Sankhya : The Indian Journal of Statistics
1989, Volume 51, Series A, Pt. 2, pp. 178–191.

BOOTSTRAPPING A FINITE STATE MARKOV CHAIN

By R. J. KULPERGER
University of Western Ontario, London, Canada
and
B. L. S. PRAKASA RAO
University of California, Davis, U.S.A.

SUMMARY. Consider a finite state discrete time ergodic Markov chain $\{X_n, n \geq 0\}$ with unknown transition matrix P and state space $S = \{1, ..., m\}$. The matrix P can be estimated by $\hat{P}_n$, the matrix of empirical transition rates. An estimaror of the stationary distribution π or the distribution of the first hitting time of some state, say m, can be obtained by using $\hat{P}_n$. However exact sampling distributions of the estimators are in general difficult to calculate. Here we use the boot strap technique to obtain approximation to these sampling distributions. It is shown that the boot-strap technique works in an asymptotic sense. Results are illustrated by some simulations.

1. Introduction

Let $X_1, ..., X_n$ be independent and identically distributed random variables with distribution function F. Denote the vector $(X_1, ..., X_n)$ by X. Let $x = (x_1, ..., x_n)$ be an observed realization of X. In general, it is difficult to find the exact distribution of a statistic $T(X)$ of X, even if F is known, except in some special cases. If F is unknown, then the exact distirbution of a random variable $R(X, F)$, depending both on X and the unknown distribution F, cannot be computed. The problem is how to find a "good" approximation to the distribution of $T(X)$ if F is known or to the distribution of $R(X, F)$ if F is unknown.

In a series of papers starting from 1979, Efron proposed the following approach known as the "bootstrap" method. It can be described in the following manner :

(*i*) Construct $\hat{F}_n$, an estimate of the probability distribution F, based on the observed relization x. Usually this is the empirical distribution function.

(*ii*) Draw a random sample of size n from a population with distribution function $\hat{F}_n$. Denote this sample by $X_i^* = x_i^*$, $1 \leqslant i \leqslant n$. Here X_i^*, $1 \leqslant i \leqslant$

AMS (1980) *subject classification :* 62M05, 62G05.
Key words and phrases : Bootstrap ; Finite state Markov chain.

n are i.i.d. with distribution F_n. The sample $x^* = (x_1^*, ..., x_n^*)$ is called the observed bootstrap sample. Let $X^* = (X_1^*, ..., X_n^*)$.

(*iii*) Approximate the distribution of $R(X, F)$ by the bootstrap distribution of $R^* = R(X^*, \hat{F}_n)$.

Efron (1979) has shown that the distribution of R^* approximates that of R under some regularity conditions for a large number of statistics of interest. Several authors have extended this approach to other statisical models.

Now suppose that $X = \{X_n\, n \geqslant 0\}$ is a homogeneous ergodic Markov chain with finite state space with transition matrix $P = (p_{ij})_{k \times k}$ and state space $S = \{1, 2, ..., k\}$, $k \geqslant 2$. Even in such a simple stochastic model, which occurs quite often in practice, it is difficult to compute the exact distribution of estimates of a hitting time or to find an expectation of the hitting time of a state in view of the complexity in the computation.

This paper shows that the bootstrap method is applicable in at least one stochastic process framework. Even for this most simple type of stochastic process, namely, a finite state Markov claim, the complexity of computation of the exact distribution for statistics like the hitting time when the transition matrix is known or unknown leads one to consider the use of the bootstrap method in a stochastic process framework.

As far as we are aware, this paper is the first work which deals with such a problem. It is hoped that this paper will give impetus to further studies of this type.

In the next section, some theoretical aspects and the asymptotic behaviour of bootstrap estimators are discussed. The last section deals with some computational aspects of our results.

Before concluding this section, it is cautioned that the bootstrap technique is not "the man for all seasons," or "the remedy for all diseases." There are situations where bootstrap methods lead to incorrect conclusions (see, for instance, Athreya, 1984). The bootstrap also does not give the proper answer for sample extremes.

2. The bootstrap method for Markov chains

Let $\{X_n, n \geqslant 0\}$ be an ergodic Markov chain with transition matrix $P = (p_{ij})$ and state spacz $S = \{1, 2, ..., k\}$, $k \geqslant 2$. Assume that $p_{ij} > 0$ for all i and j. This assumption can be dispensed with for regular Markov chains by enlarging the state space if necessary, since there exists integer $N > 1$ such that $P^N = (P_{ij}^{(N)})$ has all entries positive.

