n^{-1/2}(y/\rho)\right)\right] \times\left(\sigma(Z_1)\left[n + (y/\rho)^2\right]^{1/2}\right)^{-1} \le \xi_n(y)\Big\}, \tag{5.5}$$

which is the distribution of a sum of independent random variables (in a triangular array setting.) The Edgeworth expansion for independent random variables [Petrov (1975)] can be adapted to apply in this case [see Lo (1992) for details] yielding, for each y,

$$F^*(y|\mathbf{x}) = \Phi(\xi_n(y)) + 6^{-1}\tau(Z_1)(C_n/B_n)^3 \times\left(1 - \xi_n(y)^2\right)\phi(\xi_n(y)) + o(n^{-1/2}), \tag{5.6}$$

where

$$B_n^2 = \sum_{1\le i\le n}\left[(x_i - \bar{x})/\hat{\sigma} + n^{-1/2}(y/\rho)\right]^2 = n + (y/\rho)^2,$$

and

$$\begin{aligned} C_n^3 &= \sum_{1\le i\le n}\left[(x_i - \bar{x})/\hat{\sigma} + n^{-1/2}(y/\rho)\right]^3 \\ &= \hat{\sigma}^{-3}\sum_{1\le i\le n}(x_i - \bar{x})^3 + 3n^{1/2}(y/\rho)^2 + n^{-1/2}(y/\rho)^3 \\ &= n\tau(\hat{F}) + 3n^{1/2}(y/\rho)^2 + n^{-1/2}(y/\rho)^3. \end{aligned}$$

Elementary analysis [van Zwet (1979) and Weng (1988)] shows that (uniformly in y)

$$\begin{aligned} \Phi(\xi_n(y)) &= \Phi(y) + O(n^{-1}), \\ \left(1 - \xi_n(y)^2\right)\phi(\xi_n(y)) &= (1 - y^2)\phi(y) + O(n^{-1}). \end{aligned} \tag{5.7}$$

The proof is completed by noting that $(C_n/B_n)^3 = n^{-1/2}\tau(F_0) + o(n^{-1/2})$ a.s. F_0. □

The expansion in Theorem 5.1 is a "pointwise" Edgeworth expansion in the sense that $o(n^{-1/2})$ there depends on y. The asymptotic accuracy for bootstraps will be discussed based on this pointwise expansion. [We conjecture that the error $o(n^{-1/2})$ in (5.3) is valid uniformly in y, which is perhaps more in line with the usual Edgeworth expansion technology in this area; see van Zwet (1979) and Weng (1989).]

The coefficient of skewness of the BBC variable Z_1 [i.e., $\tau(Z_1)$] appears as a multiplier in the Edgeworth expansion of $F^*(y|\mathbf{x})$. This scale effect provides the flexibility needed to tailor-make the Z_1 so that the resulting BBC approximation is asymptotically accurate for Bayesians or frequentists.

DEFINITION 5.1. Call a BBC approximation based on $\rho(Z_1)$ and $\tau(Z_1)$ a BBC $(\rho; \tau)$ approximation.

Suppose we intend to approximate a posterior distribution of the population mean $\int xF(dx)$. If the prior distribution is the Dirichlet process prior, or a Dirichlet vector for categorical data models, the posterior distribution of the standardized population mean, $F_{n,\alpha}(y)$, admits a one-term Edgeworth expansion (uniformly in y),

$$(5.8)\quad \begin{aligned} F_{n,\alpha}(y) &= P\left\{n^{1/2}\left(\int sF(ds) - \hat{\mu}_\alpha\right)\Big/\hat{s}_\alpha \le y\mathbf{x}\Big|\right\} \\ &= \Phi(y) + 3^{-1}n^{-1/2}\tau(F_0)(1-y^2)\phi(y) + o(n^{-1/2}) \quad \text{a.s. } F_0, \end{aligned}$$

where $\hat{\mu}_\alpha$ is the posterior mean of $\int sF(ds)$, and $n^{-1}\hat{s}_\alpha^2$ is the posterior variance of $\int sF(ds)$; see Weng (1989). The one-term (pointwise) Edgeworth expansions for $F_{n,\alpha}(y)$ and $F^*(y|\mathbf{x})$ from a BBC (ρ; 2) are identical. Hence, for each y,

$$(5.9)\qquad F_{n,\alpha}(y) = F^*(y|\mathbf{x}) + o(n^{-1/2}) \quad \text{a.s. } F_0.$$

