A general theory for large sample confidence regions based on subsamples under minimal assumptions

by

Dimitris N. Politis
Department of Statistics
Purdue University

Joseph P. Romano
Department of Statistics
Stanford University

Technical Report #92-36

Department of Statistics
Purdue University

August, 1992

Research partially supported by a Purdue Research Foundation Faculty Grant and NSF Grant DMS89-57217

*

A general theory for large sample confidence regions based on subsamples under minimal assumptions

Dimitris N. Politis and Joseph P. Romano
Department of Statistics, Purdue University
and Department of Statistics, Stanford University

Abstract

In this article, the construction of confidence regions by approximating the sampling distribution of some statistic is studied. The true sampling distribution is estimated by an appropriate normalization of the values of the statistic computed over subsamples of the data. In the i.i.d. context, the method has been studied by Wu (1990) in regular situations where the statistic is asymptotically normal. The goal of the present work is to prove the method yields asymptotically valid confidence regions under minimal conditions. Essentially all that is required is that the statistic, suitably normalized, possesses a limit distribution under the true model. Unlike the bootstrap, the convergence to the limit distribution need not be uniform in any sense. In fact, our method corrects the failure of the bootstrap in the several known counterexamples to the bootstrap. The method is readily adapted to parameters of stationary time series or, more generally, homogeneous random fields. In this context, analogous results are proved under the weak hypothesis of convergence in distribution of a normalized statistic. For example, an immediate application is the construction of a confidence interval for the spectral density function of a homogeneous random field.

1. INTRODUCTION

In this article, a general theory for the construction of confidence intervals or regions is presented. The basic idea is to approximate the sampling distribution of a statistic based on the values of the statistic computed over smaller subsets of the data. For example, in the case where the data are n observations which are independent and identically distributed, a statistic is computed based on the entire data set and is recomputed over all $\binom{n}{b}$ data sets of size b. Implicit is the notion of a statistic sequence, so that the statistic is defined for samples of size n and b. These recomputed values of the statistic are suitably normalized to approximate the true sampling distribution.

The approach presented here is perhaps the most general theory for the construction of first order asymptotically valid confidence regions. That is, it will be seen that, under very weak assumptions on b, the method is valid whenever the original statistic, suitably normalized, has a limit distribution under the true model. Other methods, such as the bootstrap, require that the distribution of the statistic is somehow locally smooth as a function of the unknown model. In fact, many papers have been devoted to showing the convergence of a suitably normalized statistic to its limiting distribution is appropriately uniform as a function of the unknown model in specific situations. In contrast, no such assumption or verification of such smoothness is required in our theory. Indeed, the method here is applicable even in the several known situations which represent counterexamples to the bootstrap. To appreciate why our method behaves well under such weak assumptions, note that each subset of size b (taken without replacement from the original data) is indeed a sample of size b from the true model. Hence, it should be intuitively clear that one can at least approximate the sampling distribution of the (normalized) statistic based on a sample of size b. But, under the weak convergence hypothesis, the sampling distributions based on samples of size b and n should be close. The bootstrap, on the other hand, is based on recomputing a statistic over a sample of size n from some estimated model which is hopefully close to the true model.

The method has a clear extension to the context of a stationary time series or, more generally, a homogeneous random field. The only difference is that the statistic is computed over a smaller number of subsets of the data that retain the dependence structure of the observations. For example, if $X_1, \ldots, X_n$ represent n observations from some stationary time series, the statistic is recomputed over the $n - b + 1$ subsets of size b of the form $\{X_i, X_{i+1}, \ldots, X_{i+b-1}\}$. The obvious extension to homogeneous random fields will be described later.

The use of subsample values to approximate the variance of a statistic is well-known. The Quenouille-Tukey jackknife estimates of bias and variance based on computing a statistic over all subsamples of size $n-1$ has been well-studied and is closely related to the mean and variance of our estimated sampling distribution with $b = n - 1$. Half sampling methods have been well-studied in the context of sampling theory; see McCarthy (1969). Hartigan (1969) has introduced what Efron (1982) calls a random subsampling method, which is based on the computation of a statistic over all $2^n - 1$ nonempty subsets of the data. His method is seen to produce exact confidence limits in the special context of the symmetric location problem. Hartigan (1975) has adapted his finite sample results to a more general context of certain classes of estimators which have asymptotic normal distributions. Even in this context, the asymptotic results assume the number of subsamples used to recompute the statistic remains fixed as $n \to \infty$. Both the jackknife and random subsampling methods are similar in that they both use subsets of the data to approximate standard errors of a statistic, or perhaps even to approximate a sampling distribution. The method presented here retains the conceptual simplicity of these methods and is seen to be applicable under very minimal assumptions.

Efron's (1979) bootstrap, while sharing some similar properties to the aforementioned methods, has corrected some deficiencies in the jackknife, and has tackled the more ambitious goal of approximating an entire sampling distribution. Shao and Wu (1989) have shown that, by basing a jackknife estimate of variance on the statistic computed over subsamples with d observations deleted, many of the deficiences of the usual $d = 1$ jackknife estimate of variance can be removed. Later, Wu (1990) used these subsample values to approximate an entire sampling distribution by what he calls a jackknife histogram, but only in regular i.i.d. situations where the statistic is appropriately linear so that asymptotic normality ensues. Here, we show how these subsample values can accurately estimate a sampling distribution without any assumptions of asymptotic normality, by only assuming the existence of a limiting distribution. In summary, while the method developed in this paper is quite related to several well-studied techniques, the simplicity of our arguments has lead to asymptotic justification under the most general conditions.

In addition, we extend our results to the setting of stationary time series and homogeneous random fields. In this case, the existence of a limiting distribution and a very weak mixing condition yields asymptotically valid estimates of the true sampling distribution. In the context of a stationary time series, Carlstein (1986) has considered the problem of estimating the variance of a statistic based on the values of the statistic computed over subseries. Here, we develop consistent properties for an estimated sampling distribution under weaker assumptions.

The main drawback to our method as presented is its lack of second order correctness. However, Tu (1992) has shown how, in some situations where Edgeworth expansions are valid, the approximation of a sampling distribution based on jackknife pseudo-values can be appropriately modified to yield second order accuracy. Here, we have insisted upon a general first order theory, but Tu's work has demonstrated the possibility that our method can be adapted to yield desirable higher order properties.

In Section 2, the method is described in the context of i.i.d. observations. The main theorem is presented and several examples are given. Some comparisons with the bootstrap are drawn. In Section 3, the method is adapted to homogeneous random fields. The theorem yields such general asymptotic results under such weak assumptions, that the problem of constructing a confidence interval for the spectral density function of a homogeneous random field is an immediate application. In addition, the problem of bias reduction using the subsampling method is discussed.

2. GENERAL THEOREM IN THE I.I.D. CASE

2.1 The Basic Theorem.

Throughout this section, $X_1, \ldots, X_n$ is a sample of n independent and identically distributed random variables taking values in an arbitrary sample space S. The common probability measure generating the observations is denoted P. The goal is to construct a confidence region for some parameter $\theta(P)$. For now, assume θ is real-valued, but this can be considerably generalized to allow for the construction of confidence regions for multivariate parameters or confidence bands for functions.

Let $T_n = T_n(X_1, \ldots, X_n)$ be an estimator of $\theta(P)$. It is desired to estimate or approximate the true sampling distribution of T_n in order to make inferences about θ. Nothing is assumed about the form of the estimator, though it is natural in the i.i.d. context to assume T_n is symmetric in its arguments.

Define $J_n(P)$ to be the sampling distribution of $\tau_n(T_n - \theta(P))$ based on a sample of size n from P. Also define the corresponding cumulative distribution function:

$$J_n(x, P) = Prob_P\{\tau_n[T_n(X_1, \ldots, X_n) - \theta(P)] \leq x\}.$$

Essentially, the only assumption that we will need to construct asymptotically valid confidence intervals for $\theta(P)$ is the following.

Assumption A. $J_n(P)$ converges weakly to a limit law $J(P)$ as $n \to \infty$.

Assumption A will be required to hold for some sequence τ_n. It will be necessary, however, that τ_n is such that the limit law $J(P)$ is nondegenerate.

To describe the method studied in this section, let $Y_1, \ldots, Y_{N_n}$ be equal to the $N_n = \binom{n}{b}$ subsets of $\{X_1, \ldots, X_n\}$, ordered in any fashion. Only a very weak assumption on b will be required. In typical situations, it will be assumed that $b/n \to 0$ and $b \to \infty$ as $n \to \infty$. Now, let $S_{n,i}$ be equal to the statistic T_b evaluated at the data set Y_i. The approximation to $J_n(x, P)$ we study is defined by

$$L_n(x) = N_n^{-1} \sum_{i=1}^{N_n} 1\{\tau_b(S_{n,i} - T_n) \leq x\}. \tag{2.1}$$

The motivation behind the method is the following. For any i, Y_i is a random sample of size b from P. Hence, the *exact* distribution of $\tau_b(S_{n,i} - \theta(P))$ is $J_b(P)$. The empirical distribution of the N_n values of $\tau_b(S_{n,i} - \theta(P))$ should then serve as a good approximation to $J_n(P)$. Of course, $\theta(P)$ is unknown, so we replace $\theta(P)$ by T_n, which is asymptotically permissible because $\tau_b(T_n - \theta(P))$ is of order $\tau_b/\tau_n \to 0$.

Now, specialize to the case when $T_n = \bar{X}_n$ is the sample mean of n i.i.d. observations having mean θ. Assuming the variance of X_i is finite, so $\tau_n = n^{1/2}$. One can compute the variance of the distribution $L_n(\cdot)$, and is given by $[(n-b)/n] \cdot S_n^2$, where $S_n^2 = \sum_i (X_i - \bar{X}_n)^2/(n-1)$. In this case, this variance depends on b, due to the dependencies of the $S_{n,i}$ with each other and with T_n. So, it seems desirable to then replace τ_b in (2.1) by $\tau_b \cdot [n/(n-b)]^{1/2}$ so that the resulting approximating distribution has variance S_n^2, independent of b. In our theorems, $n/(n-b) \to 1$ so that the first order asymptotic properties are the same in either case. By making this change, however, our estimator more closely resembles that of Wu (1990), and the variance of the approximating distribution corresponds to Shao and Wu's (1989) delete $d = n - b$ jackknife estimate of variance. Henceforth, however, we do not modify τ_n in (2.1) because the above modification is only justified for linear statistics. Keep in mind, however, that such a modification may improve the approximation in some situations.

The above example also helps to explain the lack of consideration of using jackknife pseudo values to approximate a sampling distribution; a notable exception is Wu (1990). In the sample mean case, if $b = n - 1$ and if Y_i is the original sample with X_i deleted, then $(n-1)[S_{n,i} - T_n] = (\bar{X}_n - X_i)$. The empirical distribution of these values converges uniformly to the distribution of $\theta - X_1$ with probability one. Hence, the use of our technique (with Wu's scaling) results in inconsistency, except in the special case that X_1 is normally distributed. Also, no choice of rescaling can correct the problem. This example points to the failure of the use of the traditional jackknife pseudo values when only one observation is deleted at a time. See example 2.2.2 for an example where taking b of the same order as n results in inconsistency. In the opposite extreme, if $b = 1$ and $Y_i = X_i$, then $S_{n,i} - T_n = X_i - \bar{X}_n$, and the empirical distribution of these values converges uniformly to the distribution of $X_1 - \theta$ with probability one; inconsistency follows. Aside from these extreme choices of b, our theory is typically applicable when $b/n \to 0$ and $b \to \infty$.

Theorem 2.1. *Assume Assumption A. Also assume $\tau_b/\tau_n \to 0$, $b \to \infty$ and $b/n \to 0$ as $n \to \infty$. Let x be a continuity point of $J(\cdot, P)$.*

(i). Then, $L_n(x) \to J(x, P)$ in probability.

(ii). If $J(\cdot, P)$ is continuous, then

$$\sup_x |L_n(x) - J_n(x, P)| \to 0 \tag{2.2}$$

in probability.

(iii). Let $c_n(1-\alpha) = \inf\{x : L_n(x) \geq 1-\alpha\}$. Correspondingly, define $c(1-\alpha, P) = \inf\{x : J(x, P) \geq 1-\alpha\}$. If $J(\cdot, P)$ is continuous at $c(1-\alpha, P)$, then

$$Prob_P\{\tau_n[T_n - \theta(P)] \leq c_n(1-\alpha)\} \to 1-\alpha \tag{2.3}$$

as $n \to \infty$. Thus, the asymptotic coverage probability under P of the interval $[T_n - \tau_n^{-1} c_n(1-\alpha), \infty)$ is the nominal level $1-\alpha$.

