Statistica Sinica **7**(1997), 1005-1019

A BAYESIAN BOOTSTRAP FOR FINITE STATE MARKOV CHAINS

Cheng-Der Fuh and Tsai-Hung Fan

Academia Sinica and National Central University

Abstract: The Bayesian bootstrap for Markov chains is the Bayesian analogue of the bootstrap method for Markov chains. We construct a random-weighted empirical distribution, based on i.i.d. exponential random variables, to simulate the posterior distribution of the transition probability, the stationary probability, as well as the first hitting time up to a specific state, of a finite state ergodic Markov chain. The large sample theory is developed which shows that with a matrix beta prior on the transition probability, the Bayesian bootstrap procedure is second-order consistent for approximating the pivot of the posterior distributions of the transition probability. The small sample properties of the Bayesian bootstrap are also discussed by a simulation study.

Key words and phrases: Bayesian bootstrap, hitting time, Markov chain, matrix beta distribution, transition probability.

1. Introduction

The Markov chain model has become an important and a powerful tool for the statistician, engineer and economist. It provides the researcher with a modeling framework and a computationally efficient way to compute parameter estimates over a wide range of situations. Problems involved in the social sciences, in queuing networks, in adaptive control, and in information science, all have benefited from this tool. Statistical inference for Markov chains is summarized in Billingsley (1961); the Bayesian approach for Markov chains is gvien in Martin (1975), who offers an alternative method for modeling and estimation.

By using the idea of the bootstrap (cf. Efron (1979)), Kulperger and Rao (1989) consider Markov chains with finite state space; the countable state space case has been investigated by Athreya and Fuh (1992). The asymptotic validity of these models has been established by the respective authors. The small sample property for bootstrapping Markov chains is in Fuh (1993).

The Bayesian analogue of the frequentist bootstrap for i.i.d. random variables has been proposed by Rubin (1981). The idea is that, given i.i.d. random variables, $\underline{X} = (X_1, \ldots, X_n)$ having an unknown distribution function F, and a given specific functional $\theta(F, \underline{X})$ depending on both the unknown distribution function F and the given data $\underline{X}$, we are interested in accessing the "posterior" opinion of $\theta(F, \underline{X})$ given $\underline{X} = \underline{x}$. A Bayesian approach to this problem is to construct a prior distribution on F and then use the posterior distribution of $\theta(F, \underline{X})$ given $\underline{X} = \underline{x}$ to summarize the "posterior" opinion of $\theta(F, \underline{X})$ given $\underline{X} = \underline{x}$. The Bayesian bootstrap to solve this problem is a procedure based on a random distribution D_n, by replacing the jump-sizes of the empirical distribution function by the gaps of $n-1$ i.i.d. $U(0,1)$ random variables and suggesting that the conditional distribution of $\theta(D_n, \underline{X})|\underline{X} = \underline{x}$ can be used as the posterior distribution of $\theta(F, \underline{X})|\underline{X} = \underline{x}$. Recently, Lo (1987) showed that the Bayesian bootstrap and the frequentist bootstrap are first-order asymptotically equivalent. Weng (1989) considered second-order efficiency properties. Hjort (1991) and Lo (1993a) also developed the Bayesian bootstrap for censored data.

In this paper, we investigate a Bayesian analogue to the frequentist bootstrap for finite state Markov chains. It turns out that the posterior distribution of the transition probability matrix with respect to a "flat" conjugate prior is the distribution of the Bayesian bootstrap. Large sample results are also considered in terms of second-order asymptotic justification.

The model considered is an ergodic Markov chain, $\{X_n\}$, with finite state space $\{1, \ldots, k\}$ and transition probability matrix P. Let $\underline{x} = \{x_0, x_1, \ldots, x_n\}$ be a realization from this process, where $x_0 = 1$ is assumed known. Thus $\underline{x}$ is obtained under the consecutive sampling rule. The likelihood function of the transition probability matrix P is of the form

$$l(P|\underline{x}) \propto \prod_{i,j=1}^{k} p_{ij}^{n_{ij}},$$

where p_{ij} is the transition probability from state i to state j with $\sum_{j=1}^{k} p_{ij} = 1$, and n_{ij} is the number of ij transitions during the given sample $\underline{x}$ with $\sum_{j=1}^{k} n_{ij} = n_i$, the number of visits to state i in $\underline{x}$. Thus, the natural conjugate prior for P is the matrix beta distribution with density

$$\pi(P|M) \propto \prod_{i,j=1}^{k} p_{ij}^{m_{ij}-1}, \qquad (1)$$

where $M = [m_{ij}]$ is a $k \times k$ matrix such that $m_{ij} \geq 0$, for $i, j = 1, \ldots, k$, and $\sum_{j=1}^{k} p_{ij} = 1$, for $i = 1, \ldots, k$.

By a simple calculation, it can be seen that the posterior distribution a P given $\underline{x}$ is also a matrix beta distribution but with parameter $M' = [m_{ij} + n_{ij}]$. The posterior distributions of Π (the stationary probability) and T_k (the first hitting time up to state k), both being smooth functions of P, are difficult to clarify in this Bayesian framework, however. We therefore propose the Bayesian bootstrap procedure to approximate the posterior distribution of the parameters of interest.

Based on a given realization $\underline{x} = (x_0, \ldots, x_n)$ of the Markov chain $\{X_n\}$, the m.l.e. of P is

$$\hat{P}_n \equiv [\,\hat{p}_n(i,j)] = [n_{ij}/n_i]. \tag{2}$$

Instead of resampling according to $\hat{P}_n$ in the parametric bootstrap method, the Bayesian bootstrap is based on a simulation from i.i.d. standard exponential variables Z_{it} and sufficient statistics n_{ij}, which will be described precisely in Section 2. In Section 3, we give the asymptotic justification of the Bayesian bootstrap procedure. Some empirical studies that illustrate the utility of the Bayesian bootstrap procedure in small samples will be presented in the last section.