Comparing (5.8) and (5.9), we conclude that BBC (ρ; 2) (for any $\rho > 0$) approximations to the posterior distribution $F_{n,\alpha}(y)$ are more accurate than the standard normal approximation (5.8). Rubin's (1981) classic Bayesian bootstrap is a BBC (1; 2), and hence is accurate; this was previously proved by Weng (1989).

Next, we turn to BBC approximations to the sampling distribution of the sample average. If X_1 is nonlattice (and has a finite third moment), then uniformly in y,

$$(5.10)\quad \begin{aligned} F_n(y|F_0) &= P\left\{n^{1/2}(\bar{X} - m_0)\big/\sigma_0 \le y|F_0\right\} \\ &= \Phi(y) + 6^{-1}n^{-1/2}\tau(F_0)(1-y^2)\phi(y) + o(n^{-1/2}); \end{aligned}$$

see Feller (1971). A $F^*(y|\mathbf{x})$ corresponding to a BBC (ρ; 1) (for any $\rho > 0$) also has expansion (5.10), implying $F_n(y|F_0) = F^*(y|\mathbf{x}) + o(n^{-1/2})$ a.s. F_0. That is, BBC (ρ; 1) approximations are more accurate than the standard normal approximation (5.10).

Efron [(1982), Section 9.3] points out that Hartigan's (1969) random subsampling (RS) plan is a BBC with Bernoulli ($p = 1/2$) Z_i's. Hartigan's RS is a BBC (1; 1) which provides asymptotically accurate approximation to $F_n(y|F_0)$ (for a nonlattice X_1). Weng (1988) shows that the use of a gamma (4; 1) as Z_1 results in an accurate approximation to $F_n(y|F_0)$; independently, Tu and Zheng (1987) suggest the use of a gamma (4; 2) as Z_1. Both $\mathscr{L}\{Z_1\}$ = gamma(4; 1) and $\mathscr{L}\{Z_1\}$ = gamma(4; 2) correspond to a BBC(2; 1).

On the other hand, Efron's bootstrap suggests approximating $F_n(y|F_0)$ by

$$(5.11)\qquad E^*(y|\mathbf{x}) = P\left\{n^{1/2}\left(n^{-1}\sum_{1\le i\le n} X_i^* - \bar{x}\right)\Big/\hat{\sigma} \le y\mathbf{x}\Big|\right\}.$$

Here $\{X_1^*, X_2^*, \ldots, X_n^*\}$ is an i.i.d. resample from the data $\{x_1, \ldots, x_n\}$. Singh (1981) shows that, if X_1 is nonlattice (uniformly in y),

$$(5.12)\quad E^*(y|\mathbf{x}) = \Phi(y) + 6^{-1}n^{-1/2}\tau(F_0)(1-y^2)\phi(y) + o(n^{-1/2}).$$

It follows then $F_n(y|F_0) = E^*(y|\mathbf{x}) + o(n^{-1/2})$ a.s. F_0, and the bootstrap approximation to $F_n(y|F_0)$ ties with BBC (ρ; 1) approximations.

TABLE 1

	Posterior distribution	Sampling distribution
$N(0,1)$ approx.	inaccurate	inaccurate
Efron's B	inaccurate	accurate
BBC $(\rho; 1)$	inaccurate	accurate
Hartigan's RS	inaccurate	accurate
BBC $(\rho; 2)$	accurate	inaccurate
Rubin's BB	accurate	inaccurate

Table 1 summarizes the asymptotic accuracy of Efron's B, Rubin's BB, Hartigan's RS and the BBC $(\rho; \tau)$ approximations for linear functionals of F.