(iv). Assume, for every $d > 0$, $\sum_n exp\{-d[n/b]\} < \infty$ and $\tau_b(T_n - \theta(P)) \to 0$ almost surely. Then, the convergences in (i) and (ii) hold with probability one.

Proof. Let

$$U_n(x) = N_n^{-1} \sum_{i=1}^{N_n} 1\{\tau_b[S_{n,i} - \theta(P)] \leq x\}.$$

To prove (i), it suffices to show $U_n(x)$ converges in probability to $J(x, P)$ for every continuity point x of $J(x, P)$. To see why,

$$L_n(x) = N_n^{-1} \sum_i 1\{\tau_b[S_{n,i} - \theta(P)] + \tau_b[\theta(P) - T_n] \leq x\},$$

so that for every $\epsilon > 0$,

$$U_n(x-\epsilon)1(E_n) \leq L_n(x)1(E_n) \leq U_n(x+\epsilon),$$

where $1(E_n)$ is the indicator of the event $E_n \equiv \{\tau_b|\theta(P) - T_n| \leq \epsilon\}$. But, the event E_n has probability tending to one. So, with probability tending to one,

$$U_n(x-\epsilon) \leq L_n(x) \leq U_n(x+\epsilon)$$

for any $\epsilon > 0$. hence, if $x+\epsilon$ and $x-\epsilon$ are continuity points of $J(\cdot, P)$, then $U_n(x \pm \epsilon) \rightarrow J(x \pm \epsilon, P)$ in probability implies

$$J(x-\epsilon, P) - \epsilon \leq L_n(x) \leq J(x+\epsilon, P) + \epsilon$$

with probability tending to one. Now, let $\epsilon \rightarrow 0$ so that $x \pm \epsilon$ are continuity points of $J(\cdot, P)$. Therefore, it suffices to show $U_n(x) \rightarrow J(x, P)$ in probability for all continuity points x of $J(\cdot, P)$. But, $U_n(x)$ is a U-statistic of degree b. Also, $0 \leq U_n(x) \leq 1$ and $E[U_n(x)] = J_b(x, P)$. By an inequality of Hoeffding (1963) (see Serfling (1980), Theorem A, p.201): for any $t > 0$,

$$Prob_P\{U_n(x) - J_b(x, P) \geq t\} \leq exp\{-2[n/b]t^2\}. \tag{2.4}$$

One can obtain a similar inequality for $t < 0$ by considering the U-statistic $-U_n(x)$. Hence, $U_n(x) - J_b(x, P) \rightarrow 0$ in probability. The result (i) follows since $J_b(x, P) \rightarrow J(x, P)$. To prove (ii), given any subsequence $\{n_k\}$, one can extract a further subsequence $\{n_{k_j}\}$ so that $L_{n_{k_j}}(x) \rightarrow J(x, P)$ almost surely. Hence, $L_{n_{k_j}}(x) \rightarrow J(x, P)$ almost surely for all x in some countable dense set of the real line. So, $L_{n_{k_j}}$ tends weakly to $J(x, P)$ and this convergence is uniform by Polya's theorem. Hence, the result (ii) holds. The proof of (iii) is very similar to the proof of Theorem 1 of Beran (1984) given our result (i). To prove (iv), follow the same argument, using the added assumptions and the Borel-Cantelli Lemma on the inequality (2.4).

Remark 2.1. The assumptions $b/n \rightarrow 0$ and $b \rightarrow \infty$ need not imply $\tau_b/\tau_n \rightarrow 0$. For example, in the unusual case $\tau_n = \log(n)$, if $b = n^\gamma$ and $\gamma > 0$, the assumption $\tau_b/\tau_n \rightarrow 0$ is not satisfied. In regular cases, $\tau_n = n^{1/2}$, and the assumptions on b simplify to $b/n \rightarrow 0$ and $b \rightarrow \infty$. The further assumption on b in part (iv) of the Theorem will then hold, for example, if $b = n^\gamma$ for any $\gamma \in (0, 1)$. In fact, it is easy to see that it holds if $b \log(n)/n \rightarrow 0$.

Remark 2.2. The assumptions on b are as weak as possible under the weak assumptions of the theorem. However, in some cases, the choice $b = O(n)$ yields similar results; this occurs in Wu (1990), where the statistic is approximately linear with an asymptotic Gaussian distribution and $\tau_n = n^{1/2}$. This choice will not work in general; see example 2.2.2.

Assumption A is satisfied in numerous examples. Next, we offer an interesting example which illustrates the scope of our method, as it falls outside the range of $n^{1/2}$-consistent estimators and normal limits. While methods like the bootstrap are potentially applicable in this example, the validity of the bootstrap is not known. At the very least, the bootstrap method would require a somewhat tedious argument to justify its asymptotic validity.

Example 2.1.1. Optimal replacement time. Consider the problem of age replacement where replacements of a unit X occur at failure of the unit or at age t, whichever comes first. X is assumed continuous with an increasing failure rate distribution F having density f, finite mean, and $F(0) = 0$. Suppose a cost c_1 is incurred for each failed unit which is replaced and a cost $c_2 < c_1$ is suffered for each nonfailed unit which is exchanged. It is easy to see that the average cost per time unit, over an infinite time horizon, based on the strategy of preventively replacing the unit at time t is given by

$$A(t, F) = \frac{c_1 F(t) + c_2[1 - F(t)]}{\int_0^t [1 - F(x)]dx}.$$

The problem is to find $\theta(F)$ which minimizes $A(t, F)$ over t. If $r(x) = f(x)/[1 - F(x)]$ is the failure rate of F, then if $r(x)$ is assumed continuous and increasing to ∞, then $\theta(F)$ is well-defined. The optimal minimum cost is then

$$\beta(F) = (c_1 - c_2)r(\theta(F)).$$

In practice, F is unknown, so our problem is to construct a confidence interval for $\theta(F)$ based on a random sample $X_1, \ldots, X_n$ from F. Let $\hat{F}_n$ denote the empirical distribution of the data, and let T_n be a value of t minimizing $A(t, \hat{F}_n)$; that is, $T_n = \theta(\hat{F}_n)$. For the purposes of the discussion here, don't worry about problems of existence or uniqueness; see Arunkumar (1972) for a careful description. Arunkumar (1972) has shown that $n^{1/3}[T_n - \theta(F)]$ has a nondegenerate limiting distribution, so our Condition A is verified with $\tau_n = n^{1/3}$. The asymptotic distribution is the distribution of $c(F)$ times the value of t which minimizes $[W(t) - t^2]$, where $W(t)$ is a two-sided Wiener-Lévy process and the constant $c(F)$ depends on intricate properties of F such as $f(\theta(F))$. Hence, the asymptotic distribution is of little use towards the construction of confidence intervals for $\theta(F)$. Léger and Cléroux (1990) have constructed bootstrap confidence intervals for $\beta(F)$. The approach here may be used for this problem as well because $n^{1/2}[\beta(\hat{F}_n) - \beta(F)]$ has a limiting normal distribution.

2.2. Comparison With The Bootstrap.

The usual bootstrap approximation to $J_n(x, P)$ is $J_n(x, \hat{Q}_n)$, where $\hat{Q}_n$ is some estimate of P. In many (but not all) i.i.d. situations, $\hat{Q}_n$ is taken to be the empirical distribution of the sample $X_1, \ldots, X_n$. The analogous results to (2.2) and (2.3) with $L_n(x)$ replaced by $J_n(x, \hat{Q}_n)$ have been proved in many situations; see Bickel and Freedman (1981) and Beran (1984). In fact, dozens of other papers exist whose sole purpose is to prove such results in very specific situations. Our theorem immediately applies very generally with no further work.

To elaborate a little further, analogous bootstrap limit results are typically proved in the following manner. For some choice of metric (or pseudo-metric) d on the space of probability measures, it must be known that $d(P_n, P) \to 0$ implies $J_n(P_n)$ converges weakly to $J(P)$. That is, Assumption A must be strengthed so that the convergence of $J_n(P)$ to $J(P)$ is suitably locally uniform in P. In addition, the estimator $\hat{Q}_n$ must then be known to satisfy $d(\hat{Q}_n, P) \to 0$ almost surely or in probability under P. In contrast, no such strengthing of Assumption A is required in Theorem 2.1. In the known counterexamples to the bootstrap, it is precisely a certain lack of uniformity in convergence which leads to failure of the bootstrap.

In some special cases, it has been realized that a sample size trick can often remedy the inconsistency of the bootstrap. To describe how, focus on the case where $\hat{Q}_n$ is the empirical measure, denoted $\hat{P}_n$. Rather than approximating $J_n(P)$ by $J_n(\hat{P}_n)$, the suggestion is to approximate $J_n(P)$ by $J_b(\hat{P}_n)$ for some b which usually satisfies $b/n \to 0$ and $b \to \infty$. The resulting estimator $J_b(x, \hat{P}_n)$ is obviously quite similar to our $L_n(x)$ given in (2.1). In words, $J_b(x, \hat{P}_n)$ is the bootstrap approximation defined by the distribution (conditional on the original data) of $\tau_b[T_b(X_1^*, \ldots, X_b^*) - T_n]$, where $X_1^*, \ldots, X_b^*$ are chosen with replacement from $X_1, \cdots, X_n$. In contrast, $L_n(x)$ is the distribution (conditional on the data) of $\tau_b[T_b(Y_1^*, \ldots, Y_b^*) - T_n)]$, where $Y_1^*, \ldots, Y_b^*$ are chosen *without* replacement from $X_1, \ldots, X_n$. Clearly, these two approaches must be similar if b is so small that sampling with and without replacement are essentially the same. Indeed, if one resamples b numbers (or indices) from the set $\{1, \ldots, n\}$, then the chance that none of the indices is duplicated is $\Pi_{i=1}^{b-1}(1 - \frac{i}{n})$. This probability tends to 0 if $b^2/n \to 0$. (To see why, take logs and do a Taylor expansion analysis.) Hence, the following is true.

Corollary 2.1. *Under the further assumption that $b^2/n \to 0$, parts (i)- (iii) of Theorem 2.1 remain valid if $L_n(x)$ is replaced by the bootstrap approximation $J_b(x, \hat{P}_n)$.*

In spite of the Corollary, we point out that L_n is more generally valid. Indeed, without the assumption $b^2/n \to 0$, $J_b(x, \hat{P}_n)$ can be inconsistent. To see why, let P be any distribution on the real line with a density (with respect to Lebesgue measure). Consider any statistic T_n, τ_n and $\theta(P)$ satisfying Assumption A. Even the sample mean will work here. Now, modify T_n to $\tilde{T}_n$ so that the statistic $\tilde{T}_n(X_1, \ldots, X_n)$ completely misbehaves if any of the observations $X_1, \ldots, X_n$ are identical. The bootstrap approximation to the distribution of $\tilde{T}_n$ must then misbehave as well unless $b^2/n \to 0$, while the consistency of L_n remains intact.

The above example, though artificial, was designed to illustrate a point. Below, some known counterexamples to the bootstrap are reviewed.

Example 2.2.1. U-statistics of degree 2. Let $X_1, \ldots, X_n$ be i.i.d. on the line with c.d.f. F. Denote by $\hat{F}_n$ the empirical distribution of the data. Let $\theta(F) = \int \int \omega(x, y) dF(x) dF(y)$, and assume $\omega(x, y) = \omega(y, x)$. Assume $\int \omega^2(x, y) dF(x) dF(y) < \infty$. Set $\tau_n = n^{1/2}$ and $T_n = \theta(\hat{F}_n)$. Then, it is well known that $J_n(F)$ converges weakly to $J(F)$, the normal distribution with mean 0 and variance given by

$$v^2(F) = 4[\int \{\omega(x, y) dF(y)\}^2 dF(x) - \theta^2(F)].$$

Hence, our condition A is satisfied. However, in order for the bootstrap to succeed, the additional condition $\int \omega^2(x, x) dF(x) < \infty$ is required. Bickel and Freedman (1981) give a counterexample to show the inconsistency of the bootstrap without this additional condition.