Let $x = (x_0, x_1, \ldots, x_n)$ be a realization of the process observed up to time n. Estimate P by $\hat{P}_n = (\hat{p}_{ij})$, where

$$\begin{aligned} p_{ij} &= \frac{n_{ij}}{n_i} \qquad \text{if } n_i > 0 \\ &= \delta_{ij} \qquad \text{if } n_i = 0, \end{aligned} \qquad \ldots \quad (2.1)$$

$$n_{ij} = \text{observed number of transitions from state } i \text{ to state } j \text{ in } \{x_0, x_1, \ldots, x_n\}$$

and

$$n_i = \sum_{j=1}^{k} n_{ij}.$$

It is well known that

$$n^{\frac{1}{2}}(\hat{p}_{ij} - p_{ij}) \to Z_{i,j} \; 1 \leqslant i, j \leqslant k \qquad \ldots \quad (2.2)$$

in distribution, where $Z_{i,j}$ $i, j \in S$ have a joint multivariate normal distribution with mean 0 and covariance which is continuous in P (see Billingsley, 1961, or Basawa and Prakasa Rao, 1980). The rate of convergence in (2.2) is uniform in a neighbourhood of P (Sirazhdinov and Formann, 1983; Lifshits, 1978; or Nagaev, 1957). When the dependence of Z on the transition matrix P is explicitly required, write $Z_{i,j} = Z_{i,j}(P)$.

Without loss of generality, assume that the initial state for the Markov chain is 1, that is, $x_0 = 1$. Consider the distribution of the first hitting time of state k. Let

$$T = \inf\{n > 0 : X_n = k\} \qquad \ldots \quad (2.3)$$

(define $T = \infty$ if there is no such n). Let $Pr(t\,;\,Q) = Pr(T \leqslant t \mid X_0 = 1\,;\,Q)$ denote the probability that $T \leqslant t$, for $t \in \{1, 2, \ldots\}$, for a Markov chain with transition matrix, Q and initial value $X_0 = 1$.

The bootstrap method of estimation of the distribution of the hitting time T may be explained as follows. $Pr(t\,;\,\hat{P}_n)$ estimates $Pr(t\,;\,P)$, and therefore an interesting question is to find the distribution of

$$\sqrt{n}(Pr(t\,;\,\hat{P}_n) - Pr(t\,;\,P)),\, t \geqslant 1. \qquad \ldots \quad (2.4)$$

When P is known, the distribution of (2.4) can be obtained. A Monte Carlo method could be used, since the distribution is a complicated function of P.

When P is not known, the bootstrap procedure is to approximate the distribution of (2.4) by the distribution of

$$G_n(t\,;Q) = \sqrt{n}(\hat{P}r(t\,;Q_n) - Pr(t\,;Q)), \qquad \dots \quad (2.5)$$

where $\hat{Q}_n$ is given by formula (2.1), as computed from the subsample generated from a Markov chain with transition matrix $\hat{P}_n$, and Q is a transition matrix close to P, namely $Q = \hat{P}_n$. The distribution of (2.5) may be computed since $\hat{P}_n$ is a known matrix, for example by a Monte Carlo method. The problem is to justify that the distributions of (2.5) and (2.4) are close.

Adapting the same type of analysis, one can estimate $m_1(P) = E_P(T \mid X_0 = 1, P)$, the expectation of the first hitting time or also study the estimation of $\pi(P)$, where $\pi(P)$ is the stationary distribution of the process. For example if $m_1(P)$ is estimated by $m_1(\hat{P}_n)$, then the interest is in estimating the distribution of

$$\sqrt{n}(m_1(\hat{P}_n) - m_1(P)). \qquad \dots \quad (2.6)$$

For any square matrix $A = (a_{ij})$, define

$$\| A \| = \sum_{i,j} |a_{ij}|. \qquad \dots \quad (2.7)$$

Note that $\| P \| = k$ for any stochastic matrix of order k.

We shall show that the distributions of (2.5) and (2.6) have suitable continuity properties in the topology defined by the norm (2.7).

