Ann. Inst. Statist. Math.
Vol. 42, No. 2, 253–268 (1990)

BOOTSTRAP IN MARKOV-SEQUENCES BASED ON ESTIMATES OF TRANSITION DENSITY*

M. B. RAJARSHI

Department of Statistics, University of Poona, Pune 411007, India

(Received November 21, 1988; revised March 30, 1989)

Abstract. We develop a Bootstrap method in Markov-sequences. This method is based on kernel estimates of the transition density of the Markov-sequence. It is shown that the Bootstrap estimate of the variance of a statistic which is a function of means, is consistent. We also show that the Bootstrap distributions of mean-like statistics and von Mises differentiable statistics converge to appropriate normal distributions. A few simulation results are reported to illustrate the discussion.

Key words and phrases: Variance estimation, Bootstrap, non-parametrics, Markov-sequences.

1. Introduction

Efron's (1979) Bootstrap is perhaps the most important non-parametric procedure for studying the sampling distribution of a statistic. In the case of independently and identically distributed random variables, it has been shown that Bootstrap has an edge over the traditional normal approximation as well as the Jackknife method of estimation of standard error of an estimator (cf. Efron (1979), Bickel and Freedman (1981), Singh (1981), Beran (1982), Babu and Singh (1983, 1984) and Efron and Tibshirani (1986)).

Recent papers by Freedman (1981) and Bose (1986) discuss Bootstrap procedures in linear stochastic models, such as auto-regressive and moving average models. Here, residuals are regarded as proxies of unknown errors which are independently and identically distributed random variables. Resampling is done by sampling with replacement from the set of standardized residuals. Freedman (1981) shows that in linear dynamic models, Bootstrap estimators of standard errors of estimators are consistent, and that the Bootstrap approximation of distribution of an estimator converges

*Most of this work was carried out when the author was at the Department of Statistics, Pennsylvania State University.

to the usual normal distribution. Bose (1986) extends the results of Singh (1981) and Babu and Singh (1983, 1984) to auto-regressive and moving average models. Carlstein (1986) discusses a very general procedure for estimation of the variance of a statistic obtained from a strong mixing or α-mixing sequence. His procedure is based on sub-series values of the statistic, which parallel the sub-sample or psuedo-values of Hartigan (cf. Carlstein (1986)).

In this paper, we plan to discuss the Bootstrap for Markov-sequences. It is instructive to restate here the Bootstrap principle (see Efron and Tibshirani (1986), Fig. 8). Suppose that we want to study the distribution of a random variable $R(X, P)$ where X denotes the data and P denotes the unknown model. Let us assume that we have succeeded in developing a method of estimation of the model P from the data x. Let this estimate be denoted by $\hat{P}$. Then, we can estimate the distribution of $R(X, P)$ by that of $R(X^*, \hat{P})$, where X^* follows the distribution induced by $\hat{P}$. If necessary, the distribution of $R(X^*, \hat{P})$ can be approximated by Monte Carlo methods.

We now apply the Bootstrap principle to obtain non-parametric procedures in stationary Markov-sequences. (Actually, in many parametric stochastic models, exact distribution theory is either unknown or too complicated to be of any practical use. Further, traditional normal approximation may offer a poor approximation. Therefore, the parametric Bootstrap offers an important alternative to the normal approximation. Results on parametric Bootstrap can be deduced from our results (see Remarks in Section 2).) We observe that the statistical model P associated with a stationary Markov-sequence is completely specified by its transition distribution function viz., $P[X_2 \leq y | X_1 = x]$. We estimate this via estimating the transition density by suitable kernel estimates. The estimate of the transition density or distribution function is then utilized to generate Bootstrap sample paths. A distinct advantage of this approach lies in its potential for relaxing the assumption of linearity made in Freedman (1981) and Bose (1986) and thus, developing truly non-parametric methods in large samples. In this connection, it is important to point out that Bootstrap and other resampling plans may possibly play a more prominent role in stochastic models in view of their mathematical intractability: consider obtaining the asymptotic distribution theory for the two models for which simulation studies are reported in Section 4.

Let $\boldsymbol{X} = \{X_1, X_2, \dots\}$ be a first-order, strictly stationary Markov-sequence with

$$\begin{aligned} F(x) &= P[X_1 \leq x] , \\ F(x,y) &= P[X_1 \leq x,\ X_2 \leq y] , \qquad (1.1) \\ F(y|x) &= P[X_2 \leq y | X_1 = x] . \end{aligned}$$

In (1.1), we have assumed that $F(x,y)$ is an absolutely continuous distribution function on $\mathbb{R}^2$. Derivatives corresponding to functions in (1.1) are denoted by $f(x)$, $f(x,y)$ and $f(y|x)$, respectively. The conditional probability density function $f(y|x)$ would be referred to as the transition density of the sequence $\boldsymbol{X}$ and, as remarked earlier, is the essence of statistical model P of $\boldsymbol{X}$.

Now, suppose that we have observed $x_1, x_2, \ldots, x_n$, a sample path of length n. Let $k(x,y)$ be an appropriately chosen kernel on $\mathbb{R}^2$, regularity conditions of which would be specified later. Then, an estimate of $f(x,y)$ is given by

$$\hat{f}(x,y) = \frac{1}{nh^2} \sum_{i=1}^{n-1} k\left(\frac{x - x_i}{h}, \frac{y - x_{i+1}}{h}\right), \tag{1.2}$$

where $h = h(n)$ is a sequence of reals such that $h(n) \downarrow 0$ and $nh(n) \to \infty$. An obvious estimate of the transition density is given by

$$\hat{f}(y|x) = \frac{\hat{f}(x,y)}{\hat{f}(x)}; \quad \hat{f}(x) = \int \hat{f}(x,y)\,dy\,. \tag{1.3}$$

Having estimated the model P of $\boldsymbol{X}$, the Bootstrap can be performed as follows.