Interestingly, the bootstrap may fail even if $\int \omega^2(x, x) dF(x) < \infty$, stemming from the possibility that $v^2(F) = 0$. (Otherwise, Bickel and Freedman's argument justifies the bootstrap.) As an example, let $w(x, y) = xy$. In this case, $\theta(\hat{F}_n) = \bar{X}_n - S_n^2/n$, where S_n^2 is the usual unbiased sample variance. If $\theta(F) = 0$, then $v(F) = 0$. Then, $n[\theta(\hat{F}_n) - \theta(F)]$ converges weakly to $\sigma^2(F)(Z^2 - 1)$, where Z denotes a standard normal random variable and $\sigma^2(F)$ denotes the variance of F. However, it is easy to see that the bootstrap approximation to the distribution of $n[\theta(\hat{F}_n) - \theta(F)]$ has a representation $\sigma^2(F) Z^2 + 2Z\sigma(F) n^{1/2} \bar{X}_n$. Thus, failure of the bootstrap follows.

In the context of U-statistics, the possibility of using a reduced sample size in the resampling has been considered in Bretagnolle (1983); an alternative correction is given by Arcones (1990).

Example 2.2.2. Extreme order statistic. Bickel and Freedman (1981) provide the following counterexample. If $X_1, \ldots, X_n$ are i.i.d. uniform on $(0, \theta)$, then $n[\max(X_1, \ldots, X_n) - \theta]$ has a limit distribution given by the distribution of $-\theta X$, where X is exponential with mean one. Hence, Assumption A is satisfied here. However, the usual bootstrap fails. Note in Theorem 2.1 that the conditions on b (with $\tau_n = n$) reduce to $b/n \to 0$ and $b \to \infty$. In this example, at least, it is clear that we cannot assume $b/n \to c$, where $c > 0$. Indeed, $L_n(x)$ places mass b/n at 0. Thus, while it is sometimes true that, under further conditions such as Wu (1990) assumes, we can assume b is of the same order as n, this example makes it clear that we cannot in general weaken our assumptions on b without assuming further structure.

Example 2.2.3. The mean in the infinite variance case. Let $X_1, \ldots, X_n$ be i.i.d. real-valued random variables with c.d.f. F and mean $\theta(F)$. If the variance of F is not finite, the bootstrap is known to fail; see Babu (1984), Athreya (1987), Knight (1989) and Kinateder (1992). In this example, it has been realized that taking a smaller bootstrap sample size can result in consistency of the bootstrap; see Arcones (1990), who attributes the idea to an unpublished report of Athreya; also see Wu, Carlstein and Cambanis (1989) and Arcones and Giné (1989). For our method, Theorem 2.1 is generally applicable if F is in the domain of attraction of a stable law of index greater than one.

Example 2.2.4. Density estimation. Let $X_1, \ldots, X_n$ be i.i.d. real-valued random variables with density f. Suppose f is smooth and unimodal, with mode denoted by $\theta(f)$. Let $\hat{f}_{n,h}(t) = (nh)^{-1} \sum_i K[(t - X_i)/h]$, and let $\hat{\theta}_{n,h}$ denote a mode of $\hat{f}_{n,h}$. Under regularity $(nh^3)^{1/2}(\hat{\theta}_{n,h} - \theta)$ has a limiting normal distribution, so Theorem 2.1 is applicable. The optimal choice for h in this case is to take h proportional to $n^{-1/7}$. For such a choice of h, bootstrap methods based on resampling from the empirical distribution are not consistent. Indeed, bootstrap methods based on resampling from the empirical will work if the choice of bandwidth h for estimating the mode is suboptimal. In the optimal choice of bandwidth case, the bootstrap can be fixed by resampling from $\hat{f}_{n,g}$, where g is large compared with h. This phenomenon of having to resample from an appropriately oversmooth density estimate occurs in many other contexts involving functionals of a density. In short, the results depend upon delicate choices of smoothing parameters. For details, see Romano (1988).

Example 2.2.5. Superefficient estimator. Consider Hodges' famous example of a superefficient estimator. In this example, $X_1, \ldots, X_n$ are i.i.d. according the normal distribution with mean θ and variance one. Let $T_n = c\bar{X}_n$ if $|\bar{X}_n| \leq n^{-1/4}$ and $T_n = \bar{X}_n$ otherwise. Here, $c > 0$. As is well known, $n^{1/2}(T_n - \theta)$ has a limit distribution for every θ, so the conditions for our Theorem 2.1 remain applicable. However, Beran (1984) shows that the sampling distribution of $n^{1/2}(T_n - \theta)$ cannot be bootstrapped, even if one is willing to apply a parametric bootstrap!

2.3. Stochastic Approximation. Because $\binom{n}{b}$ may be large, L_n may be difficult to compute. Instead, a stochastic approximation may be employed. For example, let $I_1, \ldots I_s$ be chosen randomly with or without replacement from $\{1, 2, \ldots, N_n\}$. Then, $L_n(x)$ may be approximated by

$$\hat{L}_n(x) = s^{-1} \sum_{i=1}^{s} 1\{\tau_b(S_{n,I_i} - T_n) \leq x\}.$$

Corollary 2.2. *Under the assumptions of Theorem 2.1 and the assumption $s \to \infty$ as $n \to \infty$, the results of Theorem 2.1 are valid if $L_n(x)$ is replaced by $\hat{L}_n(x)$.*

Proof. In the case the I_i are sampled with replacement, $\sup_x |\hat{L}_n(x) - L_n(x)| \to 0$ almost surely by the Dvoretzky, Kiefer, Wolfowitz inequality; see Serfling (1980, p. 59). This result is also true in the case the I_i are sampled without replacement by a similar inequality; see Romano (1989).

An alternative approach, which also requires fewer computations, is the following. Rather than considering all $\binom{n}{b}$ subsamples of size b from $X_1, \ldots, X_n$, just use the $n-b+1$ subsamples of size b of the form $\{X_i, X_{i+1}, \ldots, X_{i+b-1}\}$. Notice, some ordering of the data is fixed and retained in the subsamples. Indeed, this is the approach that is applied in the next section for time series data. Even in the i.i.d. case, this approach may be desirable to ensure robustness against possible serial correlation. Most inferential procedures based on i.i.d. models are simply not valid (i.e., not even first order correct) if the independence assumption is violated, so it seems worthwhile to account for possible dependencies in the data if we do not sacrifice too much in efficiency.

2.4. General Parameters and Other Choices of Root.

In general, it may be desirable to approximate the sampling distribution of other roots. Below, two common choices are considered. The first generalization concerns the approximation of studentized root. The second generalization applies to general parameters which need not be real-valued. In particular, the setup is designed to handle confidence bands for functional parameters. We leave it to the reader to consider further obvious generalizations whose asymptotic validity may be justified by mimicking the proof of Theorem 2.1.

2.4.1. Studentized roots. Here, the goal is to approximate the distribution of $\tau_n[T_n - \theta(P)]/\hat{\sigma}_n$, where $\hat{\sigma}_n$ is some estimate of scale. Let $\hat{\sigma}_{n,i}$ be equal to the estimate of scale based on the ith subsample of size b from the original data. Analogous to (2.1), define

$$K_n(x) = N_n^{-1} \sum_{i=1}^{N_n} 1\{\tau_b(S_{n,i} - T_n)/\hat{\sigma}_{n,i} \leq x\}. \tag{2.5}$$

Then, the following theorem holds. The proof is similar to that of Theorem 2.1, and so it is omitted.

Theorem 2.2. *Assume Assumption A. Also assume* $\tau_b/\tau_n \to 0$, $b \to \infty$ *and* $b/n \to 0$ *as* $n \to \infty$. *Assume* $\hat{\sigma}_n \to \sigma$ *in probability, where* $\sigma = \sigma(P)$ *is a positive constant. Let* $x \cdot \sigma(P)$ *be a continuity point of* $J(\cdot, P)$.

(i). Then, $K_n(x) \to J(x \cdot \sigma(P), P)$ *in probability.*

(ii). If $J(\cdot, P)$ *is continuous, then*

$$\sup_x |L_n(x) - J_n(x \cdot \sigma(P), P)| \to 0$$

in probability.

(iii). Let $d_n(1-\alpha) = \inf\{x : K_n(x) \geq 1 - \alpha\}$. *If* $J(\cdot, P)$ *is continuous at* $c(1-\alpha, P) \cdot \sigma(P)$, *then*

$$Prob_P\{\tau_n[T_n - \theta(P)]/\hat{\sigma}_n \leq d_n(1-\alpha)\} \to 1 - \alpha$$

as $n \to \infty$. *Thus, the asymptotic coverage probability of the interval* $[T_n - \hat{\sigma}_n \tau_n^{-1} d_n(1-\alpha), \infty)$ *is the nominal level* $1 - \alpha$.

(iv). Assume, for every $d > 0$, $\sum_n exp\{-d[n/b]\} < \infty$, $\tau_b(T_n - \theta(P)) \to 0$ almost surely, and $\hat{\sigma}_n \to \sigma(P)$ almost surely. Then, the convergences in (i) and (ii) hold with probability one.

2.4.2. General Parameter Space. It is often desirable to construct confidence regions for multivariate parameters, or for parameters taking values in a function space. For example, consider the problem of constructing confidence bands for the density or distribution function, which may form the basis of a goodness of fit test. Assume $\theta(P)$ takes values in a normed linear space Θ, with norm denoted $\|\cdot\|$. Let T_n be an estimate of $\theta(P)$. Assume Assumption A, with the interpretation that $\tau_n[T_n - \theta(P)]$ has a distribution in θ. Here, Θ is endowed with an appropriate σ-field so that $\tau_n[T_n - \theta(P)]$ is measurable and an appropriate weak convergence theory ensues, though we omit such measurability issues here. Let $H_n(P)$ denote the distribution of $\tau_n\|T_n - \theta(P)\|$ under P, with corresponding c.d.f $H_n(x, P)$. Assumption A implies $H_n(P)$ converges weakly to $H(P)$, the distribution of ξ, where ξ has distribution $J(P)$. The corresponding cdf of $H(P)$ is denoted $H(x, P)$. The approximation to $H_n(x)$ we study is defined analogously to (2.1):

$$\hat{H}_n(x) = N_n^{-1} \sum_{i=1}^{N_n} 1\{\tau_b \|S_n^i - T_n\| \le x\}.$$

Theorem 2.3. *Assume Assumption A. Also assume $\tau_b/\tau_n \to 0$, $b \to \infty$ and $b/n \to 0$ as $n \to \infty$. Let x be a continuity point of $H(\cdot, P)$.*

(i). Then, $\hat{H}_n(x) \to H(x, P)$ in probability.

(ii). If $H(\cdot, P)$ is continuous, then

$$\sup_x |\hat{H}_n(x) - H_n(x, P)| \to 0$$

in probability.

(iii). Let $h_n(1-\alpha) = \inf\{x : \hat{H}_n(x) \ge 1 - \alpha\}$. Correspondingly, define $h(1-\alpha, P) = \inf\{x : H(x, P) \ge 1 - \alpha\}$. If $H(\cdot, P)$ is continuous at $h(1-\alpha, P)$, then

$$Prob_P\{\tau_n \|T_n - \theta(P)\| \le h_n(1-\alpha)\} \to 1 - \alpha$$

as $n \to \infty$. Thus, the asymptotic coverage probability of the set $\{\theta \in \Theta : \tau_n \|T_n - \theta\| \le h_n(1-\alpha)\}$ is the nominal level $1 - \alpha$.

(iv). Assume, for every $d > 0$, $\sum_n exp\{-d[n/b]\} < \infty$ and $\tau_b(T_n - \theta(P)) \to 0$ almost surely. Then, the convergences in (i) and (ii) hold with probability one.

The proof of the above theorem is similar to that of Theorem 2.1, and is omitted. Immediate applications of the theorem result in uniform confidence bands for a cumulative distribution function F, based on iid observations from F or in the case where observations are censored. The theory is also applicable to biased sampling models, including stratified sampling, enriched statified sampling, choice-based sampling, and case-control studies; these models are developed in Gill, Vardi, and Wellner (1988), where they show Assumption A is satisfied under weak assumptions. Since distributional theory is quite hard in these models, our method offers simplicity in implementation and mathematical justification.

2.5. Second Order Asymptotics and Choice of b.

Fortunately, the theory developed thus far assumes little about the choice of b. In practice, however, the implementation of the method requires a particular choice. In order to choose b optimally, higher order considerations are necessary.