2. The Bayesian Bootstrap Algorithm

Following the notation defined in the previous section, we now assume a "flat" matrix beta prior density for P, i.e. all the $m_{ij} = 0$ in (1), then the corresponding posterior distribution of P given $\underline{x}$ is

$$\pi(P|M') = \prod_{i}^{k} \frac{\Gamma(n_i)}{\prod_{j=1}^{k} \Gamma(n_{ij})} \Big(\prod_{j=1}^{k} p_{ij}^{n_{ij}-1} \Big), \tag{3}$$

where $\sum_{j=1}^{k} p_{ij} = 1$, for any $i = 1, \ldots, k$, and $\sum_{j=1}^{k} n_{ij} = n_i$. Thus, in this case, the posterior mean (which is the same as the Bayes estimate with respect to the squared error loss) of the transition probability matrix $P = [p_{ij}]$ is identical to the m.l.e. $\hat{P}_n \equiv [\,\hat{p}_n(i,j)]$ defined in (2). Note that $\hat{p}_n(i,j)$ can be written as

$$\hat{p}_n(i,j) = \frac{\sum_{t \in B_{ij}} 1}{\sum_{t=1}^{n_i} 1}, \tag{4}$$

where $B_{ij} = \{\sum_{l=1}^{j} n_{i(l-1)} + 1, \ldots, \sum_{l=1}^{j} n_{il}\}$, with $n_{i0} = 0$. (See Remark 2.1 below.) In the parametric bootstrap (cf. Athreya and Fuh (1992)), it uses $\hat{P}_n$ as a mechanism to generate bootstrap samples. Here the Bayesian bootstrap is based on simulation rather than resampling. That is, for each fixed i, j, we replace the "$1's$" in (4) by the i.i.d. standard exponential random variables "$Z_{it}'s$", and obtain

$$\hat{p}_n^*(i,j) = \frac{\sum_{t \in B_{ij}} Z_{it}}{\sum_{t=1}^{n_i} Z_{it}}.$$

The joint distribution of $\hat{P}_n^* \equiv [\,\hat{p}_n^*(i,j)]$ can be calculated analytically: For given data and fixed i, let $Q_l = Z_{il}/\sum_{t=1}^{n_i} Z_{it}$, $l = 1, 2, \ldots, n_i$. Partition Q_l into k collections, the jth having n_{ij} elements. Then the above $\hat{p}_n^*(i,j)$ is the sum of the Q_l in the jth collection, $j = 1, \ldots, k$. Therefore, the joint distribution of

$\hat{\mathbf{p}}_i^* = (\hat{p}_n^*(i,1), \ldots, \hat{p}_n^*(i,k))$ is the same as that of the $k-1$ gaps of k ordered independent U(0,1) random variables (cf. Reiss (1989)). Such joint distribution is indeed the $k-1$ variate Dirichlet $(n_{i1}, \ldots, n_{ik})$ (cf. Wilks (1962)), which is also the posterior distribution of $\mathbf{p}_i = (p_{i1}, \ldots, p_{ik})$. Thus both the Bayesian bootstrap and the posterior distribution of P turn out to be the matrix beta distribution given by (3).

Now we summarize the Bayesian bootstrap algorithm for simulating the posterior distribution of P as follows:

(1) Simulate i.i.d. standard exponential random variables, $Z_{it},\ t = 1, \ldots, n_i$, $i = 1, \ldots, k$.

(2) Replace the "$1's$" in (4) by "$Z'_{it}s$" to obtain $\hat{p}_n^*(i,j)$.

(3) Repeat the previous two steps a large number of times, say B times, to obtain $\hat{P}_{n1}^*, \ldots, \hat{P}_{nB}^*$, and use the empirical distribution based on $\hat{P}_{nt}^*, t = 1, \ldots, B$ to approximate the posterior distribution $\mathcal{L}\{P|\underline{x}\}$ of P given $\underline{x}$.

The corresponding Bayesian bootstrap method to approximate the posterior distributions of Π and T_k given $\underline{x}$ can be developed as follows:

(1) By the balance equation $\Pi P = \Pi$, we can apply $\hat{P}_{n1}^*, \ldots, \hat{P}_{nB}^*$ to get the corresponding $\hat{\Pi}_{n1}^*, \ldots, \hat{\Pi}_{nB}^*$, and use the empirical distribution of $\hat{\Pi}_{nt}^*,\ t = 1, \ldots, B$ to approximate the posterior distribution $\mathcal{L}\{\Pi|\underline{x}\}$ of Π given $\underline{x}$.

(2) The distribution function of T_k is $Pr(t;P) \equiv Pr(T_k \leq t | X_0 = 1, P) = (A^t)_{1,k}$, where $A \equiv A(P)$, the stochastic matrix which is the same as P except that the kth row is replaced by $(0, \ldots, 0, 1)$. Therefore, we use the empirical distribution of $Pr(t; \hat{P}_{nt}^*),\ t = 1, \ldots, B$ to approximate the posterior distribution $\mathcal{L}\{Pr(t;P)|\underline{x}\}$ of $Pr(t;P)$ given $\underline{x}$.

Remark 2.1. The B_{ij} defined above are obtained by first partitioning $\{1, \ldots, n\}$ into k groups of sizes $n_1, \ldots, n_k$, say $B_1, \ldots, B_k$, then dividing each B_i into k disjoint subgroups of sizes $n_{i1}, \ldots, n_{ik}$. These subgroups are denoted by B_{ij}. For example, for a 3-state Markov chain with 30 observations such that $n_{11} = 5, n_{12} = 3, n_{13} = 1, n_{21} = 3, n_{22} = 4, n_{23} = 2, n_{31} = 7, n_{32} = 2$, and $n_{33} = 3$, we have $B_{11} = \{1,2,3,4,5\}, B_{12} = \{6,7,8\}$, and $B_{13} = \{9\}$, etc; hence, $\hat{p}_n^*(1,1) = \sum_{t=1}^{5} Z_{1t} / \sum_{t=1}^{9} Z_{1t}$, $\hat{p}_n^*(1,2) = \sum_{t=6}^{8} Z_{1t} / \sum_{t=1}^{9} Z_{1t}$, and $\hat{p}_n^*(1,3) = Z_{19} / \sum_{t=1}^{9} Z_{1t}$, etc. in this case. In fact, any well defined B_{ij} such that the density of $\hat{P}_n^*$ is given by $\pi(P|M')$ will work.

