The Annals of Statistics
1994, Vol. 22, No. 4, 2031–2050

LARGE SAMPLE CONFIDENCE REGIONS BASED ON SUBSAMPLES UNDER MINIMAL ASSUMPTIONS[1]

BY DIMITRIS N. POLITIS AND JOSEPH P. ROMANO

Purdue University and Stanford University

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 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. The method is readily adapted to parameters of stationary time series or, more generally, homogeneous random fields. 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 i.i.d. observations, a statistic is computed based on the entire data set and is recomputed over all $\binom{n}{b}$ data sets of size b. These recomputed values of the statistic are suitably normalized to approximate the true sampling distribution. Under certain assumptions on b (which require $b \to \infty$ and $b/n \to 0$), 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 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

Received August 1992; revised February 1994.

[1]Research partially supported by a Purdue Research Foundation Faculty Grant and NSF Grant DMS-89-57217.

AMS 1991 *subject classifications.* Primary 60F99; secondary 62G99.

Key words and phrases. Approximate confidence limit, bootstrap, homogeneous random field, jackknife histogram, stationary, time series.

the weak convergence hypothesis, the sampling distributions based on samples of size b and n should be close.

The method extends to the context of a stationary time series or, more generally, a homogeneous random field. Here, the statistic is computed over subsets of the data that retain the dependence structure of the observations. For example, if $X_1, \ldots, X_n$ represent n observations from a 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 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 survey sampling; 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. Hartigan (1975) has adapted his finite sample results to a more general context of certain classes of estimators which have asymptotic normal distributions.

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 deficiencies of the usual 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 asymptotically linear so that asymptotic normality ensues; see Remark 2.1 for further remarks on Wu's work. A more refined analysis of Wu's method in the case of quantile estimation is presented in Shi (1991). Here, we show how these subsample values can accurately estimate a sampling distribution without any assumptions of asymptotic normality.

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. Thus, 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.* In this section, $X_1, \ldots, X_n$ is a sample of n i.i.d. 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 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 the true sampling distribution of T_n in order to make inferences about θ. 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, with corresponding c.d.f. denoted $J_n(\cdot, P)$. 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 $n \to \infty$.

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. 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

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

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 $J_n(P)$. Replacing $\theta(P)$ by T_n is permissible because $\tau_b(T_n - \theta(P))$ is of order $\tau_b/\tau_n \to 0$.

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

$$\text{Prob}_P\Big\{\tau_n\big[T_n - \theta(P)\big] \leq c_n(1-\alpha)\Big\} \to 1-\alpha, \tag{2.3}$$

and the asymptotic coverage probability under P of the interval $[T_n - \tau_n^{-1} c_n(1-\alpha), \infty)$ is $1-\alpha$.

(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* 1.

PROOF. Let

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

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,

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

so that for every $\varepsilon > 0$, $U_n(x-\varepsilon)1(E_n) \leq L_n(x)1(E_n) \leq U_n(x+\varepsilon)$, where $1(E_n)$ is the indicator of the event $E_n \equiv \{\tau_b|\theta(P) - T_n| \leq \varepsilon\}$. But, the event E_n has probability tending to 1. So, with probability tending to 1, $U_n(x-\varepsilon) \leq L_n(x) \leq U_n(x+\varepsilon)$ for any $\varepsilon > 0$. Hence , if $x+\varepsilon$ and $x-\varepsilon$ are continuity points of $J(\cdot, P)$, then $U_n(x \pm \varepsilon) \to J(x \pm \varepsilon, P)$ in probability implies

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

with probability tending to 1. Now, let $\varepsilon \to 0$ so that $x \pm \varepsilon$ are continuity points of $J(\cdot, P)$. Therefore, it suffices to show $U_n(x) \to J(x,P)$ 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 [see Serfling (1980), Theorem A, page 201]: for any $t > 0$,

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

One can obtain a similar inequality for $t < 0$ by considering the U-statistic $-U_n(x)$. Hence, $U_n(x) - J_b(x,P) \to 0$ in probability. The result (i) follows since $J_b(x,P) \to J(x,P)$. To prove (ii), given any subsequence $\{n_k\}$, one can extract a further subsequence $\{n_{k_j}\}$ so that $L_{n_{k_j}}(x) \to J(x,P)$ almost surely. Hence,

$L_{n_{k_j}}(x) \to J(x,P)$ 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 Pólya'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–Centelli lemma on the inequality (2.4). □