Observe that the distribution of the hitting time for the state k given that $X_0 = i$, $i \neq k$ is the same as the distribution of the absorption time in the modified Markov chain with transition matrix P replaced by

$$A = A(P) \qquad \dots \quad (2.8)$$

is given by P in which the last (k^{th}) row is replaced by $(0, \dots, 0, 1)$. In this chain k is an absorbing state and $J = \{1, 2, \dots, k-1\}$ is a set of transient states.

Consider the matrix valued stochastic process, indexed by $t \geqslant 1$,

$$D_n(t\,;P) = \sqrt{n}(\hat{A}_n^t - A^t), \qquad \dots \quad (2.9)$$

where $\hat{A}_n = A(\hat{P}_n)$. Notice $Pr(T \leqslant t\,;P) = (A^t)_{1,k}$, so that studying continuity properties of (2.9) will yield some appropriate continuity properties of the stochastic processes given by (2.5). The sequence $Pr(T \leq t\,; P), t \geq 1$, can be shown to be continuous in P by showing that the sequence $A^t, t \geq 1$, is continuous in A, by using Lemma A. 1.

Let H be the matrix sequence-valued function defined in the Appendix (after Definition A.1). Thus, for $t \geq 1$,

$$D_n(t, P) = H(t, \hat{A}_n, \sqrt{n}(\hat{A}_n - A)) + E(t, \hat{A}_n, \sqrt{n}(\hat{A}_n - A), A), \qquad \ldots \quad (2.10)$$

where

$$E(t, A_0, V, A) = \sum_{q=0}^{t-1} A_0^{t-1-q} V (A^q - A_0^q).$$

Let $V = V(Q) = (V_{i,j}(Q))$, where $V_{i,j}(Q) = Z_{i,j}(Q), i, j \in J$, with $Z_{i,j}(Q)$ given by (2.2), and $V_{k,j} = 0, j \in S$. Then

$$H(t, A(P), V(Q)) = \sum_{q=0}^{t-1} A^{t-1-q} V A^q, t \geq 1 \qquad \ldots \quad (2.11)$$

is a process whose distribution depends on P, and on Q through the distribution of V.

Lemma 2.1 :

(*i*) $\|H(t, \hat{A}_n, \sqrt{n}(\hat{A}_n - A)) - H(t, A, \sqrt{n}(\hat{A}_n - A))\| \leq K_1 \| \sqrt{n}(\hat{A}_n - A)\| \|\hat{A}_n - A\|$ *for some constant* $K_1 = K_1(A)$. *Since* $\hat{A}_n$ *is a* $\sqrt{n}$ *consistent estimate of* A, *the bound is* $O_p(1/\sqrt{n})$ *uniformly in* t.

(*ii*) *lim sup* $\|H(t, A(P), V(Q))\| = 0$ *in probability, uniformly in* Q *such that* $c(A(Q)) > \delta > 0$ *for some fixed* $\delta > 0$. ($t \to \infty$)

(*iii*) $\sup_{t \to \infty} E(t, A(P), V(Q), A(Q)) \to 0$ *in probability, uniformly in* Q.

Proof : (i) follows directly from the Appendix.

(ii) Let $B = (b_{i,j})$ with $b_{i,j} = 0$ if $j \neq k$ and $b_{i,k} = 1$. Then Lemma A.1 gives

$$\|A^t - B\| \leq K(1 - c(A))^t$$

for some constant K. Notice also that $VB = 0$, since $\sum_{j \in S} p_{i,j} = 1$. Also $BV = 0$ Therefore

$$\|H(t, A(P), V(Q))\| \leq \sum_{q=0}^{t-1} \|A^{t-1-q} - B\| \|V\| \|A^q - B\|$$

$$\leq K^2 \|V\| t (1 - c(A))^t \to 0 \text{ as } t \to \infty. \qquad \ldots \quad (2.12)$$