Remark 2.2. To approximate the posterior distribution of Π, we need to solve a system of equations, $\hat{\Pi}^* \hat{P}^* = \hat{\Pi}^*$, for $\hat{\Pi}^*$ as many as B times which can be handled by many computer packages such as Mathematica, IMSL, etc..

Remark 2.3. When the Markov chain is operating in the steady-state and the initial state X_0 is unknown, the distribution X_0 is $\Pi(P) = (\pi_1(P), \ldots, \pi_k(P))$, the steady-state probability associated with the transition matrix P. In this case, observation X_0 provides information about P. When P is regarded as a random matrix, the natural conjugate prior is the matrix beta-1 distribution (cf. Martin (1975)). The Bayesian bootstrap for this situation can be carried out in a similar way and will not be repeated here.

Remark 2.4. The Bayesian bootstrap incorporating prior information was suggested by Lo (1988). In the present case, a similar method can be adapted to incorporate prior information. Suppose the prior information is summarized by a set of m prior Markov chain data $\{x_0, x_1, \ldots, x_m\}$. Combine these prior data with the current data $\{x_t,\ t = m+1, \ldots, m+n\}$ to obtain an updated sample instead of $\{x_t,\ t = m+1, \ldots, m+n\}$. Here it is necessary to simulate $m+n$ i.i.d. standard exponential random variables Z_{it} in each execution of the Bayesian bootstrap algorithm. In this case, we may choose Z_{it} to be from the gamma distribution to reflect the prior distribution. Further research in this case needs to be done.

Remark 2.5. The Bayesian bootstrap described above depends on the prior distribution of the elements of P in such a manner as to reflect accurately the decision maker's state of knowledge. It would be of considerable interest, therefore, to investigate the sensitivity of the Bayesian bootstrap to relatively small changes in the prior distribution.

3. Asymptotic Justification of the Procedure

The asymptotic behavior of generalized Bayes estimators for a Markov chain with transition probability $P(\theta)$, which is a function of a one dimensional parameter θ, has been developed by Levit (1974). Here we consider a specific posterior distribution of P but with multi-dimensional parameters. We show in this section that under appropriate conditions, the Bayesian bootstrap procedure applied to the innovations yields asymptotically consistent estimators for the posterior distributions of P, Π and T_k given $\underline{x}$. The main results will be developed only for P.

Theorem 1 gives the central limit theorem for the Bayesian bootstrap distribution. The proof of Theorem 1 provides insight into the construction of the Bayesian bootstrap distribution, although it can be provided by Theorem 2 instead.

Theorem 1. *Along almost all sample sequences, for each i, j, as $n \to \infty$, we have*

$$\frac{\sqrt{n_i}(\hat{p}_n^*(i,j) - \hat{p}_{ij})}{\sqrt{\hat{p}_{ij}(1-\hat{p}_{ij})}} | \underline{x} \longrightarrow N(0,1) \quad \textit{in distribution},$$

where $\hat{p}_{ij} = \hat{p}_n(i,j)$, the posterior mean with respect to the "flat" prior (as well as the m.l.e.) of p_{ij} defined in (2).

Proof. For any fixed i, j, we define

$$S'_{n_{ij}} \equiv \sum_{t \in B_{ij}} (Z_{it} - 1), \quad \text{and} \quad S'_{n_i, n_i - n_{ij}} \equiv \sum_{t=1}^{n_i} (Z_{it} - 1) - \sum_{t \in B_{ij}} (Z_{it} - 1).$$

Then for any given $\underline{x}$,

$$\begin{aligned} &\sqrt{n_i}(\hat{p}_n^*(i,j) - \hat{p}_{ij}) = \sqrt{n_i}\Big(\sum_{t \in B_{ij}} Z_{it} / \sum_{t=1}^{n_i} Z_{it} - n_{ij}/n_i\Big) \\ &= \frac{n_i}{\sum_{t=1}^{n_i} Z_{it}} \Big[\Big(1 - \frac{n_{ij}}{n_i}\Big) \sqrt{\frac{n_{ij}}{n_i}} \frac{S'_{n_{ij}}}{\sqrt{n_{ij}}} - \frac{n_{ij}}{n_i} \sqrt{1 - \frac{n_{ij}}{n_i}} \frac{S'_{n_i, n_i - n_{ij}}}{\sqrt{n_i - n_{ij}}} \Big] \qquad (5) \\ &= \frac{n_i}{\sum_{t=1}^{n_i} Z_{it}} \Big[(1 - \hat{p}_{ij}) \sqrt{\frac{n_{ij}}{n_i}} \frac{S'_{n_{ij}}}{\sqrt{n_{ij}}} - \hat{p}_{ij} \sqrt{1 - \hat{p}_{ij}} \frac{S'_{n_i, n_i - n_{ij}}}{\sqrt{n_i - n_{ij}}} \Big] \\ &= (I) - (II) \qquad \text{(say).} \end{aligned}$$

Since $n_i / \sum_{t=1}^{n_i} Z_{it} \to 1$ almost surely, it is easy to check that for almost all sample sequences, both $\frac{(I)}{(1-\hat{p}_{ij})\sqrt{\hat{p}_{ij}}}$ and $\frac{(II)}{\hat{p}_{ij}\sqrt{1-\hat{p}_{ij}}}$ converge in distribution to the standard normal distribution. Therefore,

$$\sqrt{n_i}(\hat{p}_n^*(i,j) - \hat{p}_{ij}) / [\hat{p}_{ij}(1 - \hat{p}_{ij})]^{1/2} | \underline{x} \longrightarrow N(0,1) \quad \text{in distribution,}$$

as $n \to \infty$.

Under mild conditions of the likelihood, Johnson (1970) gives the asymptotic expansion for the posterior distribution of the parameter of interest with respect to a positive and sufficiently smooth prior. Such an expansion is also discussed in Ghosh (1994). This result shows that as the sample size n goes to infinity, the normalized posterior distribution of p_{ij}, with respect to any positive and smooth prior, converges to the standard normal distribution, i.e.

$$\sqrt{n_i}(p_{ij} - \hat{p}_{ij}) / [\hat{p}_{ij}(1 - \hat{p}_{ij})]^{1/2} | \underline{x} \longrightarrow N(0,1) \quad \text{in distribution.}$$

Applying Johnson's expansion, we have the following lemma that can be used to prove the result of the second order efficiency for the posterior of p_{ij}.