Consider the case where θ is a univariate mean. As in Section 4 of Wu (1990), consider the Edgeworth expansion of the usual studentized statistic:

$$Prob\{n^{1/2}(\bar{X}_n - \theta)/\hat{\sigma}_n \leq t\} = \Phi(t) + (2t^2 + 1)\phi(t)\gamma n^{-1/2}/6 + o(n^{-1/2}), \tag{2.6}$$

where $\gamma = E(X - \theta)^3/\sigma^3$. Then, $K_n(x)$ given in (2.5) has the expansion (see Babu and Singh (1985) and Wu (1990)):

$$K_n(t(1-f)^{1/2}) = \Phi(t) + [3t^2 - \frac{1-2f}{1-f}(t^2 - 1)](1-f)^{1/2}\phi(t)\hat{\gamma}_n b^{-1/2}/6 + o_P(b^{-1/2}), \tag{2.7}$$

where $\hat{\gamma}_n = \sum_i (X_i - \bar{X}_n)^3/n\hat{\sigma}_n^3$ and $f = b/n$. The difference between (2.6) and (2.7) is minimized if $b = O(n^{2/3})$, in which case the difference between the two expressions is of order $n^{-1/3}$. A good second order theory would demand the difference to be $o(n^{-1/2})$. Basically, the problem lies in the fact that K_n serves as a good approximation to the studentized statistic based on a sample of size b (not n), and has skewness approximately $\gamma/b^{1/2}$ instead of $\gamma/n^{1/2}$.

The above considerations can be generalized somewhat without the use of Edgeworth expansions by the following heuristic argument. Assume $J_n(x,P) = J(x,P) + n^{-\beta}c(P) + o(n^{-\beta})$ for some $\beta > 0$. Here, J_n could represent the distribution of a studentized or unstudentized root. Our approximation $L_n(x)$ serves as a good approximation to $J_b(x,P)$, with the main error due to the fact that T_n in (2.1) is not $\theta(P)$. Specifically, $L_n - J_b$ is of order b/n in probability. To appreciate why, L_n is the distribution, conditional on $X_1, \ldots, X_n$, of $\tau_b[S_{n,I} - \theta(P)] + \tau_b[\theta(P) - T_n]$, where I is uniform on $1, \ldots, N_n$. The distribution, U_n, of the first term $Z_{n,1} = \tau_n[S_{n,I} - \theta(P)]$ is a good approximation to the distribution J_b; indeed, one can show, in regular situations (by a variance calculation) that U_n differs from J_b by $O_P(n^{-1/2})$. The second term, $Z_{n,1} = \tau_n[\theta(P) - T_n]$ is of order τ_b/τ_n in probability. In regular cases, $\tau_n = n^{1/2}$, in which case the second term is order $(b/n)^{1/2}$ in probability. Hence,

$$L_n(t) = Prob\{Z_{n,1} + Z_{n,2} \leq t | X_1, \ldots, X_n\} = U_n(t - Z_{n,2}).$$

Now, assuming $n^{1/2}[T_n - \theta(P)]$ converges weakly to the normal distribution with mean 0 and variance σ^2 (or any distribution with mean 0 and finite variance), we have

$$U_n(t - Z_{n,2}) \approx \int U_n(t - (b/n)^{1/2}z) d\Phi(z\sigma) \approx \int J_b(t - (b/n)^{1/2}z) d\Phi(z/\sigma) \approx J_b(t) + O(b/n)$$

by a Taylor expansion argument, using the fact that Φ has mean 0. Thus, in regular cases, $L_n - J_n$ is order b/n in probability. Now, the difference between J_n and J_b is of order $b^{-\beta}$. In the case $\beta = 1/2$, the difference between J_n and L_n is then of order $b^{-1/2} + b/n$ in probability. The choice $b = n^{2/3}$ minimizes this order.

We would also like to point out that the choice of b depends crucially on the desired goal. Consider the use of our approximation L_n for the purposes of estimating the bias of T_n. Typically, $\tau_n = n^{1/2}$ and $E(T_n) - \theta(P) = a(P)/n + o(1/n)$. If the mean of $n^{1/2}[T_n - \theta(P)]$ is approximated by m_n, the mean of L_n, then our estimate of $E(T_n) - \theta(P)$ becomes $n^{-1/2}m_n$. But, $n^{-1/2}m_n$ has mean

$$n^{-1/2}[E(T_b) - E(T_n)] = n^{-1/2}b^{-1/2}a(P) + o((nb)^{-1/2}).$$

Hence, in order to accurately estimate the bias of T_n, we should at least require $b/n \to 1$. This is consistent with the usual jackknife estimate of bias which uses $b = n - 1$.

In summary, the optimal choice of b is difficult and future work will focus on this problem. Tu (1992) has shown how jackknife values may be used appropriately to obtain second order accuracy. Basically, Tu (1992) makes use of a normalizing transformation, and a similar approach could be applied here. A further possibility, in cases where Edgeworth expansions exist so that second order accuracy is obtainable, is to consider a k-fold convolution of our estimated sampling distribution. If k is chosen so $k \sim n/b$, then the new distribution has the appropriate skewness term. Such considerations are beyond the scope of the present work, whose goal is to establish the broad applicability of a particular methodology. In general, the optimal choice of b and construction of suitably defined pseudo-values will depend on the particular nature of the problem.

3. STATIONARY TIME SERIES AND HOMOGENEOUS RANDOM FIELDS

3.1. Some Motivation: the Simplest Example.

To fix ideas, suppose the sample $X_1, \ldots, X_n$ is known or suspected to exhibit serial dependence, and that it can generally be modeled as a stationary time series. Assume that the parameter of interest θ is the common mean EX_i, and the statistic T_n is the sample mean $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$. If the serial dependence is weak enough such that $\sum |R(k)| < \infty$, where $R(k) = Cov(X_1, X_{1+k})$, then (under regularity conditions, cf. Brockwell and Davis (1991)) $\bar{X}_n$ is asymptotically normal, i.e., $\sqrt{n}(\bar{X}_n - \theta)$ has the limiting normal $N(0, \sigma_\infty^2)$ distribution, where $\sigma_\infty^2 = Var(X_1) + 2\sum_{i=1}^{\infty} R(k)$. Note that $\sigma_\infty^2 = \lim_{n\to\infty} \sigma_n^2$, where $\sigma_n^2 = Var(\sqrt{n}\bar{X}_n) = Var(X_1) + 2\sum_{i=1}^{n}(1 - |i|/n)R(k)$.

Now to construct confidence intervals for θ using the previously mentioned Central Limit Theorem, a consistent estimate of σ_∞^2 is required. A straightforward approach would be to estimate the covariances $R(k)$, $k = 1, \ldots, n$, and plug them in the formula for σ_n^2, since $\sigma_n^2 \to \sigma_\infty^2$. However, this naive procedure is not consistent, because the estimates of $R(k)$ for k close to n are highly inaccurate; indeed, since σ_∞^2 is just a constant multiple of the spectral density of the time series evaluated at point zero, this is a well known difficulty in the literature concerning spectral estimation (cf., for example, Priestley (1981)). Apparently, based on a sample of size n we could only accurately estimate $R(k)$, for $k = 1, \ldots, b$, where $b << n$. It then follows that we can only hope to estimate well σ_b^2, and not σ_n^2, where $\sigma_b^2 = Var(X_1) + 2\sum_{i=1}^{b}(1 - |i|/b)R(k)$. But there is a natural way to estimate σ_b^2 from $X_1, \ldots, X_n$, namely to look at the sample variability of $\frac{1}{\sqrt{b}}\sum_{t=i}^{i+b-1} X_t$, for $i = 1, \ldots, n-b+1$. This is equivalent to considering the 'sample variance' estimator

$$\hat{\sigma}_b^2 = \frac{1}{n-b+1}\sum_{i=1}^{n-b+1}\left(\frac{1}{\sqrt{b}}\sum_{t=i}^{i+b-1} X_t - \sqrt{b}\bar{X}_n\right)^2$$

that was studied in Carlstein (1986) and in Politis and Romano (1993). This proposed estimator is consistent, under some moment and mixing conditions, essentially because both σ_b^2 and σ_n^2 converge asymptotically to σ_∞^2, if both b and n are assumed to tend to infinity.

In the same light, one could look at the more general problem of estimating the distribution of $\sqrt{n}(\bar{X}_n - \theta)$. But this can be done by looking at the sample variability of $\frac{1}{\sqrt{b}}\sum_{t=i}^{i+b-1} X_t$, for

$i = 1, \ldots, n - b + 1$, and defining the corresponding 'empirical' distribution

$$L_n(x) = (n - b + 1)^{-1} \sum_{i=1}^{n-b+1} 1\{\sqrt{b}(b^{-1} \sum_{t=i}^{i+b-1} X_t - \bar{X}_n) \le x\}$$

as an approximation to the sampling distribution of $\sqrt{n}(\bar{X}_n - \theta)$. Here the underlying principle is that both $\sqrt{b}(\bar{X}_b - \theta)$ and $\sqrt{n}(\bar{X}_n - \theta)$ have the same asymptotic distribution, (which just happens to be the normal $N(0, \sigma_\infty^2)$ distribution), where of course $\bar{X}_b = b^{-1} \sum_{t=1}^{b} X_t$.

Although variance estimation is intimately linked with the assumption of asymptotic normality, this more general idea of directly approximating the sampling distribution would work in a variety of different situations, including cases where asymptotic normality does not hold, where the rate of convergence is not $\sqrt{n}$, or where variance estimation is not consistent. Suppose that $T_n = T_n(X_1, \ldots, X_n)$ is the statistic of interest, where $X_1, \ldots, X_n$ is an observed stretch of a stationary time series, and that assumption A of the previous section is satisfied. Let $S_{n,i}$ be the statistic T_b evaluated at the subseries $X_{(i-1)h+1}, \ldots, X_{(i-1)h+b}$. As will be proved in the following sections, the 'empirical' distribution

$$L_n(x) = q^{-1} \sum_{i=1}^{q} 1\{\tau_b(S_{n,i} - T_n) \le x\} \tag{3.1}$$

is a consistent approximation to the limit law $J(x, P)$ under very weak assumptions; here h is some integer that may depend on n, and $q = [(n - b)/h] + 1$, where $[\cdot]$ is the integer part. In general, as will be argued later on, it is suggested to let $h = 1$, and $b \to \infty$ as $n \to \infty$.

3.2. Basic Definitions.

Suppose $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^d\}$ is a random field in d dimensions, with $d \in \mathbf{Z}^+$, i.e., a collection of random variables $X(\mathbf{t})$ taking values in a state space S, defined on a probability space $(\Omega, \mathcal{A}, P)$, and indexed by the variable $\mathbf{t} \in \mathbf{Z}^d$. The random field $\{X(\mathbf{t})\}$ is assumed to be *homogeneous*, meaning that for any set $\mathbf{E} \subset \mathbf{Z}^d$, and for any point $\mathbf{i} \in \mathbf{Z}^d$, the joint distribution of the random variables $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{E}\}$ is identical to the joint distribution of $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{E} + \mathbf{i}\}$. In the very important special case where $d = 1$, the random field $\{X(\mathbf{t})\}$ is just a stationary time series.

For two points $\mathbf{t} = (t_1, \ldots, t_d)$ and $\mathbf{u} = (u_1, \ldots, u_d)$ in $\mathbf{Z}^d$, define the sup-distance in $\mathbf{Z}^d$ by $d(\mathbf{t}, \mathbf{u}) = \sup_j |t_j - u_j|$, and for two sets $\mathbf{E}_1, \mathbf{E}_2$ in $\mathbf{Z}^d$, define $d(\mathbf{E}_1, \mathbf{E}_2) = \inf\{d(\mathbf{t}, \mathbf{u}) : \mathbf{t} \in \mathbf{E}_1, \mathbf{u} \in \mathbf{E}_2\}$.

Our goal again is to construct a confidence region for a real-valued parameter $\theta = \theta(P)$, on the basis of observing $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{E_n}\}$; $\mathbf{E_n}$ is the rectangle consisting of the points $\mathbf{t} = (t_1, t_2, \ldots, t_d) \in \mathbf{Z}^d$ such that $1 \leq t_k \leq n_k$, where $k = 1, 2, \ldots, d$, and $\mathbf{n} = (n_1, n_2, \ldots, n_d)$. The sample size is again denoted by n, although now $n \equiv \prod_{i=1}^d n_i = |\mathbf{E_n}|$, where $|\mathbf{E}|$ denotes the cardinality of the set $\mathbf{E}$.