(1) Generate a random variate x_1^* with density $\hat{f}(\cdot)$.

(2) Generate a random variate x_2^* with probability density $\hat{f}(\cdot\,|x_1^*)$.

(3) Repeat (2) to generate $x_r^* \sim \hat{f}(\cdot\,|x_{r-1}^*)$, $r = 3, 4, \ldots, n$.

(4) Repeat (1), (2) and (3) B times, where B is the chosen number of Bootstrap samples.

The rest of the Bootstrap methodology is the same as described in Efron (1979) and Efron and Tibshirani (1986). If we have reason to believe that X_1 does not follow the equilibrium distribution $f(x)$, the above Bootstrap procedure can be modified to start with (2), with $x_1^* = x_1$ for each Bootstrap sample path.

In this paper, we propose to establish that under certain conditions, the above Bootstrap "works". These conditions involve assumptions of a rather weak dependence, as defined through ergodicity and mixing coefficients.

Let $P(x,A) = P[X_2 \in A \,|\, X_1 = x]$ be the (one-step) transition function of the sequence $\boldsymbol{X}$. Let $\mathbb{B}$ denote the Borel σ-field. Let S denote the state-space of $\boldsymbol{X}$. We assume throughout this paper that the following holds:

A1: The transition function $P(x,A)$ satisfies

$$\sup_{x,x' \in S, A \in \mathbb{B}} |P(x,A) - P(x',A)| < 1\,.$$

Assumption A1 is related to the ergodicity coefficient

$$\alpha(P) = 1 - \sup_{x, x' \in S, A \in \mathbf{B}} |P(x, A) - P(x', A)| \tag{1.4}$$

which dates back to Markov. It was rediscovered by Dynkin (cf. Iosifescu and Theodorescu (1969), p. 40), and a good account of $\alpha(P)$ is available in Iosifescu and Theodorescu (1969). We notice that $0 \leq \alpha(P) \leq 1$ and Assumption A1 is equivalent to the condition that $\alpha(P) > 0$. The independence of X_1 and X_2 is equivalent to $\alpha(P) = 1$, and $\alpha(P) = 0$ if and only if for every $\varepsilon > 0$, there exist x and x' such that $P(x, A(x)) = 1$, $P(x', A(x')) = 1$, $P[x, A(x) \cap A(x')] < \varepsilon$ and $P[x', A(x) \cap A(x')) < \varepsilon$. This suggests that $\alpha(P) > 0$ is a mild condition on $\boldsymbol{X}$, since, with $\alpha(P) = 0$, the two conditional probability distributions $P(x, A)$ and $P(x', A)$ do not have any set of common support.

There are some important consequences of the Assumption A1 made above. First, from Iosifescu and Theodorescu's (1969) Theorem 2.1.35, it follows that there exists a unique probability measure π on S such that $|P^{(n)}(x, A) - \pi(A)| \leq (1 - \delta)^n$, $0 < \delta \leq 1$, where $P^{(n)}(x, A) = P[X_{n+1} \in A \mid X_1 = x]$. The distribution π is the unique stationary distribution of the Markov-sequence $\boldsymbol{X}$. Thus, if $X_1 \sim \pi$, the Markov-sequence $\boldsymbol{X}$ is a strictly stationary sequence. Further, such a Markov-sequence is a φ-mixing sequence (Lemma 2.1 below). This allows us to have uniform strong convergence of $\hat{f}(y|x)$ to $f(y|x)$, under some additional assumptions on $k(x, y)$ and $f(x, y)$ (Corollaries 2.1 and 2.2 below).

We may point out here that all our results can be generalized to an m-th order Markov-sequence also. This is particularly true, since Ruschendorf's (1977) results (which we have used in Lemma 2.2) hold for estimators of densities in higher dimensions, obtained from a φ-mixing process with $\varphi(n) = \rho^n$, $\rho < 1$.

The rest of the paper is organized as follows. In Section 2, we show that the Bootstrap estimator of the variance of mean-like statistics is a strongly consistent estimator. We also prove that, with probability one, Bootstrap distribution of such a statistic converges to an appropriate normal distribution. In Section 3, we show that Bootstrap works for estimators $T_n = T(F_n)$ (here, F_n denotes the empirical distribution function of the observations $X_1, X_2, \ldots, X_n$) which admit an expansion $T_n - \theta = (\Sigma h(X_i, F)/n) + R_n$, where F is the distribution function of X_1 and $R_n = o_p(n^{-1/2})$. Further, $n^{-1/2} \Sigma h(X_i, F)$ obeys a Central Limit Theorem. As in the case of independent observations, this class admits a large number of estimators, such as maximum likelihood, least squares, quantiles and so on. We show in this section that the Bootstrap distribution of $n^{1/2}(T(F_n^*) - T(\hat{F}_n))$ almost surely converges to an appropriate normal distribution (here, F_n^* denotes the empirical distribution of $X_1^*, X_2^*, \ldots, X_n^*$ (a Bootstrap realization), whereas $\hat{F}_n$ denotes the estimator of F, obtained from (1.3)). Also, a consistent estimator of σ^2, the variance of the asymptotic normal distribution, can be obtained based on the Bootstrap values. In Section 4, we give a few simulation results and offer some concluding remarks.

Throughout this paper, we follow the convention of denoting the Bootstrap samples by $X_1^*, X_2^*, \ldots$. The symbols P*, E*, Var* would denote conditional probability, expectation and variance given the sample. An "almost surely" statement refers to the probability measure P of $\boldsymbol{X}$.