Lemma 1. *Let $\pi(p_{ij})$ be any positive prior with continuous second derivatives for all $0 < p_{ij} < 1$; then along almost all sample sequences, for each i, j, as $n \to \infty$,*

$$P\Big(\frac{\sqrt{n_i}(p_{ij} - \hat{p}_{ij})}{\sqrt{\hat{p}_{ij}(1 - \hat{p}_{ij})}} \le t | \underline{x} \Big) = \Phi(t) - \gamma(t, \underline{x}) \frac{\phi(t)}{\sqrt{n}} + o_p(n^{-1/2}),$$

where $\Phi(t)$ and $\phi(t)$ are the c.d.f. and p.d.f. of the standard normal distribution, and

$$\gamma(t,x) = (1-2\hat{p}_{ij})/[3\sqrt{\hat{p}_{ij}(1-\hat{p}_{ij})}](t^2+2) + \sqrt{\hat{p}_{ij}(1-\hat{p}_{ij})}\pi^{(1)}(\hat{p}_{ij})/\pi(\hat{p}_{ij}), \quad (6)$$

with $\pi^{(k)}(\cdot)$ the kth derivative of $\pi(\cdot)$.

Proof. Note that the likelihood function of p_{ij} is

$$l(p_{ij}|\underline{x}) \propto p_{ij}^{n_{ij}}(1-p_{ij})^{n_i-n_{ij}}, \quad (7)$$

which, along with $\pi(p_{ij})$, satisfy the conditions given in Johnson (1970). Let

$$b(p_{ij}) = \Big[-\frac{1}{n_i}\frac{\partial^2}{\partial p_{ij}^2}\log l(p_{ij}|\underline{x})\Big]^{1/2}, \quad \text{and} \quad a_{3;n}(p_{ij}) = \frac{1}{6n_i}\frac{\partial^3}{\partial p_{ij}^3}\log l(p_{ij}|\underline{x}).$$

Then following Theorem 5.1 of Johnson (1970), the posterior distribution function of $W_{ij} = \sqrt{n_i}(p_{ij}-\hat{p}_{ij})b(\hat{p}_{ij})$ can be expanded as

$$P(W_{ij} \le t|\underline{x}) = \Phi(t) - \gamma_1(t,\underline{x})\phi(t)/\sqrt{n_i} + o_p(n_i^{-1/2}),$$

where

$$\gamma_1(t,\underline{x}) = \pi^{-1}(\hat{p}_{ij})\{b^{-3}(\hat{p}_{ij})a_{3;n}(\hat{p}_{ij})\pi(\hat{p}_{ij})(t^2+2) + b^{-1}(\hat{p}_{ij})\pi^{(1)}(\hat{p}_{ij})\}.$$

Simple algebra yields

$$b(\hat{p}_{ij}) = \Big[\frac{1}{n_i}\Big(\frac{n_{ij}}{\hat{p}_{ij}^2} + \frac{n_i-n_{ij}}{(1-\hat{p}_{ij})^2}\Big)\Big]^{1/2} = \Big(\frac{1}{\hat{p}_{ij}} + \frac{1}{1-\hat{p}_{ij}}\Big)^{1/2} = [\hat{p}_{ij}(1-\hat{p}_{ij})]^{-1/2},$$

and

$$\begin{aligned} a_{3,n_i}(\hat{p}_{ij}) &= \frac{1}{3}\frac{1}{n_i}\Big(\frac{n_{ij}}{\hat{p}_{ij}^3} - \frac{n_i-n_{ij}}{(1-\hat{p}_{ij})^3}\Big) = \frac{1}{3}\Big(\frac{1}{\hat{p}_{ij}^2} + \frac{1}{(1-\hat{p}_{ij})^2}\Big) \\ &= \frac{1}{3}(1-2\hat{p}_{ij})/[\hat{p}_{ij}(1-\hat{p}_{ij})]^2, \end{aligned}$$

which establishes the result for n sufficiently large.

Lemma 1 gives the second order approximation to the posterior with respect to smooth priors. For considering of the second order efficiency of the Bayesian bootstrap, we only need find $\gamma(t,\underline{x})$ with respect to the "flat" prior, since the resulting posterior distribution is the same as the distribution of the Bayesian bootstrap, as pointed out in Section 2. The result is given in Theorem 2.

Theorem 2. *Along almost all sample sequences, for each i,j, uniformly for $\underline{x}$, we have*

$$P\Big(\frac{\sqrt{n_i}(\hat{p}_n^*(i,j)-\hat{p}_{ij})}{\sqrt{\hat{p}_{ij}(1-\hat{p}_{ij})}} \le t|\underline{x}\Big) = \Phi(t) + \frac{1-2\hat{p}_{ij}}{3\sqrt{\hat{p}_{ij}(1-\hat{p}_{ij})}}\frac{1-t^2}{\sqrt{n}}\phi(t) + o_p(n^{-1/2})$$

Proof. Following Lemma 1, with $\pi(p_{ij}) = p_{ij}^{-1}(1-p_{ij})^{-1}$, we have $\pi^{(1)}(\hat{p}_{ij}) = -(1-2\hat{p}_{ij})/[\hat{p}_{ij}(1-\hat{p}_{ij})]^2$. This reduces (6) to

$$\gamma(t,\underline{x}) = -(1-2\hat{p}_{ij})(1-t^2)/[3\sqrt{\hat{p}_{ij}(1-\hat{p}_{ij})}].$$

Then by Lemma 1, we get

$$P\Big(\frac{\sqrt{n_i}(p_{ij}-\hat{p}_{ij})}{\sqrt{\hat{p}_{ij}(1-\hat{p}_{ij})}} \le t|\underline{x}\Big) = \Phi(t) + \frac{1-2\hat{p}_{ij}}{3\sqrt{\hat{p}_{ij}(1-\hat{p}_{ij})}}\frac{1-t^2}{\sqrt{n}}\phi(t) + o_p(n^{-1/2}).$$

Furthermore, since $\hat{p}_n^*(i,j)|\underline{x}$ and the posterior of $p_{ij}|\underline{x}$ with respect to the "flat" beta prior have the same distribution, they have of course the same expansion. The proof is therefore complete.