The random field $\{X(\mathbf{t})\}$ will be assumed to satisfy a certain weak dependence condition. Define a collection of strong mixing coefficients by

$$\alpha_X(k; l_1, l_2) \equiv \sup_{\mathbf{E}_1, \mathbf{E}_2 \subset \mathbf{Z}^d} \{|P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_i \in \mathcal{F}(\mathbf{E}_i), |\mathbf{E}_i| \leq l_i, i = 1, 2, d(\mathbf{E}_1, \mathbf{E}_2) \geq k\}$$

where $\mathcal{F}(\mathbf{E})$ is the σ-algebra generated by $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{E}\}$. A weak dependence condition is formulated if $\alpha_X(k; l_1, l_2)$ is assumed to converge to zero at some rate, as k tends to infinity, and l_1, l_2 either remain fixed or tend to infinity as well. It is interesting to note that $\alpha_X(k; l_1, l_2)$ is decreasing in each of its arguments l_1, l_2 separately. In particular, if we let $\alpha_X(k) = \alpha_X(k; \infty, \infty)$ be the usual strong mixing coefficient of Rosenblatt (1985), it is apparent that $\alpha_X(k; l_1, l_2) \leq \alpha_X(k)$. If $\alpha_X(k) \to 0$ as $k \to \infty$, then the random field $\{X(\mathbf{t})\}$ is simply said to be strong mixing.

In the case of a stationary sequence ($d = 1$), the condition of strong mixing is rather weak and is satisfied by a whole host of interesting examples (cf. Ibragimov and Rozanov (1978)). There are still many examples of strong mixing random fields in the case $d > 1$ (cf. Rosenblatt (1985)), e.g., Gaussian fields with continuous and positive spectral density function. However, an interesting class of random fields (with $d > 1$), the so-called Gibbs states (Markov field models), are not necessarily strong mixing (cf. Dobrushin (1968) for an example), but do satisfy weak dependence conditions involving the $\alpha_X(k; l_1, l_2)$ coefficients (cf. Neaderhouser (1980), Bolthausen (1982), Zhurbenko (1986), Bradley (1991)).

3.3. The General Theorem in the Case of Dependent Data.

As in Section 2, let $T_{\mathbf{n}} = T_{\mathbf{n}}(X(\mathbf{t}), \mathbf{t} \in \mathbf{E}_{\mathbf{n}})$, and let $J_{\mathbf{n}}(P)$ be the sampling distribution of $\tau_{\mathbf{n}}(T_{\mathbf{n}} - \theta(P))$. Again the only assumption that will be needed is assumption A restated here:

Assumption A. $J_{\mathbf{n}}(P)$ converges weakly to a limit law $J(P)$, as $n_i \to \infty$, for $i = 1, \ldots, d$.

Define $Y_{\mathbf{j}}$ to be the block of size b of the *consecutive* data $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{E}_{\mathbf{j},\mathbf{b},\mathbf{h}}\}$, where $\mathbf{j} = (j_1, j_2, \ldots, j_d)$ and $\mathbf{E}_{\mathbf{j},\mathbf{b},\mathbf{h}}$ is the smaller rectangle consisting of the points $\mathbf{i} = (i_1, i_2, \ldots, i_d) \in \mathbf{Z}^d$ such that $(j_k - 1)h_k + 1 \leq i_k \leq (j_k - 1)h_k + b_k$, for $k = 1, 2, \ldots, d$; $\mathbf{b} = (b_1, \ldots, b_d)$, $\mathbf{h} = (h_1, \ldots, h_d)$ are points in $\mathbf{Z}^d$ that depend in general on n and $\mathbf{E}_{\mathbf{n}}$. The point $\mathbf{b}$ indicates the shape and size of rectangle $\mathbf{E}_{\mathbf{i},\mathbf{b},\mathbf{h}}$, and the point $\mathbf{h}$ indicates the amount of 'overlap' between the rectangles $\mathbf{E}_{\mathbf{i},\mathbf{b},\mathbf{h}}$ for neighboring $\mathbf{i}$'s, i.e., the size of their intersection; for example, if $\mathbf{h} = \mathbf{b}$ there is no overlap between $\mathbf{E}_{\mathbf{i},\mathbf{b},\mathbf{h}}$ and $\mathbf{E}_{\mathbf{j},\mathbf{b},\mathbf{h}}$ for $\mathbf{i} \neq \mathbf{j}$, while if $\mathbf{h} = (1, 1, \ldots, 1)$ the overlap is the maximum possible. It will generally be assumed that either $\mathbf{h} = (1, 1, \ldots, 1)$, or that as $b_i \to \infty$, $h_i/b_i \to a_i \in (0, 1]$, for $i = 1, 2, \ldots, d$.

As before, denote $b = \prod_{i=1}^d b_i$, and $h = \prod_{i=1}^d h_i$, and observe that, with $\mathbf{E}_{\mathbf{n}}$ and n fixed, $Y_{\mathbf{j}}$ is defined only for $\mathbf{j}$ such that $1 \leq j_k \leq q_k$, where $q_k = [\frac{n_k - b_k}{h_k}] + 1$, and thus the total number of the $Y_{\mathbf{j}}$ blocks available from the data is $q = \prod_{i=1}^d q_i$. (The number q should be compared to the number N_n in the i.i.d. case of Section 2.)

Similarly to Section 2, let $S_{\mathbf{n},\mathbf{i}}$ be equal to the statistic $T_{\mathbf{b}}$ evaluated at the data set $Y_{\mathbf{i}}$. The approximation to $J_{\mathbf{n}}(x, P)$ we study is now defined by

$$L_{\mathbf{n}}(x) = q^{-1} \sum_{i_1=1}^{q_1} \sum_{i_2=1}^{q_2} \cdots \sum_{i_d=1}^{q_d} 1\{\tau_{\mathbf{b}}(S_{\mathbf{n},\mathbf{i}} - T_{\mathbf{n}}) \leq x\} \tag{3.2}$$

Theorem 3.1. *Assume Assumption A, and that* $\tau_{\mathbf{b}}/\tau_{\mathbf{n}} \to 0$, $b_i \to \infty$, *and* $n_i \to \infty$, *for* $i = 1, 2, \ldots, d$. *Also assume that* $\prod_{j=1}^d b_j/(n_j - b_j) \to 0$, *and that* $q^{-1} \sum_{k=1}^{q^\star} k^{d-1} \alpha_X(k; b, b) \to 0$, *where* $q^\star = \max_i q_i$. *Let* x *be a continuity point of* $J(\cdot, P)$. *Then conclusions (i)–(iii) of Theorem 2.1 remain true (with* n *replaced by* $\mathbf{n}$*).*

Proof. In what follows, $c_0, c_1, c_2, \ldots$ will denote some positive constants. As in the proof of Theorem 2.1, to prove (i) it suffices to show that $U_{\mathbf{n}}(x)$ converges in probability to $J(x,P)$, where

$$U_{\mathbf{n}}(x) = q^{-1} \sum_{i_1=1}^{q_1} \sum_{i_2=1}^{q_2} \cdots \sum_{i_d=1}^{q_d} 1\{\tau_{\mathbf{b}}(S_{\mathbf{n},\mathbf{i}} - \theta(P)) \leq x\}.$$

Since $EU_{\mathbf{n}}(x) = J_{\mathbf{b}}(x,P)$, and $J_{\mathbf{b}}(x,P) \to J(x,P)$ as $b_i \to \infty$, for $i = 1,2,\ldots,d$, (by assumption A), it suffices to look at $Var(U_{\mathbf{n}}(x))$. By the homogeneity of the random field $\{X(\mathbf{t})\}$ it follows that

$$Var(U_{\mathbf{n}}(x)) = q^{-1} \sum_{i_1=-q_1}^{q_1} \sum_{i_2=-q_2}^{q_2} \cdots \sum_{i_d=-q_d}^{q_d} (1 - \frac{|i_1|}{q_1})(1 - \frac{|i_2|}{q_2}) \cdots (1 - \frac{|i_d|}{q_d}) C(\mathbf{i}),$$

where $C(\mathbf{i})$ denotes the covariance between $1\{\tau_{\mathbf{b}}(S_{\mathbf{n},\mathbf{1}} - \theta(P)) \leq x\}$ and $1\{\tau_{\mathbf{b}}(S_{\mathbf{n},\mathbf{1}+\mathbf{i}} - \theta(P)) \leq x\}$; note that $C(\mathbf{i}) = C(-\mathbf{i})$. Let $\mathbf{E}_q = \{\mathbf{i} \in \mathbf{Z}^d : |i_j| \leq q_j, j = 1,2,\ldots,d\}$, and $\mathbf{E}^* = \{\mathbf{i} \in \mathbf{Z}^d : |i_j| \leq [b_j/h_j], j = 1,2,\ldots,d\}$, where $[\cdot]$ is the integer part. Then $Var(U_{\mathbf{n}}(x)) = A^* + A$, where

$$A^* = q^{-1} \sum_{\mathbf{i} \in \mathbf{E}^*} (1 - \frac{|i_1|}{q_1})(1 - \frac{|i_2|}{q_2}) \cdots (1 - \frac{|i_d|}{q_d}) C(\mathbf{i}),$$

and

$$A = q^{-1} \sum_{\mathbf{i} \in \mathbf{E}_q - \mathbf{E}^*} (1 - \frac{|i_1|}{q_1})(1 - \frac{|i_2|}{q_2}) \cdots (1 - \frac{|i_d|}{q_d}) C(\mathbf{i}).$$

Looking at A^* it is seen that it is a sum of $\prod_{j=1}^d (2[b_j/h_j] + 1) \sim 2b/h$ terms of order $O(q^{-1})$; since $q = \prod_{j=1}^d ([\frac{n_j - b_j}{h_j}] + 1) \sim \prod_{j=1}^d (n_j - b_j)/h_j$, it follows that $|A^*| = O(\prod_{j=1}^d b_j/(n_j - b_j))$.

Now by the well known mixing inequality for the covariance between two bounded random variables (cf. Roussas and Ioannides (1987)), $|C(\mathbf{i})| \leq c_0 \alpha_X(i^* h^* - b^*; b, b)$, where $i^* = \max_k |i_k|$, $b^* = \max_i b_i$, and $h^* = \min_i h_i$. Therefore,

$$|A| \leq c_0 q^{-1} \sum_{\mathbf{i} \in \mathbf{E}_q - \mathbf{E}^*} \alpha_X(i^* h^* - b^*; b, b) \leq c_1 q^{-1} d \sum_{k=[b^*/h^*]+1}^{q^*} W(k) \alpha_X(kh^* - b^*; b, b),$$

where $W(k)$ is the cardinality of the set $\{\mathbf{i} \in \mathbf{Z}^d : i_1 = k, 0 < i_j \leq i_1, j = 2,\ldots,d\}$. By a combinatorial argument it now follows that $W(k) \leq k^{d-1}$, and therefore,

$$|A| \leq c_2 \frac{1}{q} \sum_{k=[b^*/h^*]+1}^{q^*} k^{d-1} \alpha_X(kh^* - b^*; b, b).$$

It is obvious that by the imposed conditions both terms above converge to zero, and hence $Var(U_{\mathbf{n}}(x)) \to 0$, which completes the proof of (i). The proof of (ii) and (iii) is now exactly analogous to the proof of Theorem 2.1.

It should be noted that, using stronger mixing assumptions, the fourth moments of $L_{\mathbf{n}}(x)$ could be appropriately bounded and convergence with probability one (conclusion (iv) of Theorem 2.1) would hold here too. Since however our emphasis is on obtaining asymptotically valid confidence regions under minimal assumptions, this approach will not be pursued further.

The conditions of Theorem 3.1 are as weak as possible. In practice, since one gets to choose the design parameters $\mathbf{b}$ and $\mathbf{h}$ as functions of the given sample size, a realistic set of conditions would satisfy $b_i \to \infty$, with $b_i/n_i \to 0$, as $n_i \to \infty$, and either $\mathbf{h} = (1, 1, \ldots, 1)$, or that $h_i/b_i \to a_i \in [0, 1]$, for $i = 1, 2, \ldots, d$. In the most important case of maximum overlap between the rectangles, i.e., if $\mathbf{h} = (1, 1, \ldots, 1)$, the statement of the theorem simplifies and the following corollary is true.

