The Annals of Statistics
1989, Vol. 17, No. 2, 705–710

ON A SECOND-ORDER ASYMPTOTIC PROPERTY OF THE BAYESIAN BOOTSTRAP MEAN

BY CHUNG-SING WENG

State University of New York at Buffalo

It is shown that the Bayesian bootstrap approximation to the posterior distribution of the unknown mean (with respect to a Dirichlet process prior) is more accurate than both the standard normal approximation and the bootstrap approximation. It is also shown that the Bayesian bootstrap approximation to the sampling distribution of the sample average is not as accurate as the bootstrap approximation.

1. Introduction. Efron (1982), studying the small-sample similarities between the bootstrap distribution of Efron (1979) and the Dirichlet posterior distribution when sampling from a multinomial population, suggested that the bootstrap method can be used to approximate a posterior distribution. Efron's bootstrap has a Bayesian counterpart called the Bayesian bootstrap [Rubin (1981)]. In a recent article, Lo (1987) showed that, for a variety of popular functionals, the Bayesian bootstrap [Rubin (1981)] and the bootstrap [Efron (1979)] are first-order asymptotically equivalent in the sense that, for almost all sample sequences and subject to proper centerings and the $n^{1/2}$-scaling, they achieve the same limiting conditional distribution. This result indicates that the uses of the Bayesian bootstrap and the bootstrap could be interchanged, at least in the first-order asymptotic sense. An implication is that a frequentist can use the Bayesian bootstrap to approximate the sampling distribution of a statistic.

The purpose of this paper is to point out that, in the case of a mean functional, the Bayesian bootstrap and the bootstrap are different in the second-order asymptotic sense. Our study shows that it is better to use the Bayesian bootstrap to approximate a posterior distribution (with respect to the Dirichlet process prior), and the frequentist bootstrap to approximate the sampling distribution. Our theory relies on an important result of van Zwet (1979) on the Edgeworth expansion for linear combinations of order statistics.

In Section 2, we develop two-term Edgeworth expansions for the conditional distribution of the normalized Bayesian bootstrap mean given the sample and for the posterior distribution of the normalized mean functional based on a Dirichlet process prior. The second term of these expansions turn out to be identical, and they are both different from the Edgeworth expansion for the bootstrap distribution of the sample average [Singh (1981)]. A consequence is that the Bayesian bootstrap approximation to the posterior distribution of the mean based on a Dirichlet process prior is more accurate than both the usual normal approximation discussed in Lo (1987) and the bootstrap approximation suggested by Efron (1982), page 82. Another consequence of the expansions is

Received May 1987; revised August 1988.
AMS 1980 *subject classification*. Primary 62O99.
Key words and phrases. Bayesian bootstrap, Edgeworth expansion, Dirichlet process.

that the Bayesian bootstrap approximation to the sampling distribution of the sample average is only as good as the standard normal approximation, yet not as accurate as the one based on the bootstrap method [Singh (1981)].

2. The accuracy of the Bayesian bootstrap approximation to a posterior distribution of the unknown mean. Suppose F is a random distribution having a Dirichlet process prior with a finite shape measure α. Given F, let $\mathbf{X} = (X_1, X_2, \dots, X_n)$ be an iid sample from F. Ferguson (1973) showed that $\angle\{F|\mathbf{X}\}$ is a Dirichlet process with shape measure $\alpha + \Sigma_j \delta_{X_j}$, where δ_x is a point mass probability at x. A Bayesian's opinion of the unknown mean can then be summarized by $\angle\{\int xF(dx)|\mathbf{X}\}$. The description of $\angle\{\int xF(dx)|\mathbf{X}\}$ is difficult. One way out is to use a large-sample approximation, and Lo (1987) has shown that the posterior distribution of the appropriately normalized $\int xF(dx)$ has a standard normal distribution. An alternative way is to use the Bayesian bootstrap to approximate $\angle\{\int xF(dx)|\mathbf{X}\}$. Denote the order statistics of $n-1$ iid uniform (0,1) random variables by $0 = U_{0:n-1} < U_{1:n-1} < \cdots < U_{n-1:n-1} < U_{n:n-1} = 1$, and let $\Delta_{j:n} = U_{j:n-1} - U_{j-1:n-1}$, $j = 1, \dots, n$, be the n one-spacings of the $U_{j:n-1}$'s. The U's and the X's are independent. Define D_n by

$$D_n(y) = \sum_j \Delta_{j:n} \delta_{X_j}(y), \qquad -\infty < y < \infty. \tag{2.1}$$