An alternative approach to deriving the second order efficiency of the Bayesian bootstrap distribution can be carried out via an Edgeworth expansion following Weng (1989) and Lo (1993a).

According to Lemma 1, when the posterior is centered and re-scaled, respectively, at the m.l.e. and its estimated standard deviation, the second order could involve the prior so it may not be the same as that obtained in Theorem 2. However, if the posterior is normalized by the posterior mean and posterior standard deviation with respect to matrix beta conjugate priors, the second order in the asymptotic expansion is indeed consistent with that given in the Bayesian bootstrap. Such a result is given in Theorem 3.

Theorem 3. *Let $\pi^c(p_{ij}) \propto p_{ij}^{\alpha_{ij}-1}(1-p_{ij})^{\alpha_i-\alpha_{ij}-1}$, for some $0<\alpha_{ij}<1$ and $\sum_{j=1}^k \alpha_{ij} = \alpha_i$. Then along almost all sample sequences, for each i,j, as $n\to\infty$, we have*

$$P\Big(\frac{(p_{ij}-\hat{p}_{ij}^c)}{\sqrt{V^c}} \le t|x\Big) = \Phi(t) + \frac{1-2\hat{p}_{ij}}{3\sqrt{\hat{p}_{ij}(1-\hat{p}_{ij})}}\frac{1-t^2}{\sqrt{n}}\phi(t) + o_p(n^{-1/2}),$$

where $\hat{p}_{ij}^c = (n_{ij}+\alpha_{ij})/(n_i+\alpha_i)$ and $V^c = (\alpha_{ij}+n_{ij})(\alpha_i-\alpha_{ij}+n_i-n_{ij})/[(\alpha_i+n_i)^2(\alpha_i+n_i+1)]$ are the posterior mean and posterior variance of p_{ij}, respectively, with respect to π^c.

Proof. The posterior distribution of p_{ij} with respect to π^c is

$$\pi(p_{ij}|x) \propto l(p_{ij}|x)\pi^c(p_{ij}) \propto p_{ij}^{n_{ij}+\alpha_{ij}-1}(1-p_{ij})^{n_i-n_{ij}+\alpha_i-\alpha_{ij}-1},$$

where $l(p_{ij}|x)$ is defined by (7). Let $\pi_0(p_{ij}) = p_{ij}^{-1}(1-p_{ij})^{-1}$, and $L(p_{ij}) = p_{ij}^{n_{ij}+\alpha_{ij}}(1-p_{ij})^{n_i-n_{ij}+\alpha_i-\alpha_{ij}}$. Then $\pi(p_{ij}|x)$ can be written as

$$\pi(p_{ij}|x) = \frac{\pi_0(p_{ij})\exp[L(p_{ij})]}{\int \pi_0(p_{ij})\exp[L(p_{ij})]dp_{ij}} = \frac{\pi_0(p_{ij})\exp[L(p_{ij})-L(\hat{p}_{ij}^c)]}{\int \pi_0(p_{ij})\exp[L(p_{ij})-L(\hat{p}_{ij}^c)]dp_{ij}},$$

where $\hat{p}^c_{ij}$ is the posterior mean which maximizes $L(p_{ij})$ here.

Note that the posterior density of $h_1 = \sqrt{n_i}(p_{ij} - \hat{p}^c_{ij})$ is

$$\pi(h_1|x) = \frac{\pi_0(n_i^{-1/2}h_1 + \hat{p}^c_{ij})\exp[L(n_i^{-1/2}h_1 + \hat{p}^c_{ij}) - L(\hat{p}^c_{ij})]}{\int \pi_0(n_i^{-1/2}h_1 + \hat{p}^c_{ij})\exp[L(n_i^{-1/2}h_1 + \hat{p}^c_{ij}) - L(\hat{p}^c_{ij})]dh_1}. \tag{8}$$

Taking a Taylor expansion at $\hat{p}_{ij}$ for $\pi_0(\cdot)$ and $L(\cdot)$, and letting $\pi_0^{(k)}$, $L^{(k)}$ represent the corresponding kth derivatives, we get

$$\pi_0(n_i^{-1/2}h_1 + \hat{p}^c_{ij}) = \pi_0(\hat{p}^c_{ij})\Big[1 + n_i^{-1/2}h_1\frac{\pi_0^{(1)}(\hat{p}^c_{ij})}{\pi_0(\hat{p}^c_{ij})} + \frac{1}{2}n_i^{-1}h_1^2\frac{\pi_0^{(2)}(\hat{p}^c_{ij})}{\pi_0(\hat{p}^c_{ij})}\Big] + o(n_i^{-1}),$$

and

$$L(n_i^{-1/2}h_1 + \hat{p}^c_{ij}) = L(\hat{p}^c_{ij}) + \frac{1}{2}n_i^{-1}h_1^2L^{(2)}(\hat{p}^c_{ij}) + \frac{1}{6}n_i^{-3/2}h_1^3L^{(3)}(\hat{p}^c_{ij}) + o(n_i^{-1}),$$

respectively. Thus, letting $b = -n_i^{-1}L^{(2)}(\hat{p}^c_{ij})$, yields

$$\begin{aligned} &\pi_0(n_i^{-1/2}h_1 + \hat{p}^c_{ij})\exp[L(n_i^{-1/2}h_1 + \hat{p}^c_{ij}) - L(\hat{p}^c_{ij})] \\ &= \pi_0(\hat{p}^c_{ij})\exp(-h_1^2b/2)\Big\{1 + h_1\pi_0^{(1)}(\hat{p}^c_{ij})/[n_i^{1/2}\pi_0(\hat{p}^c_{ij})] + h_1^3L^{(3)}(\hat{p}^c_{ij})/(6n_i^{3/2})\Big\} \\ &\quad +o(n_i^{-1/2}), \end{aligned} \tag{9}$$

and

$$\int \pi_0(n_i^{-1/2}h_1+\hat{p}^c_{ij})\exp[L(n_i^{-1/2}h_1+\hat{p}^c_{ij})-L(\hat{p}^c_{ij})]dh_1 = \pi_0(\hat{p}^c_{ij})\sqrt{2\pi/b}+o(n_i^{-1/2}). \tag{10}$$