(ii) As in (ii)

$$\|E(t, A_0, V, A)\| \leq \sum_{q=0}^{t-1} \|V\| \|A^q - A_0^q\|$$

$$\leq \|V\| \sum_{q=0}^{t-1} \{\|A^q - B\| + \|A_0^q - B\|\}$$

$$\leq \|V\| K^2 \sum_{q=0}^{t-1} (1-c(A_0))^{t-1}$$

$$+ \|V\| K^2 \sum_{q=0}^{t-1} (1-c(A_0))^{t-1-q} (1-c(A))^q$$

$$\leq \|V\| K^2 t(1-c(A_0))^{t-1} + K^2 \|V\| \sum_{q=0}^{\left[\frac{t}{2}\right]} (1-c(A_0))^{t-1-q}$$

$$+ K^2 \|V\| \sum_{q=\left[\frac{t}{2}\right]}^{t-1} (1-c(A))^q$$

$$\leq K_1 \|V\| \left\{ t(1-c(A_0))^{t-1} + (1-c(A_0))^{\left[\frac{t}{2}\right]-1} + (1-c(A))^{\left[\frac{t}{2}\right]-1} \right\}$$

for a constant K_1. For t large, the upper bound is uniformly small in t. For $t = 0, 1, \ldots t_0$ for any finite t_0 the upper bound is small if A_0 and A are close. Thus (iii) follows.

Consider two Markov chains, both with state space S, and with transition matrices P and O respectively. Recall from (2.2) that the distribution of $Z_{i,j}(P)$ is continuous in P. The following can be proved by Lemma 2.1 and the fact that H is continuous

Let $\mathcal{L}(H(A(\hat{Q}_n)), \sqrt{n}(A(\hat{Q}_n) - A((Q)) ; Q))$ denote the law of $H(A(\hat{Q}_n)$, $\sqrt{n}(A(\hat{Q}_n) - A(Q)))$ when the matrix of the Markov chain is Q. By Lemma 2.1 (iii), the process $\{D_n(t, P) \geq 1\}$ has the same asymptotic distribution as $\{H(t, \hat{A}_n, \sqrt{n}(\hat{A}_n - A)), t \geq 1\}$.

Theorem 2.1 : *For E, an $\mathcal{L}(H(A(P), Z(P)) ; P)$ continuity set and for any positive real sequence $\delta_n \to 0$,*

$$\lim_{n\to\infty} \sup_{\|Q-P\| \leq \delta_n} | \mathcal{L}(H(A(\hat{Q}_n), \sqrt{n}(A(\hat{Q}_n) - A(Q)); Q)(E)$$

$$- \mathcal{L}(H(A(\hat{P}_n), \sqrt{n}(A(\hat{P}_n) - A(P)) ; P)(E) | = 0.$$

Proof: By Lemma 2.1, Lemma A.3 and (2.2),

$$H(t, A(\hat{Q}_n), \sqrt{n}(A(\hat{Q}_n)-A(Q)))$$
$$= H(t, A(P), \sqrt{n}(A(\hat{Q}_n)-A(Q)))$$
$$+O_p(\|A(\hat{Q}_n)-A(Q)\|)+O(\|A(Q)-A(P)\|),$$

where the $O_p(.) \to 0$ uniformly in Q near P. Since $\sqrt{n}(A(\hat{Q}_n)-A(Q)) \to Z(Q)$ in distribution, uniformly in Q in a neighbourhood of P,

$$H(t, A(P), \sqrt{n}(A(\hat{Q}_n)-A(Q))) \to H(t, A(P), Z(P))$$

jointly for finite collections of t's, where $\|Q-P\| \leqslant \delta_n$, and $\delta_n \to 0$ as $n \to \infty$. Lemma 2.1 (ii) extends this to all $t \geqslant 1$.

Theorem 2.1 says that the bootstrap method works for the problem of estimating the distribution of the estimate of the hitting time distribution given by (2.5). This is interesting since (2.5) is an infinite dimensional vector, namely an element of $\mathscr{M}_\infty$ given in the Appendix.

Next consider $\hat{\pi}_n$, the estimate of the stationary distribution of P. For a $k \times k$ stochastic matrix Q, let $F(Q)$ be the $k \times k$ matrix whose first $k-1$ columns are the first $k-1$ columns of Q, and whose k^{th} column is $\mathbf{1}$, a column vector of all 1's and of size k. Let e be a $1 \times k$ matrix of all zero's, except for a 1 in the last position.