[For any finite measure γ on the line, we denote $\gamma(-\infty, y]$ by $\gamma(y)$.] Given $\mathbf{X}$, the Bayesian bootstrap mean is $\int y D_n(dy) = \Sigma X_j \Delta_{j:n}$ with conditional mean $\mu_n = \Sigma X_j / n$ and conditional variance $\sigma_n^2/(n+1)$ where $\sigma_n^2 = \Sigma(X_j - \mu_n)^2/n$. ($\Sigma$ denotes $\Sigma_{j:1\le j \le n}$). Denote the conditional distribution function of the normalized Bayesian bootstrap mean by $F^*(\cdot|\mathbf{X})$, i.e.,

$$F^*(x|\mathbf{X}) = P\Big\{(n+1)^{1/2}\Big(\sum X_j \Delta_{j:n} - \mu_n\Big)/\sigma_n \le x|\mathbf{X}\Big\}. \tag{2.2}$$

The Bayesian bootstrap method suggests the use of $\angle\{\Sigma X_j \Delta_{j:n}|\mathbf{X}\}$ to approximate $\angle\{\int xF(dx)|\mathbf{X}\}$. The validity of this approximation has also been given in Lo (1987) in the first-order sense. The accuracy of these two approximations can be measured based on second-order Edgeworth expansions for $\angle\{\int xF(dx)|\mathbf{X}\}$ and $\angle\{\Sigma X_j \Delta_{j:n}|\mathbf{X}\}$, which will be given in the following Theorem 2.1.

Next, we define a "posterior Dirichlet process" $D_{\alpha n}$ as follows: Let $Z_1, \dots, Z_n$ be iid standard exponential random variables and $S_n = Z_1 + \cdots + Z_n$. Let $\{\mu(y); -\infty < y < \infty\}$ be a gamma process with finite shape measure α. That is, $\mu(y)$ is an independent increment process and for each y, $\mu(y)$ is a gamma $(\alpha(y), 1)$ random variable. Assume that $\{X_j\}$, $\{Z_j\}$ and $\{\mu(y)\}$ are independent. Define $D_{\alpha n}$ by

$$D_{\alpha n}(y) = \Big\{\mu(y) + \sum_j Z_j \delta_{X_j}(y)\Big\}/\{\mu(\infty) + S_n\}, \qquad -\infty < y < \infty. \tag{2.3}$$

Given $\mathbf{X}$, $\mu(y) + \Sigma_j Z_j \delta_{X_j}(y)$ is a gamma process with shape measure $\alpha + \Sigma_j \delta_{X_j}$

and hence $D_{\alpha n}|\mathbf{X}$ is a Dirichlet process with shape measure $\alpha + \sum_j \delta_{Xj}$. [Note that the $D_{\alpha n}$ here have the same posterior distribution as the $D_{\alpha n}$ in Lo (1987).]

The mean functional of the "posterior" Dirichlet process is

$$\int y D_{\alpha n}(dy) = \left[\int y\mu(dy) + \sum Z_j X_j\right] / [\mu(\infty) + S_n], \tag{2.4}$$

the posterior mean of which (given $\mathbf{X}$) is

$$\mu_{\alpha n} = \left[\int y\alpha(dy) + \sum X_j\right] / [\alpha(\infty) + n]$$

and the posterior variance

$$\sigma_{\alpha n}^2 = \left[\int y^2\alpha(dy) + \sum X_j^2\right] / \{[\alpha(\infty) + n + 1][\alpha(\infty) + n]\}$$
$$- \left[\int y\alpha(dy) + \sum X_j\right]^2 / \{[\alpha(\infty) + n + 1][\alpha(\infty) + n]^2\}.$$

Let $\Phi(x)$ and $\phi(x)$ be the standard normal distribution function and density, respectively. Denote the "true" distribution function of the sample by F_0 and let $\mu_0 = \int x F_0(dx)$, $\sigma_0^2 = \int (x - \mu_0)^2 F_0(dx) > 0$, $\mu_0^{(3)} = \int (x - \mu_0)^3 F_0(dx)$. Denote the posterior distribution of the normalized mean functional given the sample $\mathbf{X}$ by $F^\alpha(\cdot|\mathbf{X})$, i.e.,

$$F^\alpha(x|\mathbf{X}) = P\left\{\left[\int y D_{\alpha n}(dy) - \mu_{\alpha n}\right] / \sigma_{\alpha n} \le x | \mathbf{X}\right\}. \tag{2.5}$$