Putting (9) and (10) into (8), and letting $h = \sqrt{b}h_1 = \sqrt{n_ib}(p_{ij} - \hat{p}^c_{ij})$, gives the posterior density of h given x as

$$\pi(h|x)=\sqrt{\frac{1}{2\pi}}e^{-\frac{1}{2}h^2}\Big(1+n_i^{-1/2}\sqrt{b}^{-1}\frac{\pi_0^{(1)}(\hat{p}^c_{ij})}{\pi_0(\hat{p}^c_{ij})}h+\frac{1}{6}n_i^{-3/2}\sqrt{b}^{-3}L^{(3)}(\hat{p}^c_{ij})h^3\Big)+o(n_i^{-1/2}).$$

Hence,

$$\begin{aligned} P\Big(\sqrt{n_ib}(p_{ij} - \hat{p}^c_{ij}) \le t|x\Big) &= \int_{-\infty}^{t}\pi(h|x)dh \\ &= \Phi(t) + \phi(t)n_i^{-1/2}\Big[\sqrt{b}^{-1}\frac{\pi_0^{(1)}(\hat{p}^c_{ij})}{\pi_0(\hat{p}^c_{ij})} - \frac{(2+t^2)}{6}\frac{L^{(3)}(\hat{p}^c_{ij})}{n_i}\Big] + o(n_i^{-1/2}). \end{aligned}$$

Calculation yields

$$\pi_0^{(1)}(\hat{p}^c_{ij}) = -\frac{(1-2\hat{p}^c_{ij})}{[\hat{p}^c_{ij}(1-\hat{p}^c_{ij})]^2}, \quad L^{(2)}(\hat{p}^c_{ij}) = -\frac{n_i+\alpha_i}{\hat{p}^c_{ij}(1-\hat{p}^c_{ij})},$$

$$\text{and}\quad L^{(3)}(\hat{p}^c_{ij}) = \frac{2(n_i+\alpha_i)(1-2\hat{p}^c_{ij})}{[\hat{p}^c_{ij}(1-\hat{p}^c_{ij})]^2};$$

so

$$
\begin{aligned}
P(\sqrt{n_i b}(p_{ij}-\hat{p}^c_{ij})\leq t|\underline{x}) &= \Phi(t)+\sqrt{\frac{n_i}{n_i+\alpha_i}}\frac{(1-2\hat{p}^c_{ij})}{\sqrt{\hat{p}^c_{ij}(1-\hat{p}^c_{ij})}}\frac{(1-t^2)}{3\sqrt{n_i}}\phi(t)+o_p(n_i^{-1/2})\\
&= \Phi(t)+\frac{(1-2\hat{p}^c_{ij})}{\sqrt{\hat{p}^c_{ij}(1-\hat{p}^c_{ij})}}\frac{(1-t^2)}{3\sqrt{n_i}}\phi(t)+o_p(n_i^{-1/2})\\
&= \Phi(t)+\frac{(1-2\hat{p}_{ij})}{\sqrt{\hat{p}_{ij}(1-\hat{p}_{ij})}}\frac{(1-t^2)}{3\sqrt{n_i}}\phi(t)+o_p(n_i^{-1/2}),
\end{aligned}
$$

for $\hat{p}^c_{ij}=1+O(n_i^{-1})$.

Note also that the posterior variance $V^c=\hat{p}^c_{ij}(1-\hat{p}^c_{ij})/(\alpha_i+n_i+1)$, so $\sqrt{n_i b}\sqrt{V^c}=\sqrt{\alpha_i+n_i/\alpha_i+n_i+1}=1+o(n_i^{-1/2})$. Then, as $n\to\infty$, the distributions of $(p_{ij}-\hat{p}^c_{ij})/\sqrt{V^c}$ and $\sqrt{n_i b}(p_{ij}-\hat{p}^c_{ij})$ have the same expansion for given x (see Remark 2.2 of Weng (1989)), which completes the proof.

Remark 3.1. A referee pointed out that in the matrix beta case, after proper standardization, one essentially deals with sums of independent gamma random variables, and thus by using a classical Edgeworth expansion one obtains the second order results as in Weng (1989).

Remark 3.2. The posterior of p_{ij} with respect to conjugate priors can be factored out by the "flat" prior due to conjugacy. Therefore, after proper centering, the prior effect appearing to the second order is consistent with that given by the "flat" prior as seen in Theorem 3. A referee pointed out that an Edgeworth expansion based on proper centering (posterior mean and standard deviation) for any smooth prior needs to be studied, to see if the prior density affects the second order as much as in Johnson's expansion.

Remark 3.3. In Section 2, we proposed Bayesian bootstrap algorithms to simulate the posterior distributions of Π and T. It should be interesting to invesitgate the asymptotic justification of this procedure. Therefore, for the second order efficiency of the Bayesian bootstrap, we need to investigate for which functionals of P, the Edgeworth expansion of the posterior distribution is independent of the prior up to the second order, as well as the Edgeworth expansion for the Bayesian bootstrap distribution.

It might be interesting to consider other independent random variables instead of the standard exponential, which is in the realm of Bayesian bootstrap clones. In the i.i.d. case, this has been developed by Lo (1991, 1993b). A corresponding result for the Makov chain case based on simulating other independent Z_{it} needs further study.

4. Empirical Studies

A numerical illustration for the Bayesian bootstrap is given in this section. For a 3 state ergodic Markov chain with initial state $X_0 = 1$, we compare the normal approximation and the Bayesian bootstrap to approximate confidence intervals for different parameters, which includes transition probability p_{12}, stationary probabilities π_1, π_2, π_3, and hitting time T_3.

The original sample is a computer simulation from an ergodic Markov chain with the transition probability matrix

$$P = \begin{pmatrix} .3 & .4 & .3 \\ .2 & .3 & .5 \\ .4 & .4 & .2 \end{pmatrix},$$

and the stationary probability

$$\Pi = (.298,\ .364,\ .338).$$

For this small sample study, two different sample sizes $n = 50, 100$ are included. The Bayesian bootstrap sample size is the same as the original sample size. Here, we simulate 95% confidence intervals, their average lengths and empirical coverage probabilities based on 1000 replications Monte-Carlo trials. The results are presented in Tables 1 and 2. Figures 1 and 2 show the graphs of the real posterior density of $(p_{12} - \hat{p}_{12})|\underline{x}$ and the approximated posterior based on the Bayesian bootstrap procedure for the corresponding samples.