2. Properties of kernel-based Bootstrap for mean-like statistics

We start with the φ-mixing property of a Markov-sequence satisfying A1.

LEMMA 2.1. *Suppose that* $\boldsymbol{X}$ *satisfies* A1. *Then, the sequence* $\boldsymbol{X}$ *is* φ*-mixing; i.e., for* $A \in \sigma\{X_1, \ldots, X_m\}$ *and* $B \in \sigma\{X_{n+m}, X_{n+m+1}, \ldots\}$, *we have*

$$\sup_{A,B} |P(A \cap B) - P(A)P(B)| \le \varphi(n)P(A),$$

where $\varphi(n) = \rho^n$ *for some* ρ, $0 \le \rho < 1$.

PROOF. From Iosifescu and Theodorescu ((1969), p. 1), it follows that the φ-mixing coefficient between two σ-fields $\mathscr{F}_1$ and $\mathscr{F}_2$ satisfies the property that $\varphi(\mathscr{F}_1, \mathscr{F}_2) \le 1 - \alpha(\mathscr{F}_1, \mathscr{F}_2)$ where $\alpha(\mathscr{F}_1, \mathscr{F}_2)$ is the independence coefficient between $\mathscr{F}_1, \mathscr{F}_2$ (defined in a manner similar to (1.4)). Let $\alpha(P^{(n)})$ be the independence coefficient between $\sigma\{X_1, \ldots, X_m\}$ and $\sigma\{X_{m+n}, \ldots\}$. From Proposition (1.2.4) of Iosifescu and Theodorescu (1969), we have $[1 - \alpha(P^n)] \le [1 - \alpha(P)]^n$.

Let $|\boldsymbol{u}|$ denote the Euclidean norm of a vector $\boldsymbol{u}$ of appropriate dimensions.

LEMMA 2.2. *Suppose that the following assumptions hold.*

1. *The kernel* $k(x, y)$ *satisfies the following conditions.*
   (a) $k(x, y) \to 0$ *as* $|(x, y)| \to \infty$, *uniformly in* (x, y).
   (b) $k(x, y)$ *is of bounded variation on* $S \times S$.
2. $f(x, y)$ *is a uniformly continuous and bounded function on* $S \times S$.
3. *The sequence* $\boldsymbol{X}$ *is a* φ*-mixing sequence with* $\varphi(n) = \rho^n$, $\rho < 1$.
4. $\sum_{n=1}^{\infty} [n^{1/2}h^2]^{2(k+1)} < \infty$, *for some* $k \ge 3$. *Then,*

$$\sup_{x,y} |\hat{f}(x, y) - f(x, y)| \to 0, \quad a.s.$$

PROOF. Proof follows by verifying the conditions of Ruschendorf ((1977), Theorem 1, Part B).

COROLLARY 2.1. *If, in addition to the assumptions of the above lemma, $f(x)$ is uniformly continuous and bounded on S, we have*

$$\sup_x |\hat{f}(x) - f(x)| \to 0, \qquad a.s.$$

COROLLARY 2.2. *Suppose further that $f(x) \geq \delta > 0$ for each $x \in S$. Then, we have*

$$\sup_{x,y} |\hat{f}(y|x) - f(y|x)| \to 0 \qquad a.s.$$

Remark 2.1. The condition that $f(x) \geq \delta > 0$ for each $x \in S$ may appear to be restrictive. However, this is not so; this can be seen as follows. For a given $\varepsilon > 0$, one can find a subset $S(\varepsilon)$ of the state-space S such that $f(x) \geq \delta(\varepsilon)$ (> 0) for $x \in S(\varepsilon)$ and $P[X_1 \in S(\varepsilon)] > 1 - \varepsilon$. (This is more easily seen when the state-space is the real line.) We may then consider the process whose state-space is given by $S(\varepsilon)$. For all practical purposes, this reduction of the state-space would not affect sampling properties of the Bootstrap estimators, provided moments of appropriate orders exist.

COROLLARY 2.3. *Suppose that the assumptions of Corollary* 2.2 *hold. Let*

$$\text{(2.1)} \qquad \begin{aligned} \hat{f}(x_1, x_2, \ldots, x_m) &= \hat{f}(x_1) \prod_{k=1}^{m} \hat{f}(x_k | x_{k-1}) ; \\ f(x_1, x_2, \ldots, x_m) &= f(x_1) \prod_{k=1}^{m} f(x_k | x_{k-1}) . \end{aligned}$$

Then, for each m, almost surely

$$\text{(2.2)} \qquad \sup_{x_1, x_2, \ldots, x_m} |\hat{f}(x_1, x_2, \ldots, x_m) - f(x_1, x_2, \ldots, x_m)| \to 0, \qquad \textit{all in } S .$$

A similar result holds for an m-dimensional distribution function and its estimate obtained from the density estimate. Corollary 2.3 explains why the Bootstrap can be anticipated to "work"; finite dimensional distributions of $\boldsymbol{X}$ and $\boldsymbol{X}^*$ eventually agree. Therefore, weak limits of statistics $t(\cdot)$ obtained from $(X_1, X_2, \ldots, X_n)$ and $(X_1^*, X_2^*, \ldots, X_n^*)$ should also match.

The following result makes this more precise in terms of the Mallows metric (see Bickel and Freedman (1981) for a detailed discussion of the Mallows metric). Let G and H be two m-dimensional distribution functions. Suppose further that the expectation of $|\boldsymbol{u}|^p$ is finite under both G and H. The Mallows metric $d_p(G, H)$ is defined by the infimum $\{E|\boldsymbol{U} - \boldsymbol{V}|^p\}^{1/p}$ where the infimum is taken over all $\boldsymbol{U} \sim G$ and $\boldsymbol{V} \sim H$.

