The Annals of Statistics
1994, Vol. 22, No. 2, 980–994

VALIDITY OF BLOCKWISE BOOTSTRAP FOR EMPIRICAL PROCESSES WITH STATIONARY OBSERVATIONS

BY U. V. NAIK-NIMBALKAR[1] AND M. B. RAJARSHI

University of Poona

We show that the empirical process of the block-based bootstrap observations from a stationary sequence converges weakly to an appropriate Gaussian process, conditionally in probability and almost surely, depending upon the block length. This bootstrap was introduced by Künsch and later by Liu and Singh. Applications in estimation of the sampling distribution of a compactly differentiable functional are indicated.

1. Introduction. It is now well-established that, in exchangeable situations, Efron's bootstrap offers a very important nonparametric technique to study the sampling distributions of complicated statistics and pivotals. Extensions of Efron's technique to the dependent random variables have been done by Freedman (1984), Bose (1988) and Rajarshi (1990). However, these authors assume either a semiparametric model such as an autoregressive model [Freedman (1984) and Bose (1988)] or impose a structure such as a Markovian character [Rajarshi (1990)]. Carlstein (1986) offers a truly nonparametric method, based on a subseries technique for estimation of the variance of a statistic.

Recently, Künsch (1989) and Liu and Singh (1992) have developed an ingenious extension of Efron's bootstrap to general stationary sequences of observations. This technique involves selecting k blocks by a simple random sample with replacement, from $n - \ell + 1$ blocks of observations $(X_{j+1}, X_{j+2}, \ldots, X_{j+\ell})$, $j = 0, 1, \ldots, n - \ell$, where n denotes the sample size and $n = \ell k$. Under some assumptions on the rate of decay of the mixing coefficients of the stochastic sequence $\{X_n, n \geq 1\}$, and by a proper choice of $\ell = \ell(n)$, Künsch (1989) shows that the block-based bootstrap correctly estimates the sampling distribution as well as the asymptotic variance of the sample mean, when n is sufficiently large. Künsch (1989) also studies various jackknife estimators and establishes their consistency for various classes of estimators. He further demonstrates that the jackknife and bootstrap procedures work better than Carlstein's subseries technique.

Under certain conditions [see (2.2)], the empirical process of $\{X_1, X_2, \ldots, X_n\}$ converges weakly to a Gaussian process Z, sample paths of which are continuous. The covariance kernel of this Gaussian process Z would be typically un-

Received December 1990; revised January 1993.

[1]Research partially supported by the University Grants Commission (India) No. 26-1(876)/89 (SR-IV). Part of the work was done while visiting the University of Waterloo, Canada.

AMS 1991 *subject classifications.* Primary 62G08; secondary 62M99.

Key words and phrases. Bootstrap, empirical processes, weak convergence, stationary and mixing sequences, compactly differentiable functionals.

known to the statistician. Even if it is known or estimated, applying it to derive a formula for the variance of a complicated functional (or, more ambitiously, the sampling distribution of a functional) would ask for a detailed theoretical investigation compounded by computational complexities. Consider, for example, the form of the variance of the trimmed mean of a Gaussian process, as derived by Gastwirth and Rubin (1975). Thus, resampling procedures such as jackknife and bootstrap have to play a more prominent role in a statistical analysis of dependent observations than that of independently and identically distributed (i.i.d.) observations. Further, Künsch [(1989), page 1222] rightly remarks that "with dependence, parametric methods are even more dangerous than in i.i.d. situations." It is indeed a very redeeming feature of a bootstrap procedure that it allows a data analyst a purely computational device to enable her/him to answer most of the questions of interest.

In this paper, we show that, conditionally on the observations, the bootstrapped empirical process converges weakly to the process Z in probability or almost surely depending upon the order of the block size ℓ. Consequently, the bootstrap estimator of the sampling distribution of a compactly (or Hadamard) differentiable statistical functional is weakly or strongly consistent. This result further can be used to estimate the variance (when it exists) of the asymptotic distribution of a compactly differentiable statistic. Related work for i.i.d. observations can be found in Bickel and Freedman (1981), Singh (1981), Parr (1985), Shorack (1982), Swanepoel (1986), Liu, Singh and Lo (1989) and Gill (1989). The last two papers explicitly deal with compactly differentiable functions. Once the weak convergence of the bootstrap empirical process is established, our approach is similar to that of Gill (1989).

In Section 2, we give the notation and assumptions and state our main results. Section 3 deals with the preliminaries, the main result there being that the bootstrap estimator of the covariance kernel of Z is strongly consistent, *uniformly in* (s,t). In Section 4, we prove the main results.

2. Assumptions and main results. Throughout we assume that $\{X_n, n \geq 1\}$ is a stationary, strong or α-mixing process with mixing coefficients $\alpha(n)$. Let

$$\begin{aligned} U_i(s,t) &= I[s < X_i \leq t], \\ U_i(t) &= I[X_i \leq t], \\ F_n(t) &= n^{-1} \sum_{i=1}^{n} U_i(t), \qquad (2.1) \\ F(t) &= P[X_1 \leq t], \\ Z_n(t) &= \sqrt{n}\big[F_n(t) - F(t)\big]. \end{aligned}$$

Under the assumptions that X_1 has a continuous distribution and that $\Sigma(i+1)^2\alpha(i)^{1/2-\tau} < \infty$ for $\tau \in (0, \frac{1}{2})$, it is known [Deo (1973)] that

$$Z_n \Rightarrow Z, \qquad (2.2)$$

where Z is a Gaussian process with almost all continuous sample paths; the symbol $\Rightarrow$ is being used, as usual, to denote the weak convergence in the space of cadlag functions [cf. Billingsley (1968), Chapters 3 and 4, and Pollard (1984)]. Our assumptions on $\alpha(n)$'s would imply these assumptions and are described in Theorem 2.1. In particular, they also imply that the covariance kernel $\mathcal{K}(s,t)$ of the process Z is continuous in (s,t).