Note that for regular Markov chains $F(Q)$ is invertible. The vectors π and $\hat{\pi}_n$ satisfy

$$\pi F(P) = e \text{ and } \hat{\pi}_n F(\hat{P}_n) = e.$$

Therefore,

$$0 = \hat{\pi}_n F(\hat{P}_n) - \pi F(P)$$
$$= \hat{\pi}_n (F(\hat{P}_n)-F(P)) + (\hat{\pi}_n - \pi) F(P). \quad \ldots \quad (2.13)$$

Note $\sqrt{n}(F(\hat{P}_n)-F(P))$ converges uniformly, in a neighbourhood of P, to a limit distribution which is continuous in P. This, together with (2.13), gives the following theorem, which says the distribution of $\sqrt{n}(\hat{\pi}_n - \pi)$ is continuous in P, and thus the bootstrap method works for approximating the distribution of $\sqrt{n}(\hat{\pi}_n - \pi)$.

Theorem 2.2 : *Let X be the Markov chain described above, and let Q be another transition matrix such that* $\|Q-P\| \leqslant \delta$, *where* δ *is sufficiently small* *so that a Markov chain with transition matrix Q is ergodic. Let* $\hat{\pi}_n$ *be an estimate of* π *satisfying* $\hat{\pi}_n = \hat{\pi}_n \hat{P}_n$, *and let* $\mathcal{L}(\sqrt{n}(\hat{\pi}_n - \pi); P)$ *and* $\mathcal{L}(\sqrt{n}(\hat{\pi}_n - \pi); Q)$ *be the distributions of* $\sqrt{n}(\hat{\pi} - \pi)$ *when the transition matrices are P and Q respectively. Then, for given P, and for* $\delta_n \to 0$,

$$\lim_{n \to \infty} \sup_{\|Q-P\| < \delta_n} |\mathcal{L}(\sqrt{n}(\hat{\pi}_n - \pi); P)(E) - \mathcal{L}(\sqrt{n}(\hat{\pi}_n - \pi); Q)(E)| \to 0, \quad \ldots \quad (2.14)$$

where E is a $\lim_{n \to \infty} \mathcal{L}(\sqrt{n}(\hat{\pi}_n - \pi); P)$ *continuity set.*

Similarly, certain other estimators of the form $f(\hat{P}_n)$, where f is continuous in P, can be considered. One other such example is $m_i(P) = E(T \mid X_0 = i; P)$, where T is given by (2.3). In this case, the functional is given as follows. Let $R = R(P)$ be $(k-1) \times (k-1)$ matrix, with $R_{i,j} = P_{i,j}$ $i, j \in J = \{1, \ldots, k-1$. Let $\mathbf{m} = (m_1, \ldots, m_{k-1})^t$ and $\mathbf{1} = (1, \ldots, 1)^t$ be column vectors of dimension $k-1$. Then

$$m(P) = (I - R(P))^{-1} \mathbf{1}.$$

The matrix $I - R(P)$ is invertible (Kemeny and Snell, 1960). An estimator of $m(\hat{P})$ is then $m(P_n)$. The distribution to be considered is that of

$$\sqrt{n}(m(\hat{P}_n) - m(P)). \quad \ldots \quad (2.15)$$

It can be shown that the distribution of (2.15) satisfies a continuity analogue of (2.14) in Theorem 2.2. This can be done by using (2.2) and noting that (2.15) has a continuous derivative with respect to P. Therefore the bootstrapping technique works for the distribution of $m(\hat{P}_n)$.

There is a natural estimator of π, the stationary distribution of P, that is not of the form $f(\hat{P}_n)$. Let $N_i(n) = \sum_{m=0}^{n} I(X_m = i)$ be the number of visits to state i up to time n. Then $\bar{\pi}_n = (\bar{\pi}_{1,n}, \ldots, \hat{\pi}_{k,n})$ given by $\bar{\pi}_{i,n} = \frac{N_i(n)}{n}$ is an estimate of π. For any posotive regular P,

$$\sqrt{n}(\hat{\pi}_n - u) \xrightarrow{\mathcal{L}} N(0, \Sigma(P)).$$

This convergence can be shown to be uniform in a neighbourhood of P. The boot strap technique works for estimating the distribution of the distribution of $\bar{\pi}_n$.

3. Numerical examples

In this section, a Markov chain is simulated, and the bootstrap procedure is applied. The Markov chain has state space $S = \{1, 2, 3, 4, 5\}$ and transition matrix

$$P = \begin{bmatrix} .1 & .2 & .3 & .3 & .1 \\ .5 & .1 & .1 & .1 & .2 \\ .1 & .5 & .1 & .1 & .2 \\ .1 & .2 & .5 & .1 & .1 \\ .2 & .2 & .1 & .4 & .1 \end{bmatrix}.$$

The distribution of the hitting time random variable (2.3) is shown in Figure 2. The stationary distribution π and the mean hitting times $m_i = E(T \mid X_0 = i)$ are given in Table 2.