THEOREM 2.1. *Suppose that the assumptions of Corollary* 2.3 *are met and that* $E(X_i^2) < \infty$. *Then, for each* m,

$$d_2(\hat{F}[m], F[m]) \to 0, \qquad a.s.$$

where $F[m]$ *and* $\hat{F}[m]$ *denote the distribution function of* $(X_1, X_2, \ldots, X_m)$ *and its estimator derived from* (2.1) *respectively.*

PROOF. In view of Corollary 2.3 and Scheffe's theorm (cf. Billingsley (1968)), it follows that $P[\hat{F}[m] \to F[m]$ weakly] equals unity. Further, since both $\boldsymbol{X}^*$ and $\boldsymbol{X}$ are strictly stationary, it suffices to prove that

$$\mathrm{E}^*[X_1^{*2}] \to E[X_1^2] \qquad \text{a.s.}$$

$\left(\text{so that } \mathrm{E}^*\left[\sum_1^m X_i^{*2}\right] \to E\left[\sum_1^m X_i^2\right] \text{ a.s.}\right)$. However, the last convergence follows by an elementary computation. We complete the proof by an appeal to Lemma 8.3 of Bickel and Freedman (1981).

The following result is a key result in our development of statistical properties of the Bootstrap based on density estimates.

LEMMA 2.3. *For almost all sample sequences* $\boldsymbol{x}$, *the conditionally Markov-sequence* $\boldsymbol{X}^*$ *satisfies Assumption* A1. *Consequently, the process* $\boldsymbol{X}^*$ *is conditionally* φ*-mixing with* $\varphi(n) = \rho^n$, *for almost all* $\boldsymbol{x}$*'s.*

PROOF. As usual, it suffices to restrict ourselves to the sets A (in Assumption A1) which are of the type $(-\infty, y]$. Conclusion of the lemma then follows by using the almost sure uniform convergence of $\hat{F}(y|x)$ to $F(y|x)$ (uniform in x and y) and the fact that

$$\begin{aligned}\sup_{x, x', y} &|\hat{F}(y|x) - \hat{F}(y|x')| \\ &\le 2 \sup_{x, y} |\hat{F}(y|x) - F(y|x)| + \sup_{x, x', y} |F(y|x) - F(y|x')| \,.\end{aligned}$$

The mixing property follows from Lemma 2.1.

Most of the statistics we use in statistical analysis of Markov models are smooth functions of means of functions of (X_i, X_{i+1}), $i = 1, 2, \ldots, n$. For example, the first-order serial correlation is a smooth function of $\bar{X} = \Sigma X_i/n$, $\Sigma X_i^2/n$ and $\Sigma X_i X_{i+1}/n$. We therefore prove our results for mean-like statistics only. This result can be easily extended to smooth functions of such statistics (see Bickel and Freedman (1981), Section 3 and Lemma 8.10). Further note that, if the process $\{X_1, X_2, \dots\}$ is φ-mixing, so is the bivariate process $\{(X_1, X_2), (X_2, X_3), \dots\}$. Since the φ-mixing property of $\boldsymbol{X}$ and $\boldsymbol{X}^*$ plays a more prominent role in the proofs below, our method of proof can be applied to any resampling plan under which one can establish the φ-mixing property of $\boldsymbol{X}^*$ whenever $\boldsymbol{X}$ holds such a property. It follows that if a parametric Markov model is assumed to be φ-mixing for all Θ in the parameter space, the parametric Bootstrap process $\boldsymbol{X}^*$ would be φ-mixing and results on a parametric Bootstrap would follow from our results. Methods of proofs of Theorems 2.2 and 2.3 are general enough to accomodate non-Markov models also.

For notational convenience, the proof below is given only for $\bar{X}$.

THEOREM 2.2. *Suppose that* Var $(X_1) < \infty$. *Let*

$$\sigma^2 = \mathrm{Var}\,(X_1) + 2 \sum_{k=1}^{\infty} \mathrm{Cov}\,[X_1, X_{k+1}], \qquad 0 < \sigma^2 < \infty\,. \tag{2.3}$$

Then,

$$\mathrm{Var}^*\,(\sqrt{n}\,\bar{X}_n^*) \to \sigma^2 \qquad a.s.$$

PROOF. For a φ-mixing process with $\varphi(m) = \rho^m$, $0 \le \rho < 1$, the series on the right-hand side of (2.3) is known to be absolutely convergent (Iosifescu and Theodorescu (1969), Proposition 1.1.20). Let k be an integer such that $2 \sum_{j=k}^{\infty} (\rho^j)^{1/2} < \varepsilon$. In view of Theorem 2.1, with probability one, we have

(i) $\mathrm{E}^*[X_1^{*2}] \to E[X_1^2]$,
(ii) $\mathrm{E}^*[X_1^*] \to E[X_1]$,
(iii) $\mathrm{E}^*[X_1^* X_{j+1}^*] \to E[X_1 X_{j+1}]$, $j = 1, 2, \dots, k-1$.