The Computations were performed using FORTRAN programs on a Sun Sparc II workstation. The random numbers were generated using IMSL routines. All the tests were compared on the basis of the same random numbers. Samples with different sizes were nested.

The abbreviation notations will be used in the tables as follows:

NA – normal approximation BB – Bayesian bootstrap

C.I. – confidence interval A.L. – average length

C.P. – coverage probability

Table 1. The comparison of approximate confidence intervals with $n = 50$

	NA			BB		
	95% C.I.	A.L.	C.P.	95% C.I.	A.L.	C.P.
p_{12}	(.209, .561)	.352	.950	(.188, .566)	.377	.962
π_1	(.218, .414)	.196	.957	(.209, .415)	.206	.984
π_2	(.260, .450)	.190	.963	(.246, .459)	.213	.976
π_3	(.251, .424)	.174	.970	(.231, .441)	.210	.996

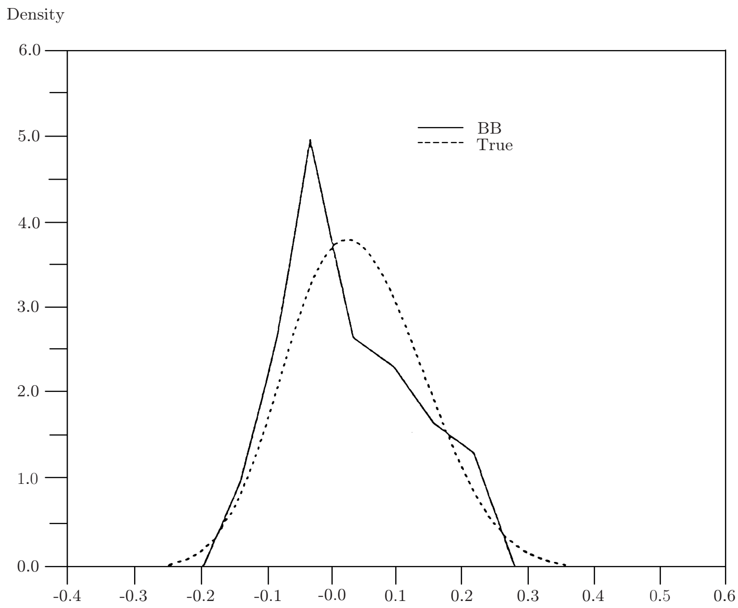

Figure 1. Comparison of the posterior density of $(p_{12} - \hat{p}_{12})|\mathbf{x}$ (dotted line) and the estimated posterior obtained by the BB procedure with $n = 50$.

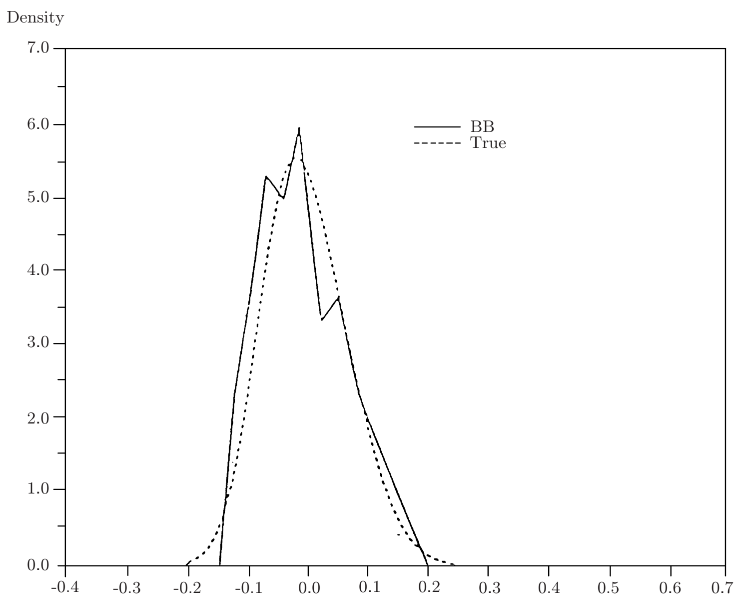

Figure 2. Comparison of the posterior density of $(p_{12} - \hat{p}_{12})|\mathbf{x}$ (dotted line) and the estimated posterior obtained by the BB procedure with $n = 100$.

Table 2. The comparison of approximate confidence intervals with $n = 100$

	NA			BB		
	95% C.I.	A.L.	C.P.	95% C.I.	A.L.	C.P.
p_{12}	(.241, .538)	.296	.956	(.234, .534)	.301	.961
π_1	(.230, .388)	.159	.962	(.226, .386)	.160	.979
π_2	(.281, .434)	.153	.962	(.272, .438)	.166	.972
π_3	(.272, .408)	.137	.971	(.257, .421)	.164	.992

Let T_3 be the first hitting time up to state 3 of the Markov chain with $X_0 = 1$ and $Pr(t; P) \equiv Pr(T_3 \leq t | X_0 = 1, P)$ be the probability that $T_3 \leq t$ for $t = 1, 2, \ldots$ For any 3×3 stochastic matrix P, let $A \equiv A(p)$ be the stochastic matrix which is the same as P except that the 3^{rd} row is replaced by $(0, 0, 1)$, so $Pr(t; P) = (A^t)_{1,3}$. The Bayesian bootstrap to estimate the posterior distribution of $Pr(t; P)$ given $\underline{x}$ is $Pr(t; \hat{P}_n^*)$.

The following abbreviation notation will be used in Figures 3 and 4 which give the comparison of the real confidence band and those approximated by the Bayesian bootstrap procedure.