TABLE 1. SIMULATED MARKOV CHAIN ($n = 50$)

1 3 5 4 4 3 2 5 4 1 2 1 1 3 5 4 3 2 1 5 4 4 3 5 4 3 4 3 4 3 2 1 2 5 4 2 1 3 3 2 2 4 3
5 1 2 1 3 2 5 4

A sample of size $n+1$ observations is simulated, where $n = 50$ and 100. $X_0 = 1$ in all the simulations. Table 1 shows the data for the case $n = 50$. For this data set, the estimated transition matrix is

$$\hat{P}_n = \begin{bmatrix} .1111 & .3333 & .4444 & .0000 & .1111 \\ .5000 & .1000 & .0000 & .1000 & .3000 \\ .0000 & .4167 & .0833 & .1667 & .3333 \\ .0909 & .0909 & .6364 & .1818 & .0000 \\ .1250 & .0000 & .0000 & .8750 & .0000 \end{bmatrix}.$$

$\hat{\pi}_n$ and $\hat{m}_n$ are given in Table 2, and $Pr(T < t\,;X_0, \hat{P}_n)$ is shown in Figure 2. The distribution of

$$K_n = \sup_{t \geq 0} |\sqrt{n}(Pr(T < t\,; X_0, \hat{P}_n) - Pr(T < t\,; X_0, P))|$$

is given for the correct or original P and for the bootstrap estimate using $\hat{P}_n$. These estimates are obtained by a Monte Carlo approximation, using 500 simulations, and taking $X_0 = 1$ in the simulations. The true and estimated distributions are almost the same, even for a small sample of size $n = 50$, in which there is an average of only 10 observations per state (see Figure 1). The .95 quantile of the bootstrapped estimate is used to obtain a confidence band for the hitting time distribution of T. These are shown in Figures 2, 3. Marginal confidence intervals for π_i and m_i are obtained by first obtaining the distributions of $\sqrt{n}(\pi_{i,n}-\hat{\pi}_i)$ and $\sqrt{n}(\hat{m}_{i,n}-m_i)$. These distributions are obtained by a Monte Carlo method, by simulating a Markov chain with transition matrices P and $\hat{P}_n$ respectively, with $X_0 = 1$. The .025 and .975 quantiles of these Monte Carlo estimates are used to obtain 95% confidence intervals for π_i and m_i. These intervals are given in Tables 2, 3. The distance between the true and bootstrap distributions is given by

$$d(\hat{F}, F) = \max\left(|\hat{F}(x)-F(x)| :-\infty < x < \infty\right)$$

where $\hat{F}$ and F are the bootstrap and true distributions of the respective statistics. These numbers are given in the last columns of Tables 2 and 3.

The bootstrap distributions and confidence intervals are very good approximations for K_n and the π_i cases, even for $n = 50$. The distribution of $\sqrt{n}(\hat{m}_{i,n}-m_i)$ is very skewed, and has a geometric tail, and is more tightly packed near zero when the mean of the distribution is nearer to zero. In the simulation presented here, $\hat{m}_{i,n}$ is smaller than m_i, and so the bootstrapped

TABLE 2. ($n=50$)

i	π_i	$\hat{\pi}_{i,n}$	True 95% C.I.	Bootstrapped 95% C.I.	$d(\hat{F},F)$
1	.2109	.1675	(.0682, .2406)	(.0570, .2005)	.098
2	.2407	.1903	(.0086, .2853)	(.0098, .2825)	.058
3	.2165	.2450	(.1630, .3285)	(.1600, .3224)	.082
4	.1859	.2430	(.1570, .3340)	(.1448, .3456)	.120
5	.1457	.1581	(.0723, .2548)	(.0070, .2204)	.150

i	m_i	$\hat{m}_{i,n}$	True 95% C.I.	Bootstrapped 95% C.I.	$d(\hat{F}, F)$
1	6.7402	4.7972	(1.8114, 14.2793)	(2.6083, 8.8450)	.148
2	6.2677	4.3763	(1.1086, 13.9189)	(2.2198, 9.2810)	.130
3	6.0787	4.0621	(0.8376, 13.8484)	(2.2306, 8.4830)	.150
4	6.6299	5.4009	(2.2211, 15.5130)	(3.5443, 9.6122)	.154