The block-based bootstrap formally operates as follows. Let $S_1, S_2, \ldots, S_k$ be i.i.d. random variables each having a uniform distribution on $\{0, 1, \ldots, n-\ell\}$. Let

$$H_j(t) = \ell^{-1} \sum_{r=1}^{\ell} U_{r+j}(t), \qquad j = 0, 1, \ldots, n-\ell. \tag{2.3}$$

Then, a bootstrap distribution function F_n^* is defined by

$$F_n^* = k^{-1} \sum_{j=1}^{k} H_{S_j}. \tag{2.4}$$

One then computes T, a functional on the space of distribution functions, at F_n^*, by realizing $S_1, S_2, \ldots, S_k$. The rest of the bootstrap methodology is the same as that of i.i.d. observations.

As is customary, we let P^* and E^* denote the conditional probability and the conditional expectation, respectively, given the data $(x_1, x_2, \ldots, x_n)$. We further define $\widetilde{F}_n$ and Z_n^* by

$$\widetilde{F}_n = E^*\left[F_n^*\right] = (n-\ell+1)^{-1} \sum_{j=0}^{n-\ell} H_j \tag{2.5}$$

and

$$Z_n^* = \sqrt{n}\left[F_n^* - \widetilde{F}_n\right]. \tag{2.6}$$

The first main result of the paper is as follows.

THEOREM 2.1. *Let the stationary and α-mixing stochastic process $\{X_n, n \geq 1\}$ satisfy the following*:

(a) $\sum (i+1)^7 \alpha(i)^{1/2-\tau} < \infty$, *for* $\tau \in (0, \frac{1}{2})$.

(b) *The random variable X_1 has a continuous distribution on $\mathbb{R}$, the real line.*

Let $\ell = \ell(n) = \mathbb{O}(n^{1/2-\varepsilon})$, with $0 < \varepsilon < \frac{1}{2}$. Then, for every subsequence $\{Z_{n_j}^*\}$, there exists a further subsequence $\{Z_{n_{j'}}^*\}$ such that, for almost all $\{x_1, x_2, \ldots\}$,

$$\{Z_{n_{j'}}^*\} \Rightarrow Z, \tag{2.7}$$

conditionally on $(x_1, x_2, \ldots, x_n)$, where Z is as defined by (2.2).

The above mode of weak convergence of Z_n^* may be described equivalently as *weak convergence in probability, given the sample*.

For almost-sure weak convergence, however, a smaller block size is required.

THEOREM 2.2. *Under assumptions* (a) *and* (b) *of Theorem* 2.1, *suppose further that* $\ell = \mathbb{O}(n^{1/2-\varepsilon})$, $\frac{1}{4} < \varepsilon < \frac{1}{2}$. *Then, for almost all* $\{x_1, x_2, \ldots\}$,

$$Z_n^* \Rightarrow Z, \tag{2.8}$$

conditionally on $(x_1, x_2, \ldots, x_n)$.

As in Gill (1989), Theorems 2.1 and 2.2 immediately yield the weak or strong consistency of the block-based bootstrap procedure for compactly differentiable statistical functionals.

3. Preliminaries. We start with moment inequalities for sums of strongly mixing random variables. Let

$$\begin{aligned} \|Y\|_p &= \left\{E\left(|Y|^p\right)\right\}^{1/p}, \\ A_{r,\nu}(\alpha) &= \sum_{i=0}^{\infty}(i+1)^{r/2-1}\left[\alpha(i)\right]^{\nu/(r+\nu)}. \end{aligned} \tag{3.1}$$

From now on, the C's will denote constants whose exact values are unimportant for the discussion and may change from line to line.

LEMMA 3.1.

(a) *Let* $\{Y_n,\ n \geq 1\}$ *be a strictly stationary* α*-mixing sequence such that* $E(Y_1) = 0$, $\|Y_1\|_{m+\nu} < \infty$ *and* $A_{m,\nu}(\alpha) < \infty$ *for some* $\nu > 0$. *Then,*

$$E\left|\sum_{i=1}^{n} Y_i\right|^m \leq K(\alpha)\left[\|Y_1\|_{m+\nu}\right]^m n^{m/2}, \tag{3.2}$$

where $K(\alpha)$ *is a constant which depends only upon the order of moments assumed and the* α*-mixing coefficients.*

(b) *If* $E|Y_1|^{4+\nu} < \infty$ *for some* $\nu > 0$, *then*

$$\begin{aligned} E\left[\sum_{i=1}^{n} Y_i\right]^4 \leq Cn\Big\{&A_{2,\nu}(\alpha)\|Y_1\|_{2+\nu}^2 + A_{4,\nu}(\alpha)\|Y_1\|_{3+\nu}^3 \\ &+ nA_{4,\nu}(\alpha)\|Y_1\|_{4+\nu}^4 + n\left[A_{2,\nu}(\alpha)\right]^2\|Y_1\|_{2+\nu}^4\Big\}. \end{aligned}$$

PROOF. Part (a) is Theorem 1 [expression (4.1)] of Yokoyama (1980). Part (b) follows from (4.2), (4.3) and (4.4) of Yokoyama's paper. □

We note that if $|Y_1| < c$, for some real c, then one may effectively take $\nu = \infty$ in (3.1), that is, the sufficient condition for (3.2) to hold is $\sum(i+1)^{(m/2)-1}\alpha(i) < \infty$, [cf. Theorem 3.2 of Yokoyama (1980)]. Further, in part (b), one may choose different ν's in the terms occurring in the upper bound.

The following lemma establishes properties of $\widetilde{F}_n$ as an estimator of F, which will be useful in the sequel.

LEMMA 3.2.

(a) $\|\widetilde{F}_n - F_n\| < C(\ell - 1)/(n - \ell + 1)$. *Consequently, if* $\ell/\sqrt{n} \to 0$, $\sqrt{n}\,\|\widetilde{F}_n - F_n\| \to 0$.