Further, in view of Lemma 2.3, we can choose an $n \ge \max\{k, n_1(\boldsymbol{x})\}$, so that, by Lemma 1 on page 170 of Billingsley (1968),

$$2 \left| \sum_{j=k}^{n-1} \mathrm{Cov}^*\,[X_1^*, X_{j+1}^*] \right| \le 2 \sum_{j=k}^{n-1} (\rho^j)^{1/2}\, \mathrm{Var}^*\,(X_1^*)$$
$$< \varepsilon\, \mathrm{Var}^*\,(X_1^*)\,.$$

Now, let $n_2(\boldsymbol{x})$ be chosen so that $n \ge n_2(\boldsymbol{x})$ implies that $|\mathrm{E}^*[X_1^*] - E[X_1]| < \varepsilon$ and $|\mathrm{Cov}^*\,[X_1^*, X_{j+1}^*] - \mathrm{Cov}\,[X_1, X_{j+1}]| < \varepsilon/[2(k+1)]$, for $j = 0, 1, 2, \dots, k-1$. Let an integer n be so chosen that $n \ge \max\{k, n_1(\boldsymbol{x}), n_2(\boldsymbol{x})\}$. Then,

$$\left| \mathrm{Var}^*(X_1^*) + 2\sum_{j=1}^{n-1} \mathrm{Cov}^*[X_1^*, X_{j+1}^*] - \mathrm{Var}(X_1) - 2\sum_{j=1}^{n-1} \mathrm{Cov}[X_1, X_{j+1}] \right|$$
$$\le |\mathrm{Var}^*(X_1^*) - \mathrm{Var}(X_1)|$$
$$+ 2\sum_{j=1}^{k-1} |\mathrm{Cov}^*[X_1^*, X_{j+1}^*] - \mathrm{Cov}[X_1, X_{j+1}]|$$
$$+ 2\sum_{j=k}^{n-1} \{|\mathrm{Cov}^*[X_1^*, X_{j+1}^*]| + |\mathrm{Cov}[X_1, X_{j+1}]|\}$$
$$\le \varepsilon + \varepsilon[\mathrm{Var}^*(X_1^*) + \mathrm{Var}(X_1)] .$$

This completes the proof (also see Remark 2.2 below).

COROLLARY 2.4. *Let $g(x,y)$ be a function on $S \times S$ such that $g(x,y)$ is $O(|(x,y)|^2)$ at (∞,∞). Let $S_n(g) = \{\Sigma g(X_i, X_{i+1})\}/n$. Then,*

$$\mathrm{Var}^* \frac{S_n^*(g)}{\sqrt{n}} \to \lim_{n\to\infty} \mathrm{Var} \frac{S_n(g)}{\sqrt{n}} \qquad a.s.$$

PROOF. Proof follows again from Lemma 8.3 of Bickel and Freedman (1981) and an argument similar to the proof of the above theorem.

Remark 2.2. The number σ^2 in (2.3) is not, in general, Var $(\sqrt{n}\,\bar{X}_n)$. The exact variance would be given by

$$\mathrm{Var}(X_1) + 2\sum_{j=1}^{n-1} \mathrm{Cov}(X_1, X_{j+1}) - \frac{2}{n}\sum_{j=1}^{n-1} j\,\mathrm{Cov}(X_1, X_{j+1}) .$$

Again, in view of the φ-mixing nature of X (with $\varphi(m) = \rho^m$), n times the third term above can be shown to be absolutely convergent. Our proof essentially ignores the third term, since, for the Bootstrapped process also, this term is eventually negligible.

To prove the central limit theorem for the Bootstrap distributions, we need explicit bounds for the error term in the normal approximation. Such bounds involve moments of sums of X_n's and are given in Tikhomirov (1980). We state his result in the form of a lemma.

LEMMA 2.4. (Tikhomirov (1980)) *Suppose that a stationary sequence X satisfies the strong mixing property viz.,*

$$\sup |P(A \cap B) - P(A)P(B)| \le \alpha(n)$$

where the supremum is taken over A in $\sigma\{X_1,\ldots, X_m\}$ and B in $\sigma\{X_{m+n},\ldots\}$.

Further suppose that

(1) $\alpha(n) \leq Ke^{-\beta n}$,

(2) $E|X_1|^3 < \infty$.

Let $E[X_1] = 0$ *and* $\sigma_n^2 = E\left(\sum_{i=1}^{n} X_i\right)^2$. *Let* $S_n = \sum_{i=1}^{n} X_i/\sigma_n$ *and* $\phi_n(t)$ *be the characteristic function of* S_n. *Let* $b = b(m)$ *denote the maximum of* $\left[E\left|\sum_{i=1}^{k} X_i\right|^3\right]^{1/3}$, *the maximum being taken over* $k = 1, 2, \ldots, m$. *Then, there exists a* T_0 *such that for* $|t| \leq T_0$,

$$\begin{aligned}
\phi_n'(t) = & -t\phi_n(t) + \theta_1 \frac{B_1 n[E|X_1|^3]^{1/3}}{\sigma_n}\left[\frac{|t|b}{\sigma_n}\right]^2 \phi_n(t) \\
& + \theta_2(t)B_2\left[\frac{|t|b}{\sigma_n}\right]^2 \sqrt{m}\,\frac{\sqrt{n}}{\sigma_n}[E|X_1|^3]^{1/3} \\
& + \frac{\theta_3(t)B_3|t|b[E|X_1|^3]^{1/3}m}{\sigma_n^2} \\
& + \frac{\theta_4(t)B_4[\alpha(m)]^{1/3}n[E|X_1|^3]^{1/3}}{\sigma_n} \\
& + \frac{\theta_5(t)n[E|X_1|^3]^{1/3}|t|b}{\sigma_n}\left[\frac{|t|b}{\sigma_n}\right]^k .
\end{aligned} \tag{2.4}$$

In (2.4), $\theta(t)$*'s refer to functions which satisfy* $|\theta(t)| \leq 1$ *and* B*'s refer to absolute constants.*

PROOF. Proof follows from Tikhomirov ((1980), Section 4 and Expression (4.6)). The constant T_0 is given by $\sigma_n/(32b(m))$.