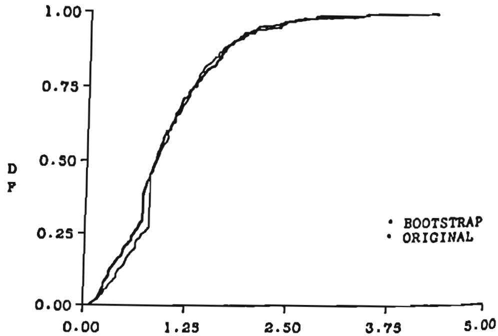
TABLE 3. ($n = 100$)

i	π_i	$\hat{\pi}_{i,n}$	True 95% C.I.	Bootstrapped 95% C.I.	$d(\hat{F},F)$
1	.2109	.1904	(.1204, .2595)	(.1300, .2700)	.056
2	.2400	.2512	(.1803, .3130)	(.1877, .3124)	.068
3	.2165	.2469	(.1803, .3103)	(.1878, .2089)	.058
4	.1859	.1803	(.1144, .2449)	(.1108, .2489)	.084
5	.1457	.1313	(.0656, .1989)	(.0797, .1991)	.104

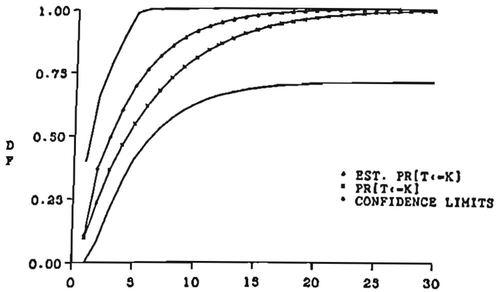
i	m_i	$\hat{m}_{i,n}$	True 95% C.I.	Bootstrapped 95% C.I.	$d(\hat{F},F)$
1	6.7402	7.4030	(5.0578, 13.2388)	(4.7865, 13.4937)	.094
2	6.2677	6.7705	(4.2024, 12.2002)	(4.0356, 12.6409)	.090
3	6.0787	6.4740	(4.0763, 12.5000)	(3.7946, 12.4375)	.114
4	6.6299	7.2940	(4.8285, 13.5610)	(4.7189, 13.2432)	.110

95% confidence interval is smaller than the true 95% C.I., as well as being shifted towards zero. In other simulations, when $\hat{m}_{i,n}$ is bigger than m_i the bootstrapped C.I. is wider than, and shifted farther away from zero than the true C.I. The bootstrap C.I.'s improve as n becomes larger.

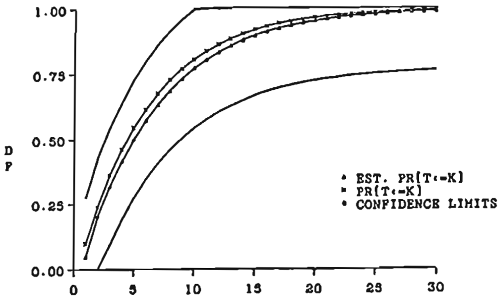
The estimator $\tilde{\pi}$ is not given in the tables, since $\tilde{\pi}$ and $\hat{\pi}$ were equal up to first 3 decimal places in the various simulation runs. A sample of size $n = 200$ was also performed, and gave similar results, with correspondingly smaller confidence bands.



Emporical Distributions Original VS Bootstrap



95% Confidence Band for $PR[T=K]$



95% Confidence Band for $PR[T=K]$

Appendix

This appendix contains a technical Lemma used in section 2.

Suppose $A=(a_{i,j})$ is a stochastic matrix on $S=\{1, 2, \ldots, k\}$, and that k is a single absorbing state. Let $J=\{1, \ldots, k-1\}$. A has an unique stationary distribution, $\gamma=(0, \ldots, 0, 1)$. It is wellknown that $A_n=(a_{i,j}^{(n)}) \rightarrow B$, where B is a stochastic matrix which has all rows equal to γ (Parzen, 1962).