(b) *If* $\sum(i+1)^{m-1}\alpha(i) < \infty$ *and if* $\ell = \mathbb{O}(n^{1/2-\varepsilon})$, $0 < \varepsilon < \frac{1}{2}$, *then* $E[\widetilde{F}_n(t) - F(t)]^{2m} < C\{(\ell-1)/(n-\ell+1) + 1/\sqrt{n}\}^{2m}$, *which is* $\mathbb{O}(n^{-m})$.

(c) *Suppose that* $\sum(i+1)^3\alpha(i)^{1/2-\tau} < \infty$, $0 < \tau < \frac{1}{2}$, $\ell = \mathbb{O}(n^{1/2-\varepsilon})$, $0 < \varepsilon < \frac{1}{2}$, *and that* X_1 *has a uniform distribution over* $[0,1]$. *Then* $\ell\|F_n - F\| \to 0$ *in probability. Further, if* $\frac{1}{6} < \varepsilon < \frac{1}{2}$, $\ell\|F_n - F\| \to 0$ *a.s. Consequently, similar results hold for* $\widetilde{F}_n$.

PROOF.

(a) From Künsch [(1989), (3.3)], it follows that

$$\widetilde{F}_n(t) - F_n(t) = \sum_{i=1}^{n} \left(a_n(i) - n^{-1}\right) U_i(t),$$

where $a_n(i)$ equals $i/[\ell(n-\ell+1)]$ for $i = 1, 2, \ldots, \ell - 1$, equals $1/(n-\ell+1)$ for $i = \ell, \ell+1, \ldots, (n-\ell+1)$ and equals $(n-i+1)/[\ell(n-\ell+1)]$ for $i = (n-\ell+2), \ldots, n$. Thus, the result follows immediately.

(b) We note that

$$\begin{aligned} &E\left[\widetilde{F}_n(t) - F(t)\right]^{2m} \\ &\quad = E\left[|\widetilde{F}_n(t) - F_n(t)| + |F_n(t) - F(t)|\right]^{2m} \\ &\quad \le C\left[(\ell-1)/(n-\ell+1) + 1/\sqrt{n}\right]^{2m} \end{aligned} \tag{3.3}$$

by the Minkowski inequality, part (a) and Lemma 3.1 (a).

(c) Convergence in probability follows easily from the fact that, under the stated conditions, $\sqrt{n}\|F_n - F\| = O_p(1)$, [cf. (2.2)]. For almost-sure convergence, arguing as in Shorack and Wellner [(1986), page 96], we see that $\ell\|F_n - F\| \le \max_{1\le i \le M(n)} \ell|F_n(i/M(n)) - i/M(n)| + \ell/M(n)$. Now we choose $M(n) = \mathbb{O}(n^{1/2-\varepsilon+\varepsilon'})$, $\varepsilon' > 0$, so that $\ell/M(n) \to 0$. Then for any $\delta > 0$,

$$\begin{aligned} &\sum P\left[\max_{1\le i\le M(n)} \ell\left|F_n\left(\frac{i}{M(n)}\right) - \frac{i}{M(n)}\right| > \delta\right] \\ &\quad \le C\sum M(n)\ell^8 E\frac{|F_n(t) - t|^8}{\delta^8} \\ &\quad \le C\sum M(n)\frac{\ell^8}{\delta^8 n^4}, \end{aligned}$$

by Lemma 3.1(a). The series converges if one can choose an $\varepsilon' > 0$ such that $M(n)\ell^8/n^4$ is $\mathbb{O}(n^{-1-\delta'})$, $\delta' > 0$. This is possible if $\varepsilon > \frac{1}{6}$. □

Our development toward the proof of Theorems 2.1 and 2.2 requires a study of the bootstrap estimator of the parameter

$$\begin{aligned}\sigma^2(s,t) &= \lim_{n\to\infty} \operatorname{Var}\left\{ \sum_{i=1}^{n} \left[U_i(s,t) - \big(F(t) - F(s)\big)\right] \Big/ \sqrt{n} \right\} \\ &= \operatorname{Var}\left[U_1(s,t)\right] + 2\sum_{i=2}^{\infty} \operatorname{Cov}\left[U_1(s,t), U_i(s,t)\right].\end{aligned}$$

Under the assumptions on the mixing coefficients as stated in Theorem 2.1, it follows that the above series converges absolutely [cf. Deo (1973), Lemma 2]. Also, if $F(t) = t$, $0 \le t \le 1$,

$$\begin{aligned}\sigma^2(s,t) &= \sum_{i=-\infty}^{\infty} E\Big\{\left[U_1(s,t) - (t-s)\right]\left[U_{1+|i|}(s,t) - (t-s)\right]\Big\} \\ &\le C \sum_{i=-\infty}^{\infty} \alpha\big(|i|\big)^{1/2-\tau} \Big\{E\left[U_1(s,t) - (t-s)\right]^p\Big\}^{2/p}\end{aligned}$$

with $p = 4/(1+2\tau)$, $2 < p < 4$ [cf. Withers (1975), Lemma 1 (f)]. Therefore,

$$\sigma^2(s,t) \le C|t-s|^b, \qquad b = 2/p,\ \tfrac{1}{2} < b < 1. \tag{3.4}$$

Further, it is clear that $\sigma^2(s,t)$ is the variance of the asymptotic normal distribution of the mean-like statistic, namely, $\sum U_i(s,t)/n$, and that $\mathcal{K}(s,t) = \operatorname{Cov}(Z(s), Z(t))$ is given by the equation

$$\sigma^2(s,t) = \sigma^2(0,t) + \sigma^2(0,s) - 2\mathcal{K}(s,t). \tag{3.5}$$

We note that $\sigma(s,t)$ defines a semimetric on $[0,1]$.