REMARK 2.1. In regular cases, $\tau_n = n^{1/2}$, and the assumptions on b simplify to $b/n \to 0$ and $b \to \infty$. The further assumption on b in part (iv) of the theorem will then hold, for example, if $b = n^\gamma$ for any $\gamma \in (0,1)$. More generally, it holds if $b\log(n)/n \to 0$. The assumptions on b are as weak as possible under the assumptions of the theorem. However, in some cases, the choice $b = O(n)$ also works, though it will not work in general as shown in Example 2.1.2. However, the case $b = O(n)$ does work in the situations analyzed by Wu (1990), where the statistic is approximately linear with an asymptotic Gaussian distribution and $\tau_n = n^{1/2}$. Specifically, Wu assumes the statistic T_n has an expansion

$$T_n(X_1,\ldots,X_n) = \theta(P) + n^{-1}\sum_{i=1}^{n}\phi_P(X_i) + R_n,$$

where $n^{1/2}R_n \to 0$ in probability. Then (2.2) follows even if $b/n \to p$ and $p > 0$; see Theorem 2 of Wu (1990). Wu's result is especially nice in that this form of asymptotic linearity is extremely weak in situations of asymptotic normality, and avoids having to invoke any differentiability hypotheses for example. Our goal here is to obtain consistency without having to assume any further structure on the statistic sequence. Although Example 2.1.2 shows that some structure must be assumed if we wish to assume only $b/n \to p$ with $p > 0$, it does not rule out the possibility that choice of b will not work in situations less extreme than that of Example 2.1.2. It would be interesting to investigate further situations where $b/n \to p$ with $p > 0$ could be used, as it would perhaps lend support to the use of the method in finite samples. Note, however, that in nice situations where the statistic is asymptotically linear, optimal choices of b will actually satisfy $b/n \to 0$ [specifically $b = O(n^{2/3})$], so that a choice of b (such as $b/n \to p$ with $p > 0$) too big should be avoided even in simple situations; see Section 2.4 below, Shao and Wu (1989) and Section 4 of Wu (1990). Finally, it is interesting to mention that in the time series context of Section 3, the choice $b/n \to p$ with $p > 0$ will not work even for linear statistics.

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.

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. Suppose a cost c_1 is incurred for each failed unit which is replaced and a cost $c_2 < c_1$ is incurred for each nonfailed unit which is exchanged. Then 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. Let $r(x) = f(x)/[1 - F(x)]$ be the failure rate of F. 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 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, $T_n = \theta(\hat{F}_n)$. To handle problems of existence or uniqueness, see Arunkumar (1972). Arunkumar (1972) has shown that $n^{1/3}[T_n - \theta(F)]$ has a nondegenerate limiting distribution, so our Assumption 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 toward 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.

EXAMPLE 2.1.2 (Extreme order statistic). Bickel and Freedman (1981) provide the following counterexample to the bootstrap. [For other counterexamples to the bootstrap, see Beran's (1984) analysis of Hodges' superefficient estimator; Bickel and Freedman's (1981) U-statistic, and the correction suggested by Arcones and Giné (1992); Bretagnolle (1983); Babu (1984); Athreya's (1987) analysis of the mean in the infinite variance case; and the variance of a quantile as discussed in Ghosh, Parr, Singh and Babu (1984). Shi (1991) shows how subsampling works for this problem under weaker conditions than the bootstrap requires. In all these examples, our Assumption A holds.] 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 1. In Theorem 2.1, the conditions on b (with $\tau_n = n$) reduce to $b/n \to 0$ and $b \to \infty$. In this example, 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. Note that, in some of the aforementioned cases where the bootstrap is known to fail, it has been realized that a smaller

bootstrap resample size can lead to consistency. In fact, under our Assumption A, this is generally true; see Politis and Romano (1993b).

2.2. *Random subsampling.* 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

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

COROLLARY 2.1. *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* $\widehat{L}_n(x)$.

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

2.3. *General parameters and other choices of root.*

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

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

Under the conditions of Theorem 2.1 and the added assumption that $\widehat{\sigma}_n \to \sigma$, where $\sigma = \sigma(P)$ is a positive constant, K_n will be a consistent estimate of the distribution of $\tau_n[T_n - \theta(P)]/\widehat{\sigma}_n$. The proof is similar to that of Theorem 2.1, so it is omitted.