Lemma A.1 : *Suppose* $c = \min\ (a_{i,k} : i \in J) > 0$ *and* $d = \max\ (a_{ij} : i, j \in J) < 1$. *Then* $A_n \to B$ *geometrically fast. In particular*

(i) $a_{i,j}^{(n)} \leqslant (1-c)^n d,\ j \in J$ *and*

(ii) $1-a_{i,k}^{(n)} \leqslant (1-c)^n (k-1) d$.

Proof : Let $j \in J$, and

$$M_j(n) = \max\ (a_{i,j}^{(n)} : i \in J).$$

Let i_0 be such that $M_j(n+1) = a_{i_0,j}^{(n+1)}$. Notice $a_{k,j}^{(n)} = 0$. Then

$$\begin{aligned} M_j(n+1) = a_{i,j}^{(n+1)} &= \Sigma_{l \in J} a_{i_0,l} a_{l,j}^{(n)} \\ &\leqslant (\Sigma_{l \in J} a_{i_0,l}) M_j(n) \\ &= (1-a_{i_0,k}) M_j(n) \leqslant (1-c) M_j(n). \end{aligned}$$

Part (*i*) now follows easily. Part (*ii*) follows from (*i*) since $\Sigma_{j \in S} a_{i,j}^{(n)} = 1$.

Notice that $c = c(A)$ and $d = d(A)$ are continuous functions of A.

The following Lemma is easy to prove by induction.

Lemma A.2 : *Suppose that* A *and* B *are* $k \times k$ *matrices. Then for* $t \geqslant 1$,

$$A^t - B^t = \sum_{q=0}^{t-1} A^{t-1-q}(A-B)B^q.$$

Definition A.1 : (i) Let $\mathcal{M}_1$ be the set of $k \times k$ stochastic matrices A such that the k^{th} row is $(0, \ldots, 0, 1)$.

(ii) Let $\mathcal{M}_2$ be the set of $k \times k$ matrices V such that $V_{kj} = 0,\ j \in S$ and $\sum_{j=1}^{k} V_{i,j} = 1$ for $i \neq k$.

(iii) Let $\mathcal{M}_\infty$ be the set of sequences $\{A_1, A_2, \ldots\}$ of $k \times k$ matrices with sup norm

$$\|A\|_\infty = \sup_{t \geqslant 1} \|A_t\|.$$

Let $H : \mathcal{M}_1 \times \mathcal{M}_2 \to \mathcal{M}_\infty$ be given by

$$H(A, V) = (H(t, A,\ V) : t \geqslant 1)$$

where

$$H(t, A, V) = \sum_{q=0}^{t-1} A^{t-1-q}\ V A^q.$$

Lemma A.3 : (i) $\|H(t, A_1, V) - H(t, A_2, V)\| \leqslant K_1 \|V\| \|A_1 - A_2\|$ where $K_1 = K_1(A_1)$ and A_2 is in an open neighbourhood of A_1.

(ii) $\|H(t, A, V_1) - H(t,\ A,\ V_2)\| \leqslant K_2 \|V_1 - V_2\|$, where $K_2 = K_2(A)$ is a finite constant depending on A.

The above lemma says that H is a continuous function.

Acknowledgments. This work has been supported by NSERC grant number A5724 for the first author. George Verveniotis is gratefully acknowledged for doing the numerical computations for Section 3. The authors also thank the referee for his careful reading of the manuscript and pointing out the omission of the error process E in (2.10) in the earlier version.

References

Athreya, K. B. (1984): Bootstrap for the mean in the infinite variance case. *IMS Bulletin*, 13, 373..

Basawa, I. V., and Prakasa Rao B. L. S. (1980): *Statistical Inference for Stochastic Processes*, Academic Press, London.

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

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

Kemeny, J. G., and Snell, J. L. (1960): *Finite Markov Chains*, Van Nostrand Reinhold, New York.

Lifshits, B. A. (1978): On the central limit theorem for Markov chains. *Th. Prob. and Appl.*, 23, 279-296.

Nagaev, S. V. (1957): Some limit theorems for stationary Markov chains. *Th. Prob. and Appl.*, 2, 378-406.

Parzen, E. (1962): *Stochastic Processes.*, Holden-Day, San Francisco.

Sirazhdinov, S. Kh., and Formanov, Sh. K. (1983): On estimates of the rate of convergence in the central limit theorem for homogeneous Markov chains. *Th. Prob. and Appl.*, 28, 229-239.

Paper received: *April*, 1987.
Revised: *February*, 1988.