Let $\widehat{\sigma}_n^2(s,t)$ denote the bootstrap estimator of $\sigma^2(s,t)$, that is,

$$\widehat{\sigma}_n^2(s,t) = \operatorname{Var}^*\left[Z_n^*(t) - Z_n^*(s)\right].$$

It follows from (3.15) of Künsch (1989) that

$$\widehat{\sigma}_n^2(s,t) = \frac{\ell}{n-\ell+1} \sum_{j=0}^{n-\ell} \left\{ \sum_{r=1}^{\ell} \frac{U_{j+r}(s,t) - \big(\widetilde{F}_n(t) - \widetilde{F}_n(s)\big)}{\ell} \right\}^2. \tag{3.6}$$

It is then easily seen that

$$\widehat{\sigma}_n^2(s,t) - E\left[\widehat{\sigma}_n^2(s,t)\right] = \frac{\ell}{n-\ell+1} \left[\sum_{j=0}^{n-\ell} V_j(s,t) + V(s,t) \right], \tag{3.7}$$

where

$$V_j(s,t) = \left\{ \ell^{-1} \sum_{r=1}^{\ell} \left[U_{j+r}(s,t) - \big(F(t) - F(s)\big) \right] \right\}^2 - E\left\{ \ell^{-1} \sum_{r=1}^{\ell} \left[U_{j+r}(s,t) - \big(F(t) - F(s)\big) \right] \right\}^2 \tag{3.8}$$

and

$$V(s,t) = -(n-\ell+1)\left\{ \left[\widetilde{F}_n(t) - \widetilde{F}_n(s)\right] - \left[F(t) - F(s)\right] \right\}^2 + (n-\ell+1)E\left\{ \left[\widetilde{F}_n(t) - \widetilde{F}_n(s)\right] - \left[F(t) - F(s)\right] \right\}^2. \tag{3.9}$$

The following result establishes the strong consistency of $\widehat{\sigma}_n^2(s,t)$. Künsch (1989) gives sufficient conditions under which the strong consistency holds. Since we aim at proving uniform consistency, our method of proof is somewhat different than that of Künsch. We therefore include a proof here.

LEMMA 3.3. *Suppose that the mixing coefficients of the process* $\{X_n,\ n \geq 1\}$ *satisfy the following conditions*:

(a) $\sum (i+1)^7 \alpha(i) < \infty$;
(b) $\sum (i+1) \alpha(i)^{1/2} < \infty$.

Further, let $\ell(n) = \mathbb{O}(n^{1/2-\varepsilon})$, $0 < \varepsilon < \frac{1}{2}$. *Then, for every fixed* (s,t), $\widehat{\sigma}_n^2(s,t) \to \sigma^2(s,t)$ *a.s.*

PROOF. From (3.8) and the Minkowski inequality, we have

$$\begin{aligned} &E\left\{ \widehat{\sigma}_n^2(s,t) - E\left[\widehat{\sigma}_n^2(s,t)\right] \right\}^4 \\ &\quad \leq \ell^4 (n-\ell+1)^{-4} \left\{ \left[E\left(\sum_{j=0}^{n-\ell} V_j \right)^4 \right]^{1/4} + \left[E(V^4)\right]^{1/4} \right\}^4, \end{aligned} \tag{3.10}$$

where we write V_j and V for $V_j(s,t)$ and $V(s,t)$, respectively.

To obtain an upper bound for $E(\sum_{j=0}^{n-\ell} V_j)^4$, we note that the process $\{V_j, j \geq 0\}$ is a strictly stationary and strongly mixing process whose α-mixing coefficients are given by

$$\alpha^*(i) \leq \begin{cases} 1, & i \leq \ell, \\ \alpha(i-\ell), & i > \ell. \end{cases}$$

We use the more precise bound for $E(\sum_{j=0}^{n-\ell} V_j)^4$, as given by part (b) of Lemma

3.1. Thus, we have

$$
(3.11)\qquad
\begin{aligned}
E\left[\sum_{j=0}^{n-\ell} V_j\right]^4 \le C(n-\ell+1)\Big\{&A_{2,\nu}(\alpha^*)\|V_1\|_{2+\nu}^2 + A_{4,\nu}(\alpha^*)\|V_1\|_{3+\nu}^3 \\
&+ (n-\ell+1)\Big[A_{4,\nu}(\alpha^*)\|V_1\|_{4+\nu}^4 + \big[A_{2,\nu}(\alpha^*)\big]^2\|V_1\|_{2+\nu}^4\Big]\Big\}.
\end{aligned}
$$

By applying Lemma 3.1(a) to $\Sigma_{r=1}^{\ell}\{U_r(s,t) - [F(t) - F(s)]\}$ with $m = 16$, we get

$$
E\left\{\sum_{r=1}^{\ell}\Big(U_r(s,t) - \big[F(t) - F(s)\big]\Big)\right\}^{16} \le C\ell^8.
$$

Using the Minkowski inequality and simplifying, we get $E[V_1^8] \le C/\ell^8$. Now choose $\nu = 4$ in (3.11) and use $\|X\|_p \le \|X\|_q$ for $p \le q$. Noting that the first ℓ mixing coefficients of the process $\{V_j,\, j \ge 0\}$ are at most unity and after some simplifying calculations, we have

$$
(3.12)\qquad E\left[\sum_{j=0}^{n-\ell} V_j\right]^4 \le C\ell^{-2}(n-\ell+1)^2.
$$

Returning to the second term of (3.7), let $W = \{[\widetilde{F}_n(t) - \widetilde{F}_n(s)] - [F_n(t) - F_n(s)]\}^2$. Since $0 \le W \le 1$, it follows that $E(V^4) = (n-\ell+1)^4 E[W - E(W)]^4 \le (n-\ell+1)^4\{E(W^4) + 6E(W^2)[E(W)]^2\}$. Now, by the Minkowski inequality and Lemma 3.2(b), we conclude that

$$
\begin{aligned}
E(W^r) &= E\Big[\widetilde{F}_n(t) - \widetilde{F}_n(s) - \big(F(t) - F(s)\big)\Big]^{2r} \\
&\le C\left[\frac{\ell-1}{n-\ell+1} + \frac{1}{\sqrt{n}}\right]^{2r}, \qquad r = 1, 2 \text{ and } 4.
\end{aligned}
$$

Substituting these bounds above, we get

$$
(3.13)\qquad E(V^4) \le C(n-\ell+1)^4\left[\frac{\ell-1}{n-\ell+1} + \frac{1}{\sqrt{n}}\right]^8.
$$

Putting (3.13) and (3.12) into (3.10) and retaining the dominant terms only, we finally conclude that

$$
(3.14)\qquad E\Big\{\widehat{\sigma}_n^2(s,t) - E\big[\widehat{\sigma}_n^2(s,t)\big]\Big\}^4 = \mathbb{O}\left(\frac{\ell^2}{(n-\ell+1)^2}\right),
$$

so that the series $\sum E\{\widehat{\sigma}_n^2(s,t) - E[\widehat{\sigma}_n^2(s,t)]\}^4$ converges uniformly in (s,t), provided $\ell(n) = \mathbb{O}(n^{1/2-\varepsilon})$.