THEOREM 2.1. (i) *If* $\int |x|^3 F_0(dx) < \infty$, *then uniformly in* x,

$$F^*(x|\mathbf{X}) = \Phi(x) - \{\mu_0^{(3)}/(3n^{1/2}\sigma_0^3)\}(x^2 - 1)\phi(x) + o(n^{-1/2}) \quad a.s.\ [F_0].$$

(ii) *If* $\int |x|^3 F_0(dx) < \infty$ *and* $\int |x|^3 \alpha(dx) < \infty$, *then uniformly in* x,

$$F^\alpha(x|\mathbf{X}) = \Phi(x) - \{\mu_0^{(3)}/(3n^{1/2}\sigma_0^3)\}(x^2 - 1)\phi(x) + o(n^{-1/2}) \quad a.s.\ [F_0].$$

PROOF. Our proof relies heavily on Edgeworth expansions for linear functions of uniform order statistics obtained by van Zwet (1979).

Statement (i) is in fact a direct consequence: Invoking Theorem 1 (or more directly Corollary 2) in van Zwet (1979), we have

$$\begin{aligned} &\sup_{x \in R} |F^*(x|\mathbf{X}) - \Phi(x) + \{\mu_n^{(3)}/(3n^{1/2}\sigma_n^3)\}(x^2 - 1)\phi(x)| \\ &\quad = O\left\{\sum (X_j - \mu_n)^4/(n^2\sigma_n^4)\right\} \quad \text{a.s. } [F_0], \end{aligned} \tag{2.6}$$

where $\mu_n^{(3)} = \sum (X_j - \mu_n)^3/n$. Furthermore,

$$\begin{aligned} \sum (X_j - \mu_n)^4/(n^2\sigma_n^4) &\le \left(\max_{1 \le j \le n} [|X_j - \mu_n|/(n^{1/2}\sigma_n)]\right) \\ &\quad \times \left(\sum |X_j - \mu_n|^3/(n^{3/2}\sigma_n^3)\right) \\ &= o(n^{-1/2}) \quad \text{a.s. } [F_0]. \end{aligned}$$

The last equality follows from

$$\max_{1\le j\le n}\left[|X_j-\mu_n|/\left(n^{1/2}\sigma_n\right)\right]\to 0 \quad \text{a.s. } [F_0]$$

and

$$\sum|X_j-\mu_n|^3/\left(n^{3/2}\sigma_n^3\right)=O(n^{-1/2}) \quad \text{a.s. } [F_0],$$

which are consequences of the assumptions that F_0 has a positive variance and a finite third moment. These assumptions also ensure that we can replace $\mu_n^{(3)}$ and σ_n by $\mu_0^{(3)}$ and σ_0, respectively, without affecting the order of the remainder term $o(n^{-1/2})$ in expansion (2.6).

A proof of statement (ii) can be obtained by modifying the arguments in van Zwet's proof of his Theorem 1 [van Zwet (1979)]. Since the random vector $\{\Delta_{j:n};\ j=1,\dots,n\}$ has the same distribution as $\{Z_j/S_n;\ j=1,\dots,n\}$, we have $\angle\{\Sigma X_j\Delta_{j:n}|\mathbf{X}\}=\angle\{\Sigma X_jZ_j/S_n|\mathbf{X}\}$. It follows that (2.2) can be expressed as

$$F^*(x|\mathbf{X})=P\left\{(n+1)^{1/2}\left(\sum X_jZ_j/S_n-\mu_n\right)/\sigma_n\le x|\mathbf{X}\right\}. \tag{2.7}$$

Furthermore, (2.4) and (2.5) imply

$$F^\alpha(x|\mathbf{X})=P\left\{\left[\left(\int y\mu(dy)+\sum X_jZ_j\right)/(\mu(\infty)+S_n)-\mu_{\alpha n}\right]\Big/\sigma_{\alpha n}\le x|\mathbf{X}\right\}. \tag{2.8}$$