– – true distribution	T – true confidence band
B – Bayesian bootstrap confidence band	A.L. – average length

Here, A.L. is computed for each $t = 1, 2, \ldots, 30$, and then the average taken

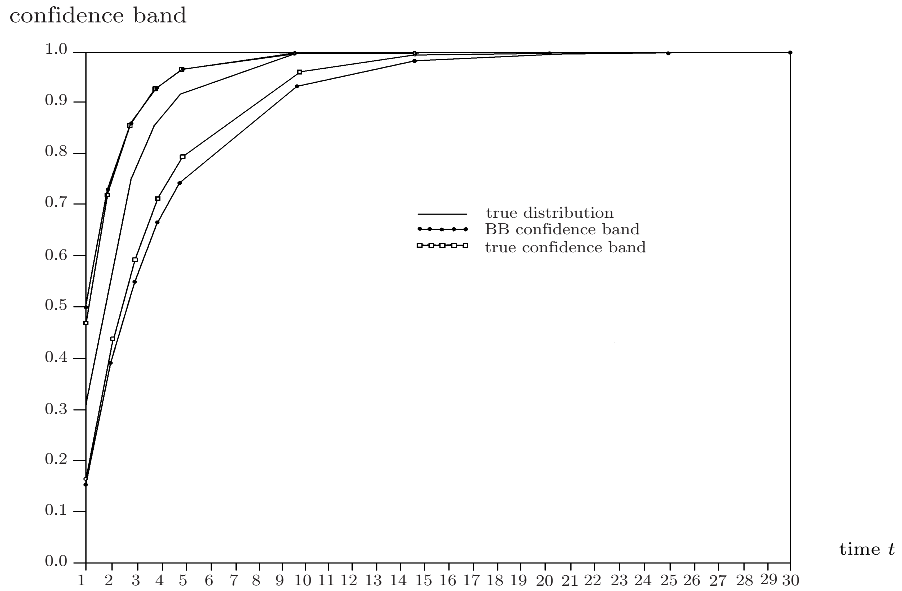

Figure 3. The BB approximate confidence band for T_3, $n = 50$ and A.L. = .154

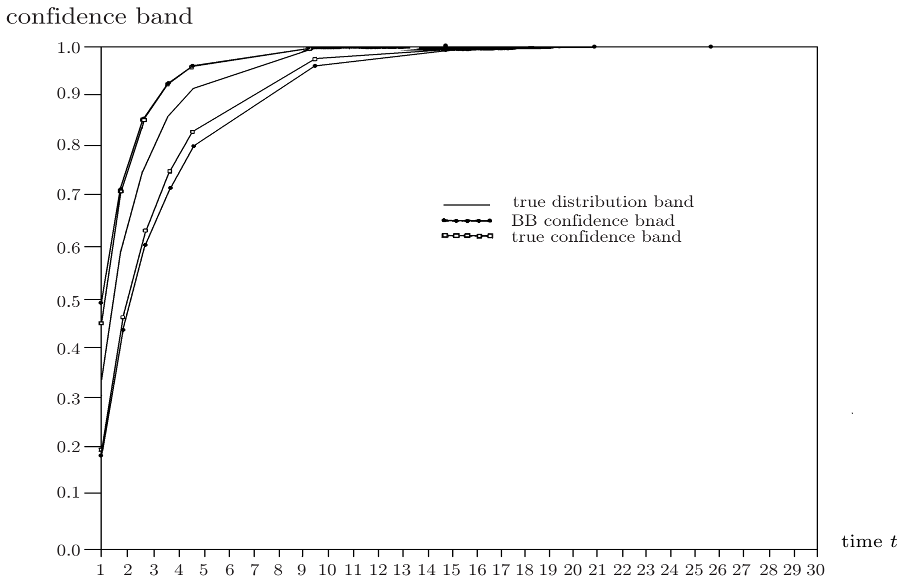

Figure 4. The BB approximate confidence band for T_3, $n = 100$ and A.L. $= .121$

Acknowledgements

The authors would like to thank Dr. Albert Y. Lo for numerous stimulating discussions and suggestions and Mr. Deng-Hsiang Lee for the computer simulation. The work of C. D. Fuh is partially supported by NSC 85-2121-M001-008 and the work of T. H. Fan is partially supported by NSC 85-2121-M008-003.

References

Athreya, K. B. and Fuh, C. D. (1992). Bootstrapping Markov chains: countable case. *J. Statist. Plann. Inference* **33**, 311-331.

Billingsley, P. (1961). *Statistical Inference for Markov Processes.* The University of Chicago Press, Chicago.

Efron, B. (1979). Bootstrap method: another look at the jackknife. *Ann. Statist.* **7**, 1-26.

Fuh, C. D. (1993). Statistical inquiry for Markov chains by bootstrap method. *Statist. Sinica* **3**, 53-66.

Ghosh, J. K. (1994) *Higher Order Asymptotics.* NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 4.

Hjort, J. F. K. (1991). Bayesian and empirical Bayesian bootstrapping. Statistical Research Report, University of Oslo.

Johnson, R. A. (1970). Asymptotic expansions associated with posterior distributions. *Ann. Math. Statist.* **41**, 851-864.

Kulperger, R. J. and Prakasa Rao, B. L. S. (1989). Bootstrapping a finite state Markov chain. *Sankhya* **51**, 178-191.

Levit, V. Y. (1974). On the behavior of generalized Bayes estimators for Markov observations. *Theory. Probab. Appl.* **19**, 332-346.

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

Lo, A. Y. (1988). A Bayesian bootstrap for a finite population. *Ann. Statist.* **16**, 1684-1695.

Lo, A. Y. (1991). Bayesian bootstrap clones and a biometry function. *Sankhya* **53**, 320-333.

Lo, A. Y. (1993a). A Bayesian bootstrap for censored data. *Ann. Statist.* **21**, 100-123.

Lo, A. Y. (1993b). A Bayesian method for weighted sampling. *Ann. Statist.* **21**, 2138-2148.

Martin, J. J. (1975). *Bayesian Decision Problems and Markov Chains.* Robert E. Krieger Publishing Company, Huntington, New York.

Reiss, R. D. (1989). *Approximate Distribution of the Order Statistics: with Applications to Nonparametric Statistics.* Springer-Verlag, Berlin.

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

Weng, C. S. (1989). On a second-order asymptotic property of the Bayesian bootstrap mean. *Ann. Statist.* **17**, 705-710.

Wilks, S. S. (1962). *Mathematical Statistics.* Wiley, New York.

Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan.

E-mail: stcheng@ccvax.sinica.edu.tw

Institute of Statistics, National Central University, Chung-li, Taiwan.

E-mail: thfan@stat1.ncu.edu.tw

(Received May 1993; accepted December 1996)