THEOREM 2.3. *Let* $E|X_1|^3 < \infty$, $\sigma_n^{*2} = V^*[\Sigma X_i^*]$. *Then, we have,*

$$\sup_z \left| \mathrm{P}^*\left\{\frac{\Sigma[X_i^* - \mathrm{E}^*(X_i^*)]}{\sigma_n^*} \leq z\right\} - \Phi(z)\right| \leq A^*(n)n^{-1/2}(\log n) \quad a.s.,$$

where $\Phi(z)$ *is the distribution function of a standard normal variate at* z *and* $A^*(n)$ *is a sequence of random variables such that* $A^*(n) \to A$ *almost surely* $0 < A < \infty$.

PROOF. For brevity, we only sketch the proof. We first notice that, almost surely, by Lemma 2.3, the conditional process $\boldsymbol{X}^*$ is φ-mixing and therefore strong-mixing also, since $\varphi(n)P(A) \leq \varphi(n)$. Further, since $\mathrm{E}^*|X_1^*|^3$

can be shown to converge to $E|X_1|^3$, there is a set of probability unity where the moment condition of Lemma 2.4 is satisfied. The following argument is applicable to the $\boldsymbol{x}$'s which belong to the intersection of these two almost-certain events.

We now apply Lemma 2.4 to the process X^*. Let $T_0^* = \sigma_n^* 32 b^*(m)$. Now, as in Tikhomirov (1980), choose m such that $[\alpha(m)]^{1/3} \leq Cn^{-2}$, $(|t| b^*(m)/\sigma^*(n))^{k-2} \leq c/n$ for a suitable k and $m = O(n^{-(2\beta+1)})$. Such a choice is possible in view of the exponential decay of $\alpha(m)$ and the fact that $\sigma^*(n) = O(\sqrt{n})$ and $b^*(m) = O(m)$. At this stage, we make a routine appeal to Lemma 2 on page 512 of Feller (1966). This establishes the bound claimed in the theorem. To see that $A^*(n) \to A$, we need to note that $A^*(n)$ is completely specified by $b^*(m)$ and $\sigma^*(n)$ and as before, $b^*(m)/m$ and $\sigma^{*2}(n)/n$ converge to the corresponding population quantities.

Remark 2.3. It is possible to prove that, under Assumption A1 and the assumptions of Lemma 2.2 and Corollary 2.2, the Bootstrap distribution of smooth functions of sample moments offers an approximation to the sampling distribution of such statistics which is superior to the usual normal approximation. This approach assumes Cramér's condition ((1.1) of Gotze and Hipp (1983)) together with the existence of higher moments resulting in Edgeworth expansions for smooth functions of sample moments. This has been discussed by Bose (1986) for auto-regressive and moving average processes. Actually, under Assumption A1 (see (1.13) of Gotze and Hipp (1983)) and Cramér's condition, smooth functions of sample moments obtained from a Markov-sequence admit Edgeworth expansions of appropriate degree. We also refer to Babu and Singh (1984) in this connection. For the sake of brevity, we have not adopted this approach here.

3. Statistics which are of the von Mises differentiable type

Throughout this section, we assume that the conditions of Section 2 continue to hold; more precisely, Assumption A1 and the assumptions for the density estimation in Lemma 2.2 and Corollary 2.2 hold.

Let F_n be the empirical distribution function of $\{X_1, X_2, \ldots, X_n\}$. We first prove a lemma which gives sufficient conditions for a statistic $T(F_n)$, obtained from a sequence of dependent observations, to have an expansion which is of the von Mises differentiable type. Let $h(x, F)$ denote $d_1 T(F; \delta_x - F)$, the Gâteaux differential of T at F in the direction of $\delta_x - F$, where δ_x is the distribution function of a random variable degenerate at x (cf. Serfling (1980), Chapter 6).

LEMMA 3.1. *Suppose that the process* $\boldsymbol{X}$ *satisfies Assumption* A1. *Suppose further that a statistic is Gâteaux differentiable so that*

$$(3.1) \qquad T(F_n) - T(F) = \frac{1}{n} \sum_{i=1}^{n} h(X_i, F) + R_n .$$

Assume that $0 < \operatorname{Var}(h(X_1, F)) < \infty$. *Then,* $R_n = o_p(n^{-1/2})$. *Consequently,* $n^{1/2}(T(F_n) - T(F))$ *has the same limiting distribution as that of* $n^{-1/2} \cdot \sum_{i=1}^{n} h(X_i, F)$, *which is given by* $N(0, \sigma^2)$, *where*

$$(3.2) \qquad \sigma^2 = \operatorname{Var}(h(X_1, F)) + 2 \sum_{j=2}^{\infty} \operatorname{Cov}(h(X_1, F), h(X_j, F)) .$$

PROOF. As noted in Lemma 2.1, the process $\boldsymbol{X}$ is φ-mixing with $\varphi_n = \rho^n$. Now, by Theorem 4 of Withers (1975), it follows that the empirical distribution function process

$$(3.3) \qquad n^{-1/2} \sum_{i=1}^{n} \{I[X_i \le t] - F(t)\} \qquad (t \in S)$$

converges weakly to a Gaussian process. Therefore, with $\|g(t)\| = \sup |g(t)|$, we have $n^{1/2}\|F_n - F\| = O_p(1)$. By Lemma B of Serfling ((1980), p. 218), it follows that $n^{1/2} R_n = o_p(1)$.

Now, in view of the φ-mixing nature of the process and the assumption that $\operatorname{Var}(h(X_1, F)) < \infty$, it follows that the series on the right-hand side of (3.3) is absolutely convergent. The Central Limit Theorem for $n^{-1/2} \sum h(X_i, F)$ follows from Theorem 1.1.2.3 of Iosifescu and Theodorescu (1969). Combining this with the fact that $n^{1/2} R_n = o_p(1)$, the proof is complete.

Let $\hat{F}[1] = \hat{F}_n$ denote the estimator of F, the distribution function of X_1, obtained from (1.3). The following theorem shows that the limiting distribution of the Bootstrapped statistic $n^{1/2}(T(F_n^*) - T(\hat{F}_n))$ essentially agrees with that of the $n^{1/2}(T(F_n) - T(F))$.