Corollary 3.1. *Assume Assumption A, and that* $\tau_{\mathbf{b}}/\tau_{\mathbf{n}} \to 0$, $b_i \to \infty$, *and* $b_i/n_i \to 0$, *as* $n_i \to \infty$, *for* $i = 1, 2, \ldots, d$. *Also set* $\mathbf{h} = (1, 1, \ldots, 1)$, *and assume that* $n^{-1} \sum_{k=1}^{n^*} k^{d-1} \alpha_X(k; b, b) \to 0$, *where* $n^* = \max_i n_i$. *Let* x *be a continuity point of* $J(\cdot, P)$. *Then conclusions (i)–(iii) of Theorem 2.1 remain true (with* n *replaced by* $\mathbf{n}$*).*

Remark 3.1. It is easy to see that if the random field is actually strong mixing, then a sufficient weak dependence condition for Corollary 3.1 to hold is that $k^{d-1}\alpha_X(k) \to 0$ as $k \to \infty$. For the case $d > 1$, a sufficient condition is that $k^{d-1}\alpha_X(k)$ converges to some finite number as $k \to \infty$, and for the important special case of a time series ($d = 1$), this sufficient condition boils down to the minimal assumption that the time series is strong mixing, i.e., that $\alpha_X(k) \to 0$ as $k \to \infty$. As a matter of fact, Theorem 3.1 limited to the time series case is remarkably similar to Theorem 2.1; we include it here as a corollary.

Corollary 3.2. *Let* $d = 1$. *Assume Assumption A, and that* $\tau_b/\tau_n \to 0$, $b \to \infty$, *and* $b/n \to 0$, *as*

$n \to \infty$. Also let $1 \le h \le c_0 b$, for some $c_0 > 0$, and assume the time series is strong mixing. Let x be a continuity point of $J(\cdot, P)$. Then conclusions (i)–(iii) of Theorem 2.1 remain true (with $L_n(\cdot)$ defined as in equation (3.1)).

Theorem 3.1 can also be extended to studentized roots and general parameter spaces, in the same manner Theorem 2.1 was generalized to Theorems 2.2 and 2.3; the details are obvious and are omitted. However, there is an interesting extension of Theorem 3.1 or Corollary 3.1 that should be mentioned here.

Suppose that instead of having a limit theorem where $n_i \to \infty$, for $i = 1, \ldots, d$, we have a modified version of Assumption A that reads:

Assumption A*. $J_{\mathbf{n}}(P)$ converges weakly to a limit law $J(P)$, as $n_i \to \infty$, for $i = 1, \ldots, d^\star$, and $n_j \to Q_j$, for $j = d^\star + 1, \ldots, d$, where $1 \le d^\star \le d$, and the Q_j's are some fixed positive integers.

This notation allows for the case of a limit theorem where not all dimensions n_i of the sample diverge to infinity; for an example of such a limit theorem in the sample mean case, see Bradley (1992). To appreciate where such a limit theorem might be useful in practice, consider the case $d = 2$, and suppose the data are observed on a very long and thin strip on the plane; that is, suppose that n_2 is small for all practical purposes, whereas n_1 is large.

Since the index set cannot be thought to extend arbitrarily in all dimensions, it seems that $d^\star$ is the 'effective' dimension, and the set-up seems equivalent to a vector–valued random field in $d^\star$ dimensions. This point of view however obscures the fact that the probability structure is shift-invariant in d dimensions, a fact that should be used in the analysis. The following corollary addresses this set-up; its proof is analogous to the proof of Theorem 3.1.

Corollary 3.3. *Assume Assumption A*, and that $\tau_{\mathbf{b}}/\tau_{\mathbf{n}} \to 0$, $b_i \to \infty$, and $b_i/n_i \to 0$, as $n_i \to \infty$, for $i = 1, 2, \ldots, d^\star$, whereas $b_j \to Q_j$, and $n_j \to Q_j$, for $j = d^\star + 1, \ldots, d$. Also set $\mathbf{h} = (1, 1, \ldots, 1)$, and assume that $n^{-1} \sum_{k=1}^{n^\star} k^{d^\star - 1} \alpha_X(k; b, b) \to 0$, where $n^\star = \max_{i=1,\ldots,d^\star} n_i$. Let x be a continuity point of $J(\cdot, P)$. Then conclusions (i)–(iii) of Theorem 2.1 remain true (with n replaced by $\mathbf{n}$).*

3.4. Variance Estimation and Bias Reduction.

3.4.1. Variance estimation and choice of the design parameters. In this section, denote by $m_{\mathbf{n}}^{(j)}, \mu_{\mathbf{n}}^{(j)}$, and $\mu^{(j)}$ the jth (noncentral) moments of distributions $L_{\mathbf{n}}(\cdot), J_{\mathbf{n}}(\cdot, P)$, and $J(\cdot, P)$ respectively, assuming $\mu_{\mathbf{n}}^{(j)}$ and $\mu^{(j)}$ exist. It follows that if in the assumptions of Theorem 3.1 we include that $m_{\mathbf{n}}^{(2)}$ converges to $\mu^{(2)}$, then the subsampling methodology can also be used for estimating the variance of the statistic $T_{\mathbf{n}}$. As a matter of fact, in the case where $T_{\mathbf{n}}$ is the sample mean or a closely related statistic, convergence of $m_{\mathbf{n}}^{(2)}$ to $\mu^{(2)}$ can actually be proven under stronger moment and mixing conditions and does not have to be included in the assumptions; see the next section for a brief review of the literature on this problem.

Note that this variance estimation might be important only in situations where $J(P)$ is a mean zero normal distribution, whose variance is the sole unknown. As demonstrated in Theorem 3.1, it is quite unnecessary to assume stronger moment and mixing conditions if the goal is just to construct asymptotically valid confidence intervals.

Nevertheless, looking at the problem of variance estimation can yield useful insights. For example, a most interesting question for practical applications is how to choose $\mathbf{b}$ and $\mathbf{h}$ as functions of $\mathbf{n}$. In the case of sample mean type statistics, it turns out that to have a most accurate (from the point of view of asymptotic mean squared error) variance estimator, one should let $\mathbf{h} = (1, 1, \ldots, 1)$, and $b \sim An^{\frac{d}{d+2}}$, (cf. Politis and Romano (1992b, 1993)); the constant $A > 0$ can in principle be calculated (or estimated) given the specifics of the problem, (see Künsch (1989) for an explicit calculation in the sample mean example for the case $d = 1$).

The variance of the variance estimator $m_{\mathbf{n}}^{(2)}$ can be shown to be of order $O(b/n)$, in the sample mean and related examples (cf. Politis and Romano (1992b)), *regardless* of choice of $\mathbf{h}$. However, taking $\mathbf{h} = (1, 1, \ldots, 1)$ is preferred because it decreases the variance of $m_{\mathbf{n}}^{(2)}$ by a constant factor. Intuitively this makes sense, since the case $\mathbf{h} = (1, 1, \ldots, 1)$ corresponds to a maximum overlap between the rectangles $\mathbf{E}_{\mathbf{i},\mathbf{b},\mathbf{h}}$, for $\mathbf{i}$ such that $1 \le i_k \le q_k$, $k = 1, \ldots, d$, which in turn (for given $\mathbf{b}$ and $\mathbf{n}$) maximizes q, the number of subsamples available from the data, making it equal to $\prod_{i=1}^{d}(n_i - b_i + 1)$. On the other hand, taking $h_i/b_i \to a_i \in [0, 1]$ would imply that a proportion of the $\prod_{i=1}^{d}(n_i - b_i + 1)$ available $S_{\mathbf{n},\mathbf{i}}$'s are thrown away when computing the 'empirical' estimate $L_{\mathbf{n}}$ and its variance.

Another insight offered by the problem of variance estimation is apparent by comparing the i.i.d. case of Section 2 and the dependent case of Section 3. The difference is that, whereas in the i.i.d. case (under some extra conditions) b can be taken of the same order as n, this *cannot* be done in the dependent case, even in the simplest setting of the mean. This is manifested by the fact that, as mentioned, the variance of the variance estimator $m_{\mathbf{n}}^{(2)}$ is of order $O(b/n)$, in contrast to the i.i.d. case where the variance of $m_{\mathbf{n}}^{(2)}$ is of order $O(1/n)$, independent of b.

To fix ideas, consider again the example of the sample mean of a stationary sequence of Section 3.1. Then the variance estimator $m_n^{(2)}$ is asymptotically equivalent to a kernel smoothed (with Bartlett's kernel) estimator of the spectral density at the origin (Künsch (1989)). It is now well-known (cf. Priestley (1981)) that the bias of $m_n^{(2)}$ is of order $O(1/b)$, and the variance of $m_n^{(2)}$ is of order $O(b/n)$; this of course implies that consistent variance estimation requires $b \to \infty$ as well as $b/n \to 0$.

3.4.2. Bias reduction. Since statistics calculated from time series and random fields are often heavily biased, the subsampling methodology could be used for bias reduction, in the same vein as the original proposition of a 'jackknife' by Quenouille (1949). To outline the method, assume that assumption A holds together with $\mu_{\mathbf{n}}^{(1)} \to \mu^{(1)}$ and $m_{\mathbf{n}}^{(1)} \to \mu^{(1)}$; usually, but not always, it will be the case that $\mu^{(1)} = 0$. Then, since $L_{\mathbf{n}}(\cdot)$ and $J_{\mathbf{n}}(\cdot, P)$ have the same limiting distribution $J(\cdot, P)$, (with first moments converging as well), one can approximate $Bias(T_{\mathbf{n}}) = ET_{\mathbf{n}} - \theta$ by a re-scaled version of the 'empirical' bias, i.e., by

$$\hat{Bias}(T_{\mathbf{n}}) = \frac{1}{\tau_{\mathbf{n}}} m_{\mathbf{n}}^{(1)} = \frac{\tau_{\mathbf{b}}}{\tau_{\mathbf{n}}}(Ave(S_{\mathbf{n},\mathbf{i}}) - T_{\mathbf{n}})$$

where $Ave(S_{\mathbf{n},\mathbf{i}}) = q^{-1} \sum_{i_1=1}^{q_1} \sum_{i_2=1}^{q_2} \cdots \sum_{i_d=1}^{q_d} S_{\mathbf{n},\mathbf{i}}$; correspondingly one can form the bias corrected estimator

$$\hat{T}_{\mathbf{n}} = T_{\mathbf{n}} - \hat{Bias}(T_{\mathbf{n}}) = (1 + \frac{\tau_{\mathbf{b}}}{\tau_{\mathbf{n}}})T_{\mathbf{n}} - \frac{\tau_{\mathbf{b}}}{\tau_{\mathbf{n}}} Ave(S_{\mathbf{n},\mathbf{i}}). \tag{3.3}$$

It is obvious that this is an asymptotic bias correction. For example, in the simplest case where $T_{\mathbf{n}}$ is the sample mean (which is unbiased), $\hat{Bias}(T_{\mathbf{n}}) \not\equiv 0$, due to edge effects; nevertheless $\hat{Bias}(T_{\mathbf{n}}) \to 0$ as it should (cf. Politis and Romano (1992a)). In the following theorem the conditions of Theorem 3.1 are strengthened to ensure that the bias correction suggested in equation

(3.3) is indeed asymptotically valid. The argument is actually most relevant when $\mu^{(1)} \neq 0$, such as in the case of an optimally smoothed spectral density estimator (see example 3.6.2). In regular, $\sqrt{n}-$ consistent cases, the bias correction (3.3) can be seen to be efficacious as well, by similar arguments as in the i.i.d. case (see Section 2.5).