Further, if $\ell(n) = o(n^{1/2})$, it follows from Deo [(1973), Lemma 1 (with $r_1 = r_2 = 4$)] that

$$\left\| E\left[\hat{\sigma}_n^2(s,t)\right] - \sigma^2(s,t) \right\| \le C \sum_{k=1}^{\infty} k \frac{\alpha(k)^{1/2}}{\ell(n)}, \tag{3.15}$$

which converges to 0 in view of assumption (b). Also see Künsch [(1989), Theorem 3.2], in this connection. This, combined with the Borel–Cantelli lemma, completes the proof. □

We now use the above lemma to prove the uniform consistency of $\hat{\sigma}_n^2(s,t)$.

THEOREM 3.1.

(a) *Let* $\ell(n) = \mathbb{O}(n^{1/2-\varepsilon})$, $0 < \varepsilon < \frac{1}{2}$. *Assume that the conditions of Lemma* 3.3 *hold and that* X_1 *is uniformly distributed on* $[0,1]$. *Then,*

$$\left\| \hat{\sigma}_n^2(s,t) - \sigma^2(s,t) \right\| \to 0 \quad \textit{in probability}.$$

(b) *Further, if* $\varepsilon > \frac{1}{4}$, *we have*

$$\left\| \hat{\sigma}_n^2(s,t) - \sigma^2(s,t) \right\| \to 0 \quad \textit{a.s.}$$

PROOF. This proof is a suitable modification of proofs of the Glivenko–Cantelli theorem. From (3.8), we have

$$V_j(s,t) = \frac{\left(\sum_{r=1}^{\ell} U_{j+r}(s,t)\right)^2 - E\left(\left(\sum_{r=1}^{\ell} U_{j+r}(s,t)\right)^2\right)}{\ell^2} - 2(t-s)\frac{\sum_{r=1}^{\ell}\{U_{j+r}(s,t) - (t-s)\}}{\ell}$$

Now consider a grid $\{(i/M(n), p/M(n)) \mid i, p = 0, 1, \dots, M(n)\}$ of the unit rectangle. Using the above expression of $V_j(s,t)$ and arguing as in Shorack and Wellner [(1986), page 96], we see that

$$\begin{aligned}
&\left\| \ell(n-\ell+1)^{-1} \sum_{j=0}^{n-\ell} V_j(s,t) \right\| \\
&\quad \le \max_{0 \le i,p \le M(n)} \left| \ell(n-\ell+1)^{-1} \sum_{j=0}^{n-\ell} V_j\left(\frac{i}{M(n)}, \frac{p}{M(n)}\right) \right| \\
&\qquad + 8\ell \max_{0 \le i \le M(n)} \left| \tilde{F}_n\left(\frac{i}{M(n)}\right) - \frac{i}{M(n)} \right| + \frac{6\ell}{M(n)} \\
&\qquad + \max_{0 \le i,p \le M(n)} \left| E\left[\hat{\sigma}_n^2\left(\frac{i}{M(n)}, \frac{p-1}{M(n)}\right)\right] - E\left[\hat{\sigma}_n^2\left(\frac{i-1}{M(n)}, \frac{p}{M(n)}\right)\right] \right|.
\end{aligned} \tag{3.16}$$

Now, the continuity of $\mathcal{K}(s,t)$ implies the uniform continuity of $\sigma^2(s,t)$ on the unit rectangle. Thus, for a given $\delta > 0$, there exists an n_0 so that, for every $n \geq n_0$, with $M(n) = \mathbb{O}(n^{1/2-\varepsilon+\varepsilon'})$, we have, for every $0 \leq i, p \leq M(n)$,

$$\left|\sigma^2\left(\frac{i}{M(n)}, \frac{p-1}{M(n)}\right) - \sigma^2\left(\frac{i-1}{M(n)}, \frac{p}{M(n)}\right)\right| < \frac{\delta}{6},$$

and $\ell/M(n) < \delta/6$. In view of this and the fact that $\|E(\hat{\sigma}_n^2(s,t)) - \sigma^2(s,t)\| \to 0$ (as seen at the end of the proof of Lemma 3.3), the last two terms on the right-hand side of (3.16) converge to zero, as $n \to \infty$. The a.s. convergence of the second term to zero follows from lemma 3.2(c). Further, for $n \geq n_0$,

$$\begin{aligned} P&\left[\max_{0 \leq i,p \leq M(n)} \left|\ell(n-\ell+1)^{-1} \sum_{j=0}^{n-\ell} V_j\left(\frac{i}{M(n)}, \frac{p}{M(n)}\right)\right| > \frac{\delta}{2}\right] \\ &\leq CM(n)^2 \frac{\ell^2}{\delta^4(n-\ell+1)^2} \end{aligned} \tag{3.17}$$

in view of (3.12).

Thus, the left-hand side of (3.16) converges to 0 in probability, so long as $0 < \varepsilon < \frac{1}{2}$.