2.3.2. *General parameter space.* It is often desirable to construct confidence regions for multivariate parameters, or for parameters taking values in 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. $H_n(x, P)$. If Assumption A holds, then $H_n(P)$ converges weakly to $H(P)$, where $H(P)$ is the

distributionof ξ when ξ has distribution $J(P)$. The corresponding c.d.f. $H(P)$ is denoted $H(x,P)$. The approximation to $H_n(x)$ we study is defined analogously to (2.1):

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

THEOREM 2.2. *Assume Assumption* A. *Also assume* $\tau_b/\tau_n \to 0$, $b \to \infty$ *and* $b/n \to 0$ *as* $n \to \infty$. *Let* x *be a continuity point of* $H(\cdot, P)$.

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

(ii) *If* $H(\cdot,P)$ *is continuous, then* $\sup_x |\widehat{H}_n(x) - H_n(x,P)| \to 0$ *in probability.*

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

$$\mathrm{Prob}_P\big\{\tau_n \|T_n - \theta(P)\| \le h_n(1-\alpha)\big\} \to 1-\alpha,$$

and the asymptotic coverage probability of the set $\{\theta \in \Theta\colon \tau_n \|T_n - \theta\| \le h_n(1-\alpha)\}$ *is* $1-\alpha$.

(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* 1.

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 i.i.d. 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. Although distributional theory is quite hard in these models, our method is justified.

2.4. *Second-order asymptotics and choice of b.* The theory developed thus far assumes $b \to \infty$ and $b/n \to 0$. In order to choose b optimally, higher-order considerations are necessary. Consider 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 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 $Z_{n,2} = \tau_n[\theta(P) - T_n]$ is of order τ_b/τ_n in probability. In regular cases, $\tau_n = n^{1/2}$, in which case $Z_{n,2}$ is $O_P(b/n)^{1/2}$. Hence,

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

If $n^{1/2}[T_n - \theta(P)]$ converges weakly to the normal distribution with mean 0 and variance σ^2,

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

by a Taylor expansion argument, using the fact that Φ has mean 0. Thus, in regular cases, $L_n - J_b$ 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. In this case, the error $L_n(t) - J_n(t, P)$ is $O_P(n^{-1/3})$, though a good second-order theory would require the error to be $o_P(n^{-1/2})$.

We would also like to point out that the choice of b depends crucially on the desired goal. Consider the use of 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

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

Hence, in order to accurately estimate the bias of T_n, we should at least require $b/n \to 1$.

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 case 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 them. 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. *Basic definitions.* Let $\mathbf{Z}$ denote the integers and $\mathbf{Z}^+$ the positive integers. Suppose $\{X(\mathbf{t}),\ \mathbf{t} \in \mathbf{Z}^d\}$ is a random field in d dimensions, with $d \in \mathbf{Z}^+$, that is, 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 $\{X(\mathbf{t}),\ \mathbf{t} \in E\}$ is identical to the joint distribution of $\{X(\mathbf{t}),\ \mathbf{t} \in \mathbf{E} + \mathbf{i}\}$. In the case $d = 1$, 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 $d(\mathbf{E}_1, \mathbf{E}_2) = \inf\{d(\mathbf{t}, \mathbf{u})\colon \mathbf{t} \in \mathbf{E}_1,\ \mathbf{u} \in \mathbf{E}_2\}$.

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 $1 \le t_k \le n_k$, 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

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

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 0 at some rate, as k tends to ∞, and l_1, l_2 either remain fixed or tend to ∞ as well. Let $\alpha_X(k) = \alpha_X(k; \infty, \infty)$ be the usual strong mixing coefficient of Rosenblatt (1985); then $\alpha_X(k; l_1, l_2) \le \alpha_X(k)$. 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)], for example, 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) and Bradley (1991)].

3.2. *The general theorem in the case of dependent data.* As in Section 2, let $T_\mathbf{n} = T_\mathbf{n}(X(\mathbf{t}),\ \mathbf{t} \in \mathbf{E_n})$ 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 the following.