Theorem 3.2. *Assume Assumption A strengthened to include* $\mu_{\mathbf{n}}^{(1)} \to \mu^{(1)}$; *assume* $\tau_{\mathbf{b}}/\tau_{\mathbf{n}} \to 0$, $b_i \to \infty$, *and* $n_i \to \infty$, *for* $i = 1, 2, \ldots, d$. *Also assume that* $\prod_{j=1}^{d} b_j/(n_j - b_j) \to 0$, *that* $E|\tilde{S}_{\mathbf{n},1}|^{2+\delta} < C$, *and that* $q^{-1}\sum_{k=1}^{q^\star} k^{d-1}\{\alpha_X(k; b, b)\}^{\delta/(2+\delta)} \to 0$, *where* δ *and* C *are two positive constants independent of* $\mathbf{n}$, $\tilde{S}_{\mathbf{n},1} \equiv S_{\mathbf{n},1}/\sqrt{Var(S_{\mathbf{n},1})}$, *and* $q^\star = \max_i q_i$. *Then* $|m_{\mathbf{n}}^{(1)} - \mu_{\mathbf{n}}^{(1)}| \to 0$ *in probability.*

Proof. First note that $Em_{\mathbf{n}}^{(1)} = \tau_{\mathbf{b}}(ES_{\mathbf{n},1} - ET_{\mathbf{n}}) = \mu_{\mathbf{b}}^{(1)} - \frac{\tau_{\mathbf{b}}}{\tau_{\mathbf{n}}}\mu_{\mathbf{n}}^{(1)} = \mu_{\mathbf{b}}^{(1)} + o(1)$, and that $|\mu_{\mathbf{b}}^{(1)} - \mu_{\mathbf{n}}^{(1)}| \to 0$, by the (strengthened) assumption A. Now

$$Var(m_{\mathbf{n}}^{(1)}) = Var(\tau_{\mathbf{b}}(Ave(S_{\mathbf{n},\mathbf{i}}) - T_{\mathbf{n}}))$$

$$= Var(\tau_{\mathbf{b}}(Ave(S_{\mathbf{n},\mathbf{i}}) - ES_{\mathbf{n},1}) - \tau_{\mathbf{b}}(T_{\mathbf{n}} - ES_{\mathbf{n},1})) = Var(\tau_{\mathbf{b}}(Ave(S_{\mathbf{n},\mathbf{i}}) - ES_{\mathbf{n},1})) + o(1),$$

because $Var(\tau_{\mathbf{b}}(T_{\mathbf{n}} - ES_{\mathbf{n},1})) \to 0$ as $\tau_{\mathbf{b}}/\tau_{\mathbf{n}} \to 0$. But

$$Var(\tau_{\mathbf{b}}(Ave(S_{\mathbf{n},\mathbf{i}}) - ES_{\mathbf{n},1})) = \tau_{\mathbf{b}}^2 q^{-1}\sum_{\mathbf{i}\in\mathbf{E}_q}(1 - \frac{|i_1|}{q_1})(1 - \frac{|i_2|}{q_2})\cdots(1 - \frac{|i_d|}{q_d})Cov(S_{\mathbf{n},1}, S_{\mathbf{n},1+\mathbf{i}})$$

and thus

$$|Var(\tau_{\mathbf{b}}(Ave(S_{\mathbf{n},\mathbf{i}}) - ES_{\mathbf{n},1}))| \le q^{-1}\sum_{\mathbf{i}\in\mathbf{E}_q} Cov(\tilde{S}_{\mathbf{n},1}, \tilde{S}_{\mathbf{n},1+\mathbf{i}}),$$

where it was taken into account that $Var(S_{\mathbf{n},1}) = O(1/\tau_{\mathbf{b}}^2)$, and $\mathbf{E}_q$ was defined in the proof of Theorem 3.1. Finally, by a similar argument to the proof of Theorem 3.1, and using the mixing inequality

$$|Cov(\tilde{S}_{\mathbf{n},1}, \tilde{S}_{\mathbf{n},1+\mathbf{i}})| \le 10C^2\{\alpha_X(i^\star h^\star - b^\star; b, b)\}^{\delta/(2+\delta)}$$

(cf. Roussas and Ioannides (1987)), it follows that $Var(m_{\mathbf{n}}^{(1)}) \to 0$ and the theorem is proved.

A bias correction identical to the one suggested in equation (3.3) can be employed in the i.i.d. set-up of section 2 as well; in that case of course, $Ave(S_{\mathbf{n},\mathbf{i}})$ should be redefined to be the average of all available $S_{n,i}$.

3.5. Comparison with other resampling methods.

As mentioned in section 3.4.1, the subsampling methodology for dependent data has been used in the past for variance estimation (Carlstein (1986), Raïs (1992), Politis and Romano (1992a,1992b, 1993)), and is closely related to other nonparametric resampling methods, such as the 'moving blocks' jackknife and bootstrap (Künsch (1989), Liu and Singh (1992), Raïs and Moore (1990), Politis and Romano (1992a)).

In the case of a stationary strong mixing sequence ($d = 1$), Carlstein (1986) used $m_n^{(2)}$, i.e., the variance of the 'empirical' $L_n(x)$, to estimate $\mu_n^{(2)}$, i.e., the variance of $\tau_n T_n$. Assuming strong enough conditions ensuring that $\tau_n = \sqrt{n}$ and T_n is asymptotically normally distributed, and specializing to the case $h = b$, (no overlap between the blocks of data used to compute $S_{n,i}$ and $S_{n,i+1}$), Carlstein showed the consistency of $m_n^{(2)}$ as an estimator of $\mu_n^{(2)}$.

Carlstein's idea was generalized in Politis and Romano (1993) to a certain class of statistics of 'linear' type that are not necessarily $\sqrt{n}$–consistent. In addition, the important case where either $h = 1$, or $h/b \to a \in (0, 1]$, was studied, and the variance estimator $m_n^{(2)}$ with $h = 1$ was shown to be more accurate than the one with $h/b \to a \in (0, 1]$. The subsampling variance estimator $m_n^{(2)}$ was also generalized to the case of homogeneous random fields ($d > 1$) by Raïs (1992) and Politis and Romano (1992b).

The fact that taking $h = 1$ is preferrable to taking $h = b$ was also discussed in Künsch (1989). As it turns out, the so-called 'moving blocks' jackknife estimate of the variance of $\tau_n T_n$ (cf. Künsch (1989), Liu and Singh (1992)) is *identical* to $m_n^{(2)}$ with $h = 1$. Let $\hat{J}_n(x, P)$ denote the 'moving blocks' bootstrap estimate of $J_n(x, P)$ (cf. Künsch (1989), Liu and Singh (1992)); as can be calculated, the variance of $\hat{J}_n(x, P)$ is approximately (up to an asymptotically negligible factor) equal to $m_n^{(2)}$ with $h = 1$, and indeed $\hat{J}_n(x, P)$ is very closely related to the 'empirical' $L_n(x)$. For an extension of the 'moving blocks' jackknife and bootstrap to the case of homogeneous random fields see Raïs and Moore (1990) and Politis and Romano (1992b).

It turns out that, for the case of the sample mean considered in Section 3.1, $\hat{J}_n(x, P)$ is a k-fold convolution of $L_n(x)$ with itself, where $k \sim [n/b]$. Since it is a necessary assumption that $b/n \to 0$ (see Section 3.4.1), it follows that $\hat{J}_n(x, P)$ will always be asymptotically normal if $L_n(x)$ is well behaved; this is proved in Künsch (1989) and Liu and Singh (1992)). Under conditions ensuring the consistency of the variance estimator $m_n^{(2)}$, the 'moving blocks' bootstrap estimate $\hat{J}_n(x, P)$ is consistent as an estimator of $J(x, P)$, *assuming* the limit distribution $J(x, P)$ is itself Gaussian.

The above discussion helps put the subsampling methodology into perspective. To summarize, in the case where the limit distribution $J(x, P)$ is normal, for example in the case of the sample mean or related statistics (differentiable statistics or statistics of the 'linear' type), variance estimation by subsampling or 'moving blocks' jackknife, and distribution estimation by subsampling or 'moving blocks' bootstrap are both applicable. The point to be made in this paper is that distribution estimation by subsampling is actually applicable in quite more general situations, for instance when asymptotic normality does not hold, or where variance estimation is not consistent. Indeed, distribution estimation by subsampling is consistent under the minimal assumptions that there is a limiting distribution $J(x, P)$, and that the data are weakly dependent (so that consistent estimation is even possible).

3.6. Some Examples.

The examples will address some rather unorthodox cases; in all standard cases of statistics from time series and random fields that possess asymptotic distributions, e.g., the sample mean, the sample autocovariances and autocorrelations, estimates of the spectral and cross–spectral density, estimates of the coherency function, etc., the subsampling methodology outlined in Section 3 is obviously applicable.

For the examples consider the case of a real valued stationary sequence ($d = 1$), in which case the notation is much simpler, although all examples have immediate analogs in the random field case. So suppose the sample $\{X_t, t = 1, \ldots, n\}$ is observed from the stationary strong mixing sequence $\{X_t, t \in \mathbf{Z}\}$.

3.6.1. Robust statistics from time series. Suppose the first marginal of the sequence $\{X_t\}$, i.e., the distribution of the random variable X_1, is symmetric and unimodal, with unknown location θ. Much of the methodology of robustness can be applied to the case of dependent data as well (cf. Gastwirth and Rubin (1975), Künsch (1984), Martin and Yohai (1986)). Under regularity conditions, the median, the trimmed mean, the Hodges-Lehmann estimator, linear combinations of order statistics, etc., all possess asymptotic distributions, and hence Theorem 3.1 is directly applicable.

As an example, consider a Gaussian strong mixing sequence $\{X_t\}$, satisfying $\sum |R(k)| < \infty$, where $R(k) = Cov(X_1, X_{1+k})$. Then (cf. Gastwirth and Rubin (1975)) the Hodges-Lehmann estimator, i.e., the median of all pairwise averages of the data, is asymptotically normal, with mean θ and variance proportional to $2n^{-1} \sum_{k=-\infty}^{\infty} \arcsin(R(k)/2)$. It is apparent that to use this asymptotic normal distribution to set confidence intervals for θ, the constant $\sum \arcsin(R(k)/2)$ should be consistently estimated which is a difficult task. To appreciate the difficulty recall than even estimating $\sum_{k=-\infty}^{\infty} R(k)$ is hard and amounts to estimation of the spectral density function at the origin. Using Theorem 3.1 to set approximate confidence intervals for θ bypasses this difficult problem.

3.6.2. The spectral density function. As before assume that $\sum |R(k)| < \infty$, and define the spectral density function f by $f(w) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} R(k) e^{-ikw}$. Fix a point $w \in [-\pi, \pi]$, and consider a kernel smoothed estimator of $f(w)$ given by $\hat{f}(w) = \frac{1}{2\pi} \sum_{k=-n}^{n} B_n(k) \hat{R}(k) e^{-ikw}$, where $\hat{R}(k) = \frac{1}{n} \sum_{t=1}^{n-k} X_t X_{t+k}$ is the usual sample autocovariance, and $B_n(k)$ is the 'lag–window'. Under regularity conditions (cf. Priestley (1981) and the references therein), there is a sequence τ_n, corresponding to a particular choice of a sequence of lag–windows $B_n(\cdot)$, such that $\tau_n(\hat{f}(w) - f(w))$ has an asymptotic normal distribution.

To fix ideas, suppose $B_n(\cdot)$ is the Parzen window (cf. Priestley (1981), p. 443); here m_n is a sequence of design parameters that should be chosen appropriately depending on n. Then, under moment and weak dependence conditions (the latter having a correspondence to conditions on the smoothness of the spectral density, cf. Ibragimov and Rozanov (1978)), it can be calculated (see Parzen (1961) or Priestley (1981), p. 462) that asymptotically

$$Bias(\hat{f}(w)) = E\hat{f}(w) - f(w) \sim \frac{b_w}{m_n^2},$$

$$Var(\hat{f}(w)) \sim \frac{\sigma_w^2 m_n}{n},$$

and that $\tau_n(\hat{f}(w) - E\hat{f}(w))$ is asymptotically normal $N(0, \sigma_w^2)$, where $\tau_n = \sqrt{n/m_n}$, and b_w and σ_w^2 are constants depending on w and on f.

By Slutsky's theorem it follows that for a choice of m_n satisfying $n^{1/5}/m_n \to 0$, as $n \to \infty$, $\sqrt{n/m_n}(\hat{f}(w) - f(w))$ is also asymptotically normal $N(0, \sigma_w^2)$. Thus Assumption A is satisfied, and the subsampling methodology can be used to set confidence intervals for $f(w)$. The same ideas are directly applicable in the case of homogeneous random fields ($d > 1$); kernel smoothed estimators of $f(w)$ for $w \in [-\pi, \pi]^d$ are formed in analogous manner, and are shown to be asymptotically normally distributed under regularity conditions (cf. Rosenblatt (1985)).