To prove the part (b), we note that the series $\sum_{n=n_0}^{\infty} [M(n)\ell]^2/(n-\ell+1)^2$ converges if one can choose a positive ε' such that $4\varepsilon - 2\varepsilon' > 1$ for a given $\varepsilon > 0$, $0 < \varepsilon < \frac{1}{2}$. This leads to the constraint on ε, as stated in the statement of the theorem. Thus, by the Borel–Cantelli lemma, the first term on the right-hand side of (3.16) converges to 0 a.s. Therefore,

$$\left\|\ell(n-\ell+1)^{-1} \sum_{j=0}^{n-\ell} V_j(s,t)\right\| \to 0 \quad \text{a.s.}$$

Now, both part (a) and part (b) follow by using Lemma 3.2(c), (3.7) and (3.15). □

COROLLARY 3.1. *Let $\mathcal{K}_n^*(s,t)$ denote the conditional covariance kernel of the process Z_n^*, that is,* $\mathcal{K}_n^*(s,t) = \mathrm{Cov}^*(Z_n^*(s), Z_n^*(t))$. *Then, under the conditions of the theorem,* $\|\mathcal{K}_n^*(s,t) - \mathcal{K}(s,t)\| \to 0$ *in probability or a.s. depending upon the choice of* ε.

PROOF. The proof easily follows from the above theorem and an expression for the bootstrap estimator similar to (3.5). □

Let

$$\Delta_n = \left\|\hat{\sigma}_n^2(s,t) - \sigma^2(s,t)\right\|. \tag{3.18}$$

COROLLARY 3.2. *Under the conditions of the theorem, there exists an $r > 0$ such that the following hold:*

(a) $n^r \Delta_n \to 0$ *in probability*;
(b) $n^r \Delta_n \to 0$ *a.s., if* $\varepsilon > \frac{1}{4}$.

PROOF. We prove (b) only; the proof of (a) follows similarly. Returning to (3.16) and multiplying each term by n^r, we see that arguments made in the proof of Theorem 3.1 are still applicable: there exist r and ε', $r > 0$, $\varepsilon' > 0$, such that (i) $4\varepsilon - 2\varepsilon' - 4r > 1$, (ii) $9\varepsilon - \varepsilon' - 8r > \frac{3}{2}$, (iii) $\varepsilon' > r$, (iv) $(b/2)(\frac{1}{2} - \varepsilon + \varepsilon') - r > 0$ and (v) $r < \frac{1}{2} - \varepsilon$ are satisfied simultaneously, if $\varepsilon > \frac{1}{4}$. Conditions (i)–(iii) correspond to the first three terms of the right-hand side of (3.16). The last condition ensures that $n^r/\ell \to 0$ and thus the term arising from the bias, namely, $n^r\|E\hat{\sigma}_n^2(s,t)] - \sigma^2(s,t)\| \to 0$ [cf. (3.15)]. The last term of the right-hand side of (3.16) is bounded above by the sum of the bias term and

$$\left|\sigma^2\left(\frac{i}{M(n)}, \frac{p-1}{M(n)}\right) - \sigma^2\left(\frac{i-1}{M(n)}, \frac{p}{M(n)}\right)\right|.$$

The last expression, by the triangle inequality on the semimetric $\sigma(s,t)$ and by (3.4), is at most equal to $2C(M(n))^{-b/2}$. Thus, n^r times the last term converges to 0, due to (iv) and (v). □

4. Proofs of the main results. To prove tightness of the bootstrap empirical process, we follow the approach based on chaining, as described in Pollard [(1984), Chapter 7]. [The chaining arguments are based on the work by Dudley, Le Cam and Giné and Zinn; for references, see Pollard (1984), page 167.]

We first assume that X_1, has a uniform distribution on [0, 1]. Let the corresponding limiting Gaussian process of (2.2) be denoted by Y. On $T = [0, 1]$, let $d(s,t) = \sigma(s,t)$ be the semimetric as defined by (3.5). Let

$$\varphi_n(\delta,\eta) = P^*\left[\sup_{d(s,t)\le\delta} \left|Z_n^*(t) - Z_n^*(s)\right| > \eta\right]. \tag{4.1}$$

To prove Theorem 2.2, we need to show that, for every $\eta > 0$, there exists a $\delta > 0$ such that $\limsup_n \varphi_n(\delta,\eta) = 0$ a.s. For Theorem 2.1, we need to show that, for every subsequence, there exists a further subsequence $\{n_k\}$ such that $\limsup_{n_k} \varphi_{n_k}(\delta,\eta) = 0$ a.s. This is established through the following lemmas, wherein we use Theorem 3.1, Corollary 3.2 and the fact that convergence in probability implies that almost-sure convergence holds over subsequences. For notational convenience, results are stated and proved in terms of the sequence itself.

LEMMA 4.1. *For every* $\lambda \in (0,1)$, *every* $\eta > 0$ *and* $\delta > 0$ *such that* $\delta^2/\eta > n^{-\varepsilon}/(2B^{-1}(\lambda))$ *and* $\delta^2 \ge \Delta_n$, *we have*

$$P^*\left[\left|Z_n^*(t) - Z_n^*(s)\right| > \eta\right] \le 2\exp\left\{-\frac{\eta^2(\lambda/2)}{2\delta^2}\right\},$$

provided $d(s,t) \le \delta$. [*Here*, $B(\cdot)$ *denotes the function on the right-hand side of Bennett's inequality*; *cf. Pollard* (1984), *page* 192)].

PROOF. We note that

$$Z_n^*(t) - Z_n^*(s) = \sqrt{\ell} \sum_{i=1}^{k} \frac{[H_{S_i}(t) - H_{S_i}(s)] - [\widetilde{F}_n(t) - \widetilde{F}_n(s)]}{\sqrt{k}}, \tag{4.2}$$

where $S_1, S_2, \ldots, S_k$ are i.i.d. random variables. From Bennett's inequality, we have

$$\begin{aligned} P^*\Big[\big|Z_n^*(t) - Z_n^*(s)\big| > \eta\Big] &\le 2\exp\left\{-\frac{1}{2}\frac{\eta^2}{\widehat{\sigma}_n^2(s,t)}B\left(\frac{\eta n^{-\varepsilon}}{\widehat{\sigma}_n^2(s,t)}\right)\right\} \\ &\le 2\exp\left\{-\frac{1}{2}\frac{\eta^2}{\Delta_n + \sigma^2(s,t)}B\left(\frac{\eta n^{-\varepsilon}}{\Delta_n + \sigma^2(s,t)}\right)\right\}, \end{aligned}$$

as $\widehat{\sigma}_n^2(s,t) \le \Delta_n + \sigma^2(s,t)$ and $\lambda B(\lambda)$ is an increasing function of λ.