ASSUMPTION A1. $J_\mathbf{n}(P)$ 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 $(j_k - 1)h_k + 1 \le i_k \le (j_k - 1)h_k + b_k$, 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 $\mathbf{E_n}$. 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, that is, 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 $\mathbf{i} \ne \mathbf{j}$, 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 $\mathbf{E_n}$ and n fixed, $Y_{\mathbf{j}}$ is defined only for $\mathbf{j}$ such that $1 \le j_k \le q_k$, where $q_k = [(n_k - b_k)/h_k] + 1$, 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

$$L_{\mathbf{n}}(x) = q^{-1} \sum_{i_1=1}^{q_1} \sum_{i_2=1}^{q_2} \cdots \sum_{i_d=1}^{q_d} 1\{\tau_{\mathbf{b}}(S_{\mathbf{n},\mathbf{i}} - T_{\mathbf{n}}) \le x\}. \tag{3.1}$$

THEOREM 3.1. *Assume Assumption* A1 *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* $q^{-1}\sum_{k=1}^{q^*} k^{d-1}\alpha_X(k;b,b) \to 0$, *where* $q^* = \max_i q_i$. *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

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

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 $\mathrm{Var}(U_{\mathbf{n}}(x))$. By the homogeneity of the random field $\{X(\mathbf{t})\}$,

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

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

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

and

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

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

$$q = \prod_{j=1}^{d} \left(\left[\frac{n_j - b_j}{h_j} \right] + 1 \right) \sim \prod_{j=1}^{d} (n_j - b_j)/h_j,$$

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,

$$|A| \leq c_0 q^{-1} \sum_{i \in \mathbf{E}_q - \mathbf{E}^*} \alpha_X(i^* h^* - b^*; b, b)$$

$$\leq c_1 q^{-1} d \sum_{k=[b^*/h^*]+1}^{q^*} W(k) \alpha_X(k h^* - b^*; b, b),$$

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,

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

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

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, that is, if $\mathbf{h} = (1, 1, \ldots, 1)$, the statement of the theorem simplifies and the following corollary is true.

COROLLARY 3.1. *Assume Assumption* A1 *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. As a matter of fact, Theorem 3.1 limited to the time series case is remarkably similar to Theorem 2.1.

COROLLARY 3.2. *Let* $d = 1$. *Assume Assumption* A1 *and that* $\tau_{\mathbf{b}}/\tau_{\mathbf{n}} \to 0$, $b \to \infty$ *and* $b/n \to 0$, *as* $n \to \infty$. *Also let* $1 \leq h \leq c_0 b$, *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* (3.1)].

Theorem 3.1 can also be extended to Studentized roots and general parameter spaces. There is an interesting extension of Theorem 3.1 that should be mentioned. 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 A1 that reads as follows.

ASSUMPTION A1*. $J_{\mathbf{n}}(P)$ converges weakly to a limit law $J(P)$, as $n_i \to \infty$, for $i = 1, \ldots, d^*$, and $n_j \to Q_j$, for $j = d^* + 1, \ldots, d$, where $1 \leq d^* \leq d$, 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 ∞; 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^* is the "effective" dimension, and the setup seems equivalent to a vector-valued random field in d^* 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 setup; its proof is analogous to the proof of Theorem 3.1.

COROLLARY 3.3. *Assume Assumption* A1* *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* $i = 1, 2, \ldots, d^*$, *whereas* $b_j \to Q_j$ *and* $n_j \to Q_j$, *for* $j = d^* + 1, \ldots, d$. *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* $n^* = \max_{i=1,\ldots,d^*} n_i$. *Let* x *be a continuity point of* $J(\cdot, P)$. *Then conclusions* (i)–(iii) *of Theorem* 2.1 *remain true* (*with* n *replaced by* $\mathbf{n}$).

3.3. *Variance estimation and bias reduction.*

3.3.1. *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.

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 $b \sim An^{d/(d+2)}$ [cf. Politis and Romano (1992b, a)]; 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 the 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 $E_{\mathbf{i}, \mathbf{b}, \mathbf{h}}$, for $\mathbf{i}$ such that $1 \leq i_k \leq q_k$, $k = 1, \ldots, d$, which in turn (for given $\mathbf{b}$ and $\mathbf{n}$) maximizes q, the number of subsamples available for 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 $m_{\mathbf{n}}^{(n)}$ 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 the sample mean. Then the variance estimator $m_{\mathbf{n}}^{(2)}$ is asymptotically equivalent to a kernel smoothed (with Bartlett's kernel) estimator of the spectral density at the origin [Künsch (1989)]. It is well known [cf. Priestley (1981)] that the bias of $m_{\mathbf{n}}^{(2)}$ is of order $O(1/b)$, and the variance of $m_{\mathbf{n}}^{(2)}$ is of order $O(b/n)$; this of course implies that consistent variance estimation requires $b \to \infty$ as well as $b/n \to 0$.