Acknowledgments. The author is indebted to V. V. Sazonov, Michael Woodroofe and a referee for helpful comments and suggestions.

REFERENCES

ALEXANDER, K. S. (1987). Central limit theorems for stochastic processes under random entropy conditions. *Probab. Theory Related Fields* **75** 351–378.

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

BROWN, M. (1972). Low density traffic streams. *Adv. in Appl. Probab.* **4** 177–192.

COOK, R. D. and MARTIN, F. B. (1974). A model for quadrat sampling with "visibility bias." *J. Amer. Statist. Assoc.* **69** 345–349.

COX, D. R. (1969). Some sampling problems in technology. In *New Developments in Survey Sampling* (N. L. Johnson and H. Smith, Jr., ed.) 606-627. Wiley, New York.

DYKSTRA, R. L. and LAUD, P. (1981). A Bayesian nonparametric approach to reliability. *Ann. Statist.* **9** 356–367.

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.

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.

GILL, R. D., VARDI, Y. and WELLNER, J. N. (1988). Large sample theory of empirical distributions in biased sampling models. *Ann. Statist.* **16** 1069–1112.

HARTIGAN, J. A. (1969). Using subsample values as typical values. *J. Amer. Statist. Assoc.* **64** 1303–1317.

LO, A. Y. (1982). Bayesian nonparametric statistical inference for Poisson point process. *Z. Wahrsch. Verw. Gebiete* **59** 55–66.

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

LO, A. Y. (1991). Bayesian bootstrap clones and a biometry function. *Sankhyā Ser. A* **53** 320–333.

LO, A. Y. (1992). A second-order property of the Bayesian bootstrap clone approximations. Statistics research report, SUNY, Buffalo.

LO, A. Y. (1993). A Bayesian bootstrap for censored data. *Ann. Statist.* **21** 100–123.

LO, A. Y. and WENG, C. S. (1989). On a class of Bayesian nonparametric estimates: II. Hazard rate estimates. *Ann. Inst. Statist. Math.* **41** 227–245.

PALMER, R. C. (1948). The dye sampling method of measuring fibre length distribution. *Journal of the Textile Institute* **39** 8–22.

PATIL, G. P. and RAO, C. R. (1977). The weighted distributions: a survey of their applications. In *Applications of Statistics* (P. R. Krishnaiah, ed.) 383–405. North-Holland, Amsterdam.

PETROV, V. V. (1975). *Sums of Independent Random Variables*. Springer, New York.

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

RAO, C. R. (1965). On discrete distributions arising out of methods of ascertainment. *Sankhyā Ser. A* **27** 311–324.

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

SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1196.

TASKAHASHI, M. (1966). Theoretical basis for cell cycle analysis. I. Labeled mitosis wave method. *Journal of Theoretical Biology* **13** 202–211.

TU, D. S. and ZHENG, Z. G. (1987). The Edgeworth expansion for the random weighting method. *Chinese J. Appl. Probab. Statist.* **3** 340–347.

VARDI, Y. (1985). Empirical distributions in selection bias models. *Ann. Statist.* **13** 178–203.

VAN ZWET, W. R. (1979). The Edgeworth expansion for linear combinations of uniform order statistics. In *Proceedings of the Second Prague Symposium on Asymptotic Statistics* (P. Mandl and M. Hušková, eds.) 93–101. North-Holland, Amsterdam.

WENG, C. S. (1988). Ph.D. dissertation, Dept. Statistics, SUNY, Buffalo.

WENG, C. S. (1989). A second-order property of the Bayesian bootstrap mean. *Ann. Statist.* **17** 705–710.

ZELEN, M. (1974). Problems in cell kinetics and the early detection of disease. In *Reliability and Biometry* 701–726. SIAM, Philadelphia.

DEPARTMENT OF STATISTICS
STATE UNIVERSITY OF NEW YORK
SCHOOL OF MEDICINE AND BIOMEDICAL SCIENCES
249 FARBER HALL
BUFFALO, NEW YORK 14214