Further, if $\sigma^2(s,t) \le \delta^2$ and $\Delta_n \le \delta^2$, the last expression cannot exceed

$$2\exp\left\{-\frac{1}{2}\frac{\eta^2}{2\delta^2}B\left(\frac{\eta n^{-\varepsilon}}{2\delta^2}\right)\right\},$$

which, in turn, is bounded above by $2\exp\{-\frac{1}{2}(\eta^2\lambda)/(2\delta^2)\}$, if $\delta^2/\eta \ge n^{-\varepsilon}/(2B^{-1}(\lambda))$. The last inequality follows since $B(0) = 1$ and B is a continuous, decreasing function. □

Let $N(\delta, d, T)$ be the covering number of T with respect to the δ-net of the semimetric d and let $J(\delta, d, T)$ be the corresponding covering integral [see Pollard (1984), page 160 for details].

LEMMA 4.2. *The covering integral $J(\delta, d, T)$ is finite and converges to* 0 *as* $\delta \to 0$.

PROOF. From (3.4), we recall that $\sigma^2(s,t) \le C|t-s|^b$, $\frac{1}{2} < b < 1$, for some positive C. We then note that $d_1(s,t) = C|t-s|^{b/2}$ is also a semimetric, as $\frac{1}{2} < b < 1$. The covering number $N(\delta, d_1, T)$ is clearly 1 plus the integer part of $(\delta/C)^{-2/b}$, which, in turn, dominates the covering number $N(\delta, d, T)$. □

Now we are in a position to apply Theorem VII.26 of Pollard [(1984), page 160] to the process Z_n^*. We take $\alpha^2 = \alpha_n^2 = \Delta_n + n^{-\varepsilon}/(2B^{-1}(\lambda))$, $\lambda = 1/4$ and $D = \sqrt{(2/\lambda)}$. In view of the above cited theorem, it remains to establish the following.

LEMMA 4.3. *For any* $\nu > 0$.

$$\limsup_{n\to\infty} P^*\left[\sup_{t\in T}\big|Z_n^*(t) - Z_n^*(t_\alpha)\big| > \nu\right] = 0 \quad a.s., \tag{4.3}$$

where $\alpha^2 = \alpha_n^2 = \Delta_n + n^{-\varepsilon}/(2B^{-1}(\lambda))$ *and* t_α *is the member of the* α*-net of* T [*with respect to* $d(s,t)$] *which is nearest to* t.

PROOF. We first observe that, given the sample $(x_1, x_2, \dots, x_n)$, $Z_n^*(t) - Z_n^*(s)$ assumes at most $(n+1)^2$ values as s and t vary in T. This is shown as follows.

For any two points t and t' with $F_n(t) = F_n(t')$, we have $I[x_r \le t] = I[x_r \le t']$, for all $r = 1, 2, \dots, n$. Thus, $H_i(t) - \widetilde{F}_n(t) = \ell^{-1}\sum_{r=1}^{\ell} I[x_{i+r} \le t] - \ell^{-1}(n-\ell+1)^{-1} \sum_{j=0}^{n-\ell}\sum_{r=1}^{\ell} I[x_{j+r} \le t] = H_i(t') - \widetilde{F}_n(t')$, for all $i = 0, 1, \dots, n-\ell$. Hence, for such two points, $H_{S_i}(t) - \widetilde{F}_n(t) = H_{S_i}(t') - \widetilde{F}_n(t')$ for all $i = 1, 2, \dots, k$. Consequently, $Z_n^*(t) = Z_n^*(t')$. Since F_n assumes only $n+1$ different values, the claim follows.

Therefore,

$$(4.4)\quad P^*\left[\sup_{t\in T} |Z_n^*(t) - Z_n^*(t_\alpha)| > \nu\right] \le (n+1)^2 \sup_{t\in T} P^*\left[|Z_n^*(t) - Z_n^*(t_\alpha)| > \nu\right].$$

Since $|H_{S_i}(t) - H_{S_i}(t_\alpha) - (\widetilde{F}_n(t) - \widetilde{F}_n(t_\alpha))| \le 2$ and

$$E^*\left[\left\{H_{S_i}(t) - H_{S_i}(t_\alpha) - \left(\widetilde{F}_n(t) - \widetilde{F}_n(t_\alpha)\right)\right\}^2\right] = \widehat{\sigma}_n^2(t, t_\alpha)\ell^{-1},$$

Bernstein's inequality [Pollard (1984), page 193] and expression (4.2) imply that

$$(4.5)\qquad P^*\left[|Z_n^*(t) - Z_n^*(t_\alpha)| > \nu\right] \le 2\exp\left\{-\frac{\frac{1}{2}\nu^2 k n^{-1}}{\widehat{\sigma}_n^2(t, t_\alpha)\ell^{-1} + \frac{2}{3}\nu n^{-1/2}}\right\}.$$

As $\widehat{\sigma}_n^2(t, t_\alpha) \le \Delta_n + \sigma^2(t, t_\alpha) \le \Delta_n + \alpha^2$, the probability in (4.5) is bounded above by

$$2\exp\left\{-\frac{\frac{1}{2}\nu^2}{\Delta_n + \alpha^2 + \frac{2}{3}\nu\ell n^{-1/2}}\right\}.$$

Now, Corollary 3.2, the definition of α and the condition on ℓ imply that, for some $r > 0$, $[\Delta_n + \alpha^2 + \frac{2}{3}\nu\ell n^{-1/2}] = \mathbb{O}(n^{-r})$. Thus,

$$P^*\left[\sup_{t\in T} |Z_n^*(t) - Z_n^*(t_\alpha)| > \nu\right] \le 2(n+1)^2 \exp(-Cn^r),$$

for some constant C. This completes the proof. □

PROOF OF THEOREM 2.1. It remains to show that the finite dimensional convergence holds almost surely over subsequences. This is done by closely following the proof of Theorem 3.5 of Künsch (1989).