3.3.2. *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 $\mathrm{Bias}(T_{\mathbf{n}}) = ET_{\mathbf{n}} - \theta$ by a rescaled version of the "empirical" bias, that is, by

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

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

$$\widehat{T}_{\mathbf{n}} = T_{\mathbf{n}} - \widehat{\text{Bias}}(T_{\mathbf{n}}) = \left(1 + \frac{\tau_{\mathbf{b}}}{\tau_{\mathbf{n}}}\right)T_{\mathbf{n}} - \frac{\tau_{\mathbf{b}}}{\tau_{\mathbf{n}}}\text{Ave}(S_{\mathbf{n},\mathbf{i}}). \tag{3.2}$$

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), $\widehat{\text{Bias}}(T_{\mathbf{n}}) \not\equiv 0$, due to edge effects; nevertheless $\widehat{\text{Bias}}(T_{\mathbf{n}}) \to 0$ 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 (3.2) 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 3.5.2).

THEOREM 3.2. *Assume Assumption* A1 *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* $E|\widetilde{S}_{\mathbf{n},1}|^{2+\delta} < C$ *and that* $q^{-1}\sum_{k=1}^{q^*} k^{d-1}\{\alpha_X(k; b, b)\}^{\delta/(2+\delta)} \to 0$, *where* δ *and* C *are two positive constants independent of* $\mathbf{n}$, $\widetilde{S}_{\mathbf{n},1} \equiv S_{\mathbf{n},1}/\sqrt{\text{Var}(S_{\mathbf{n},1})}$ *and* $q^* = \max_i q_i$. *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

$$\begin{aligned}\text{Var}(m_{\mathbf{n}}^{(1)}) &= \text{Var}\Big(\tau_{\mathbf{b}}\big(\text{Ave}(S_{\mathbf{n},\mathbf{i}}) - T_{\mathbf{n}}\big)\Big)\\ &= \text{Var}\Big(\tau_{\mathbf{b}}\big(\text{Ave}(S_{\mathbf{n},1}) - ES_{\mathbf{n},\mathbf{i}}\big) - \tau_{\mathbf{b}}(T_{\mathbf{n}} - ES_{\mathbf{n},1})\Big)\\ &= \text{Var}\Big(\tau_{\mathbf{b}}\big(\text{Ave}(S_{\mathbf{n},\mathbf{i}}) - ES_{\mathbf{n},1}\big)\Big) + o(1),\end{aligned}$$

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

$$\begin{aligned}\text{Var}\Big(\tau_{\mathbf{b}}\big(\text{Ave}(S_{\mathbf{n},1}) - ES_{\mathbf{n},\mathbf{i}}\big)\Big) = \tau_{\mathbf{b}}^2 q^{-1}\sum_{\mathbf{i}\in\mathbf{E}_q}&\left(1 - \frac{|i_1|}{q_1}\right)\left(1 - \frac{|i_2|}{q_2}\right)\cdots\left(1 - \frac{|i_d|}{q_d}\right)\\ &\times \text{Cov}(S_{\mathbf{n},1}, S_{\mathbf{n},1+\mathbf{i}})\end{aligned}$$

and thus

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

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

$$\left|\mathrm{Cov}(\widetilde{S}_{\mathbf{n},\mathbf{1}}, \widetilde{S}_{\mathbf{n},\mathbf{1}+\mathbf{i}})\right| \leq 10C^2\{\alpha_X(i^*h^* - b^*; b, b)\}^{\delta/(2+\delta)}$$

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

3.4. *Comparison with other resampling methods.* Subsampling methodology for dependent data has been used in the past for variance estimation [Carlstein (1986), Raïs (1992) and Politis and Romano (1992a, b, 1993a)] 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) and Politis and Romano (1992a)]. In the case of a stationary strong mixing sequence ($d = 1$), Carlstein (1986) used $m_n^{(2)}$ (but only in the case $h = b$). Carlstein's idea was generalized in Politis and Romano (1993a) 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 $m_{\mathbf{n}}^{(2)}$ was also generalized to the case of homogeneous random fields ($d > 1$) by Possolo (1991), Raïs (1992) and Politis and Romano (1992b).