Note however that to have a most accurate (from the point of view of asymptotic mean squared error) estimator $\hat{f}(w)$, we should choose $m_n \sim (4b_w^2/\sigma_w^2)^{1/5} n^{1/5}$. In this case, the $Bias(\hat{f}(w))$ is significant and is of the same order as $\sqrt{Var(\hat{f}(w))}$; the asymptotic distribution of $\sqrt{n/m_n}(\hat{f}(w) - f(w))$ is now normal $N(\pm\frac{1}{2}\sigma_w, \sigma_w^2)$, where the $\pm$ sign corresponds to the sign of b_w. Since b_w and σ_w^2 are generally unknown, they could either be estimated, or the choice $m_n \sim An^{1/5}$ can be made, for some constant $A > 0$; this choice would imply that the asymptotic distribution of $\sqrt{n/m_n}(\hat{f}(w) - f(w))$ is normal $N(const., \sigma_w^2)$. In other words, if m_n is of the optimal order of magnitude $n^{1/5}$, there exists an asymptotic distribution for $\tau_n(\hat{f}(w) - f(w))$, but it generally has nonzero mean. Therefore, a bias correction in the spirit of Theorem 3.2 is useful here. To ensure that we have a nonnegative estimate of $f(w)$, the positive part of the bias corrected estimator may also be taken which is also justified by asymptotic considerations.

3.6.3. Nonparametric estimation of the first marginal distribution. Let $F(\cdot)$ denote the distribution of the random variable X_1, and let $\hat{F}(x) = n^{-1}\sum_{i=1}^{n} 1\{X_i \leq x\}$. Under regularity conditions (cf. Györfi *et al.* (1989)), $\sqrt{n}(\hat{F}(x) - F(x))$ possesses a limiting normal distribution, and hence Assumption A is satisfied. Furthermore, $\sqrt{n}(\hat{F}(\cdot) - F(\cdot))$, viewed as a random function, converges weakly to a Gaussian process (cf. Deo (1973)). Looking at the sup–norm $\sup_x |\sqrt{n}(\hat{F}(x) - F(x))|$, uniform confidence bands for the unknown distribution $F(\cdot)$ can be set by the subsampling methodology, similarly to the i.i.d. case of Section 2.4.

If it is known that $F(\cdot)$ is absolutely continuous with probability density $F'(\cdot)$, then $F'(\cdot)$ may be estimated by the derivative of a smoothed version of $\hat{F}(\cdot)$. The subsampling methodology will be useful here too, although the rate τ_n is no longer $\sqrt{n}$, and the problem of bias and bias correction becomes important, in exact analogy to the example of the spectral density function.

3.6.4. The sample mean of a time series with long range dependence. Let $\bar{X}_n = n^{-1}\sum_{t=1}^{n} X_t$ be the sample mean, and $\theta = EX_0$ be the mean; this example was considered in Section 3.1 under the assumption that $\sum |R(k)| < \infty$. We now abandon this assumption, and instead suppose that, although the sequence $\{X_t\}$ is strong mixing, the mixing coefficients decrease to zero slowly enough so that the variance of $\bar{X}_n$ is not of order n^{-1}. Suppose that actually $\sum_{k=1}^{n} R(k) \sim n^{2\beta}$, and therefore $Var(\bar{X}_n) \sim \sigma_\infty^2 n^{2\beta-1}$, for some $0 < \beta < 1/2$, and $\sigma_\infty^2 > 0$. Assuming $E|X_0|^{2+\delta} < \infty$, for some $\delta > 0$, it follows from Rosenblatt (1984) that $n^{\frac{1}{2}-\beta}(\bar{X}_n - \theta)$ has an asymptotic normal $N(0, \sigma_\infty^2)$ distribution, and thus Assumption A is satisfied with $\tau_n = n^{\frac{1}{2}-\beta}$. The number β could be estimated (cf. Beran (1986), Künsch (1989)) if it is not known, as will typically be the case.

4. CONCLUSION

In this paper, we have demonstrated how the sampling distributions of normalized statistics can be estimated through the use of jackknife pseudo-values or, equivalently, the values of the statistic computed over certain subsets of the data. The applicability of such methods has been discussed in complicated i.i.d. situations and in the setting of homogeneous random fields. The viability of such methods in the context of time series and random fields is particularly important because the distribution theory of many estimators is quite complicated. Our results are powerful enough that the intricate problem of constructing a confidence interval for the spectral density function, for example, is immediate from our general results. Indeed, in all of our results, the asymptotic justification of the method studied hinges on the simple assumption of a limit distribution for the normalized statistic. Hence, the method is applicable in quite complex settings.

Future work will focus on the higher order asymptotic properties of these methods, which was somewhat discussed in Section 2.5. In particular, the choice of b remains a practical and theoretical issue, in spite of our results which support the view that the method is justified over a wide range of subsample size. As previously mentioned in Section 2.5, there are undoubtedly several possible routes to construct second order correct procedures in regular situations. Tu (1992) has presented such a scheme. Outside of the i.i.d. context, very little is known about higher order accuracy in the nonparametric analysis of time series. Our method immediately applies to most of the interesting statistics in time series, unlike bootstrap methods such as the moving blocks of Künsch (1989) and Liu and Singh (1988) or the stationary bootstrap of Politis and Romano (1991). Indeed, as in the i.i.d. case, bootstrap methods require the weak convergence of the statistic to be smooth as a function of the model, and the verification of such smoothness can be challenging even in specific situations. In contrast, the first order validity of our method is quite apparent in general with little further work. Now that there exist methods that possess minimal consistency requirements without having to invoke unrealistic model assumptions, further work should compare and refine these methods so that inferences can be valid to a high degree of accuracy in a broad range of situations.

References

Arcones, M. (1990). On the asymptotic theory of the bootstrap. Ph.D dissertation, The City University of New York.

Arcones, J. and Giné, E. (1989). The bootstrap of the mean with arbitrary bootstrap sample size. *Ann. Inst. Henri Poincaré* **25**, 457-481.

Arunkaumar, S. (1972). Nonparametric age replacement policy. *Sankhya*, Series A, **34**, 251-256.

Athreya, K. (1987). Bootstrap of the mean in the infinite variance case. *Ann. Statist.* **15**, 724-731.

Babu, J. (1984). Bootstrapping statistics with linear combinations of chi-squares as weak limit. *Sankhya* **46**, 86-93.

Babu, G. and Singh, K. (1985). Edgeworth expansions for sampling without replacement for finite populations. *J. Multivariate Anal.* **17**, 261-278.

Beran, J. (1986). Estimation, testing, and prediction for self-similar processes, Ph.D. Thesis, ETH Zürich.

Beran, R. (1984). Bootstrap methods in statistics. *Jber. d. Dt. Math.-Verein* **86**, 14-30.

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

Bolthausen, E. (1982). On the central limit theorem for stationary random fields, *Ann. Prob.*, 10, 1047-1050.

Bradley, R.C. (1991). Equivalent mixing conditions for random fields, Technical Report No. 336, Center for Stochastic Processes, Department of Statistics, University of North Carolina, Chapel Hill.

Bradley, R.C. (1992). On the spectral density and and asymptotic normality of weakly dependent random fields, *J. Theor. Prob.*, vol. 5, no. 2, 355-373.

Bretagnolle, J. (1983). Limites du bootstrap de certaines fonctionnelles. *Ann. Inst. H. Poincare.* **3**, 281-296.

Brockwell, P. and Davis, R. (1991). *Time series: theory and methods, 2nd ed.*, Springer, New York.

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.

Deo, C. (1973). A note on empirical processes of strong-mixing sequences, *Ann. Prob.*, 1, 870-875.

Dobrushin, R.L. (1968). The description of a random field by means of conditional probabilities and conditions of its regularity, *Theor. Prob. Appl.*, 13, 197-224.

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

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans.*, CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia.

Gastwirth, J. L. and Rubin, H. (1975). The behavior of robust estimators on dependent data, *Ann. Statist.*, vol. 3,, no. 5, 1070-1100.

Gill, R., Vardi, Y. and Wellner, J. (1988). Large sample theory of empirical distributions in biased sampling models. *Ann. Statist.* **16**, 1069-1112.

Györfi, L., Härdle, W., Sarda, P., and Vieu, P. (1989). *Nonparametric Curve Estimation from Time Series*, Lecture Notes in Statistics No.60, Springer-Verlag.

Hartigan, J. (1969). Using subsample values as typical values. *J. Amer. Statist. Ass.* **64**, 1303-1317.

Hartigan, J. (1975). Necessary and sufficient conditions for asymptotic joint normality of a statistic and its subsample values. *Ann. Statist.* **3**, 573-580.

Ibragimov, I.A. and Rozanov, Y.A. (1978). *Gaussian Random Processes*, Springer-Verlag.

Kinateder, J. (1992). An invariance principle applicable to the bootstrap. In *Exploring the Limits of Bootstrap*, Ed. by LePage, R. and Billard, L., John Wiley, New York, 157-181.

Knight, K. (1989). On the bootstrap of the sample mean in the infinite variance case. *Ann. Statist.* **17**, 1168-1175.

Künsch, H. (1984). Infinitesimal robustness for autoregressive processes, *Ann. Statist.*, 12, 843-863.

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

Léger, C. and Cléroux, R. (1990). Nonparametric age replacement: bootstrap confidence interval for the optimal cost. Publication 731, Département d'informatique et de recherche opérationnelle, Université de Montréal.

Liu, R. Y. and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the limits of bootstrap*, ed. by LePage and Billard, John Wiley.

Martin, R.D. and Yohai, V.J. (1986). Influence functionals for time series, *Ann. Statist.*, vol. 14, 781-818.

McCarthy, P. (1969). Pseudo-replication: half-samples. *Rev. ISI*, **37**, 239-263.

Neaderhouser, C.C. (1980) Convergence of block spins defined on random fields, *J. Statist. Phys.*, 22, 673-684.

Parzen, E. (1961). Mathematical Considerations in the Estimation of Spectra, *Technometrics*, Vol. 3, 167-190.

Politis, D. and Romano, J. (1991). The stationary bootstrap. Technical Report 365, Department of Statistics, Stanford University.

Politis, D. and Romano, J. (1992a). A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. *Ann. Statist.*, to appear.

Politis, D.N. and Romano, J.P. (1992b). Nonparametric resampling for homogeneous strong mixing random fields, Technical Report No. 396, Dept. of Statistics, Stanford University.

Politis, D.N. and Romano, J.P. (1993). On the Sample Variance of Linear Statistics Derived from Mixing Sequences, to appear in *Stoch. Proc. Appl.*.

Priestley, M. B. (1981). *Spectral Analysis and Time Series.* Academic Press, New York.

Quenouille, M. (1949). Approximate tests of correlation in time series, *J. Royal Statist. Soc., B*, 11, 68-84.

Raïs, N. and Moore, M. (1990). Bootstrap for some stationary α-mixing processes, Abstract, *INTERFACE '90*, 22nd Symposium on the Interface of Computing Science and Statistics.

Raïs, N. (1992). *Méthodes de reéchantillonage et de sous échantillonage dans le contexte spatial et pour des données dépendantes* , Ph.D. thesis, Department of Mathematics and Statistics, University of Montreal, Montreal, Canada.

Romano, J. (1988). Bootstrapping the mode. *Ann. Inst. Statist. Math.* **40**, 565-586.

Romano, J. (1989). Bootstrap and randomization tests of some nonparametric hypotheses. *Ann. Statist.* **17**, 141-159.

Rosenblatt, M. (1984). Asymptotic normality, strong mixing and spectral density estimates, *Ann. Prob.*, 12, 1167-1180.

Rosenblatt, M. (1985). *Stationary sequences and random fields*, Birkhäuser, Boston.

Roussas, G.G. and Ioannides, D. (1987). Moment Inequalities for Mixing Sequences of Random Variables, *Stoch. Analysis and Applications* , 5(1), p.61-120, Marcel Dekker.

Serfling, R. (1980). *Approximation theorems of mathematical statistics.* John Wiley, New York.

Shao, J. and Wu, J. (1989). A general theory for jackknife variance estimation. *Ann. Statist.* **17**, 1176-1197.

Tu, D. (1992). Approximating the distribution of a general standardized functional statistic with that of jackknife pseudo values. In *Exploring the Limits of Bootstrap.* Edited by LePage, R. and Billard, L. John Wiley, New York, p.279-306.

Wu, J. (1990). On the asymptotic properties of the jackknife histogram. *Ann. Statist.* **18**, 1438-1452.

Wu, J., Carlstein, E. and Cambanis, S. (1989). Bootstrapping the sample mean for data with infinite variance. Preprint.

Zhurbenko, I.G. (1986). *The Spectral Analysis of Time Series*, North-Holland, Amsterdam.