THEOREM 3.1. *Suppose that with probability one, for any* m (≥ 1),

$$(3.4) \qquad E^*[h(X_1^*, \hat{F}_n) h(X_m^*, \hat{F}_n)] \to E[h(X_1, F) h(X_m, F)] .$$

Then, for almost all sample sequences $\boldsymbol{x}$, *the conditional distribution of* $n^{1/2}(T(F_n^*) - T(\hat{F}_n))$ *converges weakly to* $N(0, \sigma^2)$.

PROOF. It is convenient to imagine that a very large segment of the Bootstrap process has been observed: let N denote the Bootstrap sample size and n denote the original sample size. As noted in Lemma 2.3, there exists a sufficiently large n such that the strictly stationary processes $\boldsymbol{X}^*$

satisfies Assumption A1, conditionally on $(X_1, X_2, \ldots, X_n)$. Further, there again exists a sufficiently large n such that $E^* h^2(X_i^*, \hat{F}_n) < \infty$, in view of the assumption (3.4). Hence, by Lemma 3.1, for such a fixed n, it follows that the result (3.3) holds (conditionally on $\boldsymbol{x}$) for the empirical distribution function process $N^{1/2}(F_N^*(t) - \hat{F}_n(t))$, $t \in S$. Consequently, again by Lemma 3.1, as N tends to ∞, $N^{1/2}(T(F_N^*) - T(\hat{F}_n))/\sigma_n^*$ converges weakly to $N(0, 1)$ conditionally on $\boldsymbol{x}$ (for almost all $\boldsymbol{x}$'s), where

$$\sigma_n^{*2} = \mathrm{Var}^* (h(X_1^*, \hat{F}_n)) + 2 \sum_{j=2}^{\infty} \mathrm{Cov}^* (h(X_1^*, \hat{F}_n), h(X_j^*, \hat{F}_n)) .$$

By making arguments similar to those in Theorem 2.2 and using condition (3.4), it follows that $\sigma_n^{*2} \to \sigma^2$ with probability one, where σ^2 is as defined by (3.2). The proof is complete.

Remark 3.1. Condition (3.4) is an analogue of corresponding conditions of Bickel and Freedman (1981) for von Mises differentiable functions in the case of independent observations. It is possible to replace the condition (3.4) by conditions which do not involve the Bootstrap process. We prefer to verify (3.4) in a given situation.

Example 3.1. (The sample quantiles) From Serfling ((1980), p. 236), it follows that $h(X_1, F) = \{p - I[X_1 \le \theta_p]\}/f(\theta_p)$, where θ_p is the p-th quantile of the distribution function F. It is easily seen $E^* h(X_1^*, T(\hat{F}_n))^2 = \hat{F}_n(\hat{\theta}_p)\{1 - \hat{F}_n(\hat{\theta}_p)\}/[\hat{f}(\hat{\theta}_p)]^2$, which converges to $F(\theta_p)\{1 - F(\theta_p)\}/[f(\theta_p)]^2$ $(= p(1-p)/[f(\theta_p)]^2)$, almost surely. It can be similarly verified that $E^*[h(X_1^*, \hat{F}_n) h(X_j^*, \hat{F}_n)] = P^*[X_1^* \le \hat{\theta}_p, X_j^* \le \hat{\theta}_p]/[\hat{f}(\hat{\theta}_p)]^2$, so that (3.4) follows from Corollary 2.2 and the strong consistency of $\hat{\theta}_p$, the p-th percentile of $\hat{F}_n$.

Example 3.2. (Trimmed mean) Condition (3.4) can be again easily verified for the trimmed mean by noting the expression for $h(x, F)$ for the trimmed mean in Serfling ((1980), p. 237).

Remark 3.2. We can extend the discussion in Example 3.1 to the estimator $\hat{\rho} = \text{median}\,\{X_2/X_1, X_3/X_2, \ldots, X_n/X_{n-1}\}$, which can be viewed as the analogue of the MPS (Median of the Pairwise Slopes) in the nonparametric regression analysis. Notice that the Bootstrap procedure would give correct answers irrespective of whether the linearity assumption $E(X_2 | X_1) = \rho X_1$ holds or not.

The above theorem cannot be used to conclude that the Bootstrap estimator of variance is consistent. This would require the additional assumption that $E(R_n^2) = o(n^{-1})$, verification of which could be quite difficult if the observations are dependent. We may refer to Shao and Wu ((1989), Sections 2 and 5) for relevant discussion in the context of independent observations. Extension of such results to the case of dependent observations is under progress and will be reported subsequently. At this stage, in the absence of verification of the condition $E(R_n^2) = o(n^{-1})$, we may recommend the use of the interquartile range of the Bootstrap values $n^{1/2}(T(F_n^*) - T(\hat{F}_n))$ for estimation of σ. This suggestion is due to Parr (1985) whose arguments are based on the fact that, unlike the variance, the interquartile range is a continuous functional on the space of distribution functions (cf. Huber (1981)). We omit the proof of the result that under the condition that $E(R_n^2) = o(n^{-1})$, the usual Bootstrap estimator of the variance of σ^2 is consistent under the set-up of the above theorem.

4. Simulation results for two Markov models

In this study, we simulated 400 samples each of size 100 from each of the following Markov models.

(1) A Bilinear model: $X_n = .75 X_{n-1} + .15 X_{n-1}\varepsilon_n + \varepsilon_n$ where ε_n is a sequence of independent $N(0,1)$ variables. We refer to Tong (1981) for a discussion of bilinear models.

(2) A Non-linear model: $X_n = .75 X_{n-1}^{(.90)} + \varepsilon_n$ where ε_n is as defined in (1) above.

To obtain data from a stationary sequence, a sample of size 300 was drawn in each case and the last 100 observations were retained.

For both the simulation studies, the same density estimate was employed to generate the Bootstrap samples. The kernel $k(x, y)$ was chosen to be $k_1(x)k_1(y)$ where $k_1(x)$ is the uniform density over $(-0.5, 0.5)$. Following Silverman (1986), the sequence $h(n)$ was taken to be $(.90)n^{-1/6}\hat{s}/1.34$, where $\hat{s}$ is the interquartile range of the sample. Each Bootstrap observation was rescaled so that the variance of the Bootstrap sample matched that of the original sample (see Silverman (1986), p. 143).

Three statistics were included in the study: the serial correlation coefficient (R), the sample mean ($\bar{x}$) and the sample median ($\tilde{x}$). Little can be derived regarding the sampling distributions of these three statistics. Therefore, the standard errors of these statistics were obtained by the simulated values of these three statistics. The table below summarizes the results of this simulation study.

Table 1 shows that Bootstrap procedure offers reasonable estimates of variances and standard errors of the three statistics chosen. The performance of Bootstrap estimators of standard errors is particularly satisfactory in terms of bias and variance. We may point out that the theoretical computations of variances of these statistics are almost impossible and may not be of any practical use. Also, the performance of Bootstrap estimators may possibly improve if better estimators of the density (such as two-stage

Table 1. Simulation study* of Bootstrap estimators of variances and standard errors of three statistics: serial correlation, mean and median.

Statistic	Variance	Standard error (s.e.)	Mean of Bootstrap estimates of variance	Mean of Bootstrap of s.e.
Bilinear model				
$\sqrt{n}\,R$	.5676	.7534	.5532 $(8.60 \times 10^{-2})^{a}$	.7212 $(3.30 \times 10^{-2})^{a}$
$\sqrt{n}\,\bar{x}$	15.3366	3.9162	15.1719 (191.772)	3.6314 (1.9846)
$\sqrt{n}\,\tilde{x}$	14.2937	3.7807	17.4508 (237.886)	3.887 (2.342)
Non-linear model				
$\sqrt{n}\,R$	1.069	1.0339	1.0294 (.2455)	.989 (5.11×10^{-2})
$\sqrt{n}\,\bar{x}$	4.048	2.012	3.888 (3.20)	1.926 (.1778)
$\sqrt{n}\,\tilde{x}$	4.955	2.226	5.22 (7.983)	2.2064 (.3525)

*: n (sample size) = 100; $B = 250$. Results are based on 400 simulations.

[a]: These figures are variances of the estimators of the variances or the standard deviations.

adaptive estimators etc.) are employed.

Acknowledgements

Thanks are due to Dr. William Harkness and the Department of Statistics, Pennsylvania State University for providing all the facilities for carrying out this work. Computer programs were written by Mr. Garth Longdon. Thanks are also due to both the referees for their helpful comments, which led to a considerable improvement over the earlier draft.

References

Babu, G. J. and Singh, K. (1983). Nonparametric inference on means using bootstrap, *Ann. Statist.*, **11**, 999–1003.

Babu, G. J. and Singh, K. (1984). On one term Edgeworth correction by Efron's bootstrap, *Sankhyā Ser. A*, **46**, 219–232.

Beran, R. (1982). Estimating sampling distribution: The bootstrap and its competitors, *Ann. Statist.*, **10**, 212–225.

Bickel, P. J. and Freedman, D. (1981). Some asymptotic theory for the bootstrap, *Ann. Statist.*, **9**, 1196–1217.

Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

Bose, A. (1986). Asymptotic study of estimators in some discrete and continuous time models, Ph.D. Thesis, Indian Statistical Institute, Calcutta (unpublished).

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence, *Ann. Statist.*, **14**, 1171–1179.

Efron, B. (1979). Bootstrap methods: Another look at Jackknife, *Ann. Statist.*, **7**, 1–26.

Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy, *Statist. Sci.*, **1**, 54–77.

Feller, W. (1966). *An Introduction to Probability Theory and Its Applications*, Vol. 2, Wiley, New York.

Freedman, D. (1981). On bootstrapping two-stage least-square estimates in stationary linear models, *Ann. Statist.*, **12**, 827–842.

Gotze, F. and Hipp, C. (1983). Asymptotic expansions for sums of weakly dependent random vectors, *Z. Wahrsch. Verw. Gebiete*, **64**, 211–239.

Huber, P. J. (1981). *Robust Statistics*, Wiley, New York.

Iosifescu, M. and Theodorescu, R. (1969). *Random Processes and Learning*, Springer, New York.

Parr, W. C. (1985). The bootstrap: Some large sample theory and connections with robustness, *Statist. Probab. Lett.*, **3**, 97–100.

Ruschendorf, L. (1977). Consistency of estimators for multivariate density functions and for the mode, *Sankhyā Ser. A*, **39**, 243–250.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Shao, J. and Wu, C. F. J. (1986). Some general theory for the jackknife, *Ann. Statist.*, **17**, 1176–1197.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London.

Singh, K. (1981). On asymptotic accuracy of Efron's bootstrap, *Ann. Statist.*, **9**, 1187–1195.

Tikhomirov, A. N. (1980). On the convergence rate in the central limit theorem for weakly dependent random variables, *Theory Probab. Appl.*, **25**, 790–809.

Tong, H. (1981). A note on a Markov bilinear stochastic process in discrete time, *J. Time Ser. Anal.*, **2**, 279–284.

Withers, C. S. (1975). Convergence of empirical processes of mixing rv's on [0, 1], *Ann. Statist.*, **3**, 1101–1108.