The fact that taking $h = 1$ is preferable 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) and Liu and Singh (1992)] is *identical* to $m_n^{(2)}$ with $h = 1$. Let $\widehat{J}_n(x, P)$ denote the "moving blocks" bootstrap estimate of $J_n(x, P)$ [cf. Künsch (1989) and Liu and Singh (1992)]; as can be calculated, the variance of $\widehat{J}_n(x, P)$ is approximately (up to an asymptotically negligible factor) equal to $m_n^{(2)}$ with $h = 1$, and indeed $\widehat{J}_n(x, P)$ is very closely related to the "empirical" $L_n(x)$.

The above discussion helps put the subsampling methodology into perspective. 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.

3.5. *Some examples.* The examples will address some unorthodox cases; in all standard cases of statistics from time series and random fields that possess asymptotic distributions, for example, the sample mean, the sample autocovariances and autocorrelations, estimates of the spectral and cross-spectral density,

estimates of the coherency function and so on, the subsampling methodology is applicable. To verify Assumption A1 in the case of the mean of a series with long-range dependence, see Rosenblatt (1984). 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 $\{X_t,\ t \in \mathbf{Z}\}$.

3.5.1. *Robust statistics from time series.* Suppose the first marginal of the sequence $\{X_t\}$, that is, 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) and Martin and Yohai (1986)]. Under regularity conditions, the median, the trimmed mean, the Hodges–Lehmann estimator, linear combinations or order statistics and so on, 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 $R(k) = \mathrm{Cov}(X_1, X_{1+k})$. Then [cf. Gastwirth and Rubin (1975)] the Hodges–Lehmann estimator, that is, 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 Σ arcsin $(R(k)/2)$ should be consistently estimated which is a difficult task. To appreciate the difficulty, recall that 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.

3.5.2. *The spectral density function.* As before, assume that $\sum|R(k)| < \infty$ and define the spectral density function f by $f(w) = (1/2\pi)\sum_{k=-\infty}^{\infty} R(k)e^{-ikw}$. Fix a point $w \in [-\pi, \pi]$ and consider a kernel smoothed estimator of $f(w)$ given by $\widehat{f}(w) = (1/2\pi)\sum_{k=-n}^{n} B_n(k)\widehat{R}(k)e^{-ikw}$, where $\widehat{R}(k) = (1/n)\sum_{t=1}^{n-k} X_t X_{t+k}$ is the usual sample autocovariance and $\widehat{B}_n(k)$ 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 $\tau_n(\widehat{f}(w) - f(w))$ has an asymptotic normal distribution. In fact, uniform confidence bands for the spectral density or the spectral distribution function may be constructed by the subsampling methodology, but requires a slight extension of Theorem 3.1 to the case of functional parameters (as in Section 2.3 in the i.i.d. case). One must appeal to the weak convergence of the spectral density process, as done in Woodroofe and van Ness (1967); the required extension will appear elsewhere. 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 an analogous manner and are shown to be asymptotically normally distributed under regularity conditions [cf. Rosenblatt (1985)]. Theorem 3.2 is

especially applicable in this context and yields bias-corrected estimates with desirable properties.

3.5.3. *Nonparametric estimation of the first marginal distribution.* Let $F(\cdot)$ denote the distribution of the random variable X_1 and let $\widehat{F}(x) = n^{-1} \sum_{i=1}^{n} 1\{X_i \leq x\}$. Under regularity conditions [cf. Györfi, Härdle, Sarda and Vieu (1989)], $\sqrt{n}(\widehat{F}(x) - F(x))$ possesses a limiting normal distribution, and hence Assumption A is satisfied. Furthermore, $\sqrt{n}(\widehat{F}(\cdot) - F(\cdot))$, viewed as a random function, converges weakly to a Gaussian process [cf. Deo (1973)]. Looking at the sup-norm $\sup_x |\sqrt{n}(\widehat{F}(x) - F(x))|$, 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 2.3. The moving blocks bootstrap can also handle this problem, though further assumptions are needed for consistency; see Naik-Nimbalkar and Rajarshi (1994) and Buhlmann (1992).

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 briefly discussed in Section 2.4. 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 2.4, 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 (1992) 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. and GINÉ, E. (1992). On the bootstrap of U and V statistics. *Ann. Statist.* **20** 655–674.