Notice that the structure of the mean functional of the "posterior" Dirichlet process $(\int y\mu(dy)+\Sigma X_jZ_j)/(\mu(\infty)+S_n)$ in (2.8) is quite similar to that of the Bayesian bootstrap mean $\Sigma X_jZ_j/S_n$ in (2.7). The differences are linear terms in the numerator and the denominator, both generated by a gamma process μ which is independent of $\{Z_j;\ j=1,\dots,n\}$ and $\mathbf{X}$. Taking into account a three-term Taylor's expansion for the characteristic function of $\int(y-1)\mu(dy)$, one can modify the method of proof of Theorem 1 in van Zwet (1979) to prove (ii). Details will not be given here, and can be obtained from the author. □

REMARK 2.1. Note that the almost surely remainder term $o(n^{-1/2})$ in the conclusions of Theorem 2.1 depends on the sample $\mathbf{X}$. The same remark applies to the almost surely remainder terms that follow.

REMARK 2.2. The posterior distribution (2.5) is normalized by the posterior mean and the posterior standard deviation, which is natural. Other choices of scaling are also possible. Define

$$F^\alpha(x|\mathbf{X})=P\left\{\left[\int yD_{\alpha n}(dy)-\mu_{\alpha n}\right]/c_n\le x|\mathbf{X}\right\},$$

where c_n is a statistic. Theorem 2.1(ii) remains valid as long as $c_n/\sigma_{\alpha n} = 1 + o(n^{-1/2})$ a.s. $[F_0]$.

The following corollary states the accuracy of the Bayesian bootstrap approximation to the posterior distribution of the mean and the inaccuracy of the usual normal approximation.

COROLLARY 2.1. *$\int|x|^3F_0(dx) < \infty$ and $\int|x|^3\alpha(dx) < \infty$ imply that a.s. $[F_0]$,*

$$\lim_{n\to\infty} \sup_{x\in R} \sqrt{n}|F^{\alpha}(x|\mathbf{X}) - F^*(x|\mathbf{X})| = 0, \tag{2.9}$$

$$\lim_{n\to\infty} \sup_{x\in R} \sqrt{n}|F^{\alpha}(x|\mathbf{X}) - \Phi(x)| = (2\pi)^{-1/2}|\mu_0^{(3)}|/\left[3\sigma_0^3\right].$$

The above corollary is the Bayesian analogue of the well known frequentist result that, for a nonlattice F_0, the bootstrap method provides a better approximation to the sampling distribution of the sample average [Singh (1981), Theorem 1(D)]. Note that Corollary 2.1 applies also to a lattice F_0. On the other hand, Theorem 2.1(ii) and Singh's result imply that, for a nonlattice F_0, the bootstrap approximation to the posterior distribution of the mean functional is only as good as the standard normal approximation.

REMARK 2.3. Theorem 2.1(i) and a two-term Edgeworth expansion for the sampling distribution of the sample mean [Feller (1971), page 539] imply that, in approximating the sampling distribution of the sample mean when sampling from a nonlattice F_0, the Bayesian bootstrap approximation is as accurate as the standard normal approximation and is inferior to the bootstrap approximation of it [Singh (1981), Theorem 1(D)].

However, a slight modification of the Bayesian bootstrap results in a modified Bayesian bootstrap such that, given the sample, a two-term Edgeworth expansion for the distribution of the modified Bayesian bootstrap mean is identical to the two-term expansion for the sampling distribution of the sample mean. The idea is to use n "four-spacings" $\{D_{j:n} = U_{4j:4n-1} - U_{4(j-1):4n-1},\ j = 1,\dots,n\}$ instead of n "one-spacings" $\{\Delta_{j:n} = U_{j:n-1} - U_{j-1:n-1},\ j = 1,\dots,n\}$ in our definition of the Bayesian bootstrap in (2.1). The first-order asymptotics remain unchanged for this modified Bayesian bootstrap. Furthermore, van Zwet's (1979) method of proof goes through with practically no changes to yield the desired two-term expansion. Yet we must point out that, in addition to being four times more expensive to implement, the choice of four-spacings depends on the functional of interest and is not universal.

Acknowledgments. This is a part of the author's Ph.D. thesis. We thank Albert Lo for suggesting the problem and for numerous stimulating conversations, Don Rubin for several helpful suggestions and Willem van Zwet and a referee for correcting an error in the first version.

REFERENCES

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.

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

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** 1187–1195.

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

DEPARTMENT OF BIOSTATISTICS, K-6-2
SCHERING-PLOUGH CORPORATION
KENILWORTH, NEW JERSEY 07033