From Corollary 3.1, it follows that, for every subsequence $\{n_j\}$, there exists a subsequence $\{n_{j'}\}$ such that $\|\mathcal{K}_{n_{j'}}^*(s,t) - \mathcal{K}(s,t)\| \to 0$ a.s. Let $T_{n_{j'}}^* = \sum_{i=1}^m a_i Z_{n_{j'}}^*(t_i)$. It follows that $\mathrm{Var}^*(T_{n_{j'}}^*)$ converges almost surely to $\mathrm{Var}(\sum_{i=1}^m a_i Y(t_i))$, the convergence being uniform in $t_1, t_2, \dots, t_m$. Thus, as in Künsch (1989), it remains to verify the Lindeberg condition for $T^*_{n_{j'}}/(\mathrm{Var}(\sum_{i=1}^m a_i Y(t_i)))^{1/2}$. This follows easily from the fact that

$$\max_{0\le j\le n-\ell}\left|\sum_{i=1}^{m} a_i \sum_{r=j+1}^{r=j+\ell}\left[U_r(t_i)-F(t_i)\right]\right| \le \ell\sum |a_i| = o(\sqrt{n}),$$

which verifies condition (B3) of Künsch (1989). This combined with the a.s. convergence of $\|\widetilde{F}_n - F\|$ to 0, implies the Lindeberg condition.

Thus, $Z^*_{n_{j'}} \Rightarrow Y$.

We now relax the assumption that X_1 has a uniform distribution on [0,1]. This is done in the following rather routine manner. Here, for convenience, we deal with the original sequence itself rather than a subsequence. Let G_n be a continuous version of F_n such that $\|G_n - F_n\| \le 1/n$, [cf. Billingsley (1968), page 104]. Since $\|G_n^{-1} - F\| \to 0$ a.s., where F is the uniform distribution function, $Z_n^*(G_n^{-1}) \Rightarrow Y$, in view of the continuous mapping theorem. Now, let F be any continuous distribution function. In view of the continuous mapping theorem and the fact that $\|G_n - F\| \to 0$ a.s., we readily see that $Z_n^*(G_n^{-1}(G_n)) = Z_n^* \Rightarrow Y(F)$ in the conditional setup of the theorem, $Y(F)$ being distributed like Z of (2.2). □

REMARK 4.1. Strong consistency of the bootstrap estimator of the sampling distribution of a compactly differentiable statistical functional T for a subsequence follows as in Gill (1989). Further, the limits do not depend upon the subsequences. This implies the weak consistency of the block-based bootstrap, that is,

$$\sup_x\left|P^*\left\{\sqrt{n}\left[T(F_n^*) - T(\widetilde{F}_n)\right] \le x\right\} - P\left\{\sqrt{n}\left[T(F_n) - T(F)\right] \le x\right\}\right| \to 0 \quad \text{in probability.}$$

PROOF OF THEOREM 2.2. The proof of Theorem 2.2 follows easily since the almost-sure finite dimensional convergence now holds (cf. Corollary 3.1), and Lemmas 4.1–4.3 are applicable to the sequence itself. □

REMARK 4.2. It needs to be pointed out that, when the distribution of $dT(F)\cdot Z$ is normal, Theorems 2.1 and 2.2 do not imply the consistency of the bootstrap estimator of the variance of the limiting normal distribution. The consistency of the bootstrap estimator may require further investigations based on additional assumptions, such as outlined in Künsch [(1989), Section 4.1]; also, see Ghosh, Parr, Singh and Babu (1984) and Rajarshi (1990). However, the theorem does imply that the interquartile range of the bootstrap distribution of $T(F_n^*)$ can be used to estimate the standard deviation of the limiting normal distribution. This suggestion is due to Parr (1985), which is based on the fact that the interquartile range, unlike the variance functional, is a continuous functional on the space of distribution functions.

Acknowledgments. Comments by both the referees have led to a considerable improvement over the first version. We thank them for pointing out major flaws in earlier versions. The short proof of Lemma 4.3 is essentially due to the second referee. Finally, we are also grateful to an Associate Editor and the Editors for their encouragement.

REFERENCES

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

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

BOSE, A. (1988). Edgeworth correction by bootstrap in autoregressions. *Ann. Statist.* **16** 1709–1722.

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

DEO, C. M. (1973). A note on empirical processes of strong-mixing sequences. *Ann. Probab.* **1** 870–875.

FREEDMAN, D. (1984). On bootstrapping two-stage least-square estimates in stationary linear models. *Ann. Statist.* **12** 827–842.

GASTWIRTH, J. L. and RUBIN, H. (1975). The behaviour of robust estimators on dependent data. *Ann. Statist.* **3** 1070–1100.

GHOSH, M., PARR, W. C., SINGH, K. and BABU, G. J. (1984). A note on bootstrapping the sample median. *Ann. Statist.* **12** 1130–1135.

GILL, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method (Part I). *Scand. J. Statist.* **16** 97–128.

KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241.

LIU, R. Y. and SINGH, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (R. Lepage and L. Billard, eds.) 225–248. Wiley, New York.

LIU, R. Y., SINGH, K. and LO, S. H. (1989). On a representation related to the bootstrap. *Sankhyā Ser. A* **51** 168–177.

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

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

RAJARSHI, M. B. (1990). Bootstrap in Markov sequences based on estimates of transition density. *Ann. Inst. Statist. Math.* **42** 253–268.

SHORACK, G. (1982). Bootstrapping robust regression. *Comm. Statist. Theory Methods* **11** 961–972.

SHORACK, G. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

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

SWANEPOEL, J. W. H. (1986). A note on proving that the (modified) bootstrap works. *Comm. Statist. Theory Methods* **15** 3193–3203.

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

YOKOYAMA, R. (1980). Moment bounds for stationary mixing sequences. *Z. Wahrsch Verw. Gebiete* **52** 45–57.

DEPARTMENT OF STATISTICS
UNIVERSITY OF POONA
PUNE 411007
INDIA