ARUNKUMAR, S. (1972). Nonparametric age replacement policy. *Sankhyā Ser. 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. *Sankhyā Ser. A* **46** 86–93.

BERAN, R. (1984). Bootstrap methods in statistics. *Jahresber. Deutsch. 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. Probab.* **10** 1047–1050.

BRADLEY, R. C. (1991). Equivalent mixing conditions for random fields. Technical Report 336, Center for Stochastic Processes, Dept. Statistics, Univ. North Carolina, Chapel Hill.

BRADLEY, R. C. (1992). On the spectral density and the asymptotic normality of weakly dependent random fields. *J. Theoret. Probab.* **5** 355–373.

BRETAGNOLLE, J. (1983). Limites du bootstrap de certaines fonctionnelles. *Ann. Inst. H. Poincaré Probab. Statist.* **3** 281–296.

BUHLMANN, P. (1992). Weak convergence of the bootstrapped multidimensional empirical process for stationary strong-mixing sequences. Research Report 68, Seminar für Statistik, ETH, CH-8092 Zurich.

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. Probab.* **1** 870–875.

DOBRUSHIN, R. L. (1968). The description of a random field by means of conditional probabilities and conditions of its regularity. *Theory Probab. 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*. SIAM, Philadelphia.

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

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

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 Statist.* **60**. Springer, New York.

HARTIGAN, J. (1969). Using subsample values as typical values. *J. Amer. Statist. Assoc.* **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, New York.

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, Dépt. d'informatique et de recherche opérationnelle, Univ. Montréal.

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.

MARTIN, R. D. and YOHAI, V. J. (1986). Influence functionals for time series. *Ann. Statist.* **14** 781–818.

MCCARTHY, P. (1969). Pseudo-replication: half-samples. *Internat. Statist. Rev.* **37** 239–263.

NAIK-NIMBALKAR, U. V. and RAJARSHI, M. B. (1994). Validity of block-wise bootstrap for empirical processes with stationary observations. *Ann. Statist.* **22** 980–994.

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

POLITIS, D. and ROMANO, J. (1991). The stationary bootstrap. Technical Report 365, Dept. Statistics, Stanford Univ.

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.* **20** 1985–2007.

POLITIS, D. N. and ROMANO, J. P. (1992b). Nonparametric resampling for homogeneous strong mixing random fields. Technical Report 396, Dept. Statistics, Stanford Univ.

POLITIS, D. N. and ROMANO, J. P. (1993a). On the sample variance of linear statistics derived from mixing sequences. *Stochastic Process. Appl.* **45** 155–167.

POLITIS, D. N. and ROMANO, J. P. (1993b). Estimating the distribution of a studentized statistic by subsampling. *Bulletin of the International Statistical Institute* **2** 315–316.

POSSOLO, A. (1991). Subsampling a random field. In *Spatial Statistics and Imaging* (A. Possolo, ed.) 286–294. IMS, Hayward, CA.

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

QUENOUILLE, M. (1949). Approximate tests of correlation in time series. *J. Roy. Statist. Soc. Ser. B* **11** 68–84.

RAÏS, N. (1992). Méthodes de reéchantillonage et de sous échantillonage dans le contexte spatial et pour des données dépendantes. Ph.D. thesis, Dép. de mathématique et de statistique, Univ. Montreal.

RAÏS, N. and MOORE, M. (1990). Bootstrap for some stationary α-mixing processes (abstract). In *Computing Science and Statistics: Proceedings of the 22nd Symposium on the Interface* (C. Page and R. Le Page, eds.). Springer, New York.

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. Probab.* **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. *Stochastic Anal. Appl.* **5** 61–120.

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

SHAO, J. and WU, C. F. J. (1989). A general theory for jackknife variance estimation. *Ann. Statist.* **17** 1176–1197.

SHI, X. (1991). Some asymptotic results for jackknifing the sample quantile. *Ann. Statist.* **19** 496–503.

TU, D. (1992). Approximating the distribution of a general standardized functional statistic with that of jackknife pseudo values. In *Exploring the Limits of Bootstrap* (R. LePage and L. Billard, eds.) 279–306. Wiley, New York.

WOODROOFE, M. and VAN NESS, J. (1967). The maximum deviation of sample spectral densities. *Ann. Math. Statist.* **38** 1558–1570.

WU, C. F. J. (1990). On the asymptotic properties of the jackknife histogram. *Ann. Statist.* **18** 1438–1